



Review

# Machine Learning in Geosciences: A Review of Complex Environmental Monitoring Applications

Maria Silvia Binetti <sup>1,2,\*</sup> , Carmine Massarelli <sup>1</sup> and Vito Felice Uricchio <sup>1,\*</sup>

<sup>1</sup> Water Research Institute, Italian National Research Council (IRSA-CNR), 70132 Bari, Italy; carmine.massarelli@ba.irsa.cnr.it

<sup>2</sup> Department of Earth and Geoenvironmental Sciences, University of Bari Aldo Moro, 70125 Bari, Italy

\* Correspondence: mariasilvia.binetti@ba.irsa.cnr.it or maria.binetti@uniba.it (M.S.B.); vito.uricchio@ba.irsa.cnr.it (V.F.U.)

**Abstract:** This is a systematic literature review of the application of machine learning (ML) algorithms in geosciences, with a focus on environmental monitoring applications. ML algorithms, with their ability to analyze vast quantities of data, decipher complex relationships, and predict future events, and they offer promising capabilities to implement technologies based on more precise and reliable data processing. This review considers several vulnerable and particularly at-risk themes as landfills, mining activities, the protection of coastal dunes, illegal discharges into water bodies, and the pollution and degradation of soil and water matrices in large industrial complexes. These case studies about environmental monitoring provide an opportunity to better examine the impact of human activities on the environment, with a specific focus on water and soil matrices. The recent literature underscores the increasing importance of ML in these contexts, highlighting a preference for adapted classic models: random forest (RF) (the most widely used), decision trees (DTs), support vector machines (SVMs), artificial neural networks (ANNs), convolutional neural networks (CNNs), principal component analysis (PCA), and much more. In the field of environmental management, the following methodologies offer invaluable insights that can steer strategic planning and decision-making based on more accurate image classification, prediction models, object detection and recognition, map classification, data classification, and environmental variable predictions.



**Citation:** Binetti, M.S.; Massarelli, C.; Uricchio, V.F. Machine Learning in Geosciences: A Review of Complex Environmental Monitoring Applications. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 1263–1280. <https://doi.org/10.3390/make6020059>

Academic Editor: Andreas Holzinger

Received: 30 April 2024

Revised: 31 May 2024

Accepted: 3 June 2024

Published: 5 June 2024



**Copyright:** © 2024 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/>).

**Keywords:** machine learning; environmental monitoring; geosciences

## 1. Introduction

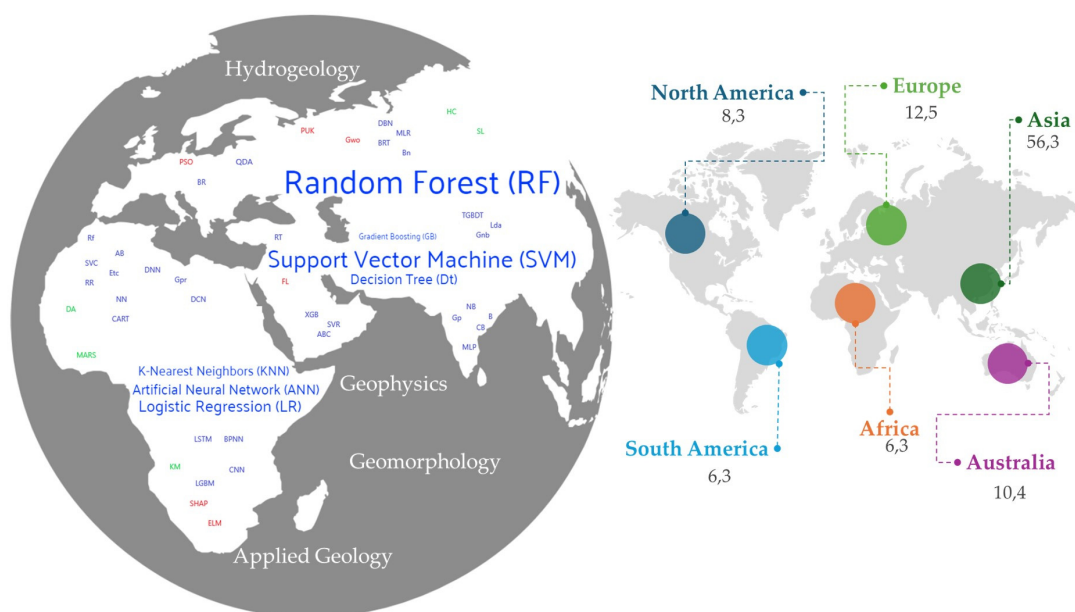
Machine learning (ML) has significantly revolutionized scientific methodology in geoscience applications by introducing automation, enhancing efficiency, enabling adaptability, ensuring security, and facilitating extensive data analytics [1]. Artificial intelligence (AI), machine learning, and deep learning (DL), highly cited contemporary technologies, are interconnected but distinct disciplines [2,3]. AI is a wider field that integrates various approaches to create intelligent systems. ML is a branch of AI that emphasizes learning from data and human-imitating algorithms, and DL is a further specialized subset of ML, focusing on the use of deep neural networks for pattern recognition [4]. This review exclusively focuses on the applications of ML within the field of geosciences.

The history of ML begins with cybernetics and the computer sciences in the early 1950s with the idea of using machines to simulate human learning processes. The primary stages between the 1950s and 1960s created the prototype of early neural networks [5]. The evolution of ML has progressed through distinct phases: rule-based systems (1960s–1970s), connectionism and backpropagation (1980s), a renaissance in the 1990s, and a deep learning resurgence in the 2010s [6]. Each phase marked significant advancements, diversification, and broader practical applications. The ML field collected substantial relevance and investment, evident in its transition from a limited number of global conferences to a

proliferation of both national and international events. This shift underscores its increasing significance and widespread interest within the scholarly community.

The application of ML covers four principal domains: prediction, feature importance extraction, anomaly detection, and discovering new materials. These categories collectively exemplify the multifaceted utility of ML methodologies in assorted analytical pursuits [7]. All predominant applications follow a uniform procedural framework: encompassing model preparation, model development, and post-model creation stages, inclusive of the interpretation and determination of applicability domains. This approach is well suited for addressing the intricate challenges in environmental monitoring. Comprehensively, environmental monitoring in geosciences contemplated the convergence of multiple disciplines in very complex data management. These disciplines are physics, geology, meteorology and atmospheric sciences, oceanography, environmental science, geomorphology, seismology, paleontology, mineralogy and petrology, geophysics, glaciology, hydrology, chemistry, biology, ecology, and anthropology.

On the global stage, scientific investigations into geosciences based on ML applications are predominantly guided by the utilization of supervised ML algorithms. The research was carried out utilizing the Clarivate site [8] by setting the following as filters: the last four-year open access scientific articles, sorted according to the first ten results by relevance, from the principal academic publishing companies specializing in scientific articles (e.g., Elsevier, Springer Nature, MDPI, IEEE, and Frontiers Media Sa) in the four geoscience fields (geophysics, geomorphology, hydrogeology, and applied geology). Upon the analysis of scientific articles, it is found that a notable 56.3% of the content originates from Asia, 12.4% from Europe, 10.4 from Australia, and 8.3% from North America, with equal results of 6.3% from South America and Africa. A substantial proportion of articles employ the supervised learning algorithm of random forest (RF), an ensemble method [9–11], a support vector machine (SVM) [12,13], logistic regression (LR) a linear model [14,15], an artificial neural network (ANN) [16–18], a decision tree (DT) [19,20], K-nearest neighbors (KNN) [21–23], and a Bayesian neural network (BNN) [24,25]. The investigation revealed an average utilization of four ML techniques reflecting the dynamic landscape of machine learning applications (Figure 1). The complete results of the frequencies and number of publications are reported in Table S1.



**Figure 1.** Word cloud graphics (on the left) displays an ML technique in order of importance (based on the font size) based on the four fields of geology (hydrology, geophysics, geomorphology, and applied geology). The word graph (on the right) represents the frequency of machine learning in geologic world publications.

The structure of the paper can be outlined as follows: Section 2 is an overview of the limits and challenges of geosciences in machine learning algorithms, while Section 3 provides a specific geoscience environmental monitoring application including quarries and discharge phenomena, coastal dunes safeguarding monitoring, illicit sea discharges, and pollution in different matrices from several sources in industrial complexes. In Section 4, the conclusion is presented and Section 5 offers a forward-looking perspective, anticipating strategic developments and innovation. The present comprehensive review systematically explores the application of machine learning (ML) algorithms within the realm of geosciences. Particular emphasis is placed on their use in environmental monitoring applications. The focus of the subsequent chapter will be to delve deeper into this specific area of application. The following table reports the nomenclature used internationally to distinguish the various ML algorithms (Table 1).

