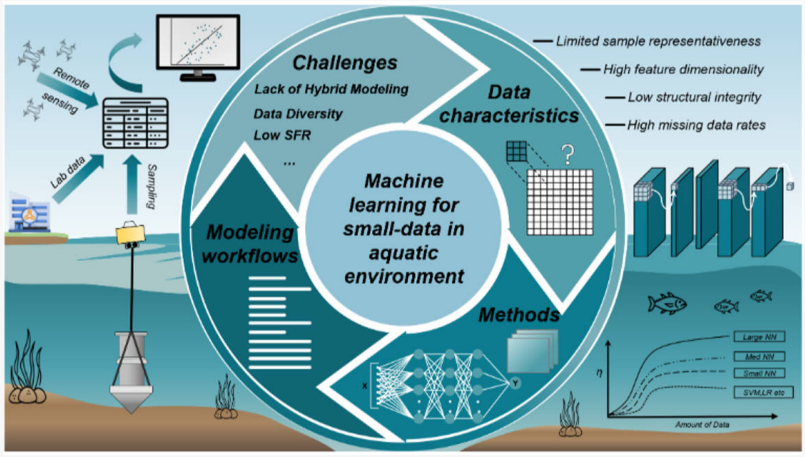
Machine learning tackles small-data challenges in aquatic environments

GA, UNITED STATES, April 21, 2026 /EINPresswire.com/ -- Machine learning (ML) has emerged as a promising tool for tackling challenges in aquatic environmental research, especially in small-data scenarios where limited sample sizes and high dimensionality typically hinder model performance. This review assesses current ML approaches applied to small datasets in aquatic systems, focusing on data characteristics, modeling strategies, and optimization frameworks. By examining the structural features of aquatic datasets, the review highlights key methodologies and presents solutions for improving predictive accuracy and robustness in data-scarce environments.



ML framework for small-data challenges in aquatic environments.

Aquatic environments are increasingly impacted by climate change and human activities, which introduce complex pollution sources and non-linear processes. Traditional modeling methods struggle to handle the high dimensionality and variability of environmental datasets. Machine learning (ML), with its capacity to identify patterns and interactions in large, complex datasets, offers a promising alternative. However, challenges remain, particularly in small-data contexts where the number of observations is limited, and the data exhibit structural inconsistencies. Based on these challenges, in-depth research is needed to develop models that can handle small-data issues effectively.

A review published (DOI: 10.1007/s11783-026-2186-9) in ENGINEERING Environment on March 17 2026, by researchers from Beijing University of Civil Engineering and Architecture and the Chinese Academy of Sciences explores how ML can be applied to small-data conditions in aquatic environments. The review systematically evaluates current approaches, comparing their adaptability and robustness in various aquatic research applications. The study provides insights into overcoming small-data limitations in environmental modeling, guiding future efforts in intelligent water governance and policy-making.

The study provides a detailed evaluation of ML techniques, particularly focusing on supervised, unsupervised, and deep learning methods. It highlights the challenges posed by high feature dimensionality, low sample sizes, and incomplete data often found in aquatic environmental research. The paper outlines several methodological advancements, including data augmentation and transfer learning, which have shown promise in overcoming the constraints of small datasets. The review emphasizes the importance of problem-oriented workflows tailored to aquatic systems and suggests that integrating data preprocessing, model construction, and evaluation can enhance the reliability of predictions. This holistic approach is vital for improving the robustness of ML models under small-data conditions.

Dr. Yulin Chen, one of the authors of the study, remarked, " ML has the potential to transform environmental modeling, particularly in areas where traditional methods have struggled. By addressing the challenges of small data, we can improve predictive models that support more informed, real-time decision-making in water management and environmental policy."

This research provides a foundation for developing more accurate and reliable ML models tailored to aquatic environmental monitoring. With applications ranging from water quality prediction to pollutant classification, the findings have significant implications for real-time environmental governance. The ability to accurately model aquatic systems, even with limited data, will be crucial for managing water resources and mitigating environmental risks. Future research will focus on refining these models, improving their scalability, and enhancing their interpretability to support informed policy decisions.

References
DOI
10.1007/s11783-026-2186-9

Original Source URL
https://doi.org/10.1007/s11783-026-2186-9

Funding information
This work was supported by the National Natural Science Foundation of China (No. 32530070), the International Partnership Program of the Chinese Academy of Sciences (No. 322GJHZ2022035MI) and STS Project of Fujian-CAS (No. 2023T3018).

Lucy Wang
BioDesign Research
email us here