



OPEN

# Habitat selection ecology of the aquatic beetle community using explainable machine learning

Marek Kruk<sup>1</sup> & Joanna Pakulnicka<sup>2</sup>

The aim of our work is to determine the importance of habitat features for the selection of the aquatic beetle community. Insects are represented by their general ecological traits such as body size, ecological element and trophic type, which are categorised into four body size ranges, four ecological groups and four trophic types. To determine the importance of habitat selection of the studied insects, we analysed the relationships between the above categories and the set of habitat features of the lake and its surroundings. Ensemble machine learning modelling (XGBoost-SHAP) revealed the mechanism of habitat feature selection in relation to the general ecological traits. We found strong interactions between the body size, ecological element and trophic type of beetles, suggesting that these general traits control the structure and functioning of the beetle community studied. The area of the lake and the features of beetle occurrence in the aquatic environment play an important but secondary role, and the importance of the characteristics of the lake's riparian zone was minimised. We found several categories of beetles as they select the number of the same habitat features. The study can provide valuable information for the practical conservation and management of lake ecosystems.

Aquatic beetles are a group of organisms that play an important role in aquatic ecosystems thanks to their biodiversity, ecological and functional (trophic) diversity. In this paper we will analyse general ecological factors that are important for the occurrence of these organisms in relation to their habitat selection. The basic ecological traits of insect communities are assessed using general characteristics that constitute this species community, primarily body size<sup>1,2</sup>, including body size and trophic type<sup>3,4</sup>, ecological element<sup>5</sup> or body size and ecological element or environment more broadly<sup>6,7</sup>. Each of the above traits can be divided into several categories covering a wide range of aquatic beetle occurrences and environmental adaptations.

Aquatic beetles are sensitive to environmental factors and show clear habitat selection<sup>8–14</sup>. They form animal communities that change as a result of environmental changes that are primarily due to human pressure. This mainly concerns changes in habitat features in a landscape rich in inland waters and the associated systems of aquatic and terrestrial ecotones. Habitat selection by aquatic beetle communities appears to be important for their abundance, richness, species and ecological structure<sup>15</sup>, may influence phylogeny<sup>16</sup> or correlate with taxonomic structure at the regional level<sup>7,17</sup>.

In the context of the cited works, we analysed the habitat selection of a regional community of aquatic beetles characterised by the basic categories of the above-mentioned general ecological traits. This analysis is based on the machine learning models and uses an extensive database from many years of research in the Olsztyn Lake District (northern Poland). The question of habitat selection of beetles in relation to their general ecological traits has not been found in the literature so far. Worth noticing, that machine learning methods have rarely been used in studies on aquatic beetle biodiversity<sup>18–21</sup>. The study of habitat selection by aquatic beetles is related to the broader context of applying the latest numerical methods of habitat selection analysis<sup>22</sup>.

The aim of this work is to analyse the relationships of general ecological traits of aquatic beetles, such as body size, ecological element and trophic type, as determinants of the species' occurrence in the environment. This analysis included the full range of categories consisting of the above-mentioned traits. Thus, four categories of beetle body size were analysed, the ecological element categories of typhophiles, lake and river species, argilophiles and eurytopes, and the trophic types of shredder, scraper and grazer, polyphages and predators. To determine the importance of the habitat selection of the insects studied, we simultaneously examined the relationships of these categories of general features with habitat features in a general sense, such as lake developmental stage, lake surface, fractal nature of the ecotone or type of reservoir surroundings, as well as with

<sup>1</sup>Chair of Applied Computer Science and Mathematical Modelling, University of Warmia and Mazury in Olsztyn, Śloneczna 54, 10–719 Olsztyn, Poland. <sup>2</sup>Department of Zoology, University of Warmia and Mazury in Olsztyn, Łódzki Sq. 3, 10–727 Olsztyn, Poland. email: mkruk@uwm.edu.pl

more specific habitat features, such as type of substrate, water pH, depth of occurrence, associated vegetation and others. These habitat features were categorised into 2 – 6 classes.

We have made a number of assumptions to fit the highly diverse reality of how aquatic beetle communities function in the mosaic of lakeland habitats. It is worth noting that these insects have considerable dispersal ability<sup>2,23</sup>), which allows them to move between land habitats and water reservoirs<sup>24–26</sup>). To follow this background, we assumed the categories of ecological traits of aquatic beetle species as a dynamic, multidimensional system of interactions between species in the biocenosis<sup>27</sup>. This provides a conceptual basis for the application of the boosting modelling technique<sup>28,29</sup>. Furthermore, we assumed the concept of free choice defined by Morris<sup>30</sup> and thus the deliberate choice of discretised habitat patches<sup>31</sup>. These accepted assumptions implies the use of the most appropriate data mining tool to mimic this natural process. Therefore, the modelling in our work is based on the multiple choice principle<sup>32</sup> and boosting method. This technique represented in our study by The eXtremeGradientBoosting (XGBoost) algorithm is based on the random selection of interactions between variables and utilises the boosting of individual variables (so-called weak learners). The aim is to obtain a model with the highest accuracy that mimics the interactions between species in the biocenosis, which counteracts the tendency to lose coherence and survival<sup>33</sup>.

To reinforce an element of interactive play between the beetle species and to visualise and interpret the model prediction results obtained, the authors added SHapley Additive exPlanations (SHAP) modelling to the boosting model. In the ensemble XGBoost-SHAP modelling, the authors assumed the existence of an adaptive game in which populations of beetle species belonging to different categories of general ecological traits participate as trophic or competing guilds<sup>2</sup>. This approach to species interactions covers the full spectrum of relationships, such as competition for habitat, including response to population density<sup>31,34</sup>, trophic relationships as well as non-antagonistic food guilds and mutualism<sup>35</sup>. A formal description of the functioning of such a system can be found in mathematical game theory and in particular in the theory of corporate game<sup>36</sup>. This approach consists of measuring the weight of each sample representing the habitat features of the sampled beetle species based on the modelling result with and without this sample. Thanks to this method, we get an insight into the interactions of the beetle species and at the same time valorise this interactivity. The authors then used SHAP modelling<sup>37</sup> to model the responses of aquatic beetle categories to habitat features, taking into account the effects of the specific interactions described above. This modelling based on binary classification and valorisation method “model with sample and without it” makes it possible to show the importances of features expressed by the Shapley value in a classification model that is built from the bottom up from the values of individual cases or observations<sup>38</sup>. The SHAP algorithm was used to calculate the effects of habitat variables on the analysed categories of body size, ecological element and trophic type. In authors intention, application the explainable machine learning tool fulfilled the requirement to use the modelling as close as possible to ecological reality<sup>39</sup>.

We investigated the importance of habitat selection of aquatic beetles in relation to their general ecological traits such as body size, ecological element and trophic type. Namely, whether this selection is more or less related to category as particular body size, e.g. > 10 mm, ecological element, e.g. argilophiles, and trophic type, e.g. predator. In the context of such a comprehensive approach to the problem of habitat selection by water beetles, we formulated a series of hypotheses that, according to the authors, could be tested thanks to the incorporation of advanced machine learning models.

We formulated hypothesis 1. that habitat features related to the aquatic and coastal environments might be less important to the beetle community than their affiliation to categories of general ecological traits. To test this hypothesis, we used a two-stage modelling approach based in the first stage on the involvement of categories of the above-mentioned ecological traits in the habitat feature selection decision process (XGBoost) and in the second stage on modelling based on the individual assessment of the impact of beetle species samples on the model (SHAP). Here we refer to the idea of Kalinkata et al. who pointed out that general traits such as body size and trophic interactions that regulate energy and nutrient flow within and between ecosystems are underestimated in ecological studies on beetles<sup>40</sup>.

Our hypotheses regarding the phenomenon of habitat selection by the studied insects are as follows. 2. Among aquatic habitat features, lake area should play an important role<sup>41</sup>, both as a feature selected for its larger and smaller surface area. 3. Detailed aquatic habitat features such as water pH, detritus content, depth of occurrence or bottom character are also important features. 4. For aquatic beetles, the group of habitat features associated with the coastal zone and aquatic vegetation does not play a major role compared to aquatic habitats. 5. It will be possible to find categories of beetles that select the same features of aquatic habitats in different general ecological traits which can be a premise to find ecological guilds<sup>42</sup> of aquatic beetles.

The study aims to contribute to a broader discussion on the usefulness of machine learning feature importance metrics as indicators of organisms' habitat selection.

## Results

### Classification of general ecological traits and habitat features

Three basic traits of beetles were selected for this study: body size, ecological element and trophic type. Each of these traits was divided into four categories (Table 1). In order to find significant habitat features influencing the above general traits of aquatic beetles, a number of discrete and continuous variables were analysed, described and classified (Table 1).

The numbers in the column “Categories (1. – 4.) and classes (1. – 6.)” in Table 1 assigned to the individual categories of ecological traits and classes of habitat features are also the nominal values of the respective category and class. They are then modelled to determine their negative (low numbers) or positive (high numbers) effects on the category of aquatic beetles under study.

General characteristics	Ecological general traits and habitat features	Data numeric character	Categories (1. – 4.) and classes (1. – 6.)	Remarks
Ecological general traits	Body size	discrete	Categories	1. 1 – 1.99 mm, 2. 2 – 4.99 mm, 3. 5 – 9.99 mm, 4. > 10 mm
	Ecological element	discrete		1. typhophiles, 2. lakes and river species and rheobionts, 3. argilophiles, 4. eurytopes
	Trophic type	discrete		1. shredders, 2. grazers and scrapers, 3. polyphages, 4. predators
Habitat general features	lake area (ha)	continuous	Classes	Mean.67 ha
	lake stadium	discrete		1. mesotrophic- $\alpha$ , 2. mesotrophic- $\beta$ , 3. eutrophic, 4. polytrophic, 5. polyhumic, 6. oligohumic
	habitat fractal	discrete		1. ecotone, 2. rare macrophytes, 3. dense macrophytes, 4. <i>Sphagnum</i> mat
	surroundings	discrete		1. open, 2. forested
Habitat detail features	sampling depth	discrete	Classes	1. < 10 cm, 2. 10–20 cm, 3. 30–40 cm, 4. > 40 cm
	water pH	continuous		Mean 7.67
	bottom	discrete		Interpretation for great lakes (> 10 ha) – 4., small lakes (< 10 ha) – 5
	detritus	discrete		1. no detritus, 2. fine, 3. fine and abundant, 4. coarse and abundant
	influxes	discrete		1. absent, 2. present
Vegetation habitat features	shore	discrete	Classes	1. steep, 2. flat
	grass	discrete		1. absent, 2. present
	elodeids	discrete		1. absent, 2. present
	nymphoids	discrete		1. absent, 2. present
	macrophytes	discrete		1. absent, 2. present

**Table 1.** General ecological traits and habitat features of aquatic beetles. Nomenclature of traits, categories, features and classes will be used in visualizations of model results.

