

Review

# Computer-Assisted Bioidentification Using Freshwater Macroinvertebrates: A Scoping Review

Lilian Dayana Cruz <sup>1,\*</sup>, Diego Mauricio Lopez <sup>1</sup>, Rubiel Vargas-Canas <sup>2</sup>, Apolinar Figueroa <sup>3</sup>  
and Juan Carlos Corrales <sup>1</sup>

<sup>1</sup> Telematics Engineering Group, Faculty of Electronics Engineering and Telecommunications, Universidad del Cauca, Popayán 190003, Colombia

<sup>2</sup> Dynamic Systems, Instrumentation and Control Research Group, Faculty of Natural, Exact and Education Sciences, Universidad del Cauca, Popayán 190003, Colombia

<sup>3</sup> Environmental Studies Group, Faculty of Natural Sciences and Education, Universidad del Cauca, Popayán 190003, Colombia

\* Correspondence: liliancruz@unicauca.edu.co

**Abstract: Background:** Evaluation and prediction of the freshwater status based on freshwater macroinvertebrates (FwM) has become valuable in bioindication because they provide a more general and accurate picture of the ecological status of water bodies over time. Recent research on bioindication through FwM has increased the use of computational technologies, mainly in the classification and data analysis stages of water quality assessment and prediction. **Objective:** This scoping review aims to provide an overview of different approaches in computer-assisted bioindication with FwM. Particularly, the objective is to identify the techniques and strategies employed for FwM automatic classification or data treatment, characterize their use in recent years, and discuss gaps and challenges to broaden the scope of bioindication as a tool for understanding real conditions in a water body. **Design:** The scoping review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) extension for scoping reviews (ScR). Scopus and Web of Science databases were used to identify articles published between 1999 and 2022. We selected 81 publications that used computational technology for automatic FwM classification or data analysis to predict water quality using biological indices. **Results and conclusions:** We identified two areas of applying computational technologies in bioindication studies with FwM. Firstly, computer-assisted technologies are used to evaluate water quality through samples already classified by human experts which correspond to 57% of the documents analyzed. The second application area is the automatic classification of FwM. In addition, we determined the main critical factors affecting strategy selection in each of the studies, such as taxonomic resolution, sample size and quality, image quality, data size, and complexity. Finally, we established the relationship between the strategies and algorithms employed in a timeline for automatic classification according to available FwM image databases. The research will allow taxonomic and related experts to better understand the role of computational technologies in FwM studies and thus increase confidence in these techniques to extend their use in bioassessment tasks.

**Keywords:** bioindicators; bioinformatics; computer-assisted biomonitoring; freshwater macroinvertebrates; informatics; machine learning



**Citation:** Cruz, L.D.; Lopez, D.M.; Vargas-Canas, R.; Figueroa, A.; Corrales, J.C. Computer-Assisted Bioidentification Using Freshwater Macroinvertebrates: A Scoping Review. *Water* **2022**, *14*, 3249. <https://doi.org/10.3390/w14203249>

Academic Editor: Celestine Iwendu

Received: 3 September 2022

Accepted: 8 October 2022

Published: 15 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Justification

Bioevaluation through freshwater macroinvertebrates (FwM) provides a more general and accurate picture of the ecological status of water bodies over time [1]. FwM is a widely known taxonomic group used in bioindication tasks to determine environmental changes in rivers or lagoons [2,3]. FwM are used because of their high sensitivity to environmental alterations, many existing species, extensive geographic distribution, long

life cycles, well-known taxonomy, sedentary nature, and their direct contact with the soil (benthos) sediments [4,5].

Macrobenthos communities are susceptible to disturbances in their habitat and human-induced sedimentation processes [5,6]. These changes affect their relative presence or absence in the environment, inducing the multiplication of dominant species that replace the ecological niche of similar species or, on the other hand, decreasing their diversity, affecting the provision of food resources for other species. The analysis of changes in these organisms' diversity, evenness, and richness allows for estimating water bodies' ecological health status [5].

The measurement of ecological conditions using freshwater macroinvertebrates is based on the morphotaxonomic classification [7] of hundreds or thousands of specimens sampled at stations located throughout a water body. According to these organisms' taxonomic resolution and ecological characteristics, it is possible to determine the environmental conditions with a high percentage of accuracy [3]. In this sense, evaluation of the ecological conditions of the water body depends on the treatment and representativeness of the samples, the accuracy in the classification results, and the analysis of the data by experts [4,8].

The use of computational technology helps to reduce time in bioindication tasks [9–11] and extends the scope of studies reducing the need for expert knowledge [12]. Moreover, computational technology also has the potential to decrease the bioassessment gap in middle- and low-resource countries [7]. Given the advantages associated with computational technologies, this work exposes a challenge from the point of view of integrating the different approaches, models, and data, which would allow improving the methods, especially for the identification and classification stages of FwM.

However, because this research area's complex and multidisciplinary nature has not been fully explored, a scoping review is necessary to examine and discuss the critical issues in the existing research on computer-assisted FwM bioindication. Moreover, strategies and trends in this study area and future perspectives for this type of research have been identified.

### 1.2. Objectives

This scoping review focuses on quantitatively and qualitatively exploring the current status of computer-assisted FwM bioindication studies, particularly on key aspects, strategies employed, and research gaps. The paper is organized according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA-ScR) checklist [13].

### 1.3. Research Questions

The research questions we sought to answer with this review were:

- RQ-1: What purposes are pursued when applying computational technology in bioindication tasks with FwM?
- RQ-2: What computer-assisted tools have been used in studies with FwM?
- RQ-3: What is the relationship between the application areas identified and the computational technologies used?
- RQ-4: What is the general approach to address FwM studies using computational technologies?

## 2. Methods

### 2.1. Eligibility Criteria

The review period was set between 1999 and 2022. It included all articles that referred only to FwM and included the use of computational techniques to directly solve a bio-identification research problem. We excluded literature that was not primary studies (reviews, book chapters, commentaries, and others) or was not related to some computational technology applied in FwM bioindication.

## 2.2. Source Information

The databases selected for the search were Scopus and Web of Science. First, the articles were selected based on the title and abstract. Then, using ScientoPy software [14], it was possible to determine the percentage of duplicate documents in the selected databases (Table 1).

**Table 1.** The number of documents returned from each database using the selected words of the structure.

Search String	Database		% Duplicates Removed
	Scopus	Web of Science	
("water" OR "water quality") AND ("benthic macroinvertebrates" OR "macroinvertebrates" OR "benthic fauna") AND ("freshwater" OR "river" OR "lake") AND ("classification" OR "software" OR "feature extraction" OR "artificial intelligence" OR "pattern recognition" OR "artificial neural network") AND NOT ("*fish*" OR "ocean*" OR "marine" OR "sea")	286	202	28%

Since the selection of FwM studies is focused on the bioassessment of the ecological status of freshwater bodies, we divided the search terms into four groups. The first group included terms related to water quality. The second group described the bioindicator, using the keywords "benthic macroinvertebrates", "macroinvertebrates", and "benthic fauna". The third group determined the place where the bioindicator is found, i.e., freshwater, rivers, or lakes. In the fourth group of keywords, we considered general computer technology terms, such as "classification", "feature extraction", "artificial intelligence", and "software", to cover all possible technologies. In order to not include studies focused on marine life, we excluded the terms fish, ocean, marine, and sea.

The word "classification" has a double connotation in this manuscript. On the one hand, it refers to one of the stages of evaluating and predicting freshwater status based on FwM. On the other hand, this word is also employed in the "automatic classification" of FwM images. Finally, keywords were included to exclude studies with macroinvertebrates in marine systems.

In this phase of the literature search, we found many recurring synonyms in the literature that could be eliminated from the final search string.

## 2.3. Search

We used the PRISMA methodology [13] to search, extract, and analyze the data. Firstly, a general reading or mapping [15] of the literature was performed to determine the keyword structure (Table 1) and identify the research questions of this scoping review. Since the selected studies on FwM were conducted for bioassessment purposes, it was necessary to include biological terms in the search strings. We then established an extraction strategy in the primary information sources for data extraction. Finally, we defined and applied inclusion and exclusion criteria and grouped the documents according to application areas, strategies encountered, and the computational technologies employed.

## 2.4. Selection of Sources of Evidence

Before performing the Scoping Review, we conducted a general mapping of the study area based on the methodology presented by Petersen et al. in [15]. This methodology allowed the selection of fundamental concepts of the study, which were grouped into two groups: (i) application areas and (ii) applied computational technologies. The choice of these concepts allowed us to adapt the PRISMA methodology [13,16] and to choose the most relevant studies. We relied on the information flow provided by the PRISMA statement for scoping reviews [17] as a literature search and extraction strategy (Figure 1).

After this, 70 studies published between 1999 and 2022 remained. Subsequently, each article was coded and classified based on a tree (Figure 2) created from technical concepts derived from [18] and specialized websites.

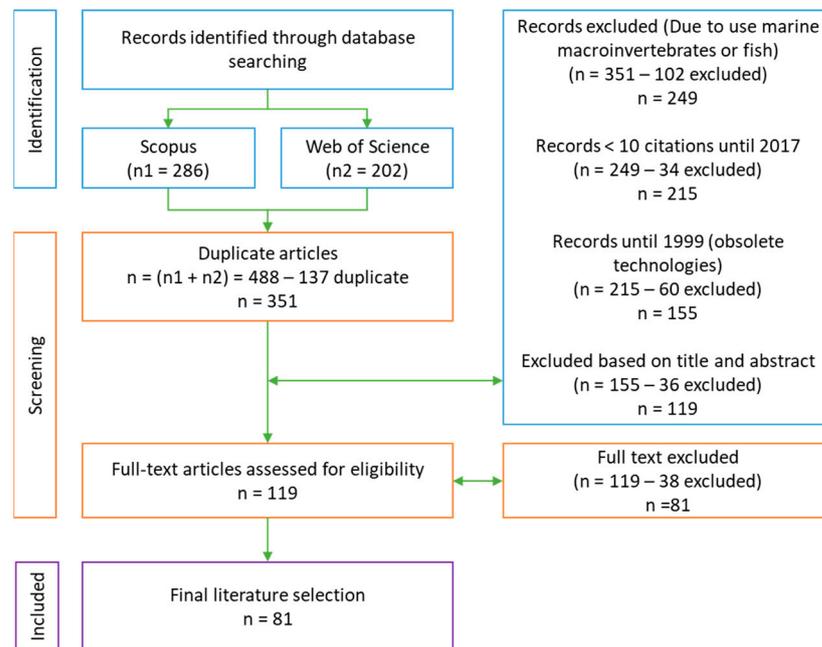


Figure 1. Extraction strategy.

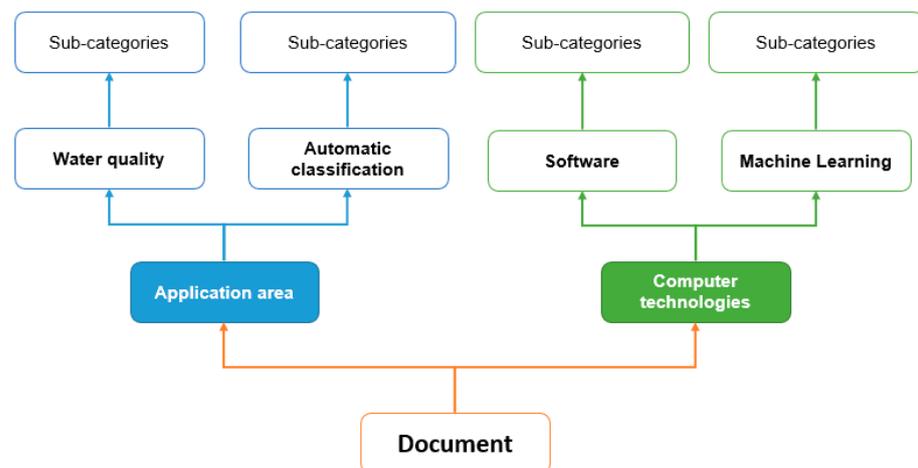


Figure 2. Categories applied to documents.

As we mentioned in Section 2.2, just as there are two types of keyword groups, we created two categories to group the studies (Figure 2). It is worth mentioning that, although there is a grouping called “automatic classification”, this group also involves water quality assessment, which is transversal to all studies.

However, the second group of categories does not imply that the grouped studies are different. We applied the categories and subcategories of “computational technologies” to the identical records. In this case, we found that there were documents that presented two or more technologies used, either for “quality” or for “classification”; therefore, the records for analyzing the technologies used were increased to include all the technologies applied.

The subcategories of the classification tree are described in Appendix A.

## 2.5. Data Charting Process

Before performing the scoping review, we conducted a general mapping of the study area based on the methodology presented by Petersen et al. in [15]. This methodology allowed the selection of fundamental concepts of the study, which were grouped into two groups; (i) application areas and (ii) applied computational technologies. The choice of these concepts allowed us to adapt the PRISMA methodology [13,16] and to choose the most relevant studies. We relied on the information flow provided by the PRISMA statement for scoping reviews [17] as a literature search and extraction strategy (Figure 1). After this, 81 studies published between 1999 and 2022 remained (Appendix E). Subsequently, each article was coded and classified based on a tree (Figure 2) created from technical concepts derived from [18] and specialized websites. The subcategories of the classification tree are described in Appendix A.

### 2.5.1. Quantitative Analysis of the Data Obtained, and the Categories Assigned

We obtained data such as titles, year, number of citations, and others, directly from the databases, and also from the abstracts of each article. We performed a descriptive statistical analysis, mainly through frequency tables.

### 2.5.2. Qualitative Analysis of Critical Factors, Methods, Computer Technologies, and Strategies Employed

We obtained data by analyzing the full text of the selected studies. The main results are timelines and graphs showing the strategies, technologies, and datasets used.

Other graphics were created to represent the relationship between extracted data and the categories assigned to each document. For example, they show how the application areas in the documents have behaved over time concerning the number of citations. Others result from qualitatively analyzing the critical factors and strategies detected in the deep full-text study. We created these graphs with R Software [19] and Bibliometrix [20].

## 2.6. Data Elements

In each phase, we used templates to record information and extract as much relevant data from the documents as possible. In addition, these templates were designed to build an overview of the scoping review that evidenced the extent, variety, and nature of the characteristics of the research area [13].

## 2.7. Synthesis of Results

The synthesis of the results was inspired by the computational method created by IBM researchers Harland Mills and Niklaus Wirth in the 1970s, called the top-down and bottom-up approaches [21]. The top-down approach begins with a general context of the data (evolution of the study area, citations, and other metadata) to describe each specific interaction between the application areas and the computational technologies used. For the bottom-up approach, we first addressed the data related to the particular strategies of each study framed around critical factors, such as decreasing data complexity, maximizing information retrieval, or selecting features in taxonomic groups.

## 3. Results

### 3.1. Characteristics of the Sources of Evidence

Within the time range, the sources of evidence are described through the number of citations, the type and number of documents, the number of keywords, and the authors' information (Table 2). According to the number of sources, we can state that in the first years, the research area was dynamic; however, between 2007 and 2017, it presented a significant decrease for the following years (Figure 3). It should be mentioned that most of the papers are concentrated in Europe, and there are no studies of this type in Latin America (Figure 4).

