

Review

Harnessing AI for smarter water management under a changing climate: A review of machine learning and deep learning applications within EU water framework directive and marine strategy framework directive

Angelica Bianconi^{a,b,c}, Elisa Furlan^{a,b,*}, Sebastiano Vascon^{b,d}, Andrea Critto^{a,b}

^a Centro Euro-Mediterraneo sui Cambiamenti Climatici, Risk Assessment and Adaptation Strategies Division, via Marco Biagi 5, Lecce, 73100, Italy

^b Department of Environmental Sciences, Informatics and Statistics, University Ca' Foscari Venice, I-30170, Venice, Italy

^c University School for Advanced Studies Pavia, Piazza della Vittoria 15, Pavia, 27100, Italy

^d European Center for Living Technology, Ca' Foscari University of Venice, Venice, Italy

ARTICLE INFO

Keywords:

Machine learning
Deep learning
Water framework directive
Marine strategy framework directive
Good environmental status
Good ecological status

ABSTRACT

The Marine Strategy Framework Directive (MSFD) and the Water Framework Directive (WFD) are key EU policies for safeguarding marine and freshwater ecosystems. They establish targets and evaluation frameworks to ensure good environmental and ecological status. This paper reviews the application of machine learning (ML) and deep learning (DL) in supporting these directives. A dual approach is used: a scientometric analysis to map the research landscape, followed by a systematic review of selected studies. The review focuses on three main areas: (i) use of spatio-temporal data, (ii) integration of climate-related scenario analysis, and (iii) application of ML and DL methods to enhance directive implementation. Results highlight the growing role of AI in water management, demonstrating its potential in handling large datasets, revealing patterns, and supporting predictive decision-making. Nonetheless, significant gaps remain. The underuse of spatio-temporal data hinders predictive accuracy, and scenario analysis methods often fail to capture the full complexity of climate impacts on aquatic systems. These limitations constrain the effectiveness of data-driven policy decisions. The paper calls for further research to better incorporate temporal dynamics in AI models and to refine scenario analysis tools. Such improvements are crucial for advancing adaptive, informed strategies in water resource management aligned with European directives.

1. Introduction

In pursuit of sustainable water management, the European Union has enacted directives that underscore the need for comprehensive strategies to protect and enhance water resources. In this setting, two key directives play a pivotal role: the Water Framework Directive (WFD) and the Marine Strategy Framework Directive (MSFD). These directives serve as fundamental frameworks that guide policies and actions aimed at safeguarding and improving the condition of both inland and marine water bodies. In particular, the WFD, established in 2000, provides a comprehensive approach to the management of surface waters, groundwater, and transitional waters within EU member states. It sets out objectives for achieving good ecological status for all water bodies, addressing issues such as pollution, over-abstraction and habitat

degradation. Furthermore, the directive emphasizes the importance of integrated river basin management and the participation of stakeholders in decision-making processes. Complementing the WFD, the MSFD, adopted in 2008, focuses on protecting the marine environment. It aims to achieve and maintain a good environmental status (GES) for European marine waters by 2020, promoting sustainable use and conservation of marine resources. The MSFD requires member states to develop marine strategies based on a preliminary assessment of the current status of their marine waters, and implement measures to address pressures, while ensuring that marine ecosystems remain healthy, productive, and resilient to environmental changes and human activities.

The intricate nature of aquatic ecosystems, coupled with the dynamic challenges posed by the complex interplay between anthropogenic activities and climate change-related pressures (Simeoni et al.,

* Corresponding author. Centro Euro-Mediterraneo sui Cambiamenti Climatici, Risk Assessment and Adaptation Strategies Division, via Marco Biagi 5, Lecce, 73100, Italy.

E-mail address: elisa.furlan@cmcc.it (E. Furlan).

<https://doi.org/10.1016/j.ocecoaman.2026.108093>

Received 8 August 2025; Received in revised form 18 December 2025; Accepted 3 January 2026

Available online 12 January 2026

0964-5691/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

2023), underscores the need for innovative approaches to effectively implement these directives. In this regard, leveraging the increasing accessibility of geospatial big data from various sources, including satellite imagery, aerial photography, geotagged social media data, Internet of Things (IoT) devices (Crampton et al., 2013; Loor and Fdez-Arroyabe, 2018; Tsou, 2015; Vo et al., 2021), has become imperative. Moreover, machine learning (ML)-based methods autonomously learn from vast datasets, enabling them to develop robust predictive capabilities and uncover patterns, correlations, and anomalies within the data. In light of these considerations, ML-based techniques offer a promising avenue for addressing the complexity of water management challenges. In environmental contexts, including assessments of ecosystem health (Kwon et al., 2024), water quality evaluation (Zhu et al., 2022), and the provision of ecosystem services (Almeida et al., 2024; Scowen et al., 2021; Willcock et al., 2018), ML and Deep Learning (DL) models, trained on geospatial datasets, are becoming increasingly prevalent and effective, revolutionizing environmental assessment practices.

This review article aims to provide a comprehensive analysis of the state-of-the-art literature, applying both scientometric and systematic review approaches, to investigate the role of ML and DL techniques in advancing the implementation of European Directives on Water Management and monitoring (i.e., WFD and MSFD). By synthesizing findings from an increasing number of research papers published between 2000 and 2025, this work underscores the applicability and effectiveness of ML and DL in various tasks, such as predicting water quality parameters, detecting eutrophication levels, monitoring habitat changes and other critical factors. Utilizing advanced algorithms and computational models, these methodologies have demonstrated the potential to transform raw sensor data into actionable information, thereby facilitating proactive decision-making and enabling timely responses to emerging challenges in aquatic ecosystem management (Deng et al., 2024; Gamb'in et al., 2021; Mart'inez-Santos et al., 2021). These advancements enhance traditional approaches by offering predictive capabilities that are invaluable for environmental management, empowering policymakers and researchers to work towards achieving good ecological and environmental status. In addition, to highlighting the primary methodologies used, this review offers insights into spatiotemporal data modelling and the development of climate scenario analysis methodologies. Both the WFD and the MSFD acknowledge the importance of considering climate-related pressures when assessing the status of aquatic and marine environments. While the WFD does not explicitly mention climate change, official guidance and implementation reports from Member States have highlighted the need to consider climate-related pressures in the assessment and management of freshwater ecosystems (Commission and for Environment, 2024). The MSFD, in contrast, explicitly identifies climate change as a pressure on marine ecosystems and encourages Member States to consider its effects when developing marine strategies (Directive, 2008/56/EC, Recitals 34, 42). For these reasons, it is essential to focus on how climate-driven alterations interact with water quality and ecosystem dynamics within the policy frameworks established by both directives. Understanding how artificial intelligence may support and enhance the assessment of water quality and the ecological status of ecosystems under changing climate conditions and extreme events, compared to conventional statistical and physical models, is crucial. Through a critical assessment of the current state of the art, key challenges have been identified, and innovative frameworks have been proposed to enhance the reliability and interpretability of AI in climate and environmental modelling. In particular, these frameworks include approaches for the analysis of spatiotemporal data, leveraging deep learning methods such as Convolutional Neural Network, Transformers, Graph Neural Network and Physics-Informed Neural Networks to capture not only complex spatial and temporal dependencies but also environmental and physical processes. These contributions provide a broader knowledge for understanding the dynamic and complex nature of water management, thereby supporting the

creation of more effective and sustainable management strategies. The paper is divided into four main sections. Section 2 provides a methodological overview of the data acquisition process for the literature review. In Section 3, key findings extracted from ML and DL applications aligned with the EU directive on water management, are presented. Section 4 engages in a focused discussion, highlighting the main strengths and limitations of the analysed studies, while proposing valuable insights for future research and applications. Finally, the main conclusions are summarized in Section 5.

2. Data acquisition and review methods

A comprehensive systematic review of the literature was conducted to gain insight into the current state of the art in scientific studies and applications of ML and DL methods to support the implementation of EU directives on water management. As illustrated in Fig. 1 the methodological approach includes three primary phases: i) data collection, ii) scientometric analysis, and iii) systematic review, as detailed in Sections 2.1 and 2.2 below.

2.1. Data collection and paper screening

In this study, the selection process to identify relevant peer-reviewed literature followed a structured approach guided by the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) method (Moher et al., 2009). The PRISMA method is a well-established and widely recognized framework for conducting systematic reviews, providing a systematic and transparent process to minimize bias and ensure the robustness of the process. To initiate the selection procedure, a series of search queries were formulated, and the open-source Scopus database (<https://www.scopus.com/>) was used to retrieve articles published between 2000 and 2025. These queries were meticulously designed to capture publications related to ML and DL within the context of the EU WFD and MSFD. An iterative process of refinement and expansion was also employed to ensure comprehensive coverage of the relevant literature, and to enhance the robustness of the review results. The formulation involved combining related keywords using appropriate Boolean operators (that is, 'AND', 'OR', 'NOT'). The initial query focused on retrieving papers relevant to the intersection of ML and DL techniques with key EU directives on waters, as well as concepts related to ecological and environmental status. Specifically, the query was set as follows: (('Machine learning' OR 'Deep learning' learning') AND ("Marine strategy framework directive" OR "Water Framework Directive" OR "MSFD" OR "WFD" OR "ecological status" OR "environmental status")), resulting in 147 articles. By refining the search to include additional terms such as "environmental quality" and "ecological quality" (key targets of the WFD and MSFD), the scope was broadened to encompass a wider array of literature that may contribute to the understanding of environmental assessment and management. This includes improving the evaluation of aquatic ecosystem conditions, applying ML and DL techniques to analyse environmental data, assessing water quality, predicting ecological trends, and supporting decision-making processes in alignment with these EU directives. The refined search ensures a more comprehensive review of studies that explore innovative approaches for monitoring, assessing, and managing the health of marine and freshwater environments. This modification resulted in 555 papers as outputs, reflecting a wider range of literature.

Furthermore, the expansion of synonymous terms for ML and DL, including "artificial intelligence" and "neural network," aimed to capture a comprehensive range of studies exploring advanced computational methods in the context of environmental policy and conservation. The final query is thus presented with this structure: (('Machine learning' OR "Deep learning" OR "Artificial intelligence" OR "Neural network") AND ("Marine strategy framework directive" OR "Water framework directive" OR "MSFD" OR "WFD" OR "ecological status" OR "environmental status" OR "environmental quality" OR "ecological quality")),

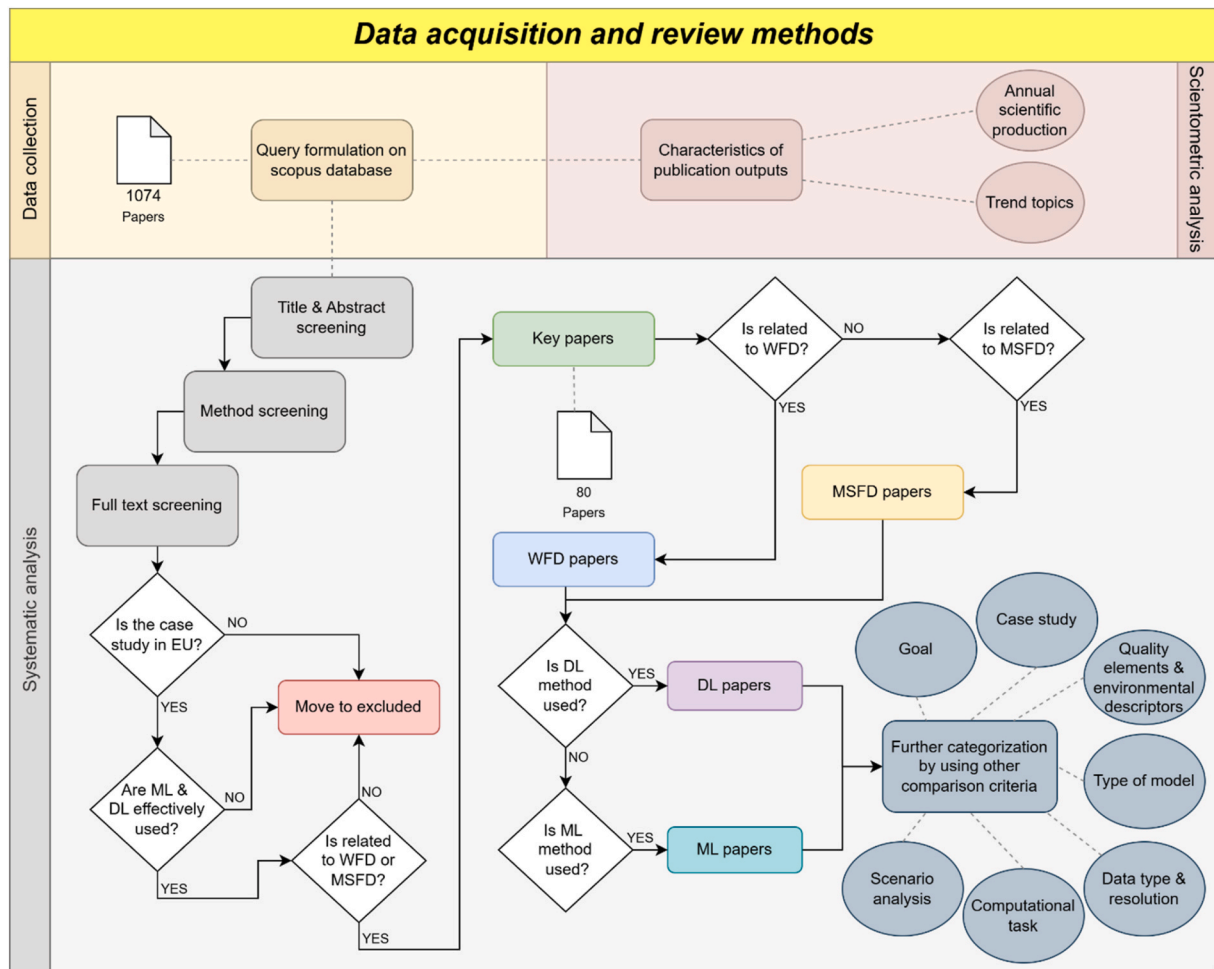


Fig. 1. Methodological approach for the evaluation of existing studies dealing with ML and DL applications to support EU directives on Waters.

yielding a total of 1074 papers. They were exported as a csv file to enable qualitative and quantitative analyses. To perform these analyses, the use of the Bibliometrix R Package (Aria and Cuccurullo, 2017; Mingers and Leydesdorff, 2015), an analytical tool highly suitable for bibliometric studies, facilitated a comprehensive examination of the literature, including how frequently articles were cited, which authors and institutions contributed most to the field and other relevant bibliographic information. Within the third step of the proposed review approach, the PRISMA methodology was used to conduct systematic screening. Initially, considering titles and abstracts, the screening process involved both the exclusion of those studies that were not directly relevant to the objective of the review, and the classification of the selected ones based on eligibility criteria. For example, at this stage, studies that either did not effectively apply ML or DL techniques, used the acronyms "MSFD" or "WFD" to refer to unrelated concepts to the marine or water frameworks, or focused on topics outside the scope, such as air quality and indoor environment, were excluded from the review. Regarding the eligibility criteria guiding the review, they aligned with the following research questions:

are ML and DL techniques used to: i) monitor water and ecosystems? ii) Assess the level of eutrophication? iii) evaluate the ecological and environmental status of water? iv) estimate biodiversity? v) Support the assessment and modelling of WFD quality elements and MSFD environmental descriptors?

