Machine Learning Applications in Water Quality Monitoring of Nansi Lake Basin: A Survey

www.surveyx.cn

Abstract

This survey paper explores the transformative role of machine learning in monitoring and assessing water quality within the Nansi Lake Basin, emphasizing its potential to enhance environmental management through data-driven insights. The integration of advanced algorithms facilitates the identification and classification of pollutants, addressing the urgent need for improved monitoring frameworks. Machine learning techniques significantly enhance the precision and efficiency of water quality assessments, offering sophisticated methodologies for predicting and managing environmental changes. The survey identifies specific algorithms effective in tackling complex environmental issues, underscoring the importance of selecting appropriate techniques based on contextual requirements. Challenges related to data availability, model accuracy, and computational demands are discussed, highlighting the need for robust data integration and innovative computational strategies. The paper also examines the integration of machine learning with traditional monitoring methods, enhancing the accuracy and efficiency of environmental assessments. Future directions include advancements in algorithm development, the integration of emerging technologies such as IoT and blockchain, and the enhancement of predictive modeling capabilities. Ethical and sustainable practices are emphasized to ensure that technological advancements align with societal values and ecological sustainability. The findings underscore the critical role of continued research and collaboration in advancing environmental monitoring, fostering interdisciplinary partnerships to unlock new opportunities for sustainable water resource management.

1 Introduction

1.1 Importance of Water Quality Monitoring in Nansi Lake Basin

Water quality monitoring in the Nansi Lake Basin is essential due to significant ecological and economic challenges stemming from industrial and agricultural pollution [1]. The region faces severe environmental degradation from various pollutants, necessitating stringent monitoring to protect human health and biodiversity. Contaminants, including priority substances and emerging pollutants, pose serious threats to aquatic ecosystems and public health [2].

Moreover, the ecological deterioration exacerbated by extensive underground coal mining underscores the urgent need for robust monitoring systems to safeguard the basin's environmental integrity [3]. The accumulation of mercury pollution from atmospheric deposition and human activities complicates the environmental landscape, highlighting the demand for advanced monitoring techniques [4].

Integrating responsible AI practices into environmental monitoring is crucial for ensuring accurate and sustainable water quality assessments, thereby supporting regulatory policies and conservation strategies. Furthermore, aligning water quality monitoring with sustainability in engineering education emphasizes the broader socio-economic implications of maintaining healthy water resources [5].

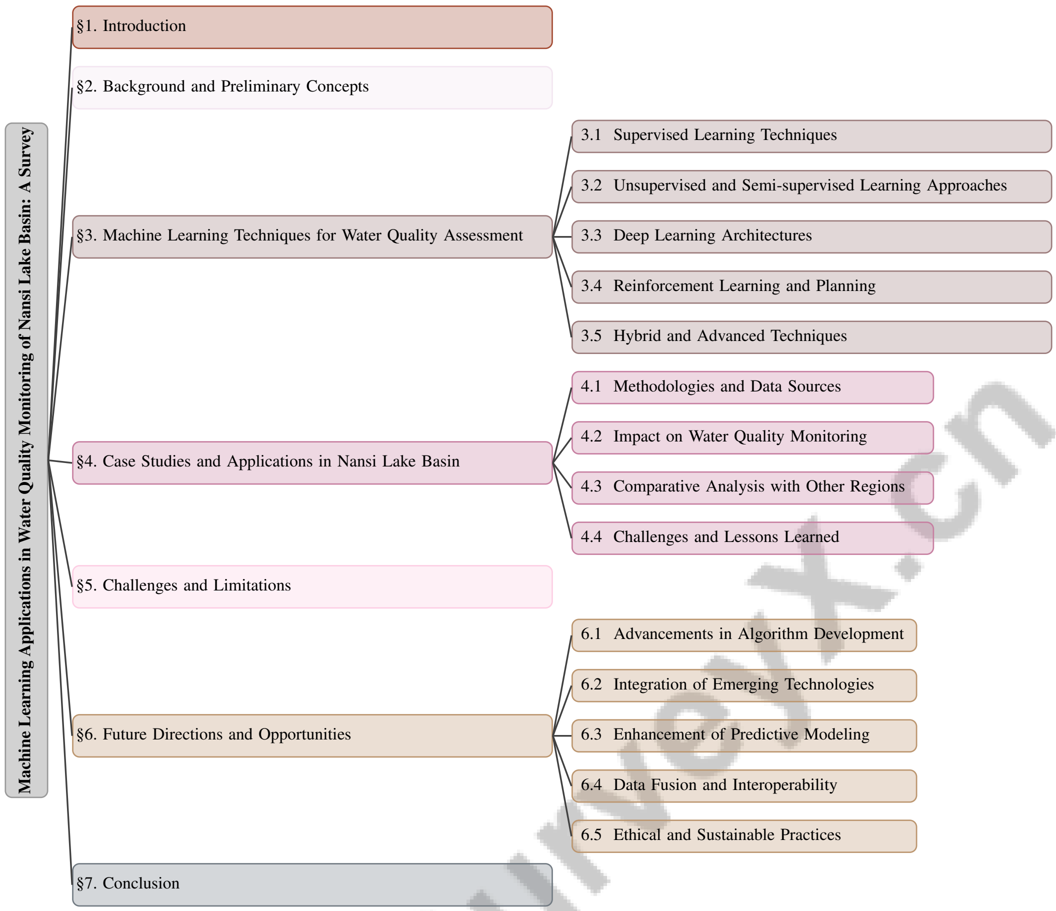

Figure 1: chapter structure

1.2 Role of Machine Learning in Enhancing Monitoring Efforts

Machine learning significantly enhances the efficiency and accuracy of water quality monitoring in the Nansi Lake Basin by employing advanced methodologies that overcome the limitations of traditional techniques. Traditional methods often struggle to detect contaminants in small water bodies; however, machine learning algorithms improve the identification and quantification of pollutants [6]. Neural networks facilitate the automatic identification of contaminants, streamlining monitoring processes and enhancing accuracy [1].

Deep learning models further expedite the detection and classification of floating objects in aquatic environments, improving monitoring speed and precision [7]. This capability is vital for timely interventions in water quality management. Additionally, machine learning algorithms that optimize sensor task scheduling contribute to efficient resource management, ensuring that monitoring systems operate with minimal energy consumption while maintaining high accuracy.

Advanced remote sensing technologies combined with machine learning yield high-resolution data that enhance environmental monitoring accuracy [1]. These technologies improve the monitoring of organic pollutants regulated by the European Union [2]. High-resolution sampling techniques further refine measurements of contaminants like labile mercury and methylmercury [4].

Machine learning's adaptability to various environmental phenomena allows continuous improvement in water quality assessments. The integration of fog computing principles in data processing exemplifies how machine learning can optimize data handling in precision agriculture, a concept applicable to water quality monitoring [8]. The exploration of major machine learning concepts in health and biomedical fields also highlights the potential for cross-disciplinary applications, enhancing the understanding and implementation of these techniques in environmental contexts [9].

1.3 Objectives of the Survey Paper

This survey aims to explore the integration of machine learning techniques for water quality assessment in the Nansi Lake Basin, with the goal of enhancing environmental monitoring and management practices. It addresses water quality improvement in the Eastern Region of Nansi Lake by analyzing trends and influencing factors related to inter-basin water diversion projects [10]. Key topics include drinking water quality, contaminants, and their health effects, emphasizing the need for improved monitoring frameworks to mitigate health risks associated with water pollutants.

A significant objective is to examine intelligent systems for monitoring aquatic environments, assessing the effectiveness of machine learning techniques in detecting anomalies, and integrating social media data into hydroinformatics to fill gaps in traditional water management data sources. The survey also investigates the Water Quality Enhanced Index (WQEI) as a machine learning technique to improve the detection of water quality parameters such as turbidity and salinity [11].

Additionally, the survey provides a comparative analysis of various deep learning models, including Artificial Neural Networks (ANN), Temporal Convolutional Networks (TCN), Long Short-Term Memory (LSTM), and Multi-Layer Perceptron (MLP), for predicting the Water Quality Index (WQI) to advance environmental monitoring and management practices [12]. It also explores innovative platforms like iEnvironment to enhance access to surface water data and models, supporting interdisciplinary research [13].

The survey systematically defines the intersection of AI and Earth observation, focusing on responsible AI practices to tackle global challenges in environmental monitoring [14]. By investigating temporal changes in water area, land use patterns, and cyanobacteria distribution in Nansi Lake, it aims to provide insights into the region's ecological dynamics [1]. Finally, the survey explores how machine learning techniques can enhance ecological data analysis for better population estimates and conservation strategies [15].

