PREMIER JOURNAL OF SCIENCE REVIEW

Predicting Tomorrow: A Review of Machine Learning's Role in Shaping Environmental Forecasts

Susan Ferebee, PhD

University of Phoenix and Purdue University Global, Tuscon, Arizona, USA

Correspondence to: Susan Ferebee, ferebees@gmail.com

Additional material is published online only. To view please visit the journal online.

Cite this as: Ferebee S. Predicting Tomorrow: A Review of Machine Learning's Role in Shaping Environmental Forecasts. Premier Journal of Science 2024;1:100008

DOI: https://doi.org/10.70389/PJS.100008

Received: 23 July 2024

Accepted: 19 August 2024

Published: 2 September 2024

Ethical approval: N/a

Consent: N/a

Funding: No industry funding

Conflicts of interest: N/a

Author contribution: Susan Ferebee – Conceptualization, Writing – original draft, review and editing

Guarantor: Susan Ferebee

Provenance and peer-review: Commissioned and externally peer-reviewed

Data availability statement: N/a

ABSTRACT

Over the past forty years, advancements in artificial intelligence (AI) and machine learning (ML) have revolutionized Earth Sciences (ES). Driven by enhanced data from Earth observations, improved communications, and increased computing power, AI and ML are now critical in addressing real-world environmental challenges. The escalating severity of climate change impacts necessitates precise and timely environmental forecasts. This article critically examines the integration of ML techniques in environmental forecasting, highlighting their role in improving predictions of weather patterns, climate change, and ecological transformations. By automating the analysis of vast datasets, ML enhances environmental predictions' accuracy, timeliness, and applicability, supporting decision-making in agriculture, disaster preparedness, and environmental management. The review discusses the current state of ML applications, evaluates their effectiveness, and identifies future research directions. It also addresses the need for standardized data protocols, improved model interpretability, and ethical considerations in leveraging ML for climate research. The article concludes with a strong call for continued investment in research and cross-disciplinary collaboration, emphasizing the ongoing importance of these efforts to fully harness ML's potential in environmental forecasting.

Keywords: Machine learning in environmental forecasting, AI climate modeling, Deep learning in earth sciences, Ethical considerations of AI in environmental monitoring, Disaster prediction and management

Introduction

Over the past forty years, the transformative power of artificial intelligence (AI) in Earth Sciences (ES) has been propelled by Earth observations, enhanced communications bandwidth, increased computing power, and AI and machine learning (ML) advancements. The application of AI to real-world challenges has intensified due to the escalating severity of weather-related disasters influenced by climate change, population growth in high-risk areas, and ongoing non-sustainable practices.[1] Industries like social media, entertainment, and retail have witnessed companies such as Google, Amazon, and Facebook, built on groundbreaking AI, becoming some of the most valuable globally. These companies, alongside academic researchers, have honed ML methodologies and developed a mature array of software tools, workflows, and best practices. Given the overwhelming amount of environmental data, ES professionals have increasingly relied on AI and ML-enhanced automation to generate new insights and provide real-time, actionable predictions, thus extending the capabilities of researchers and forecasters to serve societal needs.[2]

The American Meteorological Society (AMS) AI Committee initially focused on developing knowledge bases and expert systems. It has now shifted its attention towards utilizing robust, swift, and interpretable ML techniques, including deep learning. This shift underscores the adaptability and progress of AI in Earth Sciences.[1]

Machine learning (ML) enhances environmental forecasts' accuracy, timeliness, and applicability. Integrating ML techniques into environmental science has revolutionized our ability to analyze complex datasets, leading to more precise predictions about weather, climate change, and ecological transformations. By automating the analysis of vast and diverse data sets, ML's unparalleled speed in identifying patterns and anomalies surpasses traditional methods, drastically reducing the time it takes to deliver critical information to stakeholders. This efficiency not only improves the reliability of forecasts but also underscores the urgency and importance of our work, extending the relevance of our predictions to a broader range of practical applications, from agriculture planning to disaster preparedness, ensuring informed decisions with the most current and comprehensive data available[3].

This article will critically examine and synthesize how ML technologies are currently applied to improve the accuracy and efficiency of environmental forecasting. This article explores various ML methodologies integrated into forecasting practices, assessing their impact on predicting climatic and ecological changes. It will review research findings that demonstrate the practical benefits and potential limitations of ML in this field. Furthermore, the article seeks to identify future directions for research and application, emphasizing how ML can further enhance the predictive capabilities of environmental science, thereby aiding policymakers, scientists, and global communities in making informed decisions based on reliable forecasts. The review will also consider the ethical implications and the need for robust data handling to ensure the accuracy and fairness of ML-driven environmental predictions.

Methodology

This review article synthesizes the current research and application of machine learning (ML) in environmental forecasting. The methodology was a literature review of academic publications, government reports, industry white papers, and case studies. Sources were selected based on relevance, credibility, and contribution to the field, covering a period from 2015 to 2024 to ensure the inclusion of the most recent advancements.

