



A REVIEW OF MACHINE LEARNING TECHNIQUES APPLICATIONS IN ENVIRONMENTAL SCIENCE

EGWU L.S.¹, ENAYABA, O.F.², AJIBOYE, A.A.³
DAMOYE T⁴ OGUNDEJI, I. S⁵ AND AGBAMS P⁶.

¹Department of Animal and Environmental Biology, Prince Abubakar Audu University, Anyigba, Nigeria. ²Department of Chemistry, Federal University of Petroleum Resources, Effurun, Nigeria. ³Department of Mathematics, Purdue University. ⁴Department of Computer Science, Babcock University, Nigeria. ⁵Department of Data Science Management, Rome Business School, Italy. ⁶Department of Agricultural Economics and Extension, Abia State University, Uturu, Nigeria.

Corresponding Author: egwulinus364@gmail.com

DOI: <https://doi.org/10.70382/tijsrat.v06i9.002>

ABSTRACT

Advances in Machine Learning (ML) and Data Science (DS) hold immense potential to transform various aspects of environmental science (ES). DS, a broad field focused on extracting insights from data using techniques like statistical analysis plays a crucial role in this process. Machine learning, on the other hand, specializes in creating algorithms that

INTRODUCTION

Rapid advancements in machine learning (Haupt *et al.*, 2022) and deep learning (LeCun *et al.*, 2015) have sparked the scientific community to explore how these tools can drive scientific progress and unlocking breakthroughs that were once considered unattainable. Over decades, environmental science only revolves round how physical and chemical properties alongside with natural resources gears interactions of living organisms and the environment. In recent time, environmental data are rapidly growing into huge datasets, increasing complexity, resolution, and size. This growth creates interdisciplinary challenges for environmental scientists, requiring innovative approaches, such as data



enable computers to learn from data and make predictions. Together, these technologies can deepen our understanding of complex environmental systems, refine predictive models for climate change, support conservation efforts, and optimize resource management practices. Such scientific discovery will enhance ES to make autonomous, real-time decisions by deriving valuable insights from extensive data. By analysing large datasets, machine learning algorithms can reveal hidden patterns and insights, empowering scientists to make data-driven decisions and tackle environmental challenges more effectively. This article offers a review of the fundamental concepts of Machine Learning, Deep Learning, and Data Analytics for two groups: individuals familiar with ML who seek to expand their knowledge, and domain scientists passionate about integrating these transformative tools into their research in the environmental science profession.

Keywords: Machine Learning, Environmental Science, Climate Change, Data Science and Algorithm

processing and big data analysis, to address them effectively. Integrating diverse data from multiple sources to perform comprehensive analysis and extract meaningful insights demands a strong foundation in data science. The widespread adoption of data science techniques has greatly enhanced environmental system management, enabled scenario modelling and fostered data-driven innovation across industries. Environmental scientists are increasingly challenged to solve complex interdisciplinary problems through established and emerging data science methods ((Tharsanee *et al.*, 2020). Data science enriches environmental science by providing a practical and effective approach to tackle real-world issues (Karina, *et al.*, 2018).

To gain meaningful insights, data from environmental stressors—collected through remote sensing satellites, air and water quality sensors, weather and climate observations, and ground-based sensors that measure the magnitude of earthquakes and other geological events—must be thoroughly and efficiently analysed (Tharsanee *et al.*, 2020). Recent advancements in machine learning have empowered scientists and software engineers to address complex issues in climate variability and weather, fuelling momentum for national and international workshops (Chantry *et al.*, 2021). A key



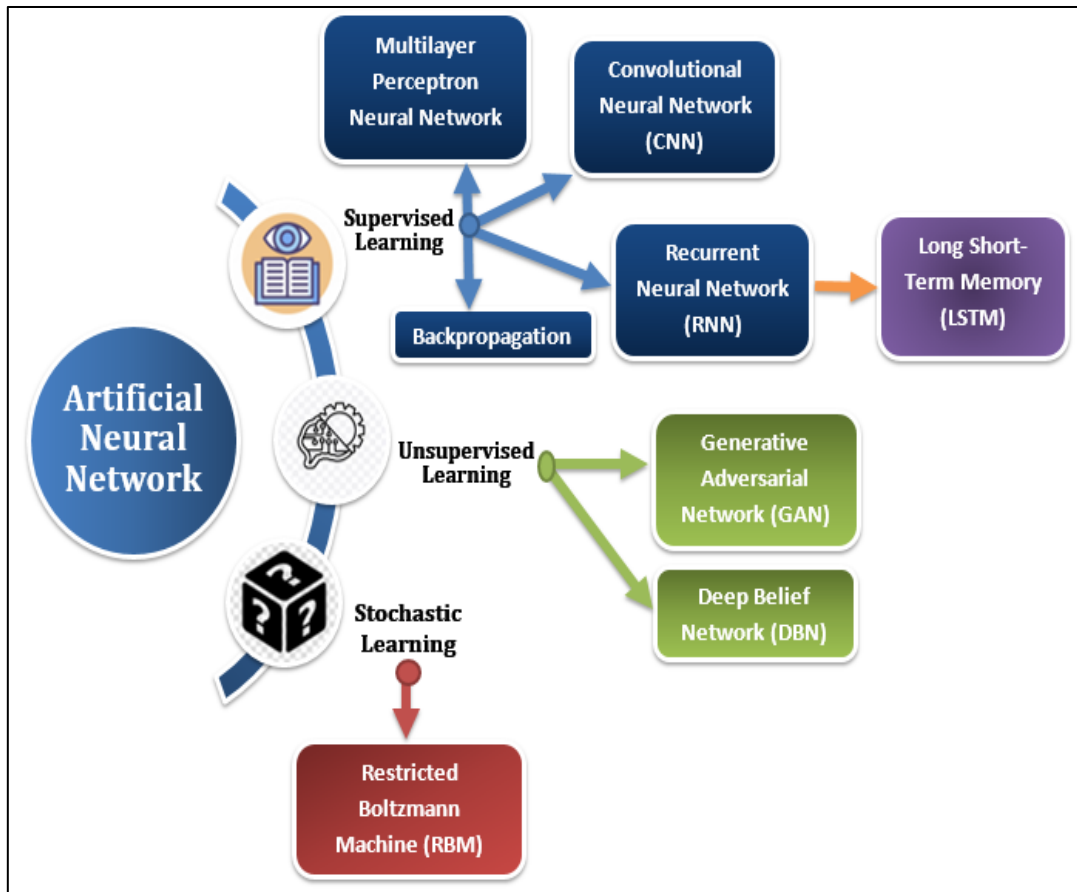
advantage of machine learning algorithms is their ability to identify trends and patterns in data autonomously. Their predictive accuracy improves as more data becomes available. Machine learning can also handle multidimensional and diverse data, even in dynamic or uncertain environments (Loussaief *et al.*, 2016; Ahn *et al.* 2016 & Brunton *et al.* 2019).