1.4 Structure of the Survey

This survey paper is organized into seven sections, each elucidating different aspects of machine learning applications in water quality monitoring within the Nansi Lake Basin. The introductory section establishes the context by emphasizing the importance of water quality monitoring and the role of machine learning in enhancing these efforts while outlining the survey's objectives, including advanced methodologies for environmental monitoring and management.

The second section, "Background and Preliminary Concepts," offers an overview of the Nansi Lake Basin's geographical and environmental significance, along with definitions of key concepts such as water pollutants and environmental monitoring. It introduces machine learning and predictive modeling, underlining their relevance to environmental monitoring.

In the third section, "Machine Learning Techniques for Water Quality Assessment," the survey examines various machine learning techniques applicable to water quality assessment, including supervised, unsupervised, and reinforcement learning. This section highlights specific algorithms used in predictive modeling of water pollutants and their effectiveness in environmental monitoring.

The fourth section, "Case Studies and Applications in Nansi Lake Basin," reviews existing studies and applications of machine learning in the Nansi Lake Basin, discussing methodologies, data sources, and outcomes while emphasizing machine learning's impact on water quality monitoring and management.

The fifth section, "Challenges and Limitations," identifies challenges and limitations associated with using machine learning for water quality monitoring in the Nansi Lake Basin, addressing issues such as data availability, model accuracy, and the integration of machine learning with traditional monitoring methods.

The sixth section, "Future Directions and Opportunities," explores potential future research directions and applications of machine learning in water quality monitoring, discussing opportunities for improving predictive modeling, enhancing data collection methods, and integrating new technologies for better water resource management.

Finally, the conclusion summarizes the key findings of the survey, highlighting machine learning's role in advancing water quality monitoring in the Nansi Lake Basin. This underscores the critical need for ongoing research and collaborative efforts to effectively tackle current challenges in environmental monitoring and resource management while capitalizing on emerging opportunities presented by advancements in social media, artificial intelligence, and machine learning technologies [16, 5, 14].The following sections are organized as shown in Figure 1.

2 Background and Preliminary Concepts

2.1 Geographical and Environmental Significance of Nansi Lake Basin

The Nansi Lake Basin, located in Shandong Province, China, is a crucial ecological and economic region, known for its biodiversity and support of agriculture and fishing. However, industrial and agricultural pollution, compounded by underground coal mining, poses significant threats to its ecological health [1, 3]. The spatiotemporal variations in water quality of the basin's tributaries, driven by economic and hydrological factors, necessitate advanced autonomous monitoring systems to manage these complexities [10, 17]. Pollution in rural areas further underscores the basin's environmental challenges [18]. High-resolution sediment sampling is vital for assessing pollution impacts, aiding in the development of strategies to mitigate anthropogenic effects on water quality and ecosystem health [4].

2.2 Key Concepts in Water Quality and Environmental Monitoring

Monitoring water quality and environmental health is essential for maintaining aquatic ecosystems. Identifying contaminants is critical for waste management and ecological restoration, with pollutants like Priority Substances and Contaminants of Emerging Concern posing challenges due to complex analyses and lack of standardized risk assessment protocols [6, 2]. Techniques like Gaussian Processes enhance spatial data analysis, offering insights into water quality variability [19]. In the Nansi Lake Basin, remote sensing reveals ecological changes, especially in coal mining areas [3]. Machine learning tools are increasingly integrated into monitoring frameworks, improving spatial relationship modeling and data source assessment [20]. These advancements are reflected in educational curricula, preparing future professionals for industry challenges [5]. Detailed datasets, such as those measuring mercury concentrations, are crucial for constructing comprehensive pollutant profiles [4], supporting robust monitoring systems to track water quality changes over time.

2.3 Introduction to Machine Learning and Predictive Modeling

Machine learning and predictive modeling are pivotal in environmental monitoring, offering advanced tools for analyzing complex datasets and forecasting changes. These methods address high-dimensional function estimation challenges, with neural networks providing fast alternatives to traditional covariance parameter estimation [19]. Convolutional neural networks improve predictions of water quality parameters like turbidity and pH using satellite data [1]. Non-stationary Gaussian Processes enhance environmental assessments by considering spatial-temporal interactions. The Deep Multi-source Spatial Prediction method exemplifies machine learning's potential, integrating diverse data sources for environmental monitoring [20]. Fog computing data aggregation optimizes data processing, reducing latency for real-time applications [8]. The use of deep learning models in mobile sensor systems highlights the importance of advanced decision-making in environmental management [21].

2.4 Integration of Geospatial Data and AI in Environmental Monitoring

Integrating geospatial data and AI in environmental monitoring enhances data collection, analysis, and interpretation. IoT systems enable real-time data collection across environmental dimensions, fostering responsive monitoring frameworks [22]. A comprehensive review of GeoAI research outlines AI's diverse applications in geospatial contexts, improving monitoring accuracy and efficiency [23]. The Confidence Region Construction for Exceedance Locations method combines geostatistics with hypothesis testing, enhancing monitoring precision by accurately identifying exceedance areas [24]. This integration of AI, IoT, and statistical methodologies offers significant opportunities for advancing environmental monitoring, addressing global challenges like climate change and disaster management. Responsible technology application requires ethical consideration, privacy protection, and bias mitigation, ensuring GeoAI benefits are maximized while maintaining scientific integrity and social good [23, 14]. By leveraging these technologies, scientists and policymakers can gain deeper ecological insights, leading to more effective management and conservation strategies.

In recent years, the application of machine learning techniques has revolutionized water quality assessment, providing innovative solutions for monitoring and management. Figure 2 illustrates the hierarchical structure of these techniques, categorizing them into supervised, unsupervised, semi-supervised, deep learning, reinforcement learning, and hybrid approaches. Each category is delineated with specific applications, models, and innovations that significantly enhance predictive capabilities and environmental management strategies. This structured overview not only clarifies the landscape of machine learning in this field but also emphasizes the importance of selecting appropriate methodologies to address diverse water quality challenges effectively.

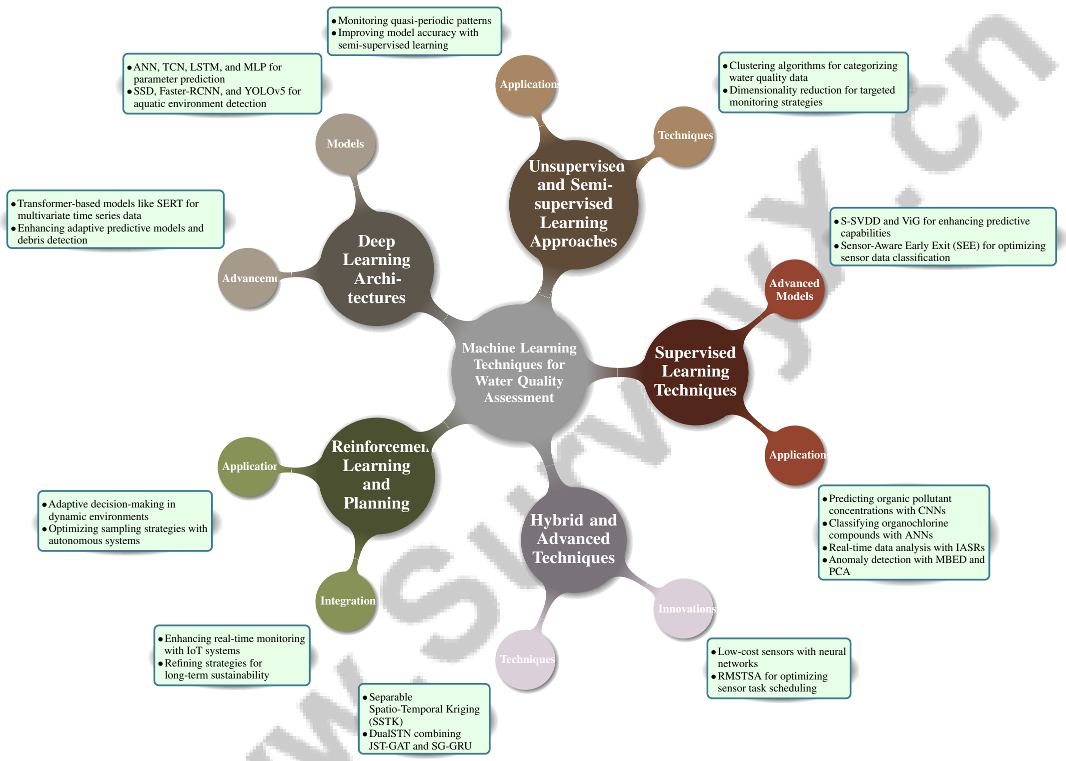

Figure 2: This figure illustrates the hierarchical structure of machine learning techniques applied to water quality assessment, categorized into supervised, unsupervised, semi-supervised, deep learning, reinforcement learning, and hybrid techniques. Each category details specific applications, models, and innovations that enhance monitoring, predictive capabilities, and environmental management.

