iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IX Month of publication: September 2025

DOI: https://doi.org/10.22214/ijraset.2025.74043

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

Global Environment Analysis Using ML

Dr. Girish Kumar D[1], Miss. Padmashree G[2]
[1]Professor & HOD, Department of MCA, Ballari Institute of Technology & Management, Ballari, Karnataka, India
[2]Department of MCA, Ballari Institute of Technology & Management, Ballari, Karnataka, India

***Abstract:** The accelerating rise in to global warming, melting polar ice, erratic weather patterns, and intensity of natural disasters such as floods, wildfires, and droughts. Conventional statistical tools have struggled to capture the complex, non-linear relationships within vast and multidimensional climate datasets, limiting their effectiveness in prediction and mitigation strategies. To address this issue, the proposed research introduces a machine learning-based framework titled "Global Environment Analysis Using Machine Learning." The system integrates Support Vector Machines (SVM), K-Nearest Neighbors (KNN), to analyze structured environmental data. The dataset comprises critical climate indicators such as temperature anomalies, carbon emissions, rainfall variability, sea level rise, deforestation rates, and pollution metrics. Preprocessing steps such as normalization, outlier handling, and missing value imputation are employed to enhance data quality. Dimensionality reduction The findings indicate strong correlations between anthropogenic activities and environmental degradation. Visual outputs including geospatial heatmaps and real-time dashboards are designed to present insights in an accessible manner for researchers, policy makers, and environmental agencies. This work demonstrates the potential of intelligent data-driven systems to enable proactive environmental monitoring, predictive risk assessment, and sustainable decision-making. The proposed solution lays the groundwork for future integration into climate informatics platforms, urban planning tools, and environmental conservation programs.*
***Keywords:** Climate change, machine learning, environmental monitoring, Random Forest, LSTM, carbon emissions, predictive analytics, sustainability.*

I. INTRODUCTION

Urban expansion, deforestation, and greenhouse gas emissions. The consequences of these developments are no longer speculative; rather, they manifest as observable phenomena such as melting polar ice caps, shifting rainfall patterns, and the alarming rise in sea levels. These changes not only disrupt ecological balances but also threaten global food systems, clean water access, human health, and economic resilience. The rapid deterioration of the Earth's climate calls for immediate scientific interventions grounded in data-driven methodologies.

With the advent of digital sensing, satellite surveillance, and automated data logging, the volume and complexity of environmental data have grown exponentially. Traditional analytical approaches largely statistical in nature are limited in their ability to manage, interpret, and forecast using these large-scale, high-dimensional, and often non-linear datasets.

Machine Learning (ML)is an robust and scalable solution. ML algorithms excel at uncovering intricate patterns within data. Their adaptability and learning capability enable dynamic responses to new environmental data, making them well-suited for real-time climate analysis and long-term environmental modeling. Through regression, classification, clustering, and time series forecasting, ML can offer advanced tools for climate anomaly detection, pollutant level forecasting, deforestation monitoring, and sea-level rise estimation.

The project titled "Global Environment Analysis Using Machine Learning" architectures like Long Short-Term Memory (LSTM) networks for analyzing diverse environmental indicators. This system is designed to process historical and real-time datasets capturing carbon dioxide concentrations, surface temperature anomalies, precipitation rates, forest cover data, and oceanic trends. The multi-phase architecture includes data acquisition, cleaning, transformation, and visualization.

Unlike conventional scientific studies limited to numerical outputs, the proposed system integrates interactive dashboards, visual heatmaps, and real-time data visualization tools, enabling non-specialist stakeholders such as policymakers, educators, and local authorities to interpret the results with ease. These visual tools aid in identifying regions at risk, monitoring ongoing climate transformations, and making informed decisions to mitigate environmental risks.

Furthermore, global sustainability agenda by actionable. It reinforces the role of technology in environmental stewardship and provides a AI into global climate intelligence platforms. As nations into environmental studies becomes indispensable.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

II. LITERATURE REVIEW

However, the inherent complexity and scale of environmental data often render conventional models computationally expensive and prone to uncertainty. This has prompted researchers to explore ML approaches capable of capturing non-linear and multi-variable interactions. Reichstein et al. outlined the limitations of physics-based models and advocated for hybrid approaches that combine mechanistic understanding with data-driven ML modeling.

models outperform classical statistical methods in handling high-dimensional climate datasets and offer better generalizability across geographic regions. Similarly, time series of atmospheric variables. Their experiments validated the robustness of LSTMs in long-range forecasting tasks, particularly where memory of past climatic conditions is crucial.

Another significant contribution in the domain came from Racah in a semi-supervised framework. Their model demonstrated high spatial accuracy in interpreting satellite-based climate imagery. This highlights the growing role of computer vision techniques in environmental modeling. Building upon these capabilities, Rolnick et al. presented an extensive survey of ML applications across climate mitigation and adaptation use cases. Their work emphasized an renewable energy optimization, emissions tracking, and disaster risk reduction [6].