Firstly, the question of "monitoring water and ecosystems" is crucial, as it addresses the core purpose of environmental monitoring, which is to track changes in water bodies and surrounding ecosystems (Lovett et al., 2007). ML and DL techniques can offer promising avenues for automating data analysis processes and detecting patterns indicative of environmental changes, enhancing predictive accuracy while enabling real-time monitoring, and facilitating large-scale data processing for more efficient and proactive environmental management. Similarly, the assessment of "the level of eutrophication" is critical due to the significant environmental impacts associated with nutrient pollution in water bodies (Smith et al., 1999). Furthermore, "evaluating the ecological and environmental status of water" encompasses a broader perspective, focusing on overall ecosystem health and environmental quality. This aligns with the overarching objectives of the WFD and MSFD, which are oriented towards achieving good ecological and environmental status of waters. Estimating "biodiversity" is essential for understanding ecosystem dynamics and identifying conservation priorities, and ML and DL approaches can help analyse species distribution patterns, habitat suitability, and other biodiversity-related data to assess the richness and abundance of species within aquatic environments. Lastly, "supporting the assessment and modelling of WFD quality elements and MSFD environmental descriptors" highlights the relevance of ML and DL techniques in regulatory frameworks such as the WFD and MSFD. In fact, clarifying the centrality of these techniques in fulfilling the objectives of

these directives is the main objective of this study. After selection based on these eligibility criteria, 80 articles are ready to be categorized according to the comparison criteria presented in Section 2.2.

2.2. Support for the implementation of EU Water directives

To effectively investigate how ML and DL techniques may support the implementation of the investigated EU directives and the achievement of their goals, an in-depth analysis was performed through a series of comparison criteria applied to the spectrum of selected key studies. This thorough investigation aimed at establishing a deeper understanding of the various methodologies applied in distinct case studies, and how these methodologies interact with diverse data types. It explored how structured and unstructured data (e.g., in situ measurements, spatial information, time-series data) are utilized within different analytical frameworks. Accordingly, the comparison criteria used for this systematic analysis are included in the following table (Table 1).

In summary, these comparison criteria provide a detailed framework for evaluating the application of AI techniques in support of WFD and MSFD. They also offer a solid foundation for subsequent discussions and analyses. The results derived from this meticulous examination will be elaborated upon and critically discussed in the following sections (Section 3 and Section 4), presenting valuable insights and contributing to the advancement of knowledge in this research field.

3. Results of the review

3.1. Characteristics of the publications

The use of scientometric analysis in this study facilitated the extraction and examination of bibliometric data derived from the initial set of 1047 selected articles. Through the implementation of the Open-Source Bibliometrix R Package, graphical representations were generated, allowing for the visualization and interpretation of the dataset. The analysis of the annual evolution of scientific production that spans from 2000 to 2025 revealed noteworthy trends (Fig. 2). Specifically, a concentrated surge in scientific output was observed within the last five years, with the peak occurring in the 2022 and 2024. This pattern underscores a significant increase in research activity within the specified time frame. Advancements in computational technology and the widespread availability of sophisticated ML.

and DL algorithms have facilitated the application of these techniques to address complex environmental challenges. This trend is further corroborated by the most frequent topics observed over the years. In recent years, there has been a notable prevalence of keywords such as "machine learning", "deep learning," "artificial neural networks" and "remote sensing" in scientific studies (Supplementary Material SM2). This trend underscores the growing emphasis on these methodologies, which can be attributed, in part, to the exponential expansion in the accessibility of open-source satellite data. The proliferation of such

Table 1
Comparison criteria and descriptions.

Criterion	Description
 Aim of the Work	Defines the primary objective and the main research questions of the study, providing clarity on the overall goal pursued and the scientific contributions
 Supported Directive	Identifies alignment with either the WFD or MSFD directives
 Case Study	Refers to the geographic area analysed in the study, particularly the specific regions delineated within the directives. The WFD case studies focus on eco-regions (as depicted in Supplementary Material SM1) and specific water types (surface waters, groundwaters, protected areas). The MSFD case studies examine four major marine regions: Baltic Sea; Mediterranean Sea (Western, Adriatic, Ionian, Central Mediterranean, Aegean-Levantine); North-East Atlantic Ocean (Greater North Sea, Celtic Seas, Bay of Biscay, Iberian Coast, Macaronesia); and Black Sea.
 Quality Elements & Environmental Descriptors	Identifies the quality elements and environmental descriptors utilized as specific parameters, indicators or targets. WFD elements include biological elements (e.g., phytoplankton biomass, fish fauna), chemical and physico-chemical parameters (e.g., nutrient concentrations, oxygen levels), hydromorphological elements (e.g., flow conditions, river continuity), and general variables (e.g., transparency, acidification). MSFD descriptors cover key marine habitat aspects, such as biodiversity, fish populations, eutrophication, sea floor integrity, and contaminants.
 Data Type & Resolution	Describes data sources, thresholds, acquisition methods, and spatial/temporal resolution
 Type of Algorithms	Differentiates between ML and DL models, specifies the model used and, if multiple models are used for comparison, specifies the best one.
 ML/DL Task	Specifies computational tasks (classification, regression, clustering)
 Integration of Scenario Analysis	Specifies whether a scenario analysis is conducted, either scenarios of future climate change, what-if or management

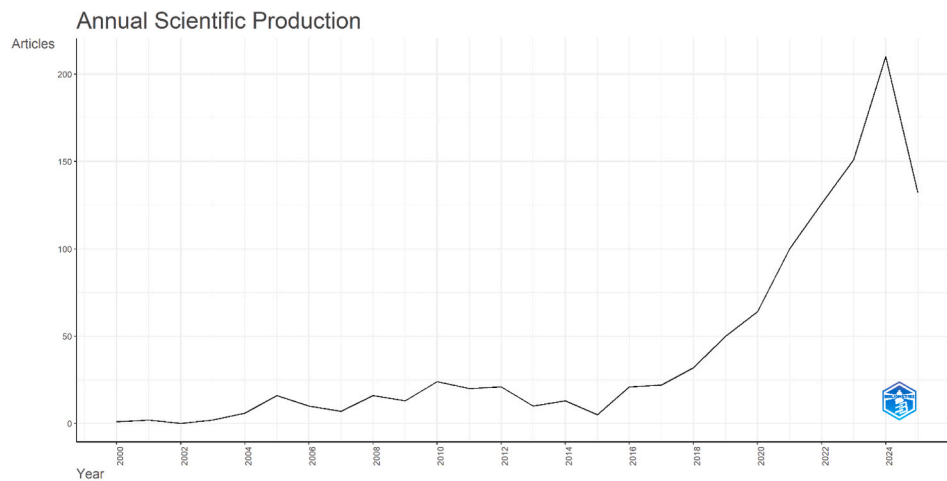


Fig. 2. Number of publications applying ML and DL methods to support WFD and MSFD in the 2000–2025 timeframe. Please note that the data for 2025 only includes analyses up to July 2025; therefore, the apparent drop is due to the partial coverage of the year.

data sources has provided researchers with unprecedented access to vast amounts of information, facilitating the application of ML, DL, and ANN techniques for environmental analysis. In addition, the scalability and flexibility of these methodologies make them well suited for processing large-scale satellite data sets and extracting valuable insights for water management and marine conservation.

3.2. Machine and deep learning support to achieve good ecological and environmental status

The systematic review, which includes 80 articles as delineated in the Supplementary Materials SM3, culminates in the articulation of

principal findings within this section. Specifically, it highlights the predominant techniques employed, their applications across different domains, and the role of scenario analysis in assessing model reliability. These insights provide a comprehensive understanding of the field's advancements and are visually summarized in the Sankey diagram in Fig. 3. The Sankey diagram is used as a synthetic and integrative visualization tool, allowing multiple dimensions of the reviewed literature to be explored simultaneously. In particular, it enables the reader to rapidly identify how study objectives, investigated quality elements or descriptors, algorithmic approaches (i.e., ML or DL), and specific models are interconnected across the two regulatory frameworks (i.e., WFD and MSFD), which would be less intuitive if presented only in tabular or

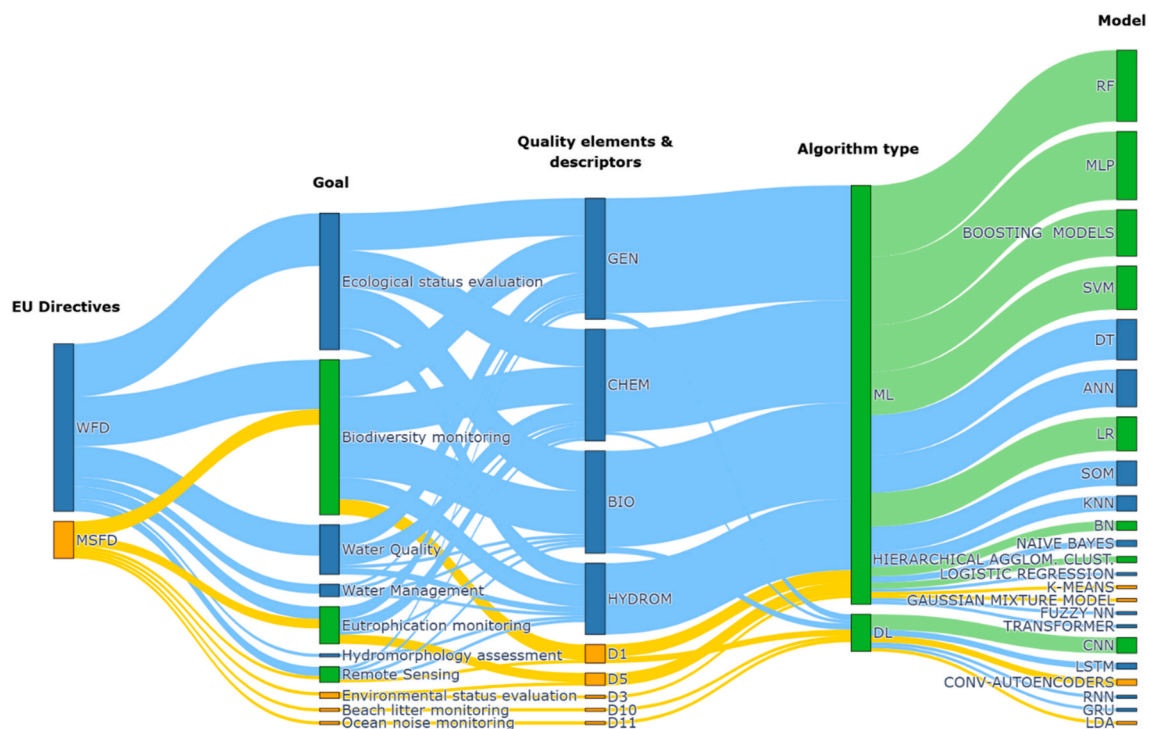


Fig. 3. Sankey diagram summarizing for the WFD and MSFD the objectives of the studies, the quality elements and descriptors investigated, the type of algorithm (whether ML or DL), and the models applied. The directives are marked by two colors (blue, WFD, and yellow, MSFD). Green bands indicate elements (goals, algorithms, or models) that are common to both directives. In particular, the green color appears when a specific model is used in studies addressing both directives — for example, Random Forest (RF) is applied in studies concerning both WFD and MSFD, hence the green band. In contrast, Decision Trees (DT) are used only in WFD-related studies, and are therefore represented with a blue band.

narrative form. A pronounced focus is directed towards the utilization of diverse modeling methodologies and the investigation of scenario analyses, elucidated prominently in Sections 3.2.1 and 3.2.2. Before entering a detailed discussion of these themes, the Sankey diagram offers an immediate overview of the distribution of research efforts and methodological choices across the two Directives. Both the relative width of the flows and the associated colors allow the reader to promptly identify dominant patterns and shared elements. It becomes readily discernible that a predominant proportion of the reviewed papers employ ML or DL methodologies, primarily in support of the WFD. Specifically, 79 % of papers is categorized as addressing WFD, with the remaining 21 % attributed to the MSFD. This imbalance is highlighted not only by the width of the flows but also by the color distribution, with blue (WFD) bands clearly prevailing over yellow (MSFD) ones. At the same time, green bands identify objectives, algorithms, or models that are common to both Directives, providing an immediate visual reference of methodological or thematic convergence. The relatively limited representation of studies pertaining to the MSFD directive can be attributed to the complexity and multidimensional nature of marine ecosystems, rendering analysis under the MSFD directive inherently more challenging and resource intensive. Furthermore, variations in data availability and accessibility may contribute to the observed asymmetry in research focus between the two directives. In terms of regionalization, the most extensively investigated marine area is the Atlantic Ocean, followed by the Mediterranean Sea and the Baltic Sea, highlighting a notable gap in research on the Black Sea region.