3 Machine Learning Techniques for Water Quality Assessment

3.1 Supervised Learning Techniques

Supervised learning techniques are pivotal in water quality assessment, enabling precise predictions of water quality parameters through the use of labeled datasets. These methods train models to discern patterns, significantly enhancing anomaly detection and monitoring effectiveness in environmental contexts [30, 31]. Convolutional Neural Networks (CNNs) effectively predict organic pollutant concentrations, as demonstrated by their use in analyzing Surface-Enhanced Raman Scattering (SERS) data [25], while Artificial Neural Networks (ANNs) classify organochlorine compounds using molecular descriptors [26].

As illustrated in Figure 3, the application of supervised learning techniques in environmental monitoring encompasses various aspects, including water quality assessment, ecological data analysis, and intelligent systems. This figure highlights key methodologies and their contributions to environmental

AI-generated, for reference only.

Method Name	Application Areas	Model Types	Functional Capabilities
MLC-SERS[25]	Environmental Monitoring	Convolutional Neural Networks	Concentration Predictions
FMDA[26]	Environmental Monitoring	Artificial Neural Networks	Classification Accuracy
IASR[27]	Aquatic Environment Monitoring	-	Anomaly Detection
MBED[28]	Environmental Monitoring	Principal Component Analysis	Real-time Event Detection
MLWC[15]	Wildlife Conservation	Machine Learning Algorithms	Species Detection
SEE[29]	Activity Monitoring	Random Forest Classifiers	Energy Consumption

Table 1: This table presents a comprehensive overview of various supervised learning methods applied in environmental monitoring, detailing their specific application areas, model types, and functional capabilities. The methods highlighted include those used in water quality assessment, wildlife conservation, and activity monitoring, showcasing their diverse applications and contributions to the field.

sciences, providing a visual representation of the interconnectedness of these techniques. Furthermore, Table 1 provides a detailed summary of supervised learning methods used in environmental monitoring, illustrating their application areas, model types, and functional capabilities.

These techniques extend to Intelligent Autonomous Surface Robots (IASRs) for real-time environmental data analysis and anomaly detection [27]. The Model-Based Event Detection (MBED) method, utilizing Principal Component Analysis (PCA), identifies deviations in sensor data, bolstering water quality monitoring [28]. Advanced models like Subspace Support Vector Data Description (S-SVDD) and Vision Graph Neural Networks (ViG) further refine predictive capabilities, demonstrating the versatility of supervised learning across diverse environmental data [32, 33].

Supervised learning also plays a crucial role in ecological data analysis, aiding in species detection and population estimation [15]. The Sensor-Aware Early Exit (SEE) method optimizes sensor data classification, enhancing energy efficiency and timely decision-making [29]. Surveys of machine learning toolkits illustrate the application of these techniques in advancing environmental monitoring systems [5].

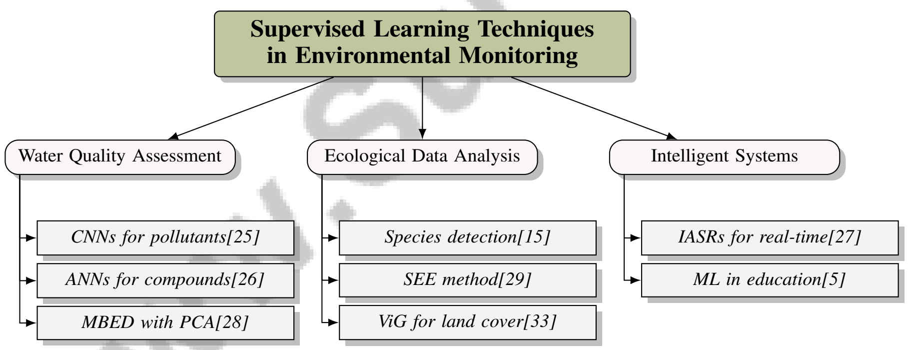

Figure 3: This figure illustrates the application of supervised learning techniques in environmental monitoring, focusing on water quality assessment, ecological data analysis, and intelligent systems. It highlights key methodologies and their contributions to environmental sciences.

3.2 Unsupervised and Semi-supervised Learning Approaches

Unsupervised and semi-supervised learning approaches are crucial for extracting patterns from water quality data, especially with limited labeled datasets. These methods, employing machine learning and image recognition technologies, enhance anomaly detection and trend identification in water quality metrics [16, 25, 34, 12, 7]. Clustering algorithms and dimensionality reduction techniques categorize water quality data, facilitating targeted monitoring strategies [35].

Unsupervised learning extends to monitoring quasi-periodic patterns, adaptable to environmental contexts for recognizing water quality changes [36]. Semi-supervised learning, leveraging both labeled and unlabeled data, improves model accuracy. The WasteMS dataset, utilizing multispectral imaging, exemplifies efficient data collection and waste differentiation [34]. Integrating unsupervised and semi-supervised learning enhances anomaly detection and supports robust monitoring frameworks, crucial for informed environmental management [37].

3.3 Deep Learning Architectures

Deep learning architectures, including feed-forward, convolutional, and recurrent networks, are essential for modeling complex water quality patterns [30]. Models like ANN, TCN, LSTM, and MLP contribute uniquely to parameter prediction and analysis [12]. Advanced algorithms such as SSD, Faster-RCNN, and YOLOv5 enhance detection capabilities for aquatic environments, crucial for real-time applications [7].

Transformer-based models like SERT, which process multivariate time series data, improve forecast accuracy and data integration [38]. Comparative studies demonstrate the superiority of deep learning in water quality pattern detection, enhancing adaptive predictive models and debris detection efficiency [39]. These advancements are pivotal for precise water quality assessments and effective resource management.

3.4 Reinforcement Learning and Planning

Reinforcement learning (RL) enhances water quality monitoring by enabling adaptive decision-making in dynamic environments. Autonomous systems with advanced sensors optimize sampling strategies, improving monitoring efficiency and accuracy [40, 41, 42]. RL excels in handling stochastic environmental systems, optimizing resource utilization for effective data capture [43].

Integrating RL with IoT systems enhances real-time monitoring, fostering proactive water quality management [22]. RL's iterative learning refines strategies for more precise assessments, crucial for long-term sustainability [44]. These capabilities highlight RL's potential in complex environmental management systems.

3.5 Hybrid and Advanced Techniques

Hybrid and advanced machine learning techniques integrate diverse methodologies to tackle complex environmental monitoring challenges. Separable Spatio-Temporal Kriging (SSTK) improves computational efficiency for large datasets [45]. DualSTN, combining JST-GAT and SG-GRU, models environmental data patterns, enhancing predictive accuracy [46].

Hybrid models integrating machine learning with ecological insights surpass traditional methods in efficiency and accuracy [15]. The S-SVDD method addresses dimensionality challenges, enhancing anomaly detection in water quality data [32]. Low-cost sensors combined with neural networks facilitate cost-effective monitoring systems [6].

The RMSTSA algorithm optimizes sensor task scheduling, maximizing information gain [21]. Introducing a fidelity score and geo-location-aware graph neural network improves data integration and reliability [20]. These innovations, including social media analytics and open science platforms like iEnvironment, enhance water quality assessments, supporting sustainable environmental practices [16, 13].

4 Case Studies and Applications in Nansi Lake Basin

4.1 Methodologies and Data Sources

Benchmark	Size	Domain	Task Format	Metric
WasteMS[34]	117	Waste Management	Semantic Segmentation	IoU, Parms
WQIB[12]	422	Water Quality Assessment	Classification	AUC
ViG[33]	590,326	Land Cover Classification	Multilabel Classification	F1-Score, Precision
HgMeHg-Benchmark[4]	12	Environmental Science	Risk Assessment	DGT-labile concentration, Risk Assessment Code

Table 2: This table provides a comprehensive overview of representative benchmarks used in environmental monitoring applications within the Nansi Lake Basin. It details the size, domain, task format, and evaluation metrics for each benchmark, highlighting the diverse methodologies employed in waste management, water quality assessment, land cover classification, and environmental science.