### Features important for aquatic beetles body size

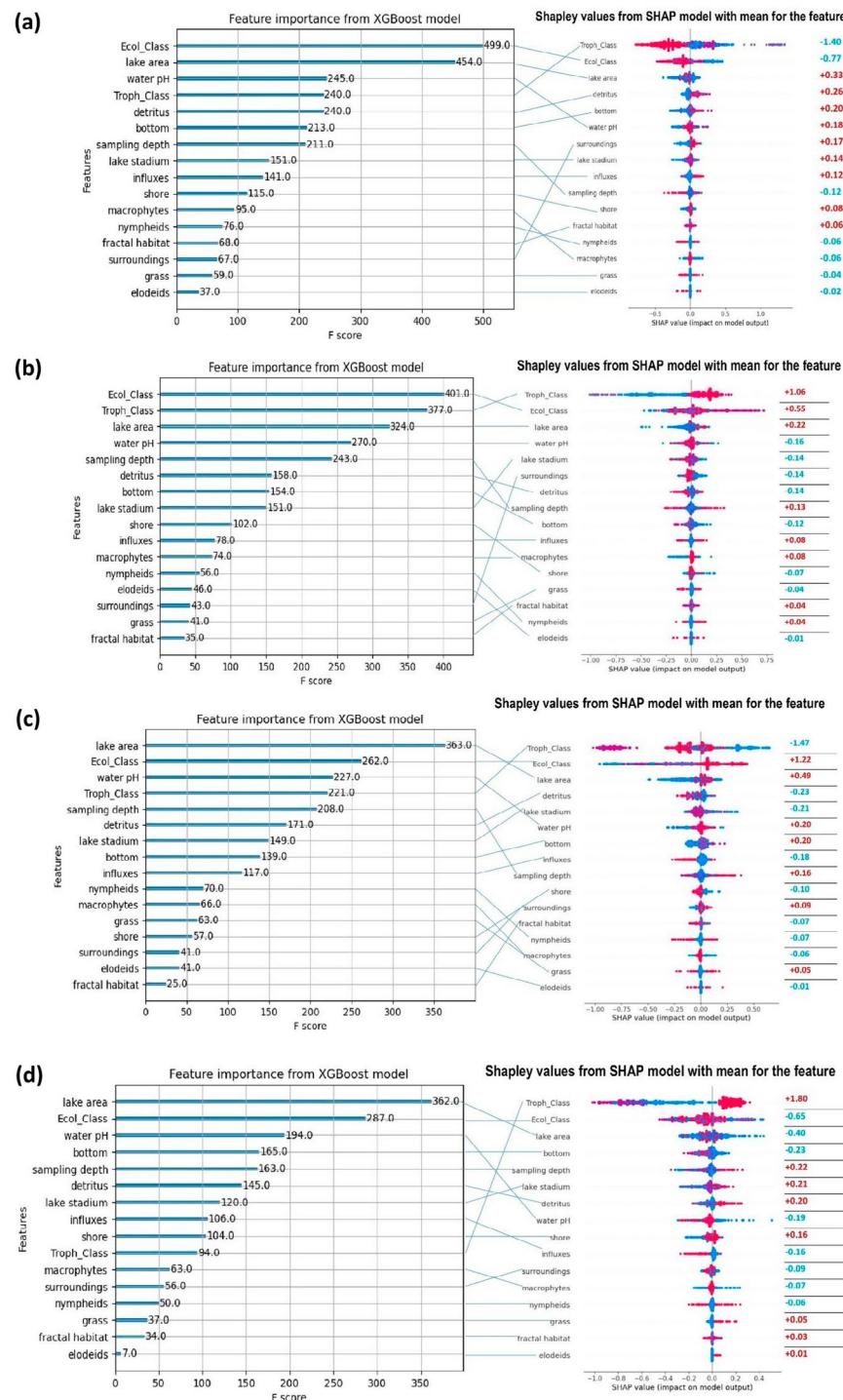
We analysed how the importance of individual general ecological traits and habitat features was distributed in the models for each of the four body size classes of aquatic beetles (see Table 1 above and Fig. 5 in Material and Methods). In this case, we extended the analysis of trait importance by predicting that a particular feature would relate more or less to the body size class of the insects studied.

Thus, for the smallest beetles (1–2.99 mm) it could be shown that the most important characteristic is their general property of ecological preference, i.e. the ecological element, while belonging to the eurytopes and argilophiles does not strengthen this category of body size. This feature is possessed by representatives of typhophiles and lake and river species (Fig. 1a). A characteristic feature of the class of these smallest beetles is also the lake surface area. Here, the larger surface area of the lakes is a feature that valorises this class in contrast to lakes with a smaller surface area. In addition to other more important habitat features, the smallest body size of the water beetles is favoured by a higher pH value of the water, the presence of detritus in it, a mud-free bottom and a shallower depth of occurrence (Fig. 1a). By comparing the place in the ranking of importance of traits measured in the XGBoost and SHAP sub-models, a clear positive influence of interspecific interaction is visible in the case of the trait trophic type, with shredders, grazers and scrapers (nominal values 1 and 2) being the categories that strengthen, by positive Shapley values, this trait (Fig. 1a). Species interactions have also radically changed the importance of the surroundings of the lake (Fig. 1a, Table 1).

For the next class of beetle body size, 3–4.99 mm, the XGBoost model distinguishes two general ecological factors, i.e. the ecological element (Ecol\_Class) and the trophic type (Troph\_Class), which play the most important role in predicting this category. It is more favoured by eurytopes and predators (nominal value for both 4 and positive Shapley value) than by typhophiles and shredders (nominal value 1) (Fig. 1b). As with the previous body size class, lakes with a larger area than those with a smaller area are a reinforcing feature. Of note is the significant increase in the importance of surroundings for this body size category in the SHAP model, although the strongest feature here is the unwooded surroundings (Fig. 1b, Table 1).

The body size class 5–9.99 mm has a distinct and important feature, namely lake size, and as in the previous classes it is reinforced by larger lakes. The importance of the ecological element is also the same as in the previous class, with argilophilous and eurytopes species being favoured. The trophic type category has gained the most importance due to the consideration of interspecific interactions, with the lower trophic levels of shredders, grazers and scrapers playing a greater role here. Aquatic beetles in this body size class favour several lake habitat features (Fig. 1c, Table 1).

The highest class of beetle body size (> 10 mm) is still dominated by the importance of the habitat feature lake area and the general trait ecological element. However, in contrast to the previous categories, in this class the strongest feature in terms of lake area is the smaller surface area of lakes, and in the case of the ecological element, the beetles with the largest body size are reinforced by typhophiles and lake and river species. The presence of beetles with the largest body sizes can also be enhanced by lower water pH and other lake



**Fig. 1.** Importance for body size, F-score of the XGBoost model, and mean Shapley values of habitat features and general ecological traits, ecological element (Ecol\_Class) and trophic type (Troph\_Class) for the differentiation of body size of aquatic beetles: (a) body size 1.0 – 2.99 mm, (b) 3.0 – 4.99 mm, (c) 5.0 – 9.99 mm, (d) > 10.0 mm. The Shapley values indicate the predictions of the mean positive (+) or negative (–) influence of the respective trait on the categories (a) – (d) of the body size of the beetles. Positive values (red print) strengthen a particular category compared to the other three. Negative values (blue print) weaken this category at the expense of strengthening the other categories. Red dots in the SHAP model diagram indicate higher values of the feature, while blue dots indicate lower values of the feature. For ecological element (Ecol\_Class) nominal values of categories: 1. typhophiles, 2. lakes and river species and rheobionts, 3. argilophiles, 4. eurytopes. For trophic type (Troph\_Class) nominal values of categories: 1. shredders, 2. grazers and scrapers, 3. polyphages, 4. predators. F-scores and mean Shapley values are averages of 5 random repetitions of XGBoost and SHAP models. Accuracy of the train and test subsets of the model, see Appendix S1: Table S2.

characteristics (Fig. 1d). Regarding the effect of the relationship between beetle species, there is a clear change in the importance of trophic type (from an intermediate feature to the most important for the category in question) and a clear decrease in the importance of water pH (Fig. 1d, Table 1).

### Features important for aquatic beetles ecological element

For beetles belonging to the ecological element typhophiles, body size is the most important feature according to the XGBoost model, and this element is reinforced by smaller individuals. An almost equally important feature is lake area, which includes smaller lakes. The third feature that is important for typhophiles is trophic type, and this feature is enhanced by shredders rather than predators. At the same time, typhophiles seem to favour shallower water and its lower pH (Fig. 2a, Table 1). Typhophiles show the greatest increase in importance due to interspecific interactions in the case of surroundings and the presence of macrophytes. They favour a forest surrounding and the absence of macrophytes (Fig. 2a, Table 1).

In the case of the ecological element lake and river species including rheobionts, body size, lake area and trophic class dominate the validity of the prediction. In contrast to the previous element, the strengthening of the characteristic here is determined by the larger body sizes of the beetles and the larger area of the lakes. Predators are most important. This ecological element is reinforced by habitat features such as higher detritus content, higher pH, shallower depth (Fig. 2b, Table 1).

For argilophiles, lake area is the most important feature, and this feature is enhanced by a larger surface. The trait of body size also stands out, and here it is reinforced by individuals with a smaller body. The trophic type feature was ranked relatively low, but it is clearly reinforced by lower levels, i.e. mainly by shredders. For the habitat variables, the discussed element is enhanced by its deeper occurrence in the water column and high pH (Fig. 2c, Table 1). For the category of argilophiles, two important features should be noted, the importance of which has increased significantly under the influence of species interactions: the trophic type and the nature of the shore (Fig. 2c).

The eurytopic element is characterised by an almost equal importance of the three features. Body size, lake area and trophic type have similar F-measures in the model and these features are strengthened for eurytopes when larger beetle sizes, larger lake areas and more trophic types represented by polyphages and predators dominate. In addition, the category of the eurytopes element is enhanced by its occurrence at greater depth and higher water pH (Fig. 2d, Table 1).

### Features important for aquatic beetles trophic types

The shredder category of the aquatic beetle community studied is strengthened primarily by smaller body sizes, belonging to eurytopes and argilophiles, and by the smaller lake area. Other habitat features important for this category include lower pH, shallower depth and muddier bottom (Fig. 3a, Table 1).

The water beetles in the grazers and scrapers category are eurytopes that favour much larger lakes and coarse-grained bottom without detritus. The trait that showed a very significant increase in interactions between beetle species in this category proved to be body size (reinforced by small size) (Fig. 3b, Table 1).

