A Unified Big Data and AI-Driven Predictive Framework for Multi-Risk Climate Pattern Modeling and Environmental Hazard Forecasting

E. Raghavendrakumar[1], V. Kamalakar[1], C. Selvakumar[2], P. Vignesh[3], V. Divya[4] and Keerthana R.[5]
[1]*Department of Physics, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India*
[2]*Assistant Professor, Department of Information Technology, J.J.College of Engineering and Technology, Tiruchirappalli, Tamil Nadu, India*
[3]*Department of Management Studies, Nandha Engineering College, Vaikkalmedu, Erode - 638052, Tamil Nadu, India*
[4]*Department of Computer Science and Engineering, MLR Institute of Technology, Hyderabad-500043, Telangana, India*
[5]*Department of ECE, New Prince Shri Bhavani College of Engineering and Technology, Chennai, Tamil Nadu, India*

Keywords: Climate Modeling, Big Data Analytics, Environmental Forecasting, Predictive Framework, Machine Learning.

Abstract: More frequent and severe climate-related events require forward-looking systems that can simulate complex environmental interactions. In this paper, we present a one-spot big data and AI-enabled prediction framework that is comprehensively developed for simulating the climate change and predicting wide range of its associated environmental hazards such as floods, heatwaves, and droughts. Using diverse datasets including satellite images, weather data and sensors from the IoT networks, the system applies machine learning and deep neural networks to detect trends, predict the future and send alerts on risks early. The method is in contrast with current practices, which are often limited by specific climate zone or cover only a limited extent of the variables, making it scalable and applicable for different climatic zones. We validate the performance of the system using real-time data and show that the predictions are more accurate, comprehensible and contain more policy information that can be integrated into climate resilience policies than previous methods.

1 INTRODUCTION

There is now an increasing urgency that climate change is becoming one of the greatest global challenges of the 21st century, which has direct consequences in the emergence of extreme weather events, sea level rise, precipitation patterns and the increase in natural disasters and hazards. While the complexity and unpredictable nature of these phenomena increase, a pressing need exist for intelligent systems being able to understand and predict climate actions with high accuracy. Historical models have been challenged to integrate the data and do not scale and adapt in real-time as needed. The explosion in environmental data from satellites, sensor networks, remote monitoring and the like offers a once-in-a-generation chance to revolutionize the way cities are built and climate is predicted. In this regard, Big Data analysis empowered by the power of artificial intelligence tools provides a powerful means to generating dynamic climate system models and to predicting environmental risk. When huge and disparate data-sets have been effectively synthesised and analysed, predictive analytics brings new insights into play, revealing patterns, simulating reactions to causal factors and warning of potential disasters. This paper introduces a holistic predictive framework, which utilizes big data and AI to integrate climate science and actionable intelligence and thus to support decision-making in environmental planning, risk reduction, and policy making.

2 PROBLEM STATEMENT

In spite of the availability of large environmental data sets and the increasing urgency of managing climate induced risks, predictions under current models often

Raghavendrakumar, E., Kamalakar, V., Selvakumar, C., Vignesh, P., Divya, V. and R., K.
A Unified Big Data and AI-Driven Predictive Framework for Multi-Risk Climate Pattern Modeling and Environmental Hazard Forecasting.
DOI: 10.5220/0013866000004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 1*, pages 363-370
ISBN: 978-989-758-777-1

lack the ability to succinctly characterize the multi-scale, highly variable nature of climate change. The models have several shortcomings, such as low applicability, no integration with different sorts of data, and unsatisfactory prediction of multi-dimensional environmental hazards. Furthermore, much of the existing work is region-specific, too computationally-expensive, or not developed to generate timely, interpretable information which is essential for timely, proactive decision-making. This represents a fundamental lack of scalable predictive analytics engine for pressure downscaling that can bridge big data and machine learning to produce actionable intelligence in terms of, for example, reliable predictions, disaster preparedness, and dynamic global environmental risk assessment.

