Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	Well-intentioned initiatives hinder understandingbiodiversity conservation: an essay on a recent deep-learning image classifier for Amazonian fishes	
Article Sub-Title		
Article CopyRight		
Journal Name	Reviews in Fish Biology and Fisheries	

Corresponding Author	FamilyName	**Ottoni**
	Particle	
	Given Name	**Felipe Polivanov**
	Suffix	
	Division	Programa de Pós-Graduação em Biodiversidade e Biotecnologia da Amazônia Legal – Rede BIONORTE
	Organization	Universidade Federal do Maranhão
	Address	São Luís, MA, 65085-580, Brazil
	Division	Laboratório de Sistemática E Ecologia de Organismos Aquáticos, Centro de Ciências de Chapadinha
	Organization	Universidade Federal do Maranhão
	Address	Chapadinha, MA, 65500-000, Brazil
	Division	Programa de Pós-Graduação em Biodiversidade e Conservação
	Organization	Universidade Federal do Maranhão
	Address	São Luís, MA, 65085-580, Brazil
	Division	Programa de Pós-Graduação em Ciências Ambientais
	Organization	Universidade Federal do Maranhão
	Address	BR 222, km 4, Boa Vista, Chapadinha, MA, 65500-000, Brazil
	Phone	
	Fax	
	Email	fpottoni@gmail.com
	URL	
	ORCID	http://orcid.org/0000-0002-9390-0918
Author	FamilyName	**Campos**
	Particle	
	Given Name	**Diego Sousa**
	Suffix	
	Division	Programa de Pós-Graduação em Biodiversidade e Biotecnologia da Amazônia Legal – Rede BIONORTE
	Organization	Universidade Federal do Maranhão
	Address	São Luís, MA, 65085-580, Brazil
	Division	Laboratório de Sistemática E Ecologia de Organismos Aquáticos, Centro de Ciências de Chapadinha
	Organization	Universidade Federal do Maranhão
	Address	Chapadinha, MA, 65500-000, Brazil
	Phone	
	Fax	
	Email	camposds1@yahoo.com.br
	URL	
	ORCID	http://orcid.org/0000-0001-8262-413X
Author	FamilyName	**Oliveira**
	Particle	de
	Given Name	**Rafael Ferreira**
	Suffix	
	Division	Laboratório de Sistemática E Ecologia de Organismos Aquáticos, Centro de Ciências de Chapadinha
	Organization	Universidade Federal do Maranhão
	Address	Chapadinha, MA, 65500-000, Brazil
	Division	Programa de Pós-Graduação em Biodiversidade e Conservação
	Organization	Universidade Federal do Maranhão
	Address	São Luís, MA, 65085-580, Brazil
	Phone	
	Fax	
	Email	raphaelmais12@gmail.com
	URL	
	ORCID	http://orcid.org/0000-0002-1659-2923

Author	FamilyName	**Oliveira Vieira**
	Particle	de
	Given Name	**Lucas**
	Suffix	
	Division	Programa de Pós-Graduação em Biodiversidade e Biotecnologia da Amazônia Legal – Rede BIONORTE
	Organization	Universidade Federal do Maranhão
	Address	São Luís, MA, 65085-580, Brazil
	Division	Laboratório de Sistemática E Ecologia de Organismos Aquáticos, Centro de Ciências de Chapadinha
	Organization	Universidade Federal do Maranhão
	Address	Chapadinha, MA, 65500-000, Brazil
	Phone	
	Fax	
	Email	lucasolivier070@gmail.com
	URL	
	ORCID	http://orcid.org/0000-0002-1448-0645
Author	FamilyName	**Bragança**
	Particle	de
	Given Name	**Pedro Henrique Negreiros**
	Suffix	
	Division	Programa de Pós-Graduação em Biodiversidade e Conservação
	Organization	Universidade Federal do Maranhão
	Address	São Luís, MA, 65085-580, Brazil
	Division	Freshwater Taxonomy Group
	Organization	South African Institute for Aquatic Biodiversity
	Address	Makhanda, 6140, Eastern Cape, South Africa
	Phone	
	Fax	
	Email	pedrobra88@gmail.com
	URL	
	ORCID	http://orcid.org/0000-0002-8357-7010
Author	FamilyName	**Guimarães**
	Particle	
	Given Name	**Erick Cristofore**
	Suffix	
	Division	Programa de Pós-Graduação em Sociedade, Natureza e Desenvolvimento, Instituto de Ciências da Educação,
	Organization	Universidade Federal do Oeste Do Pará
	Address	Santarém, PA, 68040-070, Brazil
	Phone	
	Fax	
	Email	erick.ictio@yahoo.com.br
	URL	
	ORCID	http://orcid.org/0000-0003-4480-5452
Author	FamilyName	**Katz**
	Particle	
	Given Name	**Axel Makay**
	Suffix	
	Division	Programa de Pós-Graduação em Biodiversidade e Biologia Evolutiva, Instituto de Biologia
	Organization	Universidade Federal do Rio de Janeiro
	Address	Rio de Janeiro, RJ, 21941-617, Brazil
	Phone	
	Fax	
	Email	axelmk@gmail.com
	URL	
	ORCID	http://orcid.org/0000-0002-2933-7163

Author	FamilyName	**Henschel**
	Particle	
	Given Name	**Elisabeth**
	Suffix	
	Division	Programa de Pós-Graduação em Genética, Instituto de Biologia
	Organization	Universidade Federal do Rio de Janeiro
	Address	Rio de Janeiro, RJ, 21941-617, Brazil
	Phone	
	Fax	
	Email	elisabeth.henschel@hotmail.com
	URL	
	ORCID	http://orcid.org/0000-0001-5507-7396
Author	FamilyName	**Brito**
	Particle	de
	Given Name	**Pâmella Silva**
	Suffix	
	Division	Programa de Pós-Graduação em Ciências Ambientais
	Organization	Universidade Federal do Maranhão
	Address	BR 222, km4, Boa Vista, Chapadinha, MA, 65500-000, Brazil
	Phone	
	Fax	
	Email	pamellabrito@hotmail.com
	URL	
	ORCID	http://orcid.org/0000-0003-3945-4512
Author	FamilyName	**South**
	Particle	
	Given Name	**Josie**
	Suffix	
	Division	School of Biology, Faculty of Biological Sciences
	Organization	University of Leeds
	Address	Leeds, LS2 9JT, UK
	Phone	
	Fax	
	Email	josiesouth93@gmail.com
	URL	
	ORCID	http://orcid.org/0000-0002-6339-4225

Schedule	Received	3 Jan 2024
	Revised	
	Accepted	4 Oct 2024

Abstract

The identification of fish species by non-specialists is a constant challenge for biodiversity management. In this regard, Robillard et al. developed a machine learning computer vision model to identify Amazonian fish at the genus level, with an accuracy of 97.9%. Their model aimed to make it easier for non-specialists to identify fish, allowing them to contribute to the collection and sharingof data for biodiversity management. However, when tested with a different set of fish pictures, the Classifier was unable to accurately identify fish photographs, resulting in 82% misidentification, and did not outperform what would be expected by chance, indicating that it is not suitable for the accurate identification of taxa in its current form. The results underscore the need for a balanced approach, combiningautomated tools with expert taxonomic input for accurate conservation decisions, emphasizingcaution in relying solely on Artificial Intelligence methods. While acknowledging the potential of the model, we recommend restricting its application primarily to larger fish of commercial interest or scenarios where conservation decisions are less directly affected by the model's identifications.

Keywords (separated by '-') Amazon River basin - Automated classification - Convolutional neural networks - Neotropical ichthyology - Taxonomy

Footnote Information

Rev Fish Biol Fisheries
https://doi.org/10.1007/s11160-024-09901-y

POINT-OF-VIEW

Well-intentioned initiatives hinder understanding biodiversity conservation: an essay on a recent deep-learning image classifier for Amazonian fishes

Diego Sousa Campos · Rafael Ferreira de Oliveira · Lucas de Oliveira Vieira · Pedro Henrique Negreiros de Bragança · Erick Cristofore Guimarães · Axel Makay Katz · Elisabeth Henschel · Pâmella Silva de Brito · Josie South · Felipe Polivanov Ottoni

Received: 3 January 2024 / Accepted: 4 October 2024

Abstract The identification of fish species by non-specialists is a constant challenge for biodiversity management. In this regard, Robillard et al. developed a machine learning computer vision model to identify Amazonian fish at the genus level, with an accuracy of 97.9%. Their model aimed to make it easier for non-specialists to identify fish, allowing them to contribute to the collection and sharing of data for biodiversity management. However, when tested with a different set of fish pictures, the Classifier was unable to accurately identify fish photographs, resulting in 82% misidentification, and did not outperform what would be expected by chance, indicating that it is not suitable for the accurate identification of taxa in its current form. The results underscore the need for a balanced approach, combining automated tools with expert taxonomic input for accurate conservation decisions, emphasizing caution in relying solely on

