

Review

Integration of Machine Learning and Remote Sensing for Water Quality Monitoring and Prediction: A Review

Shashank Mohan ¹ , Brajesh Kumar ²  and A. Pouyan Nejadhashemi ^{1,*} 

¹ Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA; mohansh1@msu.edu

² Department of Computer Science and Information Technology, Mahatma Jyotiba Phule Rohilkhand University, Bareilly 243006, India; bkumar@mjpru.ac.in

* Correspondence: pouyan@msu.edu

Abstract: Aquatic ecosystems play a crucial role in sustaining life and supporting key green and blue economic sectors globally. However, the growing population and increasing anthropogenic pressures are significantly degrading terrestrial water resources, threatening their ability to provide essential socioeconomic services. To safeguard these ecosystems and their benefits, it is critical to continuously monitor changes in water quality. Remote sensing technologies, which offer high-resolution spatial and temporal data over large geographic areas, including surface water bodies, have become indispensable for these monitoring efforts. They enable the observation of various physical, chemical, and biological water quality indicators, which are essential for assessing ecosystem health. Machine learning algorithms are well suited to handle the complex and often non-linear relationships between remote sensing data and water quality parameters. By integrating remote sensing with machine learning techniques, it is possible to develop predictive models that enhance the accuracy and efficiency of water quality assessments. These models can identify and predict trends in water quality, supporting timely interventions to protect aquatic ecosystems. This paper provides a thorough review of the major remote sensing techniques for estimating water quality indicators (e.g., chlorophyll-a, turbidity, temperature, total nitrogen and total phosphorous, dissolved organic, total suspended solids, dissolved oxygen, and hydrogen power). It examines how machine learning can improve water quality assessments. Additionally, it identifies key research gaps in current methodologies and suggests future directions to address challenges in water quality monitoring, aiming to improve the precision and scope of these critical efforts.

Keywords: water quality; water quality monitoring; water quality parameters; remote sensing; machine learning; deep learning



Academic Editors: Chao Zhang, Ziqi Meng and Nanshan You

Received: 25 November 2024

Revised: 18 January 2025

Accepted: 20 January 2025

Published: 26 January 2025

Citation: Mohan, S.; Kumar, B.; Nejadhashemi, A.P. Integration of Machine Learning and Remote Sensing for Water Quality Monitoring and Prediction: A Review. *Sustainability* **2025**, *17*, 998. <https://doi.org/10.3390/su17030998>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Water resources are vital for the sustainability of life on Earth. Although water bodies cover 70.8% of Earth's surface, only 3.8% is freshwater. Moreover, 68% of the freshwater is trapped in glaciers [1]. Over the years, human-induced activities such as urbanization, deforestation, industrialization, mining, and agriculture have contaminated the surface water and groundwater [2,3]. The water quality is severely affected due to human-induced contamination especially in countries where domestic sewage and chemical waste are directly discharged into waterways without proper treatment [4,5]. Additionally, the excessive use of organic fertilizers, mismanagement of garbage, and growing population are other important factors deteriorating the water quality [6]. Meanwhile, climate change

and high concentrations of nitrogen and phosphorous intensify algae bloom formation in many water bodies with higher frequency. An increase in nitrogen and phosphorous also leads to the formation of water weeds and fish deaths. All these factors are thereby negatively affecting the water quality [7] and seriously threatening ecological functions.

According to a recent United Nations (UN) report, the world population is expected to grow to 9.7 billion in 2050 [8]. Therefore, access to safe and clean water becomes progressively crucial for achieving the sustainable development goals (SDGs) of the UN. Monitoring and maintaining freshwater quality is challenging for the international community as stated in the SDGs indicator 6.3.2 on good ambient water quality [9]. It becomes important to monitor the water quality to provide crucial information to government, policy makers, and other stakeholders. Water quality parameters provide an insight into the effect of climate change and anthropogenic activities on aquatic ecosystem [10]. Accordingly, proactive measures can be used to maintain water quality and protect water bodies.

Conventional methods for water quality assessment and monitoring are based on point sampling and subsequent laboratory analysis [11]. It requires the establishment of a network of hydrometric stations at water bodies such as rivers. Water samples are collected at regular intervals, which are sent to designated laboratories for analysis. A number of parameters are determined and analyzed for water quality assessment. Point sampling-based methods for water quality assessment are effective and may provide accurate information. However, these methods are highly expensive and time-demanding [12,13]. Furthermore, hydrometric stations are non-uniformly distributed and work intermittently due to a lack of infrastructure. Some sites may be inaccessible to collect field samples. Therefore, conventional methods fail to provide broader spatial coverage, and they are not suitable for synchronous assessment for a large area, as contamination varies spatially over surface water bodies [7].

Remote sensing, as part of geospatial technology, provides an alternative platform for surface water quality assessment. It can significantly reduce the time constraints and financial expenses associated with conventional field sampling-based methods [9]. It is able to overcome accessibility challenges and spatial constraints in monitoring water quality [14]. Additionally, remote sensing has the advantage of simultaneous observation over a large area [15]. Thus, remote sensing has proved to be an efficient tool for monitoring surface water quality due to its low cost, wider spatial coverage, higher temporal frequency, lower labor cost, and reasonable accuracy. However, remote sensing can not provide direct assessment of water quality. Instead, it provides aquatic environmental information [16,17] that can be used to estimate some important parameters. Optically active components (OACs) such as chlorophyll-a, colored dissolved organic matter, and suspended particulates can be obtained from aquatic environmental variables. But there are several other non-optically active components such as dissolved oxygen and nutrients, which are important to assess water quality. Therefore, some kind of modeling is required to simulate non-optically active components.

Remote sensing imagery provides the optical reflectance of the water surface in different wavelength regions. There are different remote sensing products including single-band and multi-band imagery, which provide spectral and inter-band information. The relation between spectral information and required nutrients is indirect and non-linear [16]. The traditional empirical models are unsuitable to simulate water quality parameters due to the non-linear relationship [18]. Machine learning models are highly suitable for non-linear and non-Gaussian systems. In recent times, remotely sensed imagery is broadly used with machine learning models to quantify and monitoring water quality.

This study aims to investigate the effectiveness of integrating machine learning and remote sensing for water quality monitoring. Specifically, it explores the parameters that

can be estimated using remote sensing, assesses their adequacy for effective monitoring, identifies the most useful remote sensing products, and evaluates the machine learning techniques best suited for this purpose.

2. Methodology

This paper reviews the applications of remote sensing and machine learning for water quality assessment. Three important sub-areas are identified in this regard: (i) water quality parameters that can be estimated or simulated using spectral information, (ii) remote sensing technologies and products useful for water quality assessment, and (iii) suitable machine learning techniques. This review work uses the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [19] approach for searching, organizing, and selecting the research studies. The last search was performed on 20 September 2024 on Elsevier, Google Scholar, IEEE, and Springer databases. The search included the following keywords: Water, Quality, Monitoring, Forecasting, Prediction, Assessment, River, Lake, Reservoir, Watershed, Fresh water, Remote Sensing, Satellite, Landsat, Sentinel, MODIS, Machine Learning, Support Vector Machine, Random Forest, Boosting Algorithm, and Deep Learning. The operators “AND”, “OR”, “LIMIT-TO”, and “PUBYEAR” were used to obtain the refined results. The search was restricted to the years 2018–2024. The query used to search the databases is given as follows:

Water AND Quality AND (Monitoring OR Prediction OR Forecasting OR Assessment) AND (Lake OR River OR Reservoir OR Watershed OR Fresh Water) AND (Remote Sensing OR Landsat OR Sentinel OR MODIS) AND (Machine Learning OR Support Vector Machine OR Random Forest OR Boosting OR Deep Learning) AND (LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2021) OR LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2023) OR LIMIT-TO (PUBYEAR, 2024))

The flow chart for the PRISMA-based methodology is shown in Figure 1. Initially, a total number of 1378 documents were produced by the search query. After the preliminary screening, 552 documents were identified as duplicate entries across different databases. The duplicate documents were removed, and 826 documents were retained after the preliminary screening. The title and abstract of each document were further screened, and only 194 documents were found qualified for retrieval. The papers on ground water, agriculture, socioeconomic environment, climate change, unmanned aerial vehicles, and languages other than English were excluded. Subsequently, 15 more articles were removed because they focused more on classification than on the prediction or assessment of water quality. Further, 7 relevant documents from other sources were added. It resulted in 186 relevant articles by meticulously searching renowned databases for the literature review.

Figure 2 illustrates the research trend and growth of articles in the SCOPUS database for remote sensing and machine learning-based research. The graph shows the slow growth in the volume of articles until 2018. Therefore, the search was restricted to 2018–2024 across all the databases. It is noted that water quality analysis frequently requires interdisciplinary collaborations. This is illustrated by the subject area analysis shown in Figure 3. The major contributing subject areas are Environmental Science (23.4%), Earth and Planetary Sciences (20.8%), Computer Science (10.4%), Agriculture and Biological Sciences (10.2%), and Engineering (8.9%). Figure 4 shows the top 10 most productive countries on the basis of publications pertaining to remote sensing-based water quality research using machine learning. China is the leading country with 62 documents, followed by United States (44), India (30), Italy (12), United Kingdom (12), Australia (8), Iran (8), South Africa (8), Canada (6), and Germany (6). Based on the articles included in this work, the most prolific journals are identified and listed in Table 1. The *Remote Sensing* journal is at rank 1 with 15 articles

included in this work. It is published by MDPI (Basel, Switzerland), having an impact factor of 4.2 and a cite score of 7.3, with 173,611 citations. It is followed by the *Remote Sensing of Environment* journal (Elsevier, Amsterdam, The Netherlands) with 8 documents, an impact factor of 11.2, a cite score of 17.6, and 46,552 citations. *Science of The Total Environment* (Elsevier), *Water* (MDPI), *ISPRS Journal of Photogrammetry and Remote Sensing* (Elsevier), *IEEE Journal of Selected Topics in Applied Earth Observations and Remote* (IEEE, Piscataway, NJ, USA), *Remote Sensing Applications: Society and Environment* (Elsevier), *Sensors* (MDPI), *Journal of Environmental Management* (Elsevier), *Environmental Science and Pollution Research* (Springer, Berlin/Heidelberg, Germany) are other journals in the list. Most of these journals have good impact factors. This review work is an attempt to investigate the following research questions.

- How effective is the integration of machine learning and remote sensing for monitoring water quality?
- What parameters can be estimated using remote sensing, and are these parameters sufficient for effective water quality monitoring?
- Which remote sensing products are useful for water quality estimation?
- Which machine learning techniques are best suited for remote sensing-based water quality monitoring?

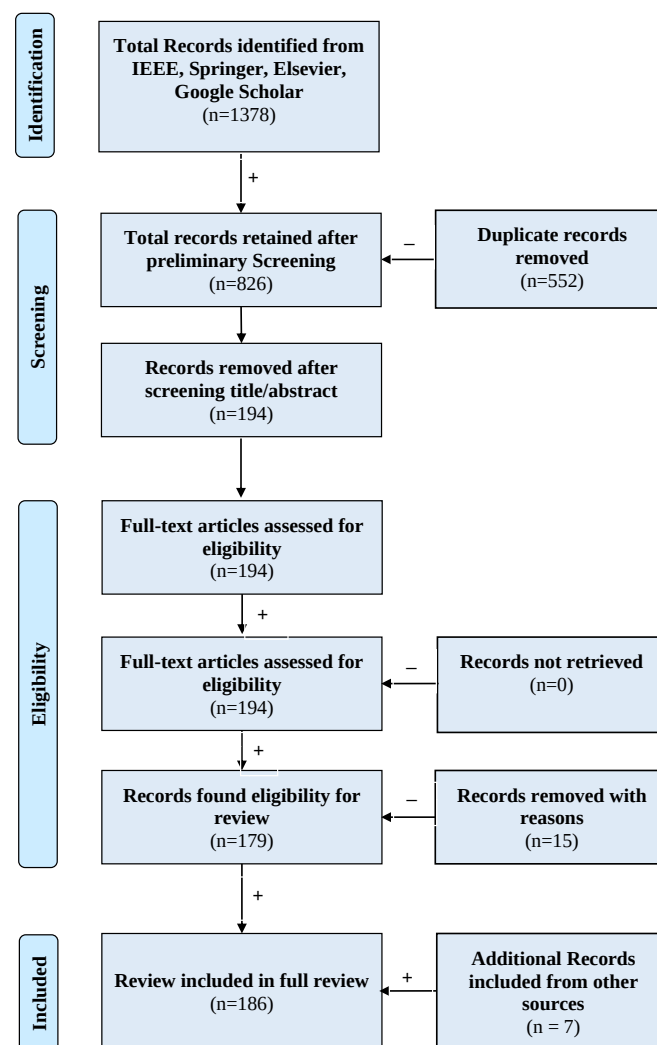


Figure 1. Procedure to explore literature using PRISMA methodology.

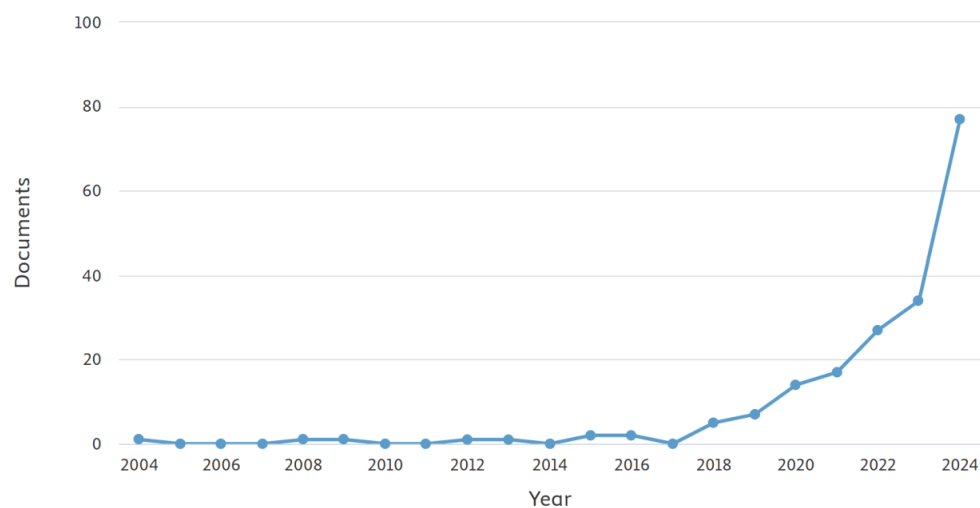


Figure 2. Research growth and historical trend of published articles.

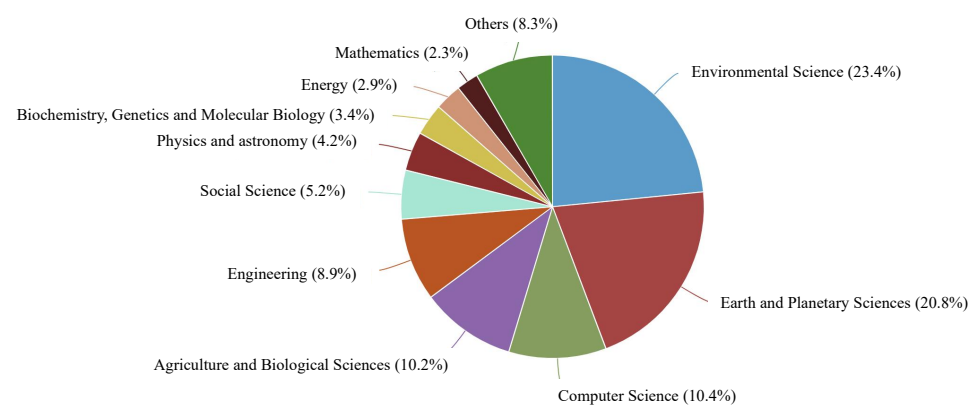


Figure 3. Contribution of different study areas in water quality assessment using remote sensing and machine learning.

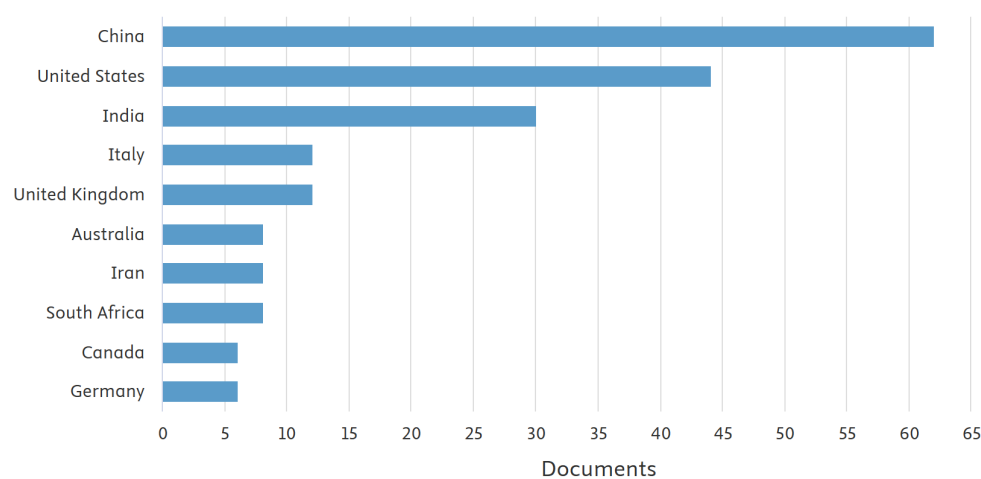


Figure 4. Most productive countries in water quality research.

Table 1. List of prolific research journals in water quality research.

#	Journal Name	Impact Factor	Cite Score	Citations	Publisher
1	<i>Remote Sensing</i>	4.2	8.3	173,611	MDPI
2	<i>Remote Sensing of Environment</i>	11.1	25.1	46,552	Elsevier
3	<i>Science of The Total Environment</i>	8.2	17.6	540,202	Elsevier
4	<i>Water</i>	3.0	5.8	89,517	MDPI
5	<i>ISPRS Journal of Photogrammetry and Remote Sensing</i>	10.6	21.0	21,901	Elsevier
6	<i>IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing</i>	4.7	9.3	27,333	IEEE
7	<i>Remote Sensing Applications: Society and Environment</i>	3.8	8.0	6053	Elsevier
8	<i>Sensors</i>	3.4	7.3	256,060	MDPI
9	<i>Journal of Environmental Management</i>	8.0	13.7	120,010	Elsevier
10	<i>Environmental Science and Pollution Research</i>	-	8.7	199,035	Springer

3. Remote Sensing Technologies for Water Quality Monitoring

Remote sensing has the potential to offer cost-effective solutions for monitoring water quality. Although remote sensing does not directly assess water quality, algorithms and mathematical models can be developed to assess water quality on remote sensing data [20]. Remote sensing offers frequent data collection across locations and time. In addition, image archives allow the analysis of past data to map water quality. Remote sensing systems are divided into two broad categories: active and passive. Active remote sensing systems use their own energy and can operate during both day and night. Passive satellites use the sun as the source of light and therefore can operate only during the day time. Satellites that provide global coverage are mostly placed in near polar sun-synchronous orbits at an altitude of 700–850 km above sea level. Various satellite missions provide high spatio-temporal resolution imagery useful for monitoring aquatic resources [9]. Spectral bands in different wavelength regions are useful to determine different water quality parameters.

Landsat imagery are widely used satellite data, and is freely available from the United States (US) Geological Survey (USGS). Landsat data can be downloaded from USGS Earth Explorer (<https://earthexplorer.usgs.gov/>, accessed on 28 December 2024) by entering the name of place or coordinates. Landsat mission was launched by National Aeronautics and Space Administration (NASA) in 1972. It was the first spaceborne digital sensor designed and developed for terrestrial environment monitoring [21]. Eight satellites have been successfully launched so far under the Landsat program, providing the longest digital terrestrial records managed and archived by USGS. Landsat imagery is widely used for water quality monitoring due to its ability to capture multiple spectral bands [22]. Landsat 5 was launched on 1 March, 1984 with a thematic mapper onboard. Landsat 5 and its successors orbit at an altitude of 705 km and have an inclination of 98.2°. Its temporal resolution is 16 days. The thematic mapper mounted on Landsat 5 provides seven-band multispectral data in the electromagnetic spectrum's visible, near-infrared, and thermal regions in the frequency range of 0.5–1.1 μm . The spatial resolution of bands 1–5 and 7 is 30 m. Band 6 has a spatial resolution of 120 m. Different spectral bands of Landsat 5 and band ratios can be used to assess various water quality parameters [23]. The major parameters include turbidity, chlorophyll-a, algal bloom concentration, water temperature, and Secchi depth. Landsat 6 was launched on 5 October 1993 but never achieved orbit. Later, Landsat 7 was launched on 15 April 1999 with a new instrument enhanced thematic mapper plus (ETM+) and a new panchromatic band. Landsat 7 has spectral capabilities similar to those of Landsat 5 with enhanced radiometric and geometric characteristics. Its ETM+ instrument has eight spectral bands [22]. Bands 1–7 have a spatial resolution of 30 m, while the spatial resolution of band 8 is 15 m. The revisit time of Landsat 7 is 16 days.