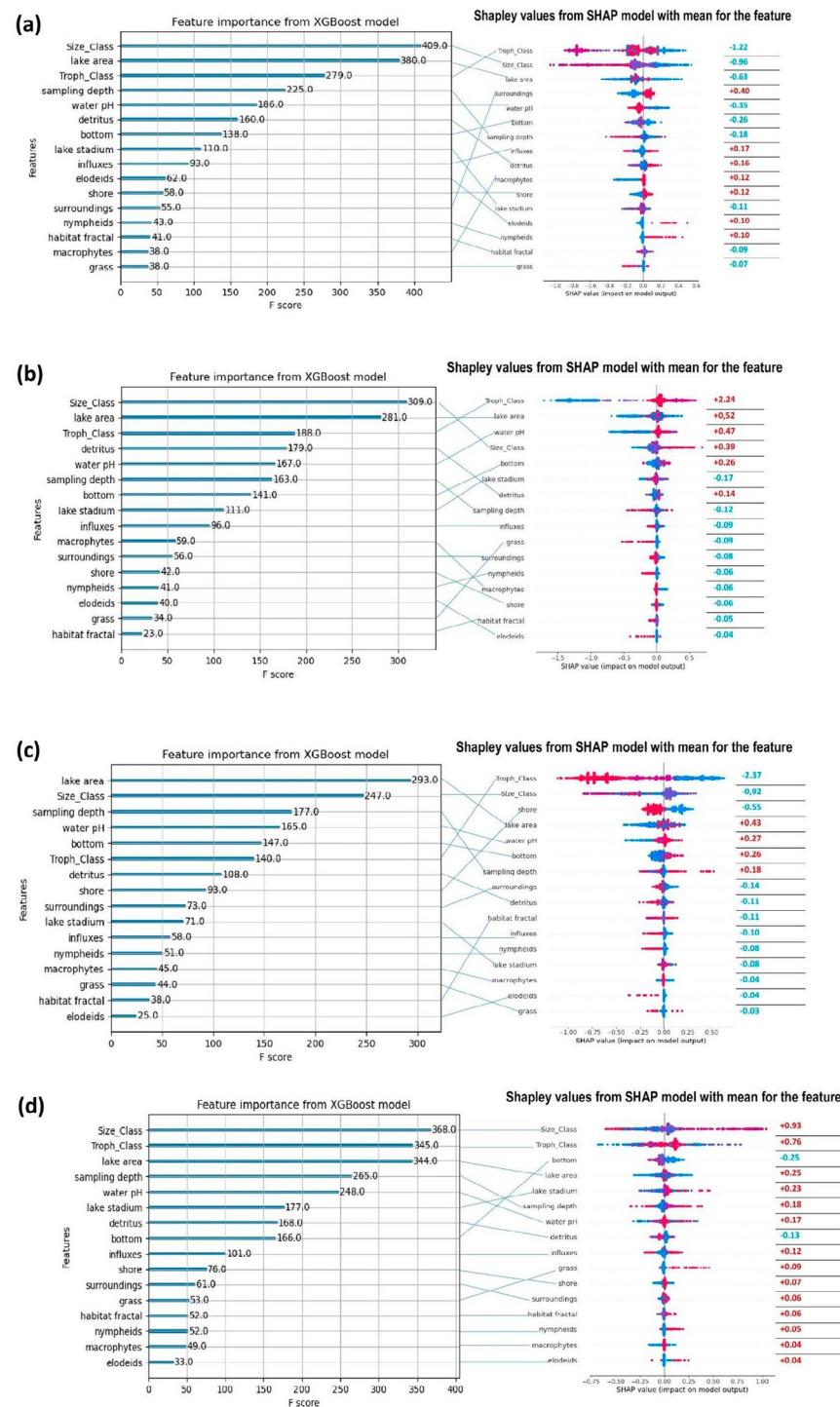
Polyphagous beetles are represented by argilophiles, eurytopes and lake and river species including rheobionts. They generally have a smaller body size and favour larger lakes. This trophic type is favoured by habitat features such as sandy and muddy bottoms, shallower depths and lack of detritus (Fig. 3c, Table 1). For species of the polyphages category in the SHAP model, the importance of water pH and the presence of influxes, increased significantly (Fig. 3c).

When it comes to the highest trophic level of predators, eurytopes are the favoured ecological element, especially from the category of largest body size. Among the smaller beetles, the importance of other ecological elements increases: typhophiles and lake and river species (Fig. 3d). An important reinforcing feature, apart from body size, is the smaller surface area of the lakes. At the same time, among the habitat features that strengthen the discussed trophic type, the following features are important: lower pH of the water and shallower depth of occurrence in the water (Fig. 3d, Table 1).

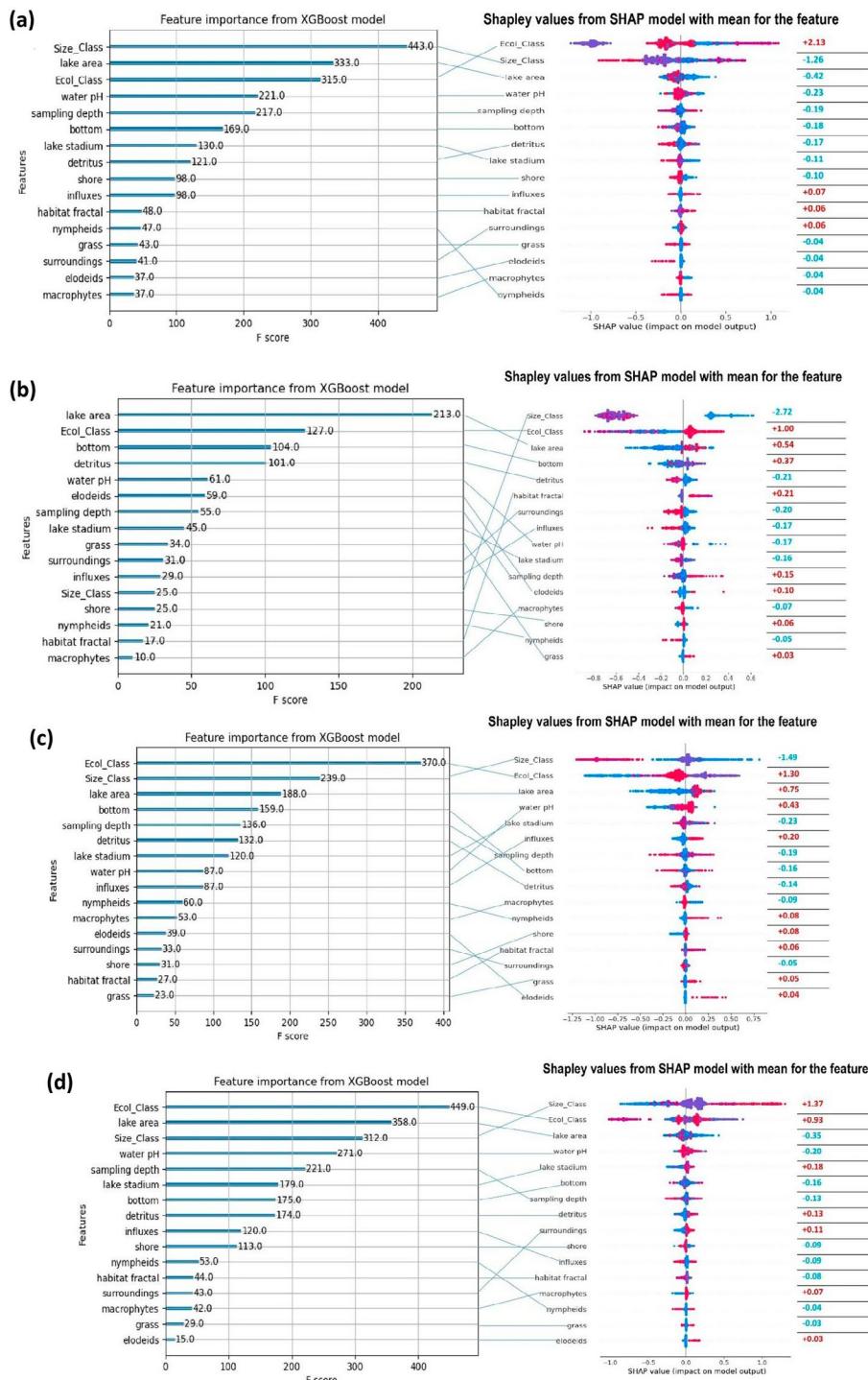
### Categories of aquatic beetles select habitat features – Synthesis

The “heat map” of average Shapley values shown in Fig. 4 was created from three analytical diagrams showing the ranking of importance of habitat features and ecological categories (Figs. 1, 2, 3). The average Shapley values calculated for ecological categories and habitat features are the results of the entire XGBoost-SHAP modelling procedure, which is described in detail in Materials and methods section. The synthetic representation of the modelling results (Fig. 4) allows a comparison of the importance of habitat features for different categories of the three general ecological traits, including the search for those habitat features that are common in terms of direction (+ or -) of impact on different ecological categories of beetles belonging to different general ecological traits. In this comparison, we did not consider habitat features that had a very low level of average Shapley values (<0.10), i.e. marginal importance to a particular category of beetle species.

Large beetles (> 10 mm), which are typhophiles and belong to the shredders or predators, opt for smaller lakes, water with a lower pH value and a muddy bottom (black frames in Fig. 4, see also Table 1). Beetles with a body size of 5–10 mm, which are also argilophiles, choose larger lakes with a higher pH, with a gravelly bottom, without detritus, with influxes and with steep banks (red frames in Fig. 4, see also Table 1). However, this last feature seems to be dominant for this ecological group of argilophiles (mean Shapley value –0.55). In turn, the smallest beetle species (1–2 mm) with a eurytopic character, which feed polyphagously, are associated with the selection of a larger surface area of lakes, a higher pH of their water and the presence of influxes (green frames in Fig. 4, see also Table 1). Beetle species with medium body size (categories 2–5 and 5–10 mm) may belong to the argilophiles or eurytopes and represent the trophic type of grazers and scrapers, and together they select larger lakes and a similar depth of occurrence (exactly sampling) of more than 30–40 cm (blue frames in Fig. 4, see also



**Fig. 2.** Importance for ecological element, F-score of the XGBoost model, and mean Shapley values of habitat features and general ecological traits, body size (Size\_Class) and trophic type (Troph\_Class) for the differentiation of the ecological element of aquatic beetles community: **(a)** typhophiles, **(b)** lake and river species incl. rheobionts, **(c)** argilophiles, **(d)** eurytopes. The Shapley values indicate the predictions of the mean positive (+) or negative (-) influence of the respective trait on the categories **(a) – (d)** of the ecological element of the beetles. Positive values (red print) strengthen a particular category compared to the other three. Negative values (blue print) weaken this category at the expense of strengthening the other categories. Red dots in the SHAP model diagram indicate higher values of the feature, while blue dots indicate lower values of the feature. For general trait body size (Size\_Class) nominal values of categories: 1. 1 – 1.99 mm, 2. 2 – 4.99 mm, 3. 5 – 9.99 mm, 4. > 10 mm. For trophic type (Troph\_Class) nominal values of categories: 1. shredders, 2. grazers and scrapers, 3. polyphages, 4. Predators. F-scores and mean Shapley values are averages of 5 random repetitions of XGBoost and SHAP models. Accuracy of the train and test subsets of the model, see Appendix S1: Table S2.



