Ecological Questions 35(2024)1 http://dx.doi.org/10.12775/EQ.2024.005

The supporting role of Artificial Intelligence and Machine/Deep Learning in monitoring the marine environment: a bibliometric analysis

Fabiana Di Ciaccio[1*]

[1]Department of Science and Technology, Parthenope University of Naples, Italy

*Corresponding author's email: fabiana.diciaccio@uniparthenope.it

Received: 5 December 2022/ Accepted: 14 December 2023

Abstract. The widespread interest towards a sustainable and effective monitoring of the environment is increasingly demanding the development of modern and more affordable technologies to support or even replace the traditional time-consuming, high-cost sampling surveys at a multi-scale level. Researchers are highly benefitting from the recent enormous progresses achieved in the Artificial Intelligence (AI) field, with Machine/Deep Learning (ML/DL) applications increasing at sight. This gives a remarkable contribution to the environmental monitoring at sea, further allowing to develop efficient, smart and low-cost solutions to support the wide variety of tasks dealing with this objective.

This study explores the global scientific literature on AI and ML/DL applications for the environmental monitoring over the last years. The VOSviewer software has been used to create maps based on the bibliographic network data: this allowed to display the relationships among scientific journals, researchers, and countries and to analyze the co-occurrence of different terms connected to the research. The resulting bibliometric analysis aims at verifying the major research interests and at providing the community with interesting findings and new perspectives on this very important topic, highlighting the great potential and flexibility of these methodologies and the excellent achievements they obtained in the last years.

Keywords: Artificial Intelligence; Bibliometric analysis; Deep Learning; Environmental monitoring; Machine Learning; Marine environment; Social network analysis; VOSviewer

1. Introduction

Oceans and seas cover more than the 70% of the planet and are of fundamental importance to the global population from an economic, social and environmental point of view. Over 3 billion people depend on marine life for their livelihood, supported by fishing activities, shipping, water resources, tourism and so on; this makes water one of the most sensitive and vulnerable resources, as it can be affected by a wide variety of anthropogenic activities and risks. For this reason, the United Nation Member States adopted the 2030 Agenda for Sustainable Development, a plan of action for people, planet and prosperity which aims at strengthening the universal peace now and into the future (Resolution et al., 2015). The 17 Sustainable Development Goals (SDGs) are at the core of this blueprint and, together with the 169 different announced targets, will encourage the realization of the society in areas of critical importance. In particular, the SDG 14 is about “Life below water” and aims at “*Conserve and sustainably use the oceans, sea and marine resources for sustainable development.*” (Desa et al., 2016).

In this context, the need to develop new effective strategies for the monitoring and protection of the environment and its ecosystems is becoming imperative. This gets particular urgency in view of the impact of anthropogenic activities, which are heavily pressing on climate change, environmental pollution, or human industry driven eutrophication, in a way that is likely to increase with the projected demographic expansion. Such impacts are traditionally monitored by surveying the biodiversity and the ecosystems evolution (Cordier et al., 2017).

Remote-sensing techniques or fixed monitoring station are generally employed to collect data at a global scale (Huang et al., 2018a), while local scale data are acquired through probes and sensors deployed from or embedded to ships (Prabowo et al., 2021). However, other than being usually costly and not always able to assure consistent sampling, these procedures prove their efficiency as long as the environmental conditions allow for accurate and safe operations, thus limiting their employment in some challenging scenarios (i.e., floods, storms, heavy seas, etc.). In the last decades robotic systems have seen an exponential development, as fundamental data-gathering tools by scientists which allow new perspectives and a greater understanding of the environment. This field is in continue evolution: researchers keep on studying increasing advanced solutions to maximize the mission’s productivity whilst keeping low operational costs and impact on the environment (Bayat et al., 2017).

Moreover, recent studies highlighted the potential of Artificial Intelligence and Machine/Deep Learning techniques in several tasks, in which they obtained exceptional results; examples can be found in medical research (Kourou et al., 2015; Shen et al., 2017), pose estimation (Mathis et al., 2018; Di Ciaccio et al., 2022), finance (Bahrammirzaee, 2010), image classification (Lu and Weng, 2007) and detection (Zhao et al., 2019), etc. Huge improvements have been achieved also in the environmental monitoring task, with many applications attracting remarkable attention; examples can be found in the assessment of water quality and marine pollution, deforestation detection, predictive modelling and biological inspections, underwater image enhancements for data analysis and so on, as will be shown in the next section.

With the aim of assessing the current state of the art on this topic and investigating the possibility and direction of developments, enhancements and future working applications, this paper presents the results of a bibliometric analysis made on the environmental monitoring at sea exploited through Artificial Intelligence and Machine/Deep Learning techniques. In the last years, this approach well performed in trends analysis of large amounts of data as it combines the bibliometrics statistical techniques (used in research quality and development assessment) with the social network analysis (SNA), which investigates social structures and relations (Otte and Rousseau, 2002). In the case of academic literature, the bibliometric analysis approach allows to obtain interesting findings based on a quantitative analysis of a topic of interest: the outcome is in fact an evaluation of the existing networks among researchers, countries, organizations or keywords dealing with the specific field of science (Pauna et al., 2019).

1.1 State of the art

This paragraph provides a brief overview on the current state of the art on the environmental monitoring operations which are currently exploited through AI and ML/DL, reporting some of the main achievements obtained by the research community in different field.

Human activities have generated an irreversible global warming process, which is inevitably influencing processes that create warming feedback loops (e.g., wildfires and melting glaciers) (Kim et al., 2022). Among the critical consequences, the acceleration in the sea-level rise (SRL) is affecting the ecosystems both on a global and a local scale: bays and estuaries are expecting a gradual loss of the coastline and saline intrusion, drastically changing the aquatic systems (e.g., rivers, lakes and groundwater) (Eidam et al., 2020). Plus, the salinization of these environments can result in economic and ecological losses (i.e., habitat reduction, species displacement, introduction of exotic species and local extinction) (Little et al., 2017). For this reason, the correct predictions of the ecosystems evolution is extremely valuable for management and decision making (Thorne et al., 2017); in particular, studies on environmental monitoring and impact assessment greatly benefit from the analysis of the benthic fauna (Borja et al., 2012), since it demonstrated to have predictable responses to changes in environmental variables such as salinity and sediments (Anderson, 2008). It is common practice to model this response considering the variation of the species' spatial distribution, recurrently addressed through Species Distribution Modeling (SDM): this approach uses species occurrence and environmental variables data to define the suitability of other areas where species may inhabit (Elith and Leathwick, 2009). In this context, Costa et al. (2022) evaluated the performance of the most recurrent SDM used for benthic macrofauna, with particular reference to the prediction of the SLR effects on the distribution of these invertebrates along the estuarine. They used a set of environmental variables layers refined to the local scale, focusing on some environmental setting as saline intrusion, reduced rainfall and increased evaporation due to the temperature increment, obtaining successful predictions with quite accurate results. In the field of benthic monitoring, Cordier et al. (2017) investigated the possibility of using supervised machine learning (SML) to build predictive models for the inference of four Biotic indices