Although ML and DS has record initial success in Environmental Science, several challenges still persist. Foremost, many ES researchers are eager to adopt these techniques but may lack the necessary expertise to apply it correctly, leading to potential misuse of the technology. Additionally, as data volume and complexity have increased, more advanced ML applications, such as deep neural networks, are being utilized to capture complex nonlinear relationships. Lastly, Applicability domain analysis of ML models is still not commonly practiced by researchers in Environmental and Science Engineering after model development, except in the case of quantitative structure–activity relationships (Gadaleta *et al.*, 2016). However, these models are often considered "black boxes," making model interpretability essential to ensure that the predictions align with core domain scientific principles. Despite growing attention to model interpretability, it is still often overlooked in ES researches (Kerckhoffs *et al.*, 2019; Song *et al.*, 2017; Pak *et al.*, 2020 & Xiao 2018).

According to Hsieh W. (2009), ML, which originates from Artificial Intelligence, has become a cutting-edge approach in data mining with significant and future potential in environmental science. These methods are used to process satellite data, predict climate trends, forecast outcomes, and analyse environmental datasets. To derive meaningful insights from this data, a modern and effective approach is required, incorporating techniques such as linear statistical analysis, time series analysis, feedforward neural networks, nonlinear optimization, generalization learning, classification models, regression models, principal component analysis, and correlation analysis. Together, these models provide an optimal framework for tackling a wide range of challenges in environmental science. This review seeks to outline future directions in ML and ES that we believe will significantly advance the field, along with brief explanations of various machine learning techniques and deep learning algorithms for their application in environmental data analysis.

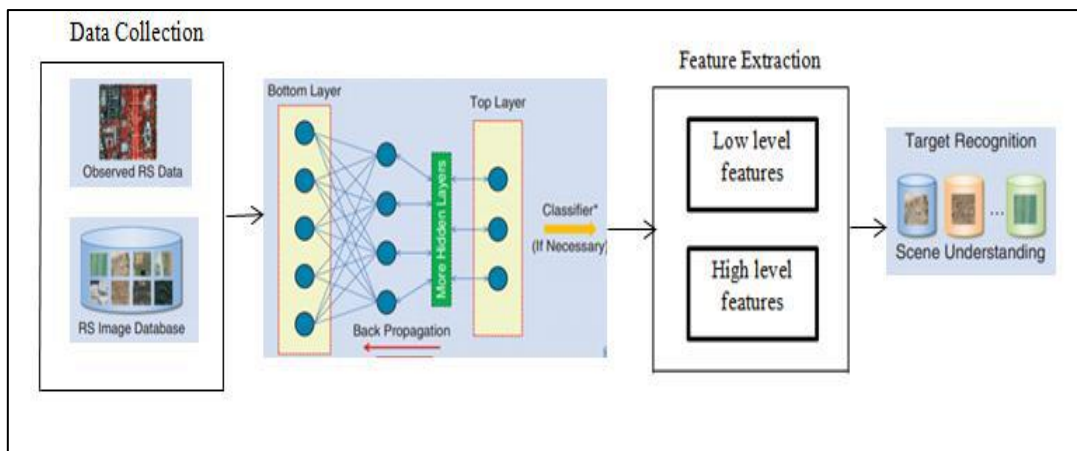
Fig 1: Graphical illustration of Deep learning





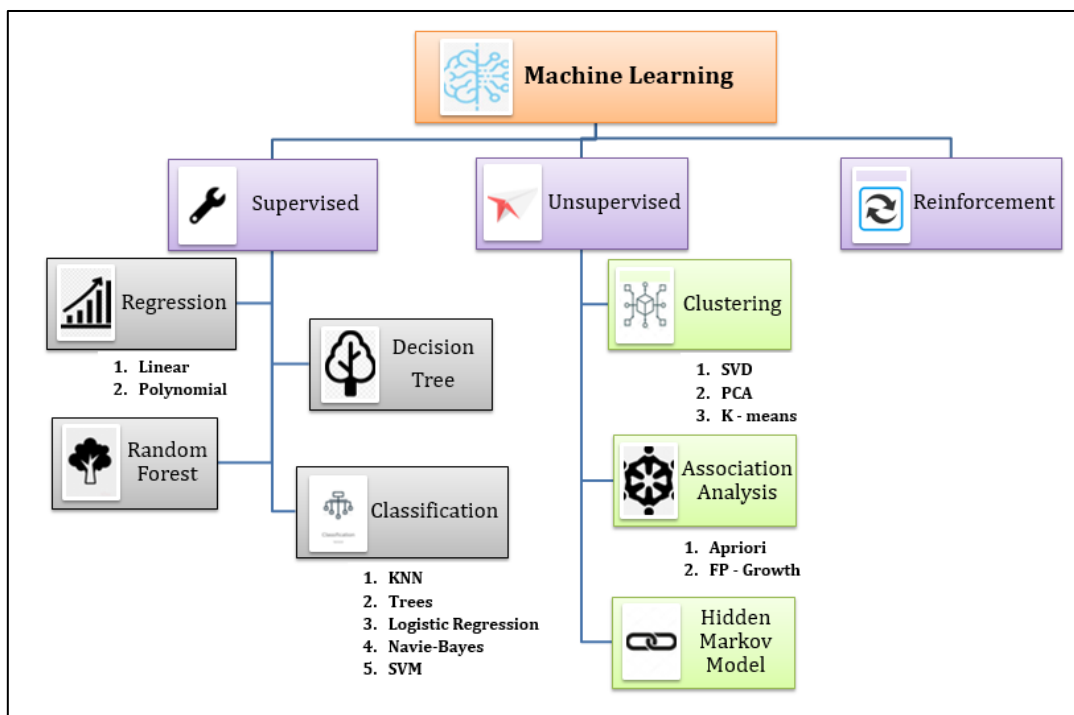
Different neural Network framework (Tharsanee et al., 2020)