D. S. Campos · L. de Oliveira Vieira · F. P. Ottoni (✉)
Programa de Pós-Graduação em Biodiversidade e Biotecnologia da Amazônia Legal – Rede BIONORTE, Universidade Federal do Maranhão, São Luís, MA 65085-580, Brazil
e-mail: fpottoni@gmail.com

D. S. Campos
e-mail: camposds1@yahoo.com.br

L. de Oliveira Vieira
e-mail: lucasolivier070@gmail.com

D. S. Campos · R. F. de Oliveira · L. de Oliveira Vieira · F. P. Ottoni
Laboratório de Sistemática E Ecologia de Organismos Aquáticos, Centro de Ciências de Chapadinha, Universidade Federal do Maranhão, Chapadinha, MA 65500-000, Brazil
e-mail: raphaelmais12@gmail.com

R. F. de Oliveira · P. H. N. de Bragança · F. P. Ottoni
Programa de Pós-Graduação em Biodiversidade e Conservação, Universidade Federal do Maranhão, São Luís, MA 65085-580, Brazil
e-mail: pedrobra88@gmail.com

P. H. N. de Bragança
Freshwater Taxonomy Group, South African Institute for Aquatic Biodiversity, Makhanda 6140, Eastern Cape, South Africa

E. C. Guimarães
Programa de Pós-Graduação em Sociedade, Natureza e Desenvolvimento, Instituto de Ciências da Educação,, Universidade Federal do Oeste Do Pará, Santarém, PA 68040-070, Brazil
e-mail: erick.ictio@yahoo.com.br

A. M. Katz
Programa de Pós-Graduação em Biodiversidade e Biologia Evolutiva, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21941-617, Brazil
e-mail: axelmk@gmail.com

E. Henschel
Programa de Pós-Graduação em Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21941-617, Brazil
e-mail: elisabeth.henschel@hotmail.com

Springer

Journal : **Medium 11160**	Article No : **9901**	Pages : **14**	MS Code : **9901**	Dispatch : **9-10-2024**

Artificial Intelligence methods. While acknowledging the potential of the model, we recommend restricting its application primarily to larger fish of commercial interest or scenarios where conservation decisions are AQ 1 less directly affected by the model's identifications.

Keywords Amazon River basin · Automated classification · Convolutional neural networks · Neotropical ichthyology · Taxonomy

Introduction

Performing accurate taxonomical assessments of freshwater fish biodiversity is a persistent challenge for conservation scientists and practitioners alike, especially in megadiverse regions such as the Amazon Basin (Olden et al. 2010; Silvano et al. 2022). Identification relies on traditional methods of collecting and identifying freshwater fish (i.e., regional inventories), which tend to be time-consuming and expensive and require high levels of training (Robillard et al. 2023). Molecular methods, such as DNA-barcoding and eDNA, have increased knowledge and allowed rapid species inventories, however, both methods rely on the availability of voucher-based reference libraries including accurately identified species (Zainal-Abidin et al. 2022). Additionally, these methods require technology and sample processing infrastructure, which are deficient in many institutions in the global south, especially in many Amazonian institutions (Robillard et al. 2023). Further, but not less important, is the little participation of non-specialists, such as fishermen, the general population, and citizen scientists in the role of documenting biodiversity. Enabling these agents to participate in collecting and sharing data would increase the likelihood of creating policies and managing decision-making (e.g., conservation measures) that better represent the stakeholders (Robillard et al. 2023).

In a recent publication, Robillard et al. (2023) present computer vision models designed for Amazonian fish identification based solely on photographs. These machine learning models, utilizing U-Net for image segmentation and a convolutional neural network (CNN) for classification at the genus level, offer a practical and reliable alternative for simplifying fish identification. The authors advocate for a cost-effective and efficient approach to species assessments, eliminating the need for specialist validation or expensive molecular barcode techniques. The models aim to seamlessly integrate data from non-specialists, addressing current barriers in fish identification. In their methodology, the authors utilized a database of 3068 photographs representing 33 fish genera from 18 families and 4 orders, collected in Loreto, Peru, between 2018 and 2019. Impressively, the study achieved a genus-level identification accuracy of 97.9%. Notably, misidentifications were predominantly linked to small tetras (Characiformes: Characidae), key components of the Amazonian ichthyofauna (Oliveira et al. 2009; Van Der Sleen and Albert 2018).

The authors assert that their openly accessible online application, the Fish Masker and Classifier (available at https://amazonian-fish-classifier.streamlit.app-permalink: https://archive.ph/OYq5a), serves as a valuable tool for non-specialists in achieving genus-level identification. The application allows users to upload pictures of live or preserved specimens under various conditions. According to the authors, the application recognizes fish pixels in the image, masks non-fish elements, and provides a taxonomic identification at the order, family, and genus levels based on their trained model. However, it is crucial to note that the performance of this machine-learning method for genus-level identification has not undergone further validation through additional tests. AQ 2

The focus on real-world applicability and the potential implications of this type of approach contribute to the ongoing discourse on the application of artificial intelligence in biodiversity research, emphasizing the crucial intersection of technological innovation and traditional taxonomic expertise in conservation decision-making (Campos et al. 2023). Therefore, this study aims to assess and validate the performance of the method proposed by Robillard

P. S. de Brito · F. P. Ottoni
Programa de Pós-Graduação em Ciências Ambientais, Universidade Federal do Maranhão, BR 222, km 4, Boa Vista, Chapadinha, MA 65500-000, Brazil
e-mail: pamellabrito@hotmail.com

J. South
School of Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK
e-mail: josiesouth93@gmail.com

et al. (2023) as an example of an innovative tool for fish identification. By employing a comprehensive approach, we seek to make a substantial contribution to the field, addressing the inherent challenges of fish identification, particularly concerning biodiversity management in tropical ecosystems. We aim to provide novel insights and discuss critical aspects of the taxonomic accuracy of the model. This study is positioned to offer valuable perspectives for both scientists and practitioners engaged in environmental conservation, emphasizing the relevance of accurate fish identification in the context of megadiverse tropical ecosystems, such as the Amazon region.

Methods

Evaluating the training dataset for the 'Amazonian Fish Masker and Classifier'

To assess the quality of the "Images used to train Amazonian fish classification model" (Dikow 2023) used in Robillard et al. (2023), we analyzed the original masked images provided. We created a custom macro function within ImageJ software (Schneider et al. 2012) for precise pixel counting. The masked images were converted to 8-bit format, to standardize the pixel values to a range between 0 and 255, where 0=black color. A threshold was applied, and any pixel with a value = zero was filtered, leaving only pixels with color information. Finally, we performed a particle analysis, counting the number of pixel aggregates in the images (code is available in Supplementary file S1). Following the pixel counting, ImageJ generated two distinct sheets as a result: one presenting values in a row for each particle – available in Supplementary file S2.1, and a second sheet with condensed values summarized by each analyzed picture (n = 3068 images) - in Supplementary file S2.2.

Testing the 'Fish Masker and Classifier' tool

To test the model provided by Robillard et al. (2023), we used 100 photographs representing 21 genera, which were also included in their model training, with specimens from river basins under Amazonian influence (Guamá, Gurupi, Turiaçu, Mearim, Munim, Preguiças, and Parnaíba river basins), as well as from the Beni and Mamoré river drainages, in the Amazon River basin. The photographed specimens and their corresponding vouchers are deposited in the following ichthyological collections: CICCAA (Coleção Ictiológica do Centro de Ciências Agrárias e Ambientais - Universidade Federal do Maranhão, Chapadinha, Brazil) and UMSS (Museo d'Orbigny - Universidad Mayor de San Simón, Cochabamba, Bolivia). The photographic database includes photographs taken under different conditions, such as color-in-life pictures taken in a photo tank and outside a tank (e.g., in the hand), pictures of preserved fish over a manually masked black background, and a white background. (Table 1 in Supplementary file S3). Photographs of *Bujurquina* spp. from the Mamoré and Beni rivers were obtained from Careaga et al. (2023), permitted by the authors.

The images were submitted to the web application 'Fish Masker and Classifier', a product developed by Robillard et al. (2023) (Fish Masker and Classifier- available at https://amazonian-fish-classifier.streamlit.app). After uploading, the application utilizes the masker model to determine the percentage of pixels classified as *fish* and to mask out the remaining pixels, rendered in black. Subsequently, we gathered this value (henceforth referred to as 'fish_pixels' in the text) and preserved both the masked image and the classifier model-generated prediction bar graph.

Since the graph lacks printed values for individual bars, we utilized the Plot Digitizer tool (accessible at https://plotdigitizer.com/app) to digitize the charts. The scale was set from 0 to 100 probability, and points on the periphery of the bars in the graph were digitized.