(BI) in fish farming industry. These models are able to infer BI values from eDNA metabarcoding data without relying on taxonomic assignments, leading to similar bioassessments as the ones obtained using traditional morphology-based macrofaunal surveys. However, using diversity metrics or composition data for the predictions on new samples may give a practical flexibility for biomonitoring surveys. The first category allows to reduce the dimensions of the dataset and consequently the computation time, performing well on samples from different geographical regions. Composition data allows instead to capture species interactions and give more importance to key taxa and specific environmental variation, thus obtaining better results for atypical samples or statistical outliers. Aquatic ecosystems and public health are often threatened by the direct and indirect effects due to the presence of algal blooms in waterbodies (Glasgow et al., 2004), which results in issues such as reduced water transparency, depletion of oxygen, and decreased biodiversity the environments (Hartnett and Nash, 2004). Also in this case, efficient monitoring techniques of algal blooms are fundamental to assess and further manage the water conditions. Currently, the chlorophyll-a (Chl-a) concentration has been used as a useful indicator for measuring the abundance and variety of phytoplankton and/or algal biomass, since all photosynthetic algae include the Chl-a (Boyer et al., 2009). In this context, Park et al. (2015) set up an Artificial Neural Network (ANN) to investigate the model performances in predicting the Chl-a concentration in two characteristically different areas in the southwest of Korea: a freshwater and an estuarine reservoirs. They proposed an early-warning protocol for managing algal blooms using a reliable model for early warning prediction of Chl-a: the use of ANN and support-vector machines (SVM) allows for the optimization of key model parameters and also for the evaluation of model-specific features based on its performance in response to different input variables. The great potential of ANN model as a forecasting tool has been demonstrated also by Palani et al. (2008), who investigated the possibility of predicting temperature, salinity, and Chl-a in Singapore coastal waters. The ANN capability has been found to be fast and reliable in learning the mechanism of convective transport of water quality variables, so they propose to use this model in parallel with physics-based ones as a new prediction tool. This could allow the identification of important parameters and the further physical/chemical monitoring with a quick water quality assessment of the seawater. In the context of sea stream analysis, some interesting applications of ML prediction models have also been proposed by Deo and Sahin (2016) and more recently by Maloney et al. (2022). The first research presents an Extreme Learning Model as a fast computational method to simulate the mean streamflow water level for three hydrological sites in Queensland. It receives as input nine predictors (rainfall, Southern Oscillation Index, Pacific Decadal Oscillation Index, ENSO Modoki Index, Indian Ocean Dipole Index, Nino 3.0, Nino 3.4, and Nino 4.0 sea surface temperatures (SSTs) and the month, to consider the flow seasonality) and its results are further validated with an ANN. Their experiments showed a good ability of the model in simulating the streamflow water level, confirmed that it could be a useful tool for environmental monitoring operations, assessing the viability of irrigation systems, analysing trends in hydrological parameters, developing flood and drought response strategies, and so on. Maloney et al. (2022) proposed instead the use of a tuned random forests ML model to predict the biological conditions of small streams in the non-tidal portion of the Chesapeake Bay watershed (mid-Atlantic United States). Their objective is to optimize the prediction while increasing the usefulness and relevance of the model output, with a particular focus on the interpretability of global and local effects; this will aid in stream restoration and management and support managers and conservation practitioners that need to be able to identify widespread pattern. Random forest models are thus one of the optimal choices when it comes to sea stream forecasting and in general to prediction, as they allow to balance accuracy and computing cost. This is confirmed by the results obtained in air quality analysis, as it is demonstrated for example Zhan et al. (2018), which used a random forest model to predict the nationwide spatiotemporal distributions of [O_3] in China to assess the exposure intensities and the consequent duration of the populations living in the different regions of the Country. A similar task is also exploited, with great results, by Freeman et al. (2018), who trained a recurrent neural network (RNN) with long short-term memory (LSTM) to predict local 8-hr averaged O_3 concentrations based on hourly air monitoring station measurements obtaining very promising results.

Pollution source detection is one of the fields which greatly benefitted from the ML/DL deployment, in particular when it comes to plastics and chemical agents. In the last century the production of plastics led in fact to a huge accumulation of debris at the shorelines and in the oceans worldwide (Barnes et al., 2009). Moreover, they usually fragment into microplastics, which disperse in the water with the consequent threatening of the biodiversity and the ecosystems all (Wright and Kelly, 2017). The importance of source detection involves also chemical agents as they may enter the marine environment from different anthropogenic sources e.g., shipping activities, industry, aquaculture, agriculture, etc. with rivers often reported as a major carrier of these chemicals (Biel-Maeso et al., 2018). In this context, Alygizakis et al. (2022) proposed an open-source workflow for untargeted detection of chemical pollutants in the marine aquatic ecosystem of the Black Sea. In particular, they used a DL convolutional neural network to detect and identify chemicals and to build their spatial distribution models, with particular attention on the percentage of those which originate from the different inflowing rivers sources. They also developed a dedicated dashboard to ease the data visualization per detected signal/compound. The obtained results confirmed the model ability at prioritizing signals of unknown compounds, which is of fundamental importance to support prioritization activities in non-target screening. A recent study of Chen and Wang (2022) proposes a new oil-spill-detection model based on the U-Net architecture (Ronneberger et al., 2015) applied on data obtained by the Sentinel-1 Polarimetric Synthetic Aperture Radar (PolSAR). The latter provides rich polarimetric information which are combined to the wind-speed information of the sea surface during the training: this allows to focus more on feature extraction and to better distinguish oil films from look-alikes. The experimental results showed high detection accuracy with good preservation of the oil-spill patches boundaries; it also allowed to validate an improved accuracy of some degrees thanks to the use of wind-speed information.

In attempting to address the problem of plastic pollution Teng et al. (2022) proposed the use of a YOLOv5 architecture (Zhu et al., 2021) able to detect, classify and localize both marine debris and marine life in images and video recordings. This image classifier has been trained and tested on a dataset of images depicting plastic bottles, plastic bags, plastic buckets, fishing nets, plastic straws, food wrappings and a fish species. Then, they applied two counting methods, i.e., the ROI line counting and the centroid tracking methods, to accurately count the items number; the latter performed slightly better thanks to its ability in tracking the labelled bounding box through all the videos. Jiang et al. (2022) exploited five ML/DL techniques to predict the presence of the Pathogenic *Vibrio spp* on microplastics debris through published datasets from an estuary and a mariculture zone in China. The experiments showed that the Deep Neural Network and the random forest techniques obtained the best predictive performance, with some differences due to the data source and sampling and to the processing. This study also gave the possibility to verify that salinity and temperature are the primary factors affecting the growth of this bacterium, constituting a useful step towards the assessing of the health impacts of microplastics on the ecosystems. Finally, it is worth mentioning the DL application of Wu et al. (2022) which proposed a multi-scale fusion generative adversarial network named Fusion Water-GAN (FW-GAN) to enhance the underwater image quality. This very complex environment is characterized by light scattering and absorption other than suspended particles and other factors which affect the visual data acquisition. In their study, they conducted qualitative and quantitative comparison experiments on real-world and synthetically distorted underwater image datasets achieving higher quantitative metrics scores and better generalization capability. This places a new step towards the improvement of the underwater visual data processing, which will further enhance the information retrieval and the environmental monitoring techniques efficiency.

Table 1. VOSviewer terminology (Van Eck and Waltman, 2018).

Term	**Description**
Items	Chosen object (e.g., publications, researchers, organizations, keywords).
Link	Relation between two items (e.g., co-occurrence of keywords).
Link strength	Positive numerical value defining the attribute of each link. In the case of co-occurrence of keywords links, higher value means high number of publications reporting the keywords.
Network	Set of items connected by their links.