**Table 1.** Algorithms' nomenclature.

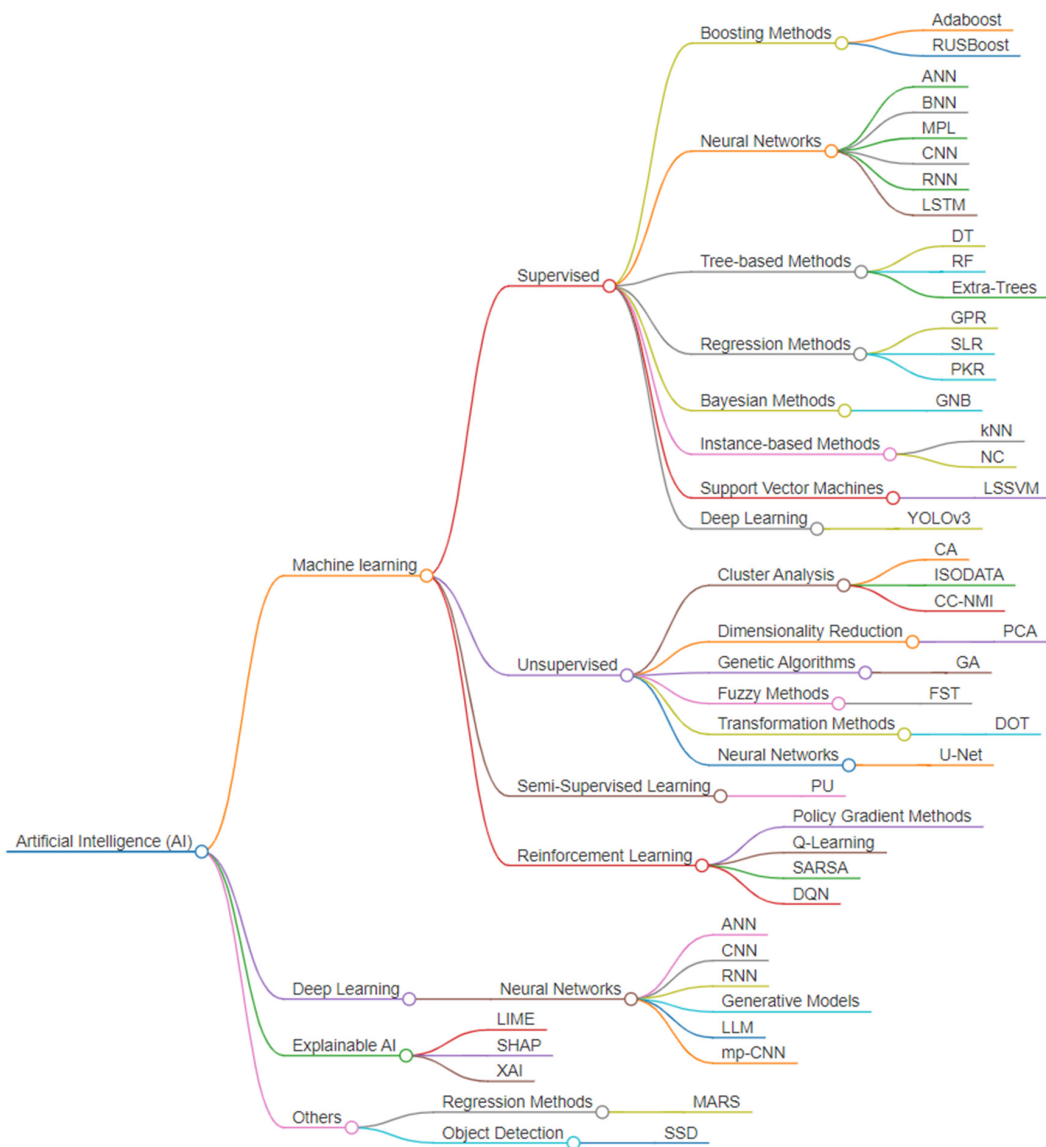
Abbreviation	Meaning	Abbreviation	Meaning
Adaboost	Adaptive Boosting	LSSVM	Vector Support Machines for Least Squares
ANN	Artificial Neural Network	LSTM	Long Short-Term Memory
BNN	Bayesian Neural Network	MARS	Multivariate Adaptive Regression Splines
CA	Cluster Analysis	ML	Machine Learning
CC-NMI	Cluster Confusion Normalized Mutual Information	mp-CNN	Multipath Convolutional Neural Network
CNN	Convolutional Neural Network	MPL	Multilayer Perceptron
DL	Deep Learning	NC	Nearest Centroid
DOT	Discrete Orthogonal Transformations	NN	Neural Network
DR	Dimensionality Reduction	OD	Object Detection
DT	Decision Tree	PCA	Principal Component Analysis
Extra-Trees	Extremely Randomized Trees	PKR	Polynomial Kernel Regression
FF	Futures Filtering	PU	Positive-Unlabeled Learning Algorithm
FPN	Feature Pyramid Network	RF	Random Forest
FST	Fuzzy Set Theory	RNN	Recurrent Neural Network
GA	Genetic Algorithm	RUSBoost	Random Under-Sampling Boosting
GNB	Gaussian Naive Bayes	SHAP	Shapley Additive Explanations
GANs	Generative Adversarial Networks	SLR	Stepwise Linear Regression
GPR	Gaussian Process Regression	SSD	Single Shot Detector Algorithm
ISODATA method	Iterative Self-Organizing Data Analysis Technique	SVM	Support Vector Machine
kNN	k-Nearest Neighbors	U-Net	Unique Architecture of the Network is a "U" Shape (CNN)
LIME	Local Interpretable Model-Agnostic Explanations	XAI	eXplainable Artificial Intelligence
LLMs	Large Language Models	YOLOv3	You Only Look Once, Version 3

## 2. Overview of the Limits and Challenges of Geosciences in Machine Learning Algorithms

In the field of geoscience environmental monitoring, conventional methodologies include a range of specified methods such as field surveys and measurements, soil–rock–water sampling and geochemical analysis, geodetic and remote sensing, and climate monitoring. These methodologies are fundamental tools for assessing and understanding environmental dynamics, providing crucial insights into various geological and ecological processes.

Traditional sampling methods still have fundamental significance, although the promising ML techniques offer notable enhancements across six distinct domains: enhanced accuracy and spatial coverage [26,27], efficiency in time and resource utilization, an improved understanding of complex models [28–30], adaptability and continual updates, automation and reduced human dependence, reliability, and validation challenges [31,32].

The implementation of ML methodologies into geosciences offers numerous potential advantages in data analysis, and ML enables the efficient analysis of large volumes of data. Before executing ML, a substantial portion of the effort is dedicated to preprocessing and data transformations, entailing tasks such as eliminating redundancy, inconsistency, noise, and heterogeneity, as well as transforming and labeling data. Dealing with big data turns out to be very advantageous, creating the opportunity to diminish reliance on human supervision by learning directly from the three key concepts characterizing these data such as volume, variety, and velocity [33,34]. Analyzing extensive datasets enhances scalability through the proficient management of large data volumes, augments adaptability by refining accuracy iteratively, and facilitates the effective management of data veracity [35]. The ability to model, optimize, and integrate multi-source data, automate complex tasks, and provide forecasts facilitates land management by providing a complete view of the processes. ML algorithms are grouped into four main applications: detecting objects and events, estimating variables, long-term forecasting variable problems, and mining relationship data [36]. In geo-monitoring, advanced methods for estimating landslide movement using drone data (UAV) have been developed, improving accuracy by 8% compared to traditional methods. In parallel, wireless sensor networks (WSNs) have been used to monitor the structural health of homes in areas at risk of ground movement. These technologies, which use artificial intelligence and the Internet of Things, represent the vanguard in remote monitoring, contributing to the prevention of harm and the safety of people [37,38]. In geoscience, machine learning methodologies outlined by Dramsch et al. (2020) are primarily categorized into developing alternative models to optimize computational efficiency, crafting models to supplement or replace human intervention, enabling previously unattainable geoscientific activities [39]. Machine learning (ML) methodologies involve supervised, unsupervised, reinforcement learning (LR), semi-supervised learning, deep learning, explainable AI, and other algorithms (Figure 2) [40].



**Figure 2.** Classification of artificial intelligence algorithms (AI). This figure illustrates the broad spectrum of AI algorithms, with a particular focus on machine learning (ML) methods. It is important to note that numerous methodologies span across multiple categories (for instance, the deep learning methodology). This overlap signifies the versatility and adaptability of these algorithms in various research and application domains.

### 2.1. Supervised ML Algorithms

Supervised learning encompasses various problem categories and techniques, including classification, regression, neural network-based approaches, ensemble methods, optimization-based techniques, object detection, feature filtering, and dimensionality reduction [40,41]. Specific algorithms and methods, such as boosting methods, neural networks, tree-based methods, regression methods, Bayesian methods, instance-based methods, support vector machines, and deep learning, are employed to address these problems.

The first supervised ML algorithm methods are boosting methods, which include adaptive boosting (AdaBoost), and random under-sampling boosting (RUSBoost). The Adaboost algorithm improves the model’s performance, and RUSBoost uses random under-sampling to resolve the class imbalance.

Neural network methods include artificial neural networks (ANNs), multilayer perceptron (MPL), convolutional neural networks (CNNs), recurrent neural networks (RNNs),



long short-term memory (LSTM), and Bayesian neural networks (BNNs). Neural network-based algorithms simulate the biological neural networks of the human brain. In artificial neural network (ANN) methods, the “neurons” act to solve some complex problems to extract trends or detect patterns [42]. In addition, convolutional neural networks (CNNs) are primarily used for image classification and recurrent neural networks (RNNs) for sequential data, such as natural language and time series, and long short-term memory (LSTM) is used to handle the gradient problem that vanishes in recurring neural networks.

