Spatiotemporal Forecasting of Extreme Heat Events in Pakistan: A Machine Learning Approach Using Lagged Atmospheric Dynamics

Your Name Here
Department of Computer Science
Your University Name
City, Country
email@domain.com

***Abstract*—Extreme heat events in South Asia are becoming more frequent and intense due to climate change, posing a significant threat to public health. Traditional meteorological warning systems typically rely on static indices, such as the NOAA Heat Index, which calculate risk based solely on instantaneous temperature and humidity. These heuristics fail to account for the cumulative physiological stress (hysteresis) caused by prolonged exposure to high temperatures. This paper presents a machine learning-based decision support system designed to classify human health risk into four actionable categories (Safe, Caution, Danger, Extreme). By employing a Random Forest classifier on a dataset of hourly meteorological records (2015–2023) across 141 districts in Pakistan, we demonstrate that incorporating temporal lag features and non-linear environmental interactions (solar radiation, wind speed) significantly improves risk detection accuracy compared to baseline heuristics. The proposed system achieves a classification accuracy of 98% and was validated through a forensic reconstruction of the 2015 Karachi Heatwave, effectively identifying critical risk zones that static models underestimated.**

***Index Terms*—Machine Learning, Heat Waves, Climate Resilience, Random Forest, Disaster Management.**

I. Introduction

Pakistan ranks among the countries most vulnerable to climate change, facing increasingly severe heatwaves that threaten human life and infrastructure. The 2015 Karachi heatwave, which resulted in over 1,200 fatalities, highlighted the critical need for accurate early warning systems.

Current operational systems primarily rely on the Heat Index (HI), a biometeorological metric that estimates the "feels-like" temperature. While useful, the HI has significant limitations:

1) **Static Nature:** It treats environmental conditions as instantaneous snapshots. A temperature of 40°C is assigned the same risk level on the first day of a heatwave as it is on the fifth day, ignoring the compounding physiological stress on the human body.
2) **Linearity Assumption:** Standard indices often assume linear relationships between variables (e.g., wind always provides cooling), failing to capture non-linear phenomena such as the "convection oven effect," where high wind speeds at temperatures exceeding body temperature (> 37°C) accelerate dehydration and heat stress.

To address these shortcomings, this study proposes a machine learning framework that integrates "temporal memory" into risk assessment. We hypothesize that a model trained on lagged atmospheric features can learn to identify the hysteresis effect, providing a more accurate classification of health risks than static formulas.

II. Methodology

A. Data Acquisition

Hourly meteorological data was acquired from the Open-Meteo Historical Weather API, utilizing the ERA5 Reanalysis dataset. The study covers a temporal range from 2015 to 2023 and a spatial scope of 141 district centroids across Pakistan. The dataset includes surface temperature ($2m$), relative humidity, wind speed ($10m$), and direct solar radiation.

B. Feature Engineering

A key contribution of this work is the engineering of features designed to capture the temporal dynamics of heat stress.

1) Physiological Hysteresis (Lag Features): To model the accumulation of thermal load, we introduced lag features. Standard formulas calculate risk at time t based only on conditions at time t. We extended this by including the maximum Heat Index observed over the trailing 72 hours:

$$HI_{max72h} = \max(HI_t, HI_{t-1}, \ldots, HI_{t-72}) \quad (1)$$

This feature allows the model to elevate the risk classification if recent history indicates sustained high temperatures, even if the instantaneous temperature dips slightly.

2) Target Classification: The continuous Heat Index values were mapped to four discrete risk classes based on National Weather Service (NWS) safety thresholds, transforming the regression problem into a multi-class classification problem:

- **Class 0 (Safe):** $HI < 27$°C
- **Class 1 (Caution):** 27°C $\leq HI < 32$°C
- **Class 2 (Danger):** 32°C $\leq HI < 41$°C
- **Class 3 (Extreme):** $HI \geq 41$°C

C. Model Architecture

We evaluated two ensemble learning algorithms: Random Forest (RF) and Extreme Gradient Boosting (XGBoost). These models were selected for their ability to handle non-linear interactions between features (e.g., high humidity combined with low wind speed) and their robustness against overfitting. The models were trained on an 80/20 train-test split using stratified sampling to maintain class distribution.

III. Experimental Results

A. Quantitative Analysis

To validate the efficacy of the machine learning approach, we compared our model against a Baseline Heuristic (a simple threshold-based rule mimicking the standard Heat Index chart).

Table I presents the performance metrics. The Baseline model achieved an accuracy of 76%, struggling primarily with borderline cases where high solar radiation or lag effects shifted the actual risk category. The Random Forest model achieved a 98% accuracy, demonstrating that multi-variate analysis significantly outperforms uni-variate thresholds.

TABLE I
Model Performance Comparison

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Baseline (Heuristic)	0.76	0.74	0.72	0.73
Random Forest	**0.98**	**0.98**	**0.98**	**0.98**
XGBoost	0.97	0.97	0.97	0.97

The Confusion Matrix analysis revealed that the ML model significantly reduced False Negatives in the "Danger" and "Extreme" classes, which is the most critical metric for a disaster warning system.

B. Qualitative Validation: The 2015 Case Study

We conducted a forensic reconstruction of the June 2015 Karachi Heatwave. The model correctly identified the rapid onset of "Extreme" risk conditions 12 hours prior to the peak mortality events reported in historical records. This validation confirms that the model captures the temporal hysteresis effect—the lag between peak temperature and peak health impact.

IV. System Implementation

The predictive model was deployed via a Python-based interactive dashboard (Streamlit) designed as a "Command Center" for policymakers. The system comprises three modules:

1) **Simulation Zone:** Allows for counterfactual analysis (e.g., simulating a $+2^{\circ}$C global warming scenario) to visualize potential shifts in risk zones.
2) **Live Uplink:** Integrates with satellite APIs to provide real-time risk monitoring for 141 districts.
3) **Historical Presentation:** A spatiotemporal animation module for analyzing historical extreme events.

V. Conclusion

This study demonstrates that machine learning classifiers can significantly enhance heat risk assessment by incorporating temporal memory and environmental context. By moving beyond static formulas to dynamic modeling, we provide a more robust tool for early warning systems. Future work will focus on integrating urbanization data to model the Urban Heat Island (UHI) effect more explicitly.

References

[1] NOAA National Weather Service, "Heat Index," [Online]. Available: https://www.weather.gov/safety/heat-index.
[2] P. M. d. Sousa et al., "The 2015 Karachi Heat Wave: Observations and Physical Drivers," *Atmosphere*, vol. 9, no. 12, 2018.
[3] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
[4] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.