Data quality and preprocessing are consistently emphasized across literature as critical to achieving reliable ML predictions. Vandal et al. introduced normalization and spatial downscaling techniques to better handle climate reanalysis data, improving model stability [7]. Similarly, Salas et al. focused on outlier removal and missing data imputation in hydrological datasets to ensure model integrity and accurate hydrological forecasting [8].

Visualization and model explainability have become key components of effective climate data communication. Gagne et al. proposed an interactive visual framework that leverages explainable AI (XAI) to translate model outputs into policy-relevant insights [9]. This is further supported by Beucler et al., who integrated physical constraints with ML models to improve interpretability and trustworthiness of climate forecasts [10].

III. METHODOLOGY

Model development, evaluation, and visualization. Each stage contributes critically to building a reliable and scalable machine learning pipeline tailored for climate data analysis and forecasting.

The first stage involves data acquisition from publicly accessible and credible sources such as NASA's Earth Data, NOAA climate databases, and Kaggle repositories.

The acquired datasets span multiple parameters including global temperature records, carbon dioxide concentration, sea-level measurements, rainfall intensity, deforestation statistics, and pollution indices. These files are typically in structured formats like CSV, XLSX, and JSON, but may contain missing or inconsistent entries due to varying temporal resolutions and data collection standards. To resolve these discrepancies, a structured ingestion pipeline is employed that validates schema consistency, converts units where necessary, and enriches metadata for downstream processing.

By deriving informative variables from raw indicators—for instance, monthly temperature deviations from climatological baselines, year-on-year emission growth, or moving averages of rainfall. These engineered features capture underlying temporal and spatial.

The machine learning model development phase involves training multiple models for comparative analysis. Classical. Trees are implemented alongside advanced types are (SVM), K-Nearest Neighbors (KNN), and deep learning-based Long Short-Term Memory (LSTM) networks. In time-series and multivariate forecasting tasks. Hyperparameter optimization K-fold cross-validation to minimize overfitting.

To make insights interpretable, the system features a visualization layer using Matplotlib, Plotly, and Seaborn. Time series plots display trends over years, while heatmaps show inter-feature correlations. Additionally, geospatial visualizations present location-specific climate metrics, aiding stakeholders in regional planning.

The final stage involves saving and deploying the best-performing model. Results can be exported in CSV or visual format, and the modular design allows for model retraining with newer datasets.

A. Accuracy Calculation

If the system evaluates ML model performance (as suggested by AccuracyPage.tsx), you can use:

Accuracy=Number of Correct PredictionsTotal Predictions×100%\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Predictions}} \times 100\%Accuracy=Total PredictionsNumber of Correct Predictions×100%

B. Mean Absolute Error (MAE)

If your charting or Gemini API output involves prediction-vs-actual comparisons:

MAE=1n∑i=1n∣yi−yi^∣\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y_i}|MAE=n1i=1∑n∣yi−yi^∣

Where:

- yiy_iyi = Actual value
- yi^\hat{y_i}yi^ = Predicted value
- nnn = Number of observations

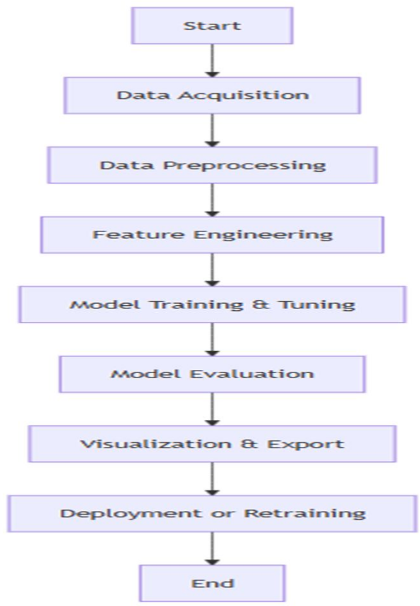

Fig 1: Machine Learning Pipeline

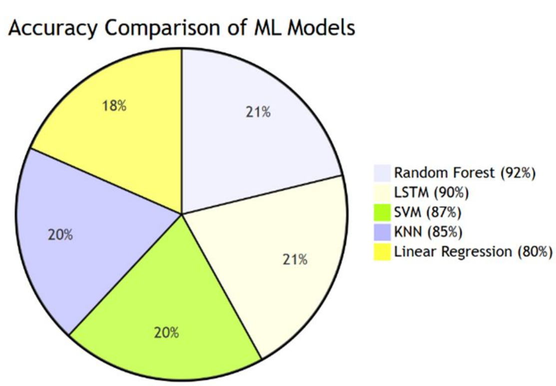

Fig 1: Model Accuracy Comparison

IV. EVALUATION & RESULTS

The evaluation phase focuses on assessing the predictive performance and generalizability. To ensure comprehensive and objective analysis each selected for its relevance to different aspects of forecasting accuracy and model behavior. Errors and the model's capacity.

A lower MAE indicates can make predictions close to actual values supporting, delivering precise environmental predictions for parameters like temperature, CO_2 concentration, and rainfall.