Fig 2: A Graphical Method for Machine Learning in Remote Sensing



A deep learning approach for analysing remote sensing data (Zhang et al., 2016)

Fig 3: Graphical Summary of ML Algorithms



Various ML Algorithms (Tharsanee et al., 2020)

Brief Overview of ML Challenges in ES

Machine learning offers significant potential in environmental science, particularly in data analysis, climate prediction, ecosystem modelling, and resource management. However, its application faces several challenges;

Data Quality and Availability: Environmental data can be sparse or incomplete, making it difficult to train accurate ML models. In some cases, data may not be standardized or may have gaps that require imputation or preprocessing. To effectively collect valid and high-quality data is the foremost issues with the application of ML in the field of ES. Low-quality data are not easily suitable for modelling. On the other hand, machine learning demands a large dataset to create robust predictive models and ensure accurate predictions (L'heureux et al., 2017). Researchers also contribute to the aforementioned challenges by often conducting experiments under different conditions, such as variations in water quality, soil or sediment properties, and catalyst or adsorbent types and loadings, even when data are available. This leads to discrepancies in the collected data. To develop predictive models that are truly robust and widely applicable, it is crucial to first build a large, consistent dataset (Shifa et al., 2021)

Model Interpretability vs Overfitting: Many ML models, particularly deep learning algorithms, operate as "black boxes," making it challenging to interpret how they arrive



at specific predictions. This lack of transparency can be problematic in environmental science, where understanding the rationale behind a model's decision is crucial. Model overfitting is a major issue in large language model development, as it affects interpretability. Overfitting occurs when a model learns the noise in the training data instead of the true patterns, leading to poor performance on unseen data and making the model's decision-making hard to understand. It results in an overly complex model that is difficult to interpret, due to increased complexity, difficulty with feature selection, reduced generalizability, and challenges in model explanation. Detecting overfitting is challenging, but methods such as feature selection can help reduce the risk of overfitting, data augmentation, cross-validation, regularization, model simplification or choice, dropout, and early stopping (Kuhn *et al.*, 2019; Bühlmann *et al.*, 2011; Srivastava, *et al.*, 2014; Yao *et al.*, 2007)

Data Underfitting: Underfitting in ML occurs when a model is too simple to capture the underlying patterns or relationships in the data. It typically happens when the model has insufficient complexity, such as having too few features, overly simplistic assumptions, or inadequate model architecture. As a result, the model fails to learn important trends in the data, leading to poor performance both on the training set and unseen data.

Key causes of underfitting include:

- Insufficient model complexity: The model lacks the capacity to learn the data's patterns.
- Inadequate feature selection: Not enough relevant features are included in the model.
- Excessive regularization: Too much regularization can prevent the model from learning sufficiently complex patterns.

To address this issue, the following factors must be taken into account:

- Increase the training time
- Boost the model's complexity
- Add more relevant features to the data
- Reduce regularization parameters
- Extend the model's training duration

Complexity of the machine learning process: Machine learning is a complex process that involves multiple stages, including data collection, preprocessing, model selection, training, evaluation, and deployment. Each stage requires careful attention to detail and can introduce challenges that affect the model's performance. The complexity arises from



the need to manage large volumes of data, select appropriate algorithms, fine-tune model parameters, and ensure generalization to unseen data. Additionally, factors such as overfitting, underfitting, model interpretability, and computational resources all contribute to the intricacy of the machine learning process

Bias in machine learning Models: From a technical perspective, bias refers to systematic errors introduced by the model or the data that lead to unfair, skewed, or inaccurate predictions. Bias can arise at various stages of the machine learning process, from data collection to model deployment, and can significantly impact the model's performance and fairness. Data contamination is a more subtle type of data leakage and can be difficult to identify without domain expertise. Data omission is a prevalent issue in scientific communities, as peer-reviewed publications often highlight only the most promising positive results, leaving out negative results and outliers that are essential for ML model performance. Additionally, data may be missing due to choices made by scientists, such as the use of specific reagents, reaction conditions, or sampling plans, or the failure to collect data that contradicts established theories. These types of anthropogenic biases in datasets can further degrade ML performance (Charidimou *et al.*, 2019). Bias of Algorithmic arises when the model's structure or loss function does not align with the intended use case. Identifying potential bias in an ML model early is critical for its successful application, beyond just environmental concerns. Mitigating bias can be achieved by enhancing the model's interpretability, allowing for the integration of domain knowledge to assess its validity. A practical strategy for detecting bias is to use an ensemble of ML models, comparing their outputs on the same set of problems. This comparison helps identify inconsistencies in performance and reveals any bias specific to a particular model (Shifa *et al.*, 2021).

Algorithmic Flaws as Data Expands: As data continues to grow, algorithms may become outdated in the future. Current models, which are considered the best, may become inaccurate and will require adjustments. Maintaining algorithms requires constant monitoring and upkeep. According to the report by Shifa *et al.* (2021), three main concerns were highlighted: (1) complete reliance on ML should be avoided, as traditional statistical tools may be more suitable in certain cases, such as when sample sizes are small; (2) it is crucial to investigate findings through experimentation or domain expertise, rather than overestimating the capabilities of ML techniques (3) Always keep in mind that not every ESE problem can be directly solved using ML tools. Transforming these problems into ones that can be effectively addressed by ML requires skilful and thoughtful design.