**Fig. 3.** Importance for trophic type, F-score of the XGBoost model, and means of Shapley values of habitat features and general ecological traits, body size (Size\_Class) and ecological element (Ecol\_Class) for distinguishing the trophic type of aquatic beetles: (a) shredders, (b) grazers and scrapers, (c) polyphages, (d) predators. The Shapley values indicate the predictions of the mean positive (+) or negative (-) influence of the respective trait on the categories (a) – (d) of the trophic type of the beetles. Positive values (red print) strengthen a particular category compared to the other three. Negative values (blue print) weaken this category at the expense of strengthening the other categories. Red dots in the SHAP model diagram indicate higher values of the feature, while blue dots indicate lower values of the feature. For general trait ecological element (Ecol\_Class) nominal values of categories: 1. typhophiles, 2. lakes and river species and rheobionts, 3. argilophiles, 4. Eurytopes. For body size (Size\_Class) nominal values of categories: 1. 1 – 1.99 mm, 2. 2 – 4.99 mm, 3. 5 – 9.99 mm, 4. > 10 mm. F-scores and mean Shapley values are averages of 5 random repetitions of XGBoost and SHAP models. Accuracy of the train and test subsets of the model, see Table S2.

Ecological traits	Body size				Ecological element				Trophic type			
Categories	1-2 mm	2-5 mm	5-10 mm	>10 mm	Tyrphoph.	L.&R. sp.	Argiloph.	Eurytopes	Shredders	Graz.&Sc.	Polyphaga	Predators
Lake area	<b>0.33</b>	<b>0.22</b>	<b>0.49</b>	<b>-0.40</b>	<b>-0.63</b>	<b>0.52</b>	<b>0.41</b>	<b>0.25</b>	<b>-0.42</b>	<b>0.54</b>	<b>0.75</b>	<b>-0.35</b>
Lake stadium	<b>0.14</b>	<b>-0.14</b>	<b>-0.21</b>	<b>0.21</b>	<b>-0.11</b>	<b>-0.17</b>	<b>-0.08</b>	<b>0.23</b>	<b>-0.11</b>	<b>-0.16</b>	<b>-0.23</b>	<b>-0.20</b>
Habitat fractal	0.06	0.04	-0.07	0.03	-0.09	-0.05	<b>-0.11</b>	0.06	0.06	<b>0.21</b>	0.06	-0.08
Surroundings	<b>0.17</b>	<b>-0.14</b>	0.09	-0.09	<b>0.40</b>	-0.08	<b>-0.14</b>	0.06	0.06	<b>-0.20</b>	-0.05	<b>0.11</b>
Sampling depth	<b>-0.12</b>	<b>0.13</b>	<b>0.16</b>	<b>0.22</b>	<b>-0.18</b>	<b>-0.12</b>	<b>0.18</b>	<b>0.18</b>	<b>-0.19</b>	<b>0.15</b>	<b>-0.19</b>	<b>-0.13</b>
Water pH	<b>0.18</b>	<b>-0.16</b>	<b>0.20</b>	<b>-0.19</b>	<b>-0.35</b>	<b>0.47</b>	<b>0.27</b>	<b>0.17</b>	<b>-0.23</b>	<b>-0.17</b>	<b>0.43</b>	<b>-0.20</b>
Bottom	<b>0.20</b>	<b>-0.12</b>	<b>0.20</b>	<b>-0.23</b>	<b>-0.26</b>	<b>0.26</b>	<b>0.26</b>	<b>-0.25</b>	<b>-0.18</b>	<b>0.37</b>	<b>-0.16</b>	<b>-0.16</b>
Detritus	<b>0.26</b>	<b>-0.14</b>	<b>-0.23</b>	<b>0.20</b>	<b>0.16</b>	<b>0.14</b>	<b>-0.11</b>	<b>-0.13</b>	<b>-0.17</b>	<b>-0.21</b>	<b>-0.14</b>	<b>0.13</b>
Influxes	<b>0.12</b>	0.08	<b>-0.18</b>	<b>-0.16</b>	<b>0.17</b>	-0.09	<b>-0.10</b>	<b>0.12</b>	0.07	<b>-0.17</b>	<b>0.20</b>	-0.09
Shore	0.08	-0.07	<b>-0.10</b>	<b>0.16</b>	<b>0.12</b>	-0.06	<b>-0.55</b>	0.07	<b>-0.10</b>	0.06	0.08	-0.09
Grass	-0.04	-0.04	0.05	0.05	-0.07	-0.09	-0.03	0.09	-0.04	0.03	0.05	-0.03
Elodeids	-0.02	-0.01	-0.01	0.01	<b>0.10</b>	-0.04	-0.04	0.04	-0.04	<b>0.10</b>	0.04	0.03
Nymphaeids	-0.06	0.04	-0.07	-0.06	<b>0.10</b>	-0.06	-0.08	0.05	-0.04	-0.05	0.08	-0.04
Macrophytes	-0.06	0.08	-0.06	-0.07	<b>0.12</b>	-0.06	-0.04	0.04	-0.04	-0.07	-0.09	0.07

**Fig. 4.** Distribution of mean Shapley values between categories of ecological traits and these categories and classes of habitat features based on Figs. 1, 2, 3 and Table 1. **Cell colours:** The importance of habitat features and categories is expressed by a colour scale: orange for positive Shapley values (the selection of a habitat feature by beetles of a certain category is favoured by higher nominal values of habitat feature classes or categories according to the classification from Table 1) and blue for negative values (favoured by lower nominal values). Four intensities of these colours were distinguished in the ranges of mean absolute Shapley values: > 1.00, 1.00 – 0.50, 0.50 – 0.20, 0.20 – 0.10 and 0.10 – 0 (Figs. 1, 2, 3). **Frame colours:** the same frame colours (black, red, green, blue) indicate that the different beetle categories (more than 2) select certain habitat features with the same ranges of nominal values of classes higher (orange cells) hb, or lower (blue cells). See detail description in paragraph below. **Font:** only higher (> 0.10) Shapley values are marked in bold and only these were considered in the presentation of habitat features for the different ecological categories of beetles.

Table 1). A whole series of ecological categories of aquatic beetles can also be found in Fig. 4, which together select a single habitat feature.

## Discussion

### Approach and modelling

In our work, we assumed a very high variability of environmental factors affecting the phenomenon of habitat selection by animals in general<sup>43</sup> and aquatic beetles in particular as objects of our research. To avoid the problem of scaling in habitat selection<sup>44</sup>, we used habitat features rather than habitat type to model the habitat selection of beetles. Most of the assumed habitat features were discrete data and the model identified habitat features based on class selection in a binary or multiclass system (Table 1). Another previously unused measure aimed at making the context of habitat feature selection by the studied organisms more realistic was the inclusion of basic ecological traits such as body size, ecological element and trophic type in the set of variables characterising habitats. This approach allows the selection of habitat features to be considered dependent on these general traits, as larger species may have an advantage in selecting a specific habitat<sup>2</sup> and show evolutionarily determined preferences for selected landscape types<sup>45</sup> or search for habitats where they expect to find food resources<sup>46</sup>. This procedure led to the achievement of model fit and high indicators of model accuracy (Appendix S1: Table S2), bringing the applied modelling closer to natural reality. In a comprehensive review of habitat selection modelling in animals<sup>22</sup>, the authors found no analogy to our approach, which is based on the study of habitat features using decision algorithms and elements of explainable machine learning.

The authors opted for modelling with a decision tree as the basic calculation algorithm and a further multiplication to a random forest as the most effective algorithm, which guarantees the highest accuracy of the models based on it and at the same time is relatively resistant to overfitting<sup>47</sup>. These properties of random forest-based models and their usefulness for XGBoost model in environmental research were confirmed by the work of Park et al<sup>48</sup>, Grbčić et al.<sup>49</sup> and Kruk et al.<sup>50,51</sup>. As the accuracy and fit results of the XGBoost models in this paper show (Appendix: Table S2), very high approximations of these models to the ecological reality contained in the database were mostly achieved for all four categories of general ecological traits (Figs. 1, 2, 3). Especially when considering the predominantly discrete nature of the data based on 2 to 6 classes and, above all, the highly variable and mobile nature of aquatic beetle species.

In this paper, for the first time in the ecological literature, we based the assessment of the phenomenon of habitat selection on differences in the calculations of the feature importance measures of the XGBoost (F) and SHAP (Shapley value) models. In the SHAP model we used an innovative method of comparative analysis that revealed the effects of interactions between aquatic beetle species on the importance of a particular general

ecological trait and a habitat feature. The first objective in using the SHAP model was to show how the selection of a particular ecological category of aquatic beetles for a particular habitat feature may be influenced by the relationships between a sample of species according to the comparative technique with the model with and without the sample<sup>37</sup>. The results of this SHAP modelling are different from XGBoost itself, as they are based on a different assumption: not on the frequency of co-occurrence of given features/traits (XGBoost), but on the evaluation of the importance of the feature/trait based on its impact on the model, resulting from the calculation of the model when it is missing for each observation (sample) with that when it is present in the model (SHAP).

In this system, traits (body size, ecological element, trophic type) are more valued in a SHAP model, while habitat features lose importance. This is the result of the individual approach (for each sample representing a species) in SHAP, i.e. the effect of the multiple interactions of each sample (species) with the other samples (species). This seems intuitively understandable and is based on observations of communities of mobile organisms. In other words, the XGBoost model itself provides information on the habitat selection of different ecological categories of beetles, while the addition of the SHAP model, i.e. the interaction between samples (species) in the modelling mechanism "with and without sample" indicates a greater importance for ecological traits the categories of other ecological traits than habitat features. Some explanations have been added to the text for clarity.

In our work, we also present a new approach to the definition of an object that selects a habitat and to the concept of "habitat selection". Namely, in this study, we have categories of general ecological traits (body size ranges, ecological groups and trophic types), not species or taxonomic groups, on the side of the habitat-selecting object. These categories of beetles do not select habitats in the sense of areas (shore, lake, littoral, etc.), but their features discretized into classes. This approach to habitat corresponds most closely to the concept of "habitat unit", which is described as "discrete analyst-defined areas in geographic space, over which the environmental variables representing habitat are quantified"<sup>22</sup>. Thanks to SHAP modelling, an assessment of the importance of a given habitat feature for its selection by the studied ecological category of beetles (e.g. with a body size > 10 mm, predators, etc.) was implemented. This is one of the main innovations in this work, which aims to show the results and conclusions of the work on more universal ecological categories in relation to species or their taxonomic groups which are marked a significant proportion of the locality attribute. The modelling used also enabled an innovative method to identify a common habitat feature selection for several different ecological categories from different general traits of the beetles studied (Fig. 4).

The authors are also aware of the shortcomings of the solutions adopted when approaching the object of research and its modelling. They mainly concern two elements. The first is the selection of habitat features and their classification. The methods for selecting habitats for analysis can certainly be more formalised and the separation and classification of their features improved. The second important element is the reliability of the modelling carried out. Despite the use of a very extensive database of 2511 datasets and the utmost care in the selection of models included in the XGBoost-SHAP model ensemble, it must be recognised that the modelling is only a simplification of natural reality. Even when including a model (SHAP), based on the individualisation of the modelling to a single sample and the construction of a model based on it (according to the bottom-up principle), the results and conclusions relate to this mathematical simplification and not to more complicated and variable real phenomena. SHAP modelling is not free from imperfections that require attention, such as: 1. SHAP values assume that the contribution of each feature is additive, which may not always be the case in real-life scenarios. 2. Shapley values are sensitive to the data used to train the model, which can lead to inconsistent results if the data is skewed or noisy. The listed potential shortcomings in database creation and modelling should be considered when conducting similar research projects in the future.