Diverse methodologies and data sources characterize machine learning applications in the Nansi Lake Basin, addressing the complexities of environmental monitoring. As illustrated in Figure 4, the hierarchical structure of these methodologies and data sources emphasizes the integration of machine learning techniques with environmental monitoring and remote sensing data. Table 2 presents a detailed summary of the benchmarks utilized in machine learning applications for environmental monitoring, illustrating the varied methodologies and data sources employed in the Nansi Lake Basin. Deep learning techniques are pivotal for land cover mapping and environmental parameter retrieval, providing a robust framework for analyzing remote sensing data and enhancing water quality assessments by capturing spatial and temporal variability [1]. The integration of high-quality sensors with AI algorithms, particularly in autonomous surface vehicles, effectively detects macro-plastics and measures water quality, demonstrating the synergy of advanced sensor technologies and machine learning [34]. Vision graph neural networks benefit from datasets like BigEarthNet, enriching training data for improved model performance [13].

In groundwater monitoring, techniques assessing radionuclide concentrations reveal long-term pollutant trends [12]. Innovations such as energy harvesting from underwater sound for machine learning inference and data transmission via backscatter communication enhance data collection sustainability in remote areas [40]. Simulation experiments assessing spatial and temporal variability are crucial for understanding aquatic dynamics [47]. Integrating social media data with traditional sources offers a comprehensive approach to water management, providing real-time insights to complement conventional monitoring data.

The Water Quality Enhanced Index (WQEI) evaluation using satellite data from commercial sources and LandSat8 over two years underscores remote sensing's effectiveness in monitoring water quality [26]. Machine learning applications in chemical data analysis, such as using Surface-Enhanced Raman Scattering (SERS) spectra for micropollutant detection, further illustrate practical applications [38]. Case studies in air quality detection, litter detection, and wildlife monitoring demonstrate machine learning's applicability to water quality monitoring in the basin [5].

These methodologies and data sources establish a robust framework for applying machine learning techniques in the Nansi Lake Basin, enhancing the ability to monitor, assess, and manage water quality. By integrating advanced satellite remote sensing data, deep learning models for debris detection, and multispectral imaging for waste segmentation, these approaches address pressing environmental challenges from industrial and agricultural activities [16, 10, 34, 1, 7].

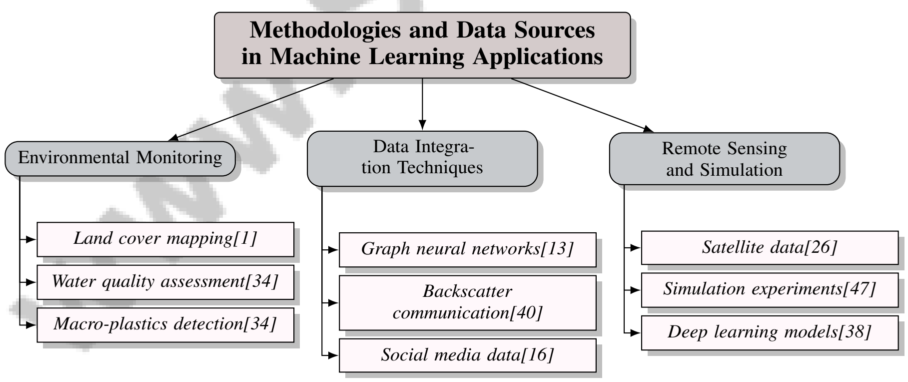

Figure 4: This figure illustrates the hierarchical structure of methodologies and data sources in machine learning applications for environmental monitoring, data integration, and remote sensing in the Nansi Lake Basin.

4.2 Impact on Water Quality Monitoring

Machine learning integration in water quality monitoring has markedly improved environmental health management in the Nansi Lake Basin. Enhanced predictive capabilities facilitate timely interventions, evident in the reduction of pollutant concentrations like Chemical Oxygen Demand (COD) and Total Phosphorus (TP) over the past two decades, reflecting a positive impact on water quality management [10]. Machine learning's transformation of marine science enhances resource management and ecological understanding, applicable to the Nansi Lake Basin's similar challenges

[48]. Real-time updates and adaptability to environmental changes improve data collection and model learning efficiency [41], crucial for responding to the basin's dynamic conditions.

Platforms like iEnvironment enhance water quality monitoring by improving data accessibility and fostering researcher collaboration [13]. Enhanced collaboration supports comprehensive environmental analyses and informed decision-making processes. Remote sensing's impact on water quality monitoring is evident through remote sensing indices, providing valuable data for assessing environmental changes and pollutant levels [1].

Machine learning techniques' integration in water quality monitoring enhances environmental assessments' accuracy and efficiency in the Nansi Lake Basin. This advancement facilitates a precise understanding of water quality dynamics and supports effective management and conservation strategies, particularly amid increasing challenges from industrial and agricultural activities. The application of machine learning, alongside satellite remote sensing data, allows comprehensive analyses of spatio-temporal changes in the lake's ecosystem, including monitoring cyanobacteria and land use patterns, significantly contributing to the protection and sustainable management of this vital freshwater resource [10, 1, 48, 7].

4.3 Comparative Analysis with Other Regions

Machine learning applications in the Nansi Lake Basin offer a comparative perspective with other regions, highlighting commonalities and unique advancements. In the Great Lakes and the Baltic Sea, advanced machine learning techniques address complex environmental challenges such as eutrophication and pollution from agricultural runoff. Techniques include predictive modeling of chemical properties and detecting harmful algal blooms, crucial for effective environmental monitoring. Moreover, machine learning enables sophisticated image recognition systems for identifying floating debris, enhancing water quality assessments and promoting sustainable management practices [34, 5, 48, 7]. These regions, like the Nansi Lake Basin, face significant ecological pressures necessitating advanced monitoring technologies.

In the Great Lakes, machine learning models predict harmful algal blooms by leveraging large datasets from satellite imagery and in-situ sensors, enhancing predictive accuracy [1]. Similarly, the Baltic Sea integrates machine learning with remote sensing data to monitor water quality parameters such as chlorophyll concentration and turbidity, providing critical insights into ecosystem health [48]. These applications parallel efforts in the Nansi Lake Basin, where remote sensing and machine learning assess water quality and predict pollutant level changes.

Distinctive differences in machine learning applications across regions stem from specific pollutants and environmental conditions, influenced by local ecological challenges, regulatory frameworks, and data availability from remote sensing technologies like hyperspectral and multispectral imaging. For instance, some areas prioritize detecting specific waste types using multispectral datasets, while others focus on broader environmental parameters such as land cover classification or atmospheric monitoring through advanced deep learning techniques. This tailored approach enhances monitoring efforts' effectiveness and addresses region-specific ecological concerns [34, 39, 35]. The Nansi Lake Basin particularly emphasizes managing industrial pollution and underground coal mining impacts, which are less prevalent in the Great Lakes and Baltic Sea regions, necessitating customized machine learning solutions for unique environmental challenges.

Moreover, integrating social media data into water quality monitoring frameworks is more pronounced in the Nansi Lake Basin, providing real-time insights that complement traditional data sources [16]. This innovative approach, less commonly adopted in other regions, offers significant potential for enhancing monitoring capabilities and engaging local communities in environmental management.

While methodologies for applying machine learning in water quality monitoring are globally similar, unique environmental conditions and challenges necessitate tailored approaches and innovative adaptations of these technologies to effectively address local needs and improve outcomes [16, 48, 35]. The Nansi Lake Basin's approach, characterized by integrating diverse data sources and focusing on specific pollutants, exemplifies machine learning's adaptability and potential in addressing global water quality issues.

4.4 Challenges and Lessons Learned

Implementing machine learning applications in the Nansi Lake Basin presents several challenges, offering valuable lessons for enhancing water quality monitoring and management. A significant challenge is data availability and quality, essential for training accurate and reliable machine learning models. The heterogeneity of data sources, including remote sensing, in-situ measurements, and social media, complicates data integration and standardization [1]. Addressing these issues requires robust data preprocessing techniques and frameworks to harmonize diverse datasets for effective model training and validation.

Another challenge involves the computational demands of machine learning models, particularly deep learning architectures, which require substantial resources for training and inference [12]. Limited access to high-performance computing infrastructure can constrain regions. To mitigate this, adopting cloud computing and distributed processing frameworks can enhance the scalability and efficiency of machine learning applications in environmental monitoring.

The complexity of environmental systems and the dynamic nature of water quality parameters further complicate predictive modeling efforts [15]. Machine learning models must adapt to temporal and spatial variations in environmental conditions, necessitating adaptive learning algorithms that continuously update and refine predictions based on new data.

Despite these challenges, lessons from machine learning applications in the Nansi Lake Basin highlight the value of integrating diverse data sources, including remote sensing and social media, to enhance the comprehensiveness and accuracy of water quality assessments [16]. This approach improves predictive capabilities and fosters greater community engagement and environmental awareness.