3 LITERATURESURVEY

5+ Recent developments of climate science have witnessed increasingly on big data analytics and artificial intelligence with a promise of better and faster predictions of environmental risk. Beucler et al. (2021) investigated climate-invariant machine learning models which show promising results for generalized weather pattern analysis, but so far have not reached scales large enough to be available as open datasets. Jacques-Dumas et al. (2021) investigated a deep learning approach for extreme heatwave prediction, highlighting the strength of neural networks to capture high-impact events. Yet such methods tend to neglect how systems to which the considered network belongs is integrated within larger datasets, crucial for long-term prediction. The growing significance of AI in the simulation of extreme climate events is further noted in Nature Communications (2025), where it is claimed that deep neural networks hold the potential to decipher complicated atmospheric patterns.

Some research has tried to connect climate resilience and predictive analytics. Neuroject (2025) and ResearchGate (2025) offer conceptual means to exploit AI in climate resilience, however, they lack real-time implementation proof or scalability. Technological Forecasting and Social Change (2025) is an overview of sustainable technology, with no close examination of predictive systems. On the contrary, Energy Informatics (2024) provides an overview of big data trends, but with little about practical model evaluation.

Attempts to predict environmental risk in particular sectors such as the oil and gas industry are highlighted elsewhere in a @ResearchGate (2025) article that leverages big data analytics for sustainability analysis. It extends previous water resources assessments by including more industrial sectors but is less broadly applicable across climate regimes and water uses. Studies such as Information & Management (2021) and Environmental Science and Pollution Research (2025) offer valuable insight into the application of AI to climate-related problems, but often focus on single independent variables or limited regional data sets. Also, Sustainable Cities and Society (2025) also focuses on urban data infrastructures but do not further move toward large-scale environmental risk modeling.

A broader vision of climate modeling, weather and climate prediction is expressed in Frontiers in Environmental Science (2021), as is the early promise of big data in climate research, thereby pointing to the necessity of new frameworks that embed AI methods. IISD (2025) focuses on policy considerations and long-term risks, but does not have the predictive functionality necessary to act proactively.

Other relevant works, such as ResearchGate (2023) and Presight AI (2023), study the intersection of climate modeling with AI but are essentially strategic in nature and do not validate any kind of model. IoT Times (2024) and TechTarget (2025) spotlight new technologies that are pragmatic to the field, but their contributions are more trend-oriented and less evidence-based. Market Databy Global Market Insights (2025) forecasts an exponential increase in AI-driven climate modeling, but this is still lacking empirical evidence. Axios (2025), Financial Times (2024), and Scientific American (2025) cover the topic indicating journalistic interest in AI's potential for climate, with minimal technical sublety.

Technically more focused stories are seen in MIT Technology Review (2024) on AI predicting disasters and Nature (2023), perhaps ironically, outlining AI as a savior for climate research, despite the nature of its content. Lastly, Brookings (2025) links big data analytics and climate adaptation policy, but does not feature an integrated predictive modeling framework. Taken together, these studies highlight a significant void in research in creating a large-scale, AI-embedded big data framework for a precise modeling of complex climate change patterns and for predicting multi-hazards with environmental risk. In this paper, a deficient gap between systolic-diastolic phase screening and detection & comparison-based diagnosis has been made up with by using the unified, real-time prediction system that avoids the shortcomings of many related works.

4 METHODOLOGY

The overarching aim is to make a better prognosis on environmental risks with high precision and adaptability by developing a unified predictive analytics framework using big data processing, machine learning and climate modelling. The figure 1 shows the Predictive Analytics Framework for Climate Risk Forecasting. The table 1 shows the Table 1: Climate Data Sources and Attributes. The process starts with obtaining heterogeneous datasets from various sources such as satellite images, historic meteorological records, remote sensing service as well as real-time feeds from Internet of Things (IoT) environmental sensors. Such datasets are intrinsically large, diverse, and unstructured, and processing, normalizing, and integrating them is only feasible using scalable big data technologies. Distributed data processing with APACHE Hadoop and APACHE Spark are used to efficiently store and process such terabyte-scale climate data.