The literature search used academic databases such as Google Scholar, PubMed, and IEEE Xplore, AI-driven research tools like Semantic Scholar and Connected Papers, as well as specialized journals in environmental science, climate research, and artificial intelligence. Keywords such as "machine learning in environmental forecasting," "AI climate modeling," "deep learning in Earth sciences," and "ethical considerations of AI in environmental monitoring" were used to identify relevant sources.

To ensure a balanced perspective, this review incorporated studies from various geographical regions, representing both developed and developing countries and considered different ML methodologies, from deep learning techniques to supervised and unsupervised learning. The findings were then critically analyzed to identify trends, challenges, and future research directions, offering a holistic view of the current literature and potential advancements in using ML for environmental forecasting.

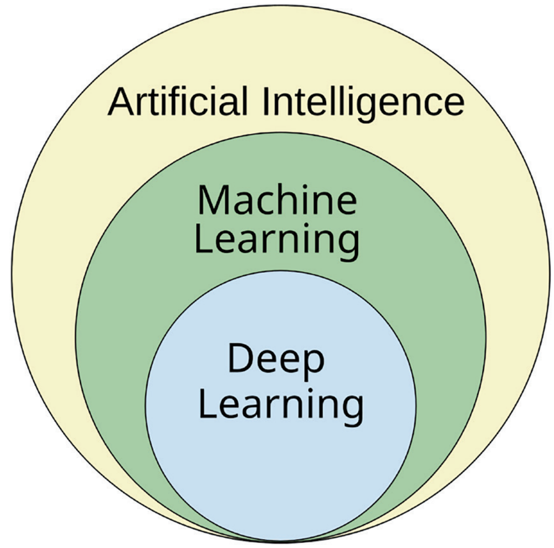

Fig 1 | AI hierarchy

Machine Learning Basics

Definition and Key Concepts

Machine learning revolves around two central inquiries: How can we develop computer systems that enhance their performance based on accumulated experience? And what are the core laws of statistics, computation, and information theory that apply to all learning entities, including computers, humans, and organizations?[4] Machine learning allows machines to learn without explicit programming and is a branch of AI.[5] Exploring machine learning is crucial not only to answer these critical scientific and engineering questions but also because of its practical implications, leading to the creation and implementation of highly effective software across numerous applications. Deep learning (DL) evolved as a type of ML and usually involves neural networks, which allows the extraction of deeper information levels.[1] Figure 1 illustrates the relationship between AI, ML, and DL.

Learning Methods

Supervised and Unsupervised ML Algorithms

Figure 2 demonstrates the differences in training data, resulting model, and application to new data during supervised and unsupervised learning in ML. A supervised learning algorithm categorizes data based on specific input features. To learn, a dataset that includes correct answers or labels guides the algorithm in recognizing patterns and relationships within the data. This training enables the algorithm to predict or classify new, unseen data. The core purpose of supervised learning is to make accurate predictions about the labels of new observations based on the knowledge it has acquired from its training dataset.[6]

In unsupervised learning, the AI learns from data that does not come pre-labeled, meaning the model must autonomously discover patterns within the raw, unlabeled input. Unsupervised learning relies heavily on trial and error to improve its understanding. Unlike supervised learning, where the model trains with known outcomes, unsupervised learning involves using various algorithms to help the model infer structures from the data it analyzes. This method significantly increases the likelihood of uncovering insights or patterns humans might need to recognize.[6]

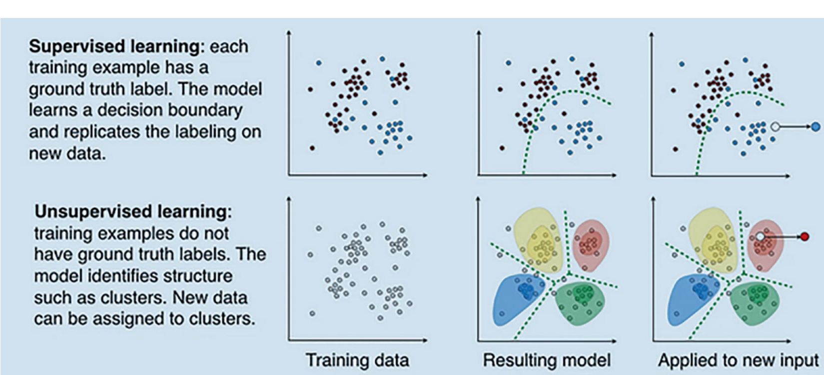

Fig 2 | Supervised and unsupervised learning

What is Environmental Data?

Environmental data encompasses various ecological parameters, such as land-use changes, deforestation, soil quality, pollution levels, and water quality. This data can be used in smart agriculture, disaster planning, cargo delivery, and transport. We will examine some of these in detail in a later section. Environmental data has traits of pressures, states, and impact.

