

Applied machine learning in exploring key features of crayfish populations

Simona R. Đuretanović^{1*}, Marija M. Jakovljević¹, Aleksandra M. Milošković², Nataša M. Kojadinović¹, Milena D. Radenković¹, Vladica M. Simić¹, Ivana Maguire³

¹ University of Kragujevac, Faculty of Science, Department of Biology and Ecology, Radoja Domanovića 12, 34000 Kragujevac, Serbia; e-mail: simona.djuretanovic@pmf.kg.ac.rs,

marija.jakovljevic@pmf.kg.ac.rs, natasa.kojadinovic@pmf.kg.ac.rs,

milena.radenkovic@pmf.kg.ac.rs, vladica.simic@pmf.kg.ac.rs,

² University of Kragujevac, Institute for Information Technologies, Jovana Cvijića bb, 34000 Kragujevac, Serbia; e-mail: aleksandra@uni.kg.ac.rs

³ University of Zagreb, Faculty of Science, Department of Biology, Roosevelt trg 6, 10000 Zagreb, Croatia; e-mail: ivana.maguire@biol.pmf.hr

* Corresponding author

DOI: 10.46793/ICCB23.184DJ

Abstract: Uniform Manifold Approximation and Projection (UMAP) is a nonlinear dimension reduction technique based on manifold learning. It is specifically designed to achieve a balance between the global and local structure when embedding data points. We applied this method to morphometric features in populations of the noble crayfish, a freshwater species recognized as both a keystone species and an ecosystem engineer, as well as an indicator of good water quality, with unquestionable cultural and economic value to humans. Our results show that the CLL parameter most contributed to the differences and grouped the investigated specimens into seven clusters, along with ROL and ABL parameters. The parameters associated with the claws also exhibited a considerable influence on differentiation.

Keywords: machine learning, Uniform Manifold Approximation and Projection (UMAP), freshwater crayfish

1. Introduction

Freshwater crayfish are physiologically, ecologically, and behaviorally remarkable animals residing in a wide range of habitats [1]. They are a very important group of aquatic invertebrates, especially in terms of ecosystem integrity retention [2]. At the same time, these invertebrates are particularly sensitive to changes in aquatic ecosystems and are among the most vulnerable, with 32% of species listed as globally threatened by the IUCN [3, 4].

This study aims to deploy the Uniform Manifold Approximation and Projection approach with the Decision Tree algorithm on the noble crayfish *Astacus astacus*

(Linnaeus, 1758) from distinct populations to explore which morphometric parameters contribute most to overall differences.

2. Materials and Methods

A total of 192 noble crayfish from ten ecosystems in Serbia, Slovenia, and Albania were caught with baited LiNi traps or collected by hand. Each specimen was measured for 21 morphometric parameters (adopted from Sint et al. 2005 [5]) and weighed to the nearest 0.01g.

We used the unsupervised dimensionality reduction learning model UMAP (i.e., Uniform Manifold Approximation and Projection). The input matrix contained 192 rows (one row represented each measured specimen) and 21 columns (each column referred to one morphometric parameter). Prior to dimensional reduction modeling, the data were Hellinger-transformed [6, 7]. In the initial stage of machine learning, the weight of crayfish specimens was extracted and further excluded from the training set.

3. Results and discussion

The obtained results (Fig. 1 and Fig. 2) showed a clear differentiation between males and females. The Decision tree algorithm extracted a CLL parameter as a key feature for ordination, which most contributed to the differences, and grouped the investigated specimens into seven clusters. The parameters ROL and ABL also contribute significantly to differentiation. This is in accordance with a previous study by Đuretanović et al. (2017) [8].

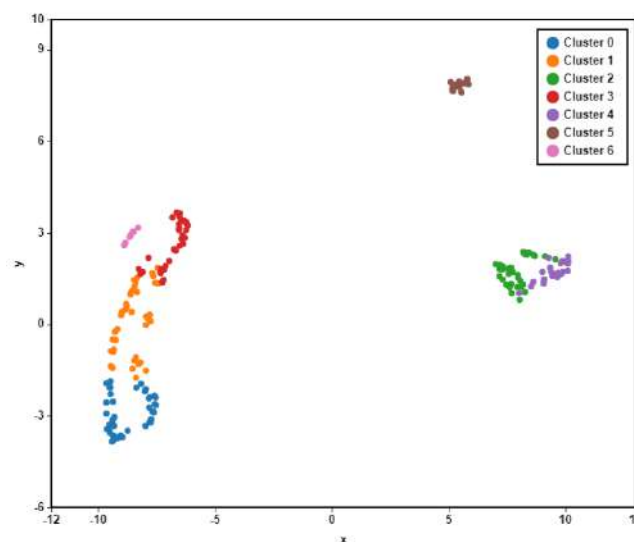


Figure 1. Ordination of morphometric parameters using the UMAP model. Colors stand for different seven clusters of parameters identified by the Decision Tree algorithm. Each point on the 2D UMAP space presents different specimens of crayfish.