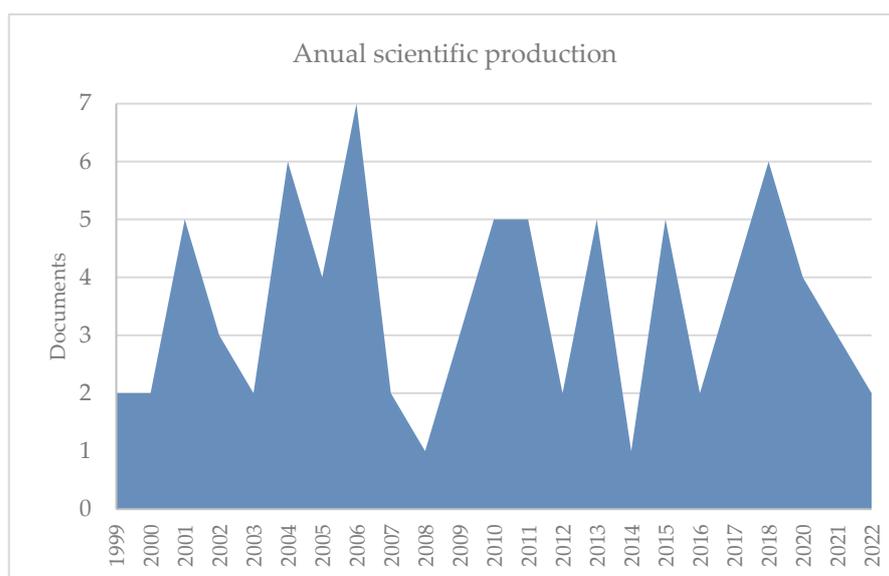
### 3.2. The Individual Outcome of the Sources of Evidence

In each phase, we used templates to record information and extract as much relevant data from the documents as possible (Table 3). In addition, these templates were designed to build an overview of the scoping review that evidenced the extent, variety, and nature of the characteristics of the research area [13].

Table 3 summarizes the most relevant data extracted from the papers analyzed. For phase 1, we took general data such as title, year, and the number of citations from the databases. On the other hand, the specific data refer to the categories and subcategories mentioned in Section 2.4 that we assigned to each article after reading the titles and abstracts. The relationship between the general data and the given categories allowed us to create frequency graphs of an application area or technology used over time.

**Table 2.** Description of sources of evidence.

Description	Results
<b>General</b>	
Timespan	1999–2022
Documents	81
<b>Document types</b>	
Article	62
Book chapter	9
Conference paper	10
<b>Citations</b>	
Average citations per document	41, 5
Average citations per year per doc	3, 54
<b>Authors</b>	
Authors	243
Authors of single-authored documents	5
Authors of multi-authored documents	263



**Figure 3.** Annual Scientific Production (chart extracted with MS Excel).

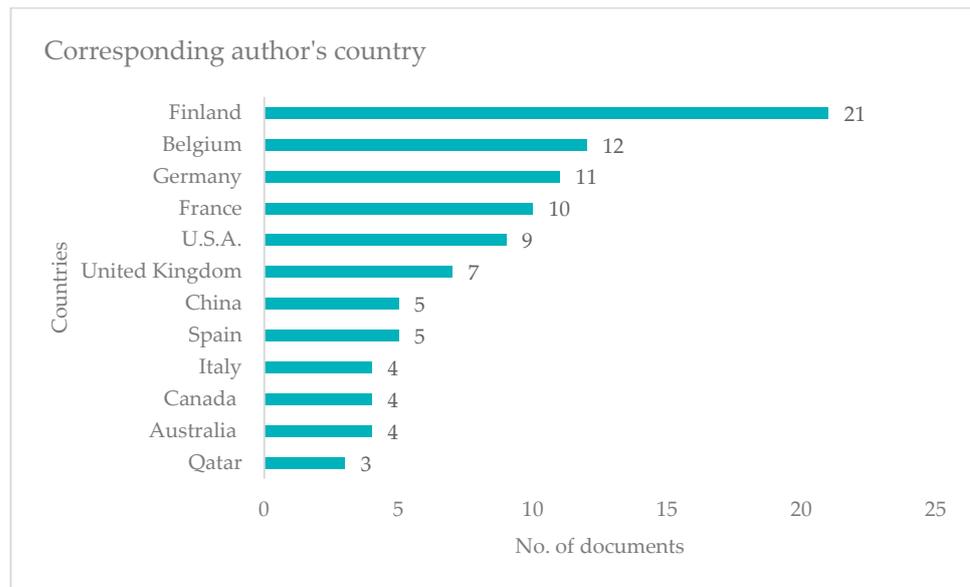


Figure 4. Most relevant countries (chart extracted with MS Excel).

Table 3. Description of chart data for individual sources of evidence.

Method	Phase	Template	Data	Description	Research Question
Top-Down	1	General data	ID, title, authors, year, citations, keywords.	Data extracted from sources.	RQ-1: What purposes are pursued when applying computational technologies in bioindication tasks with FwM?
		Specific data	Category application area, subcategory application area, computer technology.	Categories are given to each of the records according to the general review phase.	RQ-2: What computer-assisted tools have been used in studies with FwM? RQ-3: What is the relationship between the application areas and the computational technologies used?
Bottom-Up	2	Specific data	Critical factors; strategies; algorithms/computing tools; datasets and gaps.	Relevant information is extracted during the complete reading phase of each document.	RQ-4: What general approach is to address FwM studies using computational technologies?
		General data	Several articles, general strategies, a grouping of computational technologies, and datasets.	Generalization of information on a timeline (application area and computational technologies).	

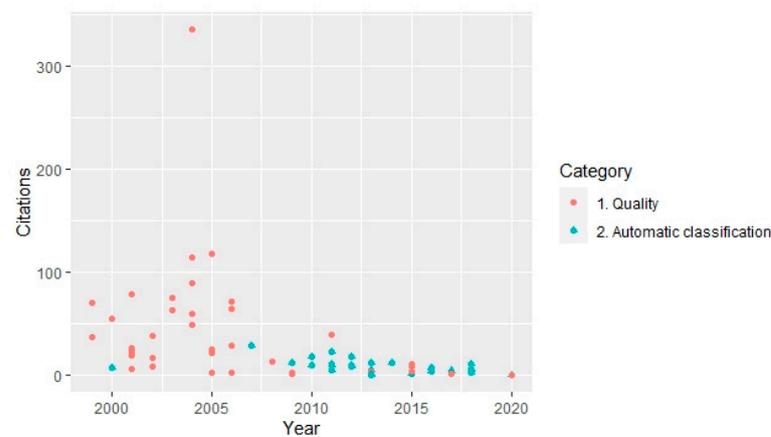
Once we had thoroughly read the articles in the second phase, we extracted general data such as the number of articles of groupings of technologies and strategies. This reading allowed us to delimit each article’s critical factors, strategies, and algorithms or computational tools. With this information, we obtained graphs of the evolution in applying technologies, architectures, computational strategies, and datasets.

3.3. Summary of Results

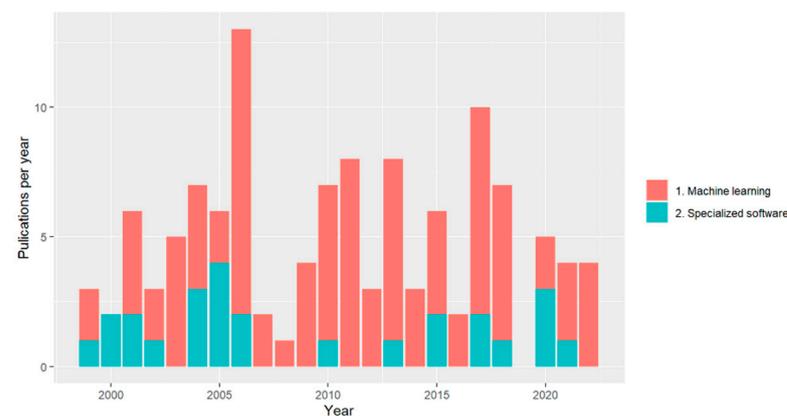
3.3.1. Phase 1: Quantitative Analysis of the Data Obtained, and Categories Assigned (RQ1, RQ2, RQ3)

The research conducted so far is grouped into two application areas: (i). Water quality assessment through data from already classified samples, and (ii). automatic classification through algorithms of images taken in the laboratory (Figure 5). Each application area is

related to the types of computational technologies employed (Figure 6) according to the category tree described in Section 2.4.



**Figure 5.** Document distribution by the application area (quality and automatic classification).



**Figure 6.** Machine learning is applied to water quality and FwM classification.

The first group has a period of concentration between 2000 and 2005. In this one, data processing and multivariate predictive analysis of organisms and populations already classified by human experts were carried out through specialized software [22–28], statistical [29–31], and automatic algorithms [32–37]. In these studies, biological indicators such as EPTC (Ephemeroptera commonly known as Mayfly; Plecoptera, commonly called stoneflies; and Trichoptera, commonly known as caddisflies taxa) [38,39], MBII (Macroinvertebrate Biotic Integrity Index) [40], or BBI (Belgium Biotic Index) [41] were used. The data used are related to the number of species, variability in sample size, biotic factors such as life cycle and morphological development, or abiotic factors such as hydromorphological conditions of water bodies and pollution [42].

Some examples of specialized software for dealing with biological indicators present in this study are:

- River Invertebrate Prediction And Classification System (RIVPACS) and its derivatives, developed by the Institute of Freshwater Ecology (IFE) of the United Kingdom [22,23].
- Integrated Assessment System for the Ecological Quality of Streams and Rivers throughout Europe using Benthic Macroinvertebrates (AQEM) was developed by the AQEM consortium of eight European countries [23–25].
- Standardization of river classifications (STAR) is a project funded by the EU and used to calibrate different results of biological surveys of rivers across Europe against the ecological quality classifications proposed by the Water Framework Directive [23,26,27].

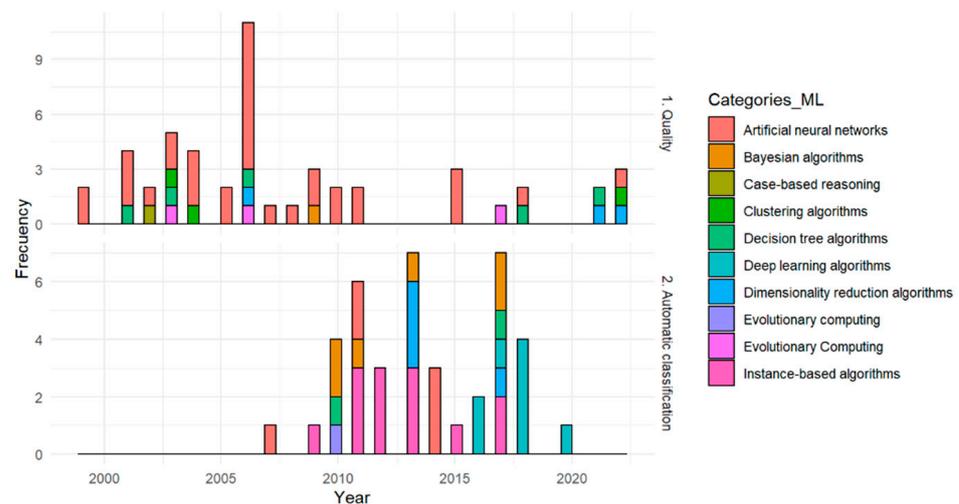
- River Pollution Diagnostic System (RPDS) maximizes information retrieval through clustering and sorting biological and environmental data (nonlinear projection of data in two-dimensional space) [28].

About statistical software, it is found:

- SYSTAT Software: Used for multivariate analysis [29].
- PRIMER-e: Although statistical, this software specializes in multivariate analysis for ecology [30].
- Splus: To obtain statistical correlations between chemical and biological variables [31].

This extensive use of specialized software is due to the increase in the number of studies and data obtained, given the new requirements for bioindication studies with FwM in policies such as the European Union Water Framework Directive created in 2000 [43].

On the other hand, regarding the most employed algorithms in water quality assessment, one finds artificial neural networks (ANNs) (Figure 7), mainly Kohonen Self-Organizing Maps (SOM). Primarily, they were implemented for decreasing data complexity [11,42]; modeling, visualization, and prediction of ecological data to FwM community dynamics [38,44–49]; recognizing patterns in complex spatio-temporal data [50]; establishing the relationship between benthic fauna and biological, environmental parameters [51]; and determining of frequency of occurrence of taxa related to physical or chemical variables [52].



**Figure 7.** Distribution of machine learning categories in the application area (quality and classification).

The second group, concentrated between 2010 and 2015, has widely used machine learning to classify species or families. Some types of ANNs, such as multi-layer perceptron (MLP), probabilistic neural network (PNN), and radial basis function networks (RBFNs) have been frequently applied to evaluate the efficiency of the automatic classification [53]. Nevertheless, in this case, the most used algorithms for automatic classification have been those based on instances such as support vector machine (SVM) and some of its multiclass variations [32–37].

In some instances, SVM algorithms have been optimized based on class selection and splitting methods such as one-vs.-one, one-vs.-all [49], or half-against-half [54–56]. Likewise, Bayesian algorithms, including random forest (RF), random Bayes forest (RBF) [52], or random Bayes array (RBA), have allowed the application of class division, in some cases, regarding taxonomic information [57–59].

Furthermore, other approaches to automatic classification have been introduced recently, especially with deep neural networks (DNNs) as convolutional neural networks (CNNs) [7,60–65], which have significantly improved the classification accuracy. However, the FwM image data needs to be increased to work extensively with these approaches.

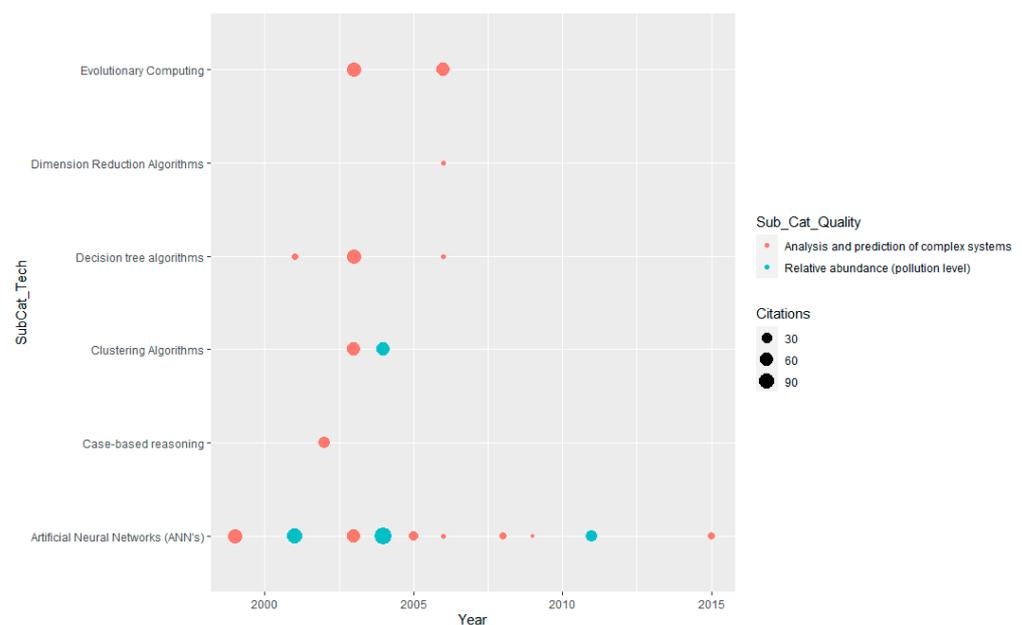
Concerning the type of features for automatic classification, they have predominantly been geometric and statistical [9,55], and deep features learned directly from the data [65]. Section 3.3.3 discusses this approach in more depth. So far, there is no evidence of extensive use of phenotypic or taxonomic criteria in automatic classification [66].

