DENGUE OUTBREAK FORECASTING USING CLIMATE DATA

A Predictive Approach to Strengthen Public Health Planning

Saurabh Rai,
3rd Year, IIT
Patna GUVI

Project 3

INTRODUCTION: CONFRONTING THE DENGUE CHALLENGE

Dengue fever, a significant public health threat, is a major mosquito-borne disease endemic across India. Its prevalence is deeply intertwined with environmental conditions.

Mosquito-Borne Disease

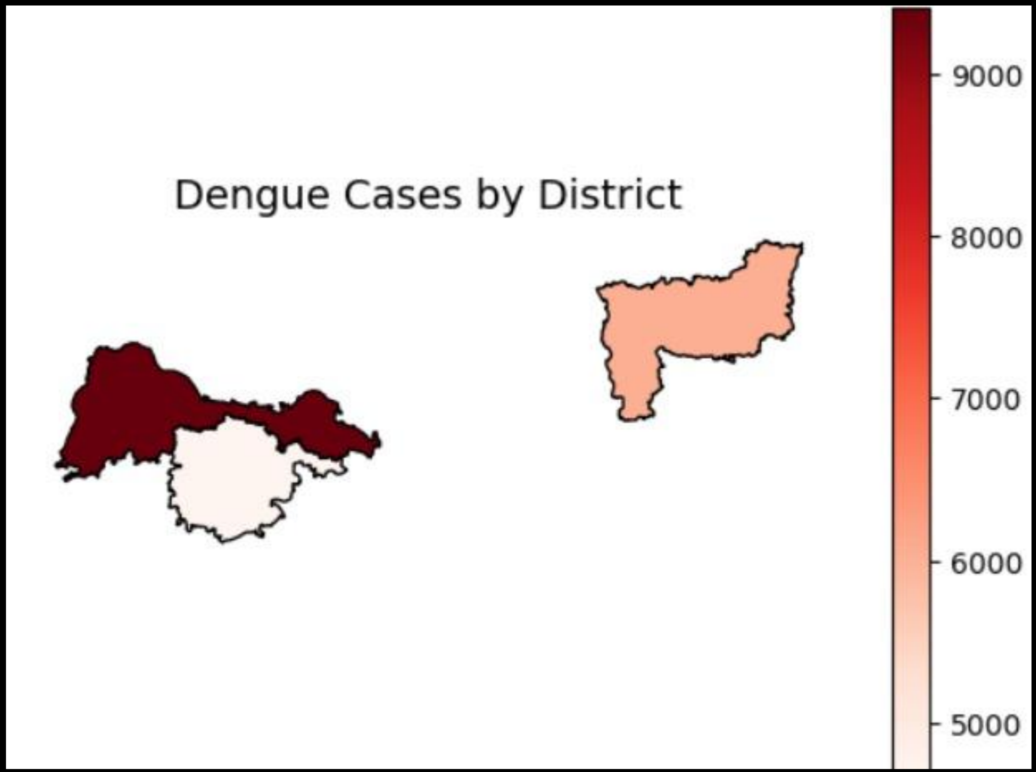
Dengue is transmitted by **Aedes aegypti** and **Aedes albopictus** mosquitoes, prevalent in tropical and subtropical regions.

Climate as a Driver

Key climate factors like temperature, rainfall, and humidity strongly influence mosquito life cycles and virus replication rates.

Our Project Aim

To develop a robust forecasting system providing early warnings, enabling proactive public health interventions.