- Pressures: Population expansion, land-use change, resource extraction, and pollution pressure the environment.
- States: Water quality, soil quality, habitat, vegetation, biodiversity, and air quality are active environmental situations.
- Impact: This refers to the impact of human actions within the environment that can lead to deforestation, diminished public health, biodiversity reduction, economic decline, and environmental damage.[9]

 DOI: https://doi.org/10.70389/PJS.100008 | Premier Journal of Science 2024;1:100008

There are several vital datasets to obtain environmental data: a) EARTHDATA: This is NASA's open dataset where data is obtained by satellites, aircraft, and field measurement; b) Climate Change Data: This is the World Bank's climate change data, which is gathered with satellites, demographic surveys, and aircraft; and c) Knowledge Network for Biodiversity (KNB): Data for this dataset is collected by field labs, research sites, and by independent researchers. It holds many types so of data related to environmental and ecological research.[10]

Data Quality and Quantity

The notion of feeding more data into machine learning models to achieve better results is prevalent in both academia and industry. However, while increasing data sometimes improves outcomes, in other instances, it does not, and not all datasets can be expanded.[7,8] A 2021 study addressed this issue by experimenting with student performance data from a Finnish vocational educational institution, comparing the performance of different machine learning algorithms in classifying students. The researchers used decision tree (DT) and random forest (RF) models, which are commonly used in explainable artificial intelligence (XAI) and other AI research fields. The evaluation metrics included accuracy, precision, recall, and overall performance. The results supported the idea that increasing the data quantity and input features improved classification results.[7]

Data features are numerical and categorical, selected from the model's learning ability. Choosing the features carefully to ensure quality data is critical when pre-processing datasets. The goal is to choose a group of features most relevant to the problem. Knowing all the data available, data engineers need to ask how much of that data they want to include. The key is to expose the model to a group of data points that provides a variation range across the data space. A small dataset might suffice if you train a model to perform image classification between horses and cows, but it will not provide the appropriate variation range. The dataset might only contain black and brown cows and may not classify an image of a white cow as a cow. This makes the argument for continually expanding the data quantity. However, small datasets can be used successfully in some situations.[8]

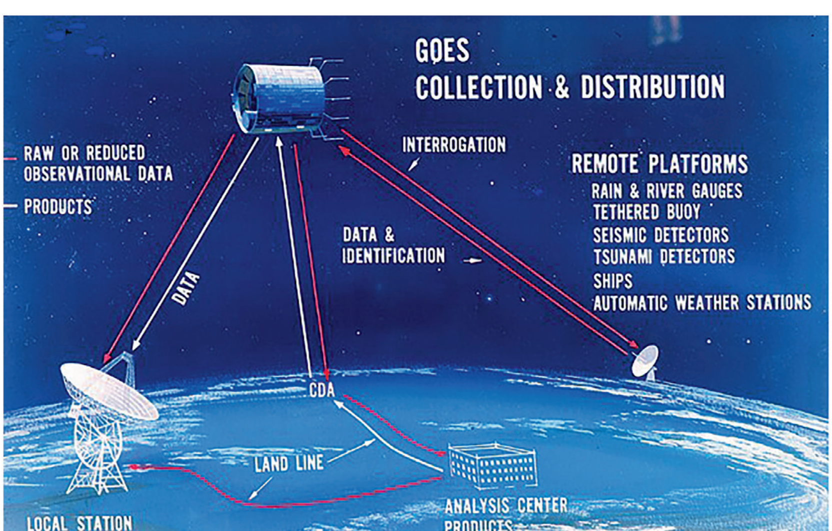

Fig 3 | GOES collection and distribution system (public domain)

Challenges of Data Acquisition in Environmental Sciences

Figure 3 diagrams the components, connections, and communication in the Geostationary Operational Environmental Satellite (GOES) Collection and Distribution System. Data sampling and analysis are the integrated components of data acquisition and are crucial to obtaining quality data. Environmental science samples include soil, water, chemicals, and other ecological and Earth elements. The sample has to be appropriately collected, stored and handled so as not to be contaminated. It is then sent to a lab for analysis, where the analyst must follow predefined protocols to preserve data quality. The analyst also has to define a reasonable level of error. Defective data can emerge from:

- Incorrect analysis protocols.
- Incorrect sampling protocols.
- Poor laboratory practices.
- Falsifying test results.[9]

Applications of Machine Learning in Environmental Forecasting

Natural Disaster Prediction and Management

Around the world, communities and governments are grappling with more frequent natural disasters and worsening extreme weather. Precision in disaster preparedness has never been more critical. Machine learning algorithms (MLAs) are transformative in boosting disaster readiness and response, and they have impressive capabilities to forecast diverse weather patterns and various natural disasters like tsunamis, heat waves, tornados, droughts, floods, and hurricanes. Natural disasters and extreme weather-related events have increased in number and severity, threatening governments, organizations, and people worldwide.[11]

DisasterNets is an example of machine learning that is applied to disaster mapping. It is a framework that allows rapid, accurate recognition of disasters. The framework has two phases: a) a space granulation stage and b) an attribute granulation stage. The framework can be applied to extensive flood mapping and earthquake-caused landslide mapping. DisasterNets relies on remote imaging during the disaster, captured by orbiting satellites like Sentinel, Landsat, and ASTER, to obtain near-real-time images from areas during the disaster. The photos are labeled with machine learning using supervised and semi/supervised deep learning models. Due to the sudden nature of disasters and the sheer volume of data, labeling becomes difficult. To address this, unsupervised pre-disaster images and domain adaptation techniques are used to segment post-disaster images without labels. These methods must overcome lighting, weather, and seasonal differences between image captures[12]. DisasterNets includes advanced networks like UCDFormer[13] and ADANet[14] to handle these complexities, using innovative approaches to align data distributions and improve disaster mapping accuracy.