As regard the inland waters, the areas comprising Iceland, the tundra, and corresponding to regions 19, 20, 21, 22, and 23 on the map provided in the [Supplementary Material SM1](#), remain largely unexplored. Moreover, surface waters, such as rivers, lakes, and coastal waters, represent the most investigated aquatic environments, covering a significant proportion (i.e. 98 %) of total research efforts. This distribution underscores the focus on areas where human activities and environmental pressures are more pronounced, reflecting a strategic approach to addressing critical ecological and management concerns. The analysed research activities have utilized AI methodologies to address a spectrum of objectives, encompassing biodiversity surveillance, assessment of ecological and environmental status, management of hydrological resources, monitoring of eutrophication phenomena, evaluation of hydro morphological characteristics, application of remote sensing techniques for detecting aquatic vegetation and biomass and identifying irrigated areas, water quality appraisal, observation of beach litter accumulation, and monitoring of oceanic acoustic disturbances. The Sankey diagram supports this analysis by visually linking each objective to the corresponding quality elements and algorithmic approaches, thereby facilitating a rapid comparison of research priorities between WFD and MSFD studies. It should be noted that biodiversity monitoring has become a prevalent focus, representing the most frequently addressed attention in studies related to both the WFD and MSFD directives. (Gur'ı et al., 2023; Lanzen et al., 2021; Marre et al., 2020; Martin-Abadal et al., 2020; Segurado et al., 2018; Kraan et al., 2024; Rakaj et al., 2025; Madhuanand et al., 2025). This is clearly reflected in the Sankey diagram by green bands associated with biodiversity monitoring, indicating that this objective is addressed under both regulatory frameworks. The use of green thus provides an immediate visual indication of shared research priorities across Directives, complementing the detailed quantitative discussion in the text. This emphasis reflects the central role of biodiversity-related components within both regulatory frameworks considered in this review. Under the WFD, biodiversity is addressed through biological quality elements used to assess the ecological status of water bodies, whereas under the MSFD biodiversity is explicitly linked to the structure, functions, and processes of marine ecosystems, as well as to their capacity to maintain resilience to human-induced environmental change. At the same time, eutrophication monitoring emerges as another significant research objective, reflecting concerns about nutrient enrichment and its detrimental

impacts on aquatic ecosystems. In this setting, within the realm of papers addressing the WFD, the predominant objective revolves around the evaluation of ecological status through indicators such as chlorophyll-a and phytoplankton density, as proxies for eutrophication (Cherif et al., 2021; Chrobak et al., 2021; Concepcion et al., 2021; Gebler et al., 2021; Tamvakis et al., 2012; Volf et al., 2023; Zennaro et al., 2023a), diatom index as proxy for monitoring the ecological status of freshwater (Apothélos-Perret-Gentil et al., 2021a; Carafa et al., 2011; Chrobak et al., 2021; Feio et al., 2020a; Szczepocka et al., 2018; Tison et al., 2007; Venkataramanan et al., 2023), the macrophyte and the macroinvertebrate indices as indicators evaluating aquatic plant communities and water quality (Gebler et al., 2014, 2017, 2018; Najafzadeh et al., 2024; Segurado et al., 2018b; Dedeker et al., 2004; Di Dato et al., 2005; Gur'ı et al., 2023b; Lock et al., 2014a; Riabchenko et al., 2016; Valerio et al., 2021b). This priority stems from the WFD's explicit mandate to assess and improve the ecological condition of water bodies, with indicators serving as proxies for ecological health. The choice to focus on indicator-based assessments aligns with the directive's regulatory framework, which emphasizes the need for standardized and comparable metrics to evaluate environmental quality across Member States. In contrast, only a single document among those addressing the MSFD (Uusitalo et al., 2022a) is dedicated to evaluating the environmental status, which stands as the directive's primary objective. With the introduction of the MSFD, new thematic areas have emerged as key research priorities, driven by the directive's descriptors. While initial studies under the MSFD primarily focused on biodiversity and eutrophication monitoring, their scope has since expanded to include emerging concerns such as the observation of beach litter and ocean noise (Sabara et al., 2020a; Scarrica et al., 2022a). This broadening of focus reflects evolving environmental challenges and underscores the increasing recognition of the multifaceted threats facing aquatic ecosystems, beyond traditional parameters such as water quality and biodiversity. The relatively early stage of research in these areas is a direct response to the introduction of new MSFD descriptors, which drive scientists to focus on understanding the impact of these new stressors. However, a lack of studies related to several MSFD descriptors emerges. Specifically, there is a deficiency of studies on Descriptor D2 on exotic species, D4 on food webs, D7 on hydrographical conditions, D8 on hazardous substances, and D9 on hazardous substances in fish. This gap can be attributed to the methodological limitations, difficulties in data collection, particularly for marine environments, where the research community has only recently begun to systematically collect and monitor ocean-related data (Le Traon et al., 2019). Given that ML and, more specifically, DL methodologies are driven by big data, it becomes almost obvious that there are limited studies available, as comprehensive datasets may only recently have become accessible or adequately compiled. Regarding WFD, all quality elements are taken into consideration as indicators. A comprehensive approach is employed to assess the ecological status of water bodies, using a broad range of indicators to capture various aspects of environmental health. The integration of these elements within the WFD reflects a structured methodology for evaluating freshwater and transitional water ecosystems, complementing the broader scope of the MSFD in marine environments.

3.2.1. Machine learning and deep learning methods and approaches

Among the methodologies employed in the examined studies, the Random Forest (RF) algorithm (Breiman, 2001) stands out as a preeminent choice due to its ease of implementation and its ability to conduct comprehensive assessments of variable importance. This attribute is particularly advantageous in environmental studies, where discerning the most influential factors shaping outcomes is essential. For example, in Segurado et al. (2018)'s investigation, the main objective is to evaluate the significance of variables pertaining to specific biotic indicators, with a focus on assessing the cumulative impacts of pressures on biotic organisms within the Sorraia basin in Portugal. To address this objective, RF is used with the permutation importance methodology to rank

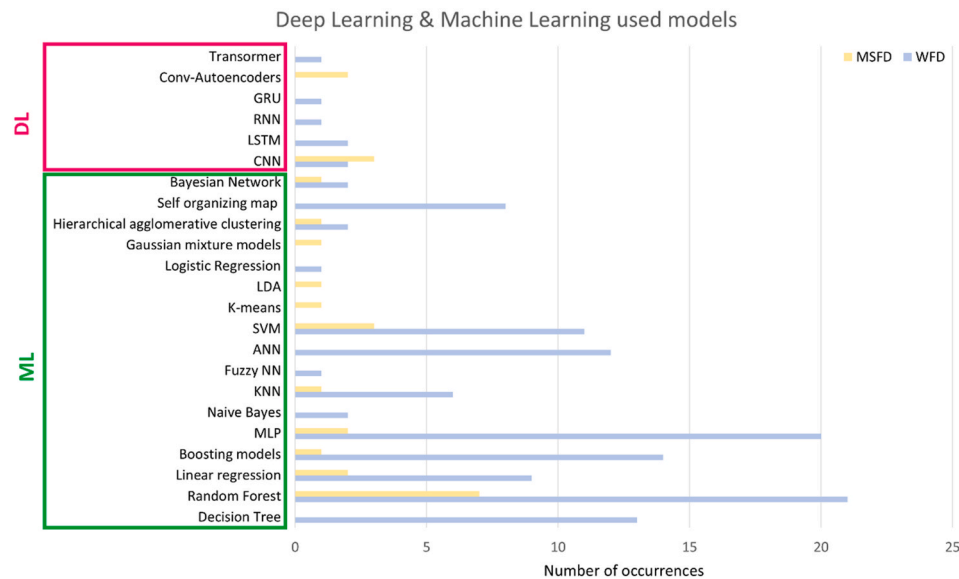


Fig. 4. Models used. Light blue bars represent models that support WFD, while yellow indicates those for MSFD. deep learning models are highlighted in magenta, while machine learning models in green.

the importance of features. Similarly, other studies (Ficsor and Csabai, 2023; Govedarica and Jakovljević, 2019; Knoll et al., 2020a; Koit et al., 2023; Leontidou et al., 2023; Segura-Méndez et al., 2023; Valerio et al., 2021a; Visser et al., 2022a) employ RF in conjunction with the permutation importance technique or with other methods not specified (e.g., **mean decrease in impurity, mean decrease in accuracy, Gini importance**) to evaluate variable importance. RF is frequently evaluated against alternative models (Arrighi and Castelli, 2023; Govedarica and Jakovljević, 2019; Sabara et al., 2020; Valerio et al., 2021; Visser et al., 2022; Yu et al., 2021), consistently demonstrating superior performance. Moreover, the multilayer perceptron (MLP) (Hastie et al., 2009) finds extensive application within this domain because of its capability to capture nonlinear relationships within datasets. This capacity accommodates the intricate interdependencies often present in the data, which in turn represent the interdependencies among various environmental variables and processes, rendering MLP an attractive option for modelling complex ecological systems. Most studies employing MLP include sensitivity analysis in the methodological approach (Dedecker et al., 2004; Gebler et al., 2014, 2017, 2021; Knoll et al., 2019; Olaya-Marín et al., 2012a; Tison et al., 2007), thus underscoring the importance of understanding the impact of individual variables on outcomes in the environmental context. It is particularly crucial when using models that, despite their ability to capture nonlinear relationships, behaves as a black box, as it is essential to understand the underlying reasons for a given outcome in order to interpret and justify the model's decisions. The bar chart in Fig. 4 emphasize the preference for decision tree (DT) (Rokach and Maimon, 2005) models, notably in the context of the WFD.

DTs are favoured because of their inherent interpretability, which allows researchers to gain insight into the decision-making processes underpinning environmental phenomena. This transparency proves invaluable in environmental research, where understanding the underlying mechanisms driving ecological dynamics is paramount to formulate effective management strategies. In addition to DTs and RF, boosting models, particularly Extreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016), are also among the most frequently employed algorithms in the context of the WFD. XGBoost combines the strengths of decision trees with boosting techniques, offering both high predictive performance and an interpretable structure, especially when coupled with explainability frameworks such as SHAP (SHapley Additive exPlanations). For instance (Politikos et al., 2024), applied an

XGBoost classifier to predict the ecological quality of Greek rivers based on a comprehensive dataset including ecological, physicochemical, geomorphological, and climatic variables. The SHAP method was used to quantify the contribution of each predictor to the model's output, revealing, among others, the critical role of nutrients in influencing ecological status, and highlighting geomorphological factors (e.g., altitude and slope) as key predictors for fish-based assessments. This approach not only provided accurate predictions but also enabled the development of a web-based what-if simulation tool to support water managers in evaluating alternative environmental scenarios. The synergy between XGBoost and SHAP thus exemplifies how advanced ML techniques can simultaneously enhance model accuracy and transparency, addressing both predictive and interpretative needs in support of the WFD. A similar use of boosting techniques is presented by (Martyszunis et al., 2024), who applied several machine-learning classifiers, including XGBoost, to assess the ecological status of unmonitored river water bodies in Poland. In contrast to approaches relying on direct ecological observations, their models were trained using anthropogenic pressure data and catchment characteristics as predictors. Among all tested algorithms, XGBoost together with RF achieved the best performance, particularly in the binary classification distinguishing water bodies with at least good ecological status from those below good status, reaching an overall accuracy of approximately 93–94 %. Model performance was evaluated not only through standard classification metrics (i.e., accuracy, precision, recall, F1-score), but also using the Probability of Misclassification (PoM), explicitly introduced to quantify classification uncertainty and support regulatory decision-making. The low PoM values obtained for XGBoost highlight its robustness and suitability for WFD reporting, reinforcing evidence that boosting-based models represent a reliable and cost-effective solution for large-scale ecological status assessment where monitoring data are scarce. Fig. 4 also illustrates that Kohonen artificial neural networks, commonly known as self-organizing maps (SOMs) (Vrieze, 2005), are frequently employed in collected studies. As unsupervised learning algorithms, SOMs excel at organizing large and heterogeneous datasets without the need for labelled examples, allowing exploration of data structure and pattern identification. Moreover, SOMs facilitate dimensionality reduction in high-dimensional environmental datasets, while preserving essential relationships between variables, thereby enabling visualization and interpretation of complex data. By transforming high-dimensional ecological data into a more manageable form, SOMs facilitate the

identification of clusters or groups of similar data points, allowing for the recognition of spatial and thematic patterns in ecological datasets. Furthermore, SOM-generated visualizations, such as heatmaps and cluster maps, enhance the ability to interpret ecological patterns and communicate findings effectively (Carafa et al., 2011; Chrobak et al., 2021a,b; Heasley et al., 2020; Park et al., 2005; Sroczynska et al., 2017; Szczepocka et al., 2018; Tison et al., 2007). Considering the spatio-temporal nature of the data, Aslan et al. (2022) explores the application of Recurrent Neural Networks (RNNs) for water quality assessment in the Venice Lagoon, a complex coastal environment. RNNs (Medsker and Jain, 2001) are a type of artificial neural network that can effectively model sequential data, making them suitable for analysing time-series water quality data, which often exhibit temporal dependencies and nonlinear patterns. The research focuses on using RNNs to predict water quality parameters such as chlorophyll-a concentration and dissolved oxygen levels in the Venice Lagoon. By training the RNNs on historical water quality data and environmental variables, the study aims to develop accurate predictive models that can help monitoring and manage water quality in the lagoon. The RNN models were also coupled with Convolutional Neural Networks (CNNs) to create hybrid models capable of capturing the structural information of the features, while simultaneously learning the temporal dependencies in the sequence. These hybrid models generally achieved the best performance, even compared to RF. CNNs (Raitoharju, 2022) have also been used in other studies (Marre et al., 2020b; Martin-Abadal et al., 2020b; Riabchenko et al., 2016; Scarrica et al., 2022b; Venkataramanan et al., 2023) for object detection in the monitoring of jellyfish and other aquatic species, demonstrating their effectiveness in analysing diverse data types ranging from images to videos. More recently, advanced deep learning architectures such as Transformers (Han et al., 2022) and LSTM (Y. Yu et al., 2019) networks have been explored in recent applications addressing ecological status modelling within the WFD framework. In particular (Krtolica et al., 2025), developed artificial neural networks (ANNs), including Transformer and LSTM models, to predict the ecological impact of xenobiotics on macroinvertebrate communities in the Danube River. Their study highlighted that these architectures outperformed traditional models in accurately predicting the concentration of specific contaminants (e.g., Fluoranthene, 2,4-Dinitrophenol), with LSTM achieving the lowest mean squared error (MSE) in several cases. These results indicate that, in the context of this specific application, deep learning models proved particularly effective in capturing intricate nonlinear relationships between chemical pollutants and biological responses, thus offering a promising direction for future ecological modelling and water quality management under the WFD.