THE PROBLEM: A GROWING BURDEN, UNTIMELY RESPONSE

Rising Dengue Burden

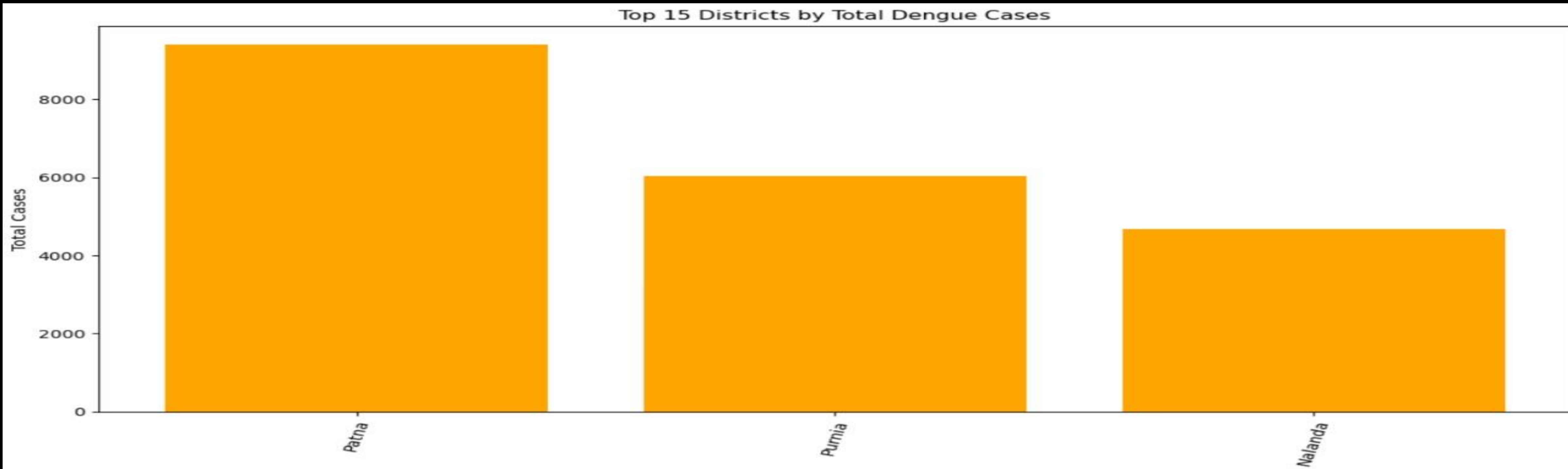
Climate variability and urbanization are escalating dengue incidence, straining healthcare systems.

Lack of Timely Alerts

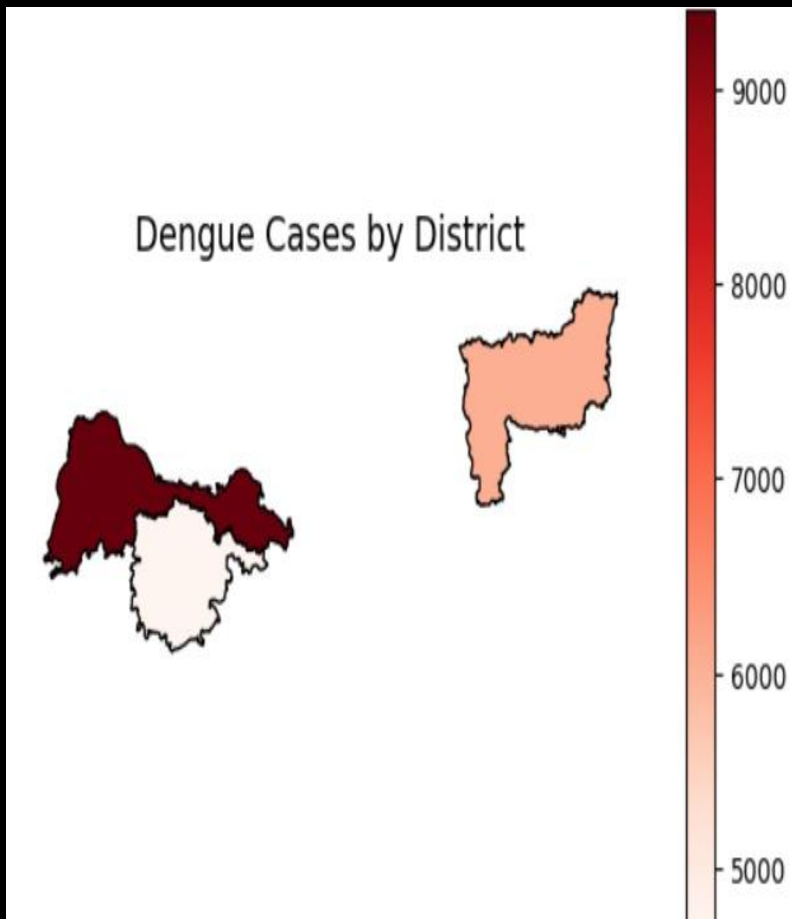
Current health systems often react to outbreaks rather than anticipating them, leading to delayed responses.

