UNIVERSITY OF HERTFORDSHIRE

School of Physics, Engineering and Computer Science

Data Science Project (7PAM2002-0509-2024)

Date: 11/06/2025

AI-Driven Climate Change Model

Name: Gaddam Sai Kumar

Student ID: 23100456

Supervisor: Hariharan Dhiamani Saravana

Climate change is one of the most pressing global challenges of our time. Increasing temperatures, rising sea levels, and extreme weather events threaten both natural ecosystems and human societies. Traditional climate modeling methods, such as physics-based simulation models, have been valuable tools for understanding climate trends. However, these models often require complex computations and lack flexibility for exploratory analysis.

Artificial Intelligence (AI) presents a promising complementary framework to climate prediction. By using machine learning algorithms on historical climate data, predictive models can be created that can aid in the interpretation of future climate situations. In addition, explainable AI (XAI) methods can provide assurance of transparency, enabling domain specialists and policy-makers to believe and read AI-based predictions.

This project aims to leverage AI to build an interpretable climate change model that predicts temperature trends and assesses climate risks using accessible, high-quality datasets.

Problem Statement

Though there are many climate models, many are vastly complicated, computationally costly, and not very transparent. Decisions on public policy and the environment demand models that are both **accurate and understandable**. Furthermore, most existing AI applications in climate modeling focus on a single task (such as forecasting) and do not integrate classification or explainability in a cohesive framework.

There is a clear need for an **integrated AI-driven pipeline** that combines accurate forecasting, risk classification, and interpretable explanations using a practical, machine learning-ready dataset.

Aim of the Project

The aim of this project is to develop an AI-driven climate change model capable of predicting temperature trends, classifying climate risk levels, and providing interpretable insights to support data-driven climate policy and planning.

Objectives

- To build and train regression models for temperature prediction using historical climate data.
- To develop classification models for categorizing climate risk based on temperature patterns.
- To apply explainable AI tools (SHAP, LIME) to interpret model predictions.
- To perform scenario analysis for future climate forecasting under different CO_2 emission pathways.

- To present findings in a transparent and accessible format for use by environmental stakeholders.

Research Questions

- **RQ1:** How effectively can AI models (Random Forest, LSTM, Transformer) predict average temperature based on climate indicators?
- **RQ2:** What are the most influential features driving climate predictions?
- **RQ3:** How can explainable AI techniques improve transparency in climate modeling?

Expected Outcomes

- A trained and validated AI model capable of predicting temperature trends.
- Classification of temperature into risk categories (Low, Moderate, High, Extreme).
- SHAP and LIME outputs to explain model decisions.
- Scenario-based climate risk analysis for selected countries.
- A Google Colab notebook documenting all code, preprocessing steps, and results for reproducibility.

Literature Review

AI is increasingly used to model and mitigate climate change impacts. Recent studies demonstrate the value of combining ensemble learning, deep learning, and explainable AI for environmental forecasting. This section reviews five key papers that inform this project's methodology.

Agrawal (2023)- Agrawal examines the role of generative AI and optimization intelligence in sustainability applications.

Key contributions:

- Highlights interpretable models like Random Forest and XGBoost for optimized, resource-aware forecasting.
- Stresses the importance of model transparency in environmental decision-making.
- Supports the use of ensemble learning for climate applications.

Amiri et al. (2024) This comprehensive survey focuses on AI strategies for climate change mitigation.

Key contributions:

- Identifies tree-based ensembles (Random Forest, Gradient Boosting) as optimal for modeling nonlinear climate relationships.
- Endorses LSTM for capturing temporal dynamics in climate data.
- Provides a taxonomy of AI approaches in climate science.

Cavus et al. (2025) Although focused on electric vehicles, this study offers insights into time-series modeling with LSTM.

Key contributions:

- Demonstrates LSTM's strength in modeling sequential, time-dependent patterns.
- Validates the use of deep learning in domains requiring predictive maintenance — applicable to climate forecasting.
- Supports the choice of LSTM in this project for temporal climate modeling.

Chakraborty et al. (2021) This study applies explainable AI (XAI) to environmental modeling.

Key contributions:

- Uses SHAP to interpret climate-related predictions.
- Demonstrates the value of scenario-based forecasting.
- Aligns with this project's use of SHAP and LIME to enhance model transparency.