3.2.2. Machine learning and deep learning task: classification, regression and clustering

The systematic analysis of the literature shows that ML and DL applications supporting the WFD and the MSFD are mainly structured around three core computational tasks: classification, regression and clustering. Among the analysed studies, classification is the most common task (40 %), followed by regression (32 %), while clustering approaches remain limited (7 %). Model comparison is performed in 31 % of the studies, with Random Forest being by far the most frequently identified best-performing algorithm, followed by boosting methods. Studies combining DL and traditional ML are rare (i.e., 2) and are exclusively regression-oriented. DL approaches are almost exclusively applied to image-based data (e.g., satellite imagery, UAV images, underwater photography) and to high-frequency time series (hourly to daily observations). In contrast, traditional ML methods dominate studies based on point-based in situ measurements, administratively reported pressures or spatially and temporally aggregated indicators.

Within the WFD context, classification is widely used to assign water bodies to discrete ecological or environmental status classes (e.g., high, good, moderate, poor, bad), directly aligning with the normative assessment schemes of the directive (Apothélos-Perret-Gentil et al.,

2021b; Etchells et al., 2006; Martyszunis et al., 2024; Ocampo-Duque et al., 2007). Algorithms such as RF, DT, SVM and boosting-based models (e.g. XGBoost) are frequently employed. In the MSFD domain, DL classifiers, especially CNNs, are widely used for image-, video- and signal-based applications, such as biodiversity monitoring, beach litter detection, and underwater acoustic classification (Marre et al., 2020c; Martin-Abadal et al., 2020c; Sabara et al., 2020c; Scarrica et al., 2022c). Model performance in classification tasks is predominantly evaluated using overall accuracy, precision, recall and F1-score, which are particularly relevant in imbalanced datasets, a common condition in environmental monitoring. Confusion matrices are routinely reported to analyse misclassification patterns among ecological status classes, providing insights into systematic biases (e.g. confusion between “good”, “moderate” and “high” status) (Chrobak et al., 2021). From a policy perspective, correctly classifying threshold exceedances or regulatory boundaries is often more critical than marginal improvements in global accuracy, reinforcing the importance of class-specific performance metrics. The probability of misclassification (PoM) can provide additional insight, serving both as an indicator of potential errors in water quality assessment and as a measure of model uncertainty, which is crucial for management and regulatory decisions (Martyszunis et al., 2024).

Regression tasks are employed when predicting continuous environmental variables such as nutrient concentrations, chlorophyll-a, dissolved oxygen, suspended sediment, or biological indices. These models are particularly useful for trend analyses, early-warning systems, and scenario-based assessments under climate change or management interventions. Within the WFD framework, regression is often used to estimate biological quality elements (e.g., diatom, macroinvertebrate, or macrophyte indices) from physicochemical, hydromorphological, and climatic predictors, capturing complex nonlinear relationships (Feio et al., 2020b; Gebler et al., 2018b; Olaya-Marín et al., 2012b). In marine and transitional waters, regression supports the quantification of eutrophication indicators and oxygen dynamics, aligning with both WFD and MSFD objectives (de Luca Lopes de Amorim et al., 2021; Yan et al., 2021b). Performance evaluation relies on error-based metrics such as RMSE, MAE, and MSE. Regression models are essential for studies integrating ML/DL with climate or management scenarios, allowing direct quantification of expected changes in water quality and ecological indicators.

Clustering and, more generally, unsupervised learning approaches are less frequently adopted than supervised tasks, but they play a strategic role in data exploration and system understanding. These methods are primarily used to identify homogeneous groups of water bodies, monitoring stations, or environmental conditions without imposing predefined classes, thus supporting typology definition and pattern discovery. SOMs are the most commonly applied clustering technique in the reviewed literature, due to their ability to handle high-dimensional environmental datasets and to provide intuitive visual representations of variable relationships. SOMs are used to identify gradients of ecological quality, group sites according to dominant pressures, and explore spatial and temporal structures in complex datasets (Heasley et al., 2020b; Kruk et al., 2016; Sroczynska et al., 2017b). Unlike supervised tasks, clustering performance is rarely assessed using standard quantitative metrics. Instead, evaluation relies on qualitative interpretation, visual inspection, and the consistency of identified clusters with known environmental gradients or expert knowledge. Despite this, clustering provides essential support for monitoring network design, identification of environmental hotspots, and the development of differentiated management strategies.

A recurring feature across the reviewed literature is the systematic comparison of multiple ML and DL algorithms applied to the same problem. Model benchmarking is a common methodological step to identify the most suitable algorithm for a given task, dataset, and management objective. This is particularly important in WFD and MSFD contexts, where robustness, interpretability, and generalization are as

crucial as predictive accuracy. For tabular environmental datasets from monitoring stations (physicochemical, biological, and hydro-morphological indicators), RF emerges as the best-performing algorithm in roughly 50 % of studies where formal comparisons are conducted. This trend is consistently reported across WFD-oriented ecological status classification, water quality assessment (e.g. (Arrighi and Castelli, 2023; Martyszunis et al., 2024; Zennaro et al., 2023; 2024), groundwater quality (Knoll et al., 2019b, 2020b; Sajib et al., 2025), and broader comparative reviews of predictive tools (Visser et al., 2022b). RF often outperforms or matches alternative algorithms in terms of accuracy, F1-score, and error metrics, while demonstrating greater stability across water bodies and ecoregions. Its strong performance is generally attributed to its ability to model nonlinear relationships, robustness to multicollinearity and noisy predictors, and suitability for small-to-medium-sized datasets typical of regulatory monitoring programs. Additionally, RF provides variable importance measures that enhance interpretability and facilitate links between model outputs and policy-relevant indicators. Boosting-based algorithms, such as XGBoost, are increasingly included in comparative analyses and, in some cases, achieve performances comparable to or slightly superior to RF (Martyszunis et al., 2024; Sajib et al., 2025). The prevalence of algorithm comparisons reflects the recognition that no single model is universally optimal; selection should be guided by data characteristics, task requirements, and policy relevance. In DL-focused studies, RF is often absent from comparisons due to the nature of the data (images, time series) rather than inherent limitations, further emphasizing that model performance is task- and data-dependent. The prevalence of algorithm comparison reflects an increasing methodological maturity within the field. While no single algorithm is universally optimal, the cumulative evidence indicates that RF represents a reliable baseline and often a reference standard for supervised learning tasks on environmental monitoring data, whereas DL approaches provide clear advantages in image-based and complex spatiotemporal scenarios.

3.2.3. Scenario analysis

Scenario analysis serves as a crucial tool in environmental research, offering information on the potential impacts of varying conditions on ecological systems. In this context, five distinct studies have conducted scenario analysis to explore the responses of biological indices to changing environmental variables. One of these studies, conducted by Valerio et al. (2021) focuses on predicting biological indices under different environmental indicators, while evaluating the sensitivity of water body status to various scenarios. The methodological framework integrates the selection of the best algorithm from RF and Gradient Boosting Regression Trees (GBRT), previously trained on a selected set of environmental variables. Variations in nutrient concentrations are considered, resulting in nine different ‘*what-if*’ scenarios. In particular, the nutrient metrics used for scenario analysis are grounded both in Spanish regulatory frameworks and percentile distributions of nutrient values observed across other European Union member states.

In the study by Olaya-Marín et al. (2012), the best-performing ML model (i.e., MLP) was used to assess the potential effects of river restoration interventions on indigenous fish populations. The authors implemented a set of ‘*what-if*’ scenarios by modifying key predictive variables related to river morphology, such as the proportion of riffles relative to river length and the length of channel segments without artificial barriers. One of these scenarios involved the simulated removal of dams within a selected river reach, a restoration measure commonly adopted to improve river ecosystem connectivity in Europe. This intervention was expected to alter hydromorphological features, particularly by increasing the ratio of riffles to river length. However, because no hydraulic model was available to explicitly simulate the resulting physical changes, the authors did not model hydrodynamics directly. Instead, they assessed the ecological response by evaluating how changes in the riffle-to-length ratio—derived from observed data collected along the studied river stretch—would affect native fish

species richness. The primary objective is to discern the potential positive impact of such restoration measures on the biodiversity of native fish species. Continuing the investigation into ‘*what-if*’ scenarios, two similar studies (Lock et al., 2014; Lock and Goethals, 2014) are dedicated to predicting the presence or absence of stoneflies in Flemish watercourses. In contrast to the methodological approach outlined in the aforementioned studies, these investigations diverge in the role of the ML model. In these studies, the ML model itself does not directly contribute to the scenario analysis. Instead, its output data serve as input for the hydrodynamic model (i.e., PEGASE water quality model), used to forecast potential enhancements in water quality. This integrated model encompasses hydrological, hydrodynamic, thermal, and biological sub-models. Within this framework, two different scenarios were tested, the first one implementing traditional policies and proposed measures, while the second one incorporating additional restoration efforts. To construct these hypothetical scenarios, various techniques were applied to assess factors such as oxygen levels, biological oxygen demand, and nutrient concentrations. In another study (Molina et al., 2013), Bayesian network analysis was employed to evaluate groundwater status under three different water management scenarios. These scenarios provide unique perspectives for analysis: the Business As Usual (BAU) scenario, the Maximum Agricultural Net Margin scenario, and the Compliance of WFD Nitrate Concentrations scenario. Each scenario represents different management strategies and their potential impact on groundwater quality. The BAU scenario reflects the historical average situation, serving as a baseline for comparison. Conversely, the maximum agricultural net profit scenario aims to maximize agricultural profitability, while the compliance scenario of WFD nitrate concentrations prioritizes meeting regulatory standards. Results of the assessment indicate varying levels of nitrate concentration and recovery time horizons in each scenario, highlighting the complex interplay between management practices and groundwater quality. Another study (Uusitalo et al., 2022b) employed Bayesian network to develop a comprehensive model suite to assess a variety of climate, nutrient loading, and fishing scenarios in the Baltic Sea. This innovative approach integrates climate projections, nutrient load scenarios, and fisheries management strategies to assess potential ecosystem responses. By combining Representative Concentration Pathways (RCP) with nutrient load scenarios according to reference conditions of the Baltic Sea Action Plan (BSAP), along with different fishing strategies, the study generated twelve distinct scenarios. These scenarios offer insight into how changes in climate, nutrient input, and fishing practices may influence the ecological dynamics of the Baltic Sea. From sustainable fishing practices to unregulated fishing scenarios, each scenario represents a unique path for ecosystem management and conservation efforts in the region. More recently, scenario-based analysis has also been employed in coastal lagoon systems. In particular, two studies have explored the Venice Lagoon under future climate conditions using a hybrid approach that combines ML techniques with deterministic models. In the first, Zennaro et al. (2023b) developed a modelling framework to simulate chlorophyll-a dynamics under future climate scenarios by integrating ML predictions with outputs from a high-resolution, 3D biogeochemical model of the lagoon. Specifically, the deterministic model was used to generate a comprehensive synthetic dataset by simulating a wide range of hydrometeorological and nutrient conditions, including future variations in temperature, salinity and nutrient inputs. ML algorithms were then trained to reproduce chlorophyll-a concentrations using this dataset, enabling rapid predictions across a wide parameter space. Once validated, the ML models were coupled with bias-corrected climate projections from CMIP5 models under RCP4.5 and RCP8.5 scenarios, allowing the generation of robust long-term scenarios linking future hydro-climatic conditions to trophic state dynamics. This approach significantly reduced computational demands while preserving the system's complexity and enabled the identification of seasonal patterns and threshold behaviours associated with eutrophication risk in response to climate forcing. In the second study (Zennaro et al., 2024),

assessed future hypoxia events in the lagoon by combining ML techniques with deterministic simulations of hydrodynamics and oxygen dynamics. Their ML models (i.e., RF, AdaBoost, and ensemble approaches) successfully classified historical hypoxia events (2008–2019) with high accuracy using water quality, meteorological, and spatio-temporal data. Integrating these models with climate projections under the RCP 8.5 scenario, they predicted a substantial increase in hypoxia frequency by 2100, with events rising from 3.5 % to 8.8 % of days annually, particularly in landward lagoon areas characterized by longer water residence times. A similar approach was adopted by (Bezák et al., 2025), who applied the XGBoost algorithm to predict suspended sediment concentrations in the Sora River catchment (Slovenia) under future climate scenarios. By combining ML predictions with outputs from a hydrological rainfall–runoff model (CemaNeige GR6J), the study projected a 15–20 % increase in sediment load under RCP4.5 and RCP8.5 scenarios by the end of the century. Additionally, the frequency of days exceeding the critical sediment concentration threshold (25 mg/L), a key water quality indicator, is expected to rise by up to 8 %. The study demonstrates the potential of ML not only to improve predictive accuracy over traditional sediment rating curves, but also to inform sediment and erosion management strategies in light of climate-driven hydromorphological changes.