Cluster	Sets of items of the map. An item can be part of only one cluster.
Number of links	Number of links between two items.
Total link strength	The total links strength of a single items with the others.

2 Bibliometric Analysis methodology

The VOSviewer software (version 1.6.16) has been used to obtain the bibliometric network analysis data which will be presented in this paper. This tool allows to create clusters-based maps through the elaboration of bibliometric data related to the chosen topic of interest downloaded by specific online databases (e.g., Web of Science, PubMed, Scopus, etc), API (e.g., Crossref, Semantich Scholar, etc) or reference manager files (i.e., RIS, EndNote and RefWorks). The analysis can be made on the basis of five different networks type (i.e., Co-authorship, Co-occurrence, Citation, Bibliographic coupling and Co-citation) and of a set of items specific for each of them (i.e., Authors, Organizations or Countries for the Co-authorship factor). Table 1 shows the main technical terms of the software. The option “Full counting" or “Fractional counting" can be chosen to assign to each link the same weight or a weight proportional to the number of elements (for instance, if a document has 10 authors, each author’s link has a weight of 1/10); documents with more than *n* authors can be ignored, where *n* stands for a variable number set by the user. As an important factor in this kind of study, a thesaurus file (in .txt format) has been uploaded to avoid the multiple display of similar words or acronyms of the same term, which are then grouped under a unique word (as in the case of the term "GIS", which combines with “Geographic Information System” and “Geographic Information Systems”). The clusters-structured resulting maps can be sized on the basis of different weight attributes ("total link strength", "number of documents", "number of citations"). To modify the level of detail of the results and the consequent number of displayed clusters, the “resolution parameter” value can be set basing on the user’s needs: high value means high resolution and hence a high number of clusters.

The analysis presented in this work uses the co-authorship, co-occurrence and citation analyses weight attributes (Table 2) to provide the networks of: (1) the co-occurrence of keywords, the co-authorship among (2) researchers and (3) countries and the citation of (4) documents and (5) scientific journals. The resolution applied in this study is set as 1, as default of VOSviewer.

Table 2. VOSviewer performed analysis (Van Eck and Waltman, 2018).

Analysis type	Description
Co-authorship	The connection between researchers or countries is made through the number of jointly authored publications.
Co-occurrence	The number of co-occurrences is defined on the basis of the number of publications in which both the keywords occur together in the title, abstract or keyword list.
Citation	Two items are linked if at least one cites the other.

2.1 Bibliographic data acquisition

The data analysed in the context of this work have been collected from the Scopus database on October 1st, 2022. The search string has been thought to allow for a comprehensive review of the scientific literature on the Environmental Monitoring at sea exploited through AI, ML and DL techniques. Particular attention has to be paid in this step as the correct choice of the terms will affect the research results: in this case, the use of acronyms greatly improved the number of output documents but most of them were not in line with the focus of the research. For this reason, the final string included the following keywords: "Artificial Intelligence" OR "Deep Learning" OR "Machine Learning" AND "Environmental monitoring" AND "marine" OR "sea*" OR "ocean*" OR "beach*" OR "underwater". The OR - AND conjunctions are logical operators, and the “*” symbol is intended as a wildcard to account for both the singular and the plural term. This research gave as output a total of 499 documents, which have been exported into a single CSV file ticking the “Citation information”, “Bibliographical information”, “Abstract & keywords” and “Include references” options. This file has then been imported into the VOSviewer software selecting the following options: i. Create a map based on bibliographic data, ii. Read data from bibliographic database files, iii. Scopus and iv. file upload, to finally proceed with the analysis of interest.

3 Intellectual landscape and discussion

This section provides a description of the actual status of scientific research on environmental monitoring operations exploited through AI-ML-DL techniques. Interesting results and findings are generated by year, authors and journals, country/region and keywords, allowing to discuss research interests, considerations and trends regarding these interesting applications for the scientific community.

3.1 Temporal trend analysis

To analyse the temporal trend of the scientific research on marine environment monitoring exploited through AI, ML and DL, the publication year of each document has been plotted: Figure 1 reports the resulting graphic. From the chronological distribution of publications, it can be seen that the first document on the topic was published in 1989: Goodenough et al. (1989) analysed the importance of multiple-source integration in predicting future environmental states. In particular, they focused on the use of Artificial Intelligence as an essential technology for the scope, able to integrate multiple sources of images, maps, and ground measurements with the knowledge of diverse experts. Research in this field remained depressed for the following twenty years until 2016, when the trend shows an exponential increase which runs in parallel with the exceptional results obtained in the computer science field: from the 18 publications in 2016 it reached 91 papers in 2021. The current year, 2022, although still in progress, is supposed to confirm this trend, with a partial number of 127 publications at three quarters of its duration.

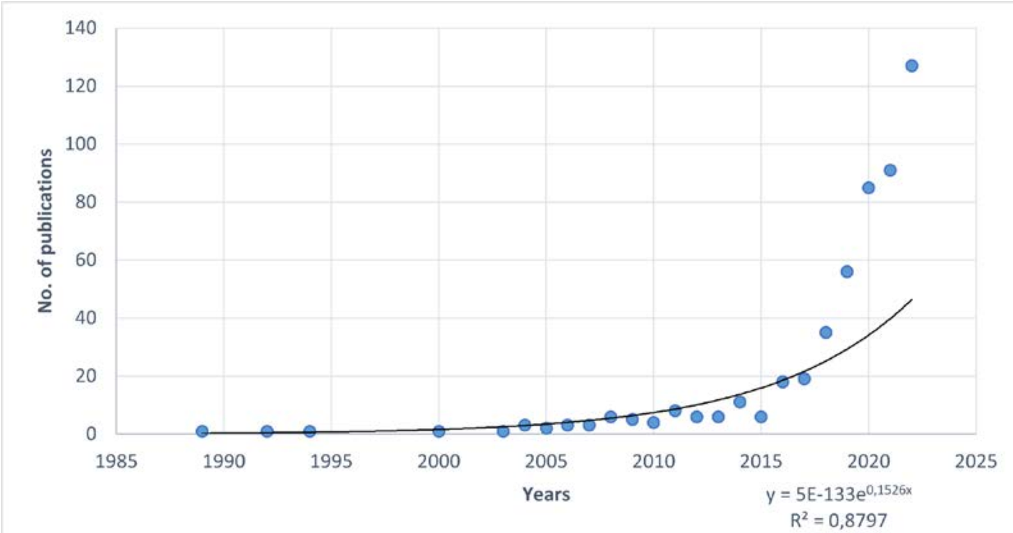

Figure 1 - Temporal trend of scientific articles published on “Environmental Monitoring" and “Artificial Intelligence". The complete research string is reported in 2.1.