Applications and Implementations of ML



Machine learning has been widely applied across multiple areas of environmental monitoring and management. These are illustrated below;

ML application into Environmental Impact Assessments

Integrating Machine Learning (ML) with Environmental Impact Assessments (EIAs) raises various ethical considerations that must be carefully addressed to ensure the fair and responsible use of technology in environmental science. A significant ethical concern is algorithmic bias, which can arise when ML models reinforce or amplify existing environmental inequalities. To address this issue, it is essential to use interpretable models like decision trees and to develop interpretability techniques that are independent of the specific model type. Embedding ethical considerations into the core design of machine learning applications lays the groundwork for responsible and sustainable technology use that fosters environmental stewardship (Yilmaz *et al.*, 2019 & Obulesu *et al.*, 2024)

ML in Agriculture: Machine learning (ML) has transformed agriculture by providing data-driven insights and automation that improve productivity, efficiency, and sustainability. Machine learning is transforming agriculture by providing tools for forecasting crop yields, optimizing farming conditions using historical data and future trends, and enabling accurate predictions and strategic planning (Sharma *et al.*, 2020). These approaches not only promote sustainable agricultural production but also improve resource efficiency, such as water and fertilizer usage, supporting environmental conservation efforts (Benos *et al.*, 2021). For example, machine learning algorithms have been applied to refine fertilizer usage, achieving a balance between enhancing crop yields and reducing environmental harm (Peng, *et al.*, 2023). Furthermore, machine learning plays a vital role in assessing soil health by analysing key soil properties to guide crop rotation and soil management practices, ensuring soil fertility and supporting healthy crop production (Mohamed *et al.*, 2023).

The effective implementation of these techniques has ushered in a new era of precision agriculture, greatly enhancing crop management and boosting yields. Research efforts such as (Maheswari *et al.*, 2023 & Kuradusenge *et al.*, 2023) emphasize the precision of machine learning models in predicting crop yields, enabling farmers and stakeholders to make well-informed decisions. The application of machine learning in disease management, as illustrated by (Prem *et al.*, 2018 & Sai *et al.*, 2023), highlights the effectiveness of convolutional neural networks (CNNs) in early detection of plant diseases, helping to minimize losses and decrease reliance on chemical pesticides. Advancements in soil analysis using machine learning, as highlighted in (Prem *et al.*, 2018;

Sai et al., 2023 & Sirsat et al., 2018), have facilitated more accurate evaluations of soil health, enhanced nutrient management and promoting soil conservation. The emergence of smart agricultural tools, ranging from plant classification to soil erosion modelling, underscores the transformative impact of these technologies (Sai et al., 2023 & Elavarasan et al., 2018). These examples emphasize the broad use of these technologies in agriculture, spanning from disease detection and yield prediction to automating harvesting and orchard navigation. By utilizing advanced algorithms, the agricultural sector can greatly boost productivity, sustainability, and its ability to adapt to climate change challenges

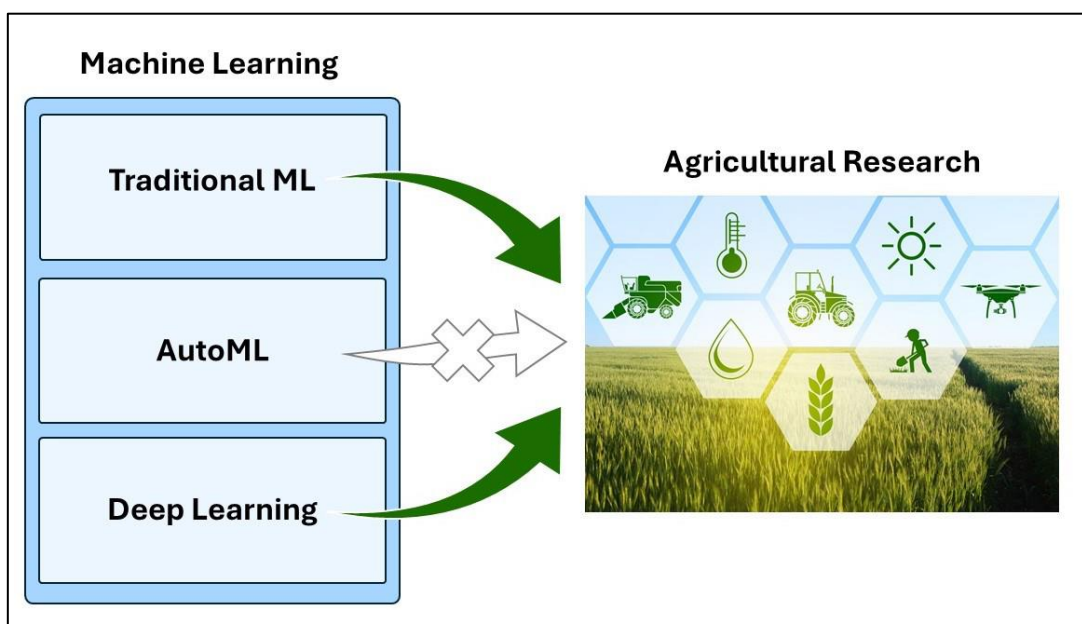


Fig 4: An illustration demonstrating the use of machine learning techniques in agricultural research, emphasizing the application of traditional ML and deep learning methods, while acknowledging the absence of AutoML implementations. (Dania et al., 2024)

Table 1 (Dania et al., 2024). Provide an overview of the applications of ML and Deep Learning techniques in agriculture

ML Technique	Agricultural Application
Decision Tree	Prediction of Crop Yields, Disease Identification, Soil Analysis
Random Forest	Prediction of Crop Yields, Disease Identification, Soil Analysis