Relationship between the application area and the computational technologies used:

- Water quality:

The most frequent application in quality assessment is analyzing and predicting complex systems and pollution measurement through species' relative abundance. The multivariate character for data management and classification defines the complexity of these studies, which generally require more specialized technological requirements.

Based on the results obtained from the relationship between these application areas and the machine learning algorithms (Figure 8), ANNs are the most widely used algorithms as a modeling method [38,44–49]. However, they are more relevant for the measurement of relative abundance. Thus, their frequent use is related to the detection of spatial patterns and prediction of contaminated areas [67] or for the inter-taxa hierarchical to evaluation of the freshwater status based on macroinvertebrates communities through SOM [42,46,47,68].



**Figure 8.** Algorithms used in the water quality category.

The other studies that employed machine learning algorithms also address the study of complex systems, among which evolutionary computation (EC) or genetic algorithms [69–71] and decision tree analysis (DTA) [69,72,73] stand out. However, it should be noted that research of this type only appears referenced until 2010 because, as shown above, in recent years, studies have focused on the automatic classification of species or families.

- Automatic classification:

A significant concentration of automatic classification works is exhibited between the years 2010 and 2015 (Figure 9). Other types of neural networks are applied during this period, such as RBFNs, the PNN, and the MLP [53]. Moreover, CNNs, in a fine-grained approach, are first used in this group and period.

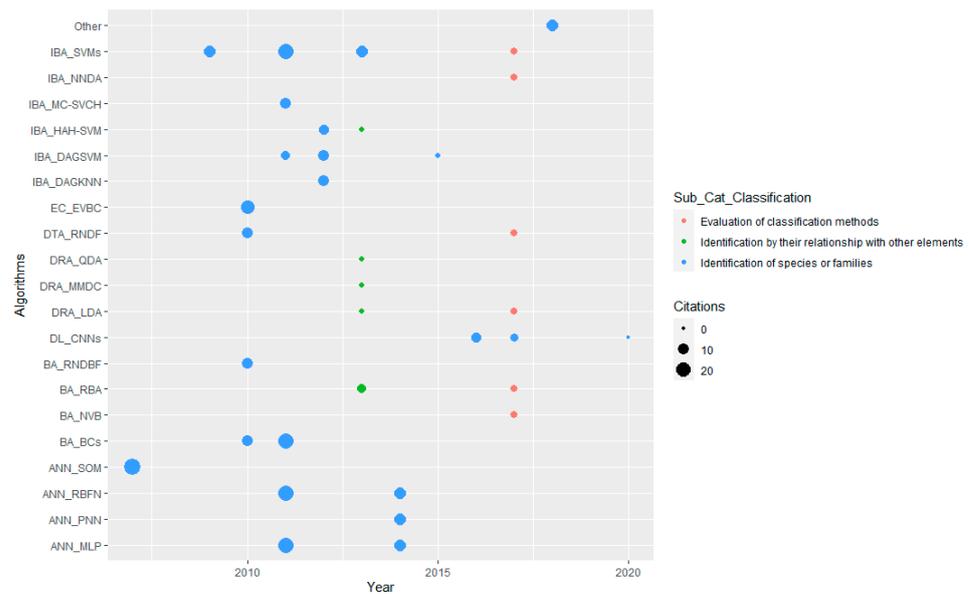


Figure 9. Algorithms used in the area of automatic classification.

### 3.3.2. Phase 2: Qualitative Analysis of Critical Factors, Methods, Computational Technologies, and Strategies Implemented

We detected common critical factors in FwM computer-assisted bioindication studies from a biological and technical point of view (Figure 10). We also obtained the strategies, software, or algorithm characteristics, and datasets used in each work. We then grouped the critical factors and application area (quality and automatic classification) with these strategies to address the evolution of the research area from a multidisciplinary perspective. The results of this grouping can be seen in Appendices C and D.

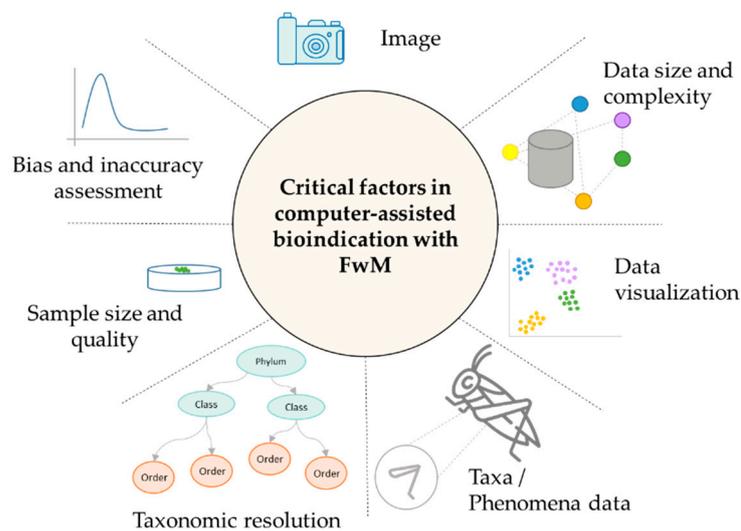


Figure 10. Critical factors in computer-assisted bioindication with FwM.

The technical aspects that generate attention are the size and complexity of the data and the bias and accuracy in the automatic classification of FwM. On the one hand, due to the large sample sizes and, on the other hand, the performance of the classification algorithms, taking into account the very particular morphological characteristics of these organisms. Taxonomic resolution is one of the biological factors most present in the literature, which represents a hierarchy of classification of organisms in which each level exhibits morphological variations about different environmental adaptations [2,25,74]. In recent years, research has focused on automatic classification, first through applying

traditional automated techniques for feature extraction, description, representation, and classification, and then through more novel approaches such as deep learning.

### 3.3.3. Characteristics of the Most Recent Studies

- Widely used deep network: CNNs [7,60,61,63–65].
- Neural network architecture: In the early years, AlexNet exhibited good results [60–62]. In [7], a ResNet architecture with 50 layers was implemented for the first time, which executes a “neuron shutdown” to improve efficiency. The results were superior in accuracy.
- Transfer learning: In [56], it was reported that the automatic classification accuracy improved by 8 to 9%. Since the images used for pre-training tasks such as ImageNet do not look like FwM, further testing is needed in future studies [7,61].
- Dimensional reduction: It was used only to reduce feature vectors obtained through deep learning and thus apply traditional classification techniques (e.g., SVM). However, they could not maintain the discriminative power of the original deep features [62].
- Morphotaxonomic classification: In [7] is the first time that morphotaxonomy is addressed regarding automatic classification methods. In this paper, the authors found that the performance of a deep classifier is more efficient on hierarchical multiclass data than traditional machine learning approaches applied on each taxonomic resolution level. Furthermore, their performance on images was superior to that of human experts. In another study [63], deep networks achieved excellent results compared to manual classification by humans on a limited set of classes.

### 3.3.4. Databases Used in Recent Years and Available for Automatic Classification Studies

FwM image databases have been very scarce. The existing ones have been used both from traditional machine learning approaches such as SVM [65] and with deep learning [63,65]. There are two open-access databases, FIN-Benthic [75] and FIN-Benthic2 [76], created in Finland under a controlled laboratory protocol described in [65]. The photos correspond to organisms in their larval stage, except for one pair, in FIN-Benthic for the adult stage. In general, the taxonomic groups are mainly present in northern Europe. Very few are present in the Neotropics. We described the databases in Appendix B.

## 4. Discussion

### 4.1. Summary of Evidence

We conducted a quantitative and qualitative analysis of the information obtained in a scoping review to describe the critical aspects of FwM computer-aided bioindication. Furthermore, we aim to establish a framework from a practical point of view, not only for the selection of computational technologies but also for the definition of new and better implementation strategies, according to the specific requirements of the research area.

#### 4.1.1. Phase 1: Quantitative Analysis of Retrieved Data and Assigned Categories

In the period established for this scoping review, the high citation points of the documents suggest two key aspects to consider: 1. Visualization of spatio-temporal distribution of macrobenthic communities [42]; and 2. multiparametric analysis through software assesses ecological status and causes of environmental degradation (AQEM) [24,25]. In both cases, the studies coincide temporarily with enacting the Water Framework Directive [43] and the latent need to expand the taxonomic level without increasing the time required for analysis. In this sense, computational technologies have adequately supported bioindication research and application growth through FwM.

Because of the increase in machine learning techniques in recent years (Figure 6), the trend in water quality assessment has been to use supervised approaches, such as neural networks, to process large volumes of data. For automatic classification, instance-based algorithms (e.g., SVM) and some multiclass variants [32–37] are the most frequent. However,

some studies point out the need for higher performance approaches that incorporate fine-grained features related to diversity, richness, dominance, and similarity of morphological features present in this type of organism [61,62,65].

The results reported in Section 3.3.1 show an insufficient number of studies on automated FwM classification. Consequently, introducing new computerized technologies in the taxonomy field has been slow. Moreover, there are no common reference points between biology and computer science, which contributes to increasing gaps in the extensive use of new technological approaches to urgent environmental problems [65].

Computer vision (CV) techniques for image classification of FwM offer a way to process large amounts of data with high efficiency and reproducibility [66]. Nevertheless, the morphological characteristics of these organisms differ from the basic, distinguishable categories used in CV problems [62]. Moreover, these differences are not always evident even for a human expert [9,62,63]. In this sense, the accuracy of automatic classification results is critical to avoid affecting biological evaluation indices and to reduce the bias caused by automation in classification methods [59].

#### 4.1.2. Phase 2: Specific Analysis of Critical Factors, Methods, Computational Technologies, and Strategies Employed (Methodological Approach)

A deep full-text analysis of the papers helped determine the critical factors for bioindication studies using some computational technology. This selection of essential factors also considered specialized literature on the treatment of phenotypic or biological traits through automatic techniques, available in [66,77–79].

To summarize, the strategies implemented by the papers approached evaluation and prediction of the freshwater status based on the FwM problem like any other, i.e., they do not consider the morphotaxonomic aspects of these organisms. However, in recent years, fine-grained approaches [61,62,65] have been presented as an option to deal with this type of feature, as they allow the localization of semantic pieces in the FwM images to facilitate feature categorization [65]. While this approach is advantageous, the macroinvertebrate morphotaxonomic classification problem is even more granular than the typical scenarios and datasets often used in fine-grained approaches [62]. After that, the structure of the models must learn from complex and abstract feature representations from the raw data [80].

On the other hand, because black-box techniques have been frequently employed to increase accuracy, end-user interpretability has been sacrificed. As a result, proposed solutions are application-centered and not user-centered [80], and consequently, experts are reluctant to use such techniques [12,55]. The above, in turn, induces low levels of confidence and, therefore, these methods are not used extensively in bioassessment tasks. The evolution of automated techniques has been related to the amount of data available. Thus, large databases were not required when more traditional approaches were applied in the early years. However, in recent years, with the use of deep learning methods, it has been necessary to increase the number of images, even implementing data enrichment and transfer learning techniques [60–63,65].

Therefore, medium-term developments will require exploring other architectures, assembling opaque and transparent machine learning techniques, and the application of interpretability techniques to find the best possible configuration. In addition, generating more interpretable models will help increase experts' confidence in the extensive use of automatic models in FwM bioindication problems. In Appendix C, we summarize all the strategies employed according to the critical factors and the application areas (quality and automatic classification). Finally, in Appendix D, we present two evolution graphs on a timeline of the strategies, the traditional machine learning techniques, and deep learning.

- Gaps in water quality

The groups selected in the samples cannot always be generalized [11] because of the dynamics of the communities. In studies with FwM, the restrictions on the dataset, the region, and the time needed to make the environmental assessment through bioindicators

will always be challenging for researchers and environmental authorities. Samples do not always contain the necessary species diversity exhibiting the expected distribution along the contamination gradient [30,42]. Therefore, there is a lack of an accurate method for error checking at each step, especially when the number of samples increases and thus the computational cost and information quality requirements [81]. In addition, sometimes, some variables are not considered, leading to prediction errors [52,82].

Since there are no clear criteria for labeling areas, about all variables considered [28] or boundaries between groups [83], studies may be prone to bias and variability in the results when some type of computational technique is employed. Moreover, the levels of taxonomic resolution chosen do not always represent the groups necessary for a complete evaluation [25] or may not even consider taxa that are sufficiently frequent but have more representative ecological traits [44].

The use of black-box approaches such as multilayer ANNs has low desirable interpretive power to gain insight into the relationships elicited by ecological phenomena [84], to detect causalities inherent in the data [45], or because not all pollution-related factors are evaluated [48]. Comparison of different methods using computational technologies' relative accuracy in FwM studies is challenging because they depend on the data sampling, frequency of occurrence of taxa, region, and types of streams [23,73].

- Gaps in automatic classification

The absence of a standardized protocol for imaging this type of organism may cause bias or overfitting in the result due to issues such as the classification threshold area and the adopted animal posture [85]. As a result of the close phylogenetic relationships between levels of taxonomic resolution, namely phylum, class, order, genus, family, and species and the ecological characteristics they exhibit, there is a need for greater availability of images at all levels [68,85]. In addition, there is an absence of better methods to create more robust databases containing more classes [55].

Before feature extraction, traditional CV approaches often required preprocessing tasks. However, this can be a difficult step with FwM images due to the morphological complexity of these organisms [32], which may evidence faults or errors that could propagate during the automatic classification process [57]. The different configurations for implementing classifiers in bioindication problems with FwM do not follow a transparent model or method [86], mainly when addressing morphotaxonomic criteria. Moreover, the methodological ambiguity in assembling these configurations does not allow generating more robust and scalable architectures [33].

- Future challenges

The ecological assessment of freshwater bodies poses a challenge from a technical and biological point of view. Therefore, it is an area that requires a multidisciplinary approach to the choice of techniques, both for estimating or predicting quality and for automatic classification. The choice of the most appropriate techniques will depend on the intersection between the knowledge accumulated by experts and the different technological solutions available. Those studies that integrate both dimensions or seek common reference points [65] will be able to address the problem better and generate solutions more in line with the needs present in this area of study.

Inter- and intraclass differences and similarities in FwM [61] represent a complex problem from a pattern recognition point of view [54]. In this sense, future technological developments will require involving morphotaxonomic criteria taking advantage of all the knowledge accumulated in biology and ecology. Despite using black-box models, future developments must be more user-centered, considering experts' resistance to automatic taxonomic identification methods [55]. Additionally, accuracy in the classification process is a determinant of reliability in environmental assessment [87] and requires special attention.

Expanding bioindication with FwM and reducing the time for morphotaxonomic classification and analysis of the resulting data can ostensibly broaden the application of this method. In addition, this would overcome the bioassessment gap in countries with scarce

and low resources [7]. In the case of LATAM, for instance, the contribution of bioassessment is still unclear, so the only criteria for water conservation are the water resource availability for consumption and economic activities. See, for example, the Colombian national water resource management plan 2009–2022 [88].