4. Discussion

The results of this review have provided insights into state-of-the-art literature on ML and DL models applications for supporting WFD and MSFD directives. This section offers a reflection on different aspects of this specific field of research. Specifically, based on the information extracted from the 80 selected articles (reported in SM3 of the Supplementary Material), this section discusses the potential, limitations, and barriers of the analysed models and their applications, providing recommendations for future research and improvements.

4.1. Can machine learning and deep learning really support the achievement of good environmental and ecological status?

ML and DL are at the forefront of technological advancements, offering significant potential to aid and enhance environmental and ecological status. In this review, their practical utility in supporting the implementation of both directives and achieving good environmental and ecological status (GES), is evidenced by numerous applications across different domains. One of the primary ways ML and DL can support GES is through improved monitoring and predictive modelling capabilities. Traditional environmental monitoring methods often face limitations in scale, frequency, and detail. However, ML and DL algorithms, particularly those leveraging large datasets from satellite imagery and remote sensing technologies, can overcome these limitations (Falanga Bolognesi et al., 2020; Huber et al., 2022; Karki et al., 2021; Oiry and Barillé, 2021). Biodiversity conservation is another critical area where ML and DL can make a substantial impact. Species distribution data used with ML techniques can predict the geographic distribution of species based on environmental variables, thereby identifying critical habitats and prioritizing areas for conservation (Gur'i et al., 2023b; Lanzen et al., 2021b; Venkataramanan et al., 2023). For example, ML algorithms can process large volumes of camera trap images and videos to identify and count marine species, reducing the labour-intensive effort required for field activities (Marre et al., 2020b; Martin-Abadal et al., 2020b). This helps in tracking population trends and assessing the effectiveness of conservation efforts. Additionally, DL models analysing underwater acoustic data (Sabara et al., 2020c) can detect and monitor various phenomena, critical to marine science and safety. By leveraging oceanic noise, these models can identify and track changes in marine biodiversity, assess environmental health, and detect natural hazards like tsunamis and underwater volcanic activity. This capability supports sustainable resource management, enhances

maritime safety through the monitoring of ship activities, and provides early warnings for natural disasters, thereby contributing significantly to both scientific research and practical applications in marine conservation and safety protocols. ML and DL also play a significant role in pollution control and waste management. Predictive models can forecast water quality levels based on a range of indicators such as nutrient concentrations, chemical pollutants, and biological markers. These models analyse vast datasets collected from remote sensing technologies, including satellite imagery and underwater sensors, to provide high quality assessments and predictions. Additionally, these models support waste management efforts by optimizing the detection and removal of marine debris and monitoring the effectiveness of pollution control strategies over time (Scarrica et al., 2022b). The integration of growing and increasingly diverse datasets into ML and DL models is therefore a key enabling step within a broader decision-support pipeline, rather than a standalone solution for achieving GES. ML- and DL-based models contribute by processing complex spatiotemporal information, extracting patterns, and generating forecasts and scenarios that can inform planners, managers, and decision-makers. By continuously incorporating new and diverse data sources, these models can enhance their predictive accuracy, detect emerging environmental threats more effectively, and provide deeper insights into ecosystem dynamics. These capabilities strengthen the evidence base available to water managers and policymakers, enabling more precise monitoring, early warning systems, and targeted intervention strategies that support sustainable water management and compliance with environmental regulations. This involves several key steps to optimise data utilization and improve model performance. For example, pre-processing is essential to transform the raw data into a format suitable for ML and DL models. This includes cleaning the data, handling missing values, normalising variables, selecting relevant features and creating synthetic datasets. Adequate pre-processing ensures that the data used in the models are accurate and reliable, which is crucial to produce meaningful results that can be effectively translated into management-relevant information and to demonstrate substantial potential in supporting the achievement of good environmental and ecological status. When embedded within science-informed and transparent decision-making frameworks, the outputs of ML and DL models support the design, prioritisation and evaluation of management measures, including monitoring programmes, protection and restoration actions, and the management of pollution sources. Indeed, their applications in monitoring, predictive modelling, biodiversity conservation and water quality control provide valuable insights and facilitate more effective environmental management strategies. As technology advances, ongoing research and development will continue to unlock new possibilities, further reinforcing the role of ML and DL as decision-support tools, driving progress towards sustainable environmental and ecological outcomes.

4.2. Harnessing advanced deep learning to capture spatiotemporal dynamics

The analysis of spatiotemporal data has gained increasing importance in the big data era with the increasing availability and importance of large spatio-temporal datasets such as maps, virtual globes, remote-sensing images, decennial census and GPS trajectories. Spatiotemporal modelling has broad applications in various research fields such as geography, climatology, biology, and social sciences. The evolution of models for analysing these data has been marked by progressive complexity and sophistication, addressing the need to accurately represent natural phenomena (Wang et al., 2020). In the selected studies, methods involving the computation of temporal metrics (Dézerald et al., 2020; Valerio et al., 2021b) and the use of SOMs to cluster similar environmental states and visualize the spatiotemporal evolution of phenomena have been employed (Carafa et al., 2011; Chrobak et al., 2021a,b; Sroczyńska et al., 2017; Szczepocka et al., 2018). These methods provide foundational insights into aquatic

monitoring but often fall short in handling large-scale, complex datasets that exhibit non-linear interactions over space and time. For instance, temporal metrics (e.g., averages, medians, maximums and minimums calculated over time), usually computed to produce the training data for baseline ML methods, can highlight changes over extended periods, but they may obscure critical short-term variations and events that are vital for environmental monitoring and management. Environmental systems are highly dynamic, and short-term events such as chemical spills or extreme weather events can have a profound impact. Averaging data over long periods can attenuate these signals, delaying detection of such events and the subsequent response. Simple models often require processes like calculating averages, which may fail to capture sudden, important changes that require immediate attention. In this setting, DL methods, recognized as powerful tools for modelling large datasets and complex spatiotemporal phenomena, offer significant advancements over traditional approaches. Firstly, DL models excel in automatic feature representation learning (Zhong et al., 2016). Unlike traditional methods, DL models can autonomously learn hierarchical feature representations from raw data. This capability is crucial for spatiotemporal data, where spatial proximity and long-term temporal correlations are intricate and challenging to capture. Moreover, DL algorithms can approximate any complex non-linear function given sufficient layers and neurons. Each layer in a DL model acts as a simple yet non-linear module, employing pooling, dropout, and activation functions to transform lower-level latent feature representations into more abstract higher-level representations. This hierarchical transformation enables DL models to learn and represent highly complex relationships within spatiotemporal data, making them well-suited for challenging spatio-temporal modelling tasks. Additionally, DL models demonstrate superior performance with large datasets. Traditional ML methods typically plateau in performance when encountering large-scale data, with minimal performance improvements as data size increases beyond a certain point. In contrast, the performance of DL methods continues to improve with the addition of more data, utilizing the available information more effectively to enhance overall performance (Wang et al., 2020). However, the adoption of DL methods in the analysed works has been limited. The combination of CNNs with RNNs has been employed in only one environmental study to capture spatiotemporal relations in the data (Aslan et al., 2022). CNNs are particularly effective in extracting spatial features from high-resolution environmental data by applying convolutional filters, while RNNs, especially Long Short-Term Memory (LSTM) networks, excel at capturing temporal dependencies by maintaining a memory of previous states, making them suitable for time series analysis. Another study (Bezák et al., 2025), however, adopts LSTM in comparison with Transformer architectures, highlighting the strengths and limitations of each model in handling temporal patterns in environmental datasets. Despite the potential of DL methods, several challenges hinder their widespread adoption in environmental spatio-temporal modelling. Key issues include data availability and quality, as high-resolution, comprehensive spatio-temporal datasets are often limited or incomplete, and the variability in data quality and coverage can impact the performance of DL models. Additionally, developing and tuning DL models requires specialized knowledge in both DL techniques and the environmental context under investigation, and the scarcity of interdisciplinary expertise can slow adoption. To address these challenges, several strategies can be employed. Data augmentation and synthesis can help fill gaps and enhance training datasets, improving model robustness and performance. Moreover, multi-modal data integration, such as combining satellite imagery with ground sensor data, can provide a richer, more comprehensive view of environmental phenomena. For instance, to mitigate the challenge of data scarcity in environmental sciences, specific constraints can be imposed (e.g., governing equations, boundary conditions) to guide the model learning more efficiently and effectively during training. This approach not only mitigates the lack of training data but also enhances the generalization capability of models and ensures the physical

plausibility of the results. This is the concept behind Physics Informed Neural Networks (PINNs), which incorporate prior knowledge of the physical laws governing the data into the learning process. PINNs overcome the issue of limited data availability (Elabid et al., 2022) and have proven to be effective and popular in geophysics (Cuomo et al., 2022; Donnelly et al., 2024; Figueredo et al., 2025; Jia et al., 2025; Omarova et al., 2023; Schuster et al., 2024). Concurrently, a more innovative approach to spatiotemporal modelling involves exploring and adopting new DL architectures, such as Graph Neural Networks (GNNs) (Gori et al., 2005), which can more effectively model the relational aspects in spatiotemporal data. GNNs can capture interactions between spatial entities over time, offering deeper insights into complex environmental processes. Although partially adopted in environmental contexts (Bianconi et al., 2025; Cachay et al., 2021; Dun et al., 2022; Lira et al., 2022; Ni et al., 2022; Oliveira Santos et al., 2023; Ouyang et al., 2021; Wang et al., 2020; Yang et al., 2022; X. Yu et al., 2021) GNNs have vast potential due to their ability to map datasets as graphs, which can represent both spatiotemporal dynamics and environmental relationships and synergies. Looking ahead, the field of spatiotemporal modelling in environmental sciences stands to benefit significantly from ongoing advancements in DL. Innovations in model architectures, data integration techniques, and computational efficiency will drive progress. By addressing current challenges and embracing new opportunities, more accurate and effective models can be developed, advancing efforts towards achieving sustainable environmental results.

4.3. Unveiling blind spots: gaps and constraints in scenario analysis

Despite the transformative potential of ML and DL models in various analytical domains, as emerged through the systematic review, their application for future climate scenario modelling (under both mid and long terms) is lacking. Scenario analysis often requires a thorough understanding of the underlying mechanisms and interactions within the model, as well as the broader system being studied. Achieving this level of understanding can require additional data collection, model validation and calibration efforts, which could be perceived as resource-intensive or technically challenging. Moreover, the interpretation and communication of scenario analysis results can be complex and nuanced. Presenting and explaining the implications of various scenarios effectively may require advanced statistical or visualization techniques, along with interdisciplinary collaboration to ensure accurate and meaningful interpretation (Tourki et al., 2013). In addition to the challenges outlined above, another significant issue that can arise in scenario analysis is the potential for inconsistency between training data used by ML models and the scenario data used for the analysis. Scenario data often represent hypothetical future conditions or alternative pathways based on assumptions about various factors such as population growth, technological development, and policy interventions (Pielke Jr and Ritchie, 2021). These assumptions may differ from the historical data used to train ML models, creating inconsistencies between the two datasets. For instance, due to the inherent uncertainty in scenario modelling, future scenario data may include values that also appear in historical records, despite representing projections rather than past observations. Similarly, scenario data might exhibit trends or relationships that contradict those found in historical data, further complicating model adaptation. Since ML algorithms rely on patterns and relationships learned from historical data to make predictions or classifications, when scenario data diverge significantly from training data, they may struggle to generalise effectively to new conditions, leading to unreliable or misleading results. This inconsistency can be a significant barrier to using future climate change scenario data in combination with ML techniques, exacerbating the challenges of uncertainty and complexity inherent in scenario analysis. Addressing the challenge of inconsistency between training and scenario data requires careful consideration and integration of both datasets throughout the modelling process. Researchers may need to develop innovative

approaches to adapt ML models to scenario conditions, such as incorporating scenario assumptions directly into the model architecture and input data. In this setting, a promising approach is the use of generative AI as a support tool for scenario analysis. Generative AI can significantly enhance the creation and communication of future climate change scenarios, providing robust support for understanding and mitigating the impacts of climate change. By leveraging large-scale datasets and advanced ML architectures, generative AI enables researchers and policymakers to explore alternative futures, assess uncertainties, and improve decision-making processes. For example, in the realm of timeseries data, Garza and Mergenthaler-Canseco (Garza and Mergenthaler-Canseco, 2023) introduced TimeGPT, a groundbreaking model for forecasting time series. It leverages efficient transfer learning to deliver accurate zero-shot predictions on previously unseen datasets. This advanced framework, built on the principles of large language models (LLMs), enables probabilistic forecasting and anomaly detection without the need for extensive dataset-specific training. This is particularly advantageous when dealing with data poor case studies. However, TimeGPT's current capabilities do not extend to spatio-temporal data, which are critical in environmental and climate sciences. Despite this limitation, TimeGPT and similar tools represent a promising avenue for generating data on future climate change scenarios. By exploiting forecasting capabilities, these tools can potentially simulate future environmental conditions consistent with historical time series data and, thus, could be used to model different climate scenarios, providing valuable insights into potential future changes and trends, which is crucial for developing adaptation strategies and policies. Beyond TimeGPT, generative AI has demonstrated considerable potential across various climate research domains. Deep generative models, such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), are being increasingly utilized to simulate high-resolution climate projections. These models enable the downscaling of coarse-grained climate data, producing more detailed regional forecasts that help communities better prepare for localized climate impacts (Harris et al., 2022; Rampal et al., 2025). Additionally, diffusion models, which draw inspiration from AI-generated imagery advancements, are being explored to generate synthetic climate data. These models can fill observational gaps and enhance historical climate reconstructions (Meuer et al., 2024; Sousa et al., 2024). Furthermore, generative AI is transforming the communication of climate scenarios to policymakers and the public. Natural language generation (NLG) models, such as ChatGPT, facilitate the translation of complex climate data into accessible reports, policy briefs, and interactive narratives. AI-generated visualizations and climate storytelling tools further enhance public engagement by presenting potential future scenarios in an intuitive and impactful way (Rane, 2023; Rane et al., 2024). The integration of generative AI into environmental research remains an area with considerable untapped potential. The adaptability of generative AI, including its ability to be fine-tuned for specific environmental research tasks and the use of retrieval-augmented generation (RAG), is crucial for earth science applications. Incorporating generative AI and RAG systems enhances efficiency, minimizes errors, and ensures more consistent results by automating repetitive tasks and mitigating the risk of AI-generated inaccuracies (Tishechkin et al., 2024). Looking ahead, the integration of generative AI into climate science is poised to evolve, offering innovative approaches for scenario modelling, risk assessment, and policy formulation. By leveraging AI's capabilities, the understanding of future climate conditions can be enhanced and more effective strategies for mitigation and adaptation in an increasingly uncertain world can emerge. Moreover, incorporating generative AI into the co-development of climate scenarios has the potential to strengthen the credibility and acceptance of ML-driven analyses. This collaborative approach will not only improve the reliability of climate projections but also ensure that the resulting insights are more robust, actionable, and aligned with diverse stakeholders' needs, paving the way for more insightful, inclusive, and effective climate solutions.