Climate Change Modeling

Scientific evidence shows that human activities, especially the emission of greenhouse gases, are significant contributors to global warming.[15] Tackling the complexities of climate change requires advanced tools and methodologies, making predictive models and environmental impact assessments increasingly important. The recent rise of AI and ML has revolutionized climate change research, significantly improving our ability to predict future climate events and assess the environmental impacts of human activities.[16] Climate change involves long-term alterations in Earth's atmospheric conditions,[17] typically caused by humans burning fossil fuels and participating in industrial and deforestation processes.[18] Consequences of climate change are extreme weather, ecosystem disruption, rising sea levels, and reduced biodiversity. The potential influence on public health, food security, and safe water adds to the urgency.[19]

Because ML does not rely on predefined equations in prediction, they can evaluate new information against existing known information and adapt to changes, continuously improving their prediction ability. Evolved pattern recognition capability allows AI and ML models to detect subtle change indicators in climate patterns for greater prediction accuracy. Advanced simulation models replicate how changes in one environmental factor, such as sea ice levels, might affect other ecological systems and weather patterns. Machine learning can also differentiate between human and naturally induced changes, increasing the ability to mitigate climate change's impact.[20]

Some examples of real-world applications of AI and ML related to climate change modeling and predicting are:

- Coral Reefs: ML models and techniques are used to identify and assess the impacts of stress factors on coral reefs (pollution, overfishing, and rising water temperatures), which informs restoration and conservation decisions.[21]
- Forests: ML and AI provide the capability to understand how tree species, climate conditions, forest biodiversity, and soil conditions contribute to carbon sequestration. This knowledge informs global climate change initiatives and sustainable forestry management.[20]
- Sea Surface and Marine Heatwave Occurrence: The ability to predict sea surface temperatures is crucial as it affects marine heatwaves (extreme sea temperatures lasting longer than five days). A marine heatwave puts pressure on aquatic ecosystems, affecting marine biodiversity. Fishing and aquaculture domains are also affected. The ability to predict marine heatwaves supports adaptive management practices. ML models improved capturing non-linear connections.[22]

Challenges and Limitations

Three key challenges exist in machine learning's role in environmental forecasting. First is the need to standardize data formats. Environmental data sources abound from sensors like river gauges, ships, weather stations, seismic detectors, tethered ocean buoys, satellites, and weather stations. These sources use different scales, units, and formats, making integration difficult. Standardizing these formats is critical to ensure improved prediction.[20,23] Standardized collection, storing, and sharing practices are needed internationally to promote and simplify information exchange between governments, researchers, and institutions.[20]

Model interpretability is another challenge, as it is imperative to understand how a model reaches specific predictions and decisions. Being unable to explain this reduces the accountability of the model, particularly in life-critical decisions. Having the trust of the public, policymakers, and scientists is essential.[20] Interpretability requires observing the inner workings of the ML model and being able to interpret the model's parameters and weights that lead to the given output. The model can then be explained in human terms.[24]

The ethics of AI and ML are crucial when life-saving decisions rely on their forecasts and predictions. It is crucial to avoid unintended consequences. Misuse of ML and AI outputs, data privacy, and algorithmic bias are all areas of ethical challenge.[25] Improved model interpretability will improve transparency and foster ethical AI and ML. AI and ML policy and regulatory frameworks must include ethical guidelines, but policy creation must catch up to technology advancement.[26]

These challenges must be addressed so AI and ML can contribute meaningfully to climate change research. This will require collaboration from scientists, policymakers, researchers, citizens, and the government. Engaging the public will be vital in these efforts.[20]

Future Trends and Directions

Some significant advancements have occurred in how ML is used in environmental sciences and climate change research. Integrating AI and ML with emerging technologies like the Internet of Things (IoT), remote sensing, and high-performance computing promises more comprehensive and real-time environmental monitoring.[27] These innovations can enhance the accuracy of extreme weather predictions, provide precise environmental impact assessments, and offer deeper insights into complex ecological interactions. The future of AI and ML in climate change research depends on fostering interdisciplinary collaboration.[16]

Technological Advances

An example of a technological advancement that integrates AI, ML, remote sensing, and IoT is the recent creation of AI-driven sensors for monitoring hazardous substances in the environment, which adds the ability to process massive datasets in real-time and identify aberrations and patterns that indicate the presence of a hazardous substance. These sensors contribute