The resulting graph presents four possible genera, each representing the probable genus of the photographed specimen, along with the corresponding probability of matching the classifier-based identification-in simpler terms, it provides a list of genera that the picture is most likely to represent. These probabilities are organized in descending order, with the top-ranked option referred to as the 'first option' and denoted as 'Class_1' in our dataset, and so forth for the subsequent options. While the sum of the four probabilities may not necessarily equal 100; however, it will never exceed this value. Therefore, the probabilities of the four identifications are considered variables with some degree of interdependence.

The dataset with the results of the simulation consisted of the labels of our pictures uploaded to the

Table 1 Results of the simulation with fish pictures in the Fish Masker and Classifier application

Genera suggestion	Count of suggestions	%	Cumulative %	Accuracy
Gymnotus	43	11,14	11,14	100%
Ancistrus	36	9,33	20,47	75%
Bunocephalus	28	7,25	27,72	–
Rineloricaria	27	6,99	34,72	–
Moenkhausia	24	6,22	40,93	40%
Otocinclus	23	5,96	46,89	–
Tetragonopterus	17	4,40	51,30	–
Bryconops	16	4,15	55,44	14%
Prochilodus	16	4,15	59,59	40%
Hyphessobrycon	14	3,63	63,21	40%
Corydoras	13	3,37	66,58	–
Erythrinus	13	3,37	69,95	–
Tatia	13	3,37	73,32	–
Astyanax	10	2,59	75,91	–
Phenacogaster	10	2,59	78,50	–
Bujurquina	9	2,33	80,83	–
Doras	9	2,33	83,16	–
Pygocentrus	9	2,33	85,49	–
Characidium	8	2,07	87,56	38%
Copella	8	2,07	89,64	–
Hemigrammus	6	1,55	91,19	25%
Oxyropsis	5	1,30	92,49	–
Pimelodella	5	1,30	93,78	–
Sorubim	5	1,30	95,08	–
Bario	4	1,04	96,11	–
Pyrrhulina	4	1,04	97,15	–
Charax	3	0,78	97,93	–
Apistogramma	2	0,52	98,45	–
Curimata	2	0,52	98,96	–
Gasteropelecus	2	0,52	99,48	–
Knodus	1	0,26	99,74	–
Tyttocharax	1	0,26	100,00	–

The genera in the first column are those suggested by the model for all options ('*Class_1*', '*Class_2*', '*Class_3*', and '*Class_4*'). The percentage is calculated by counting against the total of genera (32). Accuracy is the percentage of correct classifications for each genus. Empty cells in the accuracy column are zeros

web application, taxonomic information of Order, Family, Genus, and the genus suggested as a result also the classification status, where we verified if the identification was correct and, in the case of error, three different categories ('Order', 'Family', and 'Genus') were assigned to indicate at which taxonomical level the error was identified (Supplementary file S4).

2.2.1 The black-screen test

The amount of information available to the Classifier model in the learning phase is expected to influence the outcome of the classification. Therefore, to evaluate the response of the Classifier under controlled conditions, we performed the '*black-screen test*', which consisted of uploading the image of an all-black color (RGB = 0,0,0) rectangle to the web application and running the fish masker and classifier, collecting the outputs.

Data analysis

To determine whether the Classifier was able to correctly identify the genus (that is, '*Class_1_prob*'), we used a beta regression, via the '*betareg*' R package (Cribari-Neto and Zeileis 2010), whereupon '*Class_1_prob*' was the independent variable and '*fish_pixels*' as the predictor variable. This allows us to assess how variations in pixel composition relate to the probability of correct genus identification. We created a concordance matrix that compares the genus of the specimen depicted in each photo (previously identified by specialists) with the genus suggested by the Classifier as the primary possibility, referred to as '*Class_1*'. To evaluate the agreement in identifying fish genera using the classification model from the Robillard et al. (2023) web application, we calculated Fleiss' Kappa (Fleiss 1971). The analysis was carried out in R, using version 0.84.1. of the '*irr*' package (Gamer et al. 2019).

The datasets for the quality assessment of the training images (Supplementary file S2) and from the classification simulation (Supplementary file S4) were analyzed in Tableau Desktop Professional 2023 .2 (under Freemium Student License), to calculate the descriptive statistics and generate the plots for the masked area (%) and distribution of classification of the Classifier, including the respective probability

error by Genus, Family, and Order.

Springer

Journal : **Medium 11160**	Article No : **9901**	Pages : **14**	MS Code : **9901**	Dispatch : **9-10-2024**

Results

Quality of training dataset

The distribution of masked area percentage values in the input images of the training dataset displayed a large amount of variation concerning image quality (Dikow 2023). We believe that this discrepancy may have directly influenced the outcome in identifications by their model. The overall average of the masked area for all pictures in the "Images used to train Amazonian fish classification model" dataset (Dikow 2023) was 93.99% (Fig. 1). For certain genera (22 of 33), specific averages surpassed the overall average, which can be interpreted as a signal that the majority of the training was done with a relatively low amount of information (Fig. 1). In particular, *Tyttocharax* Fowler 1913, was the genus with the most masked area average, reaching 99.7% which means that almost the entire pictures for this genus in the training for the Classifier were composed of black pixels (non-fish).

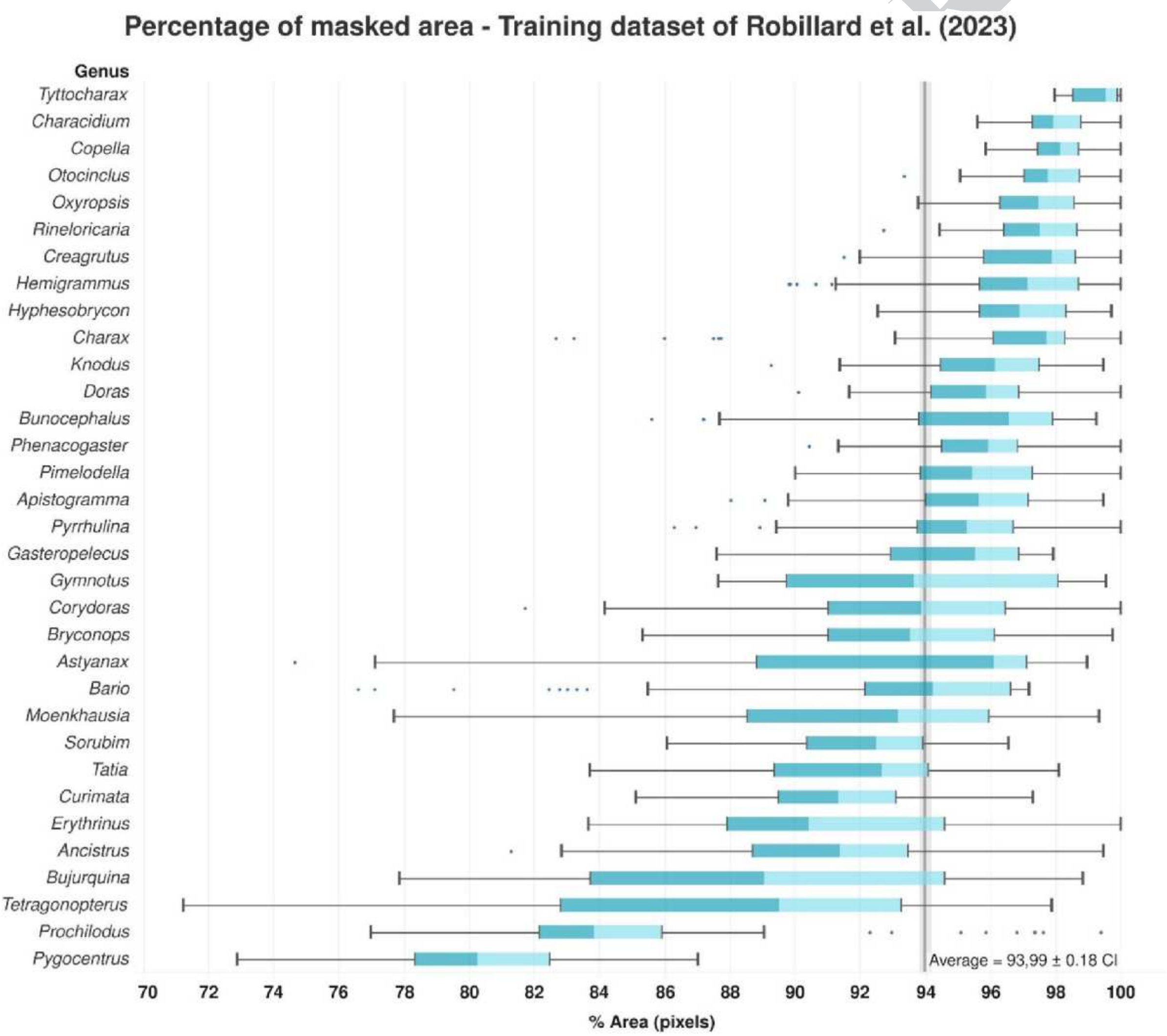

Fig. 1 Percentage of masked area values for 33 fish genera used as training dataset for the 'Amazon fish masker and classifier' model developed in Robillard et al. (2023). Genera names are sorted by average in ascending order. CI = Confidence interval 95%

Journal : **Medium 11160**	Article No : **9901**	Pages : **14**	MS Code : **9901**	Dispatch : **9-10-2024**

Simulated identification

Using fish photographs from our dataset (n = 100) with representatives of 21 genera, we tested the accuracy of the Classifier. Considering all four suggested genus identifications ('*Class_1*', '*Class_2*', '*Class_3*', and '*Class_4*'), The assigned probability values are higher for the first option, but this pattern does not differ when we look at whether the model got the identification right or wrong. Thus, there was no difference in the deviations that would justify stating that the Classifier was more convinced in each of these situations (Table 2 in Supplementary file S3).