Critical Need: Accurate Forecasts

A predictive model capable of generating weekly forecasts is vital for proactive public health planning and resource allocation.



FOUNDATION OF PREDICTION: DIVERSE DATA SOURCES



Our forecasting model is built upon a comprehensive dataset integrating three critical information streams:

1 Historical Dengue Cases

Weekly, district-wise case counts provide the core time-series data for trend analysis.

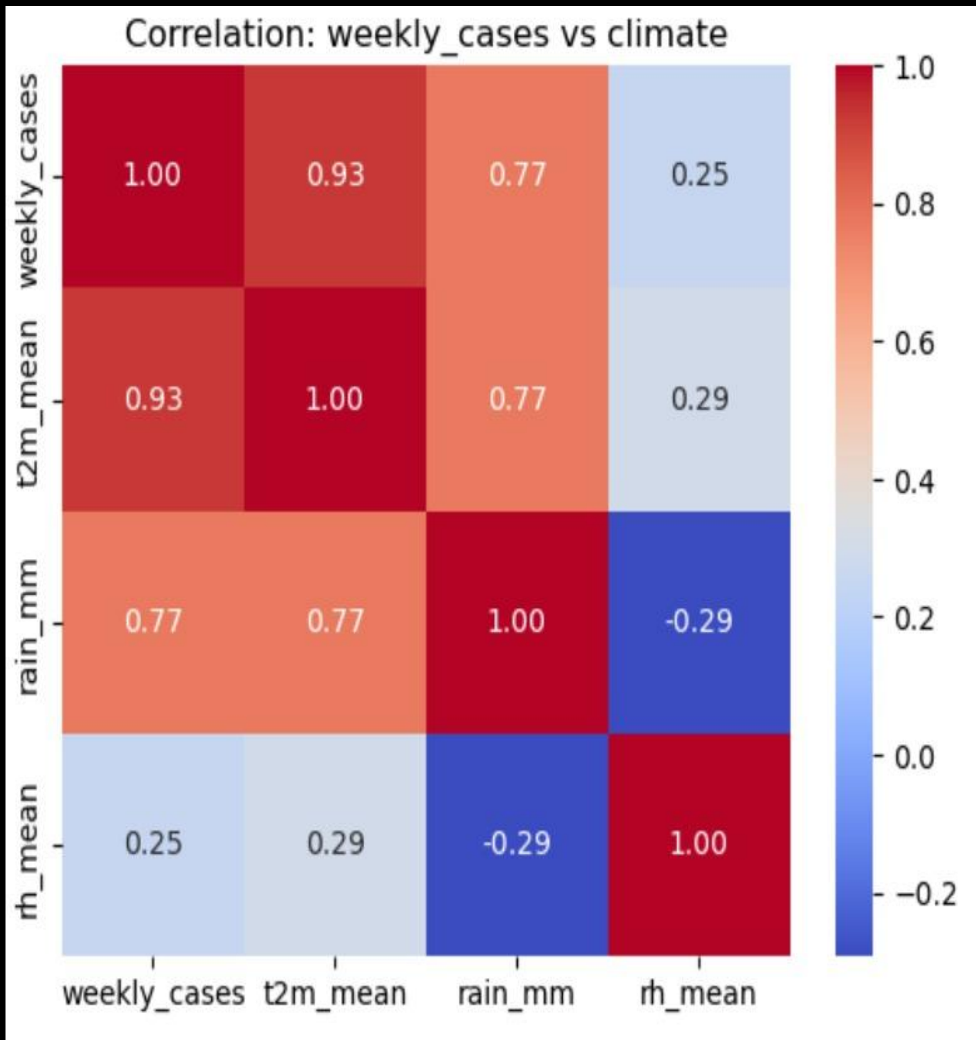
2 Climate Data (ERA5)