### Importance of ecological traits and habitat features for aquatic beetles

Body size is a fundamental general ecological trait that characterises organisms in the environment, including water beetles. For example, Bilton et al.<sup>2</sup> highlighted the importance of body size for the evolutionary success of beetles, which is particularly important for their colonisation of new aquatic habitats. Relationships between temperature and body size and the effects of climate change on the body size of ectothermic aquatic organisms have also been observed<sup>1,52</sup>. In our study, the body size of aquatic beetles was analysed to understand the influence of habitat, ecological and trophic types on the four categories of this general ecological trait. The results showed that the ecological element and lake size were the dominant feature influencing the body size of aquatic beetles, and these features were about 1.5 times more important than trophic type and other habitat features, such as water depth, bottom condition detritus thickness and developmental stage of the lake.

In beetles of the smallest size class (1–2.99 mm) belonging to typhophiles (e.g. *Haliphus fulvicollis* and *Hydraena palustris*) and lake and river species (*Oulimnius tuberculatus* and *Haliphus fluviatilis*) and larger lake surfaces, this size category increased, which is consistent with the observations of<sup>41</sup>. The larger surface area of lakes is also an important feature for the next two body size categories (3 – 10 mm). However, this trend does not apply to the largest beetles (> 10 mm), for which a positive habitat feature is the smaller size of the reservoirs (Fig. 4). These are mainly predatory typhophiles (*Acilius canaliculatus*, *Colymbetes striatus*, *Colymbetes paykulli* and *Ilybius ater*) and lake and river species (*Ilybius fuliginosus* and *Ilybius fenestratus*). The characteristic of the ecological element plays an important role for all body size categories. In this case, there is a division into categories with the smallest and largest body sizes, which are reinforced by typhophiles and lake and river species, and, on the other hand, categories with medium body sizes (3 – 10 mm), which are represented by argilophiles and eurytopes. The latter effect could be related to specific habitat features that reduce the extreme features of individuals<sup>7,53</sup>). An important feature for all body size categories is the water pH, which indicates that for body sizes 1 – 2.99 and 5 – 9.99 mm it is indicative of more alkaline waters, while for sizes 3 – 4.99 and the largest it indicates a neutral or acidic habitat. Interestingly, the largest predatory aquatic beetles are also found in dystrophic waters (Fig. 1c), where at the same time the pressure of predatory fish is limited<sup>54–56</sup>). A very clear

feature of all body size categories of beetles in the studied community is a radical change in the importance of trophic type in the SHAP model, which takes into account interspecific interactions in particular. The general ecological trait trophic type becomes a dominant feature in all body size categories, confirming the conclusions of many studies about a positive correlation between body size and occupied trophic level<sup>13,4,57</sup>.

Aquatic beetles play a key role in assessing the quality of freshwater ecosystems due to their sensitivity to environmental change<sup>5,12,56,58–63</sup> and therefore their general ecological traits are an important criterion for determining the habitat selection of these organisms. The composition of aquatic beetle communities can be influenced by various habitat features such as water pH or the stage of lake succession<sup>12,41</sup> as well as microhabitat features such as habitat size and vegetation structure<sup>64–66</sup>. These features contribute to the development of certain ecological traits over the course of evolution and, at the same time, to the emergence of distribution patterns of water beetles in different types of aquatic habitats<sup>67</sup>.

In all categories of general ecological traits, body size and lake size occupied the positions of the most important features strengthening or weakening these categories<sup>53,56,66,68,69</sup>. The importance of trophic type also increased significantly with the inclusion of interactions between beetle species in SHAP modelling. These conclusions are consistent with the results described above on the importance of body size in aquatic beetles. The modelling conducted revealed specific selection for general ecological and habitat features for different ecological elements of aquatic beetles. Thus, typhophiles favour body size, lake area and trophic type, with smaller body sizes and smaller areas of dystrophic lakes. Lower trophic levels (shredders, grazers and scrapers) strengthening this ecological element, as evidenced by the numerous occurrence of the smallest beetles, including such as: *Anacaena lutescens* and *Hydraena palustris*<sup>64,65</sup>. On the other hand, large predatory typhophiles such as *Acilius canaliculatus*, *Colymbetes striatus*, *Colymbetes paykulli* and *Ilybius ater* live in the same lakes. Thus, typhophiles generally prefer smaller lakes with habitat features such as shallows, lower pH, detritus content and muddy bottoms (Fig. 4). The body size and trophic type of typhophiles are also related to the fractal structure of the littoral zone<sup>53,65</sup> but only dystrophic lakes were analysed in the cited papers.

Features such as body size, lake area and trophic type were also most important for the elements of lake and river species. However, larger body sizes, larger lake areas and affiliation with polyphages and predators (*Gyrinus aeratus*, *Hygrotus versicolor*, *Ilybius fenestratus*, *Ilybius fuliginosus* or *Platambus maculatus*) strengthened this ecological element. In addition, the detritus content, higher pH, shallower depth of occurrence of the species, coarse-grained bottom and earlier stage of lake transformation are important for this category and confirm the good condition and stable conditions in the lakes where these species occur<sup>56</sup>.

Argilophiles clearly favour larger lakes and smaller body sizes (Fig. 4). For this category, the importance of trophic types (shredders: e.g. *Laccobius biguttatus*, *Laccobius bipunctatus*, *Laccobius minutus*, *Helophorus granularis*, *Helophorus minutus*, *Helophorus pumilio*) and predators (e.g. *Hydroglyphus pusillus*) has increased significantly in the SHAP model. The species mentioned are usually associated with bare bottoms or those sparsely covered with macrophytes<sup>53,58,59</sup>. Argilophiles are also influenced by features such as higher pH and coarse-grained bottom<sup>68</sup>.

The features that strengthen the eurytopic element are larger body sizes and higher trophic levels. They are also characterised by the importance of a larger lake area, greater water depths and higher pH. Such species are the larger Polyphaga and the larger predators: *Dytiscus marginalis*, *Colymbetes fuscus*, *Acilius sulcatus*, *Graphoderus cinereus*, *Graphoderus bilineatus*, *Cybister lateralimarginalis*.

The functioning of trophic structure in beetle communities, as in other organisms, is a factor that determines their functioning, stability and survival<sup>70</sup>. The results of modelling the categories of trophic types in aquatic beetles showed the importance of different habitat features and general ecological variables in the differentiation of trophic types in the studied beetle community. Our studies have shown that body size, affiliation to certain ecological elements and lake surface area play a dominant role in trophic differentiation (Fig. 3). In addition, habitat features such as the nature of the bottom, the depth of the beetle in the water profile and water pH value of the water were also important for the differentiation of trophic types, albeit to a lesser extent.

We have demonstrated specific trait selection for different trophic types of aquatic beetles. For example, the category of shredders was mainly strengthened by a smaller body size, belonging to the typhophiles, and a smaller lake area (Fig. 4). In addition, we found that grazers and scrapers favour larger lakes, tended to choose the role of argilophiles (e.g. *Ochthebius minimus* and eurytopes (e.g. beetles of the genus *Limnebius*) and a coarser-grained nature of the bottom. These selections for certain habitat features are consistent with the findings of<sup>71</sup>, who described the contrasting effects of different aquatic habitats on the genetic characteristics of water beetles. In the discussed category, an increase in the importance of body size as a result of interactions between species is clearly evident. This effect could indicate considerable diversity in the grazers and scrapers categories<sup>72</sup>, but with a clear dominance of smaller individuals. The same is true for the polyphages category. This group is also strengthened by smaller body sizes, but also by belonging to argilophiles (e.g. *Haliphus confinis* and *Haliphus immaculatus*), eurytopes (*Haliphus heydeni*, *Haliphus lineolatus*, *Haliphus ruficollis*) and lake and river species (*Haliphus fulvus*, *Haliphus flavicollis*, *Haliphus fluviatilis*) and a larger lake area. An interesting case for the importance of features is the predators category. Here, as with the previous category, the reinforcing categories are argilophiles and eurytropes, but, as with the shredders, a smaller lake area. However, if we consider the SHAP model, the strongest overall ecological trait reinforcing the predators is body size, and these are clearly individuals with larger body sizes, e.g. *Dytiscus marginalis*, *Dytiscus circumcinctus* and *Dytiscus dimidiatus* (Fig. 3d). This is consistent with the analysis in which the largest beetles (> 10 mm) were the scored category and predators upgraded the trophic type category (Fig. 1B, Fig. 4). It is worth drawing attention to the very species-rich group of predators, confirming previous research, e.g.<sup>7</sup>, which may be dominated by predatory species whose food resources are not other beetle species, but e.g. zooplankton or larval forms and even fish fry<sup>68,69,73</sup>.

## Conclusion

The analysis of the habitat selection of the water beetle community, understood both from the general ecological aspect and in the detailed habitat context, based on XGBoost-SHAP ensemble modelling represent novel approach in the ecological data modelling. The ensemble modelling used explains the impact of the interaction between species on the importance of general ecological and habitat features and evaluated their impact on the beetle categories. Indeed, if the beetles only “select” features based on the frequency of their association in their functioning (the XGBoost model itself), then the dominance of general ecological traits is lower and a feature of the habitat, such as the lake area, appears as an equally important factor. At the top of the hierarchy of importance of features are the water pH, the nature of the bottom, the presence of detritus in the water, the depth of beetles occurrence in the water column. However, when we take into account the interactions between the beetle species in the studied community (by addition SHAP model), we observe a clear increase in the importance of general ecological traits for all their categories. Thus, for the beetle body size categories, the ecological element and trophic type traits are the most important, for the ecological element categories the body size or trophic type traits are of highest importance, and for the position in the trophic system of beetles the body size and ecological element traits are crucial. The initial hypothesis 1. on the dominance of general ecological traits in relation to the features of the habitats in which they occur was confirmed.

The modelling confirmed hypothesis 2. In fact, the lake area proved to be the most important habitat feature for water beetles even after applying the SHAP model. In turn, for hypothesis 3, the position in the ranking of importance of lake habitat features remained quite high, essentially only in terms of lake bottom and detritus content. On the other hand, the ensemble modelling carried out negatively confirmed hypothesis 4. The selection of riparian habitat features such as the fractal character of the littoral zone, the presence of forest and the type of littoral submerged vegetation do not play a role in the selection of habitat by aquatic beetles. Regarding hypothesis 5, it has been found that it is possible to group categories with different ecological traits that are linked by the same choice of habitat features. We found four such systems covering almost all beetle categories studied and a wide range of habitat features related to the aquatic environment.

The use of ensemble modelling, which is based on the “selection” of features by the beetle species represented in the samples, and SHAP modelling, which allows an analysis of the interactions between these samples of beetle species, is an approach that allows an assessment of the environmental situation of a community of organisms at several levels. Analysing the importance of the habitat features for beetles ecological categories can provide a picture of the priority environmental variables for the stability of the ecosystem and valuable information for practical conservation and ecosystem management. The authors are planning further work in which, among other things, they will use advanced machine learning models analyse habitat selection not for ecological categories but chosen water beetle species.