3.2 Co-occurrence analysis of the keywords

The analysis of the co-occurrence of keywords produced 6542 results; the default software threshold has been applied, setting a minimum number of occurrences of the single keyword equal to 5. It resulted in 550 keywords which were grouped into 5 clusters, with the size of their indicators showing their frequency of occurrence (Fig. 2). Some brief observations can be made by analysing the clusters, each labelled by a distinct colour. The first cluster is in red and includes 143 keywords which mainly focus on the environmental monitoring of the atmosphere. Among the keywords, "emission control", "forecasting", "human", "health risk" and "particulate matter" suggest a great concern on the air quality and the several pollution sources, considering also the correlation with the human (and the biodiversity in general) health, from which the need of developing innovative and effective monitoring strategies based on the integration of new and traditional techniques. The second cluster (in green) contains 141 items, as for example "beaches", "coastal zones", "marine pollution", "object detection", "unmanned vehicle". These keywords highlight a deep connection between the coastal and sea/ocean monitoring techniques (which can vary from remote sensing to in-situ data acquisition and processing) and the recent excellent results obtained by the AI/ML/DL algorithms, with a particular focus to the marine pollution and its consequences on the environment. A total of 124 keywords are contained in the third cluster (in blue), which places more emphasis on the water analysis processes with a particular focus on its composition from a bio-chemical point of view: among the keywords: “aquatic ecosystem”, “bacteria”, “chlorophyll”, “phosphorus”, "rivers" and “water quality”. The fourth cluster (in yellow) comprises 88 keywords, in which the most predominant ones are "chemistry", "heavy metals", "sediment", “soil pollutants” and "water contamination": this clearly gives attention to the analysis of the soil composition and its pollution sources, with their consequences on the environment. Finally, the fifth cluster, in violet, is composed of 54 items and mainly focuses on the different data processing and transfer learning techniques recently developed and applied on the macro areas of interest.

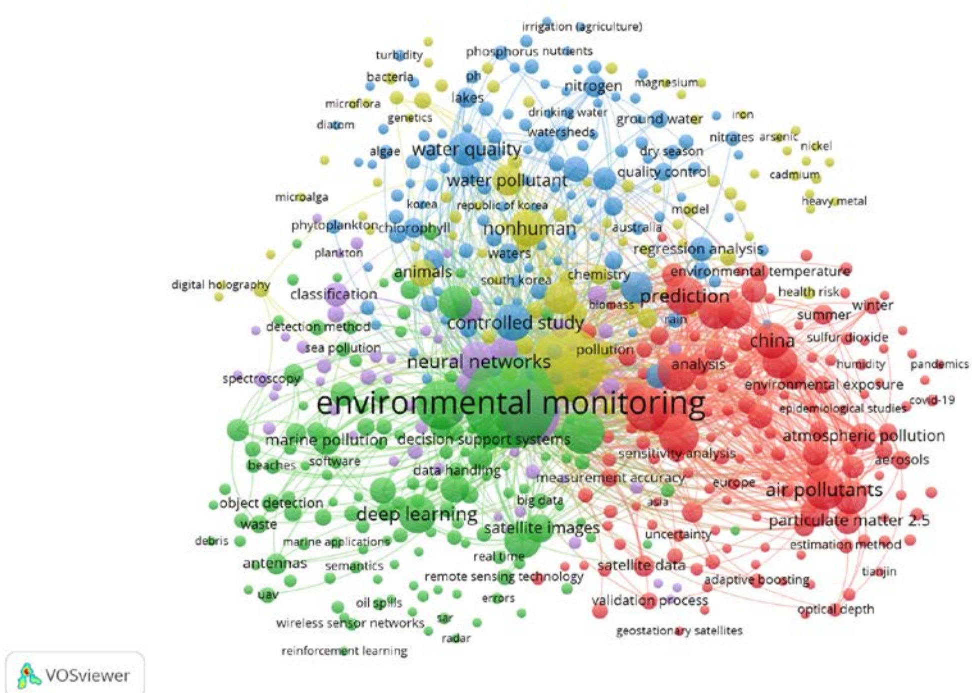

Figure 2 - Keywords network obtained from the co-occurrence analysis of the keywords.

The first 15 keywords ordered by “total link strength” are listed in Table 3, exception made for those used in the search string: keywords ranking higher by “total link strength” reflect the topics most related to the employment of ML-AI-DL for environmental monitoring. The high link strength proves the existence of a concrete interconnection between the learning algorithms of these techniques with the environment with the aim of monitoring and controlling, other than preventing, especially in the context of water and air pollution. Figure 3 reports the overlay visualization of the keywords map, showing their trends over time, highlighting the topics on which particular attention is being given in the last year: among them, the increasingly use of Deep Learning, and the monitoring of air and water pollution (due to waste, plastics and others) with its impact to the human health.

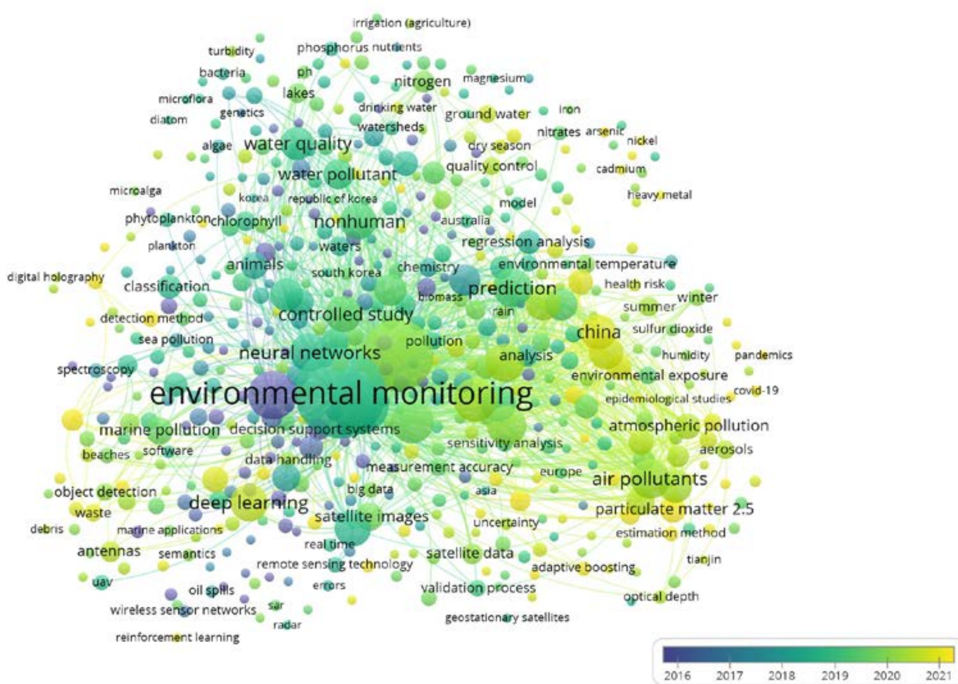

Figure 3 - Overlay visualization of the keywords network obtained from the co-occurrence analysis of the keywords.

Table 3. First 15 results of the co-occurrence analysis of the keywords, ordered by "total link strength".

Keyword	Total Link Strength	Occurrence
Learning methods	3876	191
Human	2445	109
Procedures	2218	102
Prediction	2193	86
Air pollutants	2174	90
Random forests	2133	85
Decision trees	2124	87
Controlled study	1942	75
Nonhuman	1919	76
Seasonal variation	1799	70
Concentration	1769	66
Neural Networks	1709	85
China	1693	70
Forecasting	1562	73
Particulate matter	1525	62

3.3 Co-authorship analysis of researchers' publications

Based on the bibliometric analysis, 2202 authors out of 499 documents are identified. Documents with a number of co-authors greater than 25 have been prior excluded, and the minimum threshold of an author's documents was set as one, with at least one citation. That resulted in 1851 authors, among which only the first 1000 for total link strength have been selected. Figure 4 shows the simplified author collaboration network, with each node in the map representing different authors and the links telling the exact status of co-authorship relations among authors. Overall, a sparsely connected cooperation network structure emerges. Table 4 reports the top 10 authors ranked by the total link strength: *Li Y.* is the most active author in terms of cooperation (66 links and 77 link strength), while *Wang J.* has the highest number of citations (471). To obtain a better view of the most active authors in the collaboration network, only the largest set of connected networks comprising 418 authors is analyzed, as illustrated in Fig. 5. A total number of 23 clusters is identified, among which 2 are totally separated and the other are strictly connected, highlighting a strong collaboration. The two most productive authors are in fact collocated in this area.