High-resolution data on temperature, rainfall, and humidity from the ERA5 reanalysis dataset serve as key environmental predictors.

3 Population Data

Baseline population figures account for exposure variability, normalizing case counts for better comparison.

All data were meticulously merged into a single, cohesive dataset to facilitate robust modeling.



PREPARING THE CANVAS: DATA PREPROCESSING STEPS

Before analysis, raw data underwent crucial transformations to ensure accuracy and consistency.

01

Date Standardization

Raw date formats were converted into a consistent weekly format for time-series alignment.

02

Missing Value Handling

Gaps in the dataset were addressed using a forward-fill imputation method to maintain data integrity.

03

Population Normalization

Dengue cases were adjusted by population density to ensure equitable comparison across different districts.

04

Final Dataset Refinement

The entire dataset was then validated, ensuring it was clean, complete, and ready for model training.

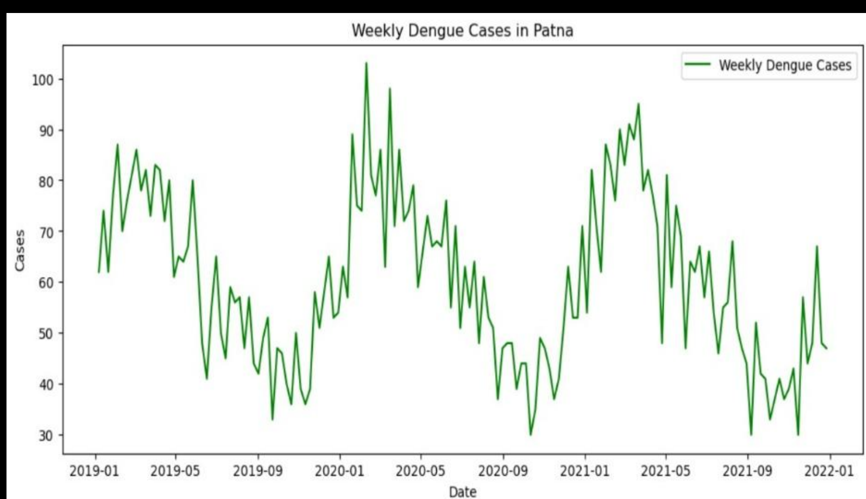
UNVEILING PATTERNS: EXPLORATORY DATA ANALYSIS (EDA)