## Material and methods

### Study area, sampling and material

We studied 25 lakes in northern Poland, in the Masurian Lake District. The analysed lakes differed in terms of surface area, depth, degree of development and differentiation of the phytolittoral zone. The characteristics of the lakes are included in the works of Pakulnicka and Zawal<sup>53,58,59</sup>.

The studies were conducted from 1998 to 2014 in spring, summer and autumn. A total of 624 samples were collected. The fauna samples were collected with a dip net covering an area of approximately 1 m<sup>2</sup>. In the pressed *Sphagnum* mats of the dystrophic lakes, 10 subsamples were taken with a 0.1 m<sup>2</sup> sieve. In total, we collected 11,799 beetles representing 146 species (Appendix S1: Table S.1). The sampling sites were chosen to represent the greatest possible diversity of littoral habitats and areas of the individual water bodies. The methods and values for the analysed variables have already been presented in previous publications<sup>53,58,59,70</sup>.

### Modelling – general scheme

It should be first declared that our modelling objects, target variables, are general ecological traits, i.e. groups of species with specific traits, related to habitat features, as predictors, characterized by classes (see Table 1). We analysed the importance of different habitat features for basic ecological traits such as body size, ecological element and trophic type in the studied aquatic beetle community using advanced machine learning classification modelling. For this purpose, we used Extreme Gradient Boosting (XGBoost) and SHapley Additive exPlanations (SHAP) algorithms. XGBoost modelling is based on decision tree and random forest techniques<sup>74</sup>. The variables used for the modelling (Table 1) consist mainly of discrete numerical data. The exceptions are two continuous variables: lake area and water pH. The distribution of their values, due to their bias towards larger lakes and lakes with alkaline water, prevented a sufficiently equal categorisation into classes. These variables were discretised as random forest components in the basic decision tree model<sup>75</sup>. The accuracy results of the XGBoost models show the stability of the modelling when using discrete and continuous data together (Appendix S1: Table S2).

The gradient boosting transforms any weak classifier into a strong classifier and reduces the residuals of previous models in the direction of the gradient to obtain a new, stronger model<sup>76</sup>. XGBoost with a decision tree explainer is the most effective tool among the tools based on boosting techniques in machine learning<sup>48</sup>. In our model, the dependent target variable (Y) is, as in the classification models, a binary classification on two classes 0 and 1, while the predictors are independent variables. For example, in the “Trophic type” model (Fig. 3a), we try to explain which predictors: 14 habitat features and other 2 traits (X variables) play an important role for category Shredders (1) compared to other categories (grazers & scrappers, polyphaga, predators) in the Trophic Type trait (0). We used a dataset of 2511 samples split for classification modelling at a ratio of 0.8 for training and 0.2 for testing the model.. We adopted the following hyperparameters of the XGBoost model:

`n_estimators=200, max_depth=10, learning_rate=0.001`. In our study, the XGBoost tool was also used as an introductory model for SHAP modelling.

To assess the importance of interactions between beetle species (at the sample level) and the tendency for a particular category of a general ecological trait to be more or less related to a habitat factor, we added SHAP model to XGBoost<sup>37</sup>. This is a game-theoretic method that uses the Shapley value metric<sup>36</sup> and is used to predict the “global” (for the whole model) and “local” (for each observation) importance of variables in binary classification tasks<sup>38</sup>. The general scheme and the idea of modelling pathway and is shown in the Fig. 5.

The runs of the XGBoost-SHAP ensemble model were repeated 5 times and the final results of modelling accuracy and feature importance are the average of the results of these repetitions. Accuracy indicators for training and test trials were calculated for the XGBoost models. Their aim was to determine the reliability of the modelling results and to identify the effects of underfitting and overfitting. The modelling was performed in Python version 3.9. We adapted the code from the Kaggle notebook “Ensemble and Model Stacking”<sup>77</sup>.

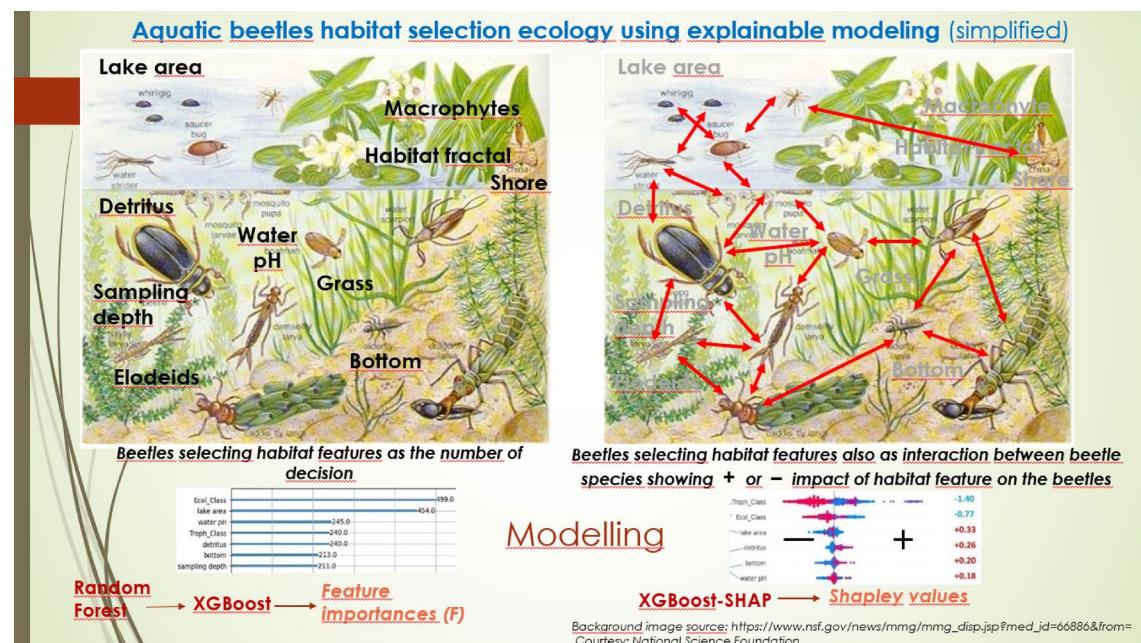
### Modelling – feature importances

To assess the importance of general ecological features (other than the one studied) and habitat features for the variability of the categories separated from body size, ecological element and trophic type, two measures from the field of advanced machine learning were used: the F measure of feature importance in the boosting analysis and the Shapley value in SHAP. Both indicate different aspects of the importance of features for the models to be evaluated.

The XGBoost algorithm uses the F-score measure to determine the importance of a variable in the model as the amount of usage of a particular variable in the boosting process, i.e. in the nodes that separate paths (splitting), in all trained iterations of random forest decision trees as a submodel of XGBoost<sup>78</sup> (Fig. 5). In other words, the F-score is the frequency of a feature being used to split the classification trees in the random forest used in our XGBoost model. In the Python script for the XGBoost model version 1.5.0. this is the default built-in method “Weight”. This is a simple, easy to interpret method. From an ecological perspective, higher F-values of variables mean that traits and habitats that occur more frequently when the beetles repeatedly select the studied category are favoured over variables with low F-values. The eXtremeGradientBoosting (XGBoost) algorithm was used to predict modelling results and assess the accuracy of the model. It confirmed the assumed categorisation of general ecological traits of water beetles and the effects of these categories on the habitat selection of these insects (see Table S2 in Supplementary Material).

XGBoost modelling for a particular feature (Figs. 1–3) consists of a multiple, random “decision” for a particular feature or trait to select the habitat features and remaining traits. The model have taken into account the co-occurrence or lack thereof in the decision tree sub-model (which is in the form of multiple trials in a random forest) as the basis for calculating the importance of the variables in the XGBoost model. In this way, the model “shuffles” the traits and features, and as a result, indicates the importance of the individual traits and features resulting from their joint occurrence in the tree model.

The SHAP model is an “overlay” to the XGBoost model and its task is to evaluate the importance of traits and features in relation to their role in the model of a particular trait, i.e. body size (Fig. 1), ecological element



**Fig. 5.** Idea and modelling scheme of the habitat selection ecology of aquatic beetles. Background image source: [www.nsf.gov/news/mmg/mmg\\_disp.jsp?med\\_id=66886&from=Courtesy: National Science Foundation](https://www.nsf.gov/news/mmg/mmg_disp.jsp?med_id=66886&from=Courtesy: National Science Foundation)

(Fig. 2) and trophic type (Fig. 3). Thus, we have three groups of models in which all habitat features and other traits (as predictors) are counted as feature importance, except for the trait that is the target of the modelling (its dependent variable).

The importance of a feature in the SHAP model, which is calculated with Shapley values<sup>36</sup>, is derived from the importance of a particular feature for the prediction of the corporate game model<sup>37</sup>, where the “win” is determined by comparing the predictions of the model with and without the variable for one class, e.g. body size > 10 mm, and “lose” for classes with smaller body sizes. For models with more than two categories, the SHAP model uses a sequential comparison of each category with the remaining categories treated as a second category<sup>78</sup>.

According to the formal definition, the method for calculating the Shapley  $\Phi_i$  value involves training the model on all feature subsets  $S \subseteq F$ , where  $F$  is the set of all features (general-ecological and habitat features of the studied beetle community) and  $S$ —the categories and classes of these features (Table 1). The method assigns an importance value to each feature, which represents the influence of this feature on the predictions of the model. To calculate this influence, the  $f_{S \cup \{i\}}$  model is trained with this feature and the  $f_S$  model is trained without this feature. Then the predictions of the two models are compared in the expression  $f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)$ , where  $x_S$  represents the values in the set  $S$ , i.e. the values of the categories and classes in each instance  $\{i\}$ , i.e. the sample of individuals of a particular beetle species. As the effect of omitting a feature depends on other features in the model, the differences of the two models (with and without the feature) are calculated for all possible subsets of  $S \subseteq F \setminus \{i\}$  (i.e. all categories and classes of all samples). This results in the effect that the individual samples interact with each other, which is reflected in a mutual interaction between the beetle species represented by these samples. Through this double model training procedure, the Shapley values are then calculated and used as feature attributes. They are a weighted average of all possible differences according to the formula<sup>37</sup>:

$$\Phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)]$$

The interpretative value of the two importance measures ( $F$  in XGBoost and Shapley value in SHAP) thus concerns other properties of the functioning of the beetle community in the environment. The  $F$ -measure in the boosting model refers to the frequency of selection of a particular feature by insects belonging to a particular category, whereas the Shapley value refers to the selection for every beetle sample of a particular feature for that category relative to other categories of a common general ecological trait. The Shapley value thus takes into account the interactions between species (competition, feeding or other relationships) of different categories of water beetles in their preference for general ecological traits and habitat features selection (Fig. 5). It was achieved by individual evaluation of the importance of every sample by the comparison the results of running the model with this sample and without it.