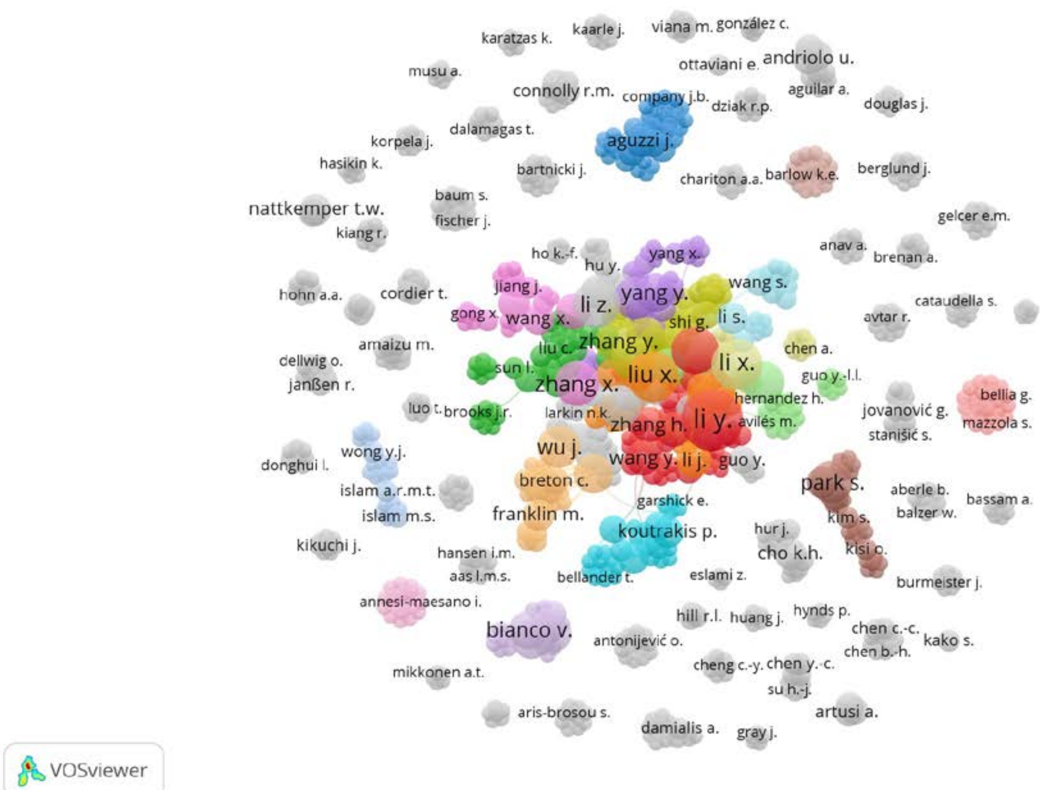

Figure 4 - Authors network obtained from the co-authorship analysis of the authors.

Table 4. First 10 items of the co-authorship analysis of authors ordered by number of documents.

Authors	Documents	Citations	Total Link Strength
Li Y.	15	226	77
Liu X.	10	69	58
Huang Y.	7	70	55
Li L.	6	120	54
Li Z.	9	454	48
Wang J.	5	471	46
Aguzzi J.	4	153	44
Wang L.	5	59	42
Wang Y.	5	248	42
Zhang Z.	5	62	41

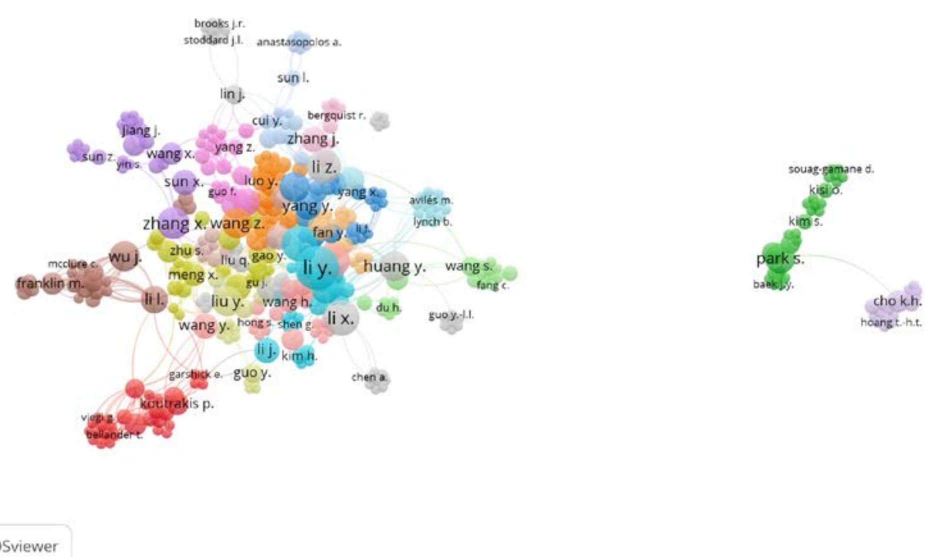

Figure 5 - Largest set of connected authors in the co-authorship network.

3.4 Co-authorship analysis of Countries

To analyse the international collaborations between influential countries/regions, a co-authorship analysis has been made: it revealed that 35 out of 92 countries published at least 5 articles on AI-ML-DL based techniques for environmental monitoring. Table 5 shows the first 10 countries ordered by total link strength. The results highlight the leading role of the *United States*, which is not only connected to more countries than the others (103), but also has the highest number of citations (2666). *China* and *United Kingdom* follows for the same categories, with *China* publishing the highest number of documents (136) together with the *United States* (115) and *Italy* (40). Figure 6 reports the 6 resulting clusters and highlights the strong interconnection among the cited Countries. It is easy to find that the willingness to cooperate is clear in all these countries despite the distance.

Table 5. First 10 items of the co-authorship analysis of Countries ordered by number of citations.

Countries	**Documents**	**Citations**	**Total Link Strength**
United States	115	2666	103
China	136	2473	80
United Kingdom	34	1278	60
Germany	27	716	43
Italy	40	920	42
Australia	37	1099	41
Spain	23	437	39
Canada	27	524	33
France	15	561	28
Switzerland	7	511	24

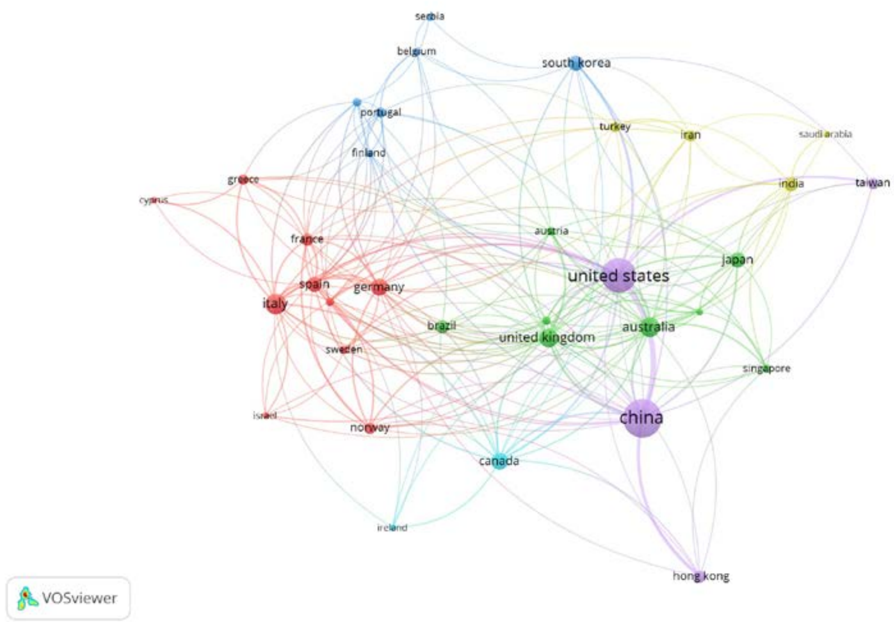

Figure 6 - Co-authorship analysis of the Countries: resulting network.