Landsat 8 was launched in 2013 carries an Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS). It provides eleven-band imagery with better radiometric, spatial, and temporal resolution [24]. The spectral bands 1–7 and 9 have a 30 m resolution. The spatial resolution of band 8 is 15 m, while bands 10 and 11 have a resolution of 100 m. Landsat 8 has a temporal resolution of 8 days. Landsat 9 is the latest satellite in the series, with improved versions of OLI and TIRS onboard, known as OLI-2 and TIRS-2, respectively. The radiometric resolution has been improved to 14-bit from 12-bit for Landsat 8 [25]. The number of bands in Landsat 9 imagery and their resolution is the same as that in Landsat 8. Both Landsat 8 and Landsat 9 data are used for water quality assessment due to their technical superiority. However, Landsat sensors cannot penetrate through clouds, limiting their effectiveness during monsoon season.

González-Márquez et al. [24] evaluated the viability of Landsat 8 to estimate water quality in the EI Guajaro Reservoir. They considered turbidity, Ph, dissolved oxygen, and electrical conductivity of the water to evaluate. The author found that reliable models can be developed using regression techniques to estimate water quality parameters using Landsat 8 data. Vakili and Amanollahi [26] determined optically inactive components including total nitrogen and total phosphorous and other parameters Chl-a, Secchi depth, and total suspended solids in Geshlagh Reservoir using Landsat 8 OLI data. They applied linear regression and artificial neural network (ANN) models to determine the parameters. ANN performed better than linear regression in estimating total phosphorous and total nitrogen with $R = 0.81$ and $R = 0.93$, respectively. Band ratio (band 3/band 2) and individual bands (3 and 4) were found to be the most suitable spectral data to estimate Chl-a. Chen et al. [27] investigated the color of water bodies in China using Landsat 8 imagery as color is an important indicator of water quality. A water color product Forel-Ule index was produced using the best available pixel composite algorithm and Google Earth Engine platform. The product obtained had shown high consistency ($R^2 = 0.90$) for surface reflectance and Secchi datasets. Guo et al. [28] used Landsat 8 imagery to monitor the water quality of Lake Simcoe by estimating Chl-a, total phosphorous (TP), and total nitrogen (TN). The authors adequately estimated the parameters with a mean absolute error of 32.57–42.58%. Markogianni et al. [29] investigated the potential of Landsat 8 OLI images to monitor the water quality of Trichonis Lake. They determined several parameters with an R value up to 0.7 and found Landsat 8 promising but challenging for monitoring fresh water quality. Al-Shaibah et al. [30] correlated Landsat 8 estimated and measured water quality parameters for Erlang Lake using regression techniques and obtained R^2 up to 0.95. Yang et al. [31] used Landsat 8 imagery to monitor the water quality of the Yangtze River in terms of Chl-a, TN, and TP. The regression-based experiments showed that Chl-a, TN, and TP were estimated with a mean absolute percentage error (MAPE) of 25.88%, 4.30%, and 8.37%, respectively. Zhang et al. [32] assessed several water quality parameters including biological oxygen demand (BOD), chemical oxygen demand (COD), permanganate of chemical oxygen demand (CODMn), and ammonia ($\text{NH}_3\text{-N}$) in addition to Chl-a, TN, and TP by incorporating Landsat 8 images and deep learning techniques. The authors analyzed changes in water quality over six years in the Yellow River. They obtained R^2 up to 0.92 between in situ and derived parameters suggesting good performance. Liao et al. [33] derived surface reflectance from Landsat imagery from different satellite missions to investigate sand mining activities in Poyang Lake from 1989 to 2021 by assessing suspended sediment levels using various regression techniques. They found that sand mining activities in Poyang Lake began after 2000 and were at their peak in 2016. Shamloo and Sima [34] demonstrated the potential of Landsat-8 imagery to monitor water quality using machine learning models in Lake Urmia. The results indicated that atmospheric corrections can improve the performance of the machine learning models.

Rajaveni et al. [35] demonstrated a strong correlation between Landsat-9 OLI reflectance and several water quality parameters. Regression analysis was used to assess the water quality of Vaigai River using Landsat-9. Band 1, Band 3, and Band 4 of OLI were found suitable for estimating different parameters.

The Sentinel program of the European Space Agency aims at different aspects of Earth observation including land, atmospheric, and oceanic monitoring. Sentinel-2 mission is primarily used for land monitoring and provides wide-swath high-resolution multispectral imagery to monitor soil, vegetation, and coastal regions. It is a bi-satellite system encompassing two satellites, Sentinel-2A and Sentinel-2B, which were launched in 2015 and 2017. Both satellites complement each other for a short revisiting period of 5 days [36]. The swath width of the system is 290 km. Sentinel-2 provides a notable advantage over some of its counterparts by offering relatively high spatial resolutions of 10, 20 and 60 m [37]. Thirteen spectral bands in multispectral imagery range from visible to short-wave infrared (443 to 2190 nm) in the wavelength region. The spatial resolution of its blue (bands 2), green (band 3), red (band 4), and near-infrared (band 8) spectral channels is 10 m [38]. Red-edge (band 5), other near-infrared (bands 6, 7, 8A), and short-wave infrared (bands 11 and 12) have a pixel size of 20 m. The ground sampling distance of coastal aerosol (band 1) and cirrus (band 10) channels is 60 m. The Sentinel-2 data are used in various applications including water quality monitoring. Sentinel data are freely available at Copernicus Data Space Ecosystem (<https://dataspace.copernicus.eu/>, accessed on 28 December 2024). The data can be obtained for a specific place by providing the location and date.

The Sentinel is one of the most widely used remote sensing products for water quality assessment. Meng et al. [37] used Sentinel-2 data to assess the water quality levels of Lake Dianchi from November 2022 to April 2023. They developed algorithms to study the relationship between water quality, area of algal bloom, and meteorological factors. They found TP to be the primary factor for algal bloom formation. Meteorological factors also have significant impact on the area of algal blooms. Zhang et al. [39] monitored the water quality in Lake Chl-a with the help of Sentinel-2 images. They retrieved Chl-a in 3067 global lakes from 2019 to 2021 with $R^2 = 0.74$. The authors also established a positive correlation between surface temperature and Chl-a. Ndou [40] investigated the water quality indicators including EC, TDS, pH, and turbidity of reservoirs in South Africa using Sentinel-2 imagery. The authors recommended red, green, shortwave infrared, and near-infrared spectral bands of Sentinel-2 to investigate EC, turbidity, TDS, and pH respectively. Zhang et al. [36] monitored and traced the water quality of seven major rivers in China in terms of permanganate index, TN, and TP using Sentinel-2 satellite images from 2016 to 2021. They obtained correlation coefficients of 0.68, 0.82, and 0.7 for permanganate index, TN, and TP, respectively. In addition, the results demonstrated that these parameters were high downstream and low upstream. Caballero and Navarro [41] evaluated the suitability of Sentinel-2 imagery for Lake Laguna water quality prediction during the Pacific typhoon season in 2020. The parameters Chl-a and total suspended matter (TSM) were estimated and compared with the pre-typhoon situation. The authors reported that the typhoon delivered a high sediment load to the lake as observed in the results. Virdis et al. [42] used 65 Sentinel-2 images to estimate the water quality parameters of tropical riverines in South East Asia from 2018 to 2021. Six parameters Chl-a, colored dissolved organic matter (CDOM), turbidity, transparency, and in situ reflectance spectra were determined and validated against observations collected at monitoring stations. The results demonstrated that remote sensing worked well for tested parameters except CDOM. It was concluded that determining CDOM in riverine tropical waters is challenging, and frequent spatial and temporal sampling is needed. Chu and He [43] retrieved Chl-a from Sentinel-2 images to assess the spatial distribution of water quality in Lake Kasumigaura. The blue–green ratio

was reported to be important for retrieving Chl-a in low concentrations, while NIR-red is suitable for high concentrations. German et al. [44] characterized the spatial and temporal distributions of Chl-a in the San Roque Reservoir in Argentina using a temporal series of Sentinel-2 from 2016 to 2019. The study found a significant fit between the measured and NIR-red bands ratio of Sentinel-2 with $R = 0.77$. A correlation between algal blooms and Chl-a level was also investigated by remote sensing data. Maimouni et al. [45] monitored low water turbidity variations in Sidi Moussa coastal lagoon Morocco using one-year multitemporal Sentinel-2 imagery. A correlation was established between the turbidity and red band information with an accuracy of $R^2 = 0.87$. The analysis showed that turbidity varies differently upstream and downstream. The turbidity map showed that turbidity is higher in the rainy season. Xie et al. [46] investigated the seasonal variation in TSM in Nanyi Lake from 2018 to 2022 using Sentinel-3 OLCI imagery. Significant regularity was observed with the spatial and temporal changes in TSM. The seasonal variations were observed owing to anthropogenic activities. Khan et al. [47] estimated the Secchi depth in the Finger Lakes using Sentinel-2 data. They trained and validated the model with separate datasets and observed R^2 up to 0.84.

The Moderate Resolution Imaging Spectroradiometer (MODIS) is an important sensor mounted on two satellites namely Terra and Aqua, which was launched by NASA [48]. Terra is timed so that it orbits the Earth north to south and passes over the equator in the morning. At the same time, Aqua orbits south to north across the equator in the afternoon. Terra-MODIS and Aqua-MODIS observe the entire Earth's surface every 1–2 days and deliver data with a swath of 2330 km. Both satellites reduce the revisit time to 0.5–1 day [49]. The MODIS instrument collects 36-band data spanning 412–14,200 nm wavelength region [50]. MODIS performs post-processing level 2 by making atmospheric corrections to the data. They are comparatively coarser data at various spatial resolutions including bands 1–2 at 250 m, bands 3–7 at 500 m, and bands 8–36 at 1000 m. Despite the coarser resolution, MODIS offers several advantages, as it visits a location twice a day, and there is no need for atmospheric corrections and mosaicing. MODIS data are made available freely by NASA through its Earth Science Data and Information System (ESDIS) project, which can be accessed at <https://www.earthdata.nasa.gov/data/instruments/modis>, (accessed on 28 December 2024).

Singh et al. [51] evaluated the potential of MODIS imagery to assess the water quality by measuring the nitrate concentration and dissolved phosphorous. The authors assessed the nitrate concentration with an accuracy of 80%. However, the MODIS-based prediction model achieved limited success with 51% accuracy in assessing the dissolved phosphorous. The results established that MODIS data can potentially be used for water quality monitoring. Xia et al. [52] developed a Forel Ule water color index (FUI)-based model for monitoring the water quality of inland lakes from 2000 to 2022. Although, MODIS provides rich data support to determine FUI, it only has three bands in the visible region. Therefore, they proposed a modified version of FUI by integrating the RGB and interpolation methods. The predictions were validated with on-site measurements, and excellent results were reported with MRE of 1.71% and root mean square error (RMSE) of 3.63°. Ariman [48] used the MODIS data to determine TN and TP for monitoring water quality in Balik Lake during 2017–2019. TP was determined more accurately than TN with $R = 0.59$. Ayana et al. [53] generated a 10-year time series data from 2000 to 2009 for total suspended solids (TSS) emissions by the inflowing rivers to Lake Tana using MODIS data. The empirical relationship was applied to validate a SWAT hydrological model. A Nash–Sutcliffe efficiency of 0.80 was observed during the validation period. Katlane et al. [54] observed the sea surface using the MODIS aqua data from 2005 to 2020 obtained via the Google Earth Engine platform. The recorded trends were validated by the measured data for the Gulf of Gabes

to analyze temperature, turbidity, and Chl-a. It was found that MODIS observations are effective in monitoring the water quality parameters. Kim et al. [55] investigated the water quality indicators including Chl-a and suspended solids on the western coast of the Korean Peninsula using the MODIS aqua data. The authors observed strong seasonal trends due to different dynamics and environmental settings. It was reported that the temperature highly influences Chl-a, while suspended solids are influenced by local tidal forcing and bottom topography. Rahat et al. [56] used MODIS data to evaluate the effect of climate change on the Ohio River water quality. The authors accurately determined TSS measurements over a large region and that remote sensing-based machine learning models can be used to study water quality with confidence. Kim et al. [57] estimated DO using MODIS data in coastal regions of Yellow Sea. They demonstrated that MODIS imagery can be used to effectively monitor DO levels in coastal environments.

4. Comprehensive Overview of Water Quality Parameters

Water quality can be determined by analyzing the physical, chemical, and biological properties of the water. Remote sensing-based methods mostly target OACs for water quality monitoring [58]. OACs can absorb and scatter the light significantly influencing the spectra recorded by the sensors in the optical region. Recently, people have also started exploring other spectrum band multispectral and hyperspectral imagery for this purpose. Some important physicochemical water quality parameters can be determined using the spectral information provided by the sensors.

4.1. Chlorophyll-a

Chlorophyll-a (Chl-a) [59] is a photosynthesis pigment responsible for green color in plants, algae, and cyanobacteria. It is one of the most widely evaluated parameters in remote sensing-based methods. Burning fossil fuels and fertilizer runoff are sources of nitrogen and phosphorous providing an impetus to algae growth in water bodies. Although not all algal blooms are harmful, blooms may contain some harmful species including cyanobacteria that are toxic to humans and animals [60]. Algae growth causes the eutrophication and degradation of water bodies. The presence of Chl-a in water bodies indicates eutrophication. The Chl-a concentration reflects light in the green (0.5 μm) and near-infrared (0.8 μm) spectral bands, which is a useful property for sensors to detect the presence of Chl-a in water. Cao et al. [61] developed a deep learning model to predict Chl-a in turbid lakes in Eastern China. The model produced satisfactory results with mean uncertainties of 28%. Khan et al. [62] developed an ensemble of adaptive boosting, random forest, and SVM to estimate Chl-a in the Finger Lakes using Sentinel data. The ensemble model observed R^2 of 0.76. Li et al. [63] integrated spectral and temporal information to retrieve Chl-a in Lake Balaton using Landsat data with RMSE 6.41 mg/m^3 . Karimi et al. [64] used semi-empirical methods to assess Chl-a concentration in Chitgar Lake using Landsat-8 and Sentinel-2 reflectance data. They developed a two-band algorithm that uses a ratio of two spectral bands to estimate Chl-a. The green and red bands ratio provided the best results for Landsat-8 imagery with $R^2 = 0.82$. Meanwhile, the red to red-edge ratio yielded the best results for Sentinel-2 data with $R^2 = 0.81$. Zhang et al. [32] applied deep learning techniques to estimate Chl-a concentration in Dongping Lake using Landsat-8 data. Markogianni et al. [29] assessed the potential of Landsat-8 products to remotely determine the Chl-a concentration and found Landsat-8 to be a promising technology for water quality assessment. Mansaray et al. [65] compared the capabilities of Landsat-8, PlanetScope, and Sentinel-2 satellites for assessing water quality parameters, including Chl-a. The authors found that Landsat-8 and Sentinel-2 better estimate the Chl-a concentration. He et al. [31] studied the Chl-a variation in Yangtze River from 2014 to 2020 with the help

of Landsat-8 images. Niroumand-Jadidi et al. [66] compared the performance of Landsat-9 and Sentinel-2 for water quality parameter estimation including Chl-a and found that Landsat-9 had higher accuracy ($R^2 = 0.89$) than Sentinel-2 ($R^2 = 0.71$). Yao et al. [67] used Chl-a as a quality indicator to analyze changes in water quality for the Miyun Reservoir from 2013 to 2022. They used optical images from multiple satellite missions and correlated the Chl-a concentration with water body erosion. A water quality contrast factor was defined to analyze water quality using a degree of erosion. Ahmed et al. [17] included Chl-a among 12 parameters to monitor the water quality of the Tigris River using multispectral satellite imagery. The authors used correlation analysis to establish a relation between Chl-a concentration and Landsat-8 spectral bands. Leggesse et al. [68] predicted Chl-a and other optically active parameters in tropical highlands of Ethiopia with the help of machine learning methods using Landsat-8 images. Guo et al. [28] included Chl-a in their study of water quality in Lake Simcoe using Landsat-8 images. Katlane et al. [54] used MODIS data to monitor Chl-a and other water quality parameters in sea surface water in Tunisian from 2005 to 2020. Kim et al. [55] studied the seasonal trends of the Chl-a concentration in the Yellow Sea using MODIS data and found the Chl-a concentration to be closely related to temperature.

4.2. Turbidity

Turbidity refers to the cloudiness or haziness of water caused by suspended particles, which can affect both living organisms and non-living components of aquatic ecosystems [69]. While not directly harmful to human health, high turbidity levels often indicate poor water quality and can conceal pathogens such as *Cryptosporidium*. To ensure water safety and effective disinfection, the World Health Organization (WHO) advises that turbidity levels remain below 1 Nephelometric Turbidity Unit (NTU) before the chlorination process [70]. Turbidity can be determined using remote sensing by utilizing light spectra such as green and red bands to form a connection between turbidity and reflectance value, enabling accurate turbidity retrieval [71]. Lioumbas et al. [72] used turbidity to monitor the water quality in the Polyphytos Reservoir. The authors measured turbidity using Sentinel data with the help of an algorithm recommended by Potes et al. [73]. Gu et al. [74] proposed a random forest ensemble model to measure river turbidity. They introduced an error-minimization-based pruning algorithm to tune the random forest to obtain higher accuracy. Devi and Mamatha [75] predicted water turbidity levels in Saki Lake from 2014 to 2017 using Landsat-8 data and regression algorithms. The decision tree-based regression model predicted water turbidity with good accuracy and obtained $R^2 = 0.776$. Zhang et al. [76] fused eight different algorithms to develop a robust model to determine Chl-a concentration and turbidity in Nansi Lake from 2016 to 2022. They used the normalized difference turbidity index (NDTI) and four other combinations of spectral bands from Sentinel data to predict turbidity. The results demonstrated that stacking model and Sentinel-2 data can reliably predict the water turbidity and Chl-a. Singh et al. [77] monitored water turbidity in the Bisalpur wetland using Landsat-8 and Landsat-9 data both in pre-monsoon and post-monsoon seasons from 2013 to 2022. They investigated the relationship of turbidity with surface temperature and Chl-a concentration. NDTI, normalized difference (NDCI), and NIR bands were applied to perform the analysis. It was found that water turbidity has a strong correlation with the surface temperature and Chl-a concentration. Ma et al. [78] investigated several regression methods to accurately predict turbidity in inland lakes. They found that gradient boosting and RF methods were effective in estimating turbidity. Maimouni et al. [45] developed a model for spatio-temporal turbidity monitoring in Sidi Moussa lagoon using Sentinel-2 data. Good accuracy was obtained for turbidity estimation with $R^2 = 0.87$ using the red band. It was observed that turbidity was lower in the dry

season as compared to the rainy season. Clermont et al. [79] assessed impact of land use, especially agricultural activities, on turbidity in lake St-Pierre. The authors developed a turbidity retrieval model-based NIR and red bands in Sentinel-2 data. The results showed that turbidity increases with agriculture perturbations. Ramesh et al. [80] predicted turbidity levels in Dooskal Lake using an ensemble deep learning model and satellite data with 92.7% accuracy. By achieving low prediction errors, they demonstrated that potential of remote sensing and machine learning to effectively monitor turbidity levels in urban areas. with 97.2% accuracy. Li et al. [81] developed several algorithms to monitor the dynamics of water turbidity across 73 lakes in China using Sentinel-2 data. Among the different algorithms, an RF-based algorithm obtained the highest accuracy at $R^2 = 0.92$. Mansaray et al. [65] investigated the utility of PlanetScope satellite data compared with Landsat-8 and Sentinel-2 data to analyze the impact of agriculture runoff on water quality by detecting turbidity along with Chl-a and Secchi depth. The investigation was carried out for 13 reservoirs in Oklahoma over three years from 2017 to 2020. The results exhibited that PlanetScope data are more suitable for water quality monitoring in small ponds and reservoirs due to better spatial resolution.

4.3. Temperature

Surface water temperature is an important parameter affecting water's biological processes and chemistry [82]. Temperature plays a crucial role in water quality by influencing various physico-chemical and microbial aspects. Studies have shown that temperature variations can impact water quality parameters such as total nitrogen, total phosphorus, dissolved oxygen, and microbial communities in water bodies and distribution systems [83–85]. Higher temperatures can lead to decreased dissolved oxygen levels, increased nutrient concentrations, changes in microbial community structures, and the potential release of pathogens into drinking water [86]. The conventional methods to study water temperature are expensive, time consuming, and unsuitable for spatially large water bodies. However, thermal infrared remote sensing can effectively estimate the water surface temperature.