In other words, interspecific relations between aquatic beetles studied refer to the results of the SHAP modelling technique, which consists of calculating the importance (Shapley value) of each observation (assigned to a sample of beetle species) from the difference between the modelling results with and without that observation. In Figs. 1–3, the SHAP modelling results (Shapley values) are presented as a sets of points (Shapley value of observations) whose location indicates the “favourability” (+) or “non-favourability” (-) of the general trait under study (e.g. body size) for a given category (e.g. > 10 mm) by the given habitat features and other ecological traits (independent variables). The colour of points on the graph indicates the nominal value of the habitat feature or trait feature category (e.g. lake area or predators).

If, for a given feature, the ranking of the Shapley value was higher than the ranking of the  $F$  value, the importance of this variable for the general ecological feature studied is characterised to a greater extent by interactions between species than just by the frequency of occurrence of a beetle category with a given feature or in a given habitat. If this rank is lower, it means that the interaction effect does not occur for a particular feature or category, i.e. it is so weak that it weakens the position of this variable in the importance ranking.

To predict the importance of one of the four categories of body size, ecological element or trophic type, the Shapley values are treated as the importance of habitat variables and general ecological traits, apart from the feature whose category is being analysed. These variables and features influence the predictions of the model with different importance and direction (+ or — sign), which is expressed by Shapley values. Variables with Shapley values with a positive sign contribute to the prediction of the gain of the tested category, while variables with a negative sign contribute to the prediction of the gain of the other three categories, i.e. contribute negatively to the prediction of the tested feature. In accordance with the principles described, four XGBoost-SHAP models were created for each category of three general ecological traits:

- Four models of body size category (1. 1–1.99 mm, 2. 2–4.99 mm, 3. 5 to 9.99 mm, 4. > 10 mm), influenced by the general ecological traits: ecological element and trophic type and 16 habitat features.
- Four models of the ecological element category (1. typhophiles, 2. lakes, river species and rheobionts, 3. argilophiles, 4. eurytopes), influenced by the general ecological traits: body size and trophic type, and 16 habitat features.
- Four models of the trophic type category (1. shredders, 2. grazers and scrapers, 3. polyphages, 4. predators), influenced by the general ecological traits body size and ecological element and 16 habitat features.

The influence of general ecological traits and habitat features on the above categories may tend to:

- Increase the impact on the importance of the tested category (positive Shapley values) and decrease the impact on the importance of the other 3 categories,

- Decrease the impact on the importance of the tested category (negative Shapley values), at the expense of increasing the impact on the importance of the other 3 categories, or.
- No clear impact on the importance of all 4 categories (Shapley values close to 0).

## Data availability

The data used in the study are available from the authors on request: joanna.pakulnicka@uwm.edu.pl, mkruk@uwm.edu.pl.

Received: 27 July 2024; Accepted: 14 November 2024

Published online: 21 November 2024

## References

1. Tseng, M. et al. Decreases in beetle body size linked to climate change and warming temperatures. *J. Anim. Ecol.* **87**, 647–659. <https://doi.org/10.1111/1365-2656.12789> (2018).
2. Bilton, D., Ribera, I. & Short, A. Water beetles as models in ecology and evolution. *Annu. Rev. Entomol.* **64**, 359–377. <https://doi.org/10.1146/annurev-ento-011118-111829> (2019).
3. Driscoll, D. A., Smith, A. L., Blight, S. & Sellar, J. Interactions among body size, trophic level, and dispersal traits predict beetle detectability and occurrence responses to fire. *Ecol. Entomol.* **45**, 300–310. <https://doi.org/10.1111/een.12798> (2020).
4. Basile, M. et al. Diversity of beetle species and functional traits along gradients of deadwood suggests weak environmental filtering. *For. Ecosyst.* **10**, 100090. <https://doi.org/10.1016/j.fecos.2023.100090> (2023).
5. Pakulnicka, J. et al. Aquatic beetles (Coleoptera) in springs situated in the valley of a small lowland river: Habitat factors vs landscape factors. *Knowl. Manag. Aquat. Ecosyst.* **417**, 29. <https://doi.org/10.1051/kmae/2016016> (2016).
6. Gillespie, M. A., Birkemoe, T. & Sverdrup-Thygeson, A. Interactions between body size, abundance, seasonality, and phenology in forest beetles. *Ecol. Evol.* **7**, 1091–1100. <https://doi.org/10.1002/ece3.2732> (2017).
7. Pakulnicka, J. & Kruk, M. Regional differences in water beetle communities networks settling in dystrophic lakes in northern Poland. *Sci. Rep.* **13**, 12699. <https://doi.org/10.1038/s41598-023-39689-z> (2023).
8. Foster G.N. & Eyre M.D. Classification ranking of water beetle communities. UK nature conservation: 1. Peterborough: Joint Nature Conservation Committee, 1–110 (1992).
9. Bosi, G. Observations on Colymbetine predation based on crop contents analysis in three species: Agabus bipustulatus, Ilybius subaeneus, Rhantus suturalis (Coleoptera: Dytiscidae). *Boll. Soc. Entomol. Ital.* **133**, 37–42 (2001).
10. Menetrey, N., Sager, L., Oertli, B. & Lachavanne, J. B. Looking for metrics to assess the trophic state of ponds. *Macroinvertebrates and amphibians. Aquat. Conserv.* **15**, 653–664. <https://doi.org/10.1002/aqc.746> (2005).
11. Pakulnicka, J. The formation of water beetle fauna in anthropogenic water bodies. *Oceanol. Hydrobiol. St.* **37**, 31–42. <https://doi.org/10.2478/v10009-007-0037-y> (2008).
12. Gioria, M., Bacaro, G. & Feehan, J. Identifying the drivers of pond biodiversity: the agony of model selection. *Community Ecol.* **11**, 179–186. <https://doi.org/10.1556/ComEc.11.2010.2.6> (2010).
13. Pakulnicka, J. et al. Development of fauna of water beetles (Coleoptera) in waters bodies of a river valley habitat factors, landscape and geomorphology. *Knowl. Manag. Aquat. Ecosyst.* **417**, 40. <https://doi.org/10.1051/kmae/2016027> (2016).
14. Heino, J. & Alahuhta, J. Knitting patterns of biodiversity, range size and body size in aquatic beetle faunas: significant relationships but slightly divergent drivers. *Ecol. Entomol.* **44**, 413–424. <https://doi.org/10.1111/een.12717> (2019).
15. Binckley, C. & Researats, W. Habitat selection determines abundance, richness and species composition of beetles in aquatic communities. *Biol. Letters* **1**, 370–374; doi.org/https://doi.org/10.1098/rsbl.2005.0310 (2005).
16. Ribera, I., Barraclough, T. & Vogler, A. The effect of habitat type on speciation rates and range movements in aquatic beetles: inferences from species-level phylogenies. *Mol. Ecol.* **10**, 721–735. <https://doi.org/10.1046/j.1365-294x.2001.01218.x> (2001).
17. Pakulnicka, J. et al. Sequentiality of beetle communities in the longitudinal gradient of a lowland river in the context of the river continuum concept. *PeerJ* **10**, e1323. <https://doi.org/10.7717/peerj.1323> (2022).
18. Wagner, R., Dapper, T. & Schmidt, H. H. The influence of environmental variables on the abundance of water insects: A comparison of ordination and artificial neural networks. *Hydrobiologia* **422**, 143–152 (2000).
19. Obach, M., Wagner, R., Werner, H. & Schmidt, H.-H. Modelling population dynamics of aquatic insects with artificial neural networks. *Ecol. Model.* **146**, 207–217; [https://doi.org/10.1016/S0304-3800\(01\)00307-6](https://doi.org/10.1016/S0304-3800(01)00307-6) (2001).
20. Wagner, R., Obach, M., Werner, H. & Schmidt, H. Artificial neural nets and abundance prediction of aquatic insects in small streams. *Ecol. Inform.* **1**, 423–430 (2006).
21. Hu, M., Jiang, S., Jia, F., Yang, X. & Li, Z. Improved Prediction of Aquatic Beetle Diversity in a Stagnant Pool by a One-Dimensional Convolutional Neural Network Using Variational Autoencoder Generative Adversarial Network-Generated Data. *Appl. Sci.-Basel.* **13**, 8841. <https://doi.org/10.3390/app13158841> (2023).
22. Northrup, J. M. et al. Conceptual and methodological advances in habitat-selection modelling: guidelines for ecology and evolution. *Ecol. Appl.* **32**, e02470; <https://doi.org/10.1002/eaap.2470> (2022).
23. Lundkvist, E., Landin, J. & Milberg, P. Diving beetle (Dytiscidae) assemblages along environmental gradients in an agricultural landscape in southeastern Sweden. *Wetlands* **21**, 48–58 (2001).
24. Biesiadka, E. & Pakulnicka, J. Water beetles (Coleoptera) in Łomżyński Landscape Park of Valley of Narew River. *Parki Narodowe i Rezerwaty Przyrody* **23**, 427–447 (2004).
25. Pakulnicka, J. & Nowakowski, J. J. The effect of hydrological connectivity on water beetles fauna in water bodies within the floodplain of a lowland river (Neman river, Belarus). *Oceanol. Hydrobiol. St.* **41**, 7–17 (2012).
26. Costea, G., Cojocaru, I. & Pusch, M. The Aquatic Beetles (Insecta: Coleoptera) assemblages in the Lower Prut Floodplain Natural Park (Romania). *Natura Montenegro* **12**, 719–736 (2013).
27. Hutchinson, G. E. Concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.* **22**, 415–427. <https://doi.org/10.1101/SQB.1957.022.01.039> (1957).
28. Shapire, R.E. & Freund, Y. Boosting: Foundations and Algorithm. <https://doi.org/10.7551/mitpress/8291.001.0001> (MIT Press, Cambridge, MA, 2012).
29. Ferrario, A. & Hämmmerli, R. On Boosting: Theory and Applications. SSRN. <http://ssrn.com/abstract=3402687> (2019).
30. Morris, D. W. Toward an ecological synthesis: a case for habitat selection. *Oecologia* **136**, 1–13 (2003).
31. Fretwell, S. D. & Lucas, H. L. On territorial behavior and other factors influencing habitat distribution in birds. *Acta Biotheor.* **19**, 16–36 (1969).
32. Mayor, S. J., Schneider, D. C., Schaefer, J. A. & Mahoney, S. P. (2009). Habitat selection at multiple scales. *Écoscience* **16**, 238–247; <http://www.jstor.org/stable/42902062> (2009).
33. Meysman, F. J. R. & Bruers, S. Ecosystem functioning and maximum entropy production: A quantitative test of hypotheses. *Philos. Trans. R. Soc. B* **365**, 1405–1416. <https://doi.org/10.1098/rstb.2009.0300> (2010).
34. Fortin, D., Morris, D. W. & McLoughlin, P. D. Habitat selection and the evolution of specialists in heterogeneous environments. *Isr. J. Ecol. Evol.* **54**, 311–328 (2008).