The tree-based methods include decision trees (DTs), extremely randomized trees (extra-trees), and random forest (RF). A decision tree (DT) is an algorithm used for regression and classification [43]. The principal idea is to divide a dataset into smaller subsets, and it is widely used for its interpretability features and ease of viewing. The extremely randomized trees (extra-trees) are a decision tree adaptation that randomly selects dividing points for each node in the tree. Despite lower accuracy, they prove to be quicker to train than traditional decision trees [9].

Regression algorithms aim to take the relationship between a variable output target and input features, facilitating the prediction of new data [40] like Gaussian process regression (GPR), stepwise linear regression (SLR), and polynomial kernel regression (PKR). GPR models input and output variable distributions using Gaussian processes. A regression model based on a genetic algorithm (GA) optimizes parameters through iterative generations, generating potential solutions and iteratively refining them to identify the optimal solution.

Gaussian naive Bayes (GNB), which pertains to Bayesian methods, assumes that features have a bell-shaped distribution, making it easier to calculate probabilities and classify data efficiently.

In the category of instance-based methods, one of the most simple and popular classifications of nonparametric variables is k-nearest neighbors (kNN) [9]. It is employed to perform mainly classifications or predictions on data grouping based on the proximity (neighborhood) of training points. The nearest centroid (NC) calculates the centroids for each class and classifies the new points based on their distance from the centroids.

Support vector machines (SVMs) are active in high-dimensional spaces by finding the optimal hyperplane that maximizes the margin between classes in the feature space robust against overfitting [44]. Vector support machines for least squares (LSSVM) integrate SVMs with least squares principles to minimize error by finding a function approximating the data.

Deep learning is recognized as belonging to the domain of supervised learning algorithms. Moreover, due to its distinctive architecture and methodology, deep learning also constitutes a distinct category within the broader landscape of machine learning techniques.

## 2.2. Unsupervised, Semi-Supervised, and Reinforcement Learning ML Algorithms

Unsupervised algorithms are used for data analysis without specific labels or targets to predict. These algorithms search for patterns or structures in the data without outward-reliant variable information and can be further subdivided into several categories including clustering algorithms, size reduction, and optimization based on set theory. The models use previously learned features to recognize the new data class entered [45,46].

Clustering algorithms are a set of techniques employed to group similar objects based on certain similarity or dissimilarity metrics, e.g., cluster analysis (CA), the iterative self-organizing data analysis technique (ISODATA) and cluster confusion normalized mutual information (CC-NMI). The ISODATA method “Iterative Self-Organizing Data Analysis Technique” is a clustering-specific algorithm that divides data into clusters built on their statistical properties, iteratively updating centroids and cluster members. CC-NMI is a measure of similarity between two cluster partitions, which considers the confusion between clusters and normalizes the result using mutual information.

In the dimensionality reduction algorithms, there is principal component analysis (PCA) and dimensionality reduction (DR). PCA is a method to reduce the dimensionality of the data while maintaining maximum variance in the original data.

A further type of optimization-based algorithm is the self-optimizing machine learning algorithm. This is an algorithm that independently gives its parameters to optimize a given performance metric. For algorithms based on set theory, there is the fuzzy set theory (FST), an extension of the classical theory. This theory assigns a grade of belonging between 0 (indicates no affinity) and 1 (indicates full affinity).

In transformation methods, the discrete orthogonal transformation (DOT) methods transform data using discrete orthogonal transformations to improve model analysis or training.

An example of a neural network is U-Net, a unique “U-shaped” neural network architecture, often used in convolutional neural networks (CNNs), designed for image segmentation and reconstruction problems.

In the domain of semi-supervised learning algorithms, the positive-unlabeled learning algorithm (PU) stands out. This approach leverages a combined dataset comprising both labeled and unlabeled data to enhance model performance. It operates under the assumption that the unlabeled data pool may encompass both positive and negative examples. This learning methodology proves particularly beneficial in scenarios characterized by an extensive repository of unlabeled data alongside a limited subset of positively labeled data.

Lastly, reinforcement learning is a paradigm of machine learning in which an agent learns to perform actions in an environment, receiving feedback through rewards or penalties, to maximize a specific goal (e.g., policy gradient methods, Q-learning, SARSA (state-action-reward-state-action), and deep Q-networks (DQNs)).

### 2.3. Deep Learning

Deep learning (DL) is a subset of machine learning that employs algorithms modeled after the brain’s structure and function. It excels in processing large datasets and uncovering complex relationships through multiple levels of abstraction. Specific algorithms and methods, such as artificial neural networks (ANNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), generative models, large language models, and multipath convolutional neural networks, are employed to address these problems.

Within convolutional neural networks (CNNs), a multitude of sophisticated techniques are employed to enhance performance and accuracy such as feature pyramid networks (FPNs), U-Net, you only look once, version 3 (YOLOv3), and single-shot detector (SSD) algorithms. U-Net carries a contraction path and an expansive path to improve its performance and accuracy in image segmentation tasks. You only look once (YOLO) is a state-of-the-art real-time object detection system. It predicts class probabilities and bounding boxes for objects directly from full images in a single pass. Utilizing a 53-layer convolutional neural network, YOLO balances speed and precision. It features bounding box prediction, multi-scale prediction, and class prediction.

Large language models (LLMs) are advanced artificial intelligence systems designed to comprehend and generate human language. They play a crucial role in numerous applications, including chatbots, virtual assistants, and sophisticated search tools.

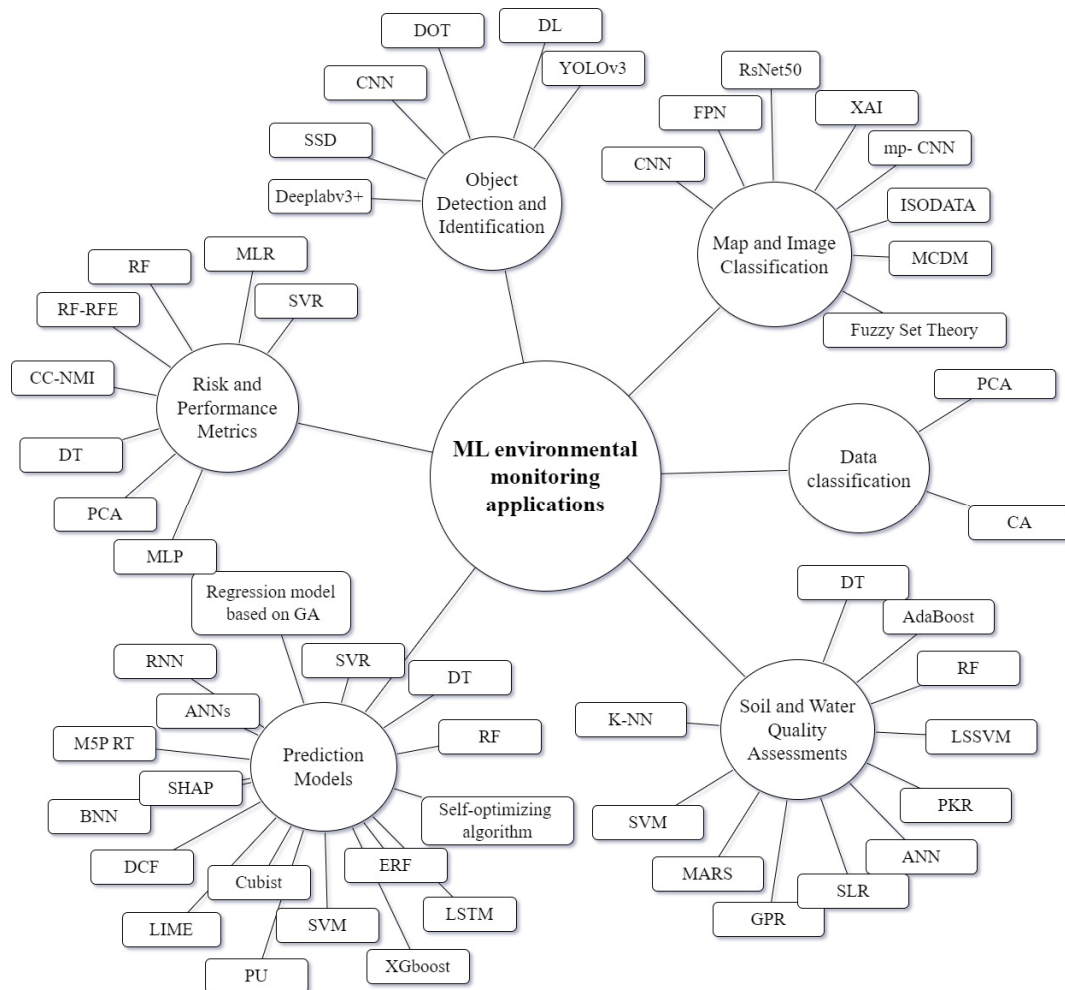
### 2.4. Explainable AI and Other Algorithms

Explainable AI encompasses methodologies aimed at enhancing interpretability in decision-making processes within artificial intelligence systems. Techniques such as LIME and SHAP provide local and individual model prediction explanations, respectively. LIME offers local interpretability, while SHAP leverages game theory principles for comprehensive insights. Regression methods, like multivariate adaptive regression splines (MARS), model relationships between variables, enhancing the understanding of complex data structures. In computer vision, object detection, exemplified by the single-shot detector (SSD) algorithms, entails identifying and localizing objects within visual data, enabling diverse applications.