The Classifier suggested a total of 32 different genera for our images. This is a 60% increase in estimation, compared to the actual number of genera (n = 21) present in our database. When considering all the outputs given in the four possible genera suggested by the Classifier, *Gymnotus* Linnaeus 1758, *Ancistrus* Kner 1854, *Bunocephalus* Kner 1855, *Rineloricaria* Bleeker 1862, *Moenkhausia* Eigenmann 1903, *Otocinclus* Cope 1871, and *Tetragonopterus* Cuvier, 1816, were the most frequently mentioned, totaling 198 occurrences, which accumulated 51,30% of all identifications in our sample (Table 1).

For the '*black screen test*' the classifier reported that 0.0% of the pixels were 'fish', as expected. However, the application still provided classifications, despite reporting that there was no fish in the image, assigning probabilities to several genera: *Tyttocharax* (39.36%), *Characidium* Reinhardt 1867 (13.74%), *Otocinclus* (9.10%), and *Hemigrammus* Gill 1858 (7.28%).

To assess the correctness of the identifications by the Classifier, we only considered the genera suggested in '*Class_1*'. The Classifier was able to correctly identify the fishes in our pictures at the genus level in only 18 of 100 photographs throughout our dataset (Supplementary file S4). For our sample of photos submitted to the Classifier, only eight out of 21 genera were correctly identified. The highest accuracy was observed for the genera *Gymnotus* (100%, n = 4) and *Ancistrus* (75%, n = 3) (Table 1).

Taking into account only the incorrect identifications (n = 82), we segmented the errors by type, order, family, or genus, always the most extensive. There was an inaccuracy of 65.85% (n = 54) at the order level (Fig. 2a). In 19.51% (n = 16) of the cases, the order was correctly classified, but there was an error at family-level identification (Fig 2.b). The model was unable to correctly classify at the genus level, although it correctly determined the order and family of the photographed specimens in 14.63% (n = 12) of the cases (Fig. 2c).

The beta regression of the relation between the '*Class_1_prob*' variable using the 'fish_pixels' as predictor showed an estimated intercept is approximately 0.791, with a standard error of 0.161. The coefficient for 'fish_pixels' = – 0.1897 was not statistically significant (p = 0.897), and the pseudo-r^2= 1,517*10^{-4} indicates that the model has a very low power to explain much of the variance in the response variable '*Class_1_prob*'. Also, the calculated Fleiss' Kappa coefficient was 0.0126 (20 subjects, 20 raters), indicating a very weak level of agreement among the raters. With an associated z-value of 1.03 and the corresponding *p*-value of 0.304, the agreement was not significantly different from what would be expected by chance.

Discussion

Quality of training dataset

The quality of the images used to train the Classifier model in Robillard et al. (2023) differs noticeably among the genera. The masker model, which extracts pixels related to the body area of the fish from the uploaded images, yields highly disparate results for the different genera of fish analyzed. This variation can be associated with factors such as the size of the fish in the photograph, light, and color intensity. In some cases, the model even completely removes most of the fish body, thus reducing the available 'fish_pixels'.

The amount of information used by the Classifier during the learning phase was little expressive in proportion since it learned from pictures with an average of only 6,01% of 'fish pixels', and for some genera, this value was even lower, less than 5% (values in Table 3, Supplementary file S2.3). This potentially hindered the model's ability to identify photographs of these genera with few pixels used to train the classification database.

Journal : **Medium 11160**	Article No : **9901**	Pages : **14**	MS Code : **9901**	Dispatch : **9-10-2024**

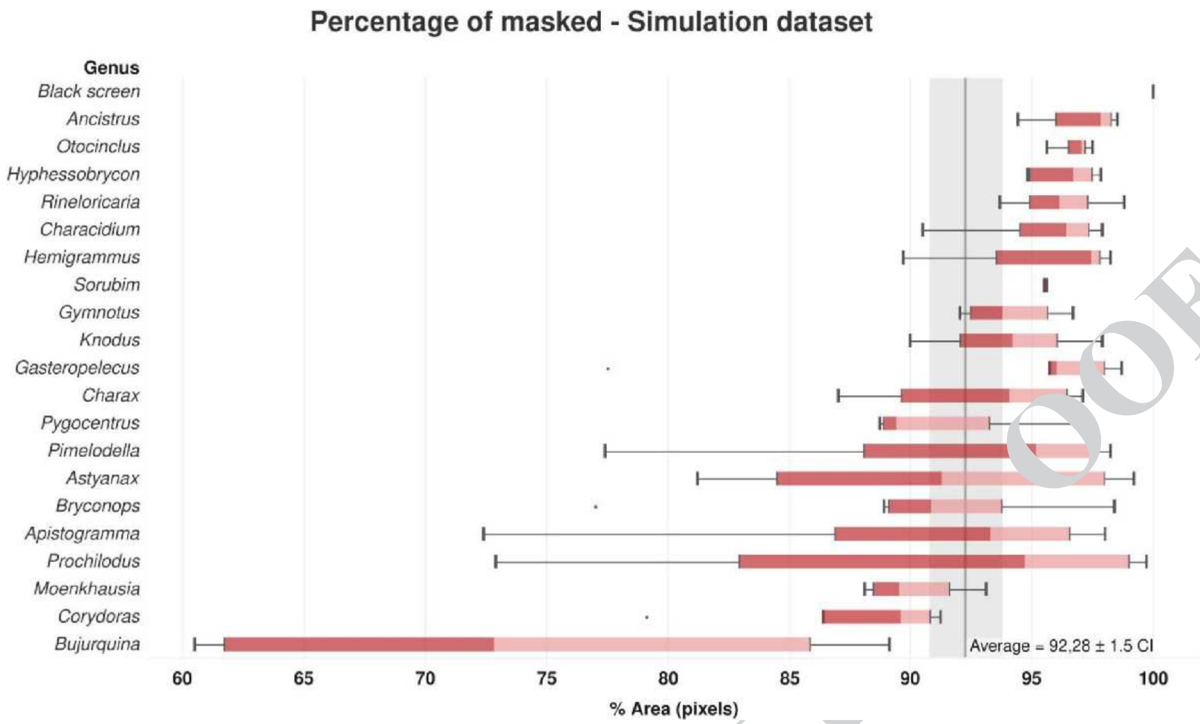

Fig. 2 Percentage of masked area values for 21 fish genera used in simulation in this essay to test the 'Amazon fish masker and classifier' model developed in Robillard et al. (2023). Genera names are sorted by average in ascending order. CI = Confidence interval 95%

Classifier limitations

Although Robillard et al. (2023) recognized that applying the model to images from geographic areas outside the northwest Amazon has not yet been explored, their model was trained and validated with photos of live specimens with varying color patterns and pictures of preserved material from verified collections. Thus, there was some degree of phenotypic variation incorporated into the development of machine learning from inception. Therefore, it would be expected that the Classifier would perform well for the same genera used to train the model, given that the training database was composed of genera widely distributed throughout the Amazon and adjacent basins (Van Der Sleen and Albert 2018).

During our tests, we found cases where the masker removed most of the fish body, leaving only part of structures such as the pectoral, pelvic, and caudal fins, as in the case of the suckermouth catfish genus *Ancistrus,* where most of the structures were removed (Table 1 in Supplementary file S3). We emphasize that although the model does not consider the presence/absence of the structures that are used to identify taxa in morphological studies, such as odontodes or fleshy tentacles, the model was still able to properly identify some genera. On the other hand, there were also instances where the masker left the fish nearly intact, but misidentifications occurred, as seen in the cases of *Apistogramma* Regan 1913*, Astyanax* Baird and Girard 1854, *and Moenkhausia* (Table 1 in Supplementary file S3).