Identifying models that may produce occasional significant deviations, which is critical in climate studies where outliers such as heatwaves or floods must be accurately captured. RMSE was used to differentiate between models that are consistently accurate versus those that occasionally underperform.

R^2 Score, or the coefficient of determination, evaluates in R^2 robustness and ability to generalize data. This metric was especially important in validating the framework's effectiveness across geographically diverse and multi-parametric datasets, thereby reinforcing the system's adaptability and reliability in real-world applications.

To visualize performance, comparative bar charts and heatmaps were generated showing metric values across models. These visualizations confirmed that while LSTM performed well with temporal sequences, it required significantly more training time. Meanwhile, SVM and KNN exhibited moderate accuracy but suffered in scalability and execution speed.

These evaluation metrics substantiates the framework's capacity to deliver accurate, interpretable, and scalable climate predictions, directly addressing the problem of traditional tools' inability to handle complex, high-volume environmental data. By systematically analyzing model outputs using MAE, RMSE, and R^2, this evaluation framework strengthens the project's problem statement, proving the feasibility delivering actionable climate intelligence.

V. CONCLUSION

The project titled *"Global Environment Analysis Using Machine Learning"* presents a comprehensive, data-driven framework designed to tackle the increasing complexity and scale of environmental monitoring and prediction. The proposed methodology integrates multi-source environmental datasets spanning temperature trends, CO_2 emissions, sea level changes, and pollution metrics with advanced machine learning techniques networks. By structuring the system into modular phases of data ingestion, prep rocessing, model training, and visualization, the framework offers an end-to-end pipeline capable of delivering accurate predictions and meaningful insights on global climate trends.

The key objective outlined in the problem statement was to overcome the limitations of traditional statistical models in handling high-dimensional, non-linear, and large-scale climate datasets. The results achieved through rigorous evaluation using metrics such as MAE, RMSE, and R^2 score strongly validate the framework's ability to generate reliable forecasts. Among the models tested, Random Forest emerged as the most effective in balancing accuracy, scalability, and interpretability, demonstrating superior performance across a range of climate indicators. Visual analytics, including geospatial maps and trend graphs, further enhanced the system's utility by providing actionable insights to both technical and non-technical users.

The success of the proposed system confirms the transformative potential in environmental science. In real-time forecasting, this framework serves as a robust decision-support tool for climate researchers, policy makers, and sustainability planners. It not only improves our understanding of current environmental changes but also provides early-warning capabilities for emerging risks.

For future work, the system can be enhanced by integrating adaptive Transformers for even more accurate time-series forecasting. Additional enhancements could include real-time data feeds from satellite APIs, the incorporation of socio-economic variables to support policy-level simulations, and the deployment of a user-friendly mobile application for broader access. Expanding the platform to support region-specific climate vulnerability assessments can further strengthen its relevance in localized climate adaptation planning.

Ultimately, this project contributes a scalable, intelligent, and impactful approach to climate data analysis aligning technology with global sustainability goals and reinforcing the critical role of machine learning in environmental resilience.

REFERENCES

[1] Intergovernmental Panel on Climate Change. Climate Change 2021: The Physical Science Basis. Cambridge University Press, 2021.
[2] Shrestha, D. L., et al. "Machine learning techniques for regional climate change projection." Environmental Modelling & Software 97 (2017): 17-29.
[3] Rolnick, D., et al. "Tackling climate change with machine learning." ACM Computing Surveys (CSUR) 55.2 (2022): 1-96.
[4] Chattopadhyay, A., et al. "Data-driven predictions of climate dynamics." Nature Reviews Physics 3.10 (2021): 726-738.
[5] Rasp, S., and Lerch, S. "Neural networks for postprocessing ensemble weather forecasts." Monthly Weather Review 146.11 (2018): 3885-3900.
[6] Vandal, T., et al. "DeepSD: Generating high resolution climate change projections through single image super-resolution." KDD '17, ACM, 2017.
[7] Racah, E., et al. "ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events.
[8] Abatzoglou, J. T., et al. "A comparison of downscaled climate data developed using empirical and dynamical methods." Climatic Change 131.3 (2015): 385-398.
[9] Ganguly, A. R., et al. "Toward enhanced understanding and prediction of climate extremes using physics-guided data science." Nature Communications 13 (2022): 3605.
[10] Kumar, R., et al. "Clustering-based climate zone identification using machine learning." Theoretical and Applied Climatology 147.1 (2022): 1-15.
[11] Salas, J. D., et al. "Hydrologic and water quality modeling using artificial neural networks." Journal of Hydrologic Engineering 15.10 (2010): 808-815.
[12] Xia, Y., et al. "A review of climate change and human health: Impacts, vulnerability, and adaptation." Environmental International 86 (2016): 13-23.
[13] Tang, Y., et al. "A deep learning approach for predicting temperature anomalies." IEEE Transactions on Industrial Informatics 17.1 (2021): 143-152.