More FwM databases that address other dimensions such as the richness and diversity of the Neotropics are urgently needed to advance this research line. It is also crucial that these image databases include water quality assessment studies. This intersection would allow validating the results obtained in automatic classification with biotic integrity indices in the context of bioassessment.

#### 4.2. Limitations

Because of the scarcity of open FwM image databases, the studies did not include contrasted results or comparative analyses of different data sources. In the area of automatic classification, this leads to limited development of the research area and possible bias in the results. On the other hand, it was necessary to make the deep full-text of the documents in the process of categorization and qualitative and quantitative analysis since there is no formal method that addresses computer-assisted bioindication for both FwM and other indicator organisms.

### 5. Conclusions

This scoping review allowed us to map the whole research area of bioindication through FwM from the point of view of computational technologies or tools. Furthermore, we analyzed the relationships between the application areas (water quality and automatic classification) and the technologies to determine the purpose for which they were employed in bioassessment, answering research questions RQ1, RQ2, and RQ3.

To further describe and understand the study area from the point of view of morpho-taxonomic features and solution approaches for automatic classification, we adopted a methodological approach to address problems of this type to solve research question RQ4. This perspective required a deeper analysis of the critical factors and strategies employed in the revised papers to understand the area's evolution over time. Other aspects included the available databases and their use in the selected period.

The integration of computer-assisted techniques and bioindication represents a challenge from a technical-biological perspective due to the diversity of complex morphological features to address. These features are related to the environment, the anthropic conditions present, and how algorithms operate with these features. For this reason, knowing how different techniques have been applied can lead to a better selection of these techniques according to the biological complexity or the creation of new approaches to face this type of problem.

This scoping review will also allow experts in areas such as artificial intelligence or computer vision to find common ground with expert taxonomists, biologists, or ecologists, enabling the creation of better computer technology by promoting more bioassessment studies of freshwater bodies across FwM.

**Author Contributions:** Conceptualization, L.D.C., D.M.L., R.V.-C., A.F. and J.C.C.; methodology, D.M.L.; software, L.D.C.; validation, L.D.C.; formal analysis, L.D.C., D.M.L.; investigation, L.D.C., D.M.L., R.V.-C., A.F. and J.C.C.; resources, D.M.L.; data curation, D.M.L. and R.V.-C.; writing—original draft preparation, L.D.C.; writing—review and editing, D.M.L. and R.V.-C.; visualization, L.D.C.; supervision, D.M.L., R.V.-C., A.F. and J.C.C.; project administration, D.M.L.; funding acquisition, D.M.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Bicentennial Excellence Doctoral Scholarship Program; stipulated in Article 45 of Law 1942 of 2018 of the Republic of Colombia.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** We would like to thank The Ministry of Science, Technology, and Innovation in Colombia for funding the research work under the “Bicentennial Excellence Doctoral Scholarship Program”. We also thank the Doctoral program in Telematics Engineering at Universidad del Cauca, Colombia, for financially supporting the publication of this work.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A

Categories applied to documents.

**Table A1.** Categories and subcategories of the application areas and technological computer.

Application Areas		Computer Technology		
Quality	Automatic Classification	Category	Subcategory	
<ul style="list-style-type: none"> <li>• Analysis and prediction of complex systems</li> <li>• Evaluation of ecological status models</li> <li>• Biotic integrity index calculation</li> <li>• Data processing</li> <li>• Ecological status about other environmental aspects</li> <li>• Sampling method evaluation</li> <li>• Modeling of fluvial ecosystems</li> <li>• Relative abundance (level of contamination)</li> </ul>	<ul style="list-style-type: none"> <li>• Biomass estimation</li> <li>• Evaluation of classification methods</li> <li>• Taxonomic groups identification</li> <li>• Identification of taxonomic groups about other environmental elements</li> </ul>	Machine Learning	<ul style="list-style-type: none"> <li>• Regression algorithms (RA)</li> <li>• Instance-based algorithms (IBA)</li> <li>• Decision tree algorithms (DTA)</li> <li>• Bayesian algorithms (BA)</li> <li>• Clustering algorithms (CA)</li> <li>• Association rule learning algorithms (ARL)</li> <li>• Artificial neural networks (ANNs)</li> <li>• Deep learning algorithms (DPA)</li> <li>• Dimension reduction algorithms (DRA)</li> <li>• Evolutionary computing (EC)</li> </ul>	
			Free software and proprietary	<ul style="list-style-type: none"> <li>• Proprietary</li> <li>• Free</li> <li>• Other</li> </ul>

### Appendix B

Description of available open-access databases.

**Table A2.** FIN-Benthic database overview.

Database	FIN-Benthic
Link	<a href="https://etsin.fairdata.fi/dataset/b6faa81a-0d91-4fbd-b010-4ebf3f3da714">https://etsin.fairdata.fi/dataset/b6faa81a-0d91-4fbd-b010-4ebf3f3da714</a> Accessed on 10 May 2020
Date of publication/data creation	11 March 2021/2016
DOI	<a href="https://doi.org/10.1016/j.imavis.2018.06.005">https://doi.org/10.1016/j.imavis.2018.06.005</a>
License	Creative Commons Attribution 4.0 International (CC BY 4.0)
Access	Open
Edit	Finnish Environment Institute
Authors	Raitoharju J., et al. (2018)
Aim	This is a dataset for automatic Fine-Grained Classification of FwM.
# image	15.074 images
Categories	64
Images per category	From 577 to 7
Content	several

**Table A2.** *Cont.*

Database	FIN-Benthic
Taxonomic groups	12 groups of images at the species level 52 groups of images at the genus level
Resolution	256 × 256 pixels
Size	10.89 GB

**Table A3.** FIN-Benthic2 database overview.

Database	FIN-Benthic2—Accessed on 10 May 2020
Link	<a href="https://etsin.fairdata.fi/dataset/a11cdc26-b9d0-4af1-9285-803d65a696a3">https://etsin.fairdata.fi/dataset/a11cdc26-b9d0-4af1-9285-803d65a696a3</a> Accessed on 10 May 2020
Date of publication	30 June 2019
DOI	<a href="https://doi.org/10.1016/j.image.2020.115917">https://doi.org/10.1016/j.image.2020.115917</a>
License	Creative Commons Attribution 4.0 International (CC BY 4.0)
Access	Open
Edit	Jenni Raitoharju
Authors	Jenni Raitoharju
Aim	This is a dataset for automatic Fine-Grained Classification of FwM
# image	460,004 images
Categories	39
Images per category	From 490 to 44,240. Maximum number per copy: 50
Content	<ul style="list-style-type: none"> <li>- Original image, variable size, png</li> <li>- 64 geometric and color scale features calculated using ImageJ</li> <li>- 2 splits: <ul style="list-style-type: none"> <li>■ Comparison data split (80% training, 20% validation)</li> <li>■ Machine learning split (Training 70%, validation 20%, testing 10%)</li> </ul> </li> </ul>
Taxonomic groups	The data comprise 7 orders, 23 families, 30 genera, and 26 species.
Resolution	For these data, it ranges from 32 × 20 pixels to 468 × 540 pixels. The images come from 3 channels
Size	11.01 GB

## Appendix C

Critical factors, strategies, computational technologies, and evolution over time.

**Table A4.** Strategies and critical factors under study quality and classification.

Critical Factors	Type of Computer Technology (SO *)	Strategies Used	Algorithm/Tool
<b>Technical</b>			
Image	Image analysis software, X-Ray tomography (Q)	The use of spatial scales and fractal indices to detect biomass/body size relationship of FwM [89] or body dimensions [85], and to determine soil affectation to micro-porosity left by the burrowing system of FwM [90].	Software: ImageJ software [89]; Adobe Premier LE; Image Tool 2.0 (University of Texas Health Science Center, San Antonio) [85].
	CNNs (C)	Image segmentation from photo-mosaicing and random ordered search approaches through CNNs, to classify fauna in image patches [60].	Framework: Google tensorflow environments; Nvidia DIGITS [60].
Bias and accuracy in the evaluation of the data	Contamination diagnostic systems, case-based reasoning (CBR), decision trees, genetic algorithms (Q)	To maximize information retrieval from clustering and to sort biological and environmental data (nonlinear projection of data in two-dimensional space) [28]. Likewise, considering experience (e.g., CBR), the data were assigned a predictive character that ostensibly improves the results [91]. Genetic algorithms for automatic variable selection in decision trees reduced the error in models used to predict the relative absence of FwM [69].	Software: RPDS [28]; PERPEST model; CBR [91], J48; WEKA [69].
	Classifiers (C)	To evaluate the performance of 6 classifiers through the one-vs-on method in tie situations, given that, in classification with FwM, it is often found with unambiguous final classes or unclassifiable regions [92].	Classifiers: k-nearest neighbors algorithm (k-NN); Linear discriminant analysis (LDA); Minimum Mahalanobis distance classifier (MMDC); Naïve Bayes (NB); Quadratic discriminant analysis (QDA); and SVM. Future extraction: ImageJ. All tests and implementations of OVO were done with Matlab 2010b together with the Bioinformatics Toolbox and Statistics Toolbox.
Data size and data complexity	Data processing software, ANNs, clustering algorithms, adaptive resonance, ecological evaluation software, classification trees, and genetic algorithms. (Q)	Among the most used strategies are: Decreasing data complexity [11,93–95]. Analysis of a large amount of data collected from samples [30]. Correlation between physico-chemical and biological variables [29,31,81]. Classification of data without prior knowledge [83]. Measuring frequency of occurrence of taxa [52]. Modeling and prediction of the richness of ecological and environmental data [82] through a selection of important predictors, information compression [45,96], and self-organization [47,48]. Data reduction through hierarchical clustering [38,42]. Obtaining ecological information and the causes of degradation through biological metrics, modeling, and prediction [24,44,73,95–97]. Automatic selection of input variables to ML algorithms [70]. Computer-assisted sampling methods and protocols address sample variability better [23] or identify temporal variation errors [27]. To recognize patterns in complex spatio-temporal data, related to migration activities [98]; biotic or abiotic parameters [50]; or delimitation of FwM communities [51,99,100]. Multivariate and multimetric analyses of FwM assemblages [101,102] The authors compared ANN configuration and performance criteria for analysing species diversity and environmental variation, considering non-linear variable relationships [103,104]	ANNs: SOM [38,50,51,83,99,102]; real-time recurrent network (RTRN) [93]; Senso-Net [11]; multi-layer-perceptrons based on the backpropagation (BP) [11,38,44,52,96]; generalized regression neural networks GRNN; linear neural networks (LNNs) [82]; multi-layer feed-forward neural network [84,97]. Algorithms: Bootstrap algorithm [81]; K-means [38,95]; decision trees (e.g., J48-WEKA); ANNs of MATLAB 5.3; genetic algorithm [70,73]; adaptive resonance theory (ART) [42]; Classification and regression trees (CART) [101]. Software: SYSTAT Software [29]; AQEM Software [24]; Framework Structural Equation Modelling (SEM) [94]; STARBUGS (STAR Bioassessment Uncertainty Software System) [27]; PRIMER [34].

Table A4. Cont.

Critical Factors	Type of Computer Technology (SO *)	Strategies Used	Algorithm/Tool
<b>Technical</b>			
Data size and data complexity	ANNs and classifiers (C)	Classification of FwM was performed through neural networks using the following strategies: Clustering of data in the classification of freshwater Macroinvertebrate communities through Self-Organizing Maps [68]. Creation of evolutionary binary classifier networks (NBCs) (scalable, adaptive, and accurate) when processing large image datasets [86]. Application of a comparison methodology based on a Neural Network architecture in which a comprehensive performance evaluation of several classifiers is carried out [33]. Feature selection in tie situations is based on nonlinear classifiers with one-vs.-one and one-vs.-all approaches, given the similarities and dissimilarities between taxonomic groups [54].	ANNs: SOM [68]. Classifiers: NBCs [86]; SVM; Bayesianian classifiers (BCs); MLPs; RBFNs [33]; multi-class SVM [54]. Future extraction: ImageJ [33,54].
	Classifiers (C)	Dimensionality reduction was applied in comparing various classification methods across feature subset selection and a variant of QDA using a random Bayes matrix (RBA); in the case of small samples with many features [58]. Besides, a directed acyclic graph model with a SVM was applied to classify FwM and six other classification methods [37].	QDA [37]; random classifier Naïve Bayes (RNB) [58]; LDA [37]. Classifiers tested: MLP [58]; k-nearest-neighbor (KNN) [37,58]; RF [58]; SVM [37,58]; RBF (although it is not a classification method, it was constructed as such with a Naïve classifier) [58]; classification tree (CT), NB, MMDC [37]. Feature extraction: ImageJ [58]. Software: VueScan(c) software; HP Scanjet4850 flatbed scanner [58].
Data visualization	ANNs, clustering algorithms, adaptive resonance (Q)	Visualization of species assemblages in a two-dimensional field [83] or predicting community dynamics [38,93]. As well as for data reduction and hierarchical clustering at different spatio-temporal scales (nonlinear data projection) [28,42,89].	ANNs: SOM [38,42,83]; backpropagation ANNs [38]; real-time recurrent network (RTRN) [93]. Specialized software: ImageJ [89]; RPDS [28]. Algorithms: K-means [38].
<b>Biologicals</b>			
Taxa/Phenomic data	Classifiers and their multiclass extensions (C)	Gradient-based feature extraction, eliminating segmentation and local feature detection on edges and other textual features of FwM [32]. Given the propagation of errors in FwM classes in the assigned proportions (automatic classification), a propagation error correction method known as confusion matrix correction in classification was proposed [57]. Optimization in classification and feature set size [34]. Comparison classifiers, improving class division by selecting taxonomic groups with similar external characteristics [56]. Comparison of classifiers with different kernel functions, analyzing tie situations and morphological properties of FwM [36].	Feature extraction: Multiple order gradient histogram (SMOGH) [32]; ImageJ [36,56]. Classifiers: SVM—extension multiclass [32] and binary [36,56]; RF, RBF; Bayes classifier (BC) [57]; decision acyclic graph support vector machines (DAGSVM) [34]; LDA; QDA; MMDC [56].
Taxonomic resolution	Ecological valuation software, ANNs(Q)	High taxonomic resolution to increase the quality of ecological assessment [25]. Mapping of structural and functional inter-taxa relationships [46]. Design of a protocol to provide a methodology for the selection of taxa and samples, which can be evaluated in a standardized way [23]	Software: Evaluation software AQEM (AAS) [25]; STAR bioassessment guidance software [23]. ANNs: SOM [46].

Table A4. Cont.

Critical Factors	Type of Computer Technology (SO *)	Strategies Used	Algorithm/Tool
<b>Biologicals</b>			
Taxonomic resolution	Classifiers and their multiclass extensions, ANNs, DNNs (C).	Using multi-order gradients for feature extraction according to FwM variations in taxonomic resolution [32]. Advanced classification and data retrieval schemes in processing large image datasets with taxonomic features [86] Comprehensive performance evaluation of various classifiers in an ANNs architecture space addresses dissimilarities between taxa [33]. Addressing dissimilarities and similarities between taxonomic groups from a pattern recognition point of view (which species can be separated and which cannot) [54]. A novel approach to class division (Scatter) in a Half-Against-Half Support Vector Machine HAH-SVM by random choice [55]. Classification of hard-to-separate classes without taking into account tie situations [35]. Application of neural networks to classification with FwM taking into account taxonomic diversity [53]. Creation of a reference database in which deep learning techniques were applied, incorporating the AlexNet architecture. The DNNs were used to extract deep features, which were used by classifiers (SVM) to evaluate their performance. [62].	Classifiers: The multi-class SVM classifier [32,62]; multiple order gradient histogram (MOGH) [32]; evolutionary binary classifier network [86]; SVM, Bayesian classifiers (BCs); HAH-SVM [55]; DAGSVM; directed acyclic graph k-nearest neighbor (DAGKNN) [35] ANNs: MLP; RBFN MLP; PNN; RBFNs [53]. Feature extraction: ImageJ [32,33,54,86]. DNNs: CNNs [62].