Indeed, for greater certainty in the identification provided by the model, it is necessary to correctly identify the elements that most directly affect the performance of the Classifier. Our hypothesis that only the number of pixels classified as fish could be a good predictor of the probability assigned by the Classifier overall is unsupported. However, the weak inverse relationship between the predictor variable 'fish_pixels' and the response '*Class_1_prob*' suggests that the Classifier performs better with images that contain less information (*i.e*. less confusion for the model to deal with).

Journal : **Medium 11160**	Article No : **9901**	Pages : **14**	MS Code : **9901**	Dispatch : **9-10-2024**

It would be expected that the application would reject the image when it does not depict a fish (0.0% fish pixels). However, the black screen test revealed that the application, even in the absence of pixels related to fish, still assigned a fish genus identification.

Accuracy

When considering misidentified individuals at the genus level, most of the errors concern specimens of the family Characidae, with the genus *Moenkhausia* being the most often suggested as the likely identification (Fig. 2c). The elevated number of mentions to *Moenkhausia* may be attributable to the nature of the model training, since the Classifier developed by Robillard et al. (2023) was trained on a substantial *Moenkhausia* dataset, encompassing 398 photos including various morphotypes, characterized by variations in morphological traits, such as scale size and color patterns, among others ("Images used to train Amazonian fish classification model" in Dikow (2023)). This morphological plasticity within the *Moenkhausia* genus may have broadened the tolerance of the model for classifying this genus, thus affecting its predictive accuracy. Moreover, this genus presents a challenging and unsettled taxonomy (non-monophyletic) due to species exhibiting variable morphology. The calculated Fleiss' Kappa concordance index reinforces the conclusion that the model's predictive capacity did not surpass what could be anticipated by chance. These findings highlight the difficulties associated with achieving a high level of agreement in accurately identifying fish genera using the current state of the proposed classification model.

Our findings notably contrast with those of Robillard et al. (2023), as they reported only 12 misclassifications out of 596 tested images, consisting of two at the order level and seven at the family level. The authors proposed enhancing accuracy by capturing a series of photos until a suitable masker outcome is achieved. However, the masker model frequently omits crucial structures for genera discrimination, especially in Characidae family, where training dataset images often lack visible caudal, dorsal, pelvic, or anal fins (Table 1 in Supplementary file S3). Despite this, our simulation and the inherent nature of the Classifier reveal its insensitivity to specific anatomical structures.

Robillard et al. (2023) noted that their model had prominent misidentifications particularly in tetras (Characidae), a key family in the Amazonian ichthyofauna (Oliveira et al. 2009; Van Der Sleen and Albert 2018). This highlights that their approach struggles with one of the most significant Amazonian fish groups. In contrast, errors at order and family levels are rare in traditional morphology-based ichthyofaunal inventories that may eventually lead to misidentifications at species, subgenus, and genus levels, especially for small and medium-sized species like tetras (Characidae), catfishes (Siluriformes), and cyprinodontiforms as these groups often exhibit uncertain taxonomy or rely on diagnostic characters not observable in field photographs or images of entire fixed specimens . Hence, the proposed model falls short of surpassing the efficiency of traditional taxonomy. Additionally, for optimal functionality, the model requires an extensive dataset encompassing varied positions, lighting, developmental stages, and colorations, live or preserved, from the majority of species in a given region.

Recommendations

Contrary to the expectation that a higher quantity of information available in the pictures could lead to increased Classifier accuracy, the disagreement between the identifications by specialists and the classification provided was insufficient, and the beta regression results did not demonstrate a significant relationship between the variable 'fish_pixels' and the probability associated with the genus suggested by the Classifier. These findings write down the necessity for further investigation and consideration of other variables that may influence the classification outcomes.

The current application lacks a criterion for rejecting images, assigning genus-level identifications regardless of whether the image depicts a fish. Implementing a simple adjustment to address this limitation is crucial for the effectiveness of approaches like the 'Amazonian Fish Classifier' for accurate fish identification. Caution is warranted when considering the use of Robillard et al. (2023) and similar automated AI image identification applications, particularly given the limitations within the highly diverse South American region. Many freshwater fish groups in South America,

Journal : Medium 11160	Article No : 9901	Pages : 14	MS Code : 9901	Dispatch : 9-10-2024

especially in the Amazon, have incomplete taxonomy, with numerous undescribed species and genera (Reis et al. 2016; Birindelli and Sidlauskas 2018; Van Der Sleen and Albert 2018).

Molecular studies expose cryptic or undescribed species, taxonomic uncertainties, novel arrangements, and proposals for genera, highlighting the unresolved nature of freshwater fish taxonomy (e.g., Benzaquem et al. 2015; Melo et al. 2016a; Melo et al. 2016b; Carvalho et al. 2018; Jacobina et al. 2018; García-Melo et al. 2019; Terán et al. 2020; Pires et al. 2021; Brito et al. 2021; Aguiar et al. 2022; Crispim-Rodrigues et al. 2023; Říčan and Říčanová, 2023). The application by Robillard et al. (2023) is ill-equipped to handle such scenarios, potentially causing confusion within the scientific community and among stakeholders due to its tendency to provide identifications for all images, including those of problematic or undescribed taxa.

It is important to emphasize that recent studies, in addition to the classic morphological examination of specimens, have increasingly incorporated molecular approaches in taxonomic descriptions—specifically, Integrative Taxonomy. This approach aims to validate diagnostic characters or reinforce hypotheses related to the existence of new taxa, especially in groups where morphological differences are not readily apparent or where diagnostic structures are small or variable (e.g., Guimarães et al. 2018; Brito et al. 2019; Guimarães et al. 2019, 2020; Santana et al. 2019; Mattox et al. 2020, 2023; Faria et al. 2021; Reia et al. 2021; Aguiar et al. 2022; Crispim-Rodrigues et al. 2023; Říčan and Říčanová, 2023; Souza et al. 2023).

In these cases, molecular data and methods serve as crucial tools for identifying species and genera, especially when taxonomy is challenging, difficult, or involves cryptic species. The identification model proposed by Robillard et al. (2023) may not accurately classify in such instances. Therefore, despite the considerable cost, discouraging the use of molecular tools for taxonomic identification is unwarranted.

As several fish genera possess diagnostic characteristics, such as delicate structures, internal features, intricate color patterns, osteological structures, gonopodial structures, tooth morphology, and subtle color patterns, some cases require additional molecular tools to resolve their taxonomy. The model by Robillard et al. (2023), relying solely on pixel patterns, does not consider these characters, diminishing its effectiveness in taxa identification.

Conservation concerns

Our primary concern with this approach is focused on the assertion that any citizen can contribute information on species identification and distribution for conservation policies and measures through this application. The scientific community should exercise caution regarding the potential misuse of such applications by non-scientists and stakeholders. For instance, it is reasonable to speculate that the photographs used to train identification models, as tested in this essay, were initially identified by specialists (taxonomists) in institutions A or B. Common sense suggests that the resulting identifications may be perceived as having the same accuracy and value as those provided by traditional taxonomy.

This can open a difficult precedent where the much-needed activity of taxonomists in the field and their identifications can be questioned because the taxa in such area were already identified by the use of an identification model that was fed by other researchers/taxonomists. Following that, wrong and ill-intentioned decisions can become more common with devastating impacts on biodiversity and conservation.

Due to the current inefficiency of the method, if broadly used without the aforementioned biases considered, it could lead to incorrect conservation decisions and impact assessments. For environmental decisions, we should always seek the input of biologists, especially taxonomists in the field or laboratory, to properly support identifications from applications. Such an approach alone is not desirable, especially for freshwater fishes that are under severe pressure from stressors (Dudgeon et al. 2006; Darwall et al. 2018; Harrison et al. 2018; Reid et al. 2019; Tickner et al. 2020; Albert et al. 2020; Ottoni et al. 2023).

At its current development stage, the tool would require several adjustments to the model parameters, so we recommend that its potential use should be limited to larger fish of commercial interest or when conservation implications are not directly affected by decisions based on the application identifications.