### 3. Environmental Monitoring Applications in Geosciences

Environmental monitoring in geosciences is fundamental to understanding and mitigating the impacts of human activities on the environment. It encompasses methodologies and strategies for identifying, analyzing, and establishing environmental parameters to gauge and quantify environmental impacts. This process relies on various testing and evaluation methodologies to furnish crucial insights into environmental conditions and potential hazard levels. This article focuses on select issues within environmental monitoring for several reasons. Firstly, it underscores the urgency and gravity of environmental concerns, given their profound implications for planetary health and human well-being. Additionally, data availability from sources such as satellites, environmental sensors, industrial registries, and other tools has facilitated the choice of pertinent environmental issues for examination. Lastly, considering the practical and socio-economic implications, the potential for substantial enhancements and the promotion of sustainable solutions contribute to tackling genuine social challenges.

The methodologies of machine learning (ML) discussed in this article will be implemented in environmental monitoring endeavors to focus on key environments, including landfills, quarries, coastal dune protection, sea discharge, and complex industrial settings (Table 2, Figure 3).



**Figure 3.** Categorization of machine learning applications in environmental monitoring. This image provides a comprehensive overview of the broad categories of environmental monitoring applications that leverage ML techniques. The categories include map and image classification, object detection and identification, prediction models, data classification, risk and performance metrics, and soil and water quality assessments.



### 3.1. Quarry and Landfill Monitoring ML Application

The problem of waste management, including unauthorized dumpsites, is a global concern [47,48]. Despite regulatory efforts, landfills have harmful effects on soil, air, water, and biodiversity [49]. The global population is expanding, resulting in a rise in waste generation [50]. The exponential escalation in waste generation has required an increased dependence on and proliferation of landfills for disposal, whether lawful or illicit [51]. Recent publications have highlighted the utilization of machine learning (ML), deep learning (DL), and heuristic models. Awadh, M. Al and Mallick J [52] merged multi-criteria decision making (MCDM), fuzzy set theory, GIS, and eXplainable Artificial Intelligence (XAI). The models provide a landfill site potential zone (LSPZ) map classification. The model employs geospatial and environmental datasets to discern candidate locations for landfill establishment. It leverages machine learning methodologies, with a focus on an optimized ensemble bagging model, to categorize various regions as prospective landfill sites [53,54]. The study utilizes SHAP (SHapley Additive exPlanations) and LIME (local interpretable model-agnostic explanations) analyses to elucidate machine learning models and enhance the comprehension of model predictions. A recent investigation introduced a machine learning model employing the positive-unlabeled (PU) learning algorithm within an ensemble framework. This model has undergone validation utilizing the PU-based random forest technique for monitoring and preventing the illegal disposal of hazardous waste (HW) [55]. Furthermore, cluster analysis [56], a statistical technique, facilitates the unsupervised grouping of set elements into classes for grouping similar classifications for regional water resource protection. An additional proposed methodology uses a machine learning method called a multipath convolutional neural network (mp-CNN), and it is used to locate waste piles in roads and roadsides. In the test phase, the model with an image classification showed excellent performance, usable in developing countries [57]. A novel method is proposed by Torres, R. N. and Fraternali with a convolutional neural network (CNN) combination of ResNet50 and feature pyramid network (FPN) methods for a risk map result [58,59]. Illegal landfill detection is formulated as a multi-scale scene classification problem, with datasets of about 3000 images with an accuracy of 88%. Leveraging the single-shot detector (SSD) algorithm, in conjunction with deep learning methodologies and remote sensing techniques, facilitates the real-time detection of objects within video streams, thereby enhancing the efficacy of dumping detection [60]. This amalgamation of advanced technologies underscores the potential for significant advancements in waste management. Moreover, a machine learning technique based on discrete orthogonal transformations (DOTs) is used. This technique is used to identify waste disposal facilities from high-resolution spatial images [61]. Lastly, YOLOv3 (you only look once, version 3) enables the real-time detection of specific objects in videos, live feeds, or images [62].

Monitoring activities as regards quarries, mines, and excavations for material extraction cause environmental problems, with potential implications for environmental degradation and a high risk of environmental damage. Below are some examples of the application of ML techniques. Larrea-Gallegos et al. 2023 presented an ML approach with an unsupervised learning algorithm (X-means) and a random forest (RF) classification model to improve strategic planning [63]. Furthermore, Fernández-Alonso et al. 2023 proposed a convolutional neural network (CNN) for the identification of mining remains [64]. The study conducted by Fissaha et al. (2023) used a Bayesian neural network (BNN) and other models like gradient boosting, K-neighbors, decision trees, and random forest to predict the blast-induced ground vibration [65]. The article evaluates additional machine learning methods such as the nearest centroid, random forest, decision trees, and Gaussian naive Bayes. Moreover, discloses a decision tree algorithm based on the parametric analysis of tunneling-induced ground settlements to understand the tunneling-induced ground subsidence [66]. This methodology can aid in the identification of historical subterranean quarries, even when their spatial coordinates have been obscured within highly urbanized locales. In a further study, a CNN, a type of deep learning model, is employed to identify deformations within a national-scale velocity field. The primary objective of the model is

to accurately detect and classify various forms of deformation. These include subsidence resulting also from coal mining activities, deformations in slate quarries, and alterations due to tunnel engineering works [67]. Another machine learning research is focused on predicting the peak particle velocity (PPV) values with a DT model. PPT is a measure of ground vibration amplitude due to blasting operations in limestone quarries by the use of the explosive charge weight per delay and the distance from the blast [68].

In conclusion, the advantages of implementing ML and AI in landfill and quarry management include enhanced accuracy, predictive capabilities, and real-time detection capabilities. Advanced ML models, such as SHAP, LIME, and PU-based random forest, offer precise classification and monitoring, while algorithms like YOLOv3 and SSD provide real-time object detection. However, the disadvantages include the complexity of implementing and interpreting sophisticated models and the dependency on high-quality, extensive datasets.

### 3.2. Coastal Dunes Preservation ML Application

The preservation of coastal dunes is paramount for safeguarding our shorelines. Coastal dunes are pivotal in combating erosion and conserving marine ecosystems [69–71]. ML techniques are employed for dune reinforcement, forecasting, monitoring, and sustainable governance. In the study conducted by Pinton et al. 2023, a regression model based on a genetic algorithm (GA) and a random forest algorithm (RF) were utilized to estimate ground elevation in coastal dunes [72]. A further exploration considers the coastal dunes along Lake Michigan's eastern shoreline to obtain an image classification from aerial images with the ISODATA classification method [73]. A further exploration uses three distinct algorithms, ANN, SVM, and RF, to employ high-resolution mapping [74]. Mohammadpoor, M. and Eshghizadeh, M. 2021 present an advanced algorithm designed for the precise extraction of dunes from Landsat satellite imagery in both terrestrial and coastal settings. K-nearest neighbors, decision trees, AdaBoost, RUS Boost, and SVM algorithms leverage intelligent techniques to accurately identify and delineate dune features [75]. Finally, there is an example of assessing wave runup and coastal dune erosion through the use of the Gaussian process (GP), a nonparametric supervised learning method [76].

Summing up, the advantages of using ML and AI in coastal dune protection include precise mapping and erosion prediction with Gaussian process models. However, the disadvantages involve challenges in adapting to environmental variability and high computational costs.

### 3.3. Water Discharges into the Sea ML Application

ML for the analysis, prediction, and comprehension of water discharges into the sea facilitates efficient marine management and environmental conservation efforts. Understanding the impact of discharges, such as wastewater, pollutants, or runoff, on marine ecosystems is imperative. The models enable the prediction of discharges, aiding in water resource planning, disaster prevention, and environmental protection. Through the study of discharges, the optimization of water usage, the prevention of shortages, and the maintenance of ecosystem balance can be achieved. ML models can accurately predict water quality parameters even with limited data, crucial for pollution control, ecosystem health, and human well-being. In their study published in 2023, Liao et al. utilized the DeepLabv3+ semantic segmentation architecture for monitoring oil spill risk in coastal areas. Their approach relied on polarimetric synthetic aperture radar (SAR) satellite imagery [77]. Magri, S. et al. 2023 developed machine learning techniques utilizing two distinct generalized linear models: stepwise linear regression (SLR) and polynomial kernel regression (PKR). These models were employed to infer seawater turbidity from Sentinel-2 imagery [78]. In an alternate investigation, various machine learning algorithms, including a support vector machine (SVMs), random forest (RF), an artificial neural network (ANN), and combined algorithms, were employed for the detection of sediment discharge in rivers using Sentinel-2 satellite imagery [79]. A recent study employed machine learning method-

ologies to reconstruct daily sea discharge. Six distinct machine learning algorithms were utilized in the analysis (RF, GPR, SVR, decision tree (DT), least squares support vector machine LSSVM, and multivariate adaptive regression spline MARS). The research aimed to accurately model and predict daily discharge patterns at sea [80]. Granata et al. 2018 developed three ML models (M5P regression tree, random forest, and support vector regression) for spring discharge forecasting. These prototypes were constructed only using historical discharge data and cumulative rainfall information [81].