DOI: https://doi.org/10.70389/PJS.100008 | Premier Journal of Science 2024;1:100008

to an improved ability to protect public health. This is achieved by synthesizing machine learning algorithms, sensors, and data processing capabilities.[28] Industrial settings and other fields susceptible to chemical leaks benefit from this technology. In addition to detecting hazardous substances, ML improves the prediction of the likelihood and severity of hazardous substance incidences, fostering proactive mitigation planning; with AI and ML, a holistic view of hazardous substance existence, location, movement, and impact exists.[29]

Cutting-edge technologies enhance air quality modeling by integrating diverse data sources, including satellite imagery and meteorological data, for more accurate predictions. Combining AI and satellite technology enables comprehensive environmental monitoring and helps identify pollution sources. Edge computing facilitates real-time environmental data analysis at its source, allowing for rapid responses to pollution incidents.[28] Additionally, hybrid models that merge physics-based simulations with machine learning improve the accuracy of pollution source identification.[30]

Over the past decade, various studies have explored machine learning to predict the effectiveness of heavy metal removal from soil. These studies have focused on using machine learning to forecast the success of various soil remediation techniques.[31] Additionally, a convolutional neural network (CNN) architecture was developed using near-infrared (NIR) spectroscopy data for deep calibration. This research aimed to assess water pollution levels from domestic and industrial sources to support appropriate agricultural irrigation practices. The study successfully established intelligent spectroscopic models with the CNN architecture, which could significantly address water recycling and agricultural conservation issues.[32]

Policy and Ethical Considerations

Despite significant advancements, several challenges remain in fully leveraging AI and ML for climate change research. Key issues include standardized data formats, improved model interpretability, ethical considerations, and integration into policy framework.[32] Researchers must develop standardized protocols for data collection and sharing to enhance the interoperability of diverse datasets. Enhancing model interpretability through innovative visualization techniques and transparent documentation will boost the reliability and trustworthiness of AI and ML models. Ethical deliberation must guide the trustworthy use of technology, ensuring positive contributions to societal and environmental well-being.

Furthermore, integrating AI and ML into policy frameworks necessitates continuous collaboration among researchers, policymakers, and legal experts across many countries. Regulatory frameworks must be flexible and adaptive to keep up with the evolving technology environment and its applications in climate change research.[33] An interdisciplinary and cross-country approach is essential. A country's level of acceptance of AI will impact collaborative efforts.[34] A recent study conducted across 28 countries provides insight into the general acceptance of AI technologies, highlighting varying readiness levels to adopt AI for diverse applications, including environmental forecasting. The study revealed that countries with high levels of AI acceptance are more likely to integrate AI and machine learning into critical areas such as disaster management and climate change mitigation.[34]

Nations with a strong digital infrastructure and AI-friendly policies, such as the United States, Japan, and several European countries, are more willing to deploy these technologies in environmental sciences. However, the purpose for which AI is used also influences acceptance. Sixty percent of Europeans, for example, do not accept the use of robots to care for children, the elderly, or the disabled. Only twenty percent of people in Ireland accept autonomous vehicles. A large percentage of U.S. citizens accept drone delivery of parcels.[34]

The approach to AI regulation and policymaking differs across countries. China's AI regulation is the Interim Measures for the Management of Generative Artificial Intelligence Services. It is an ethical model and focuses on actual liability and is limited to only generative AI. The EU's Provisional EU AI Act regulation is based on forecasted risk assessment. The Unite State's decentralized approach is based on priorities, principles, and guidelines, resulting in the Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. These policies were implemented in 2023 and 2024.[35]

Case Study: Predictive Analytics for Flood Management in Bangladesh

Background

Due to its unique geography, Bangladesh is highly susceptible to seasonal flooding. These floods affect millions and damage homes, crops, and infrastructure. Traditional methods for predicting floods often fail to provide timely and accurate forecasts, complicating preparedness and response efforts.[36]

Implementation of Machine Learning

A project was initiated to leverage machine learning (ML) for improved flood prediction. This project involved meticulously examining flood patterns and their relationship to meteorological factors. The ML models were trained using satellite imagery, historical flood data, and weather information, particularly rainfall and minimum/maximum temperatures.[36]

Methodology

Data: A dataset from Kaggle and the Bangladesh Meteorological Department included 20,544 incidents and an in-depth analysis of weather incidences in thirty-two Bangladesh districts. The researchers combined selected features from the Random Forest and Chi-square feature selection processes to form an ideal feature set for flood prediction.[36]

Model Training: Supervised ML models, specifically Random Forest and Gradient Boosting algorithms, were trained on this data to identify patterns and predict floods. DL neural networks were also used to identify concealed patterns.[36]

Results

Integrating DL models TabPFN and TabNet with traditional classifiers like RandomForest and LightGBM elevated predictive accuracy to 97.7%. The study results informed actionable insights for preparedness and mitigation approaches and confirmed the benefits of ML in disaster management.[36]

This project's success demonstrated the transformative potential of ML and DL in environmental forecasting, offering a scalable solution for regions facing similar issues. It emphasizes the importance of continued investment in ML research and cross-disciplinary collaboration to enhance predictive capabilities and improve disaster management.