Zhu et al. [87] established a correlation between temperature and other water quality parameters with the help of an ensemble model. It was demonstrated that Chl-a is sensitive to temperature, while dissolved oxygen is negatively correlated to temperature. However, turbidity was found to be insensitive to temperature. Vanhellemont [88] used various methods to estimate water surface temperature using Landsat-8 data. An open source model libRadtran was used for atmospheric correction. The results demonstrated that a single-band method is suitable for temperature retrieval. Katlane et al. [54] developed a workflow for monitoring water quality parameters including the temperature of Tunisian coastal waters using MODIS data. It was demonstrated that the proposed approach is effective for qualitative and quantitative validation. Krishnaraj and Honnasiddaih [89] developed a machine learning model for predicting temperature and several other water quality parameters in Ganga River Basin using Landsat-8 images. At different locations in the Basin, the temperature was accurately estimated with R^2 of 0.82–0.95.

4.4. Total Nitrogen and Total Phosphorous

The total nitrogen (TN) in water quality refers to the sum of all nitrogen forms present, including organic, inorganic, and ammonia nitrogen. TN is an optically inactive parameter, closely related to land use and population density [36]. Monitoring TN is crucial for assessing water quality and potential environmental impacts [90]. Total phosphorus (TP) is another optically inactive parameter, which is a crucial indicator of water quality that can significantly impact the overall health of aquatic ecosystems. The major sources of TP are largely anthropogenic including agriculture, domestic, and industrial waste waters.

TP monitoring plays a vital role in predicting water quality changes [91]. Meng et al. [37] designed a model based on AlexNet to classify water quality by analyzing the area of algal bloom in Lake Dianchi. The algal bloom extraction algorithm, correlation analysis, redundancy analysis, and random forest were also incorporated into the developed method to correlate water quality to the algal bloom area. The analytical results have established that TP is the dominant factor influencing the algal bloom area. Markogianni [29] developed an EPA classification system to determine the TP concentration in Trichonis Lake in terms of the Tropic State Index.

In most existing studies, TN and TP parameters have been estimated to predict or assess water quality. He et al. [31] used the empirical method to calculate TN and TP concentrations in the Yangtze River, using the Landsat-8 images from 2014 to 2020. As per the results, TN and TP were measured with mean absolute errors 4.3% and 8.3%, respectively, whereas, the RMSE was 0.110 mg/L and 0.01 mg/L for TN and TP, respectively. Singh et al. [51] developed methods to predict dissolved nitrogen and phosphorous in water streams using MODIS images throughout Wisconsin from 2001 to 2004. Some indices including NDVI, enhanced vegetation index (EVI), mean soil fraction (SOI), MODIS disturbance index (MDI), tasseled-cap brightness (TBI), tasseled-cap wetness index, and phenologic rate of growth (PRG), etc., were used as input variables. The parameters SOI, MDI, and TBI correlate positively with nitrogen concentration; NDVI and PRG are the most important parameters for predicting dissolved phosphorous. The method was able to predict dissolved nitrogen with good accuracy, but it is not effective for dissolved phosphorous. Ariman [48] developed two ANN-based models for estimating the TN and TP concentrations in Balik Lake using MODIS data. The dataset consists of MODIS Aqua bands from 1 to 7 and ground truth observations from 2017 to 2019. The accuracy of the model for TP prediction was better with $R = 0.56$. The authors found that MODIS is only moderately suitable for monitoring TN and TP in water. Zhang et al. [36] considered seven major rivers of Zhejiang Province in China to analyze the presence of TN with the help of Sentinel-2 images taken from 2016 to 2021. The correlation coefficient of estimation results was reported as 0.70. Vakili et al. [26] applied ANN and linear regression for TN and TP estimation in Geshlagh Reservoir using Landsat-8 OLI data. The accuracy of the ANN model was better than that of linear regression with $R = 0.93$ and 0.81 for TN and TP, respectively. Guo et al. [28] implemented multimodal deep learning models for estimating TN and TP among other water quality parameters in Lake Simcoe using Landsat-8 images. The models were trained with 70 percent LSLMP data for TN and validated with the remaining 30 percent. The results demonstrated that the implemented deep learning models were able to estimate TN and TP parameters adequately.

Zhang et al. [32] made use of the spectral properties of Landsat-8 OLI and deep learning techniques to observe water quality by measuring TN and TP along with other parameters in Dongping Lake. The method's performance was found to be good, with R^2 of 0.92 and 0.83 for TN and TP, respectively. Song et al. [92] employed random forest to measure six water quality parameters including TN and TP, in Lake Hulun using Landsat OLI images during a non-freezing period from 2016 to 2021. The model achieved high accuracy with $R^2 > 0.7$. The results demonstrated that pollution in Hulun Lake was severe during the studied period. Guo et al. [93] compared some machine learning-based regression methods for estimating TN and TP using Sentinel-2 data. The authors applied band selection techniques to identify the most appropriate spectral bands and obtained good accuracy both for TN and TP with R^2 of 0.94 and 0.88, respectively. Li et al. [94] analyzed the variations of TN and TP in Chaohu Lake using Sentinel-2 data at a large scale. They observed that TN and TP concentrations are higher in urban regions than rural regions.

4.5. Colored Dissolved Organic Matter

Colored dissolved organic matter (CDOM) is an indicator of dissolved organic carbons that assesses the carbon budget in aquatic ecosystems. As the primary sunlight absorber in surface waters and a significant photosensitizer for generating photochemically produced reactive intermediates, it aids in pollutant degradation [95]. The CDOM composition varies in different water bodies, with agricultural effluents showing a higher humification index (HIX) compared to livestock effluents (LEs), indicating differences in the biological activity and sources of CDOM [96]. Viridis et al. [42] adapted and validated C2RCC atmospheric processor for predicting water quality parameters including CDOM in riverine waters in South East Asia, using Sentinel-2 images from 2018 to 2021. However, relatively inferior performance was recorded for predicting CDOM absorption with the approach used. Li et al. [97] applied a three-dimensional fluorescence parallel factor (PARAFAC) method to detect different components of CDOM, such as human-induced and protine-like components in the Yellow River. A total of six CDOM components were found. Four human-induced components, which account for more than 80%, were derived from nonpoint erosion, whereas the protein-like components were derived from point source discharges. The result suggests that the single-parameter model is unsuitable for evaluating the trophic state in sediment-laden rivers such as the Yellow River. Madonia et al. [98] used the Northern Tyrrhenian Sea as the area of study. The PARAFAC was applied for the excitation–emission matrices (EEMs) of CDOM, distinguished into three main components: C1($\lambda_{Ex}/\lambda_{Em} = 342 \text{ nm}/435 \text{ nm}$), C2($\lambda_{Ex}/\lambda_{Em} = 281\text{--}373 \text{ nm}/460 \text{ nm}$) and C3($\lambda_{Ex}/\lambda_{Em} = 286 \text{ nm}/360 \text{ nm}$). It was found that C1 and C2 correspond to humic acid of a terrestrial origin, while C3 corresponds to tryptophan, whose fluorescence peak was detected close to sewage sites, and is strongly related to active *E. coli* cells. Qiang et al. [99] used Sentinel-3 datasets from 2016 to 2020 and developed empirical algorithms to analyze CDOM in Lake Khanka. The performance of the empirical algorithms was found to be reasonably good, with $R^2 = 0.69$.

4.6. Total Suspended Solids

Total suspended solids (TSSs) are suspended particles that do not dissolve in water under normal conditions. TSSs are also referred to as total suspended matter (TSM). Chemical, biological, or heavy metal pollutants could highly correlate to TSS or TSM. TSS is a crucial optically active water quality parameter that impacts light transmission, planktonic algae, and overall ecological health [100]. The TSS concentration can be estimated using remote sensing images, and significant correlations have been found between TSS and hydrological changes [101]. Rahat et al. [56] developed a Long Short-Term Memory model (LSTM) to analyze the spatial variability of TSS in the Ohio River using MODIS data. The performance of the model was good for most of the stations. However, the model could not perform well for specific locations due to hydrological conditions. Wang et al. [102] developed a weighted random forest-based regression model to estimate the TSM in different estuaries in China using Landsat-8 images. The experiments showed that red band and red-based band ratios and indices are sensitive to low TSM concentrations. Meanwhile, NIR and NIR-based band ratios and indices are sensitive to high TSM concentrations. The random forest-based regression model reasonably predicted TSM with R^2 of 0.90 and RMSE of 0.56 mg/L. Pahlevan et al. [103] extended the mixture density networks model to assess the consistency of water quality prediction, including TSS for multi-mission products consisting of Landsat-8, Sentinel-2, and Sentinel-3 missions. The estimation of TSS from three missions was consistent with model's uncertainty. Kupssinskü et al. [104] used machine learning techniques to estimate the TSS in Broa Reservoir in Sao Leopoldo City using Sentinel-2 images. The developed model achieved R^2 values over 0.8. Zhao et al. [105] developed a

Markov model-based method to predict water quality in terms of the TSS in Hedi Reservoir using Landsat-8 OLI data. The authors used a three-band model of TSS for regression analysis using field spectral and remote sensing data. The model achieved an acceptable accuracy of $R^2 = 0.81$. The results showed that the TSS concentration was 2.5 times higher upstream than the downstream TSS concentration. It was also observed that different seasons and rainfall also affect the TSS concentration upstream. The TSS concentration is higher in the rainy season than in the dry season. However, the TSS is negatively correlated to rainfall due to human activities. Aljoborey and Abdulhay [106] estimated TSS concentration in Mosul Dam Lake from July 2018 to September 2019 using Landsat-8 imagery. They analyzed the correlation of TSS with different spectral bands of the Landsat-8 product. It was found that the TSS is highly correlated to band 1 in the summer, band 5 has a strong correlation with the TSS in the spring, and band 6 is significantly correlated to the TSS in autumn. Al-Fahdawi et al. [107] also performed a similar study on the Al-Habbaniyah Lake using Landsat-8 imagery, but they found band 2 to be significantly correlated with TSS during the Autumn season. Jiang et al. [108] developed a semi-analytical model to estimate the TSS in clear to extremely turbid waters using satellite data over Lake Suwa and Lake Kasumigaura from 2003 to 2020. The experimental results exhibited good agreement between the estimated and in situ TSS values. Seleem et al. [109] monitored the TSM along with Chl-a in Timsah Lake using Landsat-8 and Sentinel-2 data during 2014–2020. It was observed that the TSM was higher in August 2015. Wang et al. [110] retrieved the TSS concentration from Landsat-8 images of the Pearl River Estuary in the low-flow season to the high-flow season during 1987–2015. The authors showed that remote sensing effectively analyzes the TSS concentration in estuaries and coasts under different conditions. Niroumand-Jadidi et al. [66] derived TSM using Landsat-9 and Sentinel-2 imagery for four Italian lakes encompassing oligo to eutrophic conditions. The TSM retrievals were validated with in situ matchups. The results indicated relatively high consistency among the water quality products derived from Landsat-9 and Sentinel-2 with R^2 of 0.89 and 0.71, respectively. Niroumand-Jadidi et al. [111] proposed the neural network and standard band ratio-based models for retrieving TSM using Sentinel-2 images in San Francisco Bay. The models were trained using in situ data. The results indicate that the model can retrieve TSM with $R^2 = 0.75$.

4.7. Dissolved Oxygen

Dissolved oxygen (DO) is the amount of oxygen present in the water. The solubility of oxygen in water is not considered good and depends on pressure and temperature [112]. Oxygen is introduced in surface water by aerating actions and aquatic plant photosynthesis. DO is crucial for the survival of aquatic living organisms and plants. DO is a crucial water quality parameter in various environmental settings, including aquaculture and sewage treatment. Zhu et al. [87] estimated the DO in Shenzhen Bay with the help of an ensemble machine learning model using Sentinel-2 images. The model exhibited good accuracy in estimating DO with an error of 0.02%. The Band 7 showed the highest positive correlation with DO. Tian et al. [113] tested multiple machine learning algorithms to estimate DO using Sentinel-2 images. All algorithms performed well for estimating DO with the highest R^2 of 0.90 for the XGBoost algorithm. Krishnaraj and Honnasiddaih [89] used machine learning-based regression models for predicting DO and other water quality parameters in the Ganga Basin using Landsat-8 images at different locations. The model demonstrated good accuracy with R^2 up to 0.98. The DO values were observed between 4.3 and 9.2 mg/L. González-Márquez et al. [24] analyzed the utility of Landsat-8 images for the assessment of DO in the El Guájaro Reservoir using different regression techniques. DO was assessed with R^2 of 0.93, and it was observed that bands 1, 3–4, and 7 are correlated with DO.

Al-Shaibah et al. [30] applied a regression method to assess the DO in Erlang Lake using Landsat-8 imagery. It was observed that DO is high in shallow water and low in deep waters. Kim et al. [57] studied DO variabilities in Saemangeum offshore in the Yellow Sea using MODIS data from 2003 to 2012 with multiple regression models. It was found that long-term changes in DO can be detected with satellite data.

4.8. Hydrogen Power

Hydrogen power (pH) is an important measure of water quality, which measures the acidic or basic nature of water. pH also determines the solubility and availability of nutrients and metals. Research highlights its significance in taste and odor control [114]. It serves as a fundamental indicator of water quality, influencing various chemical, biological, and physical characteristics essential for sustaining life and ecosystems [115]. Huang et al. [116] investigated the correlation of pH with other water quality parameters and showed that pH strongly correlates with DO, temperature, and salinity. Xin and Mou investigated several parameters for water quality prediction using machine learning methods and found that pH is one of the most influencing parameter in assessing water quality. González-Márquez et al. [24] performed regression analysis to determine water quality using Landsat-8 imagery and found that pH values can be reliably generated with the help of regression model and remote sensing images. Abdelmalik [117] predicted pH in Qaroun Lake with regression models using ASTER data. The authors found a high correlation between observed and predicted values with $R^2 > 0.95$. Pereira et al. [118] used genetic programming-based regression models to predict pH values in the Brazilian Pantanal wetland area using Landsat-8 data. The pH values were predicted to be in the range of 4.69–11.64 with a R^2 of 0.81.

Different remote sensing products and water quality parameters assessed using these datasets are summarized in Table 2. It can be observed from the table that Landsat and Sentinel are the most widely used remote sensing products in water quality assessment. The current research scenario in remote sensing-based water quality parameter assessment is shown in Figure 5. Chl-a, TN and TP, turbidity, and TSS are the parameters studied the most.

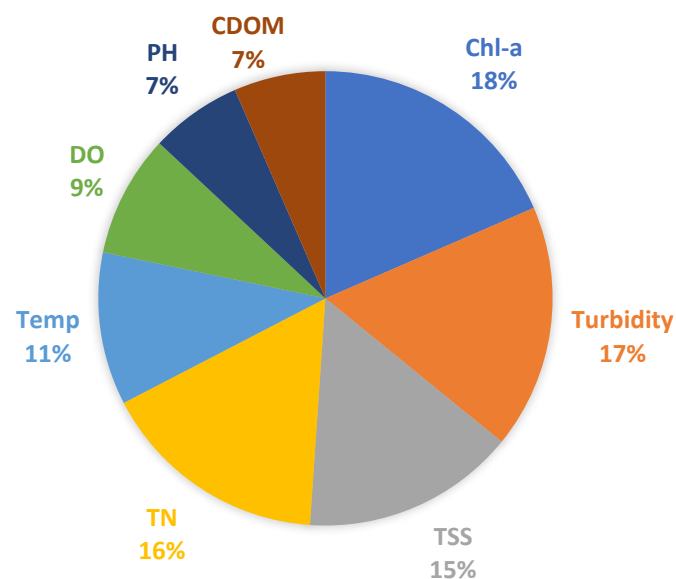


Figure 5. Current scenario of research towards water quality parameters.

Table 2. Remote sensing datasets and water quality parameters.

Satellite	Water Quality Parameter	References
Landsat	Chlorophyll-a	[17,26,28,29,31,32,34,61,63–68,72,77,103,109]
	Turbidity	[17,24,65,75,77,107]
	Temperature	[17,77,88,89,107]
	Total Nitrogen and Total Phosphorous	[26,28,29,31,32,34,92]
	Colored Dissolved Organic Matter	[32,92]
	Total Suspended Solids	[66,102,103,105–107,109,110]
	Dissolved Oxygen	[17,24,30,89,92]
	Hydrogen Power (pH)	[17,24,89,107,118]
Sentinel	Chlorophyll-a	[39,41–44,62,64–66,72,76,87,87,103,104,109,113]
	Turbidity	[42,45,65,72,74,76,78,79,81,87]
	Temperature	[87,113]
	Total Nitrogen and Total Phosphorous	[36,37,93,94]
	Colored Dissolved Organic Matter	[36,42,93,99]
	Total Suspended Solids	[41,66,103,104,109,111]
	Dissolved Oxygen	[87]
	Hydrogen Power (pH)	[104]
MODIS	Chlorophyll-a	[54,55]
	Turbidity	[54]
	Temperature	[54]
	Total Nitrogen and Total Phosphorous	[48,51]
	Total Suspended Solids	[55,56]
	Dissolved Oxygen	[57]

5. Advanced Machine Learning Approaches in Water Quality Assessment

In recent years, machine learning and deep learning have been at the forefront of classification and prediction-based tasks in various domains of applications due to their remarkable performance [119]. Specifically for water quality assessment using remotely sensed data, machine learning models are highly effective for modeling complex and non-linear relationships between water quality indicators and wavelength reflectance in multiple dimensions [16]. Due to several computing layers, deep learning has strong generalization abilities and superior learning capabilities [119]. However, machine and deep learning models are often considered black box models due to the lack of proper explanation in decision-making [120]. Despite the lack of transparency, these models are widely used due to their adaptability and predictive analytical abilities. Remote sensing-based water quality monitoring relies highly on machine learning and deep learning models. The learning methods can be broadly divided into two categories: supervised and unsupervised. The supervised learning methods require a set of samples with known labels called the training set to tune their parameters and weights for making predictions on unknown data later. The training set should contain adequate, diverse, qualified, and consistent data to produce accurate results. On the other hand, unsupervised methods do not require prior information for learning. Instead, they rely on the inherent characteristics of data to learn patterns and insights without external guidance.

5.1. Support Vector Machine

Support vector machine [121], popularly known as SVM, is a well-established and highly successful supervised machine learning method extensively used in various applications for prediction and decision-making. It can be used to perform classification as well as regression. SVM uses a kernel function to transform the input data to a higher-dimensional space. In high-dimensional space, hyperplanes are used to divide data points belonging

to different classes. For a given n -dimensional training datum $(x_i, y_i) \in \mathbb{R}^n \times \{-1, +1\}$, where y_i gives fixed labels, SVM approximates the output y_i by fitting a hyperplane [122] defined as

$$f(x) := w^T \phi(x) + d = 0 \quad (1)$$

where $w = \{w_1, w_2, \dots, w_n\}$ is the normal vector to the hyperplane, w_i is the weight for feature i , d is the displacement term between the origin and hyperplane, and ϕ is a non-linear mapping function. The distance between a point x and a hyperplane [123] is given by

$$\gamma = \frac{|w^T \phi(x) + d|}{\|w\|} \quad (2)$$

A data point with label +1 should lie on or above the following hyperplane

$$w^T \phi(x_i) + d \geq 1, y_i = +1 \quad (3)$$

while the data point labeled as -1 should lie on or below the hyperplane given as

$$w^T \phi(x_i) + d \leq -1, y_i = -1 \quad (4)$$

The data point closest to a hyperplane is called the support vector. The hyperplane representing the largest margin between the two classes is considered the best. The objective of SVM is to maximize the margin γ or maximize $\|w\|^{-1}$ equivalent to minimizing $\|w\|^2$. It also uses a cost parameter C to maximize the margin and minimize the error. Therefore, the SVM objective function becomes

$$\min : \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n |y_i - f(x_i)| \epsilon \quad (5)$$

where ϵ is the error tolerance for *epsilon*-intensive loss function $|y_i - f(x_i)|$. With the introduction of two slack variables ξ_i^{-1} and ξ_i^{+1} [124], the objective function becomes

$$\min : \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i^{-1} + \xi_i^{+1}) \quad (6)$$

subject to:

$$\begin{aligned} \forall i : y_i - f(x_i) &\leq \epsilon + \xi_i^{+1} \\ \forall i : f(x_i) - y_i &\leq \epsilon + \xi_i^{-1} \\ \forall i : \xi_i^{-1}, \xi_i^{+1} &\geq 0 \end{aligned} \quad (7)$$

The slack variables represent the degree of error in predictions.

SVM takes input in the form of a vector whose dimensions depend on the characteristics of the data. It works by finding the best possible decision boundary, i.e., hyperplane to separate data points. A trained SVM model determines the distance of new data points from the hyperplane to make decisions [122]. Tuning hyperparameters, including C and kernel parameters, is important to obtain the best performance. Multi-fold cross-validation is an effective technique to obtain the appropriate values for obtaining the appropriate hyperparameter values. SVM can effectively handle high-dimensional data such as multispectral imagery. It is robust to noise, outliers, and overfitting and exhibits good generalization performance [125]. However, SVMs have high memory requirements, as they place all data into memory. Sometimes, selecting kernel and hyperparameters tuning can be challenging for a particular problem.