The advantages of using ML and AI in sea discharge include comprehensive monitoring and improved pollution control. Models such as DeepLabv3+ and random forest accurately monitor and predict water quality parameters. Nonetheless, disadvantages include data scarcity and the complexity of modeling in diverse environments.

### 3.4. Contaminated Industrial Water and Soil Matrix ML Application

The utilization of machine learning in pollution monitoring within contaminated industrial complexes holds substantial promise for enhancing pollution monitoring and management, thereby fostering environmental sustainability. Through machine learning algorithms, the capability to forecast forthcoming pollution levels and categorize pollution sources based on gathered data is facilitated, facilitating precise intervention strategies. The remediation of polluted areas occurs in both aquatic and terrestrial environments. Machine learning (ML) techniques play a pivotal role in enhancing remediation efforts in both contexts.

Emerging technologies have showcased their capacity to enhance, simulate, and automate water treatment methodologies, surveillance, and ecological system administration. The objective is to safeguard aquatic ecosystems through the observation and identification of contaminants. With the exponential surge in aquatic environmental data, ML has emerged as a pivotal instrument for data scrutiny, categorization, and prognostication [82–85]. The eutrophication and the proliferation of chlorophyll algae in water frequently result from inadequate wastewater management and unsustainable agricultural practices. Huang, H. and Zhang, J. 2024 employed four distinct methodologies to ascertain the significant factors influencing chlorophyll-*a* (Chl-*a*) content, the support vector regression (SVR) model demonstrating superior accuracy and precise predictions [86]. A new investigation examines urban river water quality monitoring through the utilization of a self-optimizing machine learning algorithm applied to multi-source remote sensing data (satellite images, UAV images, and water samples) [87]. In addition, in Zhi, W. et al.'s study (2021), a recurrent neural network (RNN) known as long short-term memory (LSTM) is employed to forecast levels of dissolved oxygen (DO) within riverine environments [88]. Moreover, an article conducted a comparative analysis utilizing big data to assess the prediction performance and identify key water parameters in surface water quality. The study compared seven traditional and three ensemble learning models, including a decision tree (DT), random forest (RF), and deep cascade forest (DCF) [89]. Furthermore, another article aims to improve the classification of water images with a neural attention network [90]. Finally, an Indian study utilizes cluster analysis (CA) and principal component analysis (PCA) to evaluate heavy metal contamination in aquatic environments [91].

Several scholarly publications have extensively examined the application of ML techniques for the monitoring of pollutants within soil ecosystems in industrial urbanization contamination. Zhao, W. et al. 2023 present a precise prognostication framework for soil heavy metal contamination, utilizing an enhanced amalgamation of three distinct machine learning methodologies: extreme gradient boosting (XGB), random forest (RF), and an artificial neural network (ANN) [92]. A further investigation has elucidated the considerable efficacy of machine learning techniques, notably RF and cubist techniques, in leveraging environmental datasets to forecast concentrations of heavy metals in soil [93]. Moreover, a study employs RF simulations in conjunction with spatial bivariate analysis to discern the presence of heavy metal pollution in agricultural land. Spatial bivariate analysis is utilized to investigate the interplay between soil metal contamination and predominant

human activities [94]. In additional research aimed at delineating soil pollution within an arsenic-contaminated agricultural domain, four distinct automated apprehension methodologies were employed. These methodologies encompassed the support vector machine (SVM), multi-layer perceptron (MLP), random forest (RF), and extreme random forest (ERF) models. Notably, the extreme random forest (ERF) model exhibited superior performance among the studied methodologies [95,96]. And last, in Zhang, H. et al. 2020, three models, RF, ANN, and SVM, allow the source identification and spatial prediction of heavy metals in soil in a rapid urbanization area [97].

Taken together, the advantages of using ML-AI in complex industrial settings include enhanced monitoring and accurate classification. Models including extreme gradient boosting and random forest provide high accuracy in identifying contamination sources. Notwithstanding, the disadvantages include the requirement for technical expertise, high computational power, and extensive datasets.

**Table 2.** Machine learning environmental monitoring applications. This table categorizes various environmental monitoring applications (e.g., landfill, quarry, safeguarding the coastal dune, discharge into the sea, and complex industrial soil and water contamination) where ML methodologies have been utilized.

Fields	Reference	Year	Input Data	Methods	Output	Model
Landfill	[53]	2024	Spatial and environmental data	MCDM, fuzzy set theory, XAI	Map classification (landfill site potential zones (LSPZ))	ML
	[56]	2023	Spatial numerical and categorical data	SHAP, LIME, PU	Risk maps, model performance metrics	ML
	[57]	2023	Data extracted from GIS mapped with data from different sources	CA	Group similar characteristics classification	ML
	[58]	2022	Online sources and camera images	mp-CNN	Image classification	DL
	[59] [60]	2021	Satellite and imagery date	CNN, RsNet50, FPN	Image classification	DL
	[61]	2021	Unmanned aerial vehicle images	SSD, DL	Object detection	DL
	[62]	2020	Remote sensing (RS) high-resolution satellite images	DOT	Location identification and classification	Heuristic
	[63]	2020	Real-time video stream from a surveillance camera	YOLOv3	Object detection and recognition	DL
Quarries— mines	[64]	2023	Public georeferenced data	X-means, RF	Geospatial probability map to improve strategic planning	ML
	[65]	2023	Drone database imagines	CNN + (NC, RF, DT, GNB)	Object identification	DL
	[66]	2023	Environmental data	BNN + (GB, K-N, DT, RF)	Predicting the peak particle velocity (PPV) values	ML

Table 2. Cont.

Fields	Reference	Year	Input Data	Methods	Output	Model
Quarries— mines	[67]	2022	Subsidence environmental data	DT	Feature importance (relationship between tunneling-induced ground subsidence and correlated factors)	ML
	[68]	2021	Sentinel 1 data, velocity maps spanning	CNN	Detection of deformation areas	DL
	[69]	2021	Explosive charge weight per delay and the distance from the blast	DT	Predicting the peak particle velocity (PPV) values	ML
Safeguarding coastal dunes	[73]	2023	Georeferenced data (UAV-LIDAR and UAV-DAP point clouds)	Regression model based on GA	Equation for ground elevation, relative importance of predictors, interpretability	GA
				RF	Predicted ground elevation	ML
	[74]	2023	Aerial imagines	ISODATA	Image classification	ML
	[75]	2021	Multispectral data	ANN, SVM, RF	High-resolution mapping	ML
	[76]	2021	Satellite data	K-NN, DT, AdaBoost, RUSBoost, SVM	Land cover map	ML
	[77]	2019	Aerial imagines	GP	Probabilistic parameterization of wave runup	ML
Discharges into the sea	[78]	2023	Satellite data	DeepLabv3+-	Object detection	DL
	[79]	2023	Satellite data	SLR, PKR	Surface turbidity	ML
	[80]	2022	Satellite data	SVM, RF, ANN	Time-series suspended sediment discharge	ML
	[81]	2022	Upstream– downstream multi-station data	RF, GPR, SVR, DT, LSSVM, MARS	Daily averaged discharge	ML
	[82]	2018	Monthly averages of flows and monthly cumulative rainfall in the aquifer basin	M5P RT, RF, SVR	Prediction of the flow rate	ML
		[87]	2024	Chlorophyll water content	CC-NMI, PCA, DT, RF-RFE, MLR, MLP, SVR	Eutrophication prediction and risk assessments
Complex industrial contamination— Water	[88]	2023	Multi-source remote sensing data	Self-optimizing algorithm	Prediction performance of water quality parameters	ML
	[89]	2021	Dissolved oxygen	LSTM, RNN	Prediction model	ML
	[90]	2020	Water parameter sets	DT, RF, DCF	Perdition water quality	ML
	[91]	2020	Water images	CNN	Water image classification	ML
	[92]	2020	Water parameter sets	CA, PCA	Data classification	ML



Table 2. Cont.