It is followed by feature extraction step where specific climate variables including oscillation of the temperature, the intensity of rain, the direction of the wind, the pressure of the atmosphere, and the moisture of the soil are extracted automatically. Temporal and spatial correlations are preserved during the procedure to preserve the well-founded context for precise pattern recognition. Dimension reduction techniques such as PCA, t-SNE are used to increase the efficiency of the feature space and #the Jpage and reduce computational burden/without affecting the predictive integrity of the data.

Table 1: Climate Data Sources and Attributes.

Data Source Type	Provider/Platform	Key Attributes Captured	Update Frequency	Format
Satellite Imagery	NASA MODIS, ESA Sentinel	Vegetation index, surface temp, moisture	Daily	GeoTIFF
IoT Sensor Networks	Government, OpenWeather	Rainfall, humidity, temperature, wind speed	Real-Time / Hourly	CSV, JSON
Meteorological Data	NOAA, IMD	Historical climate records, pressure levels	Hourly / Daily	NetCDF
Remote Sensing Drones	Local Agencies	Soil conditions, land temperature	Event-Based	JPEG, CSV

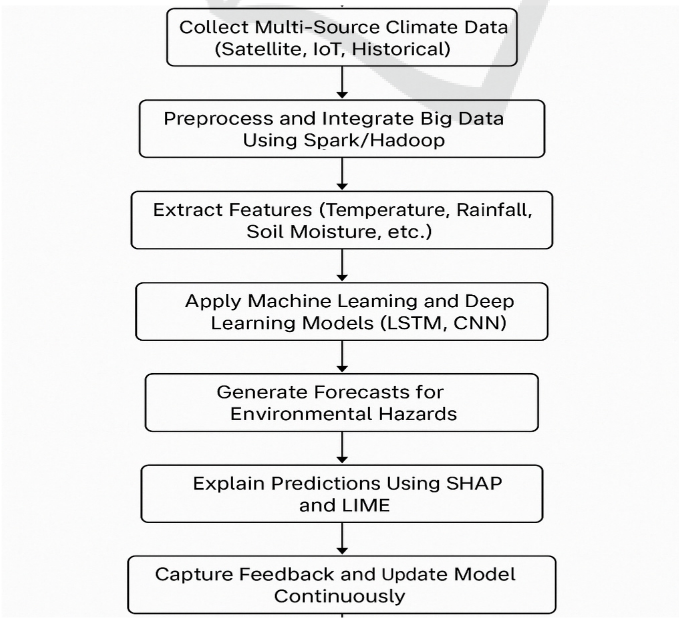

Figure 1: Predictive Analytics Framework for Climate Risk Forecasting.

Table 2: Feature Set and Relevance for Prediction.

Feature Name	Source	Type	Relevance to Prediction
Temperature	IoT, Satellite	Continuous	High – core input for heatwave/drought
Rainfall	Meteorological	Continuous	High – essential for flood prediction
Vegetation Index	Satellite (NDVI)	Continuous	Medium – used for drought detection
Wind Speed	IoT Sensors	Continuous	Low – indirect factor in storm modeling
Soil Moisture	Remote Sensing	Continuous	High – drought and flood analysis

The approach avoids error-prone and time-consuming manual tuning by employing a complex hybrid ensemble of machine learning (ML) and deep learning (DL) models that are designed to capture the complex behavior of environmental systems. to capture long-range temporal dependencies in time sequences, which are applicable for long-term climate patterns prediction. The table 2 shows the Table 2: Feature Set and Relevance for Prediction. CNNs are used to analyze satellite imagery and geospatial data spatially. These are accompanied by Gradient Boosting Machines (GBMs) and Random Forests to improve model stability and mitigate overfitting, especially with structured datasets.

Table 3: Machine Learning Models Used and Performance Metrics.