\* Application Areas: Q: Quality C: Automatic Classification.

### Appendix D

Evolution of automatic classifiers applied in FwM automatic classification studies.

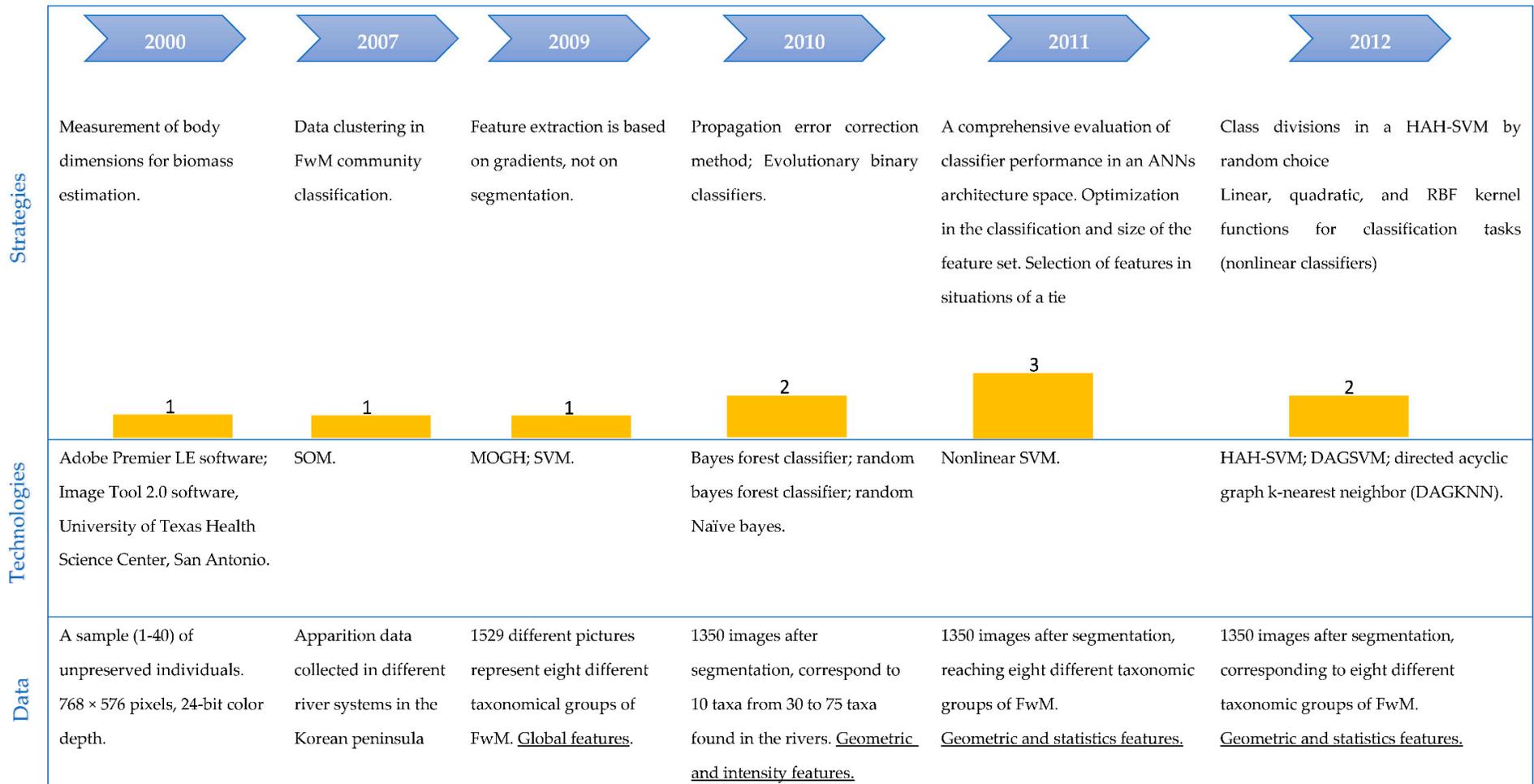
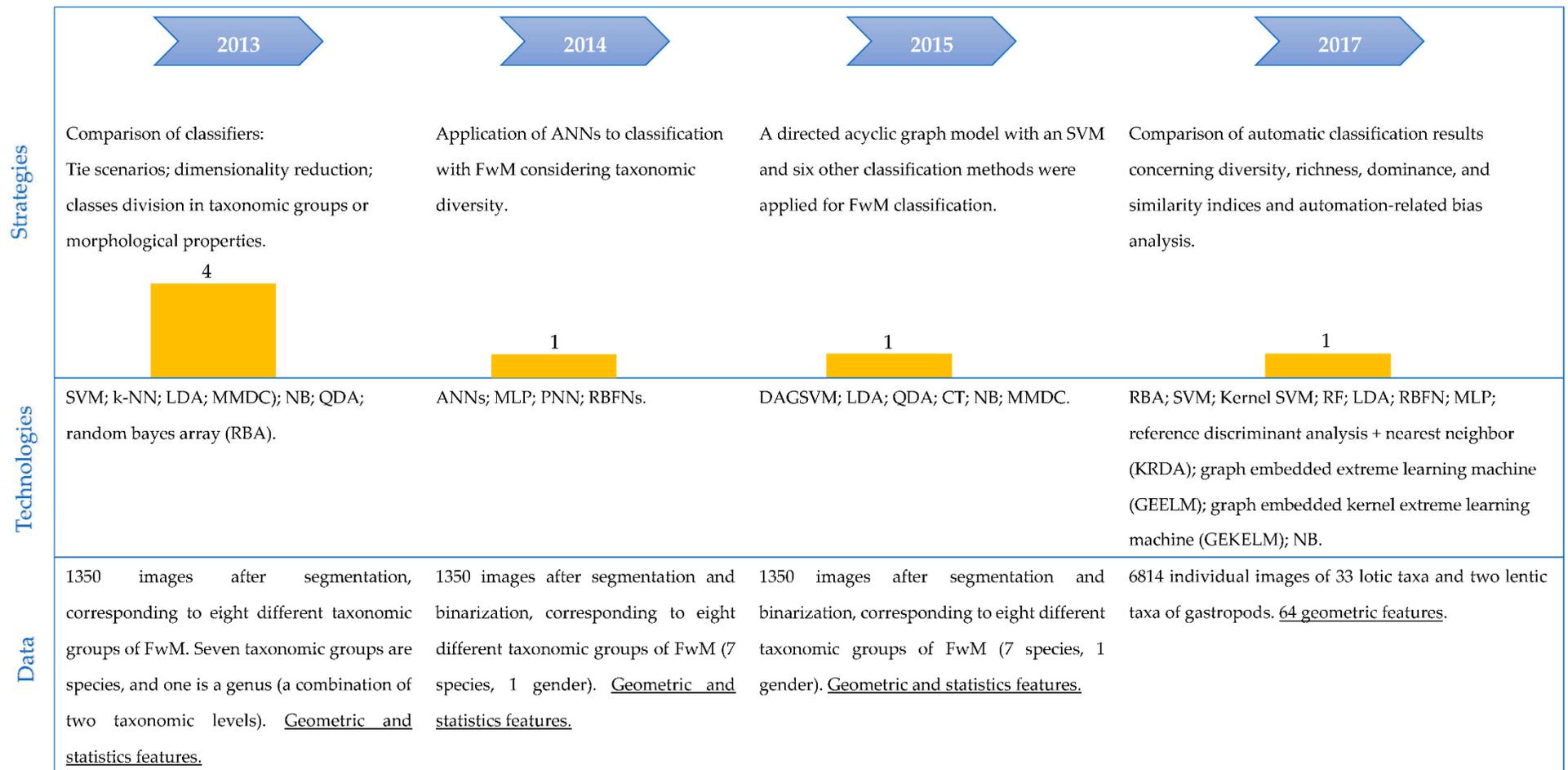


Figure A1. Evolution of the classical machine learning algorithms for FwM. 2000–2012.



**Figure A2.** Evolution of traditional automatic classifiers for FwM. 2013–2017.



Figure A3. Evolution of deep classifiers at FwM. 2016–2021.

## Appendix E

**Table A5.** List of documents extracted for the scoping review.

<b>Id</b>	<b>Type</b>	<b>Year</b>	<b>Title</b>	<b>Author</b>	<b>Publication Title</b>	<b>DOI</b>
1	Journal article	2022	A Bayesian belief network learning tool integrates multi-scale effects of riparian buffers on stream invertebrates	Forio, Marie Anne Eurie Burdon, Francis J. De Troyer, Niels Lock, Koen Witing, Felix Baert, Lotte De Saeyer, Nancy Rîşnoveanu, Geta Popescu, Cristina Kupilas, Benjamin Friberg, Nikolai Boets, Pieter Johnson, Richard K. Volk, Martin McKie, Brendan G. Goethals, Peter L.M.	Science of The Total Environment	10.1016/j.scitotenv.2021.152146
2	Journal article	2022	Influence of environmental variables on macroinvertebrate community structure in Lianhuan Lake	Dou Q., Du X., Cong Y., Wang L., Zhao C., Song D., Liu H., Huo T.	Ecology and evolution	10.1002/ece3.8553
3	Journal article	2021	Can SPEcies At Risk of pesticides (SPEAR) indices detect effects of target stressors among multiple interacting stressors?	Bray, Jonathan P. O'Reilly-Nugent, Andrew Kon Kam King, Guillaume Kaserzon, Sarit Nichols, Susan J. Nally, Ralph Mac Thompson, Ross M. Kefford, Ben J.	Science of The Total Environment	10.1016/j.scitotenv.2020.142997
4	Journal article	2021	Neural network model approach for automated benthic animal identification	Singh, Ravail Mumbarekar, Varun	ICT Express	10.1016/j.ict.2021.03.003
5	Journal article	2021	Agricultural activities compromise ecosystem health and functioning of rivers: Insights from multivariate and multimetric analyses of macroinvertebrate assemblages	Zhang Y., Leung J.Y.S., Zhang Y., Cai Y., Zhang Z., Li K.	Environmental Pollution	10.1016/j.envpol.2021.116655
6	Journal article	2020	Where does land use matter most? Contrasting land use effects on river quality at different spatial scales	Nelson Mwaijengo, G.Msigwa, A. Njau, K.N. Brendonck, L.Vanschoenwinkel, B.	Science of the Total Environment	10.1016/j.scitotenv.2019.134825
7	Journal article	2020	Application of deep learning in aquatic bioassessment: Towards automated identification of non-biting midges	Milošević, D. Milosavljević, A. Predić, B. Medeiros, A.S. Savić-Zdravković, D. Stojković Piperac, M. Kostić, T. Spasić, F. Leese, F.	Science of the Total Environment	10.1016/j.scitotenv.2019.135160
8	Journal article	2020	Tracking wireworm burrowing behaviour in soil over time using 3D X-ray computed tomography	Booth, S. Kurtz, B. de Heer, M.I. Mooney, S.J. Sturrock, C.J.	Pest Management Science	10.1002/ps.5808
9	Journal article	2020	Human experts vs. machines in taxa recognition	Johanna Ärje, Jenni Raitoharju, Alexandros Iosifidis, Ville Tirronen, Kristian Meissner, Moncef Gabbouj, Serkan Kiranyaz, Salme Kärkkäinen	Signal Processing: Image Communication	10.1016/j.image.2020.115917

Table A5. Cont.

<b>Id</b>	<b>Type</b>	<b>Year</b>	<b>Title</b>	<b>Author</b>	<b>Publication Title</b>	<b>DOI</b>
10	Journal article	2018	Benchmark database for fine-grained image classification of benthic macroinvertebrates	Raitoharju, J. Riabchenko, E. Ahmad, I. Iosifidis, A. Gabbouj, M. Kiranyaz, S. Tirronen, V. Ärje, J. Kärkkäinen, S. Meissner, K.	Image and Vision Computing	10.1016/j.imavis.2018.06.005
11	Journal article	2018	Hyperspectral Imaging of Macroinvertebrates—a Pilot Study for Detecting Metal Contamination in Aquatic Ecosystems	Salmelin, J. Pölönen, I. Puupponen, H.-H. Hämäläinen, H. Karjalainen, A.K. Väisänen, A. Vuori, K.-M.	Water, Air, and Soil Pollution	10.1007/s11270-018-3963-2
12	Journal article	2018	Evaluating freshwater macroinvertebrates from eDNA metabarcoding: A river Nalón case study	Fernández, S. Rodríguez, S. Martínez, J.L. Borrell, Y.J. Ardura, A. García-Vázquez, E.	PLoS ONE	10.1371/journal.pone.0201741
13	Journal article	2018	Can high-throughput sequencing detect macroinvertebrate diversity for routine monitoring of an urban river?	Carew, M.E. Kellar, C.R. Pettigrove, V.J. Hoffmann, A.A.	Ecological Indicators	10.1016/j.ecolind.2017.11.002
14	Book section	2018	Linking DNA Metabarcoding and Text Mining to Create Network-Based Biomonitoring Tools: A Case Study on Boreal Wetland Macroinvertebrate Communities	Compson, Z.G. Monk, W.A. Curry, C.J. Gravel, D. Bush, A. Baker, C.J.O. Al Manir, M.S. Riazanov, A. Hajibabaei, M. Shokralla, S. Gibson, J.F. Stefani, S. Wright, M.T.G. Baird, D.J.	Advances in Ecological Research	10.1016/bs.aecr.2018.09.001
15	Journal article	2018	Determining the macroinvertebrate community indicators and relevant environmental predictors of the Hun, Tai River Basin (Northeast China): A study based on community patterning	Zhang M., Muñoz-Mas R., Martínez-Capel F., Qu X., Zhang H., Peng W., Liu X.	Science of The Total Environment	10.1016/j.scitotenv.2018.04.021
16	Journal article	2017	A digital reference collection for aquatic macroinvertebrates of North America	Walters, D.M. Ford, M.A. Zuellig, R.E.	Freshwater Science	10.1086/694539
17	Journal article	2017	The effect of automated taxa identification errors on biological indices	Ärje, J. Kärkkäinen, S. Meissner, K. Iosifidis, A. Ince, T. Gabbouj, M. Kiranyaz, S.	Expert Systems with Applications	10.1016/j.eswa.2016.12.015
18	Conference paper	2017	Data enrichment in fine-grained classification of aquatic macroinvertebrates	Raitoharju, J. Riabchenko, E. Meissner, K. Ahmad, I. Iosifidis, A. Gabbouj, M. Kiranyaz, S.	IEEE Xplore	10.1109/CVAUI.2016.20

Table A5. Cont.