Fields	Reference	Year	Input Data	Methods	Output	Model
Complex industrial contamination—Soil	[93]	2023	Heavy metal content, spatial information	ANNs, RF, XGboost	Predicting soil heavy metal (HM) pollution assessment	ML
	[94]	2022	Environmental variables	RF, cubist	Prediction of heavy metals in soils	ML
	[95]	2021	Environmental variables	RF	Environmental variables predictions	ML
	[96]	2020	Environmental parameters	SVM, MLP, RF, ERF	The risk map level in the soil	ML
	[97]	2020	Soil concentrations, land use types	RF, ANN, SVM	Map spatial pattern concentration	ML

#### 4. Conclusions

In contemporary environmental monitoring, numerous ML algorithms are employed, with a preference for adapted classic models. Supervised ML methods have been predominantly favored over unsupervised approaches in recent scholarly works. RF is currently the most widely used method in this field of research. RF is favored in machine learning for its adaptability to classification and regression tasks. Its resilience against overfitting is notable, attributed to the construction of each tree using random data subsets. RF's simplicity facilitates its application and allows for ensemble integration, enhancing model efficacy in solving intricate problems and improving predictive accuracy [9–11]. The article is employed for geospatial strategic planning probability maps, object identification, and general prediction models. In various previously mentioned geoscience applications, it is predominantly utilized for monitoring water discharges and in complex industrial urbanization contamination [64,74,79,81,86,96,98,99]. The DT, following the RF methods, stands as a prevalent ML approach esteemed for its user-friendly interpretability, adaptable nature across both classification and regression tasks, and adept handling of diverse data types, including numerical and categorical variables [20]. Noteworthy for its capability to elucidate decision pathways and accommodate various data complexities, the DT consistently demonstrates robust predictive performance, often rivaling or surpassing more sophisticated methodologies such as RF. In the numerous geoscientific contexts previously cited, its primary application lies in the water industrial urbanization and associated contamination for water quality prediction [64,66,75,80,86,89]. The support vector machine (SVM) is highly esteemed in the field of machine learning due to its capacity for optimal data classification, exceptional versatility, and efficiency, all achieved without necessitating extensive parameter tuning or adjustments [74,75]. Conversely, the ANN offers the capability to approximate any computable function, facilitate pattern recognition, and address common troubleshooting challenges, leveraging the advancements in computational prowess [79,97]. SVM and ANN are largely used in concentration maps of spatial patterns for complex industrial urbanization contamination. CNN [57,64,90] and PCA [86,91] methodologies are prominently featured among the preeminent analytical approaches employed in contemporary research endeavors.

#### 5. Outlook and Future Research

Future advancements in the AI-ML domain involve improving existing models such as RF, DT, SVM, ANN, CNN, and PCA to elevate their accuracy and efficiency for adapted tasks. Simultaneously, pioneering algorithms could emerge to tackle specific environmental monitoring challenges, potentially outperforming conventional methods. Integrating diverse methodologies stands as an essential avenue for increasing model efficacy. Ensemble methods, for instance, guarantee further exploration, leveraging the complementary strengths of different techniques. The advanced sensor integration equipped with AI-ML



processing capabilities will enable collection and real-time environmental data analysis, providing a broader and more detailed coverage of monitored areas. The development of software platforms dedicated to collecting, storing, and analyzing geoscientific data, along with creating user-friendly software interfaces, will facilitate end users' effective use of these technologies.

The ML model integration and Internet of Things (IoT) sensors are paving the way for future landfill development. This combination facilitates continuous monitoring and real-time decision making, aiding policymakers in devising more effective waste management regulations. In quarry fields, using UAV and satellite imagery with ML models enables automated and comprehensive monitoring. The development of hybrid models, which amalgamate multiple ML techniques, is enhancing the accuracy and reliability of predictive models. Moreover, climate change adaptation and community involvement are revolutionizing coastal dune protection. Models are being developed to predict and adapt to the impacts of climate change on coastal dunes. Simultaneously, user-friendly tools are being created for local communities to monitor and protect their coastal areas. Sea discharge management is being employed through real-time analysis and integration with environmental policies. In complex industrial settings, smart remediation and cross-disciplinary approaches are being employed. These advancements are contributing significantly to environmental conservation and protection.

Prospective research endeavors might examine novel applications, particularly in domains where conventional methods encounter limitations. Prioritizing the development of models resilient to overfitting while maintaining interpretability could spearhead future innovations.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/make6020059/s1>, Table S1: The frequency of the machine learning algorithms in consulted geoscience literature.

**Author Contributions:** Conceptualization, M.S.B. and C.M.; methodology, M.S.B.; software, C.M.; validation, M.S.B., C.M. and V.F.U.; investigation, C.M.; resources, M.S.B.; data curation, C.M.; writing—original draft preparation, M.S.B.; writing—review and editing, C.M.; visualization, C.M.; supervision, V.F.U.; project administration, V.F.U.; funding acquisition, V.F.U. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Li, Y.E.; O'Malley, D.; Beroza, G.; Curtis, A.; Johnson, P. Machine Learning Developments and Applications in Solid-Earth Geosciences: Fad or Future? *J. Geophys. Res. Solid Earth* **2023**, *128*, e2022JB026310. [[CrossRef](#)]
2. Sören, J.; Fontoura do Rosário, Y.; Fafoutis, X. Machine Learning in Geoscience Applications of Deep Neural Networks in 4D Seismic Data Analysis. Ph.D. Thesis, Technical University of Denmark, Kongens Lyngby, Denmark, 2020.
3. Bhattacharya, S. Summarized Applications of Machine Learning in Subsurface Geosciences. In *A Primer on Machine Learning in Subsurface Geosciences*; SpringerBriefs in Petroleum Geoscience & Engineering; Springer: Berlin/Heidelberg, Germany, 2021; pp. 123–165. [[CrossRef](#)]
4. Zhang, W.; Gu, X.; Tang, L.; Yin, Y.; Liu, D.; Zhang, Y. Application of machine learning, deep learning and optimization algorithms in geoengineering and geoscience: Comprehensive review and future challenge. *Gondwana Res.* **2022**, *109*, 1–17. [[CrossRef](#)]
5. Fradkov, A.L. Early History of Machine Learning. *IFAC-PapersOnLine* **2020**, *53*, 1385–1390. [[CrossRef](#)]
6. Nilsson, N.J. *The Quest for Artificial Intelligence: A History of Ideas and Achievements*; Cambridge University Press: Cambridge, UK, 2011; pp. 1–562. [[CrossRef](#)]
7. Zhong, S.; Zhang, K.; Bagheri, M.; Burken, J.G.; Gu, A.; Li, B.; Ma, X.; Marrone, B.L.; Ren, Z.J.; Schrier, J.; et al. Machine Learning: New Ideas and Tools in Environmental Science and Engineering. *Environ. Sci. Technol.* **2021**, *55*, 12741–12754. [[CrossRef](#)]
8. Clarivate—Data, Insights and Analytics for the Innovation Lifecycle. Available online: <https://clarivate.com/> (accessed on 1 February 2024).
9. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]