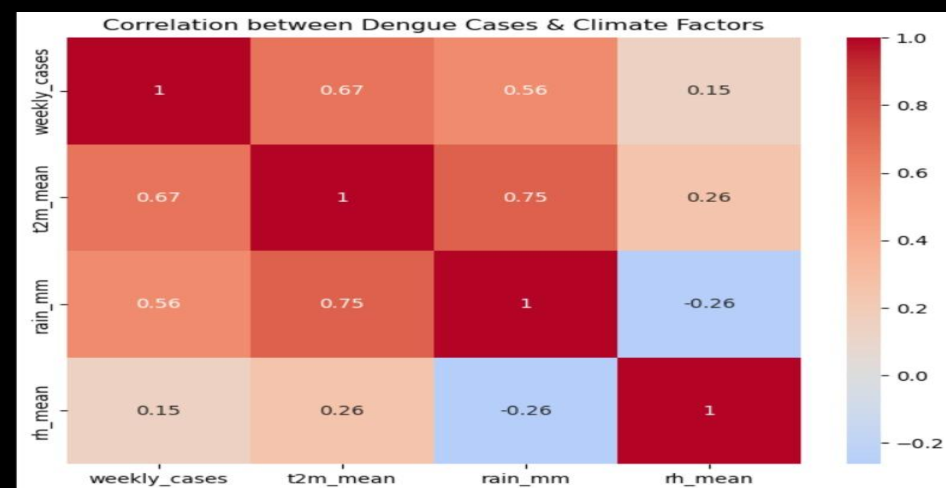
- Our initial data exploration revealed critical insights into dengue epidemiology and its relationship with climate:

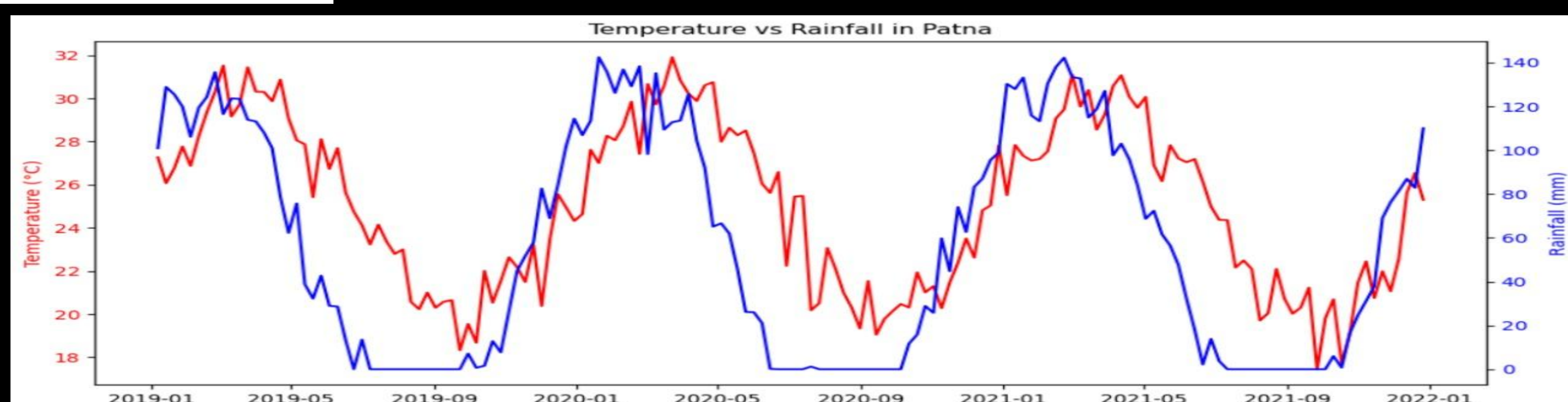
- **Seasonal Peaks:** Dengue cases consistently showed pronounced seasonal peaks, particularly during the monsoon months, correlating with increased mosquito breeding.

- **Strong Climate Correlations:** A clear positive correlation was observed: as temperature, rainfall, and humidity increased, so did dengue incidence. These climate variables are potent drivers of outbreaks.

- **Patna as a Hotspot:** Our analysis consistently identified Patna as a persistent hotspot, exhibiting higher and more regular dengue activity, indicating its vulnerability and potential for clearer predictive cycles.







THE ENGINE OF PREDICTION: ARIMAX MODEL

For our forecasting solution, we selected the ARIMAX (Autoregressive Integrated Moving Average with Exogenous Regressors) model. This powerful time-series model is ideally suited for capturing complex patterns in health data.