Future Implications and Future Research

Integrating machine learning (ML) into environmental forecasting has already shown significant promise, yet its full potential remains untapped. As AI and ML technology continue to evolve, several future implications and research directions will shape the trajectory of ML in environmental sciences.

Future Implications

As machine learning (ML) algorithms continue to evolve, they are poised to significantly enhance the accuracy and precision of environmental forecasts.[36] This progression will be crucial in managing the growing risks of climate change, including rising sea levels, increasingly severe, frequent extreme weather events, rising sea levels, and the ongoing loss of biodiversity. Improving accurate predictions will save lives and reduce economic losses by enabling more effective disaster preparedness and resource management.

The accessibility and standardization of AI and ML technologies are also expected to drive their widespread adoption in environmental monitoring on a global scale.[20] As these tools become more integrated into environmental science, we can anticipate a shift towards more comprehensive and real-time data collection. This advancement will improve our ability to monitor environmental changes and respond promptly to emerging issues. AI-driven sensors, coupled with the Internet of Things (IoT), will play a crucial role in tracking critical ecological parameters such as pollution levels, deforestation rates, and other indicators of environmental health.

The increasing reliance on ML for environmental forecasting will inevitably impact policy and decision-making processes. Governments and international organizations will increasingly turn to ML-generated insights to craft policies to mitigate climate change effects and more effectively manage natural resources. This shift underscores the need for robust regulatory frameworks that ensure AI's ethical and responsible use in environmental contexts.[32]

However, ethical and social considerations will become important as ML becomes more deeply embedded in environmental forecasting. Concerns about data privacy, algorithmic bias, and the transparency of ML models must be addressed to maintain public trust and ensure that these technologies are used equitably. The ethical deployment of AI in environmental management will be critical to maximizing its benefits while minimizing potential harm.[32]

Finally, the future of ML in environmental forecasting will depend heavily on global collaboration and the standardization of data and methodologies. Cross-border cooperation will be essential in effectively tackling global environmental challenges. Establishing standardized data formats and sharing protocols will facilitate the integration of diverse datasets, leading to more accurate and comprehensive ML models. This collaborative approach will unlock ML's full potential in addressing the complex and interconnected issues of climate change and environmental degradation.

Future Research Directions

- Interdisciplinary Research: Future research should focus on interdisciplinary collaboration between computer scientists, environmental scientists, policymakers, and ethicists. This collaboration will be crucial for developing ML models that are technically robust, socially responsible, and aligned with global environmental goals.
- Advancements in Model Interpretability: Research should prioritize enhancing ML models' interpretability in environmental forecasting. Developing techniques that allow stakeholders to understand and trust the outputs of these models will be essential for their widespread adoption in critical decision-making processes.
- Exploration of Novel ML Algorithms: Continuous exploration of new and innovative ML algorithms will be vital for tackling complex environmental challenges. Research into unsupervised learning, deep reinforcement learning, and hybrid models that combine ML with traditional simulation techniques could lead to predictive accuracy and efficiency breakthroughs.
- Integration with Emerging Technologies: Future research should explore integrating ML with emerging technologies, such as quantum computing, blockchain, and edge computing. These technologies can significantly enhance the processing power, security, and real-time capabilities of ML models, opening new frontiers in environmental forecasting.
- Addressing Ethical and Societal Impacts: Research must examine ethical and societal implications of using ML in environmental sciences. Developing frameworks that ensure fairness, transparency, and accountability in AI-driven decision-making processes will be critical to address concerns of algorithmic bias and the equitable distribution of benefits.
- Focus on Under-Researched Regions: There is a need for research that targets regions currently underrepresented in environmental ML studies, particularly in the Global South. Expanding the geographical focus of ML research will help ensure that the benefits of these technologies are

distributed globally and that solutions are relevant to the particular needs and challenges of different regions.

The future of ML in environmental forecasting is full of potential, with significant implications for science, policy, and society. By pursuing these research directions and addressing the associated challenges, we can unlock the full power of ML to address some of the most pressing environmental issues of our time.

Conclusion

Machine learning (ML) has significantly transformed environmental forecasting by enhancing predictions' accuracy, timeliness, and applicability. This technology enables the efficient analysis of vast and complex datasets, improving our ability to forecast weather events, assess environmental impacts, and understand ecological changes. Integrating ML with other advanced technologies, such as remote sensing, IoT, and high-performance computing, has expanded its capabilities, providing real-time, comprehensive environmental monitoring and more precise identification of pollution sources.