Journal : Medium 11160 | Article No : 9901 | Pages : 14 | MS Code : 9901 | Dispatch : 9-10-2024

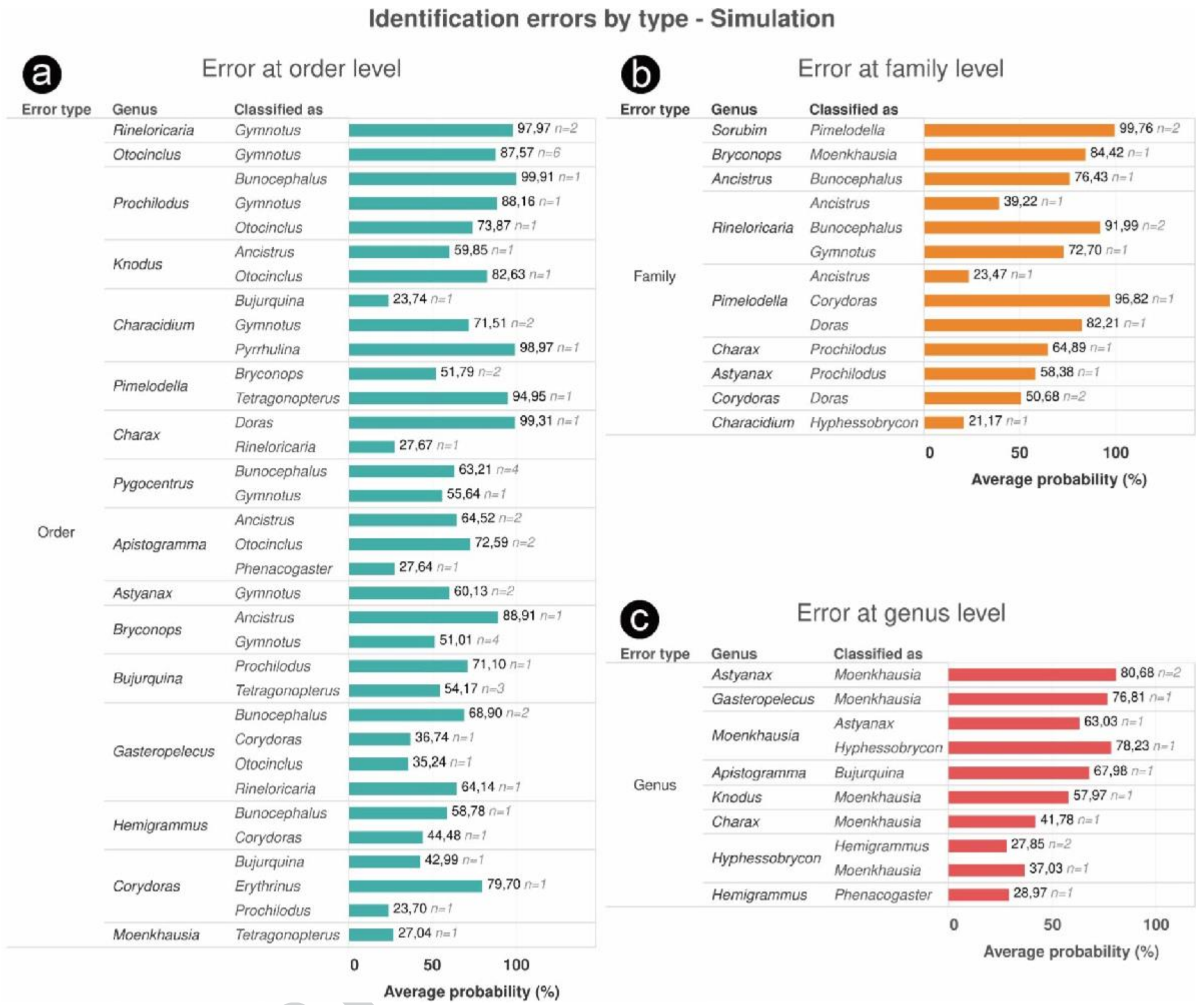

Fig. 3 Distribution of Identification Errors in the 'Amazon Fish Masker and Classifier' Model Developed in Robillard et al. (2023). **a**: the model incorrectly classified at the Order level; **b**: The model correctly identified the order but erred in Family and Genus; **c**: the model only misclassified the genus. Bars represent the probability values associated with the genus suggested by the classifier. The *n* value corresponds to the occurrence count

Conclusion

The application of automated models based on convolutional neural networks (CNN) or similar architectures for fish classification through photograph analysis holds promise. However, the success of these models is contingent upon overcoming various constraints dictated by the intended final application, and it is crucial to acknowledge their current limitations and the need for further refinement.

Upon evaluation in this study, the Robillard et al. application displayed an unsatisfactory performance, with low accuracy and an inability to surpass the null hypothesis of random identifications. The low accuracy on identifications is not beneficial and can bring more confusion to the scientific community, as well as conservation stakeholders. In addition, the potential misuse of such applications by non-scientists and stakeholders raises concerns about the reliability and validity of the data, particularly in comparison to traditional taxonomy conducted by specialists and identifications based on molecular libraries. The argument that automated classifications possess equal accuracy and value as those by taxonomists opens a Pandora's box, challenging the credibility of taxonomic

Journal : **Medium 11160**	Article No : **9901**	Pages : **14**	MS Code : **9901**	Dispatch : **9-10-2024**

work and potentially paving the way for erroneous and ill-intentioned decisions with detrimental consequences for biodiversity and conservation.

Transitioning to the broader implications, the integration of citizen-contributed information for conservation policies is desirable, however, a note of caution is sounded when considering the adoption of methods reliant on Artificial Intelligence, particularly given the potential for misuse by non-scientists and stakeholders. As we navigate the evolving landscape of technological advancements in biodiversity research, a balanced approach that integrates the strengths of both automated tools and expert taxonomic input is essential to ensure the accuracy and integrity of conservation decisions and impact assessments. Collaboration between technological innovations and traditional expertise becomes paramount in addressing the challenges posed by the dynamic and complex field of biodiversity conservation.

Acknowledgements We express our gratitude to Matías Careaga for generously granting permission to use the images of the recently described species of Bujurquina in our simulation.

Authors' contributions Diego S. Campos: Conceptualization; Methodology; Specimens' identification (taxa not specifically assigned by other authors); Specimens' photographs; Statistical Analysis; Statistical interpretation; Supplementary material; Writing—original draft and Writing—review and editing. Rafael F. Oliveira: Conceptualization; Methodology; Specimens' identification (Loricariidae); Specimens' photographs; Statistical interpretation; Supplementary material; Writing—original draft and Writing—review and editing. Lucas O. Vieira: Specimens' identification (taxa not specifically assigned by other authors); Specimens' photographs; Supplementary material; Writing—original draft and Writing – review and editing. Pedro H.N. Bragança: English writing review and Writing—original draft. Erick C. Guimarães: Specimens' identification (Characidae and taxa not specifically assigned by other authors); Financing and Writing—original draft. Axel M. Katz: Specimens' identification (taxa not specifically assigned by other authors); and Writing—original draft. Elisabeth Henschel: English writing review; Writing—original draft and Writing—review and editing. Pâmella S. Brito: Specimens' identification (Characidae and taxa not specifically assigned by other authors) and Writing—original draft. Josie South: Statistical interpretation; English writing review and Writing—original draft. Felipe P. Ottoni: Conceptualization; Specimens' photographs; Specimens' identification (Cichlidae and taxa not specifically assigned by other authors); Financing; Writing—original draft and Writing—review and editing.

Funding This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES—Finance Code 001), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Fundação Amazônia de Amparo a Estudos e Pesquisas (FAPESPA) for granting scholarships within the scope of their programs. Specifically, we acknowledge the following grants: FAPESPA grant 028/2021 to ECG, CNPq grant 307974/2021–9 to FPO, CAPES grants 88887.674455/2022–00 to DSC, and 88887.808704/2023–00 to RFO, FAPERJ grant E-26/202.006/2020 to AMK).

Data availability In adherence to transparency and reproducibility standards, we provide detailed information on the availability of data associated with this study. Repository Information: Supplementary file S1: ImageJ script used for preparation and particle count analysis of images in the Training Dataset of Robillard et al. (2023). Available at: https://figshare.com/s/4975a88e2b5909fd95c4; Supplementary file S2.1: Results of particle count analysis on images in the training dataset of Robillard et al. (2023). Available at: https://figshare.com/s/0ec926218ce08ec3c4f0; Supplementary file S2.2: Summarized results of particle count analysis on images in the training dataset of Robillard et al. (2023), with each line containing values from all particles in a single image. Available at: https://figshare.com/s/0ec926218ce08ec3c4f0; Supplementary file S2.3: Descriptive statistics for the masked area values obtained from particle count analysis on the images in the training dataset of Robillard et al. (2023). Available at: https://figshare.com/s/0ec926218ce08ec3c4f0; Supplementary file S3: Dataset of pictures used in the Simulation and results with descriptive statistics. Available at: https://figshare.com/s/f3d768ab704f50344424; Supplementary file S4: Results of fish genera identification simulation using the classifier developed in Robillard et al. (2023). Available at: https://figshare.com/s/1fca41538c799e584604. Figures and Tables Construction: Fig. 1 was constructed based on data from Supplementary files S2.1, S2.2, and S2.3; Fig. 2, Fig. 3, and Table 1 were elaborated using values from Supplementary file S4. Methods and Scripts: The ImageJ script used for image preparation and particle count analysis is provided in Supplementary file S1; Results of particle count analysis, simulations, and related data used in figure and table construction are detailed in Supplementary files S2, S3, and S4. Repository Utilization: All supplementary files have been deposited on figshare to ensure accessibility and facilitate reproducibility.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Aguiar RG, Guimarães EC, Brito PS, Santos JP, Katz AM, Dias LJBDS, Carvalho-Costa LF, Ottoni FP (2022) A new species of *Knodus* (Characiformes: Characidae), with deep genetic divergence, from the Mearim and Munim river basins, Northeastern Brazil, and evidence for hidden diversity in adjacent river basins. Neotrop Ichthyol 20(2):e210173. https://doi.org/10.1590/1982-0224-2021-0173