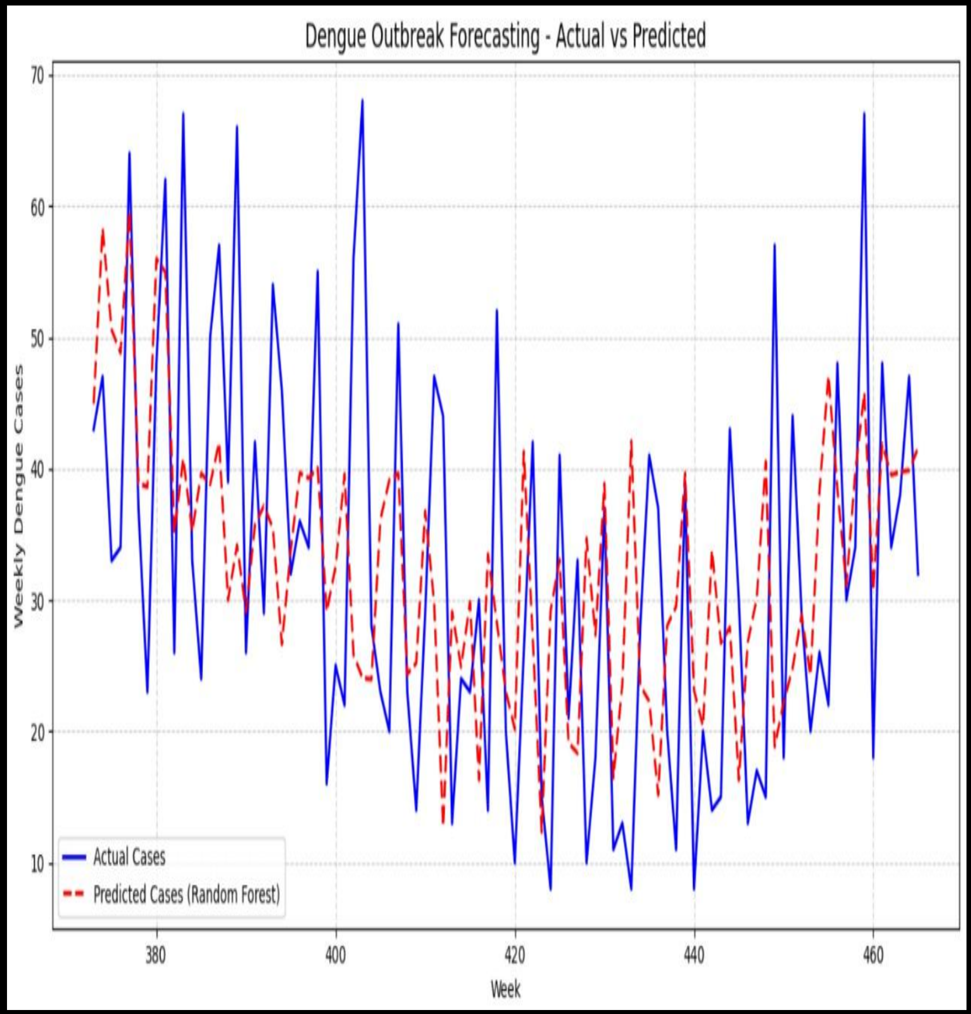
ARIMA Core

The ARIMA component effectively captures intrinsic time-series patterns: past dengue cases, underlying trends, and moving averages.