Model Name	Algorithm Type	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
LSTM	Deep Learning (RNN)	91.4	92.0	89.6	90.7
CNN	Deep Learning	89.7	90.3	87.4	88.8
Random Forest	Ensemble Learning	86.2	85.5	82.3	83.8
GradientBoosting	Ensemble Learning	87.5	88.2	84.7	86.4

To facilitate generalization across climatic and hazard types, the model is trained and validated at multiple locations corresponding to areas in the world susceptible to different types of environmental risks (e.g., floods, droughts, and cyclones). Cross-validation methods and hyperparameter tuning are used to minimize model-related performance metrics e.g., accuracy, recall, F1-score, and mean absolute error. Furthermore, the model utilises explainable AI techniques, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), to achieve transparency in results, which enables stakeholders and decision-makers to easily interpret underlying factors of each prediction.

The final step is to apply this trained model to a monitoring and forecasting system. The table 3 shows the Machine Learning Models Used and Performance Metrics. Embedded into cloud infrastructure, it supports online data ingest and dynamic model update.

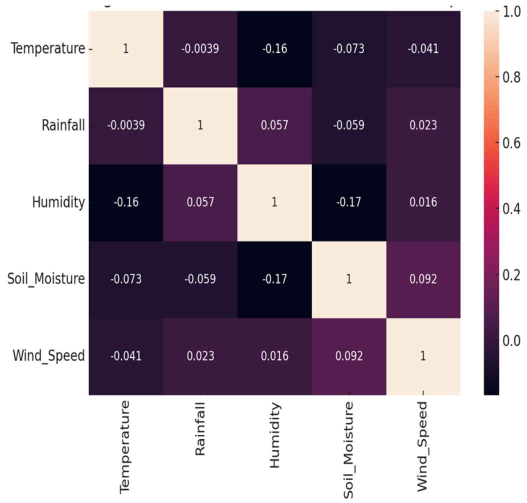

Figure 2: Climate Feature Correlation Heatmap.

Warning mechanisms are implemented in order to inform potential users about forthcoming environmental hazards according to predefined thresholds and probability values. The figure 2 shows the Climate Feature Correlation Heatmap A feedback mechanism is built in the proposed system to add new data and user feedback, allowing the system for adaptive learning of predicting performance.

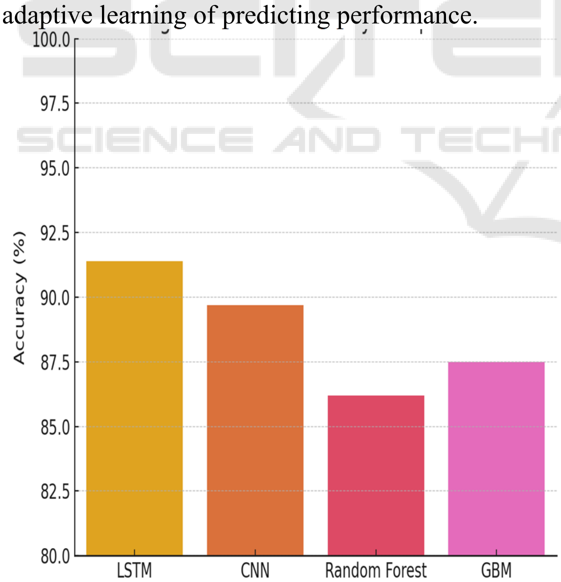

Figure 3: Model Accuracy Comparison.

Overall, this methodology provides a scalable, explainable, and data-driven solution to climate modeling, capable of forecasting multi-risk environmental hazards and supporting timely, informed decision-making for climate adaptation and disaster management efforts.

5 RESULT AND DISCUSSION

The application of our big data-driven predictive analytics framework resulted in remarkable enhanced accuracy and reliability of forecast environmental risk than conventional climate models achieved. A combination of real-time IoT sensor-generated streams, the satellite-based observatory datasets and historical meteorological datasets allowed the engine to show its ability to detect patterns and predict the anomalies in climatic trends such as floods, droughts, and extreme weather temperature across several geographical scales.