Table 6. Most highly cited articles ordered by number of citations.

Keyword	**Citations**	**Links**
Yuan Q. (2020)	353	1
Palani S. (2008)	340	5
Morawska L. (2018)	323	1
Strafoggia M. (2019)	198	3
Deo R.C. (2016)	129	0
Park Y. (2015)	128	6
Zhan Y. (2018)	127	3
Paschalidou A.L. (2011)	118	2
Huang K. (2018)	118	4
Freeman B.S. (2018)	114	1

3.5 Citation analysis of the documents

To comprehend the internal structure of the citation network on the topic, the citations analysis of publications is performed. Setting the minimum number of citations for the documents to 5, a total number of 264 items has been selected. However, only the largest set of connected items has been analysed, consisting in 18 documents as showed in Figure 7. With the help of the clustering algorithm embedded in VOSviewer, 5 clusters have been obtained. The largest one, in red, contains 5 documents among which the most cited one is Stafoggia et al. (2019) with 198 citations, published on Environment International Volume 258 124, March 2019, Pages 170-179. Entitled *Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model*, this study proposes a random forest-based model to estimate daily PM10, fine and coarse particles at 1-km2, as the particles are considered among the major death causes. The obtained results confirmed the possibility to use this method as a reliable tool for investigating long-term and short-term health effects. The second (green) and third (in blue) clusters, with 4 documents each, present two of the most cited articles: in the first case, the paper *Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment* has been published on Environmental Pollution Volume 233, February 2018, Pages 464-473 and is authored by Zhan et al. (2018): it has 150 citations. In the latter case, the paper *Predicting monthly high-resolution PM2.5 concentrations with random forest model in the North China Plain* of Huang et al. (2018b) has been published on Environmental Pollution Volume 242, Part A, November 2018, Pages 675-683 and has 137 citations. It should be highlighted that some highly cited papers do not appear in the figure because of the parameters set during the network generation. Thus, Table 6 specifies the top 10 most highly cited articles ordered by number of citations. This highlights that some of the most influencing papers on the topic do not have a strong connection with other researches (as confirmed by Deo R.C. et al (2016) with a total link strength of 0) and that in general this parameter should be improved by researchers.

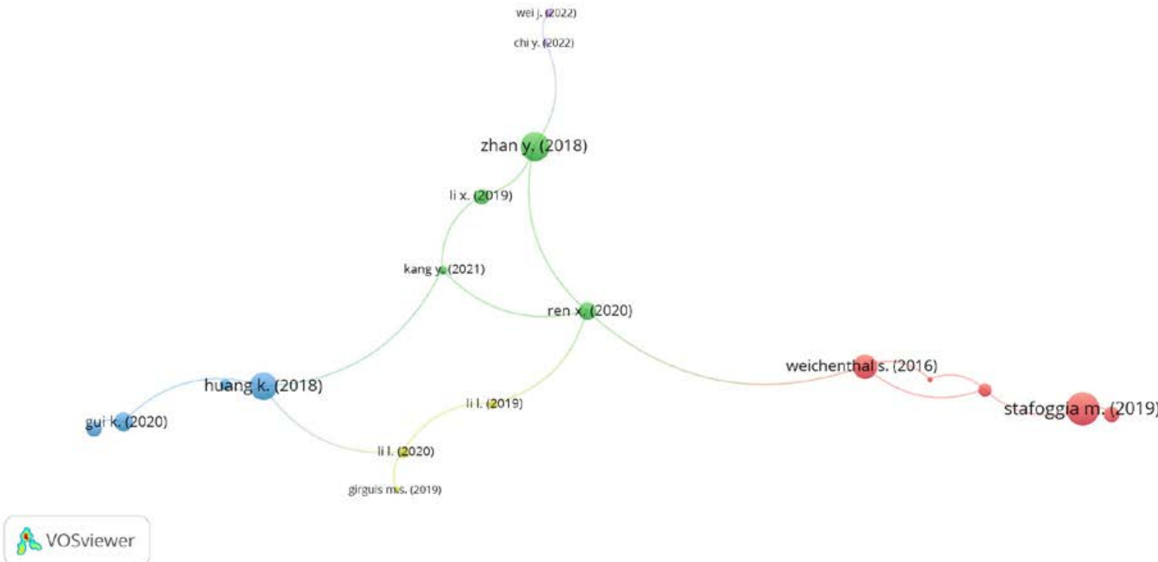

Figure 7 - Largest set of the most cited connected documents.

3.6 Citation analysis of journals

The citation analysis of journals aims at identifying the most influencing sources for the environmental monitoring exploited through AI-ML-DL techniques. The research resulted in 213 journals, among which 17 met the minimum threshold of 5 published articles. The first 10 journals ranked by number of citations are reported in Table 7, while Figure 8 shows the network of connected items, which contains a total of 15 journals. The most cited journals are *Science of the total environment* (1168 cit.), *Marine Pollution Bulletin* (888 cit.) and *Environment International* (706 cit.) while *Environmental Pollution* having the highest link strength (34) with 701 citations. These top three cited journals come from three different categories (3 different clusters), showing that they have developed their own unique research directions but, together with the others, have contributed indelibly to the development of this field.