5. Conclusion

This review comprehensively examines the state-of-the-art techniques in ML and DL employed to support European water directives, providing a comprehensive analysis blending scientometric and systematic reviews of the existing literature over the 2000–2023 time-frame. Among the papers resulting from the various queries, through a tiered selection approach based on the PRISMA method, 80 articles were identified as representative of ML- and DL-based applications supporting the implementation of the EU framework directives on waters. Over this period, since the introduction of these directives, ML techniques have predominantly been utilized, showcasing their versatility across various domains. Specifically, these techniques have been fundamental in monitoring water quality and biodiversity in inland water bodies and marine environments. Despite the abundance of literature focused on freshwaters, there is a significant lack of studies related to marine contexts, highlighting a critical gap in research that deserves to be further explored and investigated. In addition, while basic ML algorithms (e.g., RF, MLP, DT) are widely used, DL techniques are still little employed in these contexts. Drawing on these outputs, this review identified key challenges that need to be addressed in the future to provide more accurate guidance to policy makers for sustainable water management. Firstly, the potential of spatio-temporal data is not fully exploited. Current methods based on temporal metrics often fail to capture the inherent non-linearity and complex correlations within time series data. Adopting tailored DL methodologies for spatio-temporal data could significantly enhance predictive capabilities and result accuracy. The limitation of comprehensive time series data, a common barrier, can be mitigated by generating synthetic datasets to augment model training or by using models (PINNs) that can guide learning by taking into account the physical laws governing the data. Another critical gap is the inadequate analysis of future scenarios, which are essential for effective water planning and management. Enhancing the integration of projected datasets into predictive models can yield more informed and actionable insights. This forward-looking approach is crucial for anticipating and mitigating potential water management challenges. Despite challenges, such as data availability and the need for interdisciplinary expertise, both basic and advanced AI techniques are widely adopted to support compliance with European environmental directives. Regulations such as the WFD and the MSFD require continuous monitoring and assessment of water quality, pollution levels, and ecosystem health. AI-driven models help achieve these objectives by automating data analysis, supporting the systematic assessment of environmental information and enabling real-time monitoring of aquatic ecosystems. Moreover, AI serves as a valuable tool for the assessment and prediction of water management issues, offering a robust framework for analysing complex environmental systems, supporting more sustainable and informed water management practices.

CRedit authorship contribution statement

Angelica Bianconi: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Elisa Furlan:** Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Sebastiano Vascon:** Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Andrea Critto:** Validation, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This article was produced while attending the PhD program in Sustainable Development And Climate Change at the University School for Advanced Studies IUSS Pavia, Cycle XXXVIII, with the support of a scholarship financed by the Ministerial Decree no. 351 of 9th April 2022, based on the NRRP - funded by the European Union - NextGenerationEU - Mission 4 "Education and Research", Component 1 "Enhancement of the offer of educational services: from nurseries to universities" - Investment 3.4 "Advanced teaching and university skills" OR Investment 4.1 "Extension of the number of research doctorates and innovative doctorates for public administration and cultural heritage". This work has been partly funded by the European Union's Horizon Europe programmes under grant agreement No. 101112972 as part of the DesirMED project, and No. 101135343 as part of the TRANSEATION project. The authors gratefully acknowledge the support of the DoE 2023–2027 (MUR, AIS.DIP.ECCELENZA2023_27.FF project).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ocecoaman.2026.108093>.

Data availability

No data was used for the research described in the article.

References

- Almeida, B., David, J., Campos, F.S., Cabral, P., 2024. Satellite-based machine learning modelling of ecosystem services indicators: a review and meta-analysis. *Appl. Geogr.* 165, 103249.
- Apothéloz-Perret-Gentil, L., Bouchez, A., Cordier, T., Cordonier, A., Guéguen, J., Rimet, F., Vasselton, V., Pawlowski, J., 2021a. Monitoring the ecological status of rivers with diatom eDNA metabarcoding: a comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. *Mol. Ecol.* 30 (13), 2959–2968.
- Apothéloz-Perret-Gentil, L., Bouchez, A., Cordier, T., Cordonier, A., Guéguen, J., Rimet, F., Vasselton, V., Pawlowski, J., 2021b. Monitoring the ecological status of rivers with diatom eDNA metabarcoding: a comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. *Mol. Ecol.* 30 (13), 2959–2968.
- Aria, M., Cuccurullo, C., 2017. bibliometrix: an R-tool for comprehensive science mapping analysis. *J. Informetr.* 11 (4), 959–975.
- Arrighi, C., Castelli, F., 2023a. Prediction of ecological status of surface water bodies with supervised machine learning classifiers. *Sci. Total Environ.* 857, 159655.
- Arrighi, C., Castelli, F., 2023b. Prediction of ecological status of surface water bodies with supervised machine learning classifiers. *Sci. Total Environ.* 857, 159655.
- Aslan, S., Zennaro, F., Furlan, E., Critto, A., 2022. Recurrent neural networks for water quality assessment in complex coastal lagoon environments: a case study on the Venice Lagoon. *Environ. Model. Software* 154, 105403.
- Bezák, N., Lebar, K., Bai, Y., Rusjan, S., 2025. Using machine learning to predict suspended sediment transport under climate change. *Water Resour. Manag.* 1–16.
- Bianconi, A., Vascon, S., Furlan, E., Critto, A., 2025. Leveraging artificial intelligence methods to map seagrass ecosystems in Italian Seas: tackling human impact and climate change. *Environ. Model. Software*, 106678.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Cachay, S.R., Erickson, E., Buckner, A.F.C., Pokropek, E., Potosnak, W., Bire, S., Osei, S., Lütjens, B., 2021. The world as a graph: improving El Niño forecasts with graph neural networks. *ArXiv Preprint ArXiv:2104.05089*.
- Carafa, R., Faggiano, L., Real, M., Munné, A., Ginebreda, A., Guasch, H., Flo, M., Tirapu, L., von der Ohe, P.C., 2011. Water toxicity assessment and spatial pollution patterns identification in a Mediterranean River Basin District. Tools for water management and risk analysis. *Sci. Total Environ.* 409 (20), 4269–4279.
- Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794.
- Cherif, E.K., Mozetič, P., France, J., Flander-Putrlje, V., Faganeli-Pucer, J., Vodopivec, M., 2021. Comparison of in-situ chlorophyll-a time series and sentinel-3 ocean and land color instrument data in Slovenian National Waters (gulf of trieste, Adriatic Sea). *Water* 13 (14), 1903.
- Chrobak, G., Kowalczyk, T., Fischer, T.B., Chrobak, K., Szwedrański, S., Kazak, J.K., 2021a. Combining indicators for better decisions—Algorithms vs experts on lakes ecological status assessment. *Ecol. Indic.* 132, 108318.
- Chrobak, G., Kowalczyk, T., Fischer, T.B., Szwedrański, S., Chrobak, K., Kazak, J.K., 2021b. Ecological state evaluation of lake ecosystems revisited: latent variables with kSVM algorithm approach for assessment automatization and data comprehension. *Ecol. Indic.* 125, 107567.
- Commission, E., for Environment, D.-G., 2024. River Basin Management in a Changing Climate – Common Implementation Strategy for the Water Framework Directive and the Floods Directive. Publications Office of the European Union. <https://doi.org/10.2779/126657>.
- Concepcion, R., Dadios, E., Bandala, A., Caçador, I., Fonseca, V.F., Duarte, B., 2021. Applying limnological feature-based machine learning techniques to chemical state classification in marine transitional systems. *Front. Mar. Sci.* 8, 658434.
- Crampton, J.W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M.W., Zook, M., 2013. Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. *Cartogr. Geogr. Inf. Sci.* 40 (2), 130–139.
- Cuomo, S., Di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M., Piccialli, F., 2022. Scientific machine learning through physics-informed neural networks: where we are and what's next. *J. Sci. Comput.* 92 (3), 88.
- de Luca Lopes de Amorim, F., Rick, J., Lohmann, G., Wiltshire, K.H., 2021. Evaluation of machine learning predictions of a highly resolved time series of chlorophyll-a concentration. *Appl. Sci.* 11 (16), 7208.
- Dedecker, A.P., Goethals, P.L.M., Gabriels, W., De Pauw, N., 2004. Optimization of Artificial Neural Network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). *Ecol. Model.* 174 (1–2), 161–173.
- Deng, Y., Zhang, Y., Pan, D., Yang, S.X., Gharabaghi, B., 2024. Review of recent advances in remote sensing and machine learning methods for lake water quality management. *Remote Sens.* 16 (22), 4196.
- Dézerald, O., Mondy, C.P., Dembski, S., Kreutzenberger, K., Reyjol, Y., Chandresis, A., Valette, L., Brosse, S., Toussaint, A., Belliard, J., et al., 2020. A diagnosis-based approach to assess specific risks of river degradation in a multiple pressure context: insights from fish communities. *Sci. Total Environ.* 734, 139467.
- Donnelly, J., Daneshkhan, A., Abolfathi, S., 2024. Physics-informed neural networks as surrogate models of hydrodynamic simulators. *Sci. Total Environ.* 912, 168814.
- Dun, A., Yang, Y., Lei, F., 2022. Dynamic graph convolution neural network based on spatial-temporal correlation for air quality prediction. *Ecol. Inform.* 70, 101736.
- Etchells, T.A., Vellido, A., Martelli, E., Lisboa, P.J.G., Comas, J., 2006. On the Prediction of the Ecological Status of Human-Altered Streams and its Rule-Based Interpretation. *Falanga Bolognesi, S., Pasolli, E., Belfiore, O.R., De Michele, C., D'Urso, G., 2020. Harmonized landsat 8 and sentinel-2 time series data to detect irrigated areas: an application in Southern Italy. Remote Sens.* 12 (8), 1275.
- Feio, M.J., Serra, S.R.Q., Mortágua, A., Bouchez, A., Rimet, F., Vasselton, V., Almeida, S.F. P., 2020a. A taxonomy-free approach based on machine learning to assess the quality of rivers with diatoms. *Sci. Total Environ.* 722, 137900.
- Feio, M.J., Serra, S.R.Q., Mortágua, A., Bouchez, A., Rimet, F., Vasselton, V., Almeida, S.F. P., 2020b. A taxonomy-free approach based on machine learning to assess the quality of rivers with diatoms. *Sci. Total Environ.* 722, 137900.
- Fiscor, M., Csabai, Z., 2023. Machine learning model ensemble based on multi-scale predictors confirms ecological segregation and accurately predicts the occurrence of net-spinning caddisfly larvae species groups (Trichoptera: Hydropsychidae) at catchment-scale. *Ecol. Indic.* 146, 109769.
- Figueredo, M.B., de Jesus Ferreira, M., Monteiro, R.L.S., do Nascimento Silva, A., Murari, T.B., et al., 2025. Hybrid PINN-LSTM model for river temperature prediction: a physics-informed deep learning approach. *J. Comput. Commun.* 13 (6), 115–134.
- Gambini, A.F., Angelats, E., González, J.S., Miozzo, M., Dini, P., 2021. Sustainable marine ecosystems: deep learning for water quality assessment and forecasting. *IEEE Access* 9, 121344–121365.
- Garza, A., Mergenthaler-Canseco, M., 2023. TimeGPT-1. *ArXiv Preprint ArXiv:2310.03589*.
- Gebler, D., Kayzer, D., Szoszkiewicz, K., Budka, A., 2014. Artificial neural network modelling of macrophyte indices based on physico-chemical characteristics of water. *Hydrobiologia* 737, 215–224.
- Gebler, D., Kolada, A., Pasztaleniec, A., Szoszkiewicz, K., 2021. Modelling of ecological status of Polish lakes using deep learning techniques. *Environ. Sci. Pollut. Control Res.* 28, 5383–5397.
- Gebler, D., Szoszkiewicz, K., Pietruczuk, K., 2017. Modeling of the river ecological status with macrophytes using artificial neural networks. *Limnologia* 65, 46–54.
- Gebler, D., Wiegand, G., Szoszkiewicz, K., 2018. Integrating river hydromorphology and water quality into ecological status modelling by artificial neural networks. *Water Res.* 139, 395–405.
- Gori, M., Monfardini, G., Scarselli, F., 2005. A new model for learning in graph domains. *Proceedings. 2005 IEEE International Joint Conference on Neural Networks* 2, 729–734, 2005.
- Govedarika, M., Jakovljević, G., 2019. Monitoring spatial and temporal variation of water quality parameters using time series of open multispectral data. *Seventh International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2019)* 11174, 298–307.
- Gurri, R., Arranz, I., Ordeix, M., Garcia-Comas, C., 2023. Automatic image processing to determine the community size structure of riverine macroinvertebrates. *JoVE J.* 191, e64320.
- Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., et al., 2022. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* 45 (1), 87–110.
- Harris, L., McRae, A.T.T., Chantry, M., Dueben, P.D., Palmer, T.N., 2022. A generative deep learning approach to stochastic downscaling of precipitation forecasts. *J. Adv. Model. Earth Syst.* 14 (10) e2022MS003120.
- Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, vol. 2. Springer.