SVM and its variants have been used in several studies to assess water quality. Kamyab-Talesh [126] predicted the water quality index in the Sefidrud Basin using SVM from December 2007 to November 2008. The prediction accuracy regarding the determination coefficient was observed as 0.87 with a root mean square error of 0.061. Yahya [127] developed an SVM-based model with a radial basis kernel to predict Lingat River Basin's water quality. The model was used to estimate six parameters with error prediction up to 1%. Deng et al. [128] developed a SVM-based framework to analyze marine water quality in Tolo Harbour in Hong Kong. The results demonstrated the applicability of SVM for predicting the trend and magnitude of algal growth. Nazafzadeh and Niazmardi [129] proposed a modification of SVM named the multiple-kernel support vector regression algorithm to estimate the oxygen demand in Karun River. The modified algorithm predicted with correlation coefficient $R = 0.8$ and $RMSE = 4.76$ mg/L. Sillberg et al. [130] integrated an attribute-realization algorithm and SVM to classify the water quality of Chao Phraya River from 2008 to 2019 by estimating several parameters, including dissolved oxygen, fecal coliform bacteria, total coliform bacteria, nitrate, total nitrogen, TDS, turbidity, TSS, biological oxygen demand, dissolved oxygen, conductivity, salinity, ammonia, etc. Different kernel functions were tested, and the linear kernel was found to be the most suitable function for water quality prediction. Several combinations of quality parameters were also tested, and it was observed that nitrate, fecal coliform bacteria, total coliform bacteria, biological oxygen demand, dissolved oxygen, and salinity are the attributes most contributing to water quality prediction. Arias-Rodriguez [131] developed the support vector regression-based model to estimate the Chl-a, total suspended matter, turbidity, and Secchi disk depth using multiple remote sensing products. Different bands of remote sensing data were found suitable for different parameters. The highest accuracy was obtained for turbidity with $R^2 = 0.71$, and total suspended matter was the most challenging parameter to estimate with the proposed model. Batista [132] evaluated five machine learning model to classify the turbidity of Paraopeba River using Sentinel-2 images. The suitable spectral bands were selected by performing collinearity analysis. The author found SVM to be the best model with an accuracy of 96%. Jamshidzadeh et al. [133] used an enhanced version of SVM to predict a coastal aquifer's electrical conductivity and TDS. TDS was successfully estimated with $MAPE = 2\%$ to 21% and the electrical conductivity was determined with $MAPE = 3\%$ to 21% . Xi and Xue [134] combined SVM with some biologically inspired techniques to enhance the performance of SVM and tested their models by assessing the water quality of Dongting Lake. The integration of different techniques provided good testing accuracy. Dehkordi et al. [135] enhanced the capabilities of support vector regression (SVR) and other machine learning methods by applying fuzzy similarity analysis. They obtained an improvement of up to 12% in the performance of SVR model to assess the water quality parameters in Lake Houston.

5.2. Random Forest

Ensemble learning is a special machine learning paradigm that combines the outputs of multiple machine learning models to improve the robustness of predictions [136]. Different models in an ensemble can be trained with different parameters, data, and algorithms. Random forest [137] is an ensemble-based supervised machine learning model that combines decision trees. Random forest uses bagging or bootstrap aggregation to make predictions robust to noise by correlating the predictions of different decision trees [138]. A partition $P_i \in \mathbb{R}^{M_i \times N_i}$ is randomly selected from the original data $X \in \mathbb{R}^{M \times N}$, where M is the number of samples, N is the number of features, and i represents the i th partition. The subset P_i forms a tree in the forest. The prediction of the ensemble (Y_i) is obtained by

combining the predictions of individual trees. Usually, the majority vote rule is used to combine the results in the case of classification problems,

$$Y_l = \text{Mode}_{i=1 \dots N_t} Y_i \quad (8)$$

where N_t is the number of trees. The average is used to combine the results for regression problems,

$$Y_l = \frac{1}{N_t} \sum_{i=1}^{N_t} Y_i \quad (9)$$

Random forest divides the training set into multiple subsets. Each subset is used to train a different decision tree. This kind of training process ensures that trees are less correlated and diverse. For unseen data, predictions are made by the trees independently, and the final prediction is obtained by majority vote or averaging. Random forest can efficiently work with high-dimensional data and effectively handles the missing values. It is robust to overfitting and efficiently runs on large datasets. It has fewer hyperparameters to tune. There are only two important hyperparameters: number of trees and size of feature subsets [138]. Multi-fold cross-validation can be used to choose the optimal values for these parameters. However, random forest is computationally expensive and less interpretable.

Liu et al. [139] used random forest with $N_t = 100$ to determine the Chl-a concentration in Poyang Lake. The RMSE for training samples was found to be 3.19 μL , and the coefficient of determination was reported as 0.84. Wang et al. [140] developed a random forest regression model to retrieve phytoplankton density for monitoring aquatic ecosystem health using Landsat-8 OLI data. They analyzed seasonal and spatial variations in phytoplankton density in Nansi Lake from 2013 to 2023. Various band combinations were tested to estimate phytoplankton density, yielding good results with R^2 and RMSE values of 0.67 and 1.31×10^6 cells/L respectively. Alnahit et al. [141] predicted the total nitrogen, total phosphorus, and turbidity for watersheds in the Southeast Atlantic region using a random forest model. In addition, the least absolute shrinkage and selection operator (LASSO), and genetic algorithm were used to identify appropriate input parameters out of 28 climatic and catchment variables. The random forest results were compared with the results of a boosted regression tree model. In comparison, it was found that random forest is more robust to overfitting. Wang et al. [102] used random forest models to predict total suspended matter in the Liaohe River, Yellow River, Yangtze River, Hangzhou Bay, Min River, and Pearl River estuaries effectively with $R^2 = 0.90$ and RMSE of 0.56 mg/L. Mishra et al. [142] predicted pre-monsoon and post-monsoon levels of Chl-a and turbidity in Ganga River with the random forest-based model using Sentinel-2 imagery. The experimental results demonstrated R^2 up to 0.97. It was observed that the accuracy of Chl-a prediction was better for the pre-monsoon season, while it was higher for turbidity in the post-monsoon season. Ghasemi et al. [143] used the whale optimization algorithm with random forest to estimate the water turbidity in Louisiana coastal waters. The authors reported an enhancement of 0.08 in R^2 in the performance of random forest. Li et al. [94] developed RF models for measuring TN and TP concentrations in Chauho Lake using Sentinel-2 data. They estimated water quality parameters both in urban and rural regions. The model demonstrated good performance with R^2 of 0.78 in urban regions and R^2 of 0.58 in rural regions.

5.3. Boosting Algorithms

Boosting algorithms are based on a mechanism that attempts to boost the accuracy of weak learners. Several weak learners are combined to transform them into strong learners collectively. A prediction algorithm is considered a weak learner if it performs only slightly better than random guessing. Boosting is an iterative process that builds prediction models

by attempting to reduce residual errors of the preceding stage [144]. The foundation learners in boosting are weak learners with strong biases who contribute critical knowledge. The resultant model is a powerful learner with reduced bias. Boosting algorithms use weighted datasets for training. Initially, all data points are given equal weights. A set of weak learners is formed and trained iteratively. After each iteration, the misclassified data points are given higher weights so that these hard-to-classify data points can be treated more prominently in subsequent iterations. Several boosting algorithms have been developed, including gradient boosting [145], AdaBoost [146], LightGBM [147], and XGBoost [148], which use an ensemble of decision trees to devise a strong learner. Gradient boosting uses decision trees of a fixed size as base learners. During each iteration, a new weak model is trained to minimize the loss function $L(y, F(\mathbf{x}))$ of the previous stage using gradient descent. The function $F(\mathbf{x})$ maps instances $\mathbf{x}$ to corresponding outputs. The new model is added to the ensemble, and the process continues until a stopping criterion is met. Given a training set $D(\mathbf{x}_i, y_i)_{i=1}^N$, ($\mathbf{x}_i \in \mathbb{R}^M, y_i \in \mathbb{R}$) gradient boosting constructs an additive approximation of a mapping function to map instance $\mathbf{x}_i$ to its output label y_i as follows:

$$F_k(\mathbf{x}_i) = F_{k-1}(\mathbf{x}_i) + \eta_k h_k(\mathbf{x}_i) \quad (10)$$

η_k are the weights of functions $h_k(\mathbf{x}_i)$, which are ensemble models, and k is the k^{th} iteration. Initially, a constant approximation of the mapping function is obtained as

$$F_0(\mathbf{x}_i) = \arg \max_{\alpha} \sum_{i=1}^N L(y_i, \alpha) \quad (11)$$

Gradient boosting is robust to outliers but computationally expensive and prone to overfitting. XGBoost is a highly scalable ensemble based on decision tree and gradient boosting. Similar to gradient boosting, XGBoost also minimizes a loss function to build an additive expansion of the objective function. It uses a variation of the loss function L_{xgb} to control the complexity of decision trees

$$L_{xgb} = \sum_{i=1}^N L(y_i, F(\mathbf{x}_i)) + \sum_{j=1}^M \Omega(h_j) \quad (12)$$

$$\Omega(h_j) = \gamma T + \frac{1}{2} \lambda ||w||^2 \quad (13)$$

where T is the number of leaves in the tree, w represents the outputs of the leaves, Ω is the penalty term for the complexity of the model, γ controls the minimum reduction gain, and λ is responsible for the L2 regularization on leaf weights. There are some other variations of gradient boosting. LightGBM is a gradient boosting-based model that grows leafwise. In a particular instance, only one leaf from the whole tree grows, leading to reduced costs and memory requirements. However, it may be prone to overfitting sometimes. CatBoost, abbreviated for category boosting, performs well on categorical datasets. CatBoost uses symmetric trees so that all the nodes at the same level test the same condition.

Yang et al. [119] developed several machine learning models based on random forest, gradient boosting, and XGBoost to estimate Chl-a, turbidity, dissolved oxygen, etc., in karst wetlands using multi-sensor images. The results demonstrated that gradient boosting and XGBoost models achieved the highest accuracy with $R^2 = 0.75$ for turbidity and Chl-a parameters. Xin and Mou [149] designed several machine learning models based on random forest, CatBoost, XGBoost, and LGBM to predict water quality parameters including pH, TDS, turbidity, and conductivity. They found CatBoost, XGBoost, and LGBM to be the best-performing models, which were further optimized with cross-validation and super

parameter optimization. Duan et al. [150] used correlation analysis to obtain sensitive bands and band combinations to monitor suspended particulate matter in Ebinur Lake. The preferred band combinations were sent as input to machine learning models including random forest, gradient boosting, CatBoost, and XGBoost. The validation results indicated that the CatBoost model provided the best results. Garabaghi et al. [151] evaluated several ensemble classifiers including random forest, AdaBoost, and XGBoost to classify the water quality in five user classes. A feature selection algorithm was also applied on raw data to reduce the computational burden of selecting the most useful parameters. The results showed that XGBoost outperformed other classifiers with 96.96% accuracy. Leggesse et al. [68] explored six machine learning methods including SVM, artificial neural network, random forest, gradient boosting, AdaBoost, and XGBoost for monitoring turbidity, TDS, and Chl-a in large freshwater bodies from 2016 to 2022. XGBoost produced the best results for Chl-a with $R^2 = 0.78$, while the random forest outperformed other models to predict turbidity and TDS with $R^2 = 0.80$ and $R^2 = 0.79$, respectively. Shams et al. [152] tested random forest, gradient boosting, AdaBoost, and XGBoost to classify water quality. They also applied some preprocessing techniques including data imputation and normalization to make data more suitable for classifiers. The reported results showed that the increase in the gradient produced the best accuracy of 99.50% among all the methods tested. A summary of machine learning-based water quality assessment works is given in Table 3. SVM, RF, ANN, and statistical regression (SR) are effective machine learning models to analyze the water quality parameters as shown in the table.

Table 3. Summary of machine learning-based water quality assessment methods.

Reference	Method	Water Body/Location	Water Quality Parameters	Performance
[68]	RF, XGB, GB, AB, SVM, ANN	Lake Tana, Ethiopia	Chl-a, Turbidity, TDS	$R^2 = 0.78\text{--}0.80$, MARE = 0.072–0.082
[126]	SVM	Sefidrud basin, Iran	pH, DO, TDS, Temp, NO ₃ , PO ₄ , BOD, Turbidity, and FCB	$R^2 = 0.87$, RMSE = 0.061
[127]	SVM	Langat basin, Malayasia	pH, TSS, DO, AN, COD, and BOD	CC 0.979–0.998 and MSE = 0.004–0.681
[128]	SVM, ANN	Tolo Harbour, Hong Kong	Chl-a, pH, BOD, DO, TIN, PO ₄ , Temp	CC: 0.972–0.984, and RMSE = 0.66–0.765
[130]	SVM	Chao Phraya River, Thailand	NH ₃ , BOD, DO, FCB, TCB, etc.	Accuracy: 0.94, Precision: 0.84, Recall: 0.84, F1: 0.84
[131]	SVR	Trans-Mexican Volcanic Belt, Mexico	Chl-a, TSM, SDD	
[62]	SVR	Finger Lakes, USA	Chl-a	$R^2 = 0.76$, RMSE = 0.633 $\mu\text{g/L}$, MAE 0.728 $\mu\text{g/L}$
[132]	SVM	Paraopeba River, Brazil	Turbidity	Accuracy: 0.96
[135]	SVM	Lake Houston, USA	Turbidity and Specific Conductance	$R^2 = 0.55\text{--}0.90$, MAPE = 12.01–25.03
[36]	SVM, SR	Zhejiang, China	CDOM, TN, and TP	$R = 0.68\text{--}0.82$
[26]	ANN	Gheshlagh reservoir, Iran	TN and TP	$R = 0.81\text{--}0.93$
[34]	ANN	Lake Urmia, Thailand	Chl-a, TP, and Secchi depth	NSE = 0.74–0.96
[66]	ANN	San Francisco Bay	Chl-a, TSM	$R^2 = 0.71\text{--}0.89$
[48]	ANN	Balik Lake, Turkey	TN and TP	$R = 0.36\text{--}0.59$
[17]	LASSO	Tigris River, Iraq	Chl-a, Temp, EC, TDS, pH, Turbidity, Algae, and DO	$R^2 = 0.42\text{--}0.80$
[31]	SR	Yangtze River, China	Chl-a, TN, and TP	MAPE = 4.30–25.88%
[29]	SR	Trichonis Lake, Greece	Chl-a, NH ₄ -N, and TP	$R = 0.4\text{--}0.7$
[65]	SR	Oklahoma, USA	Chl-a, Turbidity, Secchi depth	$R^2 = 0.58\text{--}0.85$
[44]	SR	San Roque, Argentina	Chl-a	$R^2 = 0.77$
[31]	SR	Yangtze River, China	Chl-a, TN, and TP	RMSE = 0.475 $\mu\text{g/L}$ –0.110 mg/L, MAPE = 4.3–25.88%
[51]	SR	Wisconsin, USA	TN, TP, NO ₃ -N	$R^2 = 0.51\text{--}0.80$
[99]	DT	Lake Khanka, China/Russia	CDOM	$R^2 = 0.84$
[42]	C2X-Net	Chao Phraya River, Thailand	Chl-a, CDOM, and Turbidity	$R = 0.47\text{--}0.84$

Table 3. Cont.

Reference	Method	Water Body/Location	Water Quality Parameters	Performance
[39]	RF	Miyun Reservoir	Chl-a	$R^2 = 0.74$, RMSE = 0.42 mg/m ³ , MAE = 0.33 mg/m ³ , and MAPE = 55.56%
[74]	RF	-	Turbidity	MSE = 67.133, NMGE = 0.5763, PSNR = 29.861
[139]	RF	Yangtze River, Guojiaba, China	Chl-a	$R^2 = 0.8104$, MAPE = 6.46%
[140]	RF	Nansi Lake, China	Phytoplankton density	$R^2 = 0.67$, RMSE = 1.31×10^6 cells/L, MAE = 1.18×10^6 cells/L
[141]	RF	Yellow River, China River Ganga, India	TN, TP, Turbidity	$R^2 = 0.90$, RMSE = 0.56 mg/L
[102]	RF		TSM	$R^2 = 0.91$ –0.97, MAE = 0.59–1.13, MAPE = 2.07–7.76%
[142]	RF		Chl-a, Turbidity	$R^2 = 0.7128$ –0.8376, RMSE = 0.0029–0.8923, MAE = 0.0017–0.6757
[92]	RF	Hulun Lake, China	Chl-a, COD, DO, NH ₃ -N, TN, and TP	$R^2 = 0.86$ –0.94
[93]	RF, SVR, ANN	Tianjin, China	COD, TN, and TP	$R^2 = 0.649$ –0.844
[119]	RF, XGB, GB	Karst wetlands, China	Turbidity, DO	$R^2 = 0.45$ –0.59, RMSE = 68.67–73.71 mg/L
[150]	RF, XGB, GB, CB, DT	Ebinur Lake, China	TSM	Accuracy = 93.03–95.60%
[151]	RF, XGB, AB, LB	Büyük Menderes Basin, Turkey	WQI	$R^2 = 0.90$, RMSE = 0.56 mg/L
[102]	RF	Liaohe River, China	TSS	$R^2 = 0.7015$ –0.7927, RMSE = 2.1747–5.0617 µg/L, MAE = 1.6791–3.9776 µg/L
[76]	RF, XGB, LGB, CB	Nansi Lake, China	Chl-a and Turbidity	$R^2 = 0.776$, RMSE = 3.802, MAE = 3.246
[75]	DT	Saki Lake, India	Turbidity	$R^2 = 0.88$, RMSE = 9.90 NTU, MAE = 6.71 NTU
[78]	RF, XGB, GB, KNN	China	Turbidity	$R^2 = 0.92$
[81]	RF	China	Turbidity	Errors = 0.02–1.7%
[87]	RF, SVR, ANN	Shenzhen Bay, Hong Kong	Chl-a, Temp, Turbidity	$R^2 = 0.72$ –0.98
[89]	XGB, ANN	Ganaga Basin, India	EC, DO, pH, TDS, Temp	$R^2 = 0.31$ –0.90, RMSE = 0.0052–0.0364
[104]	SVR, RF, ANN, KNN, SR	Brazil	Chl-a, pH, and TSS	$R^2 = 0.46$ –0.78, RMSE = 0.0034–0.37 mg/L, MAPE = 8.34–38.60%
[94]	RF, ANN	Chaohu Lake, China	TN and TP	

5.4. Deep Learning Models

Deep learning techniques are machine learning techniques based on artificial neural networks that use multilayered architectures [153]. The layered architecture includes an input layer, multiple hidden layers, and one output layer. Each layer consists of several artificial neurons and activation functions that perform small computing operations [154]. The input layer accepts the raw input data, which are transformed at the hidden layers. The hidden layers extract different types of features that are useful for making predictions [155]. The final predictions are made available by the output layer. The artificial neurons are associated with weights and biases that control the influence of the input. The weights and biases are initially chosen randomly or heuristically and adjusted during the training. The activation function introduces non-linearity to the result of a neuron. The weights and biases of a deep learning model are updated iteratively to minimize errors and loss. This process aims to improve the accuracy of predictions. Deep learning techniques offer several advantages over conventional machine learning techniques. It can progressively learn features ranging from simple to abstract, eliminating the need for manual feature engineering and enabling the modeling of complex relationships in the data [156]. Deep learning achieves high accuracy with large and versatile datasets. However, deep learning techniques are computationally intensive and require large amounts of data for effective training. The training process is very complex and often requires high-end computing resources. In the absence of sufficient data, deep learning models are prone to overfitting. Several hyperparameters, such as learning rate, batch size, dropout rate, and kernel size, require significant expertise to be set to appropriate values. Usually, it is performed through trial and error due to a lack of well-defined methods and procedures.

Deep learning is often considered a black box, as it is difficult to interpret [157]. Convolutional neural network (CNN) is one of most widely used deep learning techniques [37]. The layers in a CNN model are designated as convolutional, pooling, and fully connected (Dense) as shown in Figure 6 which perform specified tasks. The convolutional layer convolves the input with a kernel filter to extract the useful features as follows:

$$O^m = g(V^m \star O^{m-1} + c^m) \quad (14)$$

where O^m is the output, V^m is the weight matrix, c^m is the bias vector at the m^{th} layer, $\star$ represents the convolution operation, and $g(\cdot)$ is the activation function. Rectified Linear Unit (ReLU) is a popular activation function in CNN models. The pooling layer is responsible for reducing the size of feature maps generated by the convolutional layer. It reduces the number of parameters in the model, lowering both the computational cost and the risk of overfitting. The pooling operation is performed by moving a kernel or filter over the feature map with a stride. The most common technique is max-pooling, which selects the maximum value from each kernel window, capturing the most prominent features as follows:

$$\text{MaxPool}(x(i, j)) = \max_{(p, q) \in K} x(i + p, j + q) \quad (15)$$

where $x(i, j)$ represents the input feature map value at position (i, j) , K is the pooling window, and $\max$ takes the maximum value in the window. The fully connected layer links every neuron in the previous layer to every neuron in the current layer, combining the extracted features to predict different classes more accurately.