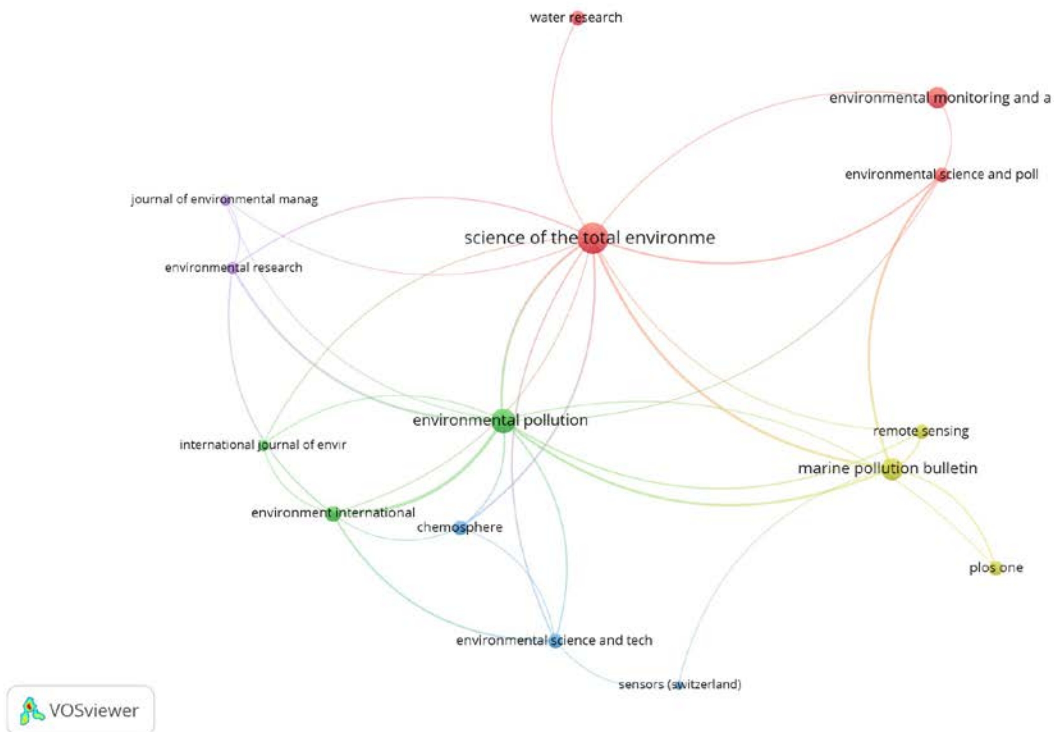

Figure 8 - Network obtained by the citation analysis of the journals.

Table 7. First 10 items of the co-authorship analysis of journals ordered by number of citations.

Source	**Documents**	**Citations**	**Total Link Strength**
Science of the Total Environment	53	1168	27
Marine Pollution Bulletin	26	888	27
Environment International	13	706	17
Environmental Pollution	31	701	34
Environmental Science and Technology	11	278	9

Environmental Monitoring and Assessment	24	271	2
Remote Sensing	11	224	6
Environmental Science and Pollution Research	11	180	10
Environmental Research	9	170	10
Plos One	11	167	3

4 Conclusions

This study investigates the use of Artificial Intelligence and Machine/Deep Learning techniques in the environmental monitoring with a particular focus to the marine ecosystems. It aims at providing useful information on the state of the art and on the current research interest with a temporal view on their evolution. The bibliometric analysis allowed for the quantification and visualization of the multi- and trans-disciplinary development obtained in the last decades, providing also new perspective on the different topics. Especially in the last years, the impressive achievements reached in the computer science field have paved the way to a variety of approaches to different disciplines and investigation techniques, which reflected also on the environmental monitoring operations. In this context, AI and ML/DL techniques demonstrated to have a huge potential and are more and more employed in several research field. Moreover, this study showed that this development occurred concurrently to the international political commitments related to the sustainable use of the global ocean. The United Nation SDGs are in fact drawing particular attention from the political and the scientific community, with the consequent race for the innovation and efficiency strategies which can guarantee zero impact on the ecosystems.

The analysed data have been downloaded from the Scopus Database by choosing a research string which included focused terms fitting with the topic. It has to be noticed that a proper selection of the research string is the core of this kind of research: in fact, it can positively affect the analysis by enhancing the number of output documents, but at the same time a careful choice should be made to avoid the inclusion of documents whose topic does not match with the study. In this context, a key role is played by the acronyms as frequently used terms in the research field, so the overall research results can greatly benefit from their use. The Scopus data have then been elaborated using the VOSviewer free software tool, which allows the creation of clusters-based maps representing the existing network among researchers, countries, organizations and keywords dealing with the topic of interest. The obtained results emphasize the ever-growing employment of AI and ML/DL methods within the several aspects of the environmental monitoring, in parallel with the continuous development and consequent enhancement of their architectures. By looking at the keywords network, different clusters can be identified, each corresponding to a different working area (i.e., human and non-human studies, water quality, pollution, remote sensing): this highlights the strong interconnection among the different tasks which cover the ecosystems and the biodiversity in all their complexity, further suggesting useful stratagems not only for environmental monitoring, but also for adaptive water resources planning, sustainable agriculture, ecosystem management and so on. The map obtained from the countries analysis shows a deep involvement in the research topics, together with a strong cooperation among them as showed by the link strength reported from the most productive countries. Finally, the outcome of the journal network could help scientists and young researchers to quickly search and choose the distribution of high-quality journals in the field, supporting their research promotion and strengthening the collaboration opportunity with other institutions.

In summary, it can be concluded that the current research on the Artificial Intelligence and Machine/Deep Learning techniques with their application to the enhancement of the environmental monitoring at sea has reached excellent achievements in the last years; promising and better results are yet to come, supported by the continual transfer, differentiation and regeneration among each diverse themes and by the collaboration among worldwide institutions, which work together for the deployment of innovative and effective strategies towards the SDG goals pursuit. Further studies will be made to deepen this research field, following its trend and improving its accuracy, with a particular focus on the research string and the use of the acronyms, as to provide a comprehensive and more detailed analysis.

Competing interests. The author has no conflicts of interest to declare that are relevant to the content of this article.

References

Alygizakis, N., Giannakopoulos, T., Thomaidis, N. S., and Slobodnik, J.: Detecting the sources of chemicals in the Black Sea using non-target screening and deep learning convolutional neural networks, Science of The Total Environment, 847, 157554, 2022.

Anderson, M. J.: Animal-sediment relationships re-visited: Characterising species' distributions along an environmental gradient using canonical analysis and quantile regression splines, Journal of Experimental Marine Biology and Ecology, 366, 16–27, 2008.

Bahrammirzaee, A.: A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems, Neural Computing and Applications, 19, 1165–1195, 2010.

Barnes, D. K., Galgani, F., Thompson, R. C., and Barlaz, M.: Accumulation and fragmentation of plastic debris in global environments, Philosophical transactions of the royal society B: biological sciences, 364, 1985–1998, 2009.

Bayat, B., Crasta, N., Crespi, A., Pascoal, A. M., and Ijspeert, A.: Environmental monitoring using autonomous vehicles: a survey of recent searching techniques, Current opinion in biotechnology, 45, 76–84, 2017.

Biel-Maeso, M., Baena-Nogueras, R. M., Corada-Fernández, C., and Lara-Martín, P. A.: Occurrence, distribution and environmental risk of pharmaceutically active compounds (PhACs) in coastal and ocean waters from the Gulf of Cadiz (SW Spain), Science of the Total Environment, 612, 649–659, 2018.

Borja, A., Basset, A., Bricker, S., Dauvin, J.-C., Elliot, M., Harrison, T., Marques, J., Weisberg, S., and West, R.: Classifying ecological quality and integrity of estuaries, 2012.

Boyer, J. N., Kelble, C. R., Ortner, P. B., and Rudnick, D. T.: Phytoplankton bloom status: Chlorophyll a biomass as an indicator of water quality condition in the southern estuaries of Florida, USA, Ecological indicators, 9, S56–S67, 2009.

Chen, Y. and Wang, Z.: Marine Oil Spill Detection from SAR Images Based on Attention U-Net Model Using Polarimetric and Wind Speed Information, International Journal of Environmental Research and Public Health, 19, 12315, 2022.

Cordier, T., Esling, P., Lejzerowicz, F., Visco, J., Ouadahi, A., Martins, C., Cedhagen, T., and Pawlowski, J.: Predicting the ecological quality status of marine environments from eDNA metabarcoding data using supervised machine learning, Environmental science & technology, 51, 9118–9126, 2017.