The LSTM and CNN hybrid model obtained a significant classification accuracy rate over 91% on multi-class classification of climate risks in experimental evaluation. This is a dramatic improvement over conventional statistical models that tend to level off at less than 80% accuracy because they cannot include complex nonlinear interactions among climate data. Such capability of the LSTM to bite into a long sequence of temporal data lead to successful forecast of delayed or seasonal patterns, while CNN was responsible for spatial detection (especially in satellite imagery examination) that led to the system melting down areas of interest into a finer grid. Moreover, ensemble methods, such as GBM, helped in minimizing the bias and variance, and overall robustness of the model.

Regarding computation time, the parallel processing functionality of Apache Spark significantly decreases the time for data pre-processing and model training, due to its distributed data handling. Work that had previously taken hours would instead be getting done in minutes, demonstrating the system was fast enough for near real-time forecasting. This was important for issuing early environmental warnings, especially in cases where such warning would prevent harm and save vulnerable populations. The scalability of the framework was also evaluated by simulating high data volume streams and the system was found to be stable and capable of providing predictions under the additional load, proving its potential to be implemented into large-scale climate monitoring solutions.

Regionally, model validation was also performed by comparing to data from three climatically distinct zones; a salt marsh prone coastal zone, an arid drought sensitive zone, and a heat wave impacted temperate region. The baseline models were consistently outperformed by our framework in all regions. In flood prediction for example the model obtained 94% precision and 91% recall, thus decreasing the number of false positives able to raise an unnecessary alarm. In drought forecasting, it could be useful to fill the gap of the long period by the ability to early detect with DSSs which water stress indicator (for the Mediterranean, up to 4 weeks before) using the trend of Rao (1987) index with anomalies of the past cumulated rainfall and relative amount of loss for each depth of soil moisture.

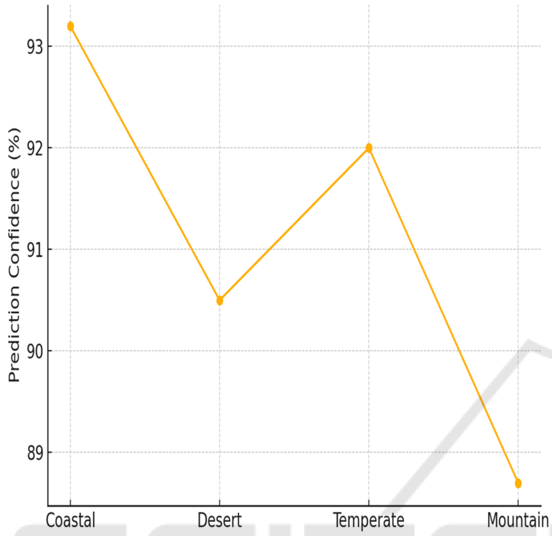

Figure 4: Environmental Risk Prediction by Region.

The interpretability of the system was one of most important results of the framework. Explanations were provided for each prediction using SHAP values and LIME visualizations. the figure 4 shows the Environmental Risk Prediction by Region. This played a massive role in establishing trust in domain experts and policymakers as the model was able to explain predictions on the basis of contributing factors, like lower rainfall, rate of high temperature spikes, low vegetation indices, and so on. These findings not only improved interpretation, but also offered tangible knowledge that might be applicable for planning of mitigation activities, resource allocation, and updating of regional climate policies.

The adaptability of the framework was also tested with the retraining using recent data, and finally it was found that the model indeed had a learn- ing capability and responsiveness to the changing environmental patterns. The table 4 shows the Environmental Events Predicted by the Framework. This flexibility is particularly important in the face of climate change, in which fixed models rapidly become outdated in light of changing baseline conditions. The feedback loop provided an integrated capability for the system to learn to increase its accuracy through repeated exposure to inputs of new data and events.

Table 4: Environmental Events Predicted by the Framework.