Additionally, developing hybrid and advanced machine learning techniques that combine ecological knowledge with algorithms has shown potential in overcoming traditional monitoring method limitations [15]. These innovative approaches enhance the precision and reliability of environmental assessments, contributing to more effective water quality management strategies.

The challenges and lessons learned from machine learning applications in the Nansi Lake Basin underscore the critical need for ongoing innovation and collaboration among researchers, policymakers, and local communities. This collaborative approach is essential for effectively addressing multifaceted environmental issues in the region, such as deteriorating water quality due to industrial and agricultural activities and climate change impacts. By leveraging advanced machine learning techniques and remote sensing data, stakeholders can gain deeper insights into water quality dynamics and ecosystem health, ultimately fostering more sustainable management practices and enhancing the resilience of local aquatic ecosystems [16, 10, 1, 39, 5].

5 Challenges and Limitations

5.1 Data Availability and Quality

The efficacy of machine learning in water quality monitoring is heavily contingent upon the availability and quality of data, which are often limited by several factors. A major issue is the dependency on extensive, high-quality labeled datasets, which are challenging to acquire due to the complex and varied nature of environmental data sources [5]. Furthermore, the lack of familiarity with machine learning tools among educators and students compounds these data-centric challenges.

The quality of input data is crucial for the accuracy of machine learning predictions; subpar data can significantly impair prediction outcomes [20]. This is particularly relevant in water quality monitoring, where diverse data integration, such as satellite imagery, is vital. However, atmospheric conditions can degrade the quality of satellite data, complicating its reliability [4]. Additionally, specific sampling techniques may not be universally applicable across different sediment types, limiting data collection and analysis [4].

Imputation methods for missing data can introduce biases, distorting datasets and hindering accurate forecasting. Thus, robust data preprocessing methods are critical for harmonizing diverse datasets for effective model training and validation. In IoT-based environmental monitoring, a significant challenge is the high energy consumption of low-power devices that often wait for complete sensor data before processing. Innovations like early exit classifiers allow processing to start with partial data, reducing energy usage by up to 50-60

Addressing data availability and quality challenges requires ongoing innovation in data collection, integration, and analysis techniques. Enhancing machine learning model robustness while effectively managing network congestion and energy consumption can significantly improve the reliability and accuracy of water quality assessments. This progress supports precise monitoring and predictive modeling of water quality indices, employing advanced techniques like Long Short-Term Memory (LSTM) networks and Random Forest regression, ultimately contributing to environmental management and conservation efforts [16, 37, 44, 12, 5].

5.2 Model Accuracy and Computational Challenges

Achieving high accuracy in machine learning models for water quality monitoring is challenged by the complex and variable nature of environmental data. A significant issue is the computational burden associated with training neural networks, which can impede real-time contaminant identification [6]. This burden is exacerbated by covariance parameter estimation demands, where existing methods struggle to efficiently manage large, heterogeneous spatial data [19].

The computational complexity of algorithms presents barriers to their application in real-time scenarios, particularly in large-scale environments where timely responses are crucial [21]. This complexity is further compounded by the need for accurate local data processing, which varies based on sensor accuracy and environmental conditions, impacting overall model reliability [8].

Moreover, the interpretability of machine learning models is a significant challenge, as the assumptions underlying statistical models can obscure insights and affect accuracy. This is particularly relevant in spatio-temporal data contexts, where proper covariance structure specification is essential for model fidelity. The computational demands of spatio-temporal kriging highlight the challenges of efficiently processing large datasets. While traditional methods can be cumbersome, advancements like the separable spatio-temporal kriging approach demonstrate that applying suitable separability assumptions can significantly reduce computational costs, allowing centralized spatial interpolation while facilitating decentralized forecasting at individual sensor locations. Additionally, alternative models, such as the SERT transformer-based model and the SST-ANN neural network, have been developed to manage missing data in multivariate spatio-temporal forecasting, further illustrating ongoing efforts to enhance computational efficiency in this area [45, 19, 49, 35, 38].

To effectively address these challenges, it is essential to develop innovative computational strategies and robust data integration techniques that leverage advanced machine learning algorithms and social media data, thereby enabling comprehensive environmental monitoring and improved management of water resources and disaster response efforts [16, 31].

As illustrated in Figure 5, the hierarchical structure of challenges and solutions in applying machine learning for environmental monitoring emphasizes key issues in model accuracy, computational demands, and innovative solutions. Enhancing the accuracy and efficiency of machine learning models can lead to more reliable water quality assessments, thereby contributing to effective environmental monitoring and management practices.

5.3 Integration with Traditional Monitoring Methods

Integrating machine learning with traditional water quality monitoring methods enhances the accuracy and efficiency of environmental assessments. By utilizing sophisticated algorithms capable of analyzing extensive datasets, machine learning improves predictive modeling for critical water quality indicators, such as the Water Quality Index (WQI), and facilitates the detection of harmful algal blooms and hypoxic conditions. This innovative approach streamlines data collection and analysis, supporting the development of more effective environmental monitoring strategies and contributing to better aquatic ecosystem management [48, 16, 34, 12, 5]. Traditional methods, which often rely on manual sampling and laboratory analysis, can be significantly enhanced by machine learning algorithms that offer automated, real-time data processing. By leveraging large datasets, machine learning techniques provide comprehensive insights into water quality dynamics, enabling precise identification and classification of pollutants.

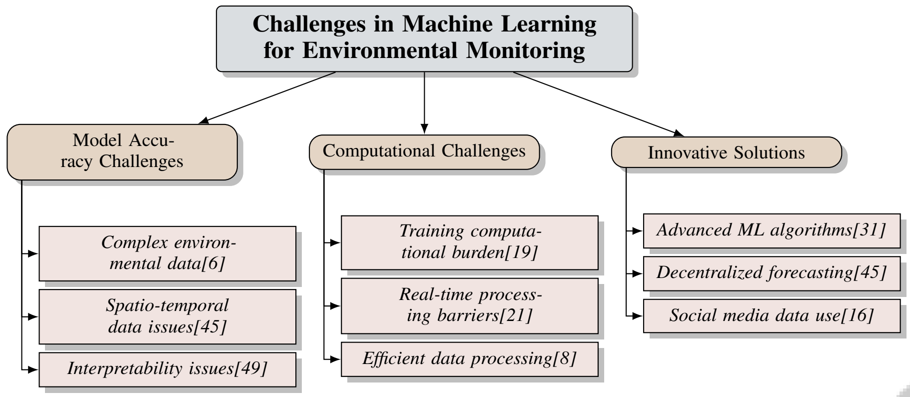

Figure 5: This figure illustrates the hierarchical structure of challenges and solutions in applying machine learning for environmental monitoring, highlighting key issues in model accuracy, computational demands, and innovative solutions.

A key advancement in this integration is the application of AI Intelligent Image Recognition methods, which utilize advanced algorithms to learn from extensive datasets, thereby improving the recognition and classification of various floating objects in aquatic environments [7]. This capability is crucial for enhancing the detection of pollutants and anomalies that traditional monitoring methods may overlook.

Machine learning models augment traditional water quality monitoring by providing advanced predictive analytics that can forecast shifts in water quality parameters, such as hypoxic conditions and harmful algal blooms, enabling proactive management and timely interventions in environmental monitoring efforts [44, 48, 12]. This predictive capability fosters proactive management strategies, reducing reliance on reactive measures characteristic of conventional methods. By integrating machine learning with existing monitoring frameworks, resource allocation can be optimized, ensuring efficient and effective monitoring efforts.

The synergy between machine learning and traditional methods facilitates the development of hybrid monitoring systems that leverage the strengths of both approaches. These systems can utilize real-time data from IoT devices and remote sensing technologies alongside traditional sampling techniques, providing a holistic view of water quality conditions. This comprehensive approach enhances environmental assessment accuracy through the integration of advanced technologies, such as satellite imagery and machine learning, while also facilitating informed decision-making processes by providing reliable data for monitoring critical environmental factors, including water quality and pollutant levels [37, 34, 5, 2].

The integration of machine learning with traditional water quality monitoring methods marks a transformative step in environmental management, enhancing predictive capabilities for critical issues, such as hypoxic conditions and harmful algal blooms, improving debris detection accuracy in aquatic environments, and facilitating efficient analysis of large datasets. This integration enables more informed decision-making in water resource conservation and management [48, 16, 34, 12, 7]. By harnessing machine learning capabilities, monitoring efforts can be made more precise and responsive, contributing to sustainable and effective water resource management.

5.4 Scalability and Operational Constraints

Implementing machine learning solutions for water quality monitoring in the Nansi Lake Basin encounters several scalability and operational constraints that must be addressed for effective and sustainable environmental management. A primary challenge is the adaptability of machine learning tools within platforms like iEnvironment, which must accommodate diverse research needs and applications [13]. This adaptability is essential for handling various datasets and analytical requirements that can vary significantly across different environmental contexts.