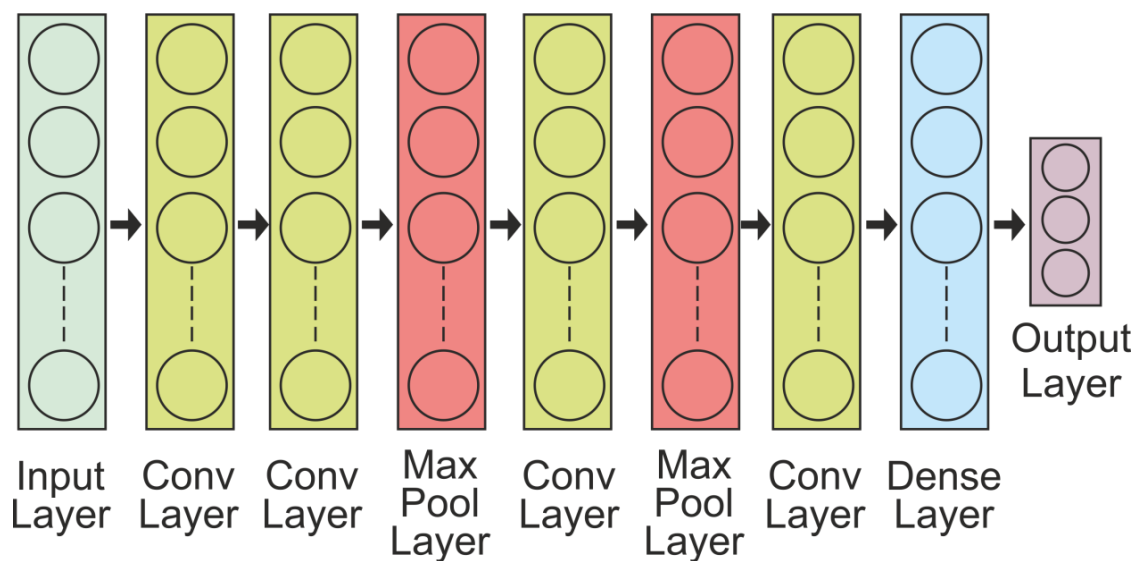


Figure 6. Deep convolutional neural network model architecture.

Recurrent neural network (RNN) [157] is a neural network that allows the output from one step to feed back as input. It is suitable to those applications where the context from previous steps is essentially required. Long-Short Term Memory (LSTM) [56,158] is an extension of RNN that uses memory cells and gates at hidden layers, which determines whether information is important or not [159]. The memory cell can store useful information, and the gates control the inflow/outflow of information [56]. The basic structure of an LSTM cell [56,158] is shown in Figure 7 that consists of a forget gate layer, input gate layer,

tanh layer, and output gate layer. The forget gate layer decides what information to delete for a cell state C_{t-1} . Mathematically, it is expressed as follows

$$\Phi(W_{f,h}h_{t-1} + W_{f,x}x_t + b_f) \quad (16)$$

where Φ is a sigmoid function, $W_{f,h}$ and $W_{f,x}$ are weights, and b_f are biases. The input layer decides how much new information should be added to the cell state and uses the sigmoid function to rank the level of updation as follows:

$$\Phi(W_{i,h}h_{t-1} + W_{i,x}x_t + b_i) \quad (17)$$

where $W_{i,h}$, $W_{i,x}$ are weights and b_i are biases. The tanh layer is responsible for deciding how much new candidate values C'_t should pass to the input layer as follows:

$$C'_t = \tanh(W_{c,h}h_{t-1} + W_{c,x}x_t + b_c) \quad (18)$$

where $W_{c,h}$ and $W_{c,x}$ are weights, and b_c are biases. The old state is updated using C'_t as follows:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C'_t \quad (19)$$

At last, the out layer decides the final cell output

$$\Phi(W_{o,h}h_{t-1} + W_{o,x}x_t + b_o) \quad (20)$$

where $W_{o,h}$ and $W_{o,x}$ are weights and b_o are biases. An LSTM model is developed by creating multiple layers of LSTM cells as shown in Figure 8. It is also possible to create hybrid models by clubbing the LSTM layers with other kind of layers such as convolutional layers. One such example is shown in Figure 9, with a CNN-LSTM hybrid model consisting of some convolutional and pooling layers followed by LSTM layers.

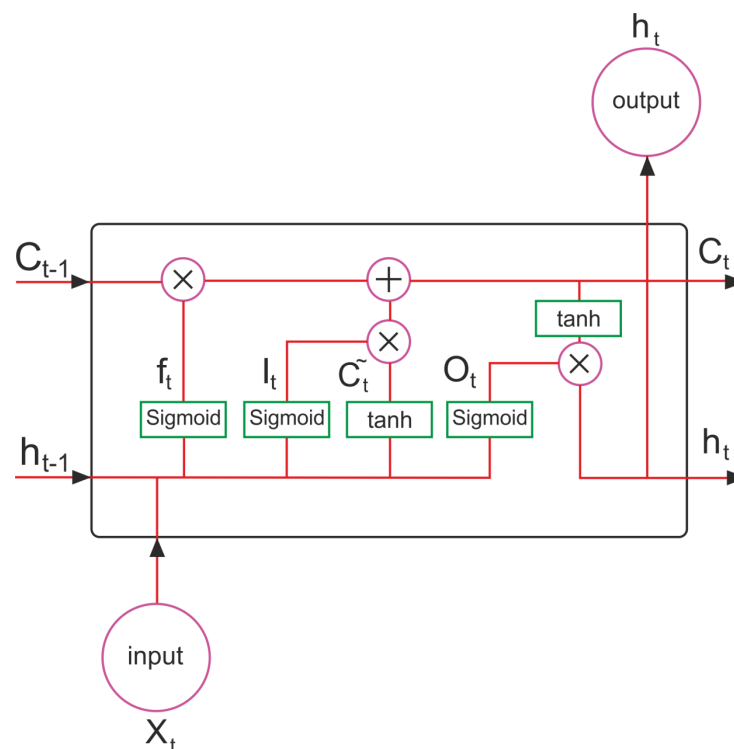


Figure 7. Basic structure of an LSTM cell.

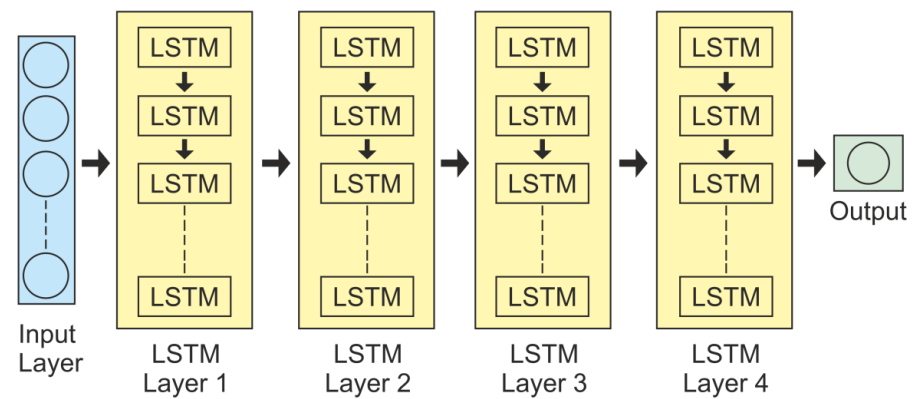


Figure 8. Architecture of a deep LSTM model.

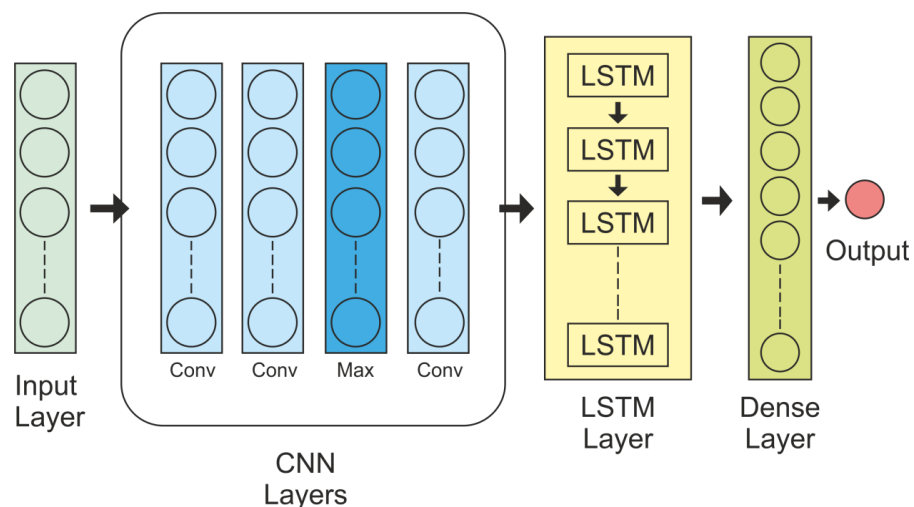


Figure 9. CNN-LSTM hybrid model architecture.

Recently, deep learning has attracted more attention in water quality monitoring. CNN and Long Short-Term Memory (LSTM) are the most successful methods used to assess water quality parameters. Guo et al. [28] employed a deep CNN model to investigate Chl-a, TN, and TP concentrations in Lake Simcoe using Landsat data. The CNN model was created with fusion of 23 layers of U-Net and 13 layers of SegNet. Meng et al. [37] designed a CNN model with five convolutional, one pooling, and one fully connected layer to investigate the water quality levels in Lake Dianchi using Sentinel-2 data. The results demonstrated that the CNN model was able to accurately estimate the water quality parameters with MAE = 32.57–42.58%. Liu et al. [158] developed an LSTM-based model to measure the drinking water quality in Yangtze River from 2016 to 2018. Their model consists of multiple hidden layers and each hidden layer contains multiple memory cells to build a deep learning model. There are 100 neurons in each LSTM layer. Adam optimizer was adopted as the LSTM optimization algorithm, and the model was trained for 100 epochs. They measured pH, DO, CDOM, turbidity, and conductivity parameters and demonstrated the effectiveness of the LSTM for water quality monitoring. Mean square error (MSE) was used to evaluate the performance of the method, and the LSTM-based method was found to be better than some models with MSE = 0.0017. Hu et al. [159] proposed an LSTM-based method to assess pH and water temperature parameters. The LSTM model consists of 15 hidden layers, and is trained with a time step of 20 and learning rate of 0.005. The experimental results show that the method observed an accuracy of 98.56% and 98.97% for pH and temperature, respectively, at the time cost of prediction of 0.273 s and 0.257 s, respectively. The author compared the results with a recurrent neural network-based

method and found that LSTM gives better accuracy at a lower time cost. Sha et al. [155] investigated CNN and LSTM models to measure the DO and TN parameters in Xin'anjiang River water. The authors developed a hybrid CNN-LSTM model for real-time water quality monitoring and found that the hybrid model performed better than standalone CNN and LSTM models. Yiqi et al. [154] compared deep learning techniques including LSTM, RNN, and gated recurrent unit (GRU) with linear regression methods and multilayer perceptron (MLP) to predict the chemical index of water. The results demonstrated that deep learning methods achieved up to 43.38% higher accuracy (R^2) than linear regression methods and MLP. Among the deep learning methods, GRU exhibited superior performance over LSTM and RNN. Barzegar et al. [160] developed deep learning models based on CNN and LSTM algorithms to estimate water quality parameters Chl-a and DO. They combined CNN and LSTM to develop a hybrid CNN-LSTM model. The CNN-LSTM hybrid model consisted of an input layer, three convolutional layers, one flattening layer, and one LSTM layer followed by an output layer. The dropout rate was set to 0.001 and learning rate was taken as 0.01 both in the convolutional and LSTM layers. The performance was evaluated in terms of various metrics including R, RMSE, MAE, RMAE, etc. The authors found that the hybrid CNN-LSTM outperformed standalone models in predicting both Chl-a and DO with $R = 0.869$ and $R = 0.97$, respectively. Khullar and Singh [161] proposed a Bi-LSTM model to predict water quality parameters in River Yamuna. The model includes an input layer, two convolutional layers, one Bi-LSTM layer, two fully connected layers, and an output layer. The output layer uses Softmax for making predictions. The model was trained for 2500 epochs with a batch size of 120 and learning rate of 0.001. They performed missing value imputation before input to improve the learning process. The experimental results showed that the predicted values were in close agreement with the actual values of the parameters. It was also observed that Bi-LSTM model is better than some conventional machine learning models with $MSE = 0.015$, $RMSE = 0.117$, $MAE = 0.115$, $MAPE = 20.32$. Zamani et al. [157] incorporated LSTM, RNN, GRU, and temporal neural network with the help of non-dominated sorting genetic algorithm (NSGA-II) into an ensemble model to predict Chl-a concentration. There are 50 neurons in RNN, LSTM, and GRU layers that use a dropout of 0.001 and ReLU activation function. The model was trained for 100 epochs with a learning rate of 0.001 and Adam optimizer. The author demonstrated that the proposed ensemble model effectively predicted the Chl-a concentration with a correlation coefficient of 0.98 and standard deviation of 0.93 in the Small Prespa Lake. Rahat et al. [56] developed an LSTM model to investigate the impact of climate change on water quality in Ohio River using MODIS data. The author demonstrated that the model was able to comprehensively predict water quality for long-term evaluation. Ramesh et al. [80] developed a deep ensemble model based on LSTM and ReLU activation to estimate turbidity levels in Dooskal Lake. The architecture consists of two LSTM layers, one ReLU layer, three Dropout layers, and two dense layers. The LSTM layer has 124 memory units and each dense layer has 32 LSTM modules. The dropout rate is set to 0.2. The model demonstrated good performance with accuracy of 97.2, precision of 94.88, recall of 86.3, and F1-score of 90.3. The deep learning-based water quality assessment methods covered in the section are summarized in Table 4.

Although deep learning achieves high accuracy with large and versatile datasets, deep learning techniques are computationally highly expensive and require a large amount of data for training. The training process is very complex and often needs high-end computing resources. In the absence of a good amount of data, deep learning models are prone to be overfit. Deep learning is often considered a black box, as it is difficult to interpret [156]. There are several hyperparameters such as learning rate, batch size, dropout,

size of kernels, etc., that need significant expertise to set appropriate values. Usually, it is performed through trial and error due to the lack of well-defined methods and procedures.

Table 4. Summary of machine learning-based water quality assessment methods.

Reference	Method	Architectural Details	Water Quality Parameters	Performance
[28]	CNN	Fusion of U-Net (23 layers) and SegNet (13 layers)	Chl-a, TN, and TP	MAE = 32.57–42.58%
[158]	LSTM	Neurons/layer: 100, Optimizer: Adam, Loss function: MSE, No. of epochs: 100	pH, DO, CDOM, turbidity, NH ₃ -N, and electrical conductivity	MSE = 0.0017–0.0020
[32]	Conv-LSTM	Convolutional layers: 3, Fully connected LSTM layers: 3, Dropout layer: 1, Activation function: ReLU	Chl-a, TN, NH ₃ -N, CODMn, TP, and BOD	R ² = 0.77–0.92
[159]	LSTM	Hidden layers: 15, Time step: 20, Learning rate: 0.0005, Training time: 10,000	pH and Temperature	RMSE = 0.0025–0.0479, MAPE = 0.0012–0.0692, MAE = 0.0027–0.0149
[37]	CNN	Convolutional layers: 5, Pooling layers: 1, Fully connected layers: 1, learning rate = 10 ^{−4} , momentum = 0.9, Epochs = 500	Water quality levels	-
[154]	GRU, LSTM, RNN	Hidden Layers: 3, Neurons/layer: 50	BOD, COD, TN, TP, NH ₄ -N	R ² = 0.90–0.94
[160]	CNN, LSTM	Convolutional layers: 3, Flattening layer: 1, LSTM layer: 1, Dropout: 0.001, Learning rate: 0.01	Chl-a and DO	R = 0.869–0.97
[161]	Bi-LSTM	Convolutional layers: 2, Flattening layers: 2, Bi-LSTM layer: 1, Optimizer: Softmax, Batch size: 120, Learning rate: 0.001, Epochs: 2500	BOD and COD	MSE = 0.015–0.107, RMSE = 0.108–0.117, MAE = 0.115–0.124, MAPE = 18.22–20.32
[157]	LSTM, RNN, GRU	Activation function: ReLU, Optimizer: Adam, Dropout = 0.001, Learning rate: 0.001, Epochs: 100	Chl-a	Correlation Coefficient: 0.98, Standard Deviation: 0.93
[80]	LSTM	LSTM layers: 2, ReLU layer: 1, Dropout layer: 3, Dense layers: 2	Turbidity	Accuracy = 97.2, Precision = 94.88, Recall = 86.3, and F1-score = 90.3
[56]	LSTM	-	TSS	R ² = 0.82, RMSE = 16.69 mg/L, MAE = 13.85 mg/L

6. Discussion

Water quality is crucial for human and plant health, aquatic life, the environment, and sustainable development. Therefore, monitoring the quality of the water in rivers, lakes, ponds, bays, seas, and oceans is important to maintain the health of the ecosystem and support population needs. Water quality monitoring has been important for a long time, and the supporting infrastructure has been developed over the years. However, conventional monitoring methods are expensive, time consuming, and unable to provide large spatial coverage [12,13,28]. Advancements in remote sensing technologies have provided an alternative approach for efficient and cost-effective water quality monitoring with potentially large spatial coverage [162]. Remote sensing utilizes sensors on board satellites or other aerial platforms to collect images and data, offering valuable insights into water bodies across extensive spatial areas [87]. Remote sensing data help to measure water quality parameters such as color, Chl-a, dissolved, total nitrogen, total potassium, temperature, and turbidity, etc.; machine learning, on the other hand involves powerful techniques to process large remote sensing datasets to extract meaningful features or patterns [163,164]. Machine learning techniques help to derive insights and measure specific quality parameters. Integrating remote sensing and machine learning offers a comprehensive, efficient, and cost-effective alternative to monitoring and assessing water quality [165]. It helps water quality assessment and monitoring in several ways.

1. Larger spatial and temporal coverage: The traditional methods need to establish hydrology stations across water bodies to collect water samples. However, establishing such stations over a large area is challenging due to cost and geographical

constraints [27]. Therefore, ground-based monitoring may sometimes be expensive and limited. Satellites orbit around Earth and cover vast geographical area in a single pass. Satellites can capture images across large swaths of water bodies, allowing to monitor even the inaccessible locations consistently. With increasing resolutions, remote sensing-based monitoring can obtain useful information at the scale of water bodies such as lakes and rivers. Satellites orbit the Earth frequently at regular intervals, leading to short-term and long-term observations in water quality [166,167]. Frequent observations are crucial for analyzing seasonal variations and temporal changes.

2. **Parameter estimation:** Several water-quality parameters that are optically active can be efficiently assessed with the help of remotely sensed imagery. These parameters include Chl-a, turbidity, TSSs, CDOM, etc. [168]. Remote sensing platforms equipped with specialized sensors measure the reflectance and absorption of particular wavelengths by water bodies to provide valuable data that can be processed to determine important water quality parameters [27]. The assessment of optically inactive parameters except temperature with the help of remote sensing techniques is a challenging task. Meanwhile, water temperature can be effectively measured by detecting infrared radiation emitted by water bodies [169]. Other parameters such as pH, nitrogen, phosphorous, and dissolved oxygen are determined indirectly using sophisticated algorithms and modeling [26].
3. **Automated monitoring:** Remote sensing data can be captured without manual intervention. To obtain better accuracy, the captured data undergo preprocessing for atmospheric corrections, geometric corrections, and sensor calibration [170]. Machine learning algorithms process the remote sensing data to assess optically active and optically inactive water quality parameters [24,26,28]. The entire process from data capture to parameter estimation is performed automatically by the sensors and computer program. These automated systems can detect anomalies in water quality anomalies and can generate alerts when needed [171]. It helps to provide timely insights into water quality conditions.
4. **Cost effectiveness:** Machine learning methods significantly rely on the availability of adequate amounts of data. The acquisition of remote sensing data has become highly cost effective in recent years due to space agencies' data accessibility initiatives [172]. Satellite missions such as Landsat and Sentinel provide low-cost or free data access, alleviating financial burdens. Some investments are required for computing machines to develop and train the machine learning algorithms. However, the overall cost involved is significantly lower than establishing the hydrology stations [173].

Water quality monitoring using remote sensing and machine learning has made significant progress over the years. However, there are several challenges that must be addressed to realize its potential effectively and with better accuracy. Some satellites do not have cloud penetrating capabilities. Data from such satellites are not suitable for all weather monitoring. Apart from that, satellites may not provide spectral bands required for specific water quality parameters. Coarser resolution is another factor of concern that limits the capabilities of remote sensing-based methods for small water bodies. Advanced sensors are required to mitigate these limitations. The fusion of data from different sensors can also improve the reliability of the method. Machine learning, on the other hand, depends on a good amount of quality data that may not always be available. A balance in data is desirable to represent all possible conditions to avoid biased predictions. In the presence of large datasets, machine learning algorithms tend to be computationally expensive. This leads to the need for high-end computing devices. Different geographical coverage should be sufficiently covered to develop a generalized model. The machine learning techniques

are often considered a black box which makes it difficult to interpret the predictions of the model. Some key research gaps in water quality monitoring are identified as follows.