Extreme Gradient Boosting	Crop Yield Estimation, Soil Health Assessment
Naive Bayes	Crop Yield Estimation, Soil Health Assessment
K-Nearest Neighbors	Crop Yield Prediction, Disease Diagnosis
Ensemble ML Models	Crop Yield Prediction
Multi-Linear Regressor	Crop Yield Prediction
RNN	Crop Yield Prediction
LSTM	Crop Yield Prediction
Support Vector Regression	Crop Yield Prediction
CNN	Crop Yield Prediction, Disease Detection
GNN	Crop Yield Prediction
U-Net	Crop Yield Prediction
ANN	Crop Yield Prediction, Disease Detection
DBSCAN	Crop Yield Prediction
Support Vector Machine	Crop Yield Prediction, Disease Detection, Smart Farming
Vision Transformers	Disease Detection
VGG-RNN Hybrid	Soil Assessment
MLP	Soil Assessment

Application of ML Techniques in Wildfire science and Management

ML, a subset of Artificial intelligence has various methods that are used for wildfire science and management. These approaches are grouped into three: supervised learning, unsupervised learning, or agent-based learning which are briefly elucidate below.

Supervised Learning: in this method, the model is trained on labelled data. This means the input data is paired with corresponding correct output labels. The algorithm learns the relationship between the input and output during training, with the goal of making accurate predictions or classifications when new, unseen data is provided. Supervised learning methods are typically used for tasks like regression (predicting continuous values) and classification (predicting discrete categories). Examples of supervised learning algorithms include linear regression, decision trees, and support vector machines (SVM)

Unsupervised Learning: Unsupervised learning is a type of machine learning where the model is trained on data without labelled responses. In this approach, the algorithm tries to identify patterns and structures in the data on its own, such as grouping similar data points (clustering) or reducing the dimensionality of the data (dimensionality

reduction). In unsupervised learning, the canonical tasks are dimensionality reduction and clustering, with relationships or patterns being extracted from the data without any guidance as to the “correct” answer.

Agent Based Learning: In this approach, an "agent" is a computational entity that makes decisions based on its observations of the environment and learns from the consequences of its actions. The agent's goal is typically to optimize its performance or achieve specific objectives, often through trial and error. This method utilizes incomplete information about the target variables. Major problem with this approach is reinforcement in which critical parts of the environment can only be observed interactively through trial and error (Sutton R. *et al.*, 1998)

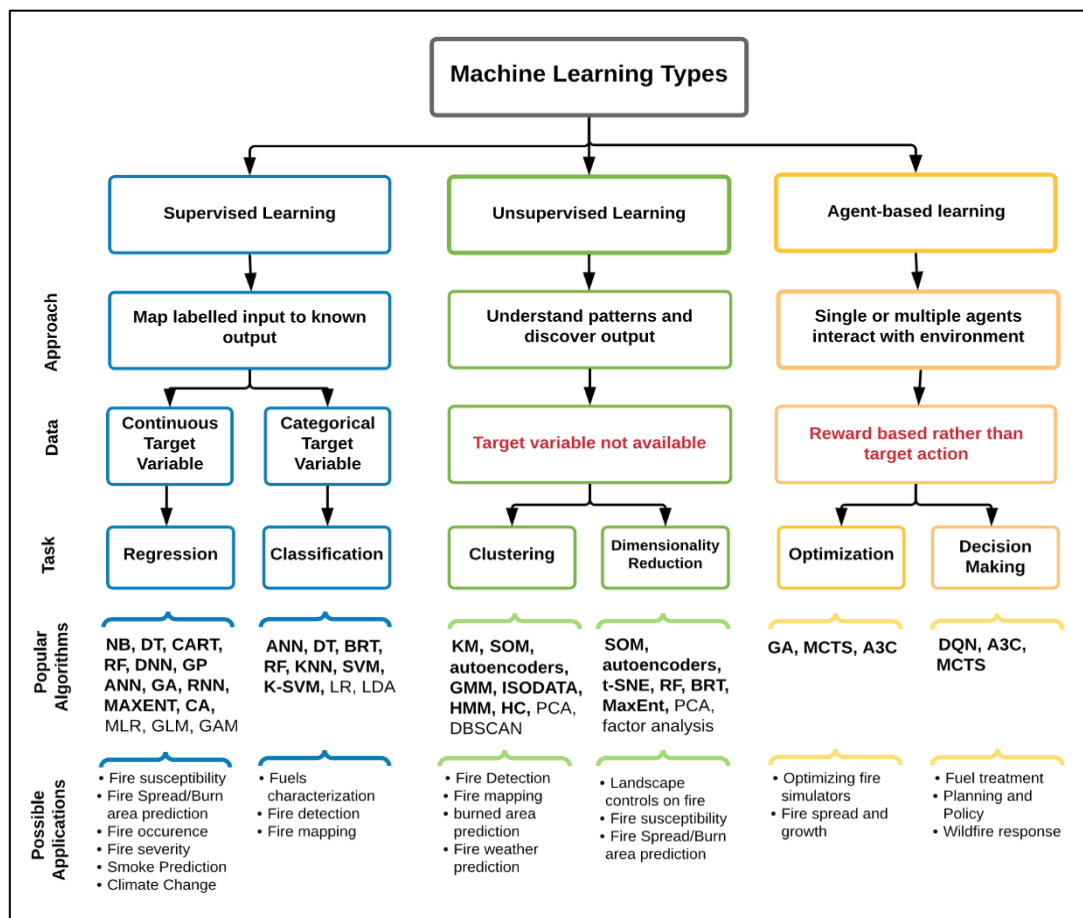


Fig 5: diagram depicting the primary types of machine learning, data types, and modeling tasks, highlighting their associations with widely used algorithms and applications in wildfire science and management. Algorithms in bold indicate core ML methods, whereas non-bolded algorithms are generally not classified as core ML (Piyush *et al.*, 2020).