Journal : **Medium 11160**	Article No : **9901**	Pages : **14**	MS Code : **9901**	Dispatch : **9-10-2024**

Albert JS, Destouni G, Duke-Sylvester SM, Magurran AE, Oberdorff T, Reis RE, Winemiller KO, Ripple WJ (2020) Scientists' warning to humanity on the freshwater biodiversity crisis. Ambio 50:85–94. https://doi.org/10.1007/s13280-020-01318-8

Benzaquem DC, Oliveira C, Batista JS, Zuanon J, Porto JIR (2015) DNA barcoding in pencilfishes (Lebiasinidae: *Nannostomus*) reveals cryptic diversity across the Brazilian Amazon. PLoS ONE 10(2):e0112217. https://doi.org/10.1371/journal.pone.0112217

Birindelli JL, Sidlauskas BL (2018) Preface: how far has Neotropical Ichthyology progressed in 20 years? Neotrop Ichthyol 16(3):e180128. https://doi.org/10.1590/1982-0224-20180128

Bragança PHN (2018) *Fluviphylax gouldingi F. walla- cei*, two new miniature killifishes from the middle and upper Rio Negro drainage. Brazilian Amazon. Spixiana 41(1):133–146

Bragança PHN, Ottoni FP, Gama CS, Henschel E (2021) Range extension of a small livebearer fish, Poecilia scalpridens (Garman, 1895) (Cyprinodontiformes, Poeciliidae): a new record for the Jari river drainage, Amapa Brazilian Amazon. Check List 17(4):1081–1087

Brito PS, Guimarães EC, Carvalho-Costa LF, Ottoni FP (2019) A new species of *Aphyocharax* Günther, 1868 (Characiformes, Characidae) from the Maracaçumé river basin, eastern Amazon. Zoosyst Evol 95(2):507–516. https://doi.org/10.3897/zse.95.36788

Brito PS, Guimarães EC, Guimarães KL, Rodrigues LR, Anjos MR, Katz AM, Carvalho-Costa LF, Ottoni FP (2021) Cryptic speciation in populations of the genus *Aphyocharax* (Characiformes: Characidae) from eastern Amazon coastal river drainages and surroundings revealed by single locus species delimitation methods. Neotrop Ichthyol 19(4):e210095. https://doi.org/10.1590/1982-0224-2021-0095

Campos DS, Oliveira RF, Vieira LO, Bragança PHN, Nunes JLS, Guimarães EC, Ottoni FP (2023) Revisiting the debate: documenting biodiversity in the age of digital and artificially generated images. Web Ecol 23:135–144. https://doi.org/10.5194/we-23-135-2023

Careaga M, Miranda G, Carvajal-Vallejos FM (2023) Description of two new species of *Bujurquina* (Teleostei: Cichlidae) from the Bolivian Amazon. Neotrop Ichthyol 21:e220093. https://doi.org/10.1590/1982-0224-2022-0093

Carvalho ML, Silva GJC, Melo S, Ashikaga FY, Shimabukuro-Dias CK, Scacchetti PC, Devidé R, Foresti F, Oliveira C (2018) The non-monotypic status of the neotropical fish genus *Hemiodontichthys* (Siluriformes, Loricariidae) evidenced by genetic approaches. Mitochondrial DNA Part a 29(8):1224–1230. https://doi.org/10.1080/24701394.2018.1431230

Cribari-Neto F, Zeileis A (2010) Beta Regression in R. J Stat Software 34(2):1–24. https://doi.org/10.18637/jss.v034.i02

Crispim-Rodrigues JL, Bernt MJ, Waltz BT, Silva GSC, Benine RC, Oliveira C, Covain R, Roxo FF (2023) A new species of mailed catfish of genus *Rhadinoloricaria* (Siluriformes: Loricariidae: Loricariinae) from Rio Negro basin. Brazil J Fish Biol 103(1):103–112. https://doi.org/10.1111/jfb.15402

Darwall W, Bremerich V, Wever A, Dell AI, Freyhof J, Gessner MO, Grossart HP, Harrison I, Irvine K, Jähnig SC, Jeschke JM, Lee JJ, Lu C, Lewandowska A, Monaghan MT, Nejstgaard JC, Patricio H, Schmidt-Kloiber A, Stuart SN, Thieme M, Tockner K, Turak E, Weyl O (2018) The Alliance for Freshwater Life: a global call to unite efforts for freshwater biodiversity science and conservation. Aquatic Conserv: Mar Freshw Ecosyst 28(4):1015–1022. https://doi.org/10.1002/aqc.2958

Dikow R (2023) Images used to train Amazonian fish classification model. Office Chief Inf off Dataset. https://doi.org/10.25573/data.17314730.v1

Dudgeon D, Arthington AH, Gessner MO, Kawabata ZI, Knowler DJ, Lévêque C, Naiman RJ, Prieur-Richard AH, Soto D, Stiassny MLJ, Sullivan CA (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. Biol Rev 81:163–182. https://doi.org/10.1017/S1464793105006950

Faria TC, Guimarães KL, Rodrigues LR, Oliveira C, Lima FC (2021) A new *Hyphessobrycon* (Characiformes: Characidae) of the *Hyphessobrycon heterorhabdus* species-group from the lower Amazon basin. Brazil Neotrop Ichthyol 19(1):e200102. https://doi.org/10.1590/1982-0224-2020-0102

Figueiredo CA, Moreira CR (2018) *Poecilia* (*Pamphorichthys*) *akroa*, a new poeciliid species (Cyprinodontiformes: Poeciliidae) from the Rio Tocantins basin. Brazil. Zootaxa 4461(3):438–444. https://doi.org/10.11646/zootaxa.4464.3.9

Fleiss JL (1971) Measuring nominal scale agreement among many raters. Psychol Bull 76(5):378–382. https://doi.org/10.1037/h0031619

Gamer MLJ, Fellows I, Singh P (2019) irr: Various Coefficients of Interrater Reliability and Agreement (R package version 0.84.1), https://CRAN.R-project.org/package=irr.

Garcia-Melo JE, Oliveira C, Silva GJC, Ochoa-Orrego LE, Pereira LHG, Maldonado-Ocampo JA (2019) Species delimitation of neotropical Characins (Stevardiinae): Implications for taxonomy of complex groups. PLoS ONE 14(6):e0216786. https://doi.org/10.1371/journal.pone.0216786

Guimaraes EC, Brito PS, Feitosa LM, Carvalho-Costa LF, Ottoni FP (2018) A new species of *Hyphessobrycon* Durbin from northeastern Brazil: evidence from morphological data and DNA barcoding (Characiformes, Characidae). ZooKeys 765:79–101. https://doi.org/10.3897/zookeys.765.23157

Guimarães EC, Brito PS, Bragança PHN, Santos JP, Katz AM, Costa LFC, Ottoni FP (2020) Integrative taxonomy reveals two new cryptic species of *Hyphessobrycon* Durbin, 1908 (Teleostei: Characidae) from the Maracaçumé and middle Tocantins River basins, Eastern Amazon region. Eur J Taxon 723:77–107. https://doi.org/10.5852/ejt.2020.723.1145

Guimarães EC, Brito PS, Feitosa LM, Costa LFC, Ottoni FP (2019) A new cryptic species of *Hyphessobrycon* Durbin, 1908 (Characiformes, Characidae) from the Eastern Amazon, revealed by integrative taxonomy. Zoosyst Evol 95(2):345–360. https://doi.org/10.3897/zse.95.34069

Journal : **Medium 11160**	Article No : **9901**	Pages : **14**	MS Code : **9901**	Dispatch : **9-10-2024**

Harrison I, Abell R, Darwall W, Thieme ML, Tickner D, Timboe I (2018) The freshwater biodiversity crisis. Science 362(6421):1369–1369. https://doi.org/10.1126/science.aav9242

Jacobina UP, Lima SMQ, Maia DG, Souza G, Batalha-Filho H, Torres RA (2018) DNA barcode sheds light on systematics and evolution of neotropical freshwater trahiras. Genetica 146:505–515. https://doi.org/10.1007/s10709-018-0043-x

Lima FC, Urbano-Bonilla A, Prada-Pedreros S (2020) A new *Hemigrammus* from the upper Río Vaupés, Colombia (Characiformes: Characidae), with a discussion on the presence of an enlarged urogenital papilla in the family. J Fish Biol 96(4):868–876. https://doi.org/10.1111/jfb.14267