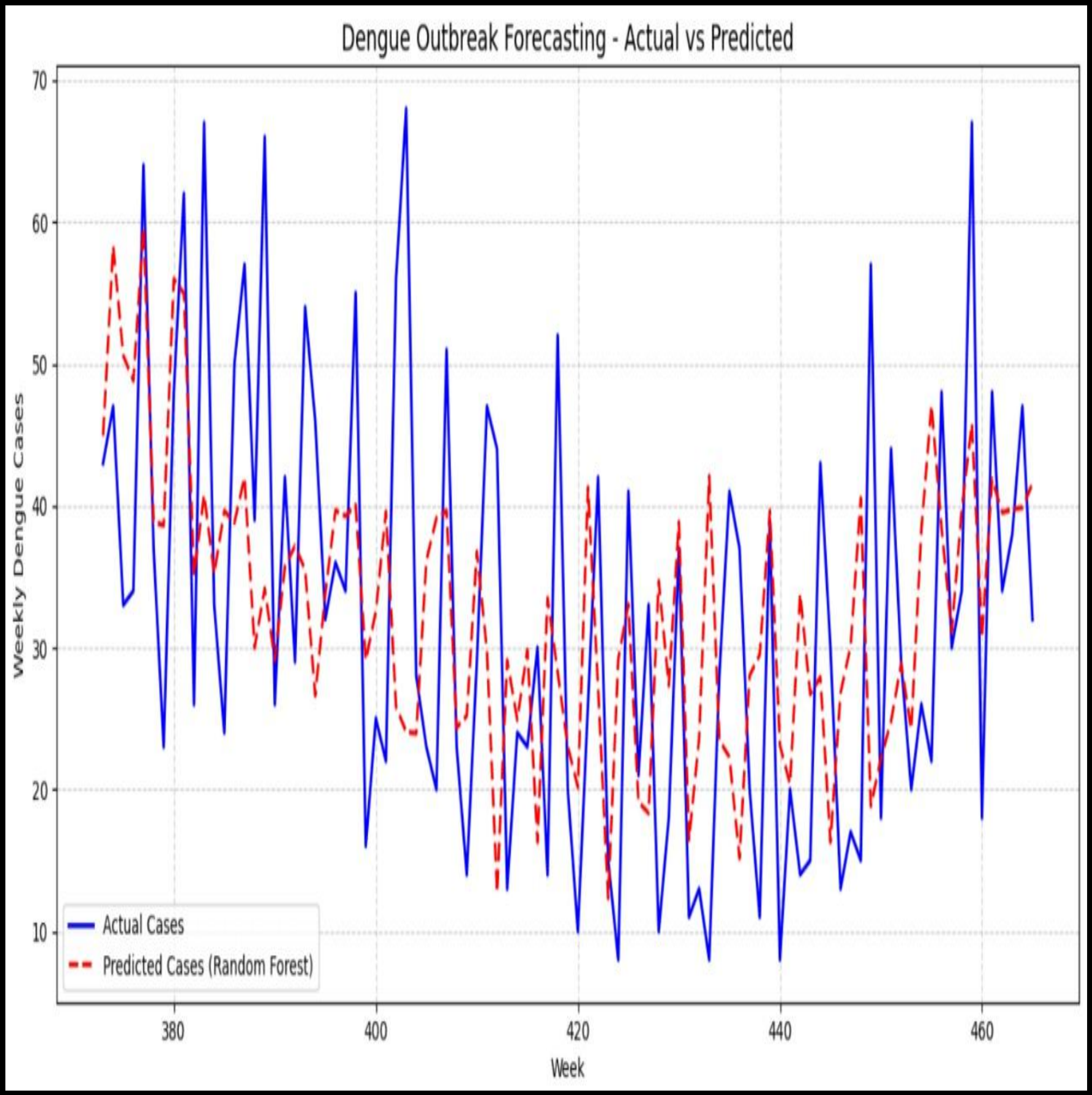
Exogenous Variables (X)

The 'X' in ARIMAX allows us to incorporate external, highly influential factors: **temperature, rainfall, and humidity**.

This hybrid approach ensures both accuracy by leveraging climate drivers and interpretability by clearly defining the influence of each component.



RIGOROUS EVALUATION: MODEL TRAINING & VALIDATION



To ensure the reliability and generalizability of our ARIMAX model, we followed a standard, robust training and validation protocol:

1

Data Split

The entire dataset was divided into an 80% training set and a 20% testing set, preserving the temporal order of the data.