1. **Robust algorithms:** The existing algorithms mostly focus on the assessing of multiple water quality parameters. However, different parameters are sensitive to different data features [174]. Therefore, refined algorithms to assess specific water-quality parameters can be developed to enhance the accuracy and precision. In addition, the algorithms are usually developed and evaluated on data for specific locations [124]. Such algorithms may not generalize well under different conditions for other locations, leading to the need for more robust algorithms.
2. **Data fusion:** Data from different remote sensing platforms vary in spectral, spatial, and temporal resolution, data formats, and collection methods. Satellites have different frequencies and visiting times. The complementary information from different sources can be integrated for better efficiency and accuracy [175]. Merging multi-source data can improve spatial and temporal resolution. The integrating of spectral bands with different wavelengths enhances the system's discriminating ability. However, combining data from different optical sensors with broad coverage remains challenging [175]. Advanced data fusion techniques can be explored to leverage the comprehensive water quality assessment strengths [154].
3. **Extended use of machine learning:** Machine learning has exhibited great potential for processing remotely sensed data for water quality assessment. However, machine learning algorithms are highly affected by the variability of data across different water bodies. Generalizing machine learning models across water bodies is difficult due to varying conditions [176]. It is desirable that models can scale well to larger geographical areas and datasets without significant performance loss. The need for high-end computing resources is another constraint with machine learning models, especially for under-resourced conditions [177]. All these issues need further consideration to develop efficient and affordable advanced methods.
4. **Atmospheric correction:** Sensors mounted on satellites and other aerial platforms record the electromagnetic radiation reflected or emitted by the objects on Earth. The radiation travels through the atmosphere of Earth. Various particles and gases in the atmosphere react with radiation, distorting the radiation or light before reaching the sensor. Scattering and absorption are the major atmospheric effects affecting electromagnetic radiation [178]. Atmospheric corrections [179,180] are applied as preprocessing steps to the remote sensing data to reduce the distortion caused by the atmosphere on the reflected values recorded by the sensor. Atmospheric corrections help to ensure that the data represent the actual characteristics of Earth's features. Atmospheric conditions vary geographically and temporally, making atmospheric correction highly challenging. Any single correction method is not universally applicable. Developing an algorithm that can adapt to varying conditions is highly desirable.

7. Conclusions

Water quality monitoring is important for studying aquatic ecosystem health and ensuring clean and safe drinking water access [9,181]. The growing population, industrialization, and climate change have adversely affected the aquatic ecology and water quality. Therefore, water quality monitoring is vital in providing crucial information to governments and policy makers on managing resources. Traditionally, water quality monitoring is performed with the help of in situ sampling at hydrology stations and laboratory analysis [12,13]. The conventional methods are expensive and do not provide the desired temporal and spatial coverage. Modern technologies, including remote sensing and machine learning, offer an alternative to the conventional approach for water quality

monitoring. Satellites and other remote sensing platforms provide important data that can be processed and analyzed with the help of machine learning methods to assess water quality parameters without in situ sampling and laboratory testing. This approach is cost effective and can be applied at large scale despite involving some challenges.

7.1. Synthesis of Key Findings

This work reviewed the integration of remote sensing and machine learning techniques for water quality monitoring. Remote sensing provides valuable data processed by machine learning techniques for assessing water quality. Data acquisition is highly crucial, as machine learning methods perform well only if adequate data are available. However, some space agencies have provided free access to satellite data in remote sensing. High spatio-temporal resolution imagery from various missions including Landsat, MODIS, and Sentinel is available for public use. The easy availability of data makes remote sensing-based methods viable. Landsat was the first mission launched for terrestrial environment monitoring. Eight satellites have been successfully launched so far under the Landsat program. Landsat-8 multispectral imagery is more popular, as it provides eleven-band imagery with better radiometric and spatial resolution. Landsat-8 carries OLI and TIRS sensors onboard. Landsat 9 is the latest satellite in the series, and it has improved versions of OLI and TIRS. However, Landsat sensors cannot penetrate through clouds, limiting their effectiveness during monsoon season. The Sentinel mission of the European Space Agency is another popular program primarily used for land monitoring and provides wide-swath high-resolution multispectral imagery to monitor soil, vegetation, and coastal regions. It is a bi-satellite system encompassing two satellites, Sentinel-2A and Sentinel-2B, which were launched in 2015 and 2017. Both satellites complement each other for a short revisiting period of 5 days. The Sentinel-2 data are used in various applications, including water quality monitoring. MODIS is another satellite-based product that is coarser in resolution up to 1000 m. It visits a location twice a day and there is no need for atmospheric corrections or mosaicing. MODIS is more useful for the monitoring of coastal areas.

A number of optically active and inactive water quality parameters can be assessed using remote sensing imagery. Chl-a is one of the most widely evaluated parameters. Landsat and Sentinel imagery monitor the Chl-a concentration in surface water, while MODIS is better for assessing the Chl-a concentration in coastal waters. Turbidity is a visual property of water indicating the presence of suspended particles in water. Turbid water may be of poor quality and contain parasites such as *Cryptosporidium*. Turbidity can be determined remotely by utilizing light spectra such as green and red bands. Temperature is also important for water quality, as it can affect the water's biological processes and chemistry. Total nitrogen and total phosphorous are optically inactive parameters that can significantly impact the overall health of aquatic ecosystems. Dissolved oxygen is crucial for the survival of aquatic living organisms and plants. pH is a fundamental indicator of water quality, influencing various chemical, biological, and physical characteristics essential for sustaining life and ecosystems. Other parameters such as CDOM, Secchi depth, TDS, and TSS can also be effectively measured using remote sensing.

Machine learning is an essential component of remote sensing-based water quality monitoring. The data captured by the sensors are provided as input to the machine learning models, which are trained for assessing different parameters using data patterns. SVM and random forest are highly successful machine learning techniques that prove to be useful in water quality prediction. Several boosting algorithms evolved over time using weaker models to form a strong learner. Deep learning techniques have recently emerged and replaced conventional machine learning techniques. Deep learning has strong generalization abilities and superior learning capabilities due to a number of computing

layers. However, machine learning and deep learning models are often considered black box models due to the lack of proper explanation involved in the decision-making. Despite the lack of transparency, these models are widely used due to their adaptability and predictive analytical abilities. Water quality monitoring using high-resolution remote sensing imagery serves as a fundamental input for machine learning and deep learning models, enabling accurate predictions.

7.2. Implications for Policy and Practice in Water Quality Assessment

Water quality is crucial for human health and biological processes in aquatic ecology. A well-designed policy and good practices are essential for public health protection and environmental conservation. Policymakers should consider aligning water resource monitoring requirements with their country-specific standards to ensure effective and context-appropriate implementation. Promoting integrated water management and related resources is important for social welfare. Industrial and agriculture pollution, personal care products, medical discharge, etc. are responsible for contaminating the water resources. Best management practices should be promoted to reduce the harmful discharge into water bodies. Comprehensive water quality monitoring at regular intervals is essential for effective water management. For remote sensing-based quality assessment, the availability of adequate data is crucial. Free access to reliable data can promote the use of modern technologies for sustainable practices. Accordingly, investments are required to modernize infrastructure and promote innovative methods to improve the reliability and temporal/spatial coverage of water quality monitoring. Considering the complexity of water quality issues, experts from different disciplines are required to work together. Therefore, interdisciplinary research and the engagement of experts from government and non-government sectors and academia is essential. Training and skill development programs are necessary for the capacity building of stakeholders for the successful implementation of advanced technologies.

7.3. Reflections on the Future of Water Quality Monitoring Techniques

Due to anthropogenic activities and other natural causes, the environment is changing rapidly, increasing stress on natural resources. On the other hand, advances in technology have changed human life. All these aspects will shape the future of water quality monitoring. Currently, remote sensing-based monitoring mostly relies on sensors on board satellites. However, advanced sensors can provide additional or complementary information. Advanced nanosensors can efficiently detect extremely low concentrations in water. Nanosensors can identify heavy metals like mercury, lead, and arsenic. They can detect bacteria, viruses, and protozoa, and trace out contaminants based on their electromagnetic characteristics to ensure water is safe for drinking and other usage. Unmanned aerial vehicles (UAVs) can capture images of water bodies at better spatial resolutions. However, UAVs cannot provide more extensive coverage. Automated underwater vehicles (AUVs) can provide another dimension to monitoring by capturing data in different aquatic environments [182]. Data from different sources can be integrated for comprehensive water quality monitoring. Developing low-cost, affordable sensors, free access to data, and use advanced computing techniques can potentially ensure sustainable and comprehensive water quality monitoring.

8. Future Directions

Future research in water quality monitoring using remote sensing focuses on developing more accurate, efficient, and robust methods for automatically assessing different water quality parameters. The key research areas in this direction are as follows.

1. **Enhanced Data Integration:** Satellites and other remote sensing platforms capture data at varying resolutions in different formats that complicate the integration of data from multiple sources. On the other hand, different sensors have different radiometric and geometric characteristics. The variability in data quality can affect the accuracy of the assessments methods. The integration methods should be able to quantify uncertainties in resultant data [183]. The integration of large datasets induces a large computational burden also. The complex challenges in data integration should lead to the development of enhanced and robust algorithms to produce reliable, consistent, and comprehensive datasets.
2. **Generalized Algorithms:** The characteristics of water bodies across the globe vary widely due to geographic conditions, environment, human activities, etc. [184]. Most algorithms are evaluated for specific water bodies and may lack broad applicability. Machine learning algorithms, which are highly influenced by the training data, may be inconsistent under varying conditions [185]. Generalized algorithms are required to monitor different types of water bodies across different regions consistently. Generalized algorithms are robust, scalable, and cost effective, and can be applied over a large area without significant adjustments.
3. **Emerging Contaminants:** Much work has been conducted on assessing some well-known water quality parameters including Chl-a, pH, total nitrogen, total phosphorous, and dissolved oxygen, etc. [28,68]. However, due to anthropogenic activities, other chemicals and microorganisms such as pharmaceuticals, industrial waste, nano-materials, personal care products, etc., enter water bodies. Such pollutants can badly affect the ecological system and human health. Future directions include working on monitoring these emerging contaminants using remote sensing data.
4. **Impact of Climate Change:** Climate changes such as rising temperature and increasing CO₂ water influence water chemistry and water quality [186]. The climate change may cause temporal and spatial variations in water quality, making the monitoring of trends challenging. The monitoring methods need to be adaptive to emerging issues and able to generate early warnings on future conditions due to climate changes.

Author Contributions: All authors contributed substantially to the manuscript. A.P.N. conceptualized the review and supervised all the stages. S.M. and B.K. designed the basic structure of the manuscript. All authors discussed and finalized the structure. S.M. wrote the manuscript. A.P.N. and B.K. edited the review. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. NASA Science: SpacePlace. Available online: <https://spaceplace.nasa.gov/> (accessed on 29 December 2023).
2. Chu, H.J.; Nguyen, M.V.; Jaelani, L.M. Satellite-Based Water Quality Mapping from Sequential Simulation with Parameter Outlier Removal. *Water Resour. Manag.* **2020**, *34*, 311–325. [CrossRef]
3. Cai, X.; Wu, L.; Li, Y.; Lei, S.; Xu, J.; Lyu, H.; Li, J.; Wang, H.; Dong, X.; Zhu, Y.; et al. Remote sensing identification of urban water pollution source types using hyperspectral data. *J. Hazard. Mater.* **2023**, *459*, 132080. [CrossRef]
4. Rajaram, T.; Das, A. Water pollution by industrial effluents in India: Discharge scenarios and case for participatory ecosystem specific local regulation. *Futures* **2008**, *40*, 56–69. [CrossRef]

5. Wang, Z.; Zhang, S.; Peng, Y.; Wu, C.; Lv, Y.; Xiao, K.; Zhao, J.; Qian, G. Impact of rapid urbanization on the threshold effect in the relationship between impervious surfaces and water quality in Shanghai, China. *Environ. Pollut.* **2020**, *267*, 115569. [CrossRef] [PubMed]
6. Song, K.; Li, L.; Wang, Z.; Liu, D.; Zhang, B.; Xu, J.; Du, J.; Li, L.; Li, S.; Wang, Y. Retrieval of total suspended matter (TSM) and chlorophyll-a (Chl-a) concentration from remote-sensing data for drinking water resources. *Environ. Monit. Assess.* **2012**, *184*, 1449–1470. [CrossRef]
7. Kowe, P.; Ncube, E.; Magidi, J.; Ndambuki, J.M.; Rwasoka, D.T.; Gumindoga, W.; Maviza, A.; de Jesus Paulo Mavaringana, M.; Kakanda, E.T. Spatial-temporal variability analysis of water quality using remote sensing data: A case study of Lake Manyame. *Sci. Afr.* **2023**, *21*, e01877. [CrossRef]
8. World Population Prospects 2022. Available online: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf (accessed on 28 December 2023).
9. Kapalanga, T.S.; Hoko, Z.; Gumindoga, W.; Chikwiramakomo, L. Remote-sensing-based algorithms for water quality monitoring in Olushandja Dam, north-central Namibia. *Water Supply* **2020**, *21*, 1878–1894. [CrossRef]
10. McCarthy, M.J.; Otis, D.B.; Méndez-Lázaro, P.; Muller-Karger, F.E. Water Quality Drivers in 11 Gulf of Mexico Estuaries. *Remote Sens.* **2018**, *10*, 255. [CrossRef]
11. Chen, C.; Tang, S.; Pan, Z.; Zhan, H.; Larson, M.; Jönsson, L. Remotely sensed assessment of water quality levels in the Pearl River Estuary, China. *Mar. Pollut. Bull.* **2007**, *54*, 1267–1272. [CrossRef] [PubMed]
12. Joshi, I.; D'Sa, E.J. Seasonal Variation of Colored Dissolved Organic Matter in Barataria Bay, Louisiana, Using Combined Landsat and Field Data. *Remote Sens.* **2015**, *7*, 12478–12502. [CrossRef]
13. Sudduth, K.A.; Jang, G.S.; Lerch, R.N.; Sadler, E.J. Long-Term Agroecosystem Research in the Central Mississippi River Basin: Hyperspectral Remote Sensing of Reservoir Water Quality. *J. Environ. Qual.* **2015**, *44*, 71–83. [CrossRef]
14. Muchini, R.; Gumindoga, W.; Togarepi, S.; Masarira, T.P.; Dube, T. Near real time water quality monitoring of Chivero and Manyame lakes of Zimbabwe. *Proc. Int. Assoc. Hydrol. Sci.* **2018**, *378*, 85–92. [CrossRef]
15. Jia, M.; Wang, Z.; Mao, D.; Ren, C.; Song, K.; Zhao, C.; Wang, C.; Xiao, X.; Wang, Y. Mapping global distribution of mangrove forests at 10-m resolution. *Sci. Bull.* **2023**, *68*, 1306–1316. [CrossRef]
16. Wu, S.; Qi, J.; Yan, Z.; Lyu, F.; Lin, T.; Wang, Y.; Du, Z. Spatiotemporal assessments of nutrients and water quality in coastal areas using remote sensing and a spatiotemporal deep learning model. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102897. [CrossRef]
17. Ahmed, W.; Mohammed, S.; El-Shazly, A.; Morsy, S. Tigris River water surface quality monitoring using remote sensing data and GIS techniques. *Egypt. J. Remote Sens. Space Sci.* **2023**, *26*, 816–825. [CrossRef]
18. Xiong, J.; Lin, C.; Cao, Z.; Hu, M.; Xue, K.; Chen, X.; Ma, R. Development of remote sensing algorithm for total phosphorus concentration in eutrophic lakes: Conventional or machine learning? *Water Res.* **2022**, *215*, 118213. [CrossRef] [PubMed]
19. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Moher, D. Updating guidance for reporting systematic reviews: Development of the PRISMA 2020 statement. *J. Clin. Epidemiol.* **2021**, *134*, 103–112. [CrossRef] [PubMed]
20. Escoto, J.E.; Blanco, A.C.; Argamosa, R.J.; Medina, J.M. Pasig river water quality estimation using an empirical ordinary least squares regression model of sentinel-2 satellite images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *XLVI-4/W6-2021*, 161–168. [CrossRef]
21. Crawford, C.J.; Roy, D.P.; Arab, S.; Barnes, C.; Vermote, E.; Hulley, G.; Gerace, A.; Choate, M.; Engebretson, C.; Micijevic, E.; et al. The 50-year Landsat collection 2 archive. *Sci. Remote Sens.* **2023**, *8*, 100103. [CrossRef]
22. Pahlevan, N.; Schott, J.R. Characterizing the relative calibration of Landsat-7 (ETM+) visible bands with Terra (MODIS) over clear waters: The implications for monitoring water resources. *Remote Sens. Environ.* **2012**, *125*, 167–180. [CrossRef]
23. Zuccari Fernandes Braga, C.; Setzer, A.W.; Drude de Lacerda, L. Water quality assessment with simultaneous Landsat-5 TM data at Guanabara Bay, Rio de Janeiro, Brazil. *Remote Sens. Environ.* **1993**, *45*, 95–106. [CrossRef]
24. González-Márquez, L.C.; Torres-Bejarano, F.M.; Torregroza-Espinosa, A.C.; Hansen-Rodríguez, I.R.; Rodríguez-Gallegos, H.B. Use of LANDSAT 8 images for depth and water quality assessment of El Guájar reservoir, Colombia. *J. S. Am. Earth Sci.* **2018**, *82*, 231–238. [CrossRef]
25. Trevisiol, F.; Mandanici, E.; Pagliarini, A.; Bitelli, G. Evaluation of Landsat-9 interoperability with Sentinel-2 and Landsat-8 over Europe and local comparison with field surveys. *ISPRS J. Photogramm. Remote Sens.* **2024**, *210*, 55–68. [CrossRef]
26. Vakili, T.; Amanollahi, J. Determination of optically inactive water quality variables using Landsat 8 data: A case study in Geshlagh reservoir affected by agricultural land use. *J. Clean. Prod.* **2020**, *247*, 119134. [CrossRef]
27. Chen, X.; Liu, L.; Zhang, X.; Li, J.; Wang, S.; Liu, D.; Duan, H.; Song, K. An Assessment of Water Color for Inland Water in China Using a Landsat 8-Derived Forel–Ule Index and the Google Earth Engine Platform. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 5773–5785. [CrossRef]
28. Guo, H.; Tian, S.; Jeanne Huang, J.; Zhu, X.; Wang, B.; Zhang, Z. Performance of deep learning in mapping water quality of Lake Simcoe with long-term Landsat archive. *ISPRS J. Photogramm. Remote Sens.* **2022**, *183*, 451–469. [CrossRef]