Lima FC, Bastos DA, Py-Daniel LHR, Ota RP (2022) A new sexually dimorphic *Hyphessobrycon* from the western Amazon basin (Characiformes: Characidae). Zootaxa 5116(2):253–266. https://doi.org/10.11646/zootaxa.5116.2.4

Mattox G, Souza C, Toledo-Piza M, Britz R, Oliveira C (2020) A new miniature species of *Priocharax* (Teleostei: Characiformes: Characidae) from the Rio Madeira drainage, Brazil, with comments on the adipose fin in characiforms. Vertebr. Zool. 70(3):417–433. https://doi.org/10.26049/VZ70-3-2020-11

Mattox GM, Britz R, Souza CS, Casas AL, Lima FC, Oliveira C (2023) Two new species of miniature tetras of the fish genus *Priocharax* from the Rio Juruá drainage, Acre, Brazil (Teleostei: Characiformes: Characidae). Can J Zool 101:248–266. https://doi.org/10.1139/cjz-2022-0136

Melo BF, Benine RC, Silva GSC, Avelino GS, Oliveira C (2016a) Molecular phylogeny of the Neotropical fish genus *Tetragonopterus* (Teleostei: Characiformes: Characidae). Mol Phylogenet Evol 94:709–717. https://doi.org/10.1016/j.ympev.2015.10.022

Melo F, Ochoa LE, Vari RP, Oliveira C (2016b) Cryptic species in the Neotropical fish genus *Curimatopsis* (Teleostei, Characiformes). Zool Scripta 45:650–658. https://doi.org/10.1111/zsc.12178

Olden JD, Kennard MJ, Leprieur F, Tedesco PA, Winemiller KO, García-Berthou E (2010) Conservation biogeography of freshwater fishes: recent progress and future challenges. Divers Distrib 16(3):496–513. https://doi.org/10.1111/j.1472-4642.2010.00655.x

Oliveira RR, Rocha MS, dos Anjos MB, Zuanon J, Py-Daniel LHR (2009) Fish fauna of small streams of the Catua-Ipixuna extractive reserve, state of Amazonas. Brazil. Check List 5(2):154–172. https://doi.org/10.15560/5.2.154

Ottoni FP, South J, Azevedo-Santos VM, Henschel E, Bragança PHN (2023) Freshwater biodiversity crisis: multidisciplinary approaches as tools for conservation. Front Environ Sci 11:246. https://doi.org/10.3389/fenvs.2023.1155608

Pires WMM, Barros MC, Fraga EC (2021) DNA Barcoding unveils cryptic lineages of *Hoplias malabaricus* from Northeastern Brazil. Braz J Biol 81(4):917–927. https://doi.org/10.1590/1519-6984.231598

Reia L, Oliveira C, Benine RC (2021) *Moenkhausia andrica* (Characiformes: Characidae): a new species from the rio Tapajós basin, Brazil, with minute fin hooklets in females. J Fish Biol 99(4):1380–1392. https://doi.org/10.1111/jfb.14847

Reid AJ, Carlson AK, Creed IF, Eliason EJ, Gell PA, Johnson PT, Kidd KA, MacCormack TJ, Olden JD, Ormerod SJ, Smol JP, Taylor WW, Tockner K, Vermaire JC, Dudgeon D, Cooke SJ (2019) Emerging threats and persistent conservation challenges for freshwater biodiversity. Biol Rev 94(3):849–873. https://doi.org/10.1111/brv.12480

Reis RE, Albert JS, Di Dario F, Mincarone MM, Petry P, Rocha LA (2016) Fish biodiversity and conservation in South America. J Fish Biol 89(1):12–47. https://doi.org/10.1111/jfb.13016

Říčan O, Říčanová Š (2023) A new highly apomorphic species of *Bujurquina* (Teleostei: Cichlidae) from a reverse flowing river in the Peruvian Amazon, with a key to the species in the genus. Eur J Taxon 870:167–201. https://doi.org/10.5852/ejt.2023.870.2127

Robillard AJ, Trizna MG, Ruiz-Tafur M, Dávila Panduro EL, Santana CD, White AE, Dikow RB, Deichmann JL (2023) Application of a deep learning image classifier for identification of Amazonian fishes. Ecol Evol 13(5):e9987. https://doi.org/10.1002/ece3.9987

Santana CD, Crampton WG, Dillman CB, Frederico RG, Sabaj MH, Covain R, Ready J, Zuanon J, de Oliveira RR, Mendes-Júnior RN, Bastos DA, Teixeira TF, Mol J, Ohara W, Castro NC, Peixoto LA, Nagamachi C, Sousa L, Montag LFA, Ribeiro F, Waddell JC, Piorski NM, Vari RP, Wosiacki WB (2019) Unexpected species diversity in electric eels with a description of the strongest living bioelectricity generator. Nat Commun 10(1):1–10. https://doi.org/10.1038/s41467-019-11690-z

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9(7):671–675

Silvano RA, Pereyra PE, Begossi A, Hallwass G (2022) Which fish is this? Fishers know more than 100 fish species in megadiverse tropical rivers. FACETS 7:988–1007. https://doi.org/10.1139/facets-2021-0136

Souza CS, Mattox GM, Vita G, Ochoa LE, Melo BF, Oliveira C (2023) Molecular species delimitation and description of a new species of *Phenacogaster* (Teleostei, Characidae) from the southern Amazon basin. ZooKeys 1164:1–21. https://doi.org/10.3897/zookeys.1164.102436

Tableau (2023) Tableau Desktop [Computer software]. Retrieved from: https://www.tableau.com/products/desktop (Date of access: 25/July/2023)

Terán GE, Benitez MF, Mirande JM (2020) Opening the Trojan horse: phylogeny of *Astyanax*, two new genera and resurrection of *Psalidodon* (Teleostei: Characidae). Zool J Linn Soc 190(4):1217–1234. https://doi.org/10.1093/zoolinnean/zlaa019

Tickner D, Opperman JJ, Abell R, Acreman M, Arthington AH, Bunn SE, Cooke SJ, Dalton J, Darwall W, Edwards G, Harrison I, Hughes K, Jones T, Leclère D, Lynch AJ, Leonard P, Mcclain ME, Muruven D, Olden JD, Ormerod SJ, Robinson J, Tharme RE, Thieme M, Tockner KT, Wright MW, Young L (2020) Bending the curve of global freshwater biodiversity loss: an emergency recovery plan. Bioscience 70(4):330–342. https://doi.org/10.1093/biosci/biaa002

Journal : **Medium 11160**	Article No : **9901**	Pages : **14**	MS Code : **9901**	Dispatch : **9-10-2024**

Van Der Sleen P, Albert JS (2018) Field guide to the fishes of the Amazon, Orinoco, and Guianas. Princeton University Press, New Jersey, p 464

Zainal-AbidinMohd.Nor DHSA, Lavoué S, Rahim MA, MohammedAkib NA (2022) Assessing a megadiverse but poorly known community of fishes in a tropical mangrove estuary through environmental DNA (eDNA) metabarcoding. Sci Rep 12(1):16346. https://doi.org/10.1038/s41598-022-19954-3

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal : **Medium 11160**	Article No : **9901**	Pages : **14**	MS Code : **9901**	Dispatch : **9-10-2024**

Journal:	**11160**
Article:	**9901**

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details Required	Author's Response
AQ6	Kindly check and confirm affiliations are correctly identified.	
AQ1	Please confirm if the author names are presented accurately and in the correct sequence (given name, middle name/initial, family name). Author 1 Given name: [Diego Sousa] Last name [Campos]. Author 2 Given name: [Rafael Ferreira] Last name [Oliveira]. Author 3 Given name: [Lucas] Middle name [de] Last name [Oliveira Vieira]. Author 4 Given name: [Pedro Henrique Negreiros] Middle name [de] Last name [Braganca]. Author 5 Given name: [Erick Cristofore] Last name [Guimarães]. Author 6 Given name: [Axel Makay] Last name [Katz]. Author 8 Given name: [Pâmella Silva] Middle name [de] Last name [Brito]. Author 10 Given name: [Felipe Polivanov] Last name [Ottoni]. Also, kindly confirm the details in the metadata are correct.	
AQ2	Kindly provide supplementary for this article.	
AQ3	Kindly provide graphical abstract.	
AQ4	Reference Guimarães et al. (2018) was mentioned in the manuscript; however, this was not included in the reference list. As a rule, all mentioned references should be present in the reference list. Please provide the reference details to be inserted in the reference list.	
AQ5	Reference Lima et al. (2020), Lima et al., 2022 Guimaraes et al. (2018), Bragança et al. (2021), Bragança (2018), Figueiredo and Moreira, 2018) was provided in the reference list; however, this was not mentioned or cited in the manuscript. As a rule, if a citation is present in the text, then it should be present in the list. Please provide the location of where to insert the reference citation in the main body text.	