ML application on Water Bodies: The rapid growth of artificial intelligence and the increasing volume of data on aquatic environments have made machine learning a vital tool for data analysis, classification, and prediction. Machine learning is a powerful tool increasingly utilized by environmental science researchers to tackle challenges in water treatment and management systems. Its applications span across water resource allocation, pollutant source tracking, real-time monitoring, prediction, pollutant concentration estimation, and the optimization of water treatment technologies (Mengyuan et al., 2022).

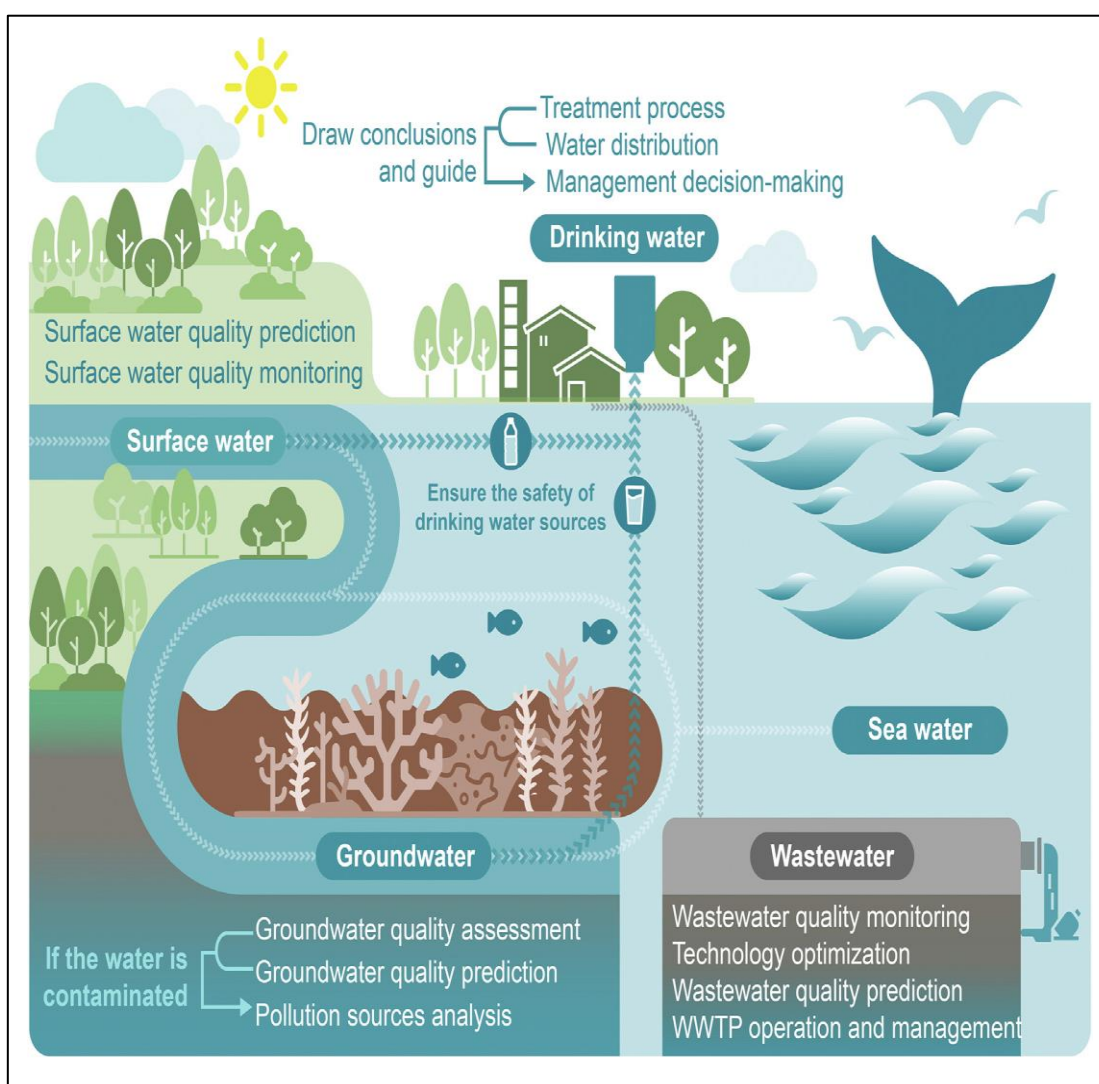


Fig. 1. Machine learning is extensively applied in water systems. WWTP: wastewater treatment plant (Mengyuan et al., 2022).

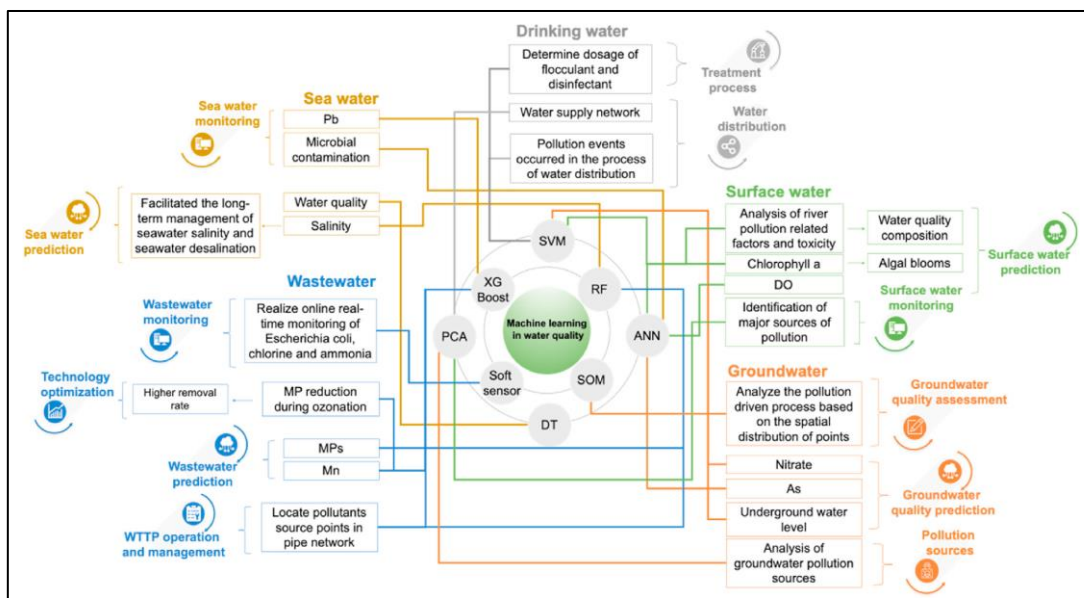


Fig: Machine learning algorithms applied across various water treatment and management systems. support vector machine, random forest, artificial neural network; SOM: self-organizing map, decision tree, principal component analysis; XGBoost: extreme gradient boosting, dissolved oxygen, micropollutant (Mengyuan et al., 2022)