Scalability issues arise from the necessity to process large volumes of data generated by IoT devices and remote sensing technologies. Integrating these data sources demands robust computational frameworks capable of managing high data throughput while ensuring accuracy and efficiency. Semantic interoperability is another critical aspect, as it enables seamless integration and analysis of data from various sources [50]. However, achieving this interoperability presents challenges, particularly in formalizing unstructured indigenous knowledge, which is vital for understanding local environmental conditions and informing monitoring strategies.

Operational constraints also encompass the need for scalable infrastructure to efficiently deploy machine learning models over extensive geographical areas, facilitating real-time environmental monitoring and management through integrated data sources, such as satellite imagery and sensor networks. These integrations are essential for addressing complex environmental challenges and supporting applications in hydrology, soil science, and wildlife monitoring [50, 22, 30, 39, 5]. Ensuring that computational resources are available for real-time data processing and that communication networks can transmit data efficiently is crucial. Additionally, the scalability of machine learning solutions is often limited by the availability of high-quality training data, which is essential for developing accurate predictive models.

Addressing these scalability and operational constraints necessitates a concerted effort to develop flexible and interoperable machine learning frameworks that can adapt to varying environmental conditions and research needs. By enhancing the scalability and operational efficiency of water quality monitoring solutions, such as the Water Quality Enhanced Index (WQEI) model and the iEnvironment platform, the accuracy and effectiveness of monitoring efforts can be improved. This advancement enables the detection of contamination levels in small water bodies with weak reflectance patterns and facilitates access to big data resources and collaborative research, contributing to informed decision-making and sustainable environmental management practices across various sectors, including oil and gas operations [11, 13].

5.5 Ethical and Social Considerations

The ethical and social implications of employing machine learning in environmental monitoring are profound and multifaceted. As these technologies become increasingly integrated into environmental management practices, it is crucial to assess their potential impacts on fundamental rights and societal structures. Current studies often lack comprehensive evaluations of all fundamental rights that may be affected by AI systems, leading to gaps in understanding the full implications of these technologies [37]. This oversight can result in unintended consequences, such as privacy violations or biases in decision-making processes, disproportionately affecting marginalized communities.

The deployment of machine learning in environmental monitoring necessitates the adoption of responsible AI practices to mitigate these risks. Such practices should include ensuring transparency in algorithmic decision-making, maintaining data privacy, and promoting inclusivity in the development and application of AI technologies [14]. By prioritizing these ethical considerations, the trustworthiness and acceptance of machine learning applications in environmental contexts can be enhanced.

Furthermore, the social implications of machine learning extend to the potential displacement of traditional monitoring roles and the need for new skill sets among environmental professionals. This shift underscores the critical necessity for educational and training initiatives that empower individuals with the skills required to effectively engage with AI technologies, particularly in rapidly evolving fields like environmental monitoring and engineering, where machine learning applications are becoming integral to decision-making processes [37, 5, 9]. Additionally, fostering community engagement and participation in the development and implementation of machine learning solutions can ensure that these technologies align with local needs and values, enhancing their social acceptability and effectiveness.

6 Future Directions and Opportunities

6.1 Advancements in Algorithm Development

Enhancing water quality assessment through algorithm development requires refining existing models and exploring new computational methodologies. Future research should focus on robust algorithms capable of handling diverse datasets and improving predictive accuracy in various environmental contexts, particularly with high-dimensional data and real-time monitoring [9]. Key developments should include optimizing inference speed in models like DualSTN, which decouples long-term and short-term environmental data patterns [46]. Integrating ecological knowledge into machine learning models will further wildlife monitoring and conservation efforts [15].

Research should also focus on complex covariance models, uncertainty quantification, and optimizing neural network architectures to enhance performance in environmental monitoring [19]. Standardizing monitoring protocols and assessing ecological risks from emerging contaminants are crucial for reliable machine learning applications across diverse ecological settings [2]. Extending the Sensor-Aware Early Exit (SEE) approach to multiple classifiers and real-world applications will improve energy efficiency and accuracy [29]. Additionally, comprehensive curricula integrating machine learning across engineering disciplines will support algorithm development for water quality assessment [5].

As illustrated in Figure 6, advancements in algorithm development for environmental monitoring can be categorized into key techniques and applications, including data integration methods, machine learning models, and ecological applications. Future studies should refine environmental monitoring ontologies, expand datasets, and improve information dissemination, such as for drought forecasting [51]. Evaluating IoT-enabled distributed data processing systems in real-time decision-making and enhancing data quality are essential for effective environmental monitoring frameworks [8]. Promoting data accessibility and methodological innovation through publicly available datasets and exploring new analytes with machine learning techniques will significantly advance water quality assessment [9].

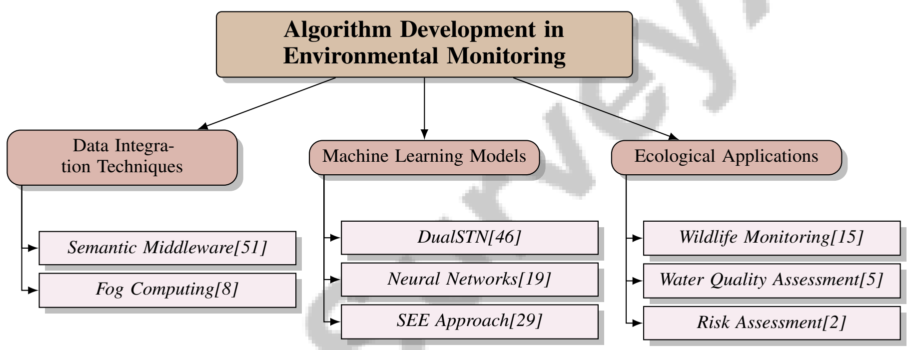

Figure 6: This figure illustrates the advancements in algorithm development for environmental monitoring, categorizing the key techniques and applications into data integration methods, machine learning models, and ecological applications.

6.2 Integration of Emerging Technologies

The integration of emerging technologies like the Internet of Things (IoT) and blockchain holds transformative potential for advancing water quality monitoring frameworks. IoT enables real-time data collection and analysis, enhancing the responsiveness and accuracy of environmental monitoring systems. Deploying IoT-enabled sensors allows continuous monitoring of parameters such as temperature, pH, and pollutant concentrations, facilitating timely interventions in water resource management [22].

Blockchain technology complements IoT by ensuring data security and transparency. Immutable records of water quality data bolster the trustworthiness of monitoring systems, crucial for regulatory compliance and environmental policies. The decentralized nature of blockchain fosters collaborative data sharing among stakeholders, promoting an integrated approach to water quality management. Platforms like iEnvironment enhance interdisciplinary partnerships and improve water resource management through collective insights [22, 16, 13].

The combination of IoT and blockchain significantly enhances the scalability and operational efficiency of water quality monitoring systems. IoT devices can be extensively deployed to monitor large geographical areas, while blockchain ensures data integrity and accessibility, essential for informed decision-making in hydrology and agriculture [16, 8, 18, 36, 22]. This integration not only improves environmental assessment precision but also supports informed decision-making by providing a comprehensive view of water quality conditions.

6.3 Enhancement of Predictive Modeling

Enhancing predictive modeling for water quality requires integrating advanced methodologies and comprehensive datasets to address aquatic environment complexities. Future research should focus on continuous modeling of environmental dynamics, refining models to incorporate spatial distance information and non-stationary models, thereby improving spatiotemporal interpolation tasks. This approach allows for sophisticated evaluations of water quality by integrating sensor-specific temporal correlation parameters, significantly enhancing predictive adaptability and precision [24, 46, 52, 12, 38].

Incorporating real-time data from IoT devices and remote sensing technologies is vital for enhancing pollution assessments and improving predictive model precision [1]. These technologies enable models to achieve higher accuracy and adaptability to varying environmental conditions, crucial for understanding health effects from low-level exposure to multiple contaminants, necessitating longitudinal studies and robust methodologies [53].

Research should also refine detection algorithms to improve robustness against environmental variations, broadening technology applicability in diverse aquatic settings [7]. Exploring additional data sources and improving model adaptability to varying environmental conditions are essential for advancing predictive modeling capabilities [20]. Expanding benchmarks to include other contaminants and testing applicability in various aquatic environments will provide a comprehensive understanding of pollutant dynamics [4].

Further, improving neural network robustness and exploring additional applications in diverse environmental monitoring scenarios will enhance predictive model versatility [6]. Developing quantum feature-empowered deep classification techniques presents promising avenues for advancing predictive modeling, improving water quality predictions [54].