10. Biau, G.; Scornet, E. A random forest guided tour. *Test* **2016**, *25*, 197–227. [[CrossRef](#)]
11. Scornet, E.; Biau, G.; Vert, J.P. Consistency of random forests. *Ann. Statist.* **2015**, *43*, 1716–1741. [[CrossRef](#)]
12. Suthaharan, S. Support Vector Machine. In *Machine Learning Models and Algorithms for Big Data Classification*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 36, pp. 207–235. [[CrossRef](#)]
13. Schölkopf, B. SVMs—A practical consequence of learning theory. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–21. [[CrossRef](#)]
14. Bisong, E. Logistic Regression. *Build. Mach. Learn. Deep Learn. Model. Google Cloud Platf.* **2019**, *1*, 243–250. [[CrossRef](#)]
15. Domínguez-Almendros, S.; Benítez-Parejo, N.; Gonzalez-Ramirez, A.R. Logistic regression models. *Allergol. Immunopathol.* **2011**, *39*, 295–305. [[CrossRef](#)]
16. Wang, S.-C. Artificial Neural Network. *Interdiscip. Comput. Java Program.* **2003**, *743*, 81–100. [[CrossRef](#)] [[PubMed](#)]
17. Zou, J.; Han, Y.; So, S.S. Overview of artificial neural networks. *Methods Mol. Biol.* **2008**, *458*, 15–23. [[CrossRef](#)] [[PubMed](#)]
18. Krogh, A. What are artificial neural networks? *Nat. Biotechnol.* **2008**, *26*, 195–197. [[CrossRef](#)] [[PubMed](#)]
19. Kingsford, C.; Salzberg, S.L. What are decision trees? *Nat. Biotechnol.* **2008**, *26*, 1011–1013. [[CrossRef](#)] [[PubMed](#)]
20. de Ville, B. Decision trees. *Wiley Interdiscip. Rev. Comput. Stat.* **2013**, *5*, 448–455. [[CrossRef](#)]
21. Zhang, Z. Introduction to machine learning: K-nearest neighbors. *Ann. Transl. Med.* **2016**, *4*, 218. [[CrossRef](#)] [[PubMed](#)]
22. Kramer, O. K-Nearest Neighbors. In *Dimensionality Reduction with Unsupervised Nearest Neighbors*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 51, pp. 13–23. [[CrossRef](#)]
23. Peterson, L.E. K-nearest neighbor. *Scholarpedia* **2009**, *4*, 1883. [[CrossRef](#)]
24. Bishop, C.M. Bayesian Neural Networks. *J. Braz. Comput. Soc.* **1997**, *4*, 61–68. [[CrossRef](#)]
25. Chang, D.T. Bayesian Neural Networks: Essentials. *arXiv* **2021**, arXiv:2106.13594.
26. Alshari, E.A.; Abdulkareem, M.B.; Gawali, B.W. Classification of land use/land cover using artificial intelligence (ANN-RF). *Front. Artif. Intell.* **2023**, *5*, 964279. [[CrossRef](#)]
27. Rolf, E.; Proctor, J.; Carleton, T.; Bolliger, I.; Shankar, V.; Ishihara, M.; Recht, B.; Hsiang, S. A generalizable and accessible approach to machine learning with global satellite imagery. *Nat. Commun.* **2021**, *12*, 4392. [[CrossRef](#)] [[PubMed](#)]
28. Hu, X.; Chu, L.; Pei, J.; Liu, W.; Bian, J. Model complexity of deep learning: A survey. *Knowl. Inf. Syst.* **2021**, *63*, 2585–2619. [[CrossRef](#)]
29. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 160. [[CrossRef](#)] [[PubMed](#)]
30. Krakauer, D.C. Unifying complexity science and machine learning. *Front. Complex Syst.* **2023**, *1*, 1235202. [[CrossRef](#)]
31. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed.; Independent Publishers Group: Chicago, IL, USA, 2022.
32. Robert, C. *Machine Learning, a Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2014; Volume 27, ISBN 9780262305242.
33. Abdalla, H.B. A brief survey on big data: Technologies, terminologies and data-intensive applications. *J. Big Data* **2022**, *9*, 107. [[CrossRef](#)] [[PubMed](#)]
34. Sabharwal, R.; Miah, S.J. A new theoretical understanding of big data analytics capabilities in organizations: A thematic analysis. *J. Big Data* **2021**, *8*, 159. [[CrossRef](#)]
35. Zhou, L.; Pan, S.; Wang, J.; Vasilakos, A.V. Machine learning on big data: Opportunities and challenges. *Neurocomputing* **2017**, *237*, 350–361. [[CrossRef](#)]
36. Karpatne, A.; Ebert-Uphoff, I.; Ravela, S.; Babaie, H.A.; Kumar, V. Machine Learning for the Geosciences: Challenges and Opportunities. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 1544–1554. [[CrossRef](#)]
37. He, H.; Ming, Z.; Zhang, J.; Wang, L.; Yang, R.; Chen, T.; Zhou, F. Robust Estimation of Landslide Displacement from Multi-temporal UAV Photogrammetry-Derived Point Clouds. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 6627–6641. [[CrossRef](#)]
38. Ragnoli, M.; Esposito, P.; Stornelli, V.; Barile, G.; Santis, E.D.; Sciarra, N. A LoRa-based Wireless Sensor Network monitoring system for urban areas subjected to landslide. In Proceedings of the 2023 8th International Conference on Cloud Computing and Internet of Things, Okinawa, Japan, 22–24 September 2023; pp. 91–97. [[CrossRef](#)]
39. Dramsch, J.S. 70 years of machine learning in geoscience in review. *Adv. Geophys.* **2020**, *61*, 1–55. [[CrossRef](#)]
40. Osisanwo, F.Y.; Akinsola JE, T.; Awodele, O.; Hinmikaiye, J.O.; Olakanmi, O.; Akinjobi, J. Supervised Machine Learning Algorithms: Classification and Comparison. *Int. J. Comput. Trends Technol.* **2017**, *48*, 128–138. [[CrossRef](#)]
41. Singh, A.; Thakur, N.; Sharma, A. A review of supervised machine learning algorithms. In Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016.
42. Abdolrasol, M.G.M.; Suhail Hussain, S.M.; Ustun, T.S.; Sarker, M.R.; Hannan, M.A.; Mohamed, R.; Ali, J.A.; Mekhilef, S.; Milad, A. Artificial neural networks based optimization techniques: A review. *Electronic* **2021**, *10*, 2689. [[CrossRef](#)]
43. Quinlan, J.R. Induction of Decision Trees. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
44. Cover, T.M.; Hart, P.E. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [[CrossRef](#)]
45. Chervonenkis, A.Y. Early history of support vector machines. *Empir. Inference Festschrift Honor Vladimir N. Vapnik* **2013**, 13–20. [[CrossRef](#)]
46. Tyagi, K.; Rane, C.; Sriram, R.; Manry, M. Unsupervised learning. *Artif. Intell. Mach. Learn. EDGE Comput.* **2022**, 33–52. [[CrossRef](#)]
47. Mahesh, B. Machine Learning Algorithms—A Review. *Int. J. Sci. Res.* **2020**, *9*, 381–386. [[CrossRef](#)]

48. Dabrowska, D.; Rykala, W.; Nourani, V. Causes, Types and Consequences of Municipal Waste Landfill Fires—Literature Review. *Sustainability* **2023**, *15*, 5713. [\[CrossRef\]](#)
49. Vaverková, M.D.; Maxianová, A.; Winkler, J.; Adamcová, D.; Podlasek, A. Environmental consequences and the role of illegal waste dumps and their impact on land degradation. *Land Use Policy* **2019**, *89*, 104234. [\[CrossRef\]](#)
50. Irvanian, A.; Ravari, S.O. Types of Contamination in Landfills and Effects on The Environment: A Review Study. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *614*, 012083. [\[CrossRef\]](#)
51. Ozbay, G.; Jones, M.; Gadde, M.; Isah, S.; Attarwala, T. Design and Operation of Effective Landfills with Minimal Effects on the Environment and Human Health. *J. Environ. Public Health* **2021**, *2021*, 6921607. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Massarelli, C. Fast detection of significantly transformed areas due to illegal waste burial with a procedure applicable to Landsat images. *Int. J. Remote Sens.* **2018**, *39*, 754–769. [\[CrossRef\]](#)
53. Al Awadh, M.; Mallick, J. A decision-making framework for landfill site selection in Saudi Arabia using explainable artificial intelligence and multi-criteria analysis. *Environ. Technol. Innov.* **2024**, *33*, 103464. [\[CrossRef\]](#)
54. Massarelli, C.; Muolo, M.R.; Uricchio, V.F.; Dongiovanni, N.; Palumbo, R. Improving environmental monitoring against the risk from uncontrolled abandonment of waste containing asbestos. The DroMEEP project. In *Geomatics Workbooks n° 12—“FOSS4G Europe Como 2015”*; FOSS4G Europe 2015: Como, Italy, 2015.
55. Massarelli, C.; Uricchio, V.F. The Contribution of Open Source Software in Identifying Environmental Crimes Caused by Illicit Waste Management in Urban Areas. *Urban Sci.* **2024**, *8*, 21. [\[CrossRef\]](#)
56. Geng, J.; Ding, Y.; Xie, W.; Fang, W.; Liu, M.; Ma, Z.; Yang, J.; Bi, J. An ensemble machine learning model to uncover potential sites of hazardous waste illegal dumping based on limited supervision experience. *Fundam. Res.* **2023**; *in press*. [\[CrossRef\]](#)
57. Massarelli, C.; Binetti, M.S.; Triozzi, M.; Uricchio, V.F. A First Step towards Developing a Decision Support System Based on the Integration of Environmental Monitoring Activities for Regional Water Resource Protection. *Hydrology* **2023**, *10*, 174. [\[CrossRef\]](#)
58. Shahab, S.; Anjum, M. Solid Waste Management Scenario in India and Illegal Dump Detection Using Deep Learning: An AI Approach towards the Sustainable Waste Management. *Sustainability* **2022**, *14*, 15896. [\[CrossRef\]](#)
59. Torres, R.N.; Fraternali, P. Learning to Identify Illegal Landfills through Scene Classification in Aerial Images. *Remote Sens.* **2021**, *13*, 4520. [\[CrossRef\]](#)
60. Torres, R.N.; Fraternali, P. AerialWaste dataset for landfill discovery in aerial and satellite images. *Sci. Data* **2023**, *10*, 63. [\[CrossRef\]](#)
61. Youme, O.; Bayet, T.; Dembele, J.M.; Cambier, C. Deep Learning and Remote Sensing: Detection of Dumping Waste Using UAV. *Procedia Comput. Sci.* **2021**, *185*, 361–369. [\[CrossRef\]](#)
62. Kazaryan, M.; Simonyan, A.; Simavoryan, S.; Ulitina, E.; Aramyan, R. Waste disposal facilities monitoring based on high-resolution information features of space images. *E3S Web Conf.* **2020**, *157*, 02029. [\[CrossRef\]](#)
63. De Carolis, B.; Ladogana, F.; MacChiarulo, N. YOLO TrashNet: Garbage Detection in Video Streams. In Proceedings of the 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), Bari, Italy, 27–29 May 2020. [\[CrossRef\]](#)
64. Larrea-Gallegos, G.; Kahhat, R.; Vázquez-Rowe, I.; Parodi, E. A machine learning approach to understand how accessibility influences alluvial gold mining expansion in the Peruvian Amazon. *Case Stud. Chem. Environ. Eng.* **2023**, *7*, 100353. [\[CrossRef\]](#)
65. Fernández-Alonso, D.; Fernández-Lozano, J.; García-Ordás, M.T. Convolutional neural networks for accurate identification of mining remains from UAV-derived images. *Appl. Intell.* **2023**, *53*, 30469–30481. [\[CrossRef\]](#)
66. Fissaha, Y.; Ikeda, H.; Toriya, H.; Adachi, T.; Kawamura, Y. Application of Bayesian Neural Network (BNN) for the Prediction of Blast-Induced Ground Vibration. *Appl. Sci.* **2023**, *13*, 3128. [\[CrossRef\]](#)
67. Liu, L.; Zhou, W.; Gutierrez, M. Mapping Tunneling-Induced Uneven Ground Subsidence Using Sentinel-1 SAR Interferometry: A Twin-Tunnel Case Study of Downtown Los Angeles, USA. *Remote Sens.* **2022**, *15*, 202. [\[CrossRef\]](#)
68. Anantrasirichai, N.; Biggs, J.; Kelevitz, K.; Sadeghi, Z.; Wright, T.; Thompson, J.; Achim, A.M.; Bull, D. Detecting Ground Deformation in the Built Environment Using Sparse Satellite InSAR Data with a Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 2940–2950. [\[CrossRef\]](#)
69. Moustafa, S.S.R.; Abdalzaher, M.S.; Yassien, M.H.; Wang, T.; Elwekeil, M.; Hafiez, H.E.A. Development of an Optimized Regression Model to Predict Blast-Driven Ground Vibrations. *IEEE Access* **2021**, *9*, 31826–31841. [\[CrossRef\]](#)
70. Beuzen, T.; Splinter, K. Machine learning and coastal processes. In *Sandy Beach Morphodynamics*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 689–710. [\[CrossRef\]](#)
71. Nordstrom, K.F. Coastal Dunes. In *Coastal Environments and Global Change*; Wiley: Hoboken, NJ, USA, 2015; pp. 178–193. [\[CrossRef\]](#)
72. Łabuz, T.A. Coastal Dunes: Changes of Their Perception and Environmental Management. In *Environmental Management and Governance. Coastal Research Library*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 8, pp. 323–410. [\[CrossRef\]](#)
73. Pinton, D.; Canestrelli, A.; Moon, R.; Wilkinson, B. Estimating Ground Elevation in Coastal Dunes from High-Resolution UAV-LIDAR Point Clouds and Photogrammetry. *Remote Sens.* **2023**, *15*, 226. [\[CrossRef\]](#)
74. Mckeehan, K.G.; Arbogast, A.F. The geography and progression of blowouts in the coastal dunes along the eastern shore of Lake Michigan since 1938. *Quat. Res.* **2023**, *115*, 25–45. [\[CrossRef\]](#)
75. Gonzalez-Moodie, B.; Daiek, S.; Lorenzo-Trueba, J.; Varde, A.S. Multispectral Drone Data Analysis on Coastal Dunes. In Proceedings of the Proceedings—2021 IEEE International Conference on Big Data, Big Data 2021, Orlando, FL, USA, 15–18 December 2021; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2021; pp. 5903–5905.