Chang & Kidman (2023) The authors discuss ethical considerations in applying AI to environmental education.

Key contributions:

- Argue for interpretable AI over black-box models.
- Emphasize transparency and trust in AI-driven climate tools.
- Reinforce the selection of explainable models such as Gradient Boosting and Random Forest in this project.

The reviewed studies collectively emphasize the value of combining:

- Ensemble models (Random Forest, Gradient Boosting) for accuracy and interpretability.
- Deep learning (LSTM) for time-series forecasting.
- Explainable AI (SHAP, LIME) for transparency in climate modeling.

However, existing research often focuses on isolated techniques. This project addresses the gap by integrating forecasting, classification, interpretability, and scenario analysis into a unified AI-driven climate modeling pipeline.

Research Gap

While many studies focus on either forecasting or interpretability, few integrate both into a single cohesive pipeline. Additionally, most AI-driven climate models do not offer built-in scenario analysis tools for future risk planning. There is also a gap in applying explainable AI methods to ensure transparency for policymakers.

Bridging the Gap

This project bridges the identified research gap by:

- Developing an integrated pipeline combining regression, classification, and interpretability.

- Using both tree-based ensembles and deep learning models for comparative insights.

- Applying SHAP and LIME to ensure model transparency.

- Conducting scenario-based forecasting to simulate future climate risks.

- Using an accessible dataset to ensure reproducibility and usability.

Dataset Details

The Kaggle Climate Change Dataset (by Bhadra Mohit) was selected for this project. Initially, ClimateSet (CMIP6) was considered, but due to its complexity (NetCDF format) and computational demands, Kaggle's dataset was chosen for its clean tabular structure, suitability for machine learning, and alignment with project timelines. This change was made after project discussions and supervisor guidance.

Dataset Overview:

- **1000 rows**, **10 columns**.

- No personal/sensitive data — fully compliant with UH ethical guidelines.

Columns include:

- Year, Country, Avg Temperature (°C), CO_2 Emissions (Tons/Capita), Sea Level Rise (mm), Rainfall (mm), Population, Renewable Energy (%), Severe Meteorological Events, Forest Area (%).

Feature Selection & EDA:

- **Seven features** selected: CO_2 Emissions, Sea Level Rise, Rainfall, Population, Renewable Energy, Extreme Weather Events, Forest Area.

- **EDA methods:** Correlation heatmaps, Random Forest feature importance.

- **Standardization:** Features standardized using StandardScaler to enhance model training.

This project proposes the development of an AI-driven climate change model that integrates forecasting, classification, and explainability. The literature review supports this approach, while the identified research gap highlights the need for an integrated pipeline combining multiple modeling elements.

The Kaggle Climate Change Dataset offers a practical, machine learning-ready foundation for building these models. Expected outcomes include accurate temperature forecasting, risk classification, transparent model explanations, and scenario-based planning outputs.

The final deliverable will include a Google Colab notebook documenting the entire modeling pipeline, which will be shared with the supervisor upon completion of final testing and validation.

References:

[1] Agrawal, K.P. (2023). Organizational sustainability of generative AI-Driven optimization intelligence. *Journal of Computer Information Systems,* pp. 1–15. https://doi.org/10.1080/08874417.2023.2286540 .

[2] Amiri, Z., Heidari, A. and Navimipour, N.J. (2024). Comprehensive survey of Artificial intelligence Techniques and Strategies for climate change mitigation. *Energy,* 308, p. 132827. https://doi.org/10.1016/j.energy.2024.132827 .

[3] Cavus, M., Dissanayake, D. and Bell, M. (2025). Next generation of electric vehicles: AI-Driven approaches for predictive maintenance and battery management. *Energies*, 18(5), p. 1041. https://doi.org/10.3390/en18051041 .

[4] Chakraborty, D. et al. (2021). Scenario-based prediction of climate change impacts on building cooling energy consumption with explainable artificial intelligence. *Applied Energy*, 291, p. 116807. https://doi.org/10.1016/j.apenergy.2021.116807 .

[5] Chang, C.-H. and Kidman, G. (2023). The rise of generative artificial intelligence (AI) language models - challenges and opportunities for geographical and environmental education. *International Research in Geographical and Environmental Education*, 32(2), pp. 85–89. https://doi.org/10.1080/10382046.2023.2194036 .