Multimodal Model for Extreme Weather Prediction and Socioeconomic Impact Assessment in Kenya

**1. Introduction**  
Extreme weather events, such as droughts, floods, and heatwaves, pose significant risks to Kenya’s economy, infrastructure, and vulnerable populations. Predicting these events accurately and estimating their socioeconomic impact can support early warning systems, disaster preparedness, and policy interventions. This research proposes a lightweight, multimodal machine learning model that integrates time-series climate data, socioeconomic vulnerability indices, and climate teleconnections to enhance extreme weather prediction and impact assessment.

**2. Background of the Study**  
Kenya experiences recurrent extreme weather events, significantly impacting agricultural productivity, water resources, and human settlements. Climate change has intensified these events, leading to increased displacement and economic losses. Existing weather forecasting models primarily focus on meteorological data without integrating socioeconomic factors, which are crucial for assessing the broader impact of extreme weather. This study aims to bridge this gap by incorporating both climate and socioeconomic data into a unified prediction framework.

**3. Problem Statement**  
Despite advances in climate modeling, Kenya still faces challenges in localized extreme weather prediction and impact estimation due to data gaps, computational costs, and lack of integrated approaches. Current weather forecasting models, such as Numerical Weather Prediction (NWP) systems (e.g., GFS, WRF, ECMWF), statistical models (e.g., XGBoost, ARIMA), and satellite-based models (e.g., ERA5, TRMM), have notable limitations in Kenya. NWP models are computationally expensive and struggle with localized forecasts due to sparse observational data. Statistical models often rely on historical trends and lack real-time adaptability, while satellite-based models face resolution constraints and data latency issues. Studies, including those by Bauer et al. (2015) and reports from ICPAC and KMD, highlight these challenges, underscoring the need for an integrated machine-learning approach to improve extreme weather prediction and impact assessment.

**4. Objectives**

1. Develop a machine learning model that predicts extreme weather events using time-series climate data.
2. Integrate static socioeconomic vulnerability indices to estimate economic loss and displacement risk.
3. Implement a dynamic risk mapping system for intervention prioritization.
4. Compare the proposed model’s performance with traditional methods such as XGBoost.

**5. Scope of the Study**  
This study will focus on Kenya, using historical climate and socioeconomic data to train and evaluate the model. The primary climate datasets will be sourced from ERA5 reanalysis, while socioeconomic data will be obtained from WorldPop, EM-DAT, and KNBS reports. The research will cover prediction of key extreme weather events, including floods, droughts, and heatwaves, and will estimate their economic impact at a regional level. The study will not cover hyperlocal weather forecasting or sector-specific impacts, such as individual business losses.

**6. Methodology**

**6.1 Data Collection**

* **Climate Data:** ERA5 reanalysis data (daily temperature, precipitation, wind speed).
* **Socioeconomic Data:** WorldPop (population density), EM-DAT (disaster history), and KNBS reports.
* **Climate Teleconnections (Optional):** ENSO, NAO, and Indian Ocean Dipole (IOD) indices.

**6.2 Data Preprocessing**

* **Time-Series Data:** 30-day rolling window, normalization.
* **Socioeconomic Data:** Grid-aligned, merged into a single dataset.
* **Climate Teleconnections:** Simple time-aligned scalar values.

**6.3 Model Architecture**

**Feature Extractors**

* **Time-Series Branch:** LSTM processes 30-day sequences of temperature, precipitation, and wind speed (output: 64-dim embedding).
* **Static Socioeconomic Branch:** MLP processes vulnerability indices (output: 16-dim embedding).
* **Climate Context Branch (Optional):** MLP processes ENSO/NAO/IOD indices (output: 8-dim embedding).

**Fusion Mechanism**

* Concatenation of LSTM, static, and climate context embeddings.
* Joint Dense Layers: Dense(64) → ReLU → Dense(1) for binary classification.
* Additional Regression Output: Dense(1) for economic loss estimation.

**6.4 Training and Evaluation**

* **Loss Functions:** Binary cross-entropy for classification, mean squared error for regression.
* **Optimizer:** Adam with early stopping (patience=5).
* **Metrics:** AUC-ROC, F1-Score (classification); MAE, R² (regression).
* **Baseline Comparison:** XGBoost trained on time-series averages and static features.

**6.5 Deployment Strategy**

* **Daily Risk Scores:** Probability of extreme weather events.
* **Impact Estimates:** Predicted economic loss for regions.
* **Risk Maps:** Geospatial visualization of high-risk zones.
* **API Integration:** Real-time updates for early warning systems.

**7. Expected Outcomes**

1. A computationally efficient model capable of predicting extreme weather events with higher accuracy than traditional methods.
2. Improved socioeconomic impact estimation for disaster preparedness.
3. An interactive risk mapping system to guide policy and intervention efforts.
4. Open-source implementation for accessibility and real-world adoption.

**8. Significance and Impact**  
This research aligns with Kenya’s disaster risk reduction policies and enhances predictive capabilities for climate resilience. By integrating machine learning with multimodal data sources, it offers a scalable approach for early warnings, ultimately reducing economic losses and protecting vulnerable communities.