76. Mohammadpoor, M.; Eshghizadeh, M. Introducing an intelligent algorithm for extraction of sand dunes from Landsat satellite imagery in terrestrial and coastal environments. *J. Coast. Conserv.* **2021**, *25*, 3. [[CrossRef](#)]
77. Beuzen, T.; Goldstein, E.B.; Splinter, K.D. Ensemble models from machine learning: An example of wave runup and coastal dune erosion. *Nat. Hazards Earth Syst. Sci.* **2019**, *19*, 2295–2309. [[CrossRef](#)]
78. Liao, L.; Zhao, Q.; Song, W. Monitoring of Oil Spill Risk in Coastal Areas Based on Polarimetric SAR Satellite Images and Deep Learning Theory. *Sustainability* **2023**, *15*, 14504. [[CrossRef](#)]
79. Magri, S.; Ottaviani, E.; Prampolini, E.; Federici, B.; Besio, G.; Fabiano, B. Application of machine learning techniques to derive sea water turbidity from Sentinel-2 imagery. *Remote Sens. Appl. Soc. Environ.* **2023**, *30*, 100951. [[CrossRef](#)]
80. Mohsen, A.; Kovács, F.; Kiss, T. Remote Sensing of Sediment Discharge in Rivers Using Sentinel-2 Images and Machine-Learning Algorithms. *Hydrology* **2022**, *9*, 88. [[CrossRef](#)]
81. Thanh, H.V.; Van Binh, D.; Kantoush, S.A.; Nourani, V.; Saber, M.; Lee, K.K.; Sumi, T. Reconstructing Daily Discharge in a Megadelta Using Machine Learning Techniques. *Water Resour. Res.* **2022**, *58*, e2021WR031048. [[CrossRef](#)]
82. Granata, F.; Saroli, M.; De Marinis, G.; Gargano, R. Machine learning models for spring discharge forecasting. *Geofluids* **2018**, *2018*, 8328167. [[CrossRef](#)]
83. 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](#)]
84. Huang, R.; Ma, C.; Ma, J.; Huangfu, X.; He, Q. Machine learning in natural and engineered water systems. *Water Res.* **2021**, *205*, 117666. [[CrossRef](#)] [[PubMed](#)]
85. Lowe, M.; Qin, R.; Mao, X. A Review on Machine Learning, Artificial Intelligence, and Smart Technology in Water Treatment and Monitoring. *Water* **2022**, *14*, 1384. [[CrossRef](#)]
86. Zhu, Y.; Yeung, C.H.; Lam, E.Y. Microplastic pollution monitoring with holographic classification and deep learning. *J. Phys. Photonics* **2021**, *3*, 024013. [[CrossRef](#)]
87. Huang, H.; Zhang, J. Prediction of chlorophyll a and risk assessment of water blooms in Poyang Lake based on a machine learning method. *Environ. Pollut.* **2024**, *347*, 123501. [[CrossRef](#)] [[PubMed](#)]
88. Chen, P.; Wang, B.; Wu, Y.; Wang, Q.; Huang, Z.; Wang, C. Urban river water quality monitoring based on self-optimizing machine learning method using multi-source remote sensing data. *Ecol. Indic.* **2023**, *146*, 109750. [[CrossRef](#)]
89. Zhi, W.; Feng, D.; Tsai, W.P.; Sterle, G.; Harpold, A.; Shen, C.; Li, L. From hydrometeorology to river water quality: Can a deep learning model predict dissolved oxygen at the continental scale? *Environ. Sci. Technol.* **2021**, *55*, 2357–2368. [[CrossRef](#)] [[PubMed](#)]
90. Chen, K.; Chen, H.; Zhou, C.; Huang, Y.; Qi, X.; Shen, R.; Liu, F.; Zuo, M.; Zou, X.; Wang, J.; et al. Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Res.* **2020**, *171*, 115454. [[CrossRef](#)] [[PubMed](#)]
91. Wu, Y.; Zhang, X.; Xiao, Y.; Feng, J. Attention Neural Network for Water Image Classification under IoT Environment. *Appl. Sci.* **2020**, *10*, 909. [[CrossRef](#)]
92. Kumar, V.; Sharma, A.; Kumar, R.; Bhardwaj, R.; Kumar Thukral, A.; Rodrigo-Comino, J. Assessment of heavy-metal pollution in three different Indian water bodies by combination of multivariate analysis and water pollution indices. *Hum. Ecol. Risk Assess. Int. J.* **2020**, *26*, 1–16. [[CrossRef](#)]
93. Zhao, W.; Ma, J.; Liu, Q.; Dou, L.; Qu, Y.; Shi, H.; Sun, Y.; Chen, H.; Tian, Y.; Wu, F. Accurate Prediction of Soil Heavy Metal Pollution Using an Improved Machine Learning Method: A Case Study in the Pearl River Delta, China. *Environ. Sci. Technol.* **2023**, *57*, 17751–17761. [[CrossRef](#)] [[PubMed](#)]
94. Azizi, K.; Ayoubi, S.; Nabiollahi, K.; Garosi, Y.; Gislum, R. Predicting heavy metal contents by applying machine learning approaches and environmental covariates in west of Iran. *J. Geochemical Explor.* **2022**, *233*, 106921. [[CrossRef](#)]
95. Yang, S.; Taylor, D.; Yang, D.; He, M.; Liu, X.; Xu, J. A synthesis framework using machine learning and spatial bivariate analysis to identify drivers and hotspots of heavy metal pollution of agricultural soils. *Environ. Pollut.* **2021**, *287*, 117611. [[CrossRef](#)] [[PubMed](#)]
96. Jia, X.; Cao, Y.; O'Connor, D.; Zhu, J.; Tsang, D.C.W.; Zou, B.; Hou, D. Mapping soil pollution by using drone image recognition and machine learning at an arsenic-contaminated agricultural field. *Environ. Pollut.* **2021**, *270*, 116281. [[CrossRef](#)]
97. Jia, X.; O'Connor, D.; Shi, Z.; Hou, D. VIRS based detection in combination with machine learning for mapping soil pollution. *Environ. Pollut.* **2021**, *268*, 115845. [[CrossRef](#)] [[PubMed](#)]
98. Zhang, H.; Yin, S.; Chen, Y.; Shao, S.; Wu, J.; Fan, M.; Chen, F.; Gao, C. Machine learning-based source identification and spatial prediction of heavy metals in soil in a rapid urbanization area, eastern China. *J. Clean. Prod.* **2020**, *273*, 122858. [[CrossRef](#)]
99. Chen, H.; Chen, A.; Xu, L.; Xie, H.; Qiao, H.; Lin, Q.; Cai, K. A deep learning CNN architecture applied in smart near-infrared analysis of water pollution for agricultural irrigation resources. *Agric. Water Manag.* **2020**, *240*, 106303. [[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.