- Heasley, E.L., Millington, J.D.A., Clifford, N.J., Chadwick, M.A., 2020a. A waterbody typology derived from catchment controls using self-organising maps. *Water* 12 (1), 78. DOI, 10, w12010078.
- Heasley, E.L., Millington, J.D.A., Clifford, N.J., Chadwick, M.A., 2020b. A waterbody typology derived from catchment controls using self-organising maps. *Water* 12 (1), 78. DOI, 10, w12010078.
- Huber, S., Hansen, L.B., Nielsen, L.T., Rasmussen, M.L., Sølvsteen, J., Berglund, J., von Friesen, C., Danbolt, M., Envall, M., Infantes, E., et al., 2022. Novel approach to large-scale monitoring of submerged aquatic vegetation: a nationwide example from Sweden. *Integrated Environ. Assess. Manag.* 18 (4), 909–920.
- Jia, L., Yen, N., Pei, Y., 2025. A physics-coupled deep learning framework for hydrodynamic diffusion modeling in Watershed systems: integrating spatiotemporal networks and environmental constraints. *IEEE Access*.
- Karki, S., Bermejo, R., Wilkes, R., Monagail, M. Mac, Daly, E., Healy, M., Hanafin, J., McKinstry, A., Mellander, P.-E., Fenton, O., et al., 2021. Mapping spatial distribution and biomass of intertidal Ulva blooms using machine learning and earth observation. *Front. Mar. Sci.* 8, 633128.
- Knoll, L., Breuer, L., Bach, M., 2019a. Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning. *Sci. Total Environ.* 668, 1317–1327.
- Knoll, L., Breuer, L., Bach, M., 2019b. Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning. *Sci. Total Environ.* 668, 1317–1327.
- Knoll, L., Breuer, L., Bach, M., 2020a. Nation-wide estimation of groundwater redox conditions and nitrate concentrations through machine learning. *Environ. Res. Lett.* 15 (6), 64004.
- Knoll, L., Breuer, L., Bach, M., 2020b. Nation-wide estimation of groundwater redox conditions and nitrate concentrations through machine learning. *Environ. Res. Lett.* 15 (6), 64004.
- Koito, O., Retiche, J., Bikse, J., Terasmaa, J., Tarros, S., Abreldaal, P., Babre, A., Hunt, M., Pärn, J., Vainu, M., et al., 2023. Hydrochemical signatures of springs for conceptual model development to support monitoring of transboundary aquifers. *Groundw. Sustain. Dev.* 21, 100927.
- Kraan, C., Haslob, H., Probst, W.N., Stelzenmüller, V., Rehren, J., Neumann, H., 2024. Thresholds of seascape fauna composition along gradients of human pressures and natural conditions to inform marine spatial planning. *Sci. Total Environ.* 914, 169940.
- Krtolica, I., Kamenko, I., Paunović, M., Raković, M., Atanacković, A., Talanov, M., Popović, N., 2025. Use of Artificial Neural Networks (ANNs) to assess xenobiotics in a river catchment using macroinvertebrates as bioindicators. *Aquat. Toxicol.* 107466.
- Kruk, A., Cieplucha Michałand Zi\keba, G., Tybulec, S., Tsydel, M., Marszał, L., Błońska, D., Galicka, W., Przybylski, M., 2016. Recovery of fish fauna in the upper Warta River, Poland: long-term (1963–2012) monitoring study. *Ecol. Inform.* 33, 109–118.
- Kwon, Y.S., Kang, H., Pyo, J., 2024. Estimation of aquatic ecosystem health using deep neural network with nonlinear data mapping. *Ecol. Inform.*, 102588.
- Lanzen, A., Mendibil, I., Borja, A., Alonso-Sáez, L., 2021a. A microbial mandala for environmental monitoring: predicting multiple impacts on estuarine prokaryote communities of the Bay of Biscay. *Mol. Ecol.* 30 (13), 2969–2987.
- Lanzen, A., Mendibil, I., Borja, A., Alonso-Sáez, L., 2021b. A microbial mandala for environmental monitoring: predicting multiple impacts on estuarine prokaryote communities of the Bay of Biscay. *Mol. Ecol.* 30 (13), 2969–2987.
- Le Traon, P.Y., Reppucci, A., Fanjul, E.A., Aouf, L., Behrens, A., Belmonte, M., Bentamy, A., Bertino, L., Brando, V.E., Kreiner, M.B., Benkiran, M., Carval, T., Ciliberti, S.A., Claustre, H., Clementi, E., Coppini, G., Cossarini, G., De Alfonso Alonso-Muñoyerro, M., Delamarche, A., et al., 2019. From observation to information and users: the Copernicus marine service perspective. In: *Frontiers in Marine Science*, vol. 6. Frontiers Media S.A. <https://doi.org/10.3389/fmars.2019.00234>. Issue May.
- Leontidou, K., Rubel, V., Stoeck, T., 2023. Comparing quantile regression spline analyses and supervised machine learning for environmental quality assessment at coastal marine aquaculture installations. *PeerJ* 11, e15425.
- Lira, H., Mart\i, L., Sanchez-Pi, N., 2022. A graph neural network with spatio-temporal attention for multi-sources time series data: an application to frost forecast. *Sensors* 22 (4), 1486.
- Lock, K., Adriaens, T., Goethals, P., 2014. Effect of water quality on blackflies (Diptera: Simuliidae) in Flanders (Belgium). *Limnologia* 44, 58–65.
- Lock, K., Goethals, P.L.M., 2014. Predicting the occurrence of stoneflies (Plecoptera) on the basis of water characteristics, river morphology and land use. *J. Hydroinform.* 16 (4), 812–821.
- Loor, J.S., Fdez-Arroyabe, P., 2018. Aerial and satellite imagery and big data: blending Old technologies. *Big Data for Remote Sensing: Visualization, Analysis and Interpretation: Digital Earth and Smart Earth* 39.
- Lovett, G.M., Burns, D.A., Driscoll, C.T., Jenkins, J.C., Mitchell, M.J., Rustad, L., Shanley, J.B., Likens, G.E., Haeuber, R., 2007. Who needs environmental monitoring? *Front. Ecol. Environ.* 5 (5), 253–260.
- Madhuanand, L., Philippart, C.J.M., Nijland, W., de Jong, S.M., Bijleveld, A.I., Addink, E. A., 2025. Optimizing predictions of environmental variables and species distributions on tidal flats by combining Sentinel-2 images and their deep-learning features with OBIA. *Int. J. Rem. Sens.* 46 (2), 811–834.
- Marre, G., Braga, C.D.A., Ienco, D., Luque, S., Holon, F., Deter, J., 2020a. Deep convolutional neural networks to monitor coralligenous reefs: operationalizing biodiversity and ecological assessment. *Ecol. Inform.* 59, 101110.
- Marre, G., Braga, C.D.A., Ienco, D., Luque, S., Holon, F., Deter, J., 2020b. Deep convolutional neural networks to monitor coralligenous reefs: operationalizing biodiversity and ecological assessment. *Ecol. Inform.* 59, 101110.
- Martin-Abadal, M., Ruiz-Frau, A., Hinz, H., Gonzalez-Cid, Y., 2020a. Jellytoring: real-time jellyfish monitoring based on deep learning object detection. *Sensors* 20 (6), 1708.
- Martin-Abadal, M., Ruiz-Frau, A., Hinz, H., Gonzalez-Cid, Y., 2020b. Jellytoring: real-time jellyfish monitoring based on deep learning object detection. *Sensors* 20 (6), 1708.
- Martin-Abadal, M., Ruiz-Frau, A., Hinz, H., Gonzalez-Cid, Y., 2020c. Jellytoring: real-time jellyfish monitoring based on deep learning object detection. *Sensors* 20 (6), 1708.
- Mart\inez-Santos, P., Aristizábal, H.F., D\iaz-Alcaide, S., Gómez-Escalonilla, V., 2021. Predictive mapping of aquatic ecosystems by means of support vector machines and random forests. *J. Hydrol.* 595, 126026.
- Martyszunis, A., Loga, M., Przędziecki, K., 2024. Using machine learning for the assessment of ecological status of unmonitored waters in Poland. *Sci. Rep.* 14 (1), 24509.
- Medsker, L.R., Jain, L., et al., 2001. Recurrent neural networks. *Design and Applications* 5 (64–67), 2.
- Meuer, J., Witte, M., Finn, T.S., Timmreck, C., Ludwig, T., Kadow, C., 2024. Latent diffusion model for generating ensembles of climate simulations. *ArXiv Preprint ArXiv:2407.02070*.
- Mingers, J., Leydesdorff, L., 2015. A review of theory and practice in scientometrics. *Eur. J. Oper. Res.* 246 (1), 1–19.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., PRISMA Group*, t., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann. Intern. Med.* 151 (4), 264–269.
- Molina, J.-L., Pulido-Velázquez, M., Llopis-Albert, C., Peña-Haro, S., 2013. Stochastic hydro-economic model for groundwater quality management using Bayesian networks. *Water Sci. Technol.* 67 (3), 579–586.
- Najafzadeh, M., Ahmadi-Rad, E.S., Gebler, D., 2024. Ecological states of watercourses regarding water quality parameters and hydromorphological parameters: deriving empirical equations by machine learning models. *Stoch. Environ. Res. Risk Assess.* 38 (2), 665–688.
- Ni, Q., Wang, Y., Fang, Y., 2022. GE-STDGN: a novel spatio-temporal weather prediction model based on graph evolution. *Appl. Intell.* 52 (7), 7638–7652.
- Ocampo-Duque, W., Schuhmacher, M., Domingo, J.L., 2007. A neural-fuzzy approach to classify the ecological status in surface waters. *Environ. Pollut.* 148 (2), 634–641.
- Oiry, S., Barillé, L., 2021. Using sentinel-2 satellite imagery to develop microphytobenthos-based water quality indices in estuaries. *Ecol. Indic.* 121, 107184.
- Olaya-Mar\in, E.J., Mart\inez-Capel, F., Costa, R.M.S., Alcaraz-Hernández, J.D., 2012a. Modelling native fish richness to evaluate the effects of hydromorphological changes and river restoration (Júcar River Basin, Spain). *Sci. Total Environ.* 440, 95–105.
- Olaya-Mar\in, E.J., Mart\inez-Capel, F., Costa, R.M.S., Alcaraz-Hernández, J.D., 2012b. Modelling native fish richness to evaluate the effects of hydromorphological changes and river restoration (Júcar River Basin, Spain). *Sci. Total Environ.* 440, 95–105.
- Olaya-Mar\in, E.J., Mart\inez-Capel, F., Costa, R.M.S., Alcaraz-Hernández, J.D., 2012c. Modelling native fish richness to evaluate the effects of hydromorphological changes and river restoration (Júcar River Basin, Spain). *Sci. Total Environ.* 440, 95–105.
- Oliveira Santos, V., Costa Rocha, P.A., Scott, J., Van Griensven Thé, J., Gharabaghi, B., 2023. Spatiotemporal air pollution forecasting in houston-TX: a case study for ozone using deep graph neural networks. *Atmosphere* 14 (2), 308.
- Omarova, P., Amirgaliyev, Y., Kozbakova, A., Ataniyazova, A., 2023. Application of physics-informed neural networks to river silting simulation. *Appl. Sci.* 13 (21), 11983.
- Ouyang, X., Yang, Y., Zhang, Y., Zhou, W., 2021. Spatial-temporal dynamic graph convolution neural network for air quality prediction. 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1–8.
- Park, Y.S., Verdonchot, P.F.M., Chon, T.S., Gevrey, M., Lek, S., 2005. Prediction of macroinvertebrate diversity of freshwater bodies by adaptive learning algorithms. In: *Modelling Community Structure in Freshwater Ecosystems*. Springer, pp. 189–205.
- Pielke Jr, R., Ritchie, J., 2021. Distorting the view of our climate future: the misuse and abuse of climate pathways and scenarios. *Energy Res. Social Sci.* 72, 101890.
- Politikos, D., Stefanidis, K., Varlas, G., Papadopoulos, A., Dimitriou, E., 2024. Predicting the ecological quality of rivers: a machine learning approach and a what-if scenarios tool. *Environ. Model. Assess.* 29 (6), 1059–1077.
- Raitoharju, J., 2022. Convolutional neural networks. In: *Deep Learning for Robot Perception and Cognition*. Elsevier, pp. 35–69.
- Rakaj, A., Morroni, L., Flori, G., Grosso, L., Pinsino, A., Ventura, D., Fianchini, A., Pellegrini, D., Regoli, F., Scardi, M., 2025. Towards automatization in morphometric analysis of sea urchin embryos for ecotoxicological applications and near-future scenarios simulation. *Ecol. Indic.* 175, 113533.
- Rampal, N., Gibson, P.B., Sherwood, S., Abramowitz, G., Hobeichi, S., 2025. A reliable generative adversarial network approach for climate downscaling and weather generation. *J. Adv. Model. Earth Syst.* 17 (1) e2024MS004668.