The future of ML in environmental sciences looks promising, with ongoing research focusing on refining predictive models and integrating them into policy frameworks. As ML methodologies continue to evolve, they will become increasingly vital in addressing the complexities of climate change. Technological advancements will further enhance the precision and reliability of environmental forecasts, aiding in disaster preparedness, sustainable resource management, and ecological conservation. Collaborative efforts between researchers, policymakers, and industry experts will be crucial to advancing ML applications in environmental science and effectively leveraging these technologies.

To harness the full potential of ML in environmental forecasting, continued investment in research and development is essential. Cross-disciplinary collaboration is vital for creating standardized protocols, improving model interpretability, and addressing ethical considerations. By fostering a collaborative environment, stakeholders can ensure that ML technologies contribute positively to societal and environmental well-being. It is imperative to support initiatives that promote the integration of ML into policy frameworks, facilitating the development of agile and adaptable regulatory structures that can keep pace with technological advancements.

In conclusion, machine learning offers unprecedented opportunities for advancing environmental science and climate change research. By investing in research, fostering collaboration, and integrating ML into policy frameworks, we can unlock its full potential and make informed decisions that benefit our planet and future generations.

References

1 Haupt SE, Gagne DJ, Hsieh W, et al. The history and practice of AI in the environmental sciences. Bulletin of the American Meteorological Society. 2022; 103(5): E1351–E1370. https://doi.org/10.1175/BAMS-D-20-0234.1
2 MacLeod N. Artificial intelligence & machine learning in the earth sciences. Acta Geologica Sinica. 2019; 93(supp.1): 48–51. https://doi.org/10.1111/1755-6724.14241
3 Zhong S, Zhang K, Bagheri M, et al. Machine learning: New ideas and tools in environmental science and engineering. Environ Sci Technol. 2021;55(19): 12741–12754. https://doi.org/10.1021/acs.est.1c01339.
4 Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science. 2015; 349(6245): 255–260. https://doi.org/10.1126/science.aaa8415
5 Wang J. Encyclopedia of data science and machine learning. IGI Global; 2023. Understanding machine learning concepts; p. 1007–1022. https://doi.org/10.4018/978-1-7998-9220-5
6 Abubakar A, Almeida CF, Gemignani M. Review of artificial intelligence-based failure detection and diagnosis methods for solar photovoltaic systems. Machines. 2021; 9(12): 328. https://doi.org/10.3390/machines9120328
7 Kariluoto A, Parnanen A, Kultanen J, Soininen J, Abrahamson P. Quality of data in machine learning. Paper presented at: The 21st IEEE International Conference on Software Quality, Reliability, and Security; 2021;6–8; Hainan Island, China. https://doi.org/10.1109/QRS-C55045.2021.00040
8 Monolith. Data quality and quantity for machine learning [Internet]. Monolith; 2024. Available from https://www.monolithai.com/blog/data-quality-and-quantity-for-machine-learning
9 Zhang C. Fundamentals of environmental sampling and analysis. Wiley; 2007. Introduction to environmental data acquisition; p. 1–10. https://doi.org/10.1002/0470120681
10 Singh A. A short introduction to environmental data. Medium; 2020. Available from https://medium.com/swlh/a-short-introduction-to-environmental-data-e243da4af019
11 Jain H, Dhupper R, Shrivastava A, et al. We leverage machine learning algorithms for improved disaster preparedness and response through accurate weather patterns and natural disaster prediction. Front Environ Sci, 2023; 11. https://doi.org/10.3389/fenvs.2023.1194918
12 Xu Q, Shi Y, Zhu, X. DISASTERNETS: Embedding machine learning in disaster mapping. Paper presented at: IEEE International Geoscience and Remote Sensing Symposium; 2023;16–21; Pasadena, CA. https://doi.org/10.1109/IGARSS52108.2023.10282003
13 Xu Q, Shi Y, Guo J, Ouyang C. UCDFormer: Unsupervised change detection using a transformer-driven image translation. Paper presented at: IEEE Transactions on Geoscience and Remote Sensing PP (99):1–1. https://doi.org/10.1109/TGRS.2023.3305334
14 Xu Q, Ouyang C, Jiang T, et al. Mffenet and adanet: A robust deep transfer learning method and its application in high precision and fast cross-scene recognition of earthquake-induced landslides. Landslides, 2022; 19(3): 1–31. https://doi.org/10.1007/s10346-022-01847-1
15 Mikhaylov A, Moiseev N, Aleshin K, et al. Global climate change and greenhouse effect. Entrepreneurship and Sustainability Issues, 2020; 7(4): 2897.
16 Leal FW, Wall T, Mucova SA, et al. Deploying artificial intelligence for climate change adaptation. Technological Forecasting and Social Change, 2022; 180: 121662. https://doi.org/10.1016/j.techfore.2022.121662
17 Mashwani ZU. Environment, climate change and biodiversity. Environment, climate, plant and vegetation growth, 2020; 473–501. https://doi.org/10.1007/978-3-030-49732-3_19
18 Kumari N, Solanki H, Kumar A. Climate change: A burning issue for the world. Medicine, 2020; 35: 501–507.
19 Gomez-Zavaglia A, Mejuto JC, Simal-Gandara J. Mitigation of emerging implications of climate change on food production systems. Food Research International, 2020;134:109256. https://doi.org/10.1016/j.foodres.2020.109256
20 Hamdan A, Ibekwe KI, Etukudoh EA, et al. AI and machine learning in climate change research: A review of predictive models and environmental impact. World Journal of Advanced Research and Reviews, 2024; 12(1): 1999–2008. https://doi.org/10.30574/wjarr.2024.21.1.0257
21 Mandal A, Ghosh AR. AI-driven surveillance of the health and disease status of ocean organisms: A review. Aquaculture International, 2023; 32(1): 1–12. https://doi.org/10.1007/s10499-023-01192-7
22 Bonino G, Galimberti G, Masina S, et al. Machine learning methods to predict sea surface temperature and marine heatwave