<b>Id</b>	<b>Type</b>	<b>Year</b>	<b>Title</b>	<b>Author</b>	<b>Publication Title</b>	<b>DOI</b>
19	Book section	2017	A variable length chromosome genetic algorithm approach to identify species distribution models useful for freshwater ecosystem management	Gobeyn, S. Goethals, P.L.M.		
20	Conference paper	2016	Deep learning for benthic fauna identification	Marburg, A. Bigham, K.	IEEE Xplore	10.1109/OCEANS.2016.7761146
21	Conference paper	2016	Learned vs. engineered features for fine-grained classification of aquatic macroinvertebrates	Riabchenko, E. Meissner, K. Ahmad, I. Iosifidis, A. Tirronen, V. Gabbouj, M. Kiranyaz, S.	IEEE Xplore	10.1109/ICPR.2016.7899975
22	Journal article	2015	Directed acyclic graph support vector machines in classification of benthic macroinvertebrate samples	Joutsijoki, H. Siemala, M. Juhola, M.	Artificial Intelligence Review	10.1007/s10462-014-9425-3
23	Journal article	2015	Invertebrate diversity in relation to chemical pollution in an Umbrian stream system (Italy)	Pallottini, M. Goretti, E. Gaino, E. Selvaggi, R. Cappelletti, D. Céréghino, R.	Comptes Rendus—Biologies	10.1016/j.crvi.2015.04.006
24	Journal article	2015	Inferring landscape-scale land-use impacts on rivers using data from mesocosm experiments and artificial neural networks	Magierowski, R.H. Read, S.M. Carter, S.J.B. Warfe, D.M. Cook, L.S. Lefroy, E.C. Davies, P.E.	PLoS ONE	10.1371/journal.pone.0120901
25	Journal article	2015	Prediction of Glossosoma biomass spatial distribution in Valley Creek by field measurements and a three-dimensional turbulent open-channel flow model	Morris, M. Mohammadi, M.H. Day, S. Hondzo, M. Sotiropoulos, F.	Water Resources Research	10.1002/2014WR015887
26	Journal article	2015	Application of a self-organizing map and canonical correspondence analysis in modern benthic foraminiferal communities: A case study from the pearl river estuary, China	Li T., Xiang R., Li T.	Journal of Foraminiferal Research	10.2113/gsjfr.45.3.305
27	Journal article	2014	Evaluating the performance of artificial neural networks for the classification of freshwater benthic macroinvertebrates	Joutsijoki, H. Meissner, K. Gabbouj, M. Kiranyaz, S. Raitoharju, J. Ärje, J. Kärkkäinen, S. Tirronen, V. Turpeinen, T. Juhola, M.	Ecological Informatics	10.1016/j.ecoinf.2014.01.004

Table A5. Cont.

<b>Id</b>	<b>Type</b>	<b>Year</b>	<b>Title</b>	<b>Author</b>	<b>Publication Title</b>	<b>DOI</b>
28	Conference paper	2013	Half-against-half structure in classification of benthic macroinvertebrate images	Joutsijoki, H.	IEEE Xplore	10.1109/EMBC.2013.6610333
29	Journal article	2013	Breaking the curse of dimensionality in quadratic discriminant analysis models with a novel variant of a Bayes classifier enhances automated taxa identification of freshwater macroinvertebrates	Ärje, J. Kärkkäinen, S. Turpeinen, T. Meissner, K.	Environmetrics	10.1002/env.2208
30	Journal article	2013	Comparison of species sensitivity distributions based on population or individual endpoints	Beaudouin, R. Péry, A.R.R.	Environmental Toxicology and Chemistry	10.1002/etc.2148
31	Journal article	2013	Kernel selection in multi-class support vector machines and its consequence to the number of ties in majority voting method	Joutsijoki, H. Juhola, M.	Artificial Intelligence Review	10.1007/s10462-011-9281-3
32	Conference paper	2013	An Application of One-vs-One Method in Automated Taxa Identification of Macroinvertebrates	Joutsijoki, Henry	2013 Fourth Global Congress on Intelligent Systems	10.1109/GCIS.2013.26
33	Conference paper	2012	Half-Against-Half Multi-Class Support Vector Machines in classification of benthic macroinvertebrate images	Joutsijoki, H.	IEEE Xplore	10.1109/ICCISci.2012.6297281
34	Book section	2012	DAGSVM vs. DAGKNN: An experimental case study with benthic macroinvertebrate dataset	Joutsijoki, H. Juhola, M.	Lecture Notes in Computer Science	10.1007/978-3-642-31537-4_35
35	Conference paper	2011	Automated benthic macroinvertebrate identification with Decision Acyclic Graph Support Vector Machines	Joutsijoki, H. Juhola, M.	Intelligent Systems and Control/Computational Bioscience—2011	10.2316/P.2011.742-041
36	Journal article	2011	Water toxicity assessment and spatial pollution patterns identification in a Mediterranean River Basin District. Tools for water management and risk analysis	Carafa, R. Faggiano, L. Real, M. Munné, A. Ginebreda, A. Guasch, H. Flo, M. Tirapu, L. der Ohe, P.C.V.	Science of the Total Environment	10.1016/j.scitotenv.2011.06.053

Table A5. Cont.

<b>Id</b>	<b>Type</b>	<b>Year</b>	<b>Title</b>	<b>Author</b>	<b>Publication Title</b>	<b>DOI</b>
37	Book section	2011	Comparing the one-vs-one and one-vs-all methods in benthic macroinvertebrate image classification	Joutsijoki, H. Juhola, M.	Computer Science	10.1007/978-3-642-23199-5_30
38	Journal article	2011	Classification and retrieval on macroinvertebrate image databases	Kiranyaz, S. Ince, T. Pulkkinen, J. Gabbouj, M. Ärje, J. Kärkkäinen, S. Tirronen, V. Juhola, M. Turpeinen, T. Meissner, K.	Computers in Biology and Medicine	10.1016/j.combiomed.2011.04.008
39	Journal article	2011	Development and application of a hybrid model to analyze spatial distribution of macroinvertebrates under flow regulation in the Lijiang River	Chen Q., Yang Q., Lin Y.	Ecological Informatics	10.1016/j.ecoinf.2011.08.001
40	Conference paper	2010	Network of evolutionary binary classifiers for classification and retrieval in macroinvertebrate databases	Kiranyaz, S. Gabbouj, M. Pulkkinen, J. Ince, T. Meissner, K.	IEEE Xplore	10.1109/ICIP.2010.5651161
41	Conference paper	2010	Statistical classification and proportion estimation—An application to a macroinvertebrate image database	Ärje, J. Kärkkäinen, S. Meissner, K. Turpeinen, T.	IEEE Xplore	10.1109/MLSP.2010.5588324
42	Journal article	2010	Automated processing and identification of benthic invertebrate samples	Lytle, David A. Martínez-Muñoz, Gonzalo Zhang, Wei Larios, Natalia Shapiro, Linda Paasch, Robert Moldenke, Andrew Mortensen, Eric N. Todorovic, Sinisa Dietterich, Thomas G.	Journal of the North American Benthological Society	10.1899/09-080.1
43	Journal article	2010	A heuristic approach to predicting water beetle diversity in temporary and fluctuating waters	Gutiérrez-Estrada J.C., Bilton D.T.	Ecological Modelling	10.1016/j.ecolmodel.2010.03.007
44	Journal article	2010	Selecting variables for habitat suitability of Asellus (Crustacea, Isopoda) by applying input variable contribution methods to artificial neural network models	Mouton A.M., Dedecker A.P., Lek S., Goethals P.L.M.	Environmental Modeling & Assessment	10.1007/s10666-009-9192-8
45	Conference paper	2009	Analysing biological, chemical and geomorphological interactions in rivers using Structural Equation Modelling	Bizzi, S. Surridge, B. Lerner, D.N.	World IMACS Congress	

Table A5. Cont.

<b>Id</b>	<b>Type</b>	<b>Year</b>	<b>Title</b>	<b>Author</b>	<b>Publication Title</b>	<b>DOI</b>
46	Journal article	2009	Modeling of the hierarchical structure of freshwater macroinvertebrates using artificial neural networks	Rico, C. Paredes, M. Fernández, N.	Acta Biologica Colombiana	
47	Book section	2009	Multiple Order Gradient Feature for Macro-Invertebrate Identification Using Support Vector Machines	Tirronen, Ville Caponio, Andrea Haanpää, Tomi Meissner, Kristian	Adaptive and Natural Computing Algorithms	10.1007/978-3-642-04921-7_50
48	Journal article	2008	Implementation of artificial neural networks (ANNs) to analysis of inter-taxa communities of benthic microorganisms and macroinvertebrates in a polluted stream	Kim, B. Lee, S.-E. Song, M.-Y. Choi, J.-H. Ahn, S.-M. Lee, K.-S. Cho, E. Chon, T.-S. Koh, S.-C.	Science of the Total Environment	10.1016/j.scitotenv.2007.09.009
49	Journal article	2007	Self-organizing mapping of benthic macroinvertebrate communities implemented to community assessment and water quality evaluation	Song, M.-Y. Hwang, H.-J. Kwak, I.-S. Ji, C.W. Oh, Y.-N. Youn, B.J. Chon, T.-S.	Ecological Modelling	10.1016/j.ecolmodel.2006.04.027
50	Journal article	2007	Applications of artificial neural networks predicting macroinvertebrates in freshwaters	Goethals P.L.M., Dedecker A.P., Gabriels W., Lek S., De Pauw N.	Aquatic Ecology	10.1007/s10452-007-9093-3
51	Book section	2006	Development and application of predictive river ecosystem models based on classification trees and artificial neural networks	Goethals, P. Dedecker, A. Gabriels, W. De Pauw, N.	Ecological Informatics: Scope, Techniques and Applications	10.1007/3-540-28426-5_8
52	Journal article	2006	Errors and uncertainty in bioassessment methods—Major results and conclusions from the STAR project and their application using STARBUGS	Clarke, R.T. Hering, D.	Hydrobiologia	10.1007/s10750-006-0079-2
53	Journal article	2006	Estimates and comparisons of the effects of sampling variation using ‘national’ macroinvertebrate sampling protocols on the precision of metrics used to assess ecological status	Clarke, R.T. Davy-Bowker, J. Sandin, L. Friberg, N. Johnson, R.K. Bis, B.	Hydrobiologia	10.1007/s10750-006-0076-5
54	Journal article	2006	Genetic algorithms for optimisation of predictive ecosystems models based on decision trees and neural networks	D’heygere, T. Goethals, P.L.M. De Pauw, N.	Ecological Modelling	10.1016/j.ecolmodel.2005.11.005

Table A5. Cont.

<b>Id</b>	<b>Type</b>	<b>Year</b>	<b>Title</b>	<b>Author</b>	<b>Publication Title</b>	<b>DOI</b>
55	Book section	2006	Modelling Ecological Interrelations in Running Water Ecosystems with Artificial Neural Networks	Schleiter, I. M. Obach, M. Wagner, R. Werner, H. Schmidt, H. -H. Borchardt, D.	Ecological Informatics	10.1007/3-540-28426-5_9
56	Journal article	2006	Development of an in-stream migration model for <i>Gammarus pulex</i> L. (Crustacea, Amphipoda) as a tool in river restoration management	Dedecker A.P., Goethals P.L.M., D'Heygere T., De Pauw N.	Aquatic Ecology	10.1007/s10452-005-9022-2
57	Journal article	2006	Patterning of impoundment impact on chironomid assemblages and their environment with use of the self-organizing map (SOM)	Penczak T., Kruk A., Grzybkowska M., Dukowska M.	Acta Oecologica	10.1016/j.actao.2006.05.007
58	Book section	2005	A neural network approach to the prediction of benthic macroinvertebrate fauna composition in rivers	Di Dato, P. Mancini, L. Tancioni, L. Scardi, M.	Modelling Community Structure in Freshwater Ecosystems	10.1007/3-540-26894-4_14
59	Journal article	2005	Application of artificial neural network models to analyse the relationships between <i>gammarus pulex</i> L. (Crustacea, Amphipoda) and river characteristics	Dedecker, A.P. Goethals, P.L. D'heygere, T. Gevrey, M. Lek, S. De Pauw, N.	Environmental Monitoring and Assessment	10.1007/s10661-005-8221-6
60	Journal article	2005	Effects of metal pollution on soil macroinvertebrate burrow systems	Nahmani, J. Capowiez, Y. Lavelle, P.	Biology and Fertility of Soils	10.1007/s00374-005-0865-4
61	Journal article	2005	Does macrophyte fractal complexity drive invertebrate diversity, biomass and body size distributions?	McAbendroth, L. Ramsay, P.M. Foggo, A. Rundle, S.D. Bilton, D.T.	Oikos	10.1111/j.0030-1299.2005.13804.x
62	Journal article	2004	Hierarchical community classification and assessment of aquatic ecosystems using artificial neural networks	Park, Y.-S. Chon, T.-S. Kwak, I.-S. Lek, S.	Science of the Total Environment	10.1016/j.scitotenv.2004.01.014
63	Journal article	2004	Assigning macroinvertebrate tolerance classifications using generalised additive models	Yuan, L.L.	Freshwater Biology	10.1111/j.1365-2427.2004.01206.x
64	Journal article	2004	Assessment methodology for southern siliceous basins in Portugal	Pinto, P. Rosado, J. Morais, M. Antunes, I.	Hydrobiologia	10.1023/B:HYDR.0000025266.86493.a2
65	Journal article	2004	Overview and application of the AQEM assessment system	Hering, D. Moog, O. Sandin, L. Verdonschot, P.F.M.	Hydrobiologia	10.1023/B:HYDR.0000025255.70009.a5

Table A5. Cont.

<b>Id</b>	<b>Type</b>	<b>Year</b>	<b>Title</b>	<b>Author</b>	<b>Publication Title</b>	<b>DOI</b>
66	Journal article	2004	The effect of taxonomic resolution on the assessment of ecological water quality classes	Schmidt-Kloiber, A. Nijboer, R.C.	Hydrobiologia	10.1023/B:HYDR.0000025270.10807.10
67	Journal article	2004	Optimization of Artificial neural network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium)	Dedecker A.P., Goethals P.L.M., Gabriels W., De Pauw N.	Ecological Modelling	10.1016/j.ecolmodel.2004.01.003
68	Journal article	2003	Use of genetic algorithms to select input variables in decision tree models for the prediction of benthic macroinvertebrates	D'heygere, T. Goethals, P.L.M. De Pauw, N.	Ecological Modelling	10.1016/S0304-3800(02)00260-0
69	Journal article	2003	Predicting the species richness of aquatic insects in streams using a limited number of environmental variables	Céréghino, R. Park, Y.-S. Compin, A. Lek, S.	Journal of the North American Benthological Society	10.2307/1468273
70	Book section	2002	River Pollution Diagnostic System (RPDS)—Computer-based analysis and visualisation for bio-monitoring data	O'Connor, M.A. Walley, W.J.		10.2166/wst.2002.0045
71	Journal article	2002	Comparison of Artificial Neural Network (ANN) Model Development Methods for Prediction of Macroinvertebrate Communities in the Zwalm River Basin in Flanders, Belgium.	Dedecker, A.P. Goethals, P.L. De Pauw, N.	TheScientificWorldJournal	10.1100/tsw.2002.79
72	Journal article	2002	Perpest model, a case-based reasoning approach to predict ecological risks of pesticides	Van den Brink, P.J. Roelsma, J. van Nes, E.H. Scheffer, M. Brock, T.C.M.	Environmental Toxicology and Chemistry	10.1002/etc.5620211132
73	Journal article	2001	Patterning and short-term predictions of benthic macroinvertebrate community dynamics by using a recurrent artificial neural network	Chon, T.-S. Kwak, I.-S. Park, Y.-S. Kim, T.-H. Kim, Y.	Ecological Modelling	10.1016/S0304-3800(01)00305-2

Table A5. Cont.