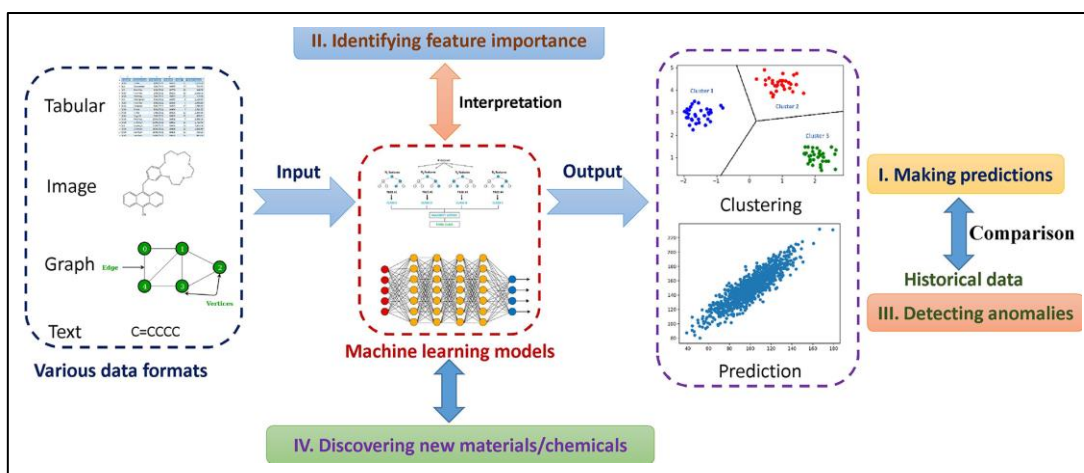


Fig Pictorial representation of four ML applications in Environmental Science and Engineering (ESE) (Shifa et al., 2021)

Methodology

We adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines outlined by Liberati et al. (2009) to identify relevant articles and



conduct the review. Our search strategy involved using Google to access the "Web of Science" and "ScienceDirect" databases, employing various combinations of relevant keywords such as machine learning, environmental science and climate change.

Authors contribution

This study was conducted through collaboration among all the authors

Declaration of Competing Interests

The authors have declared no existing competing interests

Concluding remark

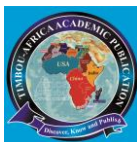
This paper has highlighted the efficacy of machine learning (ML) and its transformative role for environmental professionals in environmental science, showcasing how it continues to grow as a solution for addressing environmental issues. The importance of ML in this field cannot be overemphasized due to its futuristic potential. Our findings emphasize the widespread application of this critical technology, spanning traditional ML techniques to advanced approaches. Additionally, the incorporation of Automated Machine Learning (AutoML) offers significant, yet largely untapped, potential.

Despite its promise, the adoption of ML in environmental science faces challenges. Addressing these limitations will require greater collaboration among data scientists, environmental researchers, and policymakers to enhance model transparency and usability. By embracing these technologies, the environmental science community can contribute to more sustainable solutions for pressing global challenges.

Researchers eager to utilize these powerful tools should first master the fundamentals of their application to avoid discrepancies in findings. Finally, while ML is invaluable, it should not be solely relied upon. Traditional methods and experimental approaches must be integrated alongside ML to ensure accuracy and reliability in environmental research.

Reference

- Dania T., Xiuquan W., & Morteza M (2024). A Review of Machine Learning Techniques in Agroclimatic Studies. 14(3), 481; <https://doi.org/10.3390/agriculture14030481>
- Mengyuan Zhu, Jiawei Wang, Xiao Yang, Yu Zhang, Linyu Zhang, Hongqiang Ren, Bing Wu, Lin Ye (2022). A review of the application of machine learning in water quality evaluation. *Eco-Environment & Health*; 107–116
- Charidimou, A.; Zonneveld, H. I.; Shams, S.; Kantarci, K.; Shoamanesh, A.; Hilal, S.; Yates, P. A.; Boulouis, G.; Na, H. K.; Pasi, M.; et al. APOE and cortical superficial siderosis in CAA: Metaanalysis and potential mechanisms. *Neurology* 2019, 93 (4), e358– e371
- Obulesu Varikunta, Mr. A. Sarveswara Reddy, Mr. K. Sathish (2024). Incorporating Machine Learning into Environmental Impact Assessments for Sustainable Development. <https://www.researchgate.net/publication/380464590>
- Sharma, R.; Kamble, S.S.; Gunasekaran, A.; Kumar, V.; Kumar, A. A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. *Comput. Oper. Res.* 2020, 119, 104926. [[Google Scholar](#)] [[CrossRef](#)]
- Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors* 2021, 21, 3758. [[Google Scholar](#)] [[CrossRef](#)] [[PubMed](#)]
- Peng, W.; Karimi Sadaghiani, O. A review on the applications of machine learning and deep learning in agriculture section for the production of crop biomass raw materials. *Energy Sources Part A Recover. Util. Environ. Eff.* 2023, 45, 9178–9201. [[Google Scholar](#)] [[CrossRef](#)]