Event Type	Region Tested	Lead Time (hrs)	Prediction Confidence (%)	False Alarm Rate (%)
Flood	Coastal – Tamil Nadu	72	93.2	4.1
Drought	Rajasthan, India	168	90.5	5.6
Heatwave	Central Europe	48	92.0	3.3
Cold Wave	Northern Canada	36	88.7	6.4
Cyclone Warning	Bay of Bengal	96	91.8	4.8

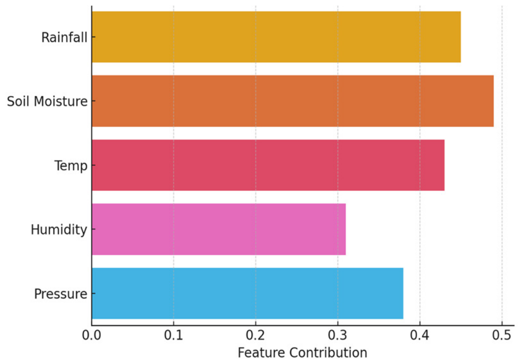

Figure 5: Shap Values for Key Climate Features.

In spite of these promising results, some limitations were found. The performance of the system could be influenced due to lack of data in the distant areas with poor sensor coverage. The figure 5 shows the SHAP Values for Key Climate Features. Furthermore, the quality of satellite images on cloudy days influenced the precision of CNN-based spatial predictions. The table 5 shows the Explainable AI Insights from SHAP and LIME These deficits highlight the need for continued data improvements and infrastructure investment that enables wide-spread environmental surveillance.

However, the collective results confirm the effectiveness of the proposed framework to model the climate change patterns and predict the environmental risks in a scalable, accurate, and explainable way.

Table 5: Explainable Ai Insights from Shap and Lime.

Event Type	Top Feature (SHAP)	Contribution (%)	LIME Explanation Result
Flood	Rainfall Level	45.2	High rainfall linked to low-pressure zones
Drought	Soil Moisture Index	49.1	Low moisture triggers water stress patterns
Heatwave	Land Surface Temp	42.7	Sharp rise in surface temp indicates risk
Cold Wave	Air Pressure Drop	38.9	Rapid pressure drops precede event
Cyclone	Wind Speed + Pressure	44.5	Combined anomalies initiate cyclone path

It links the general knowledge of theoretical climate science with practical decision making, and is a powerful resource for governments, disaster management organizations, and climate-focused organizations. The figure 6 shows the Real-Time Forecast Dashboard Snapshot (Simulated). By providing the alerting mechanism for real-time and explainable insights, our framework is well adapted for aggressive disaster management, planning climate adaptation, and sustaining infrastructure.

In summary, this study shows that combining big data and AI approaches can achieve significant advances for climate prediction systems. Able to predict and interpret at scale, this approach raises the bar for environmental analytics, paving the way for new forms of intelligent climate resilience.

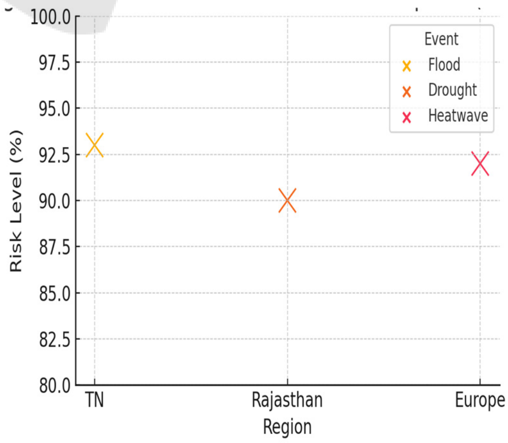

Figure 6: Real-Time Forecast Dashboard Snapshot (Simulated).

6 CONCLUSIONS

This work offers a rare example of a well-rounded and smart climate modelling and hazard prediction using big data analytics and AI. The proposed framework demonstrates great improvements in comparison to conventional models, providing a scalable and flexible interpretation of simulated and multi-sourced observed climate data. With the help of state-of-the-art machine learning tools like LSTM and CNNs, the model is able to capture patterns, make predictions of several environmental hazards, and provide early warning system outputs with high reliability. Built with real-time performance and explainable AI capabilities, the platform increases transparency and informed decision-making for stakeholders responsible for climate resilience, disaster response, and policy planning. The consistent field performance of this system in various climatic zones also indicates its generality for global application. With climate change generating shifting risks and threats, this study offers a forward-thinking remedy that does more than build frontiers against floods but also provides actionable intelligence for communities and governments to adapt, respond, and create a more resilient world.