Additionally, exploring reinforcement learning approaches for optimal sensor placement can optimize resource deployment, ensuring predictive models are informed by relevant data [21].

By focusing on advanced research directions such as transformer-based models like SERT for handling spatio-temporal sensor data with missing values, identifying exceedance regions through confidence regions in spatio-temporal processes, and applying spatial regression-based transfer learning to enhance prediction accuracy, predictive models can be significantly strengthened. These approaches address the complexities and uncertainties of dynamic environmental systems, improving our capacity to monitor and respond to critical issues like climate change, biodiversity loss, and pollution [38, 24, 49].

6.4 Data Fusion and Interoperability

Data fusion and interoperability are essential for enhancing machine learning applications in environmental monitoring, particularly for water quality assessment. These concepts involve integrating diverse datasets from multiple sources, enabling comprehensive analyses that improve the accuracy and reliability of predictive models. Integrating data from various sensors, remote sensing technologies, and IoT devices is crucial for constructing comprehensive environmental monitoring systems that effectively respond to rapidly changing conditions. This capability allows for fine-grained, real-time collection and analysis of environmental parameters, applicable in areas ranging from precision agriculture to urban flood monitoring. Advanced techniques such as distributed learning, UAV-assisted data relay, and real-time sensor management algorithms enhance adaptability and operational efficiency, improving decision-making and response strategies in dynamic environments [22, 18, 21, 8].

Interoperability is vital for seamless data integration and analysis from different systems, ensuring effective machine learning applications. Establishing common data standards and protocols facilitates information exchange between disparate systems, enhancing cross-disciplinary analyses and insights from complex datasets. This capability is particularly important in environmental monitoring, where data is collected from diverse sources, including surface water assessments and multispectral imaging, which utilize various formats and resolutions for evaluating contaminants, assessing chemical status, and enabling effective risk management and regulatory compliance [34, 13, 2].

Integrating data fusion techniques with machine learning models synthesizes information from heterogeneous sources, providing a more comprehensive understanding of environmental conditions.

By combining satellite imagery, ground-based sensors, and advanced monitoring technologies, researchers can significantly enhance the spatial and temporal resolution of water quality assessments. This approach improves contamination detection levels in water bodies, including those with weak reflectance signals, and allows for efficient monitoring of large areas, such as lakes and wetlands. For instance, the Water Quality Enhanced Index (WQEI) model has demonstrated accurate assessments of water turbidity using readily available satellite data, serving as a valuable tool for managing freshwater resources across various sectors, including oil and gas operations. Additionally, multispectral imaging technologies facilitate efficient identification of environmental issues, such as waste in lakeside areas, further emphasizing the potential of combining diverse data sources for comprehensive water quality monitoring and environmental protection [34, 1, 11]. This comprehensive approach not only enhances predictive model precision but also supports informed decision-making processes.

Furthermore, integrating Trustworthy AI assessments with fundamental rights considerations highlights the ethical and technical dimensions of data fusion and interoperability [37]. By ensuring data integration processes adhere to ethical standards and respect fundamental rights, more reliable and socially acceptable machine learning applications can be developed.

6.5 Ethical and Sustainable Practices

Incorporating ethical and sustainable practices within machine learning applications for environmental monitoring is essential to align technological advancements with societal values and ecological sustainability. These practices are particularly vital in deploying IoT systems, where standardization and accessibility are crucial for broader acceptance [22]. Establishing standardized protocols will enhance the integration of IoT technologies into environmental monitoring frameworks, maximizing their utility and impact.

Within GeoAI applications, there is a notable gap in research addressing ethical considerations and potential biases that can significantly affect machine learning model performance and fairness [23]. Addressing these gaps is critical for developing trustworthy AI systems that minimize bias and promote equitable outcomes in environmental assessments.

The utilization of social media data in water management underscores the importance of ethical and sustainable practices, particularly concerning data privacy and community engagement [16]. Fostering transparency and inclusivity in data collection processes can build trust among stakeholders, enhancing monitoring efforts' effectiveness.

Moreover, responsible AI practices in Earth observation emphasize the need for ethical and sustainable approaches in machine learning applications [14]. These practices involve not only technical considerations but also engaging local communities in data collection and decision-making processes, ensuring that AI technologies align with the needs and values of those directly affected by environmental monitoring activities [15].

7 Conclusion

Machine learning emerges as a pivotal force in advancing water quality monitoring in the Nansi Lake Basin, offering transformative capabilities in environmental management through enhanced data analysis. By improving the precision and efficiency of water quality assessments, these techniques provide robust methodologies for predicting and managing environmental changes. The deployment of advanced algorithms facilitates the identification and classification of hazardous substances, highlighting the critical need for enhanced monitoring frameworks to address environmental challenges effectively.

The selection of specific machine learning algorithms tailored to the unique challenges of the Nansi Lake Basin is crucial, as it ensures the optimization of predictive models across diverse ecological contexts. This strategic approach enhances the accuracy and reliability of environmental assessments, reinforcing the importance of context-specific applications.

Furthermore, the survey underscores the significance of continued research and interdisciplinary collaboration in the evolution of environmental monitoring. By leveraging emerging technologies and fostering cooperative efforts among stakeholders, there is potential to overcome existing challenges and explore innovative pathways for sustainable water resource management. Such collaborative

AI-generated, for reference only.

initiatives are essential for developing resilient and adaptive monitoring systems, contributing to the long-term ecological health and sustainability of the Nansi Lake Basin and similar environments.

References

[1] Xingkai Wang, Wenlong Yu, Rui Wang, Bin Li, and Xiangyang Cao. Analysis of spatio-temporal change characteristics of water environment in nansi lake based on satellite remote sensing monitoring. In *SHS Web of Conferences*, volume 178, page 02020. EDP Sciences, 2023.

[2] João CG Sousa, Ana R Ribeiro, Marta O Barbosa, M Fernando R Pereira, and Adrián MT Silva. A review on environmental monitoring of water organic pollutants identified by eu guidelines. *Journal of hazardous materials*, 344:146–162, 2018.

[3] Hu Liu, Yan Jiang, Rafal Misa, Junhai Gao, Mingyu Xia, Axel Preusse, Anton Sroka, and Yue Jiang. Ecological environment changes of mining areas around nansi lake with remote sensing monitoring. *Environmental Science and Pollution Research*, 28:44152–44164, 2021.

[4] Min Zhang, Cai Li, Xin Ma, Liyuan Yang, and Shiming Ding. Evaluating the mercury distribution and release risks in sediments using high-resolution technology in nansi lake, china. *Journal of Soils and Sediments*, 21(10):3466–3478, 2021.

[5] Andrew Schulz, Suzanne Stathatos, Cassandra Shriver, and Roxanne Moore. Utilizing online and open-source machine learning toolkits to leverage the future of sustainable engineering, 2023.

[6] Jose de Jesus Rubio, Jose Alberto Hernandez-Aguilar, Francisco Jacob Avila-Camacho, Juan Manuel Stein-Carrillo, and Adolfo Melendez-Ramirez. Sistema sensor para el monitoreo ambiental basado en redes neuronales, 2019.

[7] Jingyu Zhang, Ao Xiang, Yu Cheng, Qin Yang, and Liyang Wang. Research on detection of floating objects in river and lake based on ai intelligent image recognition, 2024.

[8] Grigore Stamatescu, Cristian Dragana, Iulia Stamatescu, Loretta Ichim, and Dan Popescu. Iot-enabled distributed data processing for precision agriculture, 2019.

[9] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.

[10] Yuhang Zhao, Shiqiang Wu, Jiangyu Dai, Xiufeng Wu, Yu Zhang, Rui Yang, Chenhui Wu, and Kedong Feng. Growth in the economic and hydraulic engineering capacities drives water quality improvement in the eastern region of nansi lake, china. *Frontiers in Environmental Science*, 10:931642, 2022.

[11] Satish Kumar, Rui Kou, Henry Hill, Jake Lempges, Eric Qian, and Vikram Jayaram. In-situ water quality monitoring in oil and gas operations, 2023.

[12] Muhammad Ismail, Farkhanda Abbas, Shahid Munir Shah, Mahmoud Aljawarneh, Lachhman Das Dhomeja, Fazila Abbas, Muhammad Shoaib, Abdulwahed Fahad Alrefaei, and Mohammed Fahad Albeshr. Performance evaluation of deep learning models for water quality index prediction: A comparative study of lstm, tcn, ann, and mlp, 2024.

[13] Paulo Alencar, Donald Cowan, and Doug Mulholland. The ienvironment platform: Developing an open science software platform for integrated environmental monitoring and modeling of surface water, 2018.