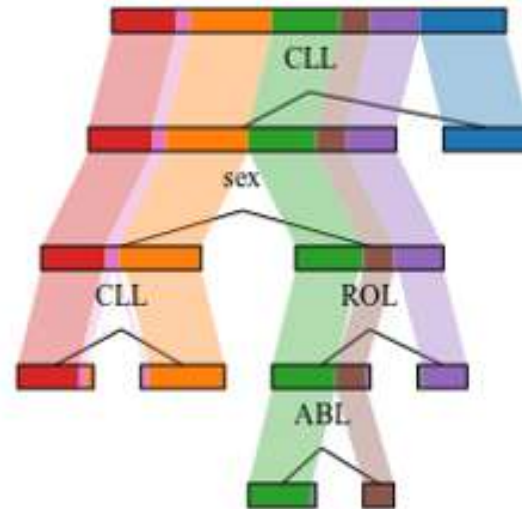


Figure 2. The Decision Tree algorithm was used to determine the most significant parameters for clustering.

Although the Decision Tree determines the key parameters (Fig. 2) it should be noted that the parameters related to the claws also contribute to the differentiation.

3. Conclusions

UMAP technique aims to retain as much variation from the original dataset as possible, surpassing the limitations of traditional reduction techniques. Data mining using the Decision Tree algorithm has proven to be an effective tool, demonstrating a highly predictive performance in exploring complex datasets and identifying representative variables for previously divided clusters.

Acknowledgment

This research was funded by the Serbian Ministry of Science, Technological Development and Innovation under grant no. 451-03-47/2023-01/200122.

References

- [1] J. Kubec, A. Kouba, M. Buřič., *Communication, behaviour, and decision making in crayfish: a review*, Zoologischer Anzeiger, 278 (2019) 28-37.
- [2] M.V. Alvanou, D.K. Papadopoulos, A. Lattos, I. Georgoulis, K. Feidantsis, A.P. Apostolidis, B. Michaelidis, I.A. Giantsis., *Biology, distribution, conservation status and stocking perspective of freshwater crayfish in Greece: An updated review*, Aquaculture Research, 53, 15 (2022) 5115-5128.
- [3] N.I. Richman, M. Böhm, S.B. Adams, F. Alvarez, E.A. Bergey, J.J. Bunn, Q. Burnham, J. Cordeiro, J. Coughran, K.A. Crandall, K.L. Dawkins, *Multiple drivers of decline in the global*

- status of freshwater crayfish (Decapoda: Astacidea)*, Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1662) (2015) 20140060.
- [4] L.M. Bland, *Global correlates of extinction risk in freshwater crayfish*, Animal Conservation, 20(6) (2017) 532–542.
- [5] D. Sint, J. Dalla Via, L. Füreder., *Morphological variations in Astacus astacus L. and Austropotamobius pallipes (Lereboullet) populations*, Bulletin Français de la Pêche et de la Pisciculture, 376-377 (2005) 637-652.
- [6] M. Jakovljević, M. Nikolić, N. Kojadinović, S. Đuretanović, M. Radenković, T. Veličković, V. Simić., *Population Characteristics of Spirlin Alburnoides bipunctatus (Bloch, 1782) in Serbia (Central Balkans): Implications for Conservation*, Diversity 15, (2023) 616.
<https://doi.org/10.3390/d15050616D>.
- [7] Dj. Milošević, A.S. Medeiros, M. Stojković Piperac, D. Cvijanović, J. Soininen, A. Milosavljević, B. Predić., *The application of Uniform Manifold Approximation and Projection (UMAP) for unconstrained ordination and classification of biological indicators in aquatic ecology*, Science of The Total Environment 815 (2022) 152365.
- [8] S. Đuretanović, M. Jaklič, A. Milošković, N. Radojković, M. Radenković, V. Simić, I. Maguire., *Morphometric variations among Astacus astacus populations from different regions of the Balkan Peninsula*, Zoomorphology, 136 (2017) 19-27.