29. Markogianni, V.; Kalivas, D.; Petropoulos, G.P.; Dimitriou, E. An Appraisal of the Potential of Landsat 8 in Estimating Chlorophyll-a, Ammonium Concentrations and Other Water Quality Indicators. *Remote Sens.* **2018**, *10*, 1018. [\[CrossRef\]](#)
30. Al-Shaibah, B.; Liu, X.; Zhang, J.; Tong, Z.; Zhang, M.; El-Zeiny, A.; Faichia, C.; Hussain, M.; Tayyab, M. Modeling Water Quality Parameters Using Landsat Multispectral Images: A Case Study of Erlong Lake, Northeast China. *Remote Sens.* **2021**, *13*, 1603. [\[CrossRef\]](#)
31. He, Y.; Jin, S.; Shang, W. Water Quality Variability and Related Factors along the Yangtze River Using Landsat-8. *Remote Sens.* **2021**, *13*, 2241. [\[CrossRef\]](#)
32. Zhang, H.; Xue, B.; Wang, G.; Zhang, X.; Zhang, Q. Deep Learning-Based Water Quality Retrieval in an Impounded Lake Using Landsat 8 Imagery: An Application in Dongping Lake. *Remote Sens.* **2022**, *14*, 4505. [\[CrossRef\]](#)
33. Liao, K.; Song, Y.; Nie, X.; Liu, L.; Qi, S. Suspended Sediment Concentrate Estimation from Landsat Imagery and Hydrological Station in Poyang Lake Using Machine Learning. *IEEE Access* **2024**, *12*, 85411–85422. [\[CrossRef\]](#)
34. Shamloo, A.; Sima, S. Investigating the potential of remote sensing-based machine-learning algorithms to model Secchi-disk depth, total phosphorus, and chlorophyll-a in Lake Urmia. *J. Great Lakes Res.* **2024**, *50*, 102370. [\[CrossRef\]](#)
35. Rajaveni, S.; Muniappan, N.; Nandhu, M.; Madhavan, V.S.; Kumar, T.P. Assessment of Surface Water Quality Based on Landsat 9 Operational Land Imager Combined with GIS and IOT. *J. Indian Soc. Remote Sens.* **2024**, *52*, 139–151. [\[CrossRef\]](#)
36. Zhang, Y.; He, X.; Lian, G.; Bai, Y.; Yang, Y.; Gong, F.; Wang, D.; Zhang, Z.; Li, T.; Jin, X. Monitoring and spatial traceability of river water quality using Sentinel-2 satellite images. *Sci. Total Environ.* **2023**, *894*, 164862. [\[CrossRef\]](#)
37. Meng, H.; Zhang, J.; Zheng, Z.; Song, Y.; Lai, Y. Classification of inland lake water quality levels based on Sentinel-2 images using convolutional neural networks and spatiotemporal variation and driving factors of algal bloom. *Ecol. Inform.* **2024**, *80*, 102549. [\[CrossRef\]](#)
38. Casal, G. Assessment of Sentinel-2 to monitor highly dynamic small water bodies: The case of Louro lagoon (Galicia, NW Spain). *Oceanologia* **2022**, *64*, 88–102. [\[CrossRef\]](#)
39. Zhao, D.; Huang, J.; Li, Z.; Yu, G.; Shen, H. Dynamic monitoring and analysis of chlorophyll-a concentrations in global lakes using Sentinel-2 images in Google Earth Engine. *Sci. Total Environ.* **2024**, *912*, 169152. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Ndou, N. Geostatistical inference of Sentinel-2 spectral reflectance patterns to water quality indicators in the Setumo dam, South Africa. *Remote Sens. Appl. Soc. Environ.* **2023**, *30*, 100945. [\[CrossRef\]](#)
41. Caballero, I.; Navarro, G. Monitoring cyanoHABs and water quality in Laguna Lake (Philippines) with Sentinel-2 satellites during the 2020 Pacific typhoon season. *Sci. Total Environ.* **2021**, *788*, 147700. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Viridis, S.G.; Xue, W.; Winijkul, E.; Nitivattananon, V.; Punpukdee, P. Remote sensing of tropical riverine water quality using sentinel-2 MSI and field observations. *Ecol. Indic.* **2022**, *144*, 109472. [\[CrossRef\]](#)
43. Chu, H.J.; He, Y.C. Remote sensing water quality inversion using sparse representation: Chlorophyll-a retrieval from Sentinel-2 MSI data. *Remote Sens. Appl. Soc. Environ.* **2023**, *31*, 101006. [\[CrossRef\]](#)
44. Germán, A.; Shimoni, M.; Beltramone, G.; Rodríguez, M.I.; Muchiut, J.; Bonansea, M.; Scavuzzo, C.M.; Ferral, A. Space-time monitoring of water quality in an eutrophic reservoir using SENTINEL-2 data - A case study of San Roque, Argentina. *Remote Sens. Appl. Soc. Environ.* **2021**, *24*, 100614. [\[CrossRef\]](#)
45. Maimouni, S.; Moufkari, A.A.; Daghor, L.; Fekri, A.; Oubraim, S.; Lhissou, R. Spatiotemporal monitoring of low water turbidity in Moroccan coastal lagoon using Sentinel-2 data. *Remote Sens. Appl. Soc. Environ.* **2022**, *26*, 100772. [\[CrossRef\]](#)
46. Xie, Y.; Zhou, Y.; Tao, Z.; Shao, W.; Yang, M. Remote Sensing Inversion of the Total Suspended Matter Concentration in the Nanyi Lake Based on Sentinel-3 OLCI Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 10380–10389. [\[CrossRef\]](#)
47. Khan, R.M.; Salehi, B.; Niroumand-Jadidi, M.; Mahdianpari, M. Global vs Local Random Forest Model for Water Quality Monitoring: Assessment in Finger Lakes Using Sentinel-2 Imagery and Gloria Dataset. In Proceedings of the IGARSS 2024—2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 7–12 July 2024; IEEE: New York, NY, USA, 2024; pp. 4389–4392.
48. Arıman, S. Determination of inactive water quality variables by MODIS data: A case study in the Kızılırmak Delta-Balik Lake, Turkey. *Estuar. Coast. Shelf Sci.* **2021**, *260*, 107505. [\[CrossRef\]](#)
49. DeVisser, M.H.; Messina, J.P. Exploration of sensor comparability: A case study of composite MODIS Aqua and Terra data. *Remote Sens. Lett.* **2013**, *4*, 599–608. [\[CrossRef\]](#)
50. Kang, E.; Park, S.; Kim, M.; Yoo, C.; Im, J.; Song, C.K. Direct aerosol optical depth retrievals using MODIS reflectance data and machine learning over East Asia. *Atmos. Environ.* **2023**, *309*, 119951. [\[CrossRef\]](#)
51. Singh, A.; Jakubowski, A.R.; Chidister, I.; Townsend, P.A. A MODIS approach to predicting stream water quality in Wisconsin. *Remote Sens. Environ.* **2013**, *128*, 74–86. [\[CrossRef\]](#)
52. Xia, K.; Wu, T.; Li, X.; Wang, S.; Shen, Q. A new method for accurate inversion of Forel-Ule index using MODIS images—Revealing the water color evolution in China's large lakes and reservoirs over the past two decades. *Water Res.* **2024**, *255*, 121560. [\[CrossRef\]](#) [\[PubMed\]](#)

53. Ayana, E.K.; Worqlul, A.W.; Steenhuis, T.S. Evaluation of stream water quality data generated from MODIS images in modeling total suspended solid emission to a freshwater lake. *Sci. Total Environ.* **2015**, *523*, 170–177. [[CrossRef](#)] [[PubMed](#)]
54. Katlane, R.; El Kilani, B.; Dhaoui, O.; Kateb, F.; Chehata, N. Monitoring of sea surface temperature, chlorophyll, and turbidity in Tunisian waters from 2005 to 2020 using MODIS imagery and the Google Earth Engine. *Reg. Stud. Mar. Sci.* **2023**, *66*, 103143. [[CrossRef](#)]
55. Kim, H.C.; Son, S.; Kim, Y.H.; Khim, J.S.; Nam, J.; Chang, W.K.; Lee, J.H.; Lee, C.H.; Ryu, J. Remote sensing and water quality indicators in the Korean West coast: Spatio-temporal structures of MODIS-derived chlorophyll-a and total suspended solids. *Mar. Pollut. Bull.* **2017**, *121*, 425–434. [[CrossRef](#)]
56. Rahat, S.H.; Steissberg, T.; Chang, W.; Chen, X.; Mandavya, G.; Tracy, J.; Wasti, A.; Atreya, G.; Saki, S.; Bhuiyan, M.A.E.; et al. Remote sensing-enabled machine learning for river water quality modeling under multidimensional uncertainty. *Sci. Total Environ.* **2023**, *898*, 165504. [[CrossRef](#)] [[PubMed](#)]
57. Kim, Y.H.; Son, S.; Kim, H.C.; Kim, B.; Park, Y.G.; Nam, J.; Ryu, J. Application of satellite remote sensing in monitoring dissolved oxygen variabilities: A case study for coastal waters in Korea. *Environ. Int.* **2020**, *134*, 105301. [[CrossRef](#)] [[PubMed](#)]
58. Tian, D.; Zhao, X.; Gao, L.; Liang, Z.; Yang, Z.; Zhang, P.; Wu, Q.; Ren, K.; Li, R.; Yang, C.; et al. Estimation of water quality variables based on machine learning model and cluster analysis-based empirical model using multi-source remote sensing data in inland reservoirs, South China. *Environ. Pollut.* **2024**, *342*, 123104. [[CrossRef](#)]
59. P. A. Brivio, C.G.; Zilioli, E. Determination of chlorophyll concentration changes in Lake Garda using an image-based radiative transfer code for Landsat TM images. *Int. J. Remote Sens.* **2001**, *22*, 487–502. [[CrossRef](#)]
60. Svirčev, Z.; Simeunović, J.; Subakov-Simić, G.; Krstić, S.; Pantelić, D.; Dulić, T. Cyanobacterial Blooms and their Toxicity in Vojvodina Lakes, Serbia. *Int. J. Environ. Res.* **2013**, *7*, 745–758. [[CrossRef](#)]
61. Cao, Z.; Ma, R.; Pahlevan, N.; Liu, M.; Melack, J.M.; Duan, H.; Xue, K.; Shen, M. Evaluating and optimizing VIIRS retrievals of chlorophyll-a and suspended particulate matter in turbid lakes using a machine learning approach. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4211417. [[CrossRef](#)]
62. Khan, R.M.; Salehi, B.; Mahdianpari, M. Machine learning methods for water quality monitoring over Finger Lakes using Sentinel-2. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: New York, NY, USA, 2022; pp. 6316–6319.
63. Li, H.; Blix, K.; Somogyi, B.; Tóth, V.R. Retrieving Chlorophyll-A Concentration for Lake Balaton with Landsat Based on GEE. In Proceedings of the IGARSS 2023—2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023; IEEE: New York, NY, USA, 2023; pp. 460–463.
64. Karimi, B.; Hashemi, S.H.; Aghighi, H. Application of Landsat-8 and Sentinel-2 for retrieval of chlorophyll-a in a shallow freshwater lake. *Adv. Space Res.* **2024**, *74*, 117–129. [[CrossRef](#)]
65. Mansaray, A.S.; Dzialowski, A.R.; Martin, M.E.; Wagner, K.L.; Gholizadeh, H.; Stoodley, S.H. Comparing PlanetScope to Landsat-8 and Sentinel-2 for Sensing Water Quality in Reservoirs in Agricultural Watersheds. *Remote Sens.* **2021**, *13*, 1847. [[CrossRef](#)]
66. Niroumand-Jadidi, M.; Bovolo, F.; Bresciani, M.; Gege, P.; Giardino, C. Water quality retrieval from landsat-9 (OLI-2) imagery and comparison to sentinel-2. *Remote Sens.* **2022**, *14*, 4596. [[CrossRef](#)]
67. Yao, J.; Sun, S.; Zhai, H.; Feger, K.H.; Zhang, L.; Tang, X.; Li, G.; Wang, Q. Dynamic monitoring of the largest reservoir in North China based on multi-source satellite remote sensing from 2013 to 2022: Water area, water level, water storage and water quality. *Ecol. Indic.* **2022**, *144*, 109470. [[CrossRef](#)]
68. Leggesse, E.S.; Zimale, F.A.; Sultan, D.; Enku, T.; Srinivasan, R.; Tilahun, S.A. Predicting Optical Water Quality Indicators from Remote Sensing Using Machine Learning Algorithms in Tropical Highlands of Ethiopia. *Hydrology* **2023**, *10*, 110. [[CrossRef](#)]
69. Pisanti, A.; Magri, S.; Ferrando, I.; Federici, B. Sea water turbidity analysis from sentinel-2 images: Atmo-spheric correction and bands correlation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *XLVIII-4/W1-2022*, 371–378. [[CrossRef](#)]
70. Stevenson, M.; Bravo, C. Advanced turbidity prediction for operational water supply planning. *Decis. Support Syst.* **2019**, *119*, 72–84. [[CrossRef](#)]
71. Miglino, D.; Jomaa, S.; Rode, M.; Isgro, F.; Cathleen Saddi, K.; Manfreda, S. The use of optical camera for river turbidity monitoring. In Proceedings of the EGU23, the 25th EGU General Assembly, Vienna, Austria and Online, 23–28 April 2023. [[CrossRef](#)]
72. Lioumbas, J.; Christodoulou, A.; Katsiapi, M.; Xanthopoulou, N.; Stournara, P.; Spahos, T.; Seretoudi, G.; Mentis, A.; Theodoridou, N. Satellite remote sensing to improve source water quality monitoring: A water utility's perspective. *Remote Sens. Appl. Soc. Environ.* **2023**, *32*, 101042. [[CrossRef](#)]
73. Potes, M.; Rodrigues, G.; Penha, A.M.; Novais, M.H.; Costa, M.J.; Salgado, R.; Morais, M.M. Use of Sentinel 2 – MSI for water quality monitoring at Alqueva reservoir, Portugal. *Proc. Int. Assoc. Hydrol. Sci.* **2018**, *380*, 73–79. [[CrossRef](#)]
74. Gu, K.; Zhang, Y.; Qiao, J. Random forest ensemble for river turbidity measurement from space remote sensing data. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 9028–9036. [[CrossRef](#)]
75. Devi, P.D.; Mamatha, G. Machine learning approach to predict the turbidity of Saki Lake, Telangana, India, using remote sensing data. *Meas. Sens.* **2024**, *33*, 101139. [[CrossRef](#)]

76. Zhang, J.; Meng, F.; Fu, P.; Jing, T.; Xu, J.; Yang, X. Tracking changes in chlorophyll-a concentration and turbidity in Nansi Lake using Sentinel-2 imagery: A novel machine learning approach. *Ecol. Infor.* **2024**, *81*, 102597. [[CrossRef](#)]
77. Singh, R.; Saritha, V.; Pande, C.B. Monitoring of wetland turbidity using multi-temporal Landsat-8 and Landsat-9 satellite imagery in the Bisalpur wetland, Rajasthan, India. *Environ. Res.* **2024**, *241*, 117638. [[CrossRef](#)] [[PubMed](#)]
78. Ma, Y.; Song, K.; Wen, Z.; Liu, G.; Shang, Y.; Lyu, L.; Du, J.; Yang, Q.; Li, S.; Tao, H.; et al. Remote sensing of turbidity for lakes in northeast China using Sentinel-2 images with machine learning algorithms. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 9132–9146. [[CrossRef](#)]
79. Clermont, M.; Kinnard, C.; Dubé-Richard, D.; Campeau, S.; Bordeleau, P.A.; de Grandpré, A.; Ziyad, J.; Roy, A. Using remote sensing to assess how intensive agriculture impacts the turbidity of a fluvial lake floodplain. *J. Great Lakes Res.* **2023**, *49*, 102240. [[CrossRef](#)]
80. Ramesh, J.V.N.; Patibandla, P.R.; Shanbhog, M.; Ambala, S.; Ashraf, M.; Kiran, A. Ensemble deep learning approach for turbidity prediction of Dooskal Lake using remote sensing data. *Remote Sens. Earth Syst. Sci.* **2023**, *6*, 146–155. [[CrossRef](#)]
81. Li, Y.; Li, S.; Song, K.; Liu, G.; Wen, Z.; Fang, C.; Shang, Y.; Lyu, L.; Zhang, L. Sentinel-3 OLCI observations of Chinese lake turbidity using machine learning algorithms. *J. Hydrol.* **2023**, *622*, 129668. [[CrossRef](#)]
82. Alcântara, E.H.; Stech, J.L.; Lorenzetti, J.A.; Bonnet, M.P.; Casamitjana, X.; Assireu, A.T.; de Moraes Novo, E.M.L. Remote sensing of water surface temperature and heat flux over a tropical hydroelectric reservoir. *Remote Sens. Environ.* **2010**, *114*, 2651–2665. [[CrossRef](#)]
83. Li, H.Y.; Xu, J.; Xu, R.Q. The effect of temperature on the water quality of lake. *Adv. Mater. Res.* **2013**, *821*, 1001–1004. [[CrossRef](#)]
84. Xu, L.; Li, H.; Liang, X.; Yao, Y.; Zhou, L.; Cui, X. Water quality parameters response to temperature change in small shallow lakes. *Phys. Chem. Earth, Parts A/B/C* **2012**, *47*, 128–134. [[CrossRef](#)]
85. Calero Preciado, C.; Boxall, J.; Soria-Carrasco, V.; Martínez, S.; Douterelo, I. Implications of climate change: How does increased water temperature influence biofilm and water quality of chlorinated drinking water distribution systems? *Front. Microbiol.* **2021**, *12*, 658927. [[CrossRef](#)]
86. Mohammed, H.; Tornyeviadzi, H.M.; Seidu, R. Modelling the impact of water temperature, pipe, and hydraulic conditions on water quality in water distribution networks. *Water Pract. Technol.* **2021**, *16*, 387–403. [[CrossRef](#)]
87. Zhu, X.; Guo, H.; Huang, J.J.; Tian, S.; Xu, W.; Mai, Y. An ensemble machine learning model for water quality estimation in coastal area based on remote sensing imagery. *J. Environ. Manag.* **2022**, *323*, 116187. [[CrossRef](#)]
88. Vanhellemont, Q. Automated water surface temperature retrieval from Landsat 8/TIRS. *Remote Sens. Environ.* **2020**, *237*, 111518. [[CrossRef](#)]
89. Krishnaraj, A.; Honnasiddaiah, R. Remote sensing and machine learning based framework for the assessment of spatio-temporal water quality in the Middle Ganga Basin. *Environ. Sci. Pollut. Res.* **2022**, *29*, 64939–64958. [[CrossRef](#)]
90. Aguilar, A.C.; Cerón-Vivas, A.; Altuve, M. Multivariate prediction of nitrogen concentration in a stream using regression models. *Environ. Earth Sci.* **2021**, *80*, 363. [[CrossRef](#)]
91. Wu, J.; Zhang, J.; Tan, W.; Lan, H.; Zhang, S.; Xiao, K.; Wang, L.; Lin, H.; Sun, G.; Guo, P. Application of time serial model in water quality predicting. *Comput. Mater. Contin.* **2023**, *74*, 67–82. [[CrossRef](#)]
92. Song, W.; Yinglan, A.; Wang, Y.; Fang, Q.; Tang, R. Study on remote sensing inversion and temporal-spatial variation of Hulun lake water quality based on machine learning. *J. Contam. Hydrol.* **2024**, *260*, 104282. [[CrossRef](#)]
93. Guo, H.; Huang, J.J.; Chen, B.; Guo, X.; Singh, V.P. A machine learning-based strategy for estimating non-optically active water quality parameters using Sentinel-2 imagery. *Int. J. Remote Sens.* **2021**, *42*, 1841–1866. [[CrossRef](#)]
94. Li, J.; Wang, J.; Wu, Y.; Cui, Y.; Yan, S. Remote sensing monitoring of total nitrogen and total phosphorus concentrations in the water around Chaohu Lake based on geographical division. *Front. Environ. Sci.* **2022**, *10*, 1014155. [[CrossRef](#)]
95. Altare, N.; Vione, D. Photochemical Implications of Changes in the Spectral Properties of Chromophoric Dissolved Organic Matter: A Model Assessment for Surface Waters. *Molecules* **2023**, *28*, 2664. [[CrossRef](#)]
96. Qi, G.; Zhang, B.; Tian, B.; Yang, R.; Baker, A.; Wu, P.; He, S. Characterization of Dissolved Organic Matter from Agricultural and Livestock Effluents: Implications for Water Quality Monitoring. *Int. J. Environ. Res. Public Health* **2023**, *20*, 5121. [[CrossRef](#)]
97. Li, D.; Pan, B.; Han, X.; Li, J.; Zhu, Q.; Li, M. Assessing the potential to use CDOM as an indicator of water quality for the sediment-laden Yellow river, China. *Environ. Pollut.* **2021**, *289*, 117970. [[CrossRef](#)]
98. Madonia, A.; Caruso, G.; Piazzolla, D.; Bonamano, S.; Piermattei, V.; Zappalà, G.; Marcelli, M. Chromophoric Dissolved Organic Matter as a tracer of fecal contamination for bathing water quality monitoring in the northern Tyrrhenian Sea (Latium, Italy). *J. Mar. Sci. Eng.* **2020**, *8*, 430. [[CrossRef](#)]
99. Qiang, S.; Song, K.; Shang, Y.; Lai, F.; Wen, Z.; Liu, G.; Tao, H.; Lyu, Y. Remote Sensing Estimation of CDOM and DOC with the Environmental Implications for Lake Khanka. *Remote Sens.* **2023**, *15*, 5707. [[CrossRef](#)]
100. Adjovu, G.E.; Stephen, H.; James, D.; Ahmad, S. Measurement of total dissolved solids and total suspended solids in water systems: A review of the issues, conventional, and remote sensing techniques. *Remote Sens.* **2023**, *15*, 3534. [[CrossRef](#)]