Costa, Y., Martins, I., Carvalho, G., and Barros, F.: Sea-level rise effects on macrozoobenthos distribution within an estuarine gradient using Species Distribution Modeling, Ecological Informatics, 71, 101816, 2022.

Deo, R. C. and S¸ahin, M.: An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland, Environmental monitoring and assessment, 188, 1–24, 2016.

Desa, U. et al.: Transforming our world: The 2030 agenda for sustainable development, 2016.

Di Ciaccio, F., Russo, P., and Troisi, S.: DOES: A Deep Learning-Based Approach to Estimate Roll and Pitch at Sea, IEEE Access, 10, 29307–29321, 2022.

Eidam, E., Sutherland, D., Ralston, D., Dye, B., Conroy, T., Schmitt, J., Ruggiero, P., and Wood, J.: Impacts of 150 years of shoreline and bathymetric change in the Coos Estuary, Oregon, USA, Estuaries and Coasts, pp. 1–19, 2020.

Elith, J. and Leathwick, J. R.: Species distribution models: ecological explanation and prediction across space and time, Annual Review of Ecology, Evolution and Systematics, 40, 677–697, 2009.

Freeman, B. S., Taylor, G., Gharabaghi, B., and Thé, J.: Forecasting air quality time series using deep learning, Journal of the Air & Waste 354 Management Association, 68, 866–886, 2018.

Glasgow, H. B., Burkholder, J. M., Reed, R. E., Lewitus, A. J., and Kleinman, J. E.: Real-time remote monitoring of water quality: a review of current applications, and advancements in sensor, telemetry, and computing technologies, Journal of experimental marine biology and ecology, 300, 409–448, 2004.

Goodenough, D. G., Robson, M. A., and Fung, K. B.: Expert Systems and Environnental Change-Information Extraction from Remote Sensing Data, in: 12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium,, vol. 2, pp. 823–828, IEEE, 1989.

Hartnett, M. and Nash, S.: Modelling nutrient and chlorophyll_a dynamics in an Irish brackish waterbody, Environmental Modelling & Software, 19, 47–56, 2004.

Huang, C., Chen, Y., Zhang, S., and Wu, J.: Detecting, extracting, and monitoring surface water from space using optical sensors: A review, Reviews of Geophysics, 56, 333–360, 2018a.

Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., Gu, D., and Liu, Y.: Predicting monthly high-resolution PM2. 5 concentrations with random forest model in the North China Plain, Environmental pollution, 242, 675–683, 2018b.

Jiang, J., Zhou, H., Zhang, T., Yao, C., Du, D., Zhao, L., Cai, W., Che, L., Cao, Z., and Wu, X. E.: Machine learning to predict dynamic changes of pathogenic Vibrio spp. abundance on microplastics in marine environment, Environmental Pollution, 305, 119257, 2022.

Kim, S.-K., Shin, J., An, S.-I., Kim, H.-J., Im, N., Xie, S.-P., Kug, J.-S., and Yeh, S.-W.: Widespread irreversible changes in surface temperature and precipitation in response to CO2 forcing, Nature Climate Change, 12, 834–840, 2022.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., and Fotiadis, D. I.: Machine learning applications in cancer prognosis and prediction, Computational and structural biotechnology journal, 13, 8–17, 2015.

Little, S., Wood, P., and Elliott, M.: Quantifying salinity-induced changes on estuarine benthic fauna: The potential implications of climate change, Estuarine, Coastal and Shelf Science, 198, 610–625, 2017.

Lu, D. and Weng, Q.: A survey of image classification methods and techniques for improving classification performance, International journal of Remote sensing, 28, 823–870, 2007.

Maloney, K. O., Buchanan, C., Jepsen, R. D., Krause, K. P., Cashman, M. J., Gressler, B. P., Young, J. A., and Schmid, M.: Explainable machine learning improves interpretability in the predictive modeling of biological stream conditions in the Chesapeake Bay Watershed, USA, Journal of Environmental Management, 322, 116068, 2022.

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., and Bethge, M.: DeepLabCut: markerless pose estimation of user-defined body parts with deep learning, Nature neuroscience, 21, 1281–1289, 2018.

Otte, E. and Rousseau, R.: Social network analysis: a powerful strategy, also for the information sciences, Journal of information Science, 28, 441–453, 2002.

Palani, S., Liong, S.-Y., and Tkalich, P.: An ANN application for water quality forecasting, Marine pollution bulletin, 56, 1586–1597, 2008.

Park, Y., Cho, K. H., Park, J., Cha, S. M., and Kim, J. H.: Development of early-warning protocol for predicting chlorophyll-a concentration using machine learning models in freshwater and estuarine reservoirs, Korea, Science of the Total Environment, 502, 31–41, 2015.

Pauna, V., Buonocore, E., Renzi, M., Russo, G., and Franzese, P.: The issue of microplastics in marine ecosystems: A bibliometric network analysis, Marine Pollution Bulletin, 149, 110612, 2019.

Prabowo, A. R., Tuswan, T., and Ridwan, R.: Advanced development of sensors' roles in maritime-based industry and research: From field monitoring to high-risk phenomenon measurement, Applied Sciences, 11, 3954, 2021.

Resolution, G. A. et al.: Transforming our world: the 2030 Agenda for Sustainable Development, UN Doc. A/RES/70/1 (September 25, 2015), 2015.

Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241, Springer, 2015.

Shen, D., Wu, G., and Suk, H.-I.: Deep learning in medical image analysis, Annual review of biomedical engineering, 19, 221, 2017.

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., De Hoogh, K., De'Donato, F., Gariazzo, C., Lyapustin, A., Michelozzi, P., Renzi, M., et al.: Estimation of daily PM10 and PM2. 5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model, Environment international, 124, 170–179, 2019.

Teng, C., Kylili, K., and Hadjistassou, C.: Deploying deep learning to estimate the abundance of marine debris from video footage, Marine 400 Pollution Bulletin, 183, 114049, 2022.

Thorne, K. M., Elliott-Fisk, D. L., Freeman, C. M., Bui, T.-V. D., Powelson, K. W., Janousek, C. N., Buffington, K. J., and Takekawa, J. Y.: Are coastal managers ready for climate change? A case study from estuaries along the Pacific coast of the United States, Ocean & Coastal Management, 143, 38–50, 2017.

Van Eck, N. and Waltman, L.: Manual for VOSviewer version 1.6. 8, CWTS Meaningful Metrics. Universiteit Leiden, 2018.

Wright, S. L. and Kelly, F. J.: Plastic and human health: a micro issue?, Environmental science & technology, 51, 6634–6647, 2017.

Wu, J., Liu, X., Lu, Q., Lin, Z., Qin, N., and Shi, Q.: FW-GAN: Underwater image enhancement using generative adversarial network with multi-scale fusion, Signal Processing: Image Communication, 109, 116855, 2022.

Zhan, Y., Luo, Y., Deng, X., Grieneisen, M. L., Zhang, M., and Di, B.: Spatiotemporal prediction of daily ambient ozone levels across China using random forest for human exposure assessment, Environmental Pollution, 233, 464–473, 2018.

Zhao, Z.-Q., Zheng, P., Xu, S.-t., and Wu, X.: Object detection with deep learning: A review, IEEE transactions on neural networks and learning systems, 30, 3212–3232, 2019.

Zhu, X., Lyu, S., Wang, X., and Zhao, Q.: TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2778–2788, 2021.