35. Hay, M. E. et al. Mutualism and Aquatic Community Structure: The Enemy of My Enemy Is My Friend. *Annu. R. Ecol. Evol. S.* **35**, 175–197. <https://doi.org/10.1146/annurev.ecolsys.34.011802.132357> (2004).
36. Shapley, L.S. A Value for n-Person Games, in *Contributions to the Theory of Games II*: (eds. Kuhn, H. W. and Tucker, A. W.) 315–317 (Princeton University Press, 1953).
37. Lundberg, S. M. & Lee, S. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30**, 4765–4774 (2017).
38. Štrumbelj, E. & Kononenko, I. An efficient explanation of individual classifications using game theory. *J. Mach. Learn. Res.* **11**, 1–18; <http://dl.acm.org/citation.cfm?id=1756006.1756007> (2010).
39. Yu, Q., Ji, W., Prihodko, L., Anchang, J. Y. & Hanan, N. P. Study becomes insight: Ecological learning from machine learning. *Methods Ecol. Evol.* **12**, 2117–2128. <https://doi.org/10.1111/2041-210X.13686> (2021).
40. Kalinkat, G., Jochum, M., Brose, U. & Dell, I. A. Body size and the behavioral ecology of insects: linking individuals to ecological communities. *Curr. Opin. Insect Sci.* **9**, 24–30. <https://doi.org/10.1016/j.cois.2015.04.017> (2015).
41. Bloechl, A., Koenemann, S., Philipp, B. & Melber, A. Abundance, diversity and succession of aquatic coleoptera and heteroptera in a cluster of artificial ponds in the north german lowlands. *Limnologica* **40**, 215–225. <https://doi.org/10.1016/j.limno.2009.08.001> (2010).
42. Simberloff, D. & Dayan, T. The Guild Concept and the Structure of Ecological Communities. *Annu. Rev. Ecol. Syst.* **22**, 115–143; <https://doi.org/10.1146/annurev.es.22.110191.000555> (1991).
43. Matthiopoulos, J., Hebblewhite, M., Aarts, G. & Fieberg, J. Generalized functional responses for species distributions. *Ecology* **92**, 583–589 (2011).
44. Beyer, H. L. et al. The interpretation of habitat preference metrics under use–availability designs. *Philos. T. Roy. Soc. B* **365**, 2245–2254 (2010).
45. Turchin, P. Translating foraging movements in heterogeneous environments into the spatial distribution of foragers. *Ecology* **72**, 1253–1266 (1991).
46. Owen-Smith, N., Fryxell, J. M. & Merrill, E. H. Foraging theory upscaled: the behavioural ecology of herbivore movement. *Philos. Trans. R. Soc. B-Biol S.* **365**, 2267–2278 (2010).
47. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. 22 ACM SIGKDD Conference on Knowledge, Discovery and Data mining, 12–17 August, San Francisco. doi.org/<https://doi.org/10.1145/2939672.2939785> (2016).
48. Park, J. et al. Interpretation of ensemble learning to predict water quality using explainable artificial intelligence. *Sci. Total Environ.* **832**, 155070. <https://doi.org/10.1016/j.scitotenv.2022.155070> (2022).
49. Grbčić, L. et al. Coastal water quality prediction based on machine learning with feature interpretation and spatio-temporal analysis. *Environ. Model. Softw.* **155**, 105458. <https://doi.org/10.1016/j.envsoft.2022.105458> (2022).
50. Kruk, M., Artiemjew, P. & Paturej, E. The application of game theory-based machine learning modelling to assess climate variability effects on the sensitivity of lagoon ecosystem parameters. *Ecol. Inf.* **6**, 101462. <https://doi.org/10.1016/j.ecoinf.2021.101462> (2021).
51. Kruk, M., Goździejewska, A. M. & Artiemjew, P. Predicting the effects of winter water warming in artificial lakes on zooplankton and its environment using combined machine learning models. *Sci. Rep.* **12**, 16145. <https://doi.org/10.1038/s41598-022-20604-x> (2022).
52. Daufresne, M., Lengfellner, K., & Sommer, U. Global warming benefits the small in aquatic ecosystems. *P. Natl. Acad. Sci.-Biol.* **106**, 12788–12793; doi.org/<https://doi.org/10.1073/pnas.0902080106> (2209).
53. Pakulnicka, J. & Zawal, A. Model of disharmonic succession of dystrophic lakes based on aquatic beetle fauna (Coleoptera). *Mar. Freshw. Res.* **69**, 1–17. <https://doi.org/10.1071/MF170502019> (2019).
54. Šiling, R. & Urbanič, G. Do lake littoral benthic invertebrates respond differently to eutrophication, hydromorphological alteration, land use and fish stocking? *Knowl. Manag. Aquat. Ecosyst.* **417**, 35. <https://doi.org/10.1051/KMAE/2016022> (2016).
55. Šigutová, H. et al. Specialization directs habitat selection responses to a top predator in semiaquatic but not aquatic taxa. *Sci Rep.* **11**, 18928. <https://doi.org/10.1038/s41598-021-98632-2> (2021).
56. Hansen, L.J. & Kreiling, A.-K. Small Islands, Small Ponds, Small Communities—Water Beetles and Water Boatmen in the Faroe Islands. *Insects* **13**, 923; doi.org/<https://doi.org/10.3390/insects13100923> (2022).
57. Pintar, M. & Resetarits, W. Match and mismatch: integrating consumptive effects of predators, prey traits, and habitat selection in colonizing aquatic insects. *Ecol. Evol.* **11**, 1902–1917. <https://doi.org/10.1002/ece3.7181> (2021).
58. Pakulnicka, J. & Zawal, A. Effect of changes in the fractal structure of a littoral zone in the course of lake succession on the abundance, body size sequence and biomass of beetles. *PeerJ* **6**, e5662. <https://doi.org/10.7717/peerj.56622018> (2018).
59. Pakulnicka, J. & Zawal, A. Community changes in water beetle fauna as evidence of the succession of harmonic lakes. *Fundam. Appl. Limnol.* **191**, 299–321; <https://doi.org/10.1127/fal/2018/11422018> (2018b).
60. Deacon, C., Samways, M. J. & Pryke, J. S. Artificial reservoirs complement natural ponds to improve pondscape resilience in conservation corridors in a biodiversity hotspot. *PLoS ONE* **13**, e0204148. <https://doi.org/10.1371/journal.pone.0204148> (2018).
61. Matsushima, R. & Yokoi, T. Flight capacities of three species of diving beetles (Coleoptera: Dytiscidae) estimated in a flight mill. *Aquat. Insects* **41**, 332–338. <https://doi.org/10.1080/01650424.2020.1804065> (2020).
62. Roth, N., Zoder, S., Zaman, A. A., Thorn, S. & Schmidl, J. Long-term monitoring reveals decreasing water beetle diversity, loss of specialists and community shifts over the past 28 years. *Insect Conserv. Diver.* **13**, 140–150. <https://doi.org/10.1111/icad.12411> (2020).
63. Martínez-Román, N., Epele, L. B., Manzo, L. M., Grech, M. G. & Archangelsky, M. Beetle mania: Understanding pond aquatic beetles diversity patterns through a multiple-facet approach. *Helicon.* **9**, e19666. <https://doi.org/10.1016/j.heliyon.2023.e19666> (2023).
64. Verberk, W.C.E.P., van Duinen, G.J.A., Peeters, T.M.J. & Esselink, H. Importance of variation in water-types for water beetle fauna (Coleoptera) in Korenburgerveen, a bog remnant in the Netherlands, in *Proceedings of the Section Experimental and Applied Entomology of the Netherlands Entomological Society (NEV)* (ed. Bruin J.), 12, 2002, Amsterdam, Netherlands. pp. 121–128 (2002).
65. Tokeshi, M. & Arakaki, S. Habitat complexity in aquatic systems: fractals and beyond. *Hydrobiologia* **685**, 27–47. <https://doi.org/10.1007/S10750-011-0832-Z> (2012).
66. Sheth, S.D., Padhye, A.D. & Ghate, H.V. Effect of environment on functional traits of co-occurring water beetles. *Ann. Limnol.- Int. J. Lim.* **57**, 2; doi.org/<https://doi.org/10.1051/limn/2020030> (2021).
67. Pakulnicka, J. et al. Relationships within aquatic beetle (Coleoptera) communities in the light of ecological theories. *Fund. Appl. Limnol.* **183**, 249–258; <http://dx.doi.org/https://doi.org/10.1127/1863-9135/2013/0413> (2013).
68. Frelik, A. & Pakulnicka, J. Relations between the structure of benthic macro-invertebrates and the composition of adult water beetle diets from the Dytiscidae family. *Environ. Entomol.* **44**, 1348–1357. <https://doi.org/10.1093/EE/NV113> (2015).
69. Frelik, A., Koszalka, J. & Pakulnicka, J. Trophic relations between adult water beetles from the Dytiscidae family and fly larvae from the Chironomidae family. *Biologia* **71**, 931–940. <https://doi.org/10.1515/BIOLOG-2016-0115> (2016).
70. Didham, R., Lawton, J., Hammond, P. & Eggleton, P. Trophic structure stability and extinction dynamics of beetles (coleoptera) in tropical forest fragments. *Philos. T. Roy. Soc. B* **353**, 437–451. <https://doi.org/10.1098/rstb.1998.0221> (1998).
71. Fujisawa, T., Vogler, A. & Baracough, T.G. Ecology has contrasting effects on genetic variation within species versus rates of molecular evolution across species in water beetles. *P. Roy. Soc. B-Biol. Sci.* **282**, 20142476; dx.doi.org/<https://doi.org/10.1098/rspb.2014.2476> (2015).
72. Thanee, I. & Phalaraksh, C. Diversity of Aquatic Insects and Their Functional Feeding Group from Anthropogenically Disturbed Streams in Mae Sot District, Tak Province. *Thailand. Chiang Mai J. Sci.* **39**, 399–409 (2012).

73. Klecka, J. & Boukal, D.S. Who eats whom in a pool? A comparative study of prey selectivity by predatory aquatic insects. *PLoS One* **7**, e37741; <https://doi.org/10.1371/journal.pone.0037741>. Epub 2012 Jun 5. PMID: 22679487; PMCID: PMC3367957 (2012).
74. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32; dx.doi.org/<https://doi.org/10.1023/A:1010933404324> (2001).
75. Zhang, Y. & Cheung, Y.-M. Discretizing Numerical Attributes in Decision Tree for Big Data Analysis. *IEEE International Conference on Data Mining Workshop, Shenzhen, China* **2014**, 1150–1157. <https://doi.org/10.1109/ICDMW.2014.103> (2014).
76. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
77. Kirpal, E. Ensembles and model stacking. Kaggle. <https://www.kaggle.com/eshaan90/ensembles-and-model-stacking> (2019).
78. Lundberg, S.M., Erion, G.G. & Lee, S. Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv*, 1802.03888 [cs.LG] <https://doi.org/10.48550/arXiv.1802.03888> (2018).

## Acknowledgements

The authors are grateful for the University of Warmia and Mazury in Olsztyn (Poland) for several own grants which allow the field studies and collection of unique biological material in the period 1997 – 2014 years.

## Author contributions

M.K. contributed to conceptualization, methodology (data base adaptation, machine learning modelling), software, formal analysis, writing—review and editing, and visualization. J.P. contributed to conceptualization, methodology (field studies, data base creation), biological interpretation of modelling results.

## Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-80083-0>.

**Correspondence** and requests for materials should be addressed to M.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024