101. Sutherland, C.W. *Spectral Analysis of Total Suspended Solids Mixtures for Solids Composition Determination*; Louisiana State University and Agricultural & Mechanical College: Baton Rouge, LA, USA 2006.
102. Wang, X.; Wen, Z.; Liu, G.; Tao, H.; Song, K. Remote estimates of total suspended matter in China's main estuaries using Landsat images and a weight random forest model. *ISPRS J. Photogramm. Remote Sens.* **2022**, *183*, 94–110. [\[CrossRef\]](#)
103. Pahlevan, N.; Smith, B.; Alikas, K.; Anstee, J.; Barbosa, C.; Binding, C.; Bresciani, M.; Cremella, B.; Giardino, C.; Gurlin, D.; et al. Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. *Remote Sens. Environ.* **2022**, *270*, 112860. [\[CrossRef\]](#)
104. Silveira Kupssinskü, L.; Thomassim Guimarães, T.; Menezes de Souza, E.; C. Zanotta, D.; Roberto Veronez, M.; Gonzaga, L., Jr.; Mauad, F.F. A method for chlorophyll-a and suspended solids prediction through remote sensing and machine learning. *Sensors* **2020**, *20*, 2125. [\[CrossRef\]](#) [\[PubMed\]](#)
105. Zhao, J.; Zhang, F.; Chen, S.; Wang, C.; Chen, J.; Zhou, H.; Xue, Y. Remote sensing evaluation of total suspended solids dynamic with Markov model: A case study of inland reservoir across administrative boundary in South China. *Sensors* **2020**, *20*, 6911. [\[CrossRef\]](#) [\[PubMed\]](#)
106. Aljoborey, A.D.A.; Abdulhay, H.S. Estimating total dissolved solids and total suspended solids in Mosul dam lake in situ and using remote sensing technique. *Period. Eng. Nat. Sci.* **2019**, *7*, 1755–1767. [\[CrossRef\]](#)
107. Al-Fahdawi, A.A.; Rabee, A.M.; Al-Hirmizy, S.M. Water quality monitoring of Al-Habbaniyah Lake using remote sensing and in situ measurements. *Environ. Monit. Assess.* **2015**, *187*, 367. [\[CrossRef\]](#)
108. Jiang, D.; Matsushita, B.; Pahlevan, N.; Gurlin, D.; Lehmann, M.K.; Fichot, C.G.; Schalles, J.; Loisel, H.; Binding, C.; Zhang, Y.; et al. Remotely estimating total suspended solids concentration in clear to extremely turbid waters using a novel semi-analytical method. *Remote Sens. Environ.* **2021**, *258*, 112386. [\[CrossRef\]](#)
109. Seleem, T.; Bafi, D.; Karantzia, M.; Parcharidis, I. Water Quality Monitoring Using Landsat 8 and Sentinel-2 Satellite Data (2014–2020) in Timsah Lake, Ismailia, Suez Canal Region (Egypt). *J. Indian Soc. Remote Sens.* **2022**, *50*, 2411–2428. [\[CrossRef\]](#)
110. Wang, C.; Li, W.; Chen, S.; Li, D.; Wang, D.; Liu, J. The spatial and temporal variation of total suspended solid concentration in Pearl River Estuary during 1987–2015 based on remote sensing. *Sci. Total Environ.* **2018**, *618*, 1125–1138. [\[CrossRef\]](#) [\[PubMed\]](#)
111. Niroumand-Jadidi, M.; Bovolo, F. Temporally transferable machine learning model for total suspended matter retrieval from sentinel-2. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *3*, 339–345. [\[CrossRef\]](#)
112. Bozorg-Haddad, O.; Delpasand, M.; Loáiciga, H.A. Water quality, hygiene, and health. In *Economical, Political, and Social Issues in Water Resources*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 217–257.
113. Tian, S.; Guo, H.; Xu, W.; Zhu, X.; Wang, B.; Zeng, Q.; Mai, Y.; Huang, J.J. Remote sensing retrieval of inland water quality parameters using Sentinel-2 and multiple machine learning algorithms. *Environ. Sci. Pollut. Res.* **2023**, *30*, 18617–18630. [\[CrossRef\]](#) [\[PubMed\]](#)
114. Adams, H.; Burlingame, G.; Ikehata, K.; Furatian, L.; Suffet, I. The effect of pH on taste and odor production and control of drinking water. *AQUA—Water Infrastruct. Ecosyst. Soc.* **2022**, *71*, 1278–1290. [\[CrossRef\]](#)
115. Zangmo, P.; Singh, D.; Raza, Y.; Narayan, R. A review on water Quality Parameter. *J. Emerg. Technol. Innov. Res.* **2021**, *8*, 607–613.
116. Huang, H.; Feng, R.; Zhu, J.; Li, P. Prediction of pH value by multi-classification in the Weizhou Island area. *Sensors* **2019**, *19*, 3875. [\[CrossRef\]](#) [\[PubMed\]](#)
117. Abdelmalik, K. Role of statistical remote sensing for Inland water quality parameters prediction. *Egypt. J. Remote Sens. Space Sci.* **2018**, *21*, 193–200. [\[CrossRef\]](#)
118. Pereira, O.J.; Merino, E.R.; Montes, C.R.; Barbiero, L.; Rezende-Filho, A.T.; Lucas, Y.; Melfi, A.J. Estimating water pH using cloud-based landsat images for a new classification of the Nhecolândia Lakes (Brazilian Pantanal). *Remote Sens.* **2020**, *12*, 1090. [\[CrossRef\]](#)
119. Yang, W.; Fu, B.; Li, S.; Lao, Z.; Deng, T.; He, W.; He, H.; Chen, Z. Monitoring multi-water quality of internationally important karst wetland through deep learning, multi-sensor and multi-platform remote sensing images: A case study of Guilin, China. *Ecol. Indic.* **2023**, *154*, 110755. [\[CrossRef\]](#)
120. Hassija, V.; Chamola, V.; Mahapatra, A.; Singal, A.; Goel, D.; Huang, K.; Scardapane, S.; Spinelli, I.; Mahmud, M.; Hussain, A. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cogn. Comput.* **2024**, *16*, 45–74. [\[CrossRef\]](#)
121. Vapnik, V. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999. [\[CrossRef\]](#) [\[PubMed\]](#)
122. Li, Q.; Lei, T.; Cheng, Y.; Wei, X.; Sun, D.W. Predicting wheat gluten concentrations in potato starch using GPR and SVM models built by terahertz time-domain spectroscopy. *Food Chem.* **2024**, *432*, 137235. [\[CrossRef\]](#) [\[PubMed\]](#)
123. Çakir, M.; Yilmaz, M.; Oral, M.A.; Özgür Kazancı, H.; Oral, O. Accuracy assessment of RFerns, NB, SVM, and kNN machine learning classifiers in aquaculture. *J. King Saud Univ. -Sci.* **2023**, *35*, 102754. [\[CrossRef\]](#)
124. Nasir, N.; Kansal, A.; Alshaltone, O.; Barneih, F.; Sameer, M.; Shanableh, A.; Al-Shamma'a, A. Water quality classification using machine learning algorithms. *J. Water Process Eng.* **2022**, *48*, 102920. [\[CrossRef\]](#)
125. Cervantes, J.; Garcia-Lamont, F.; Rodríguez-Mazahua, L.; Lopez, A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* **2020**, *408*, 189–215. [\[CrossRef\]](#)

126. Kamyab-Talesh, F.; Mousavi, S.F.; Khaledian, M.; Yousefi-Falakdehi, O.; Norouzi-Masir, M. Prediction of water quality index by support vector machine: A case study in the Sefidrud Basin, Northern Iran. *Water Resour.* **2019**, *46*, 112–116. [\[CrossRef\]](#)
127. Abobakr Yahya, A.S.; Ahmed, A.N.; Binti Othman, F.; Ibrahim, R.K.; Afan, H.A.; El-Shafie, A.; Fai, C.M.; Hossain, M.S.; Ehteram, M.; Elshafie, A. Water Quality Prediction Model Based Support Vector Machine Model for Ungauged River Catchment under Dual Scenarios. *Water* **2019**, *11*, 1231. [\[CrossRef\]](#)
128. Deng, T.; Chau, K.W.; Duan, H.F. Machine learning based marine water quality prediction for coastal hydro-environment management. *J. Environ. Manag.* **2021**, *284*, 112051. [\[CrossRef\]](#)
129. Najafzadeh, M.; Niazmardi, S. A novel multiple-kernel support vector regression algorithm for estimation of water quality parameters. *Nat. Resour. Res.* **2021**, *30*, 3761–3775. [\[CrossRef\]](#)
130. Sillberg, C.V.; Kullavanijaya, P.; Chavalparit, O. Water Quality Classification by Integration of Attribute-Realization and Support Vector Machine for the Chao Phraya River. *J. Ecol. Eng.* **2021**, *22*, 70–86. [\[CrossRef\]](#) [\[PubMed\]](#)
131. Arias-Rodriguez, L.F.; Duan, Z.; Díaz-Torres, J.d.J.; Basilio Hazas, M.; Huang, J.; Kumar, B.U.; Tuo, Y.; Disse, M. Integration of Remote Sensing and Mexican Water Quality Monitoring System Using an Extreme Learning Machine. *Sensors* **2021**, *21*, 4118. [\[CrossRef\]](#) [\[PubMed\]](#)
132. Batista, L.V. Turbidity classification of the Paraopeba River using machine learning and Sentinel-2 images. *IEEE Lat. Am. Trans.* **2022**, *20*, 799–805. [\[CrossRef\]](#)
133. Jamshidzadeh, Z.; Ehteram, M.; Shabanian, H. Bidirectional Long Short-Term Memory (BiLSTM) - Support Vector Machine: A new machine learning model for predicting water quality parameters. *Ain Shams Eng. J.* **2024**, *15*, 102510. [\[CrossRef\]](#)
134. Xi, Z.; Xue, Y. A Comparative Study of SVMs Model Optimized by Machine Learning Methods in Water Quality Assessment of Dongting Lake. In Proceedings of the 2023 4th International Conference on Information Science, Parallel and Distributed Systems (ISPDS), Guangzhou, China, 14–16 July 2023; IEEE: New York, NY, USA, 2023; pp. 608–611.
135. Dehkordi, A.T.; Zoj, M.J.V.; Mehran, A.; Jafari, M.; Chegoonian, A.M. Fuzzy similarity analysis of effective training samples to improve machine learning estimations of water quality parameters using Sentinel-2 remote sensing data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 5121–5136. [\[CrossRef\]](#)
136. Maritz, J.S. *Empirical Bayes Methods with Applications*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2018.
137. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
138. Izquierdo-Verdiguier, E.; Zurita-Milla, R. An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *88*, 102051. [\[CrossRef\]](#)
139. Liu, H.; Yu, T.; Hu, B.; Hou, X.; Zhang, Z.; Liu, X.; Liu, J.; Wang, X.; Zhong, J.; Tan, Z.; et al. UAV-Borne Hyperspectral Imaging Remote Sensing System Based on Acousto-Optic Tunable Filter for Water Quality Monitoring. *Remote Sens.* **2021**, *13*, 4069. [\[CrossRef\]](#)
140. Wang, W.; Chen, J.; Fang, L.; A, Y.; Ren, S.; Men, J.; Wang, G. Remote sensing retrieval and driving analysis of phytoplankton density in the large storage freshwater lake: A study based on random forest and Landsat-8 OLI. *J. Contam. Hydrol.* **2024**, *261*, 104304. [\[CrossRef\]](#)
141. Alnahit, A.O.; Mishra, A.K.; Khan, A.A. Stream water quality prediction using boosted regression tree and random forest models. *Stoch. Environ. Res. Risk Assess.* **2022**, *36*, 2661–2680. [\[CrossRef\]](#)
142. Mishra, A.; Ohri, A.; Singh, P.K.; Gaur, S.; Bhattacharjee, R. Water quality hotspot identification using a remote sensing and machine learning approach: A case study of the River Ganga near Varanasi. *Adv. Space Res.* **2024**, *74*, 5604–5618. [\[CrossRef\]](#)
143. Ghasemi, H.; Dehkordi, A.T.; Jafari, M.; Zoj, M.J.V. Coastal Water Quality Retrieval Based on Random Forest Coupled with Whale Optimization Algorithm using Sentinel-2 Data from Google Earth Engine. In Proceedings of the 2023 9th International Conference on Signal Processing and Intelligent Systems (ICSPIS), Bali, Indonesia, 14–15 December 2023; IEEE: New York, NY, USA, 2023; pp. 1–5.
144. Sipper, M.; Moore, J.H. AddGBoost: A gradient boosting-style algorithm based on strong learners. *Mach. Learn. Appl.* **2022**, *7*, 100243. [\[CrossRef\]](#)
145. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [\[CrossRef\]](#)
146. Rätsch, G.; Onoda, T.; Müller, K.R. Soft margins for AdaBoost. *Mach. Learn.* **2001**, *42*, 287–320. [\[CrossRef\]](#)
147. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30.
148. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
149. Xin, L.; Mou, T. Research on the application of multimodal-based machine learning algorithms to water quality classification. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 9555790. [\[CrossRef\]](#)

150. Duan, P.; Zhang, F.; Liu, C.; Tan, M.L.; Shi, J.; Wang, W.; Cai, Y.; Kung, H.T.; Yang, S. High-resolution planetscope imagery and machine learning for estimating suspended particulate matter in the Ebinur Lake, Xinjiang, China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *16*, 1019–1032. [[CrossRef](#)]
151. Garabaghi, F.H.; Benzer, S.; Benzer, R. Performance evaluation of machine learning models with ensemble learning approach in classification of water quality indices based on different subset of features. *Res. Sq.* **2022**, preprints. [[CrossRef](#)]
152. Shams, M.Y.; Elshewey, A.M.; El-kenawy, E.S.M.; Ibrahim, A.; Talaat, F.M.; Tarek, Z. Water quality prediction using machine learning models based on grid search method. *Multimed. Tools Appl.* **2024**, *83*, 35307–35334. [[CrossRef](#)]
153. Baek, S.S.; Pyo, J.; Chun, J.A. Prediction of water level and water quality using a CNN-LSTM combined deep learning approach. *Water* **2020**, *12*, 3399. [[CrossRef](#)]
154. Jiang, Y.; Li, C.; Sun, L.; Guo, D.; Zhang, Y.; Wang, W. A deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks. *J. Clean. Prod.* **2021**, *318*, 128533. [[CrossRef](#)]
155. Sha, J.; Li, X.; Zhang, M.; Wang, Z.L. Comparison of forecasting models for real-time monitoring of water quality parameters based on hybrid deep learning neural networks. *Water* **2021**, *13*, 1547. [[CrossRef](#)]
156. Marcus, G. Deep Learning: A Critical Appraisal. *arXiv* **2018**, arXiv:1801.00631.
157. Zamani, M.G.; Nikoo, M.R.; Jahanshahi, S.; Barzegar, R.; Meydani, A. Forecasting water quality variable using deep learning and weighted averaging ensemble models. *Environ. Sci. Pollut. Res.* **2023**, *30*, 124316–124340. [[CrossRef](#)]
158. Liu, P.; Wang, J.; Sangaiah, A.K.; Xie, Y.; Yin, X. Analysis and prediction of water quality using LSTM deep neural networks in IoT environment. *Sustainability* **2019**, *11*, 2058. [[CrossRef](#)]
159. Hu, Z.; Zhang, Y.; Zhao, Y.; Xie, M.; Zhong, J.; Tu, Z.; Liu, J. A water quality prediction method based on the deep LSTM network considering correlation in smart mariculture. *Sensors* **2019**, *19*, 1420. [[CrossRef](#)]
160. Barzegar, R.; Aalami, M.T.; Adamowski, J. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 415–433. [[CrossRef](#)]
161. Khullar, S.; Singh, N. Water quality assessment of a river using deep learning Bi-LSTM methodology: Forecasting and validation. *Environ. Sci. Pollut. Res.* **2022**, *29*, 12875–12889. [[CrossRef](#)]
162. Bangira, T.; Matongera, T.N.; Mabhaudhi, T.; Mutanga, O. Remote sensing-based water quality monitoring in African reservoirs, potential and limitations of sensors and algorithms: A systematic review. *Phys. Chem. Earth, Parts A/B/C* **2023**, *134*, 103536. [[CrossRef](#)]
163. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine learning in geosciences and remote sensing. *Geosci. Front.* **2016**, *7*, 3–10. [[CrossRef](#)]
164. Ball, J.E.; Anderson, D.T.; Chan, C.S. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [[CrossRef](#)]
165. Saha, A.; Pal, S.C. Application of machine learning and emerging remote sensing techniques in hydrology: A state-of-the-art review and current research trends. *J. Hydrol.* **2024**, *632*, 130907. [[CrossRef](#)]
166. Arabi, B.; Salama, M.S.; Pitarch, J.; Verhoef, W. Integration of in-situ and multi-sensor satellite observations for long-term water quality monitoring in coastal areas. *Remote Sens. Environ.* **2020**, *239*, 111632. [[CrossRef](#)]
167. Coffey, M.M.; Nezhin, N.P.; Bartlett, N.; Pasakarnis, T.; Lewis, T.N.; DiGiacomo, P.M. Satellite imagery as a management tool for monitoring water clarity across freshwater ponds on Cape Cod, Massachusetts. *J. Environ. Manag.* **2024**, *355*, 120334. [[CrossRef](#)] [[PubMed](#)]
168. Topp, S.N.; Pavelsky, T.M.; Jensen, D.; Simard, M.; Ross, M.R. Research trends in the use of remote sensing for inland water quality science: Moving towards multidisciplinary applications. *Water* **2020**, *12*, 169. [[CrossRef](#)]
169. Shi, K.; Han, J.C.; Wang, P. Near real-time retrieval of lake surface water temperature using Himawari-8 satellite imagery and machine learning techniques: A case study in the Yangtze River Basin. *Front. Environ. Sci.* **2024**, *11*, 1335725. [[CrossRef](#)]
170. Kamusoko, C.; Kamusoko, C. Pre-processing. In *Remote Sensing Image Classification in R*; Springer: Singapore, 2019; pp. 25–66.
171. Chen, Z.; Zhao, S. Automatic monitoring of surface water dynamics using Sentinel-1 and Sentinel-2 data with Google Earth Engine. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *113*, 103010. [[CrossRef](#)]
172. Mashala, M.J.; Dube, T.; Mudereri, B.T.; Ayisi, K.K.; Ramudzuli, M.R. A systematic review on advancements in remote sensing for assessing and monitoring land use and land cover changes impacts on surface water resources in semi-arid tropical environments. *Remote Sens.* **2023**, *15*, 3926. [[CrossRef](#)]
173. Adjovu, G.E.; Stephen, H.; James, D.; Ahmad, S. Overview of the application of remote sensing in effective monitoring of water quality parameters. *Remote Sens.* **2023**, *15*, 1938. [[CrossRef](#)]
174. Zhu, M.; Wang, J.; Yang, X.; Zhang, Y.; Zhang, L.; Ren, H.; Wu, B.; Ye, L. A review of the application of machine learning in water quality evaluation. *Eco-Environ. Health* **2022**, *1*, 107–116. [[CrossRef](#)] [[PubMed](#)]
175. Singh, M.K.; Mohan, S.; Kumar, B. Fusion of hyperspectral and LiDAR data using sparse stacked autoencoder for land cover classification with 3D-2D convolutional neural network. *J. Appl. Remote Sens.* **2022**, *16*, 034523. [[CrossRef](#)]

176. del Castillo, A.F.; Garibay, M.V.; Díaz-Vázquez, D.; Yebra-Montes, C.; Brown, L.E.; Johnson, A.; Garcia-Gonzalez, A.; Gradilla-Hernández, M.S. Improving river water quality prediction with hybrid machine learning and temporal analysis. *Ecol. Inform.* **2024**, *82*, 102655. [[CrossRef](#)]
177. Xu, J.; Mo, Y.; Zhu, S.; Wu, J.; Jin, G.; Wang, Y.G.; Ji, Q.; Li, L. Assessing and predicting water quality index with key water parameters by machine learning models in coastal cities, China. *Heliyon* **2024**, *10*, e33695. [[CrossRef](#)]
178. Elachi, C.; Van Zyl, J.J. *Introduction to the Physics and Techniques of Remote Sensing*; John Wiley & Sons: Hoboken, NJ, USA, 2021.
179. Chander, G.; Markham, B.L.; Helder, D.L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens. Environ.* **2009**, *113*, 893–903. [[CrossRef](#)]
180. Wang, D.; Tang, B.-H.; Li, Z.-H. Evaluation of five atmospheric correction algorithms for multispectral remote sensing data over plateau lake. *Ecol. Inform.* **2024**, *82*, 1574–9541 [[CrossRef](#)]
181. Kirschke, S.; Avellán, T.; Bärlund, I.; Bogardi, J.J.; Carvalho, L.; Chapman, D.; Dickens, C.W.; Irvine, K.; Lee, S.; Mehner, T.; et al. Capacity challenges in water quality monitoring: Understanding the role of human development. *Environ. Monit. Assess.* **2020**, *192*, 298. [[CrossRef](#)] [[PubMed](#)]
182. Sánchez, P.J.B.; Papaelias, M.; Márquez, F.P.G. Autonomous underwater vehicles: Instrumentation and measurements. *IEEE Instrum. Meas. Mag.* **2020**, *23*, 105–114. [[CrossRef](#)]
183. Hasan, F.; Medley, P.; Drake, J.; Chen, G. Advancing Hydrology through Machine Learning: Insights, Challenges, and Future Directions Using the CAMELS, Caravan, GRDC, CHIRPS, PERSIANN, NLDAS, GLDAS, and GRACE Datasets. *Water* **2024**, *16*, 1904. [[CrossRef](#)]
184. Wanganeo, A.; Wanganeo, R.R.N.; Srivastava, R.K.N.; Kumar, P. Variations in physico-chemical characteristics of water bodies placed at different geographical coordinates in Antarctica. *Polar Sci.* **2018**, *18*, 48–56. [[CrossRef](#)]
185. Sarker, I.H. Machine learning: Algorithms, real-world applications and research directions. *SN Comput. Sci.* **2021**, *2*, 160. [[CrossRef](#)]
186. Akhtar, N.; Syakir Ishak, M.I.; Bhawani, S.A.; Umar, K. Various natural and anthropogenic factors responsible for water quality degradation: A review. *Water* **2021**, *13*, 2660. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.