[14] Pedram Ghamisi, Weikang Yu, Andrea Marinoni, Caroline M. Gevaert, Claudio Persello, Sivasakthy Selvakumaran, Manuela Girotto, Benjamin P. Horton, Philippe Rufin, Patrick Hostert, Fabio Pacifici, and Peter M. Atkinson. Responsible ai for earth observation, 2024.

[15] Devis Tuia, Benjamin Kellenberger, Sara Beery, Blair R. Costelloe, Silvia Zuffi, Benjamin Risse, Alexander Mathis, Mackenzie W. Mathis, Frank van Langevelde, Tilo Burghardt, Roland Kays, Holger Klinck, Martin Wikelski, Iain D. Couzin, Grant van Horn, Margaret C. Crofoot, Charles V. Stewart, and Tanya Berger-Wolf. Seeing biodiversity: perspectives in machine learning for wildlife conservation, 2021.

[16] Yufeng Yu, Yuelong Zhu, Dingsheng Wan, Qun Zhao, Kai Shu, and Huan Liu. Applications of social media in hydroinformatics: A survey, 2019.

[17] Luis Miguel Díaz, Samuel Yanes Luis, Alejandro Mendoza Barrionuevo, Dame Seck Diop, Manuel Perales, Alejandro Casado, Sergio Toral, and Daniel Gutiérrez. Towards an autonomous surface vehicle prototype for artificial intelligence applications of water quality monitoring, 2024.

[18] Vukan Ninkovic, Dejan Vukobratovic, and Dragisa Miskovic. Uav-assisted distributed learning for environmental monitoring in rural environments, 2024.

[19] Florian Gerber and Douglas W. Nychka. Fast covariance parameter estimation of spatial gaussian process models using neural networks, 2020.

[20] Dazhou Yu, Xiaoyun Gong, Yun Li, Meikang Qiu, and Liang Zhao. Self-consistent deep geometric learning for heterogeneous multi-source spatial point data prediction, 2024.

[21] Kun Qian and Christian G. Claudel. Real-time mobile sensor management framework for city-scale environmental monitoring, 2020.

[22] Vatsala Nundloll, Barry Porter, Gordon Blair, Jack Cosby, Bridget Emmett, Ben Winterbourn, Graham Dean, Philip Beattie, Rory Shaw, Davey Jones, Dave Chadwick, Mike Brown, Wayne Shelley, and Izhar Ullah. The design and deployment of an end-to-end iot infrastructure for the natural environment, 2019.

[23] Anasse Boutayeb, Iyad Lahsen-cherif, and Ahmed El Khadimi. A comprehensive geoai review: Progress, challenges and outlooks, 2024.

[24] Joshua P. French and Stephan R. Sain. Spatio-temporal exceedance locations and confidence regions, 2013.

[25] Vishnu Jayaprakash, Jae Bem You, Chiranjeevi Kanike, Jinfeng Liu, Christopher McCallum, and Xuehua Zhang. Determination of trace organic contaminant concentration via machine classification of surface-enhanced raman spectra, 2024.

[26] Maciej Przybyłek, Waldemar Studziński, Alicja Gackowska, and Jerzy Gaca. The use of fast molecular descriptors and artificial neural networks approach in organochlorine compounds electron ionization mass spectra classification, 2019.

[27] Y. E. Shishkin and Aleksandr N. Grekov. Automated environmental monitoring intelligent system based on compact autonomous robots for the sevastopol bay, 2021.

[28] Jayant Gupchup, Andreas Terzis, Randal Burns, and Alex Szalay. Model-based event detection in wireless sensor networks, 2009.

[29] Dina Hussein, Lubah Nelson, and Ganapati Bhat. Sensor-aware classifiers for energy-efficient time series applications on iot devices, 2024.

[30] Nicholas G. Polson and Vadim O. Sokolov. Deep learning, 2018.

[31] Özer Çelik. A research on machine learning methods and its applications. *Journal of Educational Technology and Online Learning*, 1(3):25–40, 2018.

[32] Sertac Kilickaya, Mete Ahishali, Fahad Sohrab, Turker Ince, and Moncef Gabbouj. Hyperspectral image analysis with subspace learning-based one-class classification, 2023.

[33] Luca Colomba and Paolo Garza. Vigeo: an assessment of vision gnns in earth observation, 2024.

[34] Qinfeng Zhu, Ningxin Weng, Lei Fan, and Yuanzhi Cai. Enhancing environmental monitoring through multispectral imaging: The wastems dataset for semantic segmentation of lakeside waste, 2024.

[35] Utsav B. Gewali, Sildomar T. Monteiro, and Eli Saber. Machine learning based hyperspectral image analysis: A survey, 2019.

[36] René Brandborg Sørensen, Jimmy Jessen Nielsen, and Petar Popovski. Machine learning methods for monitoring of quasi-periodic traffic in massive iot networks, 2020.

[37] Marjolein Boonstra, Frédérick Bruneault, Subrata Chakraborty, Tjitske Faber, Alessio Gallucci, Eleanore Hickman, Gerard Kema, Heejin Kim, Jaap Kooiker, Elisabeth Hildt, Annegret Lamadé, Emilie Wiinblad Mathez, Florian Möslein, Genien Pathuis, Giovanni Sartor, Marijke Steege, Alice Stocco, Willy Tadema, Jarno Tuimala, Isabel van Vledder, Dennis Vetter, Jana Vetter, Magnus Westerlund, and Roberto V. Zicari. Lessons learned in performing a trustworthy ai and fundamental rights assessment, 2024.

[38] Amin Shoari Nejad, Rocío Alaiz-Rodríguez, Gerard D. McCarthy, Brian Kelleher, Anthony Grey, and Andrew Parnell. Sert: A transfomer based model for spatio-temporal sensor data with missing values for environmental monitoring, 2023.

[39] Qiangqiang Yuan, Huanfeng Shen, Tongwen Li, Zhiwei Li, Shuwen Li, Yun Jiang, Hongzhang Xu, Weiwei Tan, Qianqian Yang, Jiwen Wang, et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote sensing of Environment*, 241:111716, 2020.

[40] Genevieve Flaspohler, Nicholas Roy, and Yogesh Girdhar. Near-optimal irrevocable sample selection for periodic data streams with applications to marine robotics, 2018.

[41] Kai-Chieh Ma, Lantao Liu, and Gaurav S. Sukhatme. Informative planning and online learning with sparse gaussian processes, 2016.

[42] Yoonchang Sung, Zhiang Chen, Jnaneshwar Das, and Pratap Tokekar. A survey of decision-theoretic approaches for robotic environmental monitoring, 2023.

[43] Ahsnaul Haque, Md Naseef-Ur-Rahman Chowdhury, Hamdy Soliman, Mohammad Sahinur Hossen, Tanjim Fatima, and Imtiaz Ahmed. Wireless sensor networks anomaly detection using machine learning: A survey, 2023.

[44] Vu-Anh Le, Haruko Murakami Wainwright, Hansell Gonzalez-Raymat, and Carol Eddy-Dilek. Machine learning algorithms to assess site closure time frames for soil and groundwater contamination, 2024.

[45] M. Lambardi di San Miniato, R. Bellio, L. Grassetti, and P. Vidoni. Separable spatio-temporal kriging for fast virtual sensing, 2022.

[46] Junfeng Hu, Yuxuan Liang, Zhencheng Fan, Li Liu, Yifang Yin, and Roger Zimmermann. Decoupling long- and short-term patterns in spatiotemporal inference, 2024.

[47] Sahil Garg, Amarjeet Singh, and Fabio Ramos. Learning non-stationary space-time models for environmental monitoring, 2018.

[48] Balamurugan Sadaiappan, Preethiya Balakrishnan, Vishal CR, Neethu T Vijayan, Mahendran Subramanian, and Mangesh U Gauns. Applications of machine learning in chemical and biological oceanography, 2023.

[49] Daisuke Murakami, Mami Kajita, and Seiji Kajita. Spatial regression-based transfer learning for prediction problems, 2022.

[50] A. K. Akanbi and M. Masinde. Semantic interoperability middleware architecture for heterogeneous environmental data sources, 2018.

[51] A. K. Akanbi and M. Masinde. A framework for accurate drought forecasting system using semantics-based data integration middleware, 2017.

[52] James V. Zidek, Gavin Shaddick, and Carolyn G. Taylor. Reducing estimation bias in adaptively changing monitoring networks with preferential site selection, 2014.

[53] DK Banerjee. Water pollution and human health. 2017.

[54] Chia-Hsiang Lin, Po-Wei Tang, and Alfredo R. Huete. Quantum feature-empowered deep classification for fast mangrove mapping, 2025.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.