Id	Type	Year	Title	Author	Publication Title	DOI
74	Journal article	2001	Spatial analysis of stream invertebrates distribution in the Adour-Garonne drainage basin (France), using Kohonen self-organizing maps	Céréghino, R. Giraudel, J.L. Compin, A.	Ecological Modelling	10.1016/S0304-3800(01)00304-0
75	Journal article	2001	Bioindication of chemical and hydromorphological habitat characteristics with benthic macro-invertebrates based on Artificial Neural Networks	Schleiter, I.M. Obach, M. Borchardt, D. Werner, H.	Aquatic Ecology	10.1023/A:1011433529239
76	Journal article	2001	Stream acidification in South Germany—Chemical and biological assessment methods and trends	Braukmann, U.	Aquatic Ecology	10.1023/A:1011452014258
77	Journal article	2001	River restoration simulations by ecosystem models predicting aquatic macroinvertebrate communities based on J48 classification trees.	Goethals, P. Gasparyan, K. De Pauw, N.	Mededelingen (Rijksuniversiteit te Gent. Fakulteit van de Landbouwkundige en Toegepaste Biologische Wetenschappen)	hdl.handle.net/1854/LU-147855
78	Journal article	2000	Application of an image analysis system to the determination of biomass (ash free dry weight) of pond macroinvertebrates	Bernardini, V. Solimini, A.G. Carchini, G.	Hydrobiologia	10.1023/A:1004153703748
79	Journal article	2000	The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams	Doberstein, C.P. Karr, J.R. Conquest, L.L.	Freshwater Biology	10.1046/j.1365-2427.2000.00575.x
80	Journal article	1999	The influence of data transformations on biological monitoring studies using macroinvertebrates	Thorne, R.S.J. Williams, W.P. Cao, Y.	Water Research	10.1016/S0043-1354(98)00247-4
81	Journal article	1999	Modelling water quality, bioindication and population dynamics in lotic ecosystems using neural networks	Schleiter, I.M. Borchardt, D. Wagner, R. Dapper, T. Schmidt, K.-D. Schmidt, H.-H. Werner, H.	Ecological Modelling	10.1016/S0304-3800(99)00108-8

## References

1. Džeroski, S. Applications of symbolic machine learning to ecological modelling. *Ecol. Model.* **2001**, *146*, 263–273. [[CrossRef](#)]
2. Roldán, G.P. *Guía Para el Estudio de los Macroinvertebrados Acuáticos del Departamento de Antioquia*; Universidad de Antioquia: Medellín, Colombia, 1988.
3. Markert, B.A.; Breure, A.M.; Zechmeister, H.G. (Eds.) *Bioindicators & Biomonitors: Principles, Concepts, and Applications*; Elsevier: Amsterdam, The Netherlands; Boston, MA, USA, 2003.
4. Domínguez, E.; Fernández, H.R. (Eds.) *Macroinvertebrados Bentónicos Sudamericanos: Sistemática y Biología*; Fundación Miguel Lillo: Tucumán, Argentina, 2009.
5. Liu, Z.; Chen, M.; Li, Y.; Huang, Y.; Fan, B.; Lv, W.; Yu, P.; Wu, D.; Zhao, Y. Different effects of reclamation methods on macrobenthos community structure in the Yangtze Estuary, China. *Mar. Pollut. Bull.* **2018**, *127*, 429–436. [[CrossRef](#)] [[PubMed](#)]
6. Lv, W.; Huang, Y.; Liu, Z.; Yang, Y.; Fan, B.; Zhao, Y. Application of macrobenthic diversity to estimate ecological health of artificial oyster reef in Yangtze Estuary, China. *Mar. Pollut. Bull.* **2016**, *103*, 137–143. [[CrossRef](#)] [[PubMed](#)]
7. Milošević, D.; Milosavljević, A.; Predić, B.; Medeiros, A.S.; Savić-Zdravković, D.; Piperac, M.S.; Kostić, T.; Spasić, F.; Leese, F. Application of deep learning in aquatic bioassessment: Towards automated identification of non-biting midges. *Sci. Total Environ.* **2019**, *711*, 135160. [[CrossRef](#)]
8. Zuarth, C.A.G.; Vallarino, A.; Jiménez, J.C.P.; Pfeng, A.M.L. *Bioindicadores: Guardianes de Nuestro Futuro Ambiental*; El Colegio de la Frontera Sur (ECOSUR), Instituto Nacional de Ecología y Cambio Climático (INECC): Primera, Mexico, 2014.
9. Arje, J.; Tirronen, V.; Raitoharju, J.; Meissner, K.; Ainen, S.K. Can humans be replaced by computers in taxarecognition? *Eur. Young Stat.Meet.* **2017**, 27–34.
10. Medellín, R.A.; Viquez-R, L.R. *Los Murciélagos Como Bioindicadores de la Perturbación Ambiental. Bioindicadores: Guardianes de Nuestro Futuro Ambiental*; S y G: Mexico, 2014; pp. 521–542.
11. Schleiter, I.M.; Borchardt, D.; Wagner, R.; Dapper, T.; Schmidt, K.-D.; Schmidt, H.-H.; Werner, H. Modelling water quality, bioindication and population dynamics in lotic ecosystems using neural networks. *Ecol. Model.* **1999**, *120*, 271–286. [[CrossRef](#)]
12. Joutsijoki, H. Half-Against-Half Structure with SVM and k-NN Classifiers in Benthic Macroinvertebrate Image Classification. *JCP* **2014**, *9*, 454–462. [[CrossRef](#)]
13. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.J.; Horsley, T.; Weeks, L.; et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. [[CrossRef](#)]
14. Ruiz-Rosero, J.; Ramirez-Gonzalez, G.; Viveros-Delgado, J. Software survey: ScientoPy, a scientometric tool for topics trend analysis in scientific publications. *Scientometrics* **2019**, *121*, 1165–1188. [[CrossRef](#)]
15. Petersen, K.; Feldt, R.; Mujtaba, S.; Mattsson, M. Systematic Mapping Studies in Software Engineering. In Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE), Gothenburg, Sweden, 13–15 June 2008. [[CrossRef](#)]
16. Kitchenham, B.; Brereton, O.P.; Budgen, D.; Turner, M.; Bailey, J.; Linkman, S. Systematic literature reviews in software engineering—A systematic literature review. *Inf. Softw. Technol.* **2009**, *51*, 7–15. [[CrossRef](#)]
17. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* **2009**, *6*, e1000097. [[CrossRef](#)]
18. Mitchell, T.M. *Machine Learning*; McGraw-Hill: New York, NY, USA, 1997.
19. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
20. Aria, M.; Cuccurullo, C. bibliometrix: An R-tool for comprehensive science mapping analysis. *J. Informetr.* **2017**, *11*, 959–975. [[CrossRef](#)]
21. Mills, H.D. *Mathematical Foundations for Structured Programming*; The Harlan D. Mills Collection: Knoxville, TN, USA, 1972.
22. Wright, J.F. (Ed.) *Assessing the Biological Quality of Fresh Waters: RIVPACS And Other Techniques; Invited Contributions from an International Workshop Held in Oxford, UK an 16–18 September 1997*; Freshwater Biological Assoc: Ambleside, UK, 2000.
23. Clarke, R.T.; Davy-Bowker, J.; Sandin, L.; Friberg, N.; Johnson, R.; Bis, B. Estimates and comparisons of the effects of sampling variation using 'national' macroinvertebrate sampling protocols on the precision of metrics used to assess ecological status. *Hydrobiologia* **2006**, *566*, 477–503. [[CrossRef](#)]
24. Hering, D.; Moog, O.; Sandin, L.; Verdonschot, P.F. Overview and application of the AQEM assessment system. *Hydrobiologia* **2004**, *516*, 1–20. [[CrossRef](#)]
25. Schmidt-Kloiber, A.; Nijboer, R.C. The effect of taxonomic resolution on the assessment of ecological water quality classes. *Hydrobiologia* **2004**, *516*, 269–283. [[CrossRef](#)]
26. Furse, M.T.; Hering, D.; Brabec, K.; Buffagni, A.; Sandin, L.; Verdonschot, P.F.M. The ecological status of European rivers: Evaluation and intercalibration of assessment methods. *Hydrobiologia* **2006**, *566*, 1–2. [[CrossRef](#)]
27. Clarke, R.T.; Hering, D. Errors and uncertainty in bioassessment methods—major results and conclusions from the STAR project and their application using STARBUGS. *Hydrobiologia* **2006**, *566*, 433–439. [[CrossRef](#)]
28. O'Connor, M.; Walley, W. River Pollution Diagnostic System (RPDS)-computer-based analysis and visualisation for bio-monitoring data. *Water Sci. Technol.* **2002**, *46*, 17–23. [[CrossRef](#)]

29. Braukmann, U. Stream acidification in South Germany—chemical and biological assessment methods and trends. *Aquat. Ecol.* **2001**, *35*, 207–232. [[CrossRef](#)]
30. Thorne, R.S.; Williams, W.; Cao, Y. The influence of data transformations on biological monitoring studies using macroinvertebrates. *Water Res.* **1999**, *33*, 343–350. [[CrossRef](#)]
31. Yuan, L.L. Assigning macroinvertebrate tolerance classifications using generalised additive models. *Freshw. Biol.* **2004**, *49*, 662–667. [[CrossRef](#)]
32. Tirronen, V.; Caponio, A.; Haanpää, T.; Meissner, K. Multiple order gradient feature for Macroinvertebrate identification using support vector machines. In *Adaptive and Natural Computing Algorithms*; Kolehmainen, M., Toivanen, P., Beliczynski, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5495, pp. 489–497. [[CrossRef](#)]
33. Kiranyaz, S.; Ince, T.; Pulkkinen, J.; Gabbouj, M.; Ärje, J.; Kärkkäinen, S.; Tirronen, V.; Juhola, M.; Turpeinen, T.; Meissner, K. Classification and retrieval on macroinvertebrate image databases. *Comput. Biol. Med.* **2011**, *41*, 463–472. [[CrossRef](#)] [[PubMed](#)]
34. Joutsijoki, H.; Juhola, M. *Automated Benthic Macroinvertebrate Identification with Decision Acyclic Graph Support Vector Machines*; Presented in Intelligent Systems and Control: Cambridge, UK, 2011. [[CrossRef](#)]
35. Joutsijoki, H.; Juhola, M. DAGSVM vs. DAGKNN: An Experimental Case Study with Benthic Macroinvertebrate Dataset. In *Machine Learning and Data Mining in Pattern Recognition*; Perner, P., Ed.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7376, pp. 439–453. [[CrossRef](#)]
36. Joutsijoki, H.; Juhola, M. Kernel selection in multi-class support vector machines and its consequence to the number of ties in majority voting method. *Artif. Intell. Rev.* **2011**, *40*, 213–230. [[CrossRef](#)]
37. Joutsijoki, H.; Siermala, M.; Juhola, M. Directed acyclic graph support vector machines in classification of benthic macroinvertebrate samples. *Artif. Intell. Rev.* **2014**, *44*, 215–233. [[CrossRef](#)]
38. Céréghino, R.; Park, Y.-S.; Compin, A.; Lek, S. Predicting the species richness of aquatic insects in streams using a limited number of environmental variables. *J. N. Am. Benthol. Soc.* **2003**, *22*, 442–456. [[CrossRef](#)]
39. Paller, M.H.; Martin, F.D.; Wike, L.D.; Specht, W.L. Factors Influencing the Accuracy of a Macroinvertebrate Bioassessment Protocol in South Carolina Coastal Plain Streams. *J. Freshw. Ecol.* **2007**, *22*, 23–32. [[CrossRef](#)]
40. Snyder, C.D.; Hitt, N.P.; Smith, D.R.; Daily, J.P. Evaluating bioassessment designs and decision thresholds using simulation techniques. In *Application of Threshold Concepts in Natural Resource Decision Making*; Springer: New York, NY, USA, 2014; Volume 9781489980410, pp. 157–197. [[CrossRef](#)]
41. Callanan, M.; Baars, J.-R.; Kelly-Quinn, M. Critical influence of seasonal sampling on the ecological quality assessment of small headwater streams. *Hydrobiologia* **2008**, *610*, 245–255. [[CrossRef](#)]
42. Park, Y.-S.; Chon, T.-S.; Kwak, I.-S.; Lek, S. Hierarchical community classification and assessment of aquatic ecosystems using artificial neural networks. *Sci. Total Environ.* **2004**, *327*, 105–122. [[CrossRef](#)]
43. Council of the European Communities. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. *Off. J. Eur. Communities* **2000**, *L327*, 1–72.
44. Di Dato, P.; Mancini, L.; Tancioni, L.; Scardi, M. A neural network approach to the prediction of benthic macroinvertebrate fauna composition in rivers. In *Modelling Community Structure in Freshwater Ecosystems*; Springer: Berlin, Germany, 2005; pp. 147–157. [[CrossRef](#)]
45. Schleiter, I.M.; Obach, M.; Wagner, R.; Werner, H.; Schmidt, H.-H.; Borchardt, D. Modelling ecological interrelations in running water ecosystems with artificial neural networks. In *Ecological Informatics*; Recknagel, F., Ed.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 169–186. [[CrossRef](#)]
46. Kim, B.; Lee, S.-E.; Song, M.-Y.; Choi, J.-H.; Ahn, S.-M.; Lee, K.-S.; Cho, E.; Chon, T.-S.; Koh, S.-C. Implementation of artificial neural networks (ANNs) to analysis of inter-taxa communities of benthic microorganisms and macroinvertebrates in a polluted stream. *Sci. Total Environ.* **2008**, *390*, 262–274. [[CrossRef](#)]
47. Rico, C.; Paredes, M.; Fernández, N. Modeling of the hierarchical structure of freshwater Macroinvertebrates using artificial neural networks. *Acta Biológica Colomb.* **2009**, *14*, 71–96.
48. Carafa, R.; Faggiano, L.; Real, M.; Munné, A.; Ginebreda, A.; Guasch, H.; Flo, M.; Tirapu, L.; von der Ohe, P.C. Water toxicity assessment and spatial pollution patterns identification in a Mediterranean River Basin District. Tools for water management and risk analysis. *Sci. Total Environ.* **2011**, *409*, 4269–4279. [[CrossRef](#)]
49. Pallottini, M.; Goretti, E.; Gaino, E.; Selvaggi, R.; Cappelletti, D.; Céréghino, R. Invertebrate diversity in relation to chemical pollution in an Umbrian stream system (Italy). *Comptes Rendus. Biol.* **2015**, *338*, 511–520. [[CrossRef](#)]
50. Penczak, T.; Kruk, A.; Grzybkowska, M.; Dukowska, M. Patterning of impoundment impact on chironomid assemblages and their environment with use of the self-organizing map (SOM). *Acta Oecologica* **2006**, *30*, 312–321. [[CrossRef](#)]
51. Li, T.; Xiang, R. Application of a self-organizing map and canonical correspondence analysis in modern benthic foraminiferal communities: A case study from the pearl river estuary, China. *J. Foraminifer. Res.* **2015**, *45*, 305–318. [[CrossRef](#)]
52. Dedecker, A.P.; Goethals, P.L.; De Pauw, N. Comparison of Artificial Neural Network (ANN) Model Development Methods for Prediction of Macroinvertebrate Communities in the Zwalm River Basin in Flanders, Belgium. *Sci. World J.* **2002**, *2*, 96–104. [[CrossRef](#)]
53. Joutsijoki, H.; Meissner, K.; Gabbouj, M.; Kiranyaz, S.; Raitoharju, J.; Ärje, J.; Kärkkäinen, S.; Tirronen, V.; Turpeinen, T.; Juhola, M. Evaluating the performance of artificial neural networks for the classification of freshwater benthic macroinvertebrates. *Ecol. Inform.* **2014**, *20*, 1–12. [[CrossRef](#)]