- Mohamed, S.A.; Metwaly, M.M.; Metwalli, M.R.; AbdelRahman, M.A.E.; Badreldin, N. Integrating Active and Passive Remote Sensing Data for Mapping Soil Salinity Using Machine Learning and Feature Selection Approaches in Arid Regions. *Remote Sens.* 2023, 15, 1751. [Google Scholar] [CrossRef]
- Maheswari, M.U.; Ramani, R. A Comparative Study of Agricultural Crop Yield Prediction Using Machine Learning Techniques. In Proceedings of the 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 17–18 March 2023; IEEE: Piscataway, NJ, USA, 2023; Volume 1, pp. 1428–1433. [Google Scholar]
- Kuradusenge, M.; Hitimana, E.; Hanyurwimfura, D.; Rukundo, P.; Mtonga, K.; Mukasine, A.; Uwitonze, C.; Ngabonziza, J.; Uwamahoro, A. Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize. *Agriculture* 2023, 13, 225. [Google Scholar] [CrossRef]
- F. Yilmaz, “Performance and Environmental Impact Assessment of a geothermal-assisted combined plant for multi-generation products,” *Sustainable Energy Technologies and Assessments*, vol. 46, p. 101291, 2021. doi:10.1016/j.seta.2021.10129
- Haupt, S. E., and Coauthors, 2022: The history and practice of AI in the environmental sciences. *Bull. Amer. Meteor. Soc.*, 103, E1351–E1370, <https://doi.org/10.1175/BAMS-D-20-0234.1>
- LeCun, Y., Y. Bengio, and G. Hinton, 2015: Deep learning. *Nature*, 521, 436–444, <https://doi.org/10.1038/nature1>
- Shifa Zhong, Kai Zhang, Majid Bagheri, Joel G. Burken, April Gu, Baikun Li, Xingmao Ma, Babetta L. Marrone, Zhiyong Jason Ren, Joshua Schrier, Wei Shi, Haoyue Tan, Tianbao Wang, Xu Wang, Bryan M. Wong, Xusheng Xiao, Xiong Yu, Jun-Jie Zhu, and Huichun Zhang (2021). Machine Learning: New Ideas and Tools in Environmental Science and Engineering, <https://doi.org/10.1021/acs.est.1c01339>
- Karina Gibert, Jeffery S.Horsburgh, Ioannis N Athanasiadis and Geoff Holmes 2018 *Environmental Data Science*
- L’heureux, A.; Grolinger, K.; Elyamany, H. F.; Capretz, M. A. Machine learning with big data: Challenges and approaches. *IEEE Access* 2017, 5, 7776–7797
- Gadaleta, D.; Mangiatordi, G. F.; Catto, M.; Carotti, A.; Nicolotti, O. Applicability domain for QSAR models: where theory meets reality. *International Journal of Quantitative Structure-Property Relationships (IJQSPR)* 2016, 1 (1), 45–63.
- Tharsanee R.M., Soundariya R.S., Vishnupriya B. (2020). Machine Learning and Data Analytics for Environmental Science: A Review, Prospects and Challenges, doi:10.1088/1757-899X/955/1/012107
- Loussaief, S.; Abdelkrim, A. In Machine Learning Framework for Image Classification, 2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), 2016; IEEE, 2016; pp 58–61.
- Ahn, W.-Y.; Ramesh, D.; Moeller, F. G.; Vassileva, J. Utility of machine-learning approaches to identify behavioral markers for substance use disorders: impulsivity dimensions as predictors of current cocaine dependence. *Frontiers in Psychiatry* 2016, DOI: 10.3389/fpsy.2016.00034.
- Brunton, S. L.; Kutz, J. N. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*; Cambridge University Press, 2019
- Chantry, M., H. Christensen, P. Dueben, and T. Palmer, 2021: Opportunities and challenges for machine learning in weather and climate modelling: Hard, medium and soft AI. *Philos. Trans. Roy. Soc.*, A379, 20200083, <https://doi.org/10.1098/rsta.2020.0083>
- L. Zhang, L. Zhang and B. Du 2016 *IEEE Geoscience and Remote Sensing Magazine* 4(2) pp. 22-40.
- Kerckhoffs, J.; Hoek, G.; Portengen, L. t.; Brunekreef, B.; Vermeulen, R. C. Performance of prediction algorithms for modeling outdoor air pollution spatial surfaces. *Environ. Sci. Technol.* 2019, 53 (3), 1413–1421.
- Pak, U.; Ma, J.; Ryu, U.; Ryom, K.; Juhyok, U.; Pak, K.; Pak, C. Deep learning-based PM_{2.5} prediction considering the spatiotemporal correlations: A case study of Beijing, China. *Sci. Total Environ.* 2020, 699, 133561.
- Song, R.; Keller, A. A.; Suh, S. Rapid Life-Cycle Impact Screening Using Artificial Neural Networks. *Environ. Sci. Technol.* 2017, 51 (18), 10777–10785.
- Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer, 2013; Vol. 26. (141) Shorten, C.; Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *Journal of Big Data* 2019, 6 (1), 60.
- Bühlmann, P.; Van De Geer, S. *Statistics for High-Dimensional Data: Methods, Theory and Applications*; Springer Science & Business Media, 2011. (143) Rexstad, E.; Innis, G. S. Model simplification three applications. *Ecol. Modell.* 1985, 27 (1–2), 1–13.
- Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 2014, 15 (1), 1929–1958.
- Yao, Y.; Rosasco, L.; Caponnetto, A. On early stopping in gradient descent learning. *Constructive Approximation* 2007, 26 (2), 289–315.
- Sutton, R.S., and Barto, A.G. 1998. *Introduction to reinforcement learning*. Vol. 135. MIT Press, Cambridge, Mass., U.S.A.
- Piyush Jain Sean C.P. Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D. Flannigan (2020). A review of machine learning applications in wildfire science and management. <https://doi.org/10.1139/er-2020-0019>
- Xiao, Q.; Chang, H. H.; Geng, G.; Liu, Y. An ensemble machine-learning model to predict historical PM_{2.5} concentrations in China from satellite data. *Environ. Sci. Technol.* 2018, 52 (22), 13260–13269.