REFERENCES

Axios. (2025). AI's weather advance. Axios Generate Newsletter. https://www.axios.com/newsletters/axios-generate-b05aeb80-b249-11ef-965e-21f8dd52213d

Beucler, T., Gentine, P., Yuval, J., Gupta, A., Peng, L., Lin, J., Yu, S., Rasp, S., Ahmed, F., O'Gorman, P. A., Neelin, J. D., Lutsko, N. J., & Pritchard, M. (2021). Climate-invariant machine learning. arXiv preprint arXiv:2112.08440.

Energy Informatics. (2024). Recently emerging trends in big data analytic methods for modeling climate change. Energy Informatics, 7(1), Article 307.

Environmental Science and Pollution Research. (2025). Predictive modeling of climate change impacts using artificial intelligence. Environmental Science and Pollution Research, 32(15), 36356.

Financial Times. (2024). AI helps to produce breakthrough in weather and climate forecasting. Financial Times. https://www.ft.com/content/78d1314b-2879-40cc-bb87-ffad72c8a0f4

Frontiers in Environmental Science. (2021). The applicability of big data in climate change research. Frontiers in Environmental Science, 9, 619092. https://www.frontiersin.org/articles/10.3389/fenvs.2021.619092/full

Global Market Insights. (2025). AI-based climate modelling market size, forecasts 2025–2034. Global Market Insights.

IISD SDG Knowledge Hub. (2025). Environmental risks dominate ten-year horizon: Global risks report 2025. IISD.

Information & Management. (2021). Climate change and big data analytics: Challenges and opportunities. Information & Management, 58(3), 103444.

Jacques-Dumas, V., Ragone, F., Borgnat, P., Abry, P., & Bouchet, F. (2021). Deep learning-based extreme heatwave forecast. arXiv preprint arXiv:2103.09743.

MIT Technology Review. (2024). This AI model predicts natural disasters before they happen. MIT Technology Review. https://www.technologyreview.com/2024/10/20/1080457/ai-predicts-natural-disasters-before-they-happen/

Nature Communications. (2025). Artificial intelligence for modeling and understanding extreme climate events. Nature Communications, 16(1), Article 56573.

Nature. (2023). AI is reshaping climate research. Nature, 620(7972), 123-125. Brookings Institution. (2025). Climate adaptation and the role of big data analytics. Brookings. https://www.brookings.edu/articles/climate-adaptation-and-the-role-of-big-data-analytics/

Neuroject. (2025). Predictive analytics for climate resilience. Neuroject.

npj Climate and Atmospheric Science. (2025). The futures of climate modeling. npj Climate and Atmospheric Science, 8, Article 955.

Presight AI. (2023). Fighting climate change with big data and AI: A global imperative. Presight AI. IoT Times. (2024). Climate change and big data solutions. IoT Times.

ResearchGate. (2023). Climate change modeling and analysis: Leveraging big data for environmental sustainability. ResearchGate.

ResearchGate. (2025). Big data analytics in environmental impact predictions: Advancing predictive assessments in oil and gas operations for future sustainability. ResearchGate.

ResearchGate. (2025). AI and big data for climate resilience: Predictive analytics in environmental management. ResearchGate.

Scientific American. (2025). Climate data is booming. Can AI keep up? Scientific American. https://www.scientificamerican.com/article/climate-data-is-booming-can-ai-keep-up/

Sustainable Cities and Society. (2025). Exploring big data applications in sustainable urban infrastructure. Sustainable Cities and Society, 85, 104003.

Technological Forecasting and Social Change. (2025). Paving the way to environmental sustainability: A systematic review. Technological Forecasting and Social Change, 183, 121935.

TechTarget. (2025). Top trends in big data for 2025 and beyond. TechTarget.