54. Joutsijoki, H.; Juhola, M. Comparing the one-vs-one and one-vs-all methods in benthic macroinvertebrate image classification. In *Machine Learning and Data Mining in Pattern Recognition*; Perner, P., Ed.; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6871, pp. 399–413. [[CrossRef](#)]
55. Joutsijoki, H. Half-against-half multi-class support vector machines in classification of benthic Macroinvertebrate images. In Proceedings of the 2012 International Conference on Computer & Information Science (ICIS), Kuala Lumpur, Malaysia, 12–14 June 2012; pp. 414–419. [[CrossRef](#)]
56. Joutsijoki, H. Half-against-half structure in classification of benthic macroinvertebrate images. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 3646–3649. [[CrossRef](#)]
57. Arje, J.; Karkkainen, S.; Meissner, K.; Turpeinen, T. Statistical classification and proportion estimation—An application to a macroinvertebrate image database. In Proceedings of the 2010 IEEE International Workshop on Machine Learning for Signal Processing, Kittila, Finland, 29 August 2010; pp. 373–378. [[CrossRef](#)]
58. Ärje, J.; Kärkkäinen, S.; Turpeinen, T.; Meissner, K. Breaking the curse of dimensionality in quadratic discriminant analysis models with a novel variant of a Bayes classifier enhances automated taxa identification of freshwater macroinvertebrates: Automated identification of macroinvertebrates. *Environmetrics* **2013**, *24*, 248–259. [[CrossRef](#)]
59. Ärje, J.; Kärkkäinen, S.; Meissner, K.; Iosifidis, A.; Ince, T.; Gabbouj, M.; Kiranyaz, S. The effect of automated taxa identification errors on biological indices. *Expert Syst. Appl.* **2017**, *72*, 108–120. [[CrossRef](#)]
60. Marburg, A.; Bigham, K. Deep learning for benthic fauna identification. In Proceedings of the OCEANS 2016 MTS/IEEE Monterey, OCE 2016, Monterey, CA, USA, 19–23 September 2016. [[CrossRef](#)]
61. Raitoharju, J.; Riabchenko, E.; Meissner, K. Data enrichment in fine-grained classification of aquatic Macroinvertebrates 2017. In Proceedings of the 2016 ICPR 2nd Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI), Cancun, Mexico, 4 December 2016; pp. 43–48. [[CrossRef](#)]
62. Raitoharju, J.; Riabchenko, E.; Ahmad, I.; Iosifidis, A.; Gabbouj, M.; Kiranyaz, S.; Tirronen, V.; Ärje, J.; Kärkkäinen, S.; Meissner, K. Benchmark database for fine-grained image classification of benthic macroinvertebrates. *Image Vis. Comput.* **2018**, *78*, 73–83. [[CrossRef](#)]
63. Ärje, J.; Raitoharju, J. Human experts vs. machines in taxa recognition. *Signal Processing: Image Commun.* **2019**, *87*, 115917. [[CrossRef](#)]
64. Singh, R.; Mumbarekar, V. Neural network model approach for automated benthic animal identification. *ICT Express* **2021**. [[CrossRef](#)]
65. Riabchenko, E.; Meissner, K. Learned vs. engineered features for fine-grained classification of aquatic macroinvertebrates. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 2276–2281. [[CrossRef](#)]
66. Lürig, M.D.; Donoughe, S.; Svensson, E.I.; Porto, A.; Tsuboi, M. Computer Vision, Machine Learning, and the Promise of Phenomics in Ecology and Evolutionary Biology. *Front. Ecol. Evol.* **2021**, *9*, 642774. [[CrossRef](#)]
67. Olawoyin, R.; Nieto, A.; Grayson, R.L.; Hardisty, F.; Oyewole, S. Application of artificial neural network (ANN)–self-organizing map (SOM) for the categorization of water, soil and sediment quality in petrochemical regions. *Expert Syst. Appl.* **2013**, *40*, 3634–3648. [[CrossRef](#)]
68. Song, M.-Y.; Hwang, H.-J.; Kwak, I.-S.; Ji, C.W.; Oh, Y.-N.; Youn, B.J.; Chon, T.-S. Self-organizing mapping of benthic macroinvertebrate communities implemented to community assessment and water quality evaluation. *Ecol. Model.* **2007**, *203*, 18–25. [[CrossRef](#)]
69. D’Heygere, T.; Goethals, P.L.; De Pauw, N. Use of genetic algorithms to select input variables in decision tree models for the prediction of benthic macroinvertebrates. *Ecol. Model.* **2003**, *160*, 291–300. [[CrossRef](#)]
70. D’Heygere, T.; Goethals, P.L.; De Pauw, N. Genetic algorithms for optimisation of predictive ecosystems models based on decision trees and neural networks. *Ecol. Model.* **2006**, *195*, 20–29. [[CrossRef](#)]
71. Gobeyn, S.; Goethals, P.L.M. A variable length chromosome genetic algorithm approach to identify species distribution models useful for freshwater ecosystem management. *IFIP Adv. Inf. Commun. Technol.* **2017**, *507*, 208. [[CrossRef](#)]
72. Goethals, P.; Gasparyan, K.; De Pauw, N. River restoration simulations by ecosystem models predicting aquatic macroinvertebrate communities based on J48 classification trees. *Meded. Rijksuniv. Te Gent Fak. Van Landbouwk. Toegep. Biol. Wet.* **2001**, *66*, 213–217.
73. Goethals, P.; Dedeker, A.; Gabriels, W.; de Pauw, N. Development and application of predictive river ecosystem models based on classification trees and artificial neural networks. In *Ecological Informatics*; Recknagel, F., Ed.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 151–167. [[CrossRef](#)]
74. Li, F.; Yan, Y.; Zhang, J.; Zhang, Q.; Niu, J. Taxonomic, functional, and phylogenetic beta diversity in the Inner Mongolia grassland. *Glob. Ecol. Conserv.* **2021**, *28*, e01634. [[CrossRef](#)]
75. Raitoharju, J. FIN-Benthic. Finnish Environment Institute. 2021. Available online: <http://urn.fi/urn:nbn:fi:csc-kata20170615164351941516> (accessed on 13 May 2022).
76. Raitoharju, J. FIN-Benthic2. Jenni Raitoharju, Finland. June 2019. Available online: <https://etsin.fairdata.fi/dataset/a11cdc26-b9d0-4af1-9285-803d65a696a3> (accessed on 13 May 2022).

77. Bertin, E.; Marcelpoil, R.; Chassery, J.-M. Morphological algorithms based on Voronoi and Delaunay graphs: Microscopic and medical applications. In Proceedings of the Image Algebra and Morphological Image Processing III, San Diego, CA, USA, 1 June 1992; pp. 356–367. [[CrossRef](#)]
78. van Tonder, G.; Ejima, Y. The patchwork engine: Image segmentation from shape symmetries. *Neural Networks* **2000**, *13*, 291–303. [[CrossRef](#)]
79. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour Detection and Hierarchical Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 898–916. [[CrossRef](#)] [[PubMed](#)]
80. Angermueller, C.; Pärnamaa, T.; Parts, L.; Stegle, O. Deep learning for computational biology. *Mol. Syst. Biol.* **2016**, *12*, 878. [[CrossRef](#)] [[PubMed](#)]
81. Doberstein, C.P.; Karr, J.R.; Conquest, L.L. The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. *Freshw. Biol.* **2000**, *44*, 355–371. [[CrossRef](#)]
82. Schleiter, I.M.; Obach, M.; Borchardt, D.; Werner, H. Bioindication of chemical and hydromorphological habitat characteristics with benthic macro-invertebrates based on Artificial Neural Networks. *Aquat. Ecol.* **2001**, *35*, 147–158. [[CrossRef](#)]
83. Céréghino, R.; Giraudel, J.; Compin, A. Spatial analysis of stream invertebrates distribution in the Adour-Garonne drainage basin (France), using Kohonen self organizing maps. *Ecol. Model.* **2001**, *146*, 167–180. [[CrossRef](#)]
84. Dedecker, A.P.; Goethals, P.L.M.; D’Heygere, T.; Gevrey, M.; Lek, S.; De Pauw, N. Application Of Artificial Neural Network Models To Analyse The Relationships Between Gammarus pulex L. (Crustacea, Amphipoda) And River Characteristics. *Environ. Monit. Assess.* **2005**, *111*, 223–241. [[CrossRef](#)]
85. Bernardini, V.; Solimini, A.G.; Carchini, G. Application of an image analysis system to the determination of biomass (ash free dry weight) of pond macroinvertebrates. *Hydrobiologia* **2000**, *439*, 179–182. [[CrossRef](#)]
86. Kiranyaz, S.; Gabbouj, M.; Pulkkinen, J.; Ince, T.; Meissner, K. Network of evolutionary binary classifiers for classification and retrieval in macroinvertebrate database. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 2257–2260. [[CrossRef](#)]
87. Joutsijoki, H.; Juhola, M. A comparison of classification methods in automated taxa identification of benthic macroinvertebrates. *Int. J. Data Sci.* **2017**, *2*, 273. [[CrossRef](#)]
88. Ministry of Environment, Housing and Territorial Development. National Policy for the Integrated Management of the Water Resource. 2010. Available online: [https://www.minambiente.gov.co/images/GestionIntegraldelRecursoHidrico/pdf/Presentaci%C3%B3n\\_Pol%C3%ADtica\\_Nacional\\_-\\_Gesti%C3%B3n\\_libro\\_pol\\_nal\\_rec\\_hidrico.pdf](https://www.minambiente.gov.co/images/GestionIntegraldelRecursoHidrico/pdf/Presentaci%C3%B3n_Pol%C3%ADtica_Nacional_-_Gesti%C3%B3n_libro_pol_nal_rec_hidrico.pdf) (accessed on 22 August 2020).
89. McAbendroth, L.; Ramsay, P.M.; Foggo, A.; Rundle, S.D.; Bilton, D.T. Does macrophyte fractal complexity drive invertebrate diversity, biomass and body size distributions? *Oikos* **2005**, *111*, 279–290. [[CrossRef](#)]
90. Nahmani, J.; Capowicz, Y.; Lavelle, P. Effects of metal pollution on soil macroinvertebrate burrow systems. *Biol. Fertil. Soils* **2005**, *42*, 31–39. [[CrossRef](#)]
91. Brink, P.J.V.D.; Roelsma, J.; van Nes, E.H.; Scheffer, M.; Brock, T.C.M. Perpest model, a case-based reasoning approach to predict ecological risks of pesticides. *Environ. Toxicol. Chem.* **2002**, *21*, 2500–2506. [[CrossRef](#)]
92. Joutsijoki, H. An application of one-vs-one method in automated taxa identification of Macroinvertebrates. In Proceedings of the 2013 Fourth Global Congress on Intelligent Systems, Hong Kong, China, 13 December 2013; pp. 125–130. [[CrossRef](#)]
93. Chon, T.-S.; Kwak, I.-S.; Park, Y.-S.; Kim, T.-H.; Kim, Y. Patterning and short-term predictions of benthic macroinvertebrate community dynamics by using a recurrent artificial neural network. *Ecol. Model.* **2001**, *146*, 181–193. [[CrossRef](#)]
94. Bizzi, S.; SurrIDGE, B.; Lerner, D.N. *Analysing Biological, Chemical and Geomorphological Interactions in Rivers Using Structural Equation Modelling*; University of Western: Crawley, Australia, 2009; pp. 1837–1843.
95. Pinto, P.; Rosado, J.; Morais, M.; Antunes, I. Assessment Methodology for Southern Siliceous Basins in Portugal. In *Integrated Assessment of Running Waters in Europe*; Hering, D., Verdonschot, P.F.M., Moog, O., Sandin, L., Eds.; Springer: Dordrecht, The Netherlands, 2004; pp. 191–214. [[CrossRef](#)]
96. Dedecker, A.P.; Goethals, P.L.; Gabriels, W.; De Pauw, N. Optimization of Artificial Neural Network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). *Ecol. Model.* **2004**, *174*, 161–173. [[CrossRef](#)]
97. Mouton, A.M.; Dedecker, A.P.; Lek, S.; Goethals, P.L.M. Selecting Variables for Habitat Suitability of Asellus (Crustacea, Isopoda) by Applying Input Variable Contribution Methods to Artificial Neural Network Models. *Environ. Model. Assess.* **2009**, *15*, 65–79. [[CrossRef](#)]
98. Dedecker, A.P.; Goethals, P.L.M.; D’Heygere, T.; De Pauw, N. Development of an in-stream migration model for Gammarus pulex L. (Crustacea, Amphipoda) as a tool in river restoration management. *Aquat. Ecol.* **2005**, *40*, 249–261. [[CrossRef](#)]
99. Zhang, M.; Muñoz-Mas, R.; Martínez-Capel, F.; Qu, X.; Zhang, H.; Peng, W.; Liu, X. Determining the macroinvertebrate community indicators and relevant environmental predictors of the Hun-Tai River Basin (Northeast China): A study based on community patterning. *Sci. Total Environ.* **2018**, *634*, 749–759. [[CrossRef](#)]
100. Chen, Q.; Yang, Q.; Lin, Y. Development and application of a hybrid model to analyze spatial distribution of macroinvertebrates under flow regulation in the Lijiang River. *Ecol. Inform.* **2011**, *6*, 407–413. [[CrossRef](#)]
101. Zhang, Y.; Leung, J.Y.; Zhang, Y.; Cai, Y.; Zhang, Z.; Li, K. Agricultural activities compromise ecosystem health and functioning of rivers: Insights from multivariate and multimetric analyses of macroinvertebrate assemblages. *Environ. Pollut.* **2021**, *275*, 116655. [[CrossRef](#)]

102. Dou, Q.; Du, X.; Cong, Y.; Wang, L.; Zhao, C.; Song, D.; Liu, H.; Huo, T. Influence of environmental variables on macroinvertebrate community structure in Lianhuan Lake. *Ecol. Evol.* **2022**, *12*, e8553. [[CrossRef](#)]
103. Goethals, P.L.M.; Dedecker, A.P.; Gabriels, W.; Lek, S.; De Pauw, N. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. *Aquat. Ecol.* **2007**, *41*, 491–508. [[CrossRef](#)]
104. Gutiérrez-Estrada, J.C.; Bilton, D.T. A heuristic approach to predicting water beetle diversity in temporary and fluctuating waters. *Ecol. Model.* **2010**, *221*, 1451–1462. [[CrossRef](#)]