occurrence: A case study of the Mediterranean Sea. European Geosciences Union, 2024;20(2): 417–432. https://doi.org/10.5194/os-20-417-2024
23 Cheval S, Micu D, Dumitrescu A, et al. Meteorological and ancillary data resources for climate research in urban areas. Climate, 2020; 8(3): 37.
24 AWS. Interpretability versus explainability [Internet]. AWS; 2024. Available from https://docs.aws.amazon.com/whitepapers/latest/model-explainability-aws-ai-ml/interpretability-versus-explainability.html
25 Stahl BC. Artificial intelligence for a better future: An ecosystem perspective on the ethics of AI and emerging digital technologies. Springer Nature; 2021. Ethical issues of AI; pp. 35–53.
26 Cowls J, Tsamados A, Taddeo M, et al. The AI gambit: Leveraging artificial intelligence to combat climate change – opportunities, challenges, and recommendations. AI and Society, 2021; 38: 283–307.
27 Kachouel A, Kaushik M, Ali MA. Internet of Things-enabled food and plant sensors to empower sustainability. Advanced Intelligent Systems, 2023; 5(12): 2300321. https://doi.org/10.1002/aisy.202300321
28 Popescu SM, Mansoor S, Wani, OA, et al. Artificial intelligence and IoT-driven technologies for environmental pollution monitoring and management. Front Environ Sci, 2024; 12: 1336088. https://doi.org/10.3389/fenvs.2024.1336088
29 Shi S, Hou M, Gu Z, et al. Estimation of heavy metal content in soil based on machine learning models. Land. 2022; 11(7): 1037. https://doi.org/10.3390/land11071037
30 Anthony D, Siriwardana H, Ashvini S, et al. Trends in marine pollution mitigation technologies: Scientometric analysis of published literature (1990–2022). Regional Stud Mar Sci, 2023; 66: 103156. https://doi.org/10.1016/j.rsma.2023.103156
31 Baghat SK, Tung T, Yaseen ZM. Development of artificial intelligence for modeling wastewater heavy metal removal: State of the art, application assessment and possible future research. J Clean Prod, 2020; 250:119473. https://doi.org/10.1016/j.jclepro.2019.119473
32 Cave A, Brun NC, Sweeney F, et al. Big data: How to realize the promise. Clin Pharma & Ther, 2020; 107(4): 753–761. https://doi.org/ 10.1002/cpt.1736
33 ElZarrad MK, Lee AY, Purcell R, et al. Advancing an agile regulatory ecosystem to respond to the rapid development of innovative technologies. Clin Transl Sci, 2022. https://doi.org/ 10.1111/cts.13267
34 Vu HT, Lim J. Effects of country and individual factors on public acceptance of artificial intelligence and robotics technologies: A multilevel SEM analysis of 28-country survey data. Behaviour & Information Technology, 2022; 41(7): 1515–1528. https://doi.org/10.1080/0144929X.2021.1884288
35 Comunale M, Manera A. The economic impacts and the regulation of AI: A review of the academic literature and policy actions. IMF WP/24/65, 2024: 1–66.
36 Rakin FI, Ahmmed T, Kabir R, et al. Predictive analytics for floods in Bangladesh: A comparative exploration of machine learning and deep learning classifiers. Paper presented at 26th International Conference on Computer and Information Technology (ICCIT); 2023; 13–15. https://doi,org/10.1109/ICCIT60459.2023.10441401

List of Image Sources

1. Figure 1: https://upload.wikimedia.org/wikipedia/commons/thumb/1/1b/AI_hierarchy.svg/640px-AI_hierarchy.svg.png
2. Figure 2: https://commons.wikimedia.org/w/index.php?search=unsupervised+machine+learning&title= Special:MediaSearch&go=Go&type=image
3. Figure 3: https://upload.wikimedia.org/wikipedia/commons/thumb/0/03/Graphic_of_GOES_satellite_data_collection_and_distribution_Satellite_shown_isan_Applications_Technology_Satellite_%282297-1%29.jpg/640px-Graphic_of_GOES_satellite_data_collection_and_distribution_Satellite_shown_isan_Applications_Technology_Satellite_%282297-1%29.jpg