- Rane, N., 2023. Roles and Challenges of ChatGPT and Similar Generative Artificial Intelligence for Achieving the Sustainable Development Goals (SDGs). Available at: SSRN 4603244.
- Rane, N., Choudhary, S., Rane, J., 2024. Contribution of ChatGPT and Similar Generative Artificial Intelligence for Enhanced Climate Change Mitigation Strategies. Available at: SSRN 4681720.
- Riabchenko, E., Meissner, K., Ahmad, I., Iosifidis, A., Tirronen, V., Gabbouj, M., Kiranyaz, S., 2016. Learned vs. engineered features for fine-grained classification of aquatic macroinvertebrates. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 2276–2281.
- Rokach, L., Maimon, O., 2005. Decision Trees. *Data Mining and Knowledge Discovery Handbook*, pp. 165–192.
- Sabara, R., Soares, C., Zabel, F., Oliveira, J.V., Jesus, S.M., 2020a. Automatic acoustic target detection and classification off the Coast of Portugal. *Global Oceans 2020: Singapore–US Gulf Coast* 1–9.
- Sabara, R., Soares, C., Zabel, F., Oliveira, J.V., Jesus, S.M., 2020b. Automatic acoustic target detection and classification off the Coast of Portugal. *Global Oceans 2020: Singapore–US Gulf Coast* 1–9.
- Sabara, R., Soares, C., Zabel, F., Oliveira, J.V., Jesus, S.M., 2020c. Automatic acoustic target detection and classification off the Coast of Portugal. *Global Oceans 2020: Singapore–US Gulf Coast* 1–9.
- Sajbi, A.M., Bamal, A., Diganta, M.T.M., Ashekuzzaman, S.M., Rahman, A., Olbert, A.I., Uddin, M.G., 2025. Novel groundwater quality index (GWQI) model: a reliable approach for the assessment of groundwater. *Results Eng.* 25, 104265.
- Scarrica, V.M., Aucelli, P.P.C., Cagnazzo, C., Casolaro, A., Fiore, P., La Salandra, M., Rizzo, A., Scardino, G., Scicchitano, G., Staiano, A., 2022a. A novel beach litter analysis system based on UAV images and Convolutional Neural Networks. *Ecol. Inform.* 72, 101875. <https://doi.org/10.1016/j.ecoinf.2022.101875>.
- Scarrica, V.M., Aucelli, P.P.C., Cagnazzo, C., Casolaro, A., Fiore, P., La Salandra, M., Rizzo, A., Scardino, G., Scicchitano, G., Staiano, A., 2022b. A novel beach litter analysis system based on UAV images and Convolutional Neural Networks. *Ecol. Inform.* 72, 101875.
- Scarrica, V.M., Aucelli, P.P.C., Cagnazzo, C., Casolaro, A., Fiore, P., La Salandra, M., Rizzo, A., Scardino, G., Scicchitano, G., Staiano, A., 2022c. A novel beach litter analysis system based on UAV images and Convolutional Neural Networks. *Ecol. Inform.* 72, 101875. <https://doi.org/10.1016/j.ecoinf.2022.101875>.
- Schuster, G.T., Chen, Y., Feng, S., 2024. Review of physics-informed machine-learning inversion of geophysical data. *Geophysics* 89 (6), T337–T356.
- Scowen, M., Athanasiadis, I.N., Bullock, J.M., Eigenbrod, F., Willcock, S., 2021. The current and future uses of machine learning in ecosystem service research. *Sci. Total Environ.* 799, 149263.
- Segurado, P., Almeida, C., Neves, R., Ferreira, M.T., Branco, P., 2018a. Understanding multiple stressors in a Mediterranean basin: combined effects of land use, water scarcity and nutrient enrichment. *Sci. Total Environ.* 624, 1221–1233.
- Segurado, P., Almeida, C., Neves, R., Ferreira, M.T., Branco, P., 2018b. Understanding multiple stressors in a Mediterranean basin: combined effects of land use, water scarcity and nutrient enrichment. *Sci. Total Environ.* 624, 1221–1233.
- Segura-Méndez, F.J., Pérez-Sánchez, J., Senent-Aparicio, J., 2023. Evaluating the riparian forest quality index (QBR) in the Luchena River by integrating remote sensing, machine learning and GIS techniques. *Ecohydrol. Hydrobiol.* 23 (3), 469–483.
- Simeoni, C., Furlan, E., Pham, H.V., Critto, A., de Juan, S., Trégarot, E., Cornet, C.C., Meesters, E., Fonseca, C., Botelho, A.Z., Krause, T., N'Guetta, A., Cordova, F.E., Failler, P., Marcomini, A., 2023. Evaluating the combined effect of climate and anthropogenic stressors on marine coastal ecosystems: insights from a systematic review of cumulative impact assessment approaches. In: *Science of the Total Environment*, vol. 861. Elsevier B.V. <https://doi.org/10.1016/j.scitotenv.2022.160687>.
- Smith, V.H., Tilman, G.D., Nekola, J.C., 1999. Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environ. Pollut.* 100 (1–3), 179–196.
- Sousa, T., Ries, B., Gueff, N., 2024. Data augmentation in earth observation: a diffusion model approach. *ArXiv Preprint ArXiv:2406.06218*.
- Sroczynska, K., Claro, M., Kruk, A., Wojtal-Frankiewicz, A., Range, P., Ch'\'icharo, L., 2017a. Indicator macroinvertebrate species in a temporary Mediterranean river: recognition of patterns in binary assemblage data with a Kohonen artificial neural network. *Ecol. Indic.* 73, 319–330.
- Sroczynska, K., Claro, M., Kruk, A., Wojtal-Frankiewicz, A., Range, P., Ch'\'icharo, L., 2017b. Indicator macroinvertebrate species in a temporary Mediterranean river: recognition of patterns in binary assemblage data with a Kohonen artificial neural network. *Ecol. Indic.* 73, 319–330.
- Szczepocka, E., Nowicka-Krawczyk, P., Kruk, A., 2018. Deceptive ecological status of urban streams and rivers—evidence from diatom indices. *Ecosphere* 9 (7), e02310.
- Tamvakis, A., Miritzis, J., Tsiertsis, G., Spyropoulou, A., Spatharis, S., 2012. Effects of meteorological forcing on coastal eutrophication: modeling with model trees. *Estuar. Coast Shelf Sci.* 115, 210–217.
- Tishechkin, D., Dubovik, A., Koriagin, A., Khudorozhkov, R., 2024. Improve geoscience interpretation, reporting, and transparency with generative AI. *Fourth EAGE Digitalization Conference & Exhibition 2024* (1), 1–5.
- Tison, J., Park, Y.-S., Coste, M., Wasson, J.G., Rimet, F., Ector, L., Delmas, F., 2007. Predicting diatom reference communities at the French hydrosystem scale: a first step towards the definition of the good ecological status. *Ecol. Model.* 203 (1–2), 99–108.
- Tourkij, Y., Keisler, J., Linkov, I., 2013. Scenario analysis: a review of methods and applications for engineering and environmental systems. *Environ. Syst. Decis.* 33, 3–20.
- Tsou, M.-H., 2015. Research challenges and opportunities in mapping social media and Big Data. *Cartogr. Geogr. Inf. Sci.* 42 (Suppl. 1), 70–74.
- Uusitalo, L., Blenckner, T., Puntilla-Dodd, R., Skyttä, A., Jernberg, S., Voss, R., Müller-Karulis, B., Tomczak, M.T., Möllmann, C., Peltonen, H., 2022a. Integrating diverse model results into decision support for good environmental status and blue growth. *Sci. Total Environ.* 806, 150450. <https://doi.org/10.1016/j.scitotenv.2021.150450>.
- Uusitalo, L., Blenckner, T., Puntilla-Dodd, R., Skyttä, A., Jernberg, S., Voss, R., Müller-Karulis, B., Tomczak, M.T., Möllmann, C., Peltonen, H., 2022b. Integrating diverse model results into decision support for good environmental status and blue growth. *Sci. Total Environ.* 806, 150450. <https://doi.org/10.1016/j.scitotenv.2021.150450>.
- Valerio, C., De Stefano, L., Mart'\'inez-Muñoz, G., Garrido, A., 2021a. A machine learning model to assess the ecosystem response to water policy measures in the Tagus River Basin (Spain). *Sci. Total Environ.* 750, 141252.
- Valerio, C., De Stefano, L., Mart'\'inez-Muñoz, G., Garrido, A., 2021b. A machine learning model to assess the ecosystem response to water policy measures in the Tagus River Basin (Spain). *Sci. Total Environ.* 750, 141252.
- Venkataramanan, A., Faure-Giovagnoli, P., Regan, C., Heudre, D., Figus, C., Usseglio-Polatera, P., Pradalier, C., Laviale, M., 2023. Usefulness of synthetic datasets for diatom automatic detection using a deep-learning approach. *Eng. Appl. Artif. Intell.* 117, 105594.
- Visser, H., Evers, N., Bontsema, A., Rost, J., de Niet, A., Vethman, P., Mylius, S., van der Linden, A., van den Roovaart, J., van Gaalen, F., et al., 2022a. What drives the ecological quality of surface waters? A review of 11 predictive modeling tools. *Water Res.* 208, 117851.
- Visser, H., Evers, N., Bontsema, A., Rost, J., de Niet, A., Vethman, P., Mylius, S., van der Linden, A., van den Roovaart, J., van Gaalen, F., et al., 2022b. What drives the ecological quality of surface waters? A review of 11 predictive modeling tools. *Water Res.* 208, 117851.
- Vo, D.T., Nguyen, X.P., Nguyen, T.D., Hidayat, R., Huynh, T.T., Nguyen, D.T., 2021. A review on the internet of thing (IoT) technologies in controlling ocean environment. *Energy Sources, Part A Recovery, Util. Environ. Eff.* 1–19.
- Volf, G., Zutinić, P., Gligora Udović, M., Kulaš, A., Mustafić, P., 2023. Describing and simulating phytoplankton of a small and shallow reservoir using decision trees and rule-based models. *Environ. Monit. Assess.* 195 (4), 508.
- Vrieze, O.J., 2005. Kohonen network. *Artificial Neural Networks: An Introduction to ANN Theory and Practice* 83–100.
- Wang, S., Cao, J., Philip, S.Y., 2020a. Deep learning for spatio-temporal data mining: a survey. *IEEE Trans. Knowl. Data Eng.* 34 (8), 3681–3700.
- Wang, S., Li, Y., Zhang, J., Meng, Q., Meng, L., Gao, F., 2020b. Pm2. 5-gnn: a domain knowledge enhanced graph neural network for pm2. 5 forecasting. *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, pp. 163–166.
- Willcock, S., Mart'\'inez-López, J., Hooftman, D.A.P., Bagstad, K.J., Balbi, S., Marzo, A., Prato, C., Scianrello, S., Signorello, G., Voigt, B., et al., 2018. Machine learning for ecosystem services. *Ecosyst. Serv.* 33, 165–174.
- Yan, Y., Zhang, Y., Ren, J., Hadjal, M., Mcke, D., Kao, F., Durrani, T., 2021. Estimation of chlorophyll concentration for environment monitoring in Scottish marine water. *International Conference in Communications, Signal Processing, and Systems*, pp. 582–587.
- Yang, X., Zhang, F., Sun, P., Li, X., Du, Z., Liu, R., 2022. A spatio-temporal graph-guided convolutional LSTM for tropical cyclones precipitation nowcasting. *Appl. Soft Comput.* 124, 109003.
- Yu, X., Shi, S., Xu, L., 2021. A spatial-temporal graph attention network approach for air temperature forecasting. *Appl. Soft Comput.* 113, 107888.
- Yu, Y., Si, X., Hu, C., Zhang, J., 2019. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* 31 (7), 1235–1270.
- Zennaro, F., Furlan, E., Canu, D., Alcazar, L.A., Rosati, G., Solidoro, C., Critto, A., 2024a. Hypoxia extreme events in a changing climate: machine learning methods and deterministic simulations for future scenarios development in the Venice Lagoon. *Mar. Pollut. Bull.* 208, 117028.
- Zennaro, F., Furlan, E., Canu, D., Alcazar, L.A., Rosati, G., Solidoro, C., Critto, A., 2024b. Hypoxia extreme events in a changing climate: machine learning methods and deterministic simulations for future scenarios development in the Venice Lagoon. *Mar. Pollut. Bull.* 208, 117028.
- Zennaro, F., Furlan, E., Melaku Canu, D., Aveytua Alcazar, L., Rosati, G., Aslan, S., Solidoro, C., Critto, A., 2023a. Evaluation of lagoon eutrophication potential under climate change conditions: a novel water quality machine learning and biogeochemical-based framework. *EGU General Assembly Conference Abstracts*. EGU–8409.
- Zennaro, F., Furlan, E., Melaku Canu, D., Aveytua Alcazar, L., Rosati, G., Aslan, S., Solidoro, C., Critto, A., 2023b. Evaluation of lagoon eutrophication potential under climate change conditions: a novel water quality machine learning and biogeochemical-based framework. *EGU General Assembly Conference Abstracts*. EGU–8409.
- Zennaro, F., Furlan, E., Melaku Canu, D., Aveytua Alcazar, L., Rosati, G., Aslan, S., Solidoro, C., Critto, A., 2023c. Evaluation of lagoon eutrophication potential under climate change conditions: a novel water quality machine learning and biogeochemical-based framework. *EGU General Assembly Conference Abstracts*. EGU–8409.
- Zhong, G., Wang, L.-N., Ling, X., Dong, J., 2016. An overview on data representation learning: from traditional feature learning to recent deep learning. *The Journal of Finance and Data Science* 2 (4), 265–278.
- Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., Ye, L., 2022. A review of the application of machine learning in water quality evaluation. *Eco-Environ. Health* 1 (2), 107–116.