2

Validation Metrics

Model performance was rigorously assessed using standard metrics: **Root Mean Squared Error (RMSE)** and **Mean Absolute Error (MAE)**.

3

Dual Training

The model was trained on both the national-level data and specifically on the Patna dataset to evaluate localized accuracy.

NATIONAL PERFORMANCE: CAPTURING OUTBREAK PEAKS

The ARIMAX model demonstrated strong predictive capabilities at the national level, effectively identifying periods of high dengue incidence.

26.47

RMSE Score

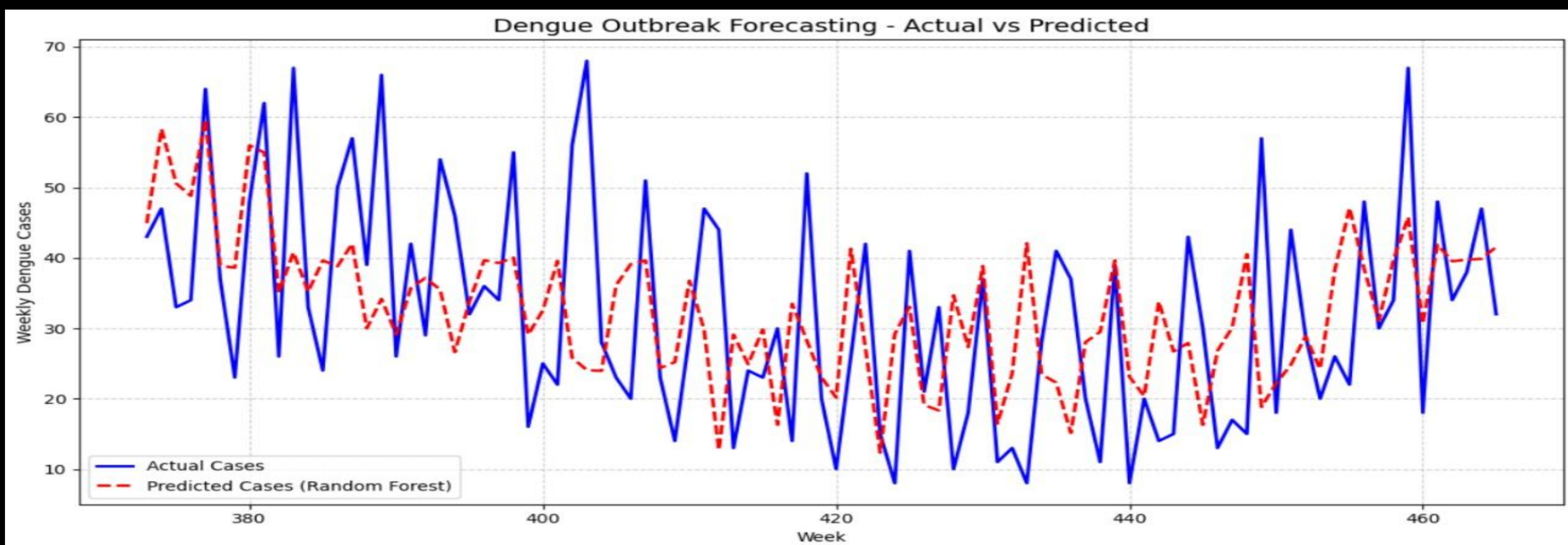
Indicating the average magnitude of the errors between predicted and actual values.

21.94

MAE Score

Representing the average absolute difference between forecast and observed values.

Crucially, the model successfully captured the significant outbreak peaks, providing a valuable early warning tool for public health authorities across the country.



PATNA FOCUS: ENHANCED ACCURACY FOR TARGETED INTERVENTION

Focusing the model on Patna yielded even more precise predictions, underscoring the benefits of localized data and clearer seasonal patterns:

10.07

RMSE Score

A significantly lower RMSE indicates higher accuracy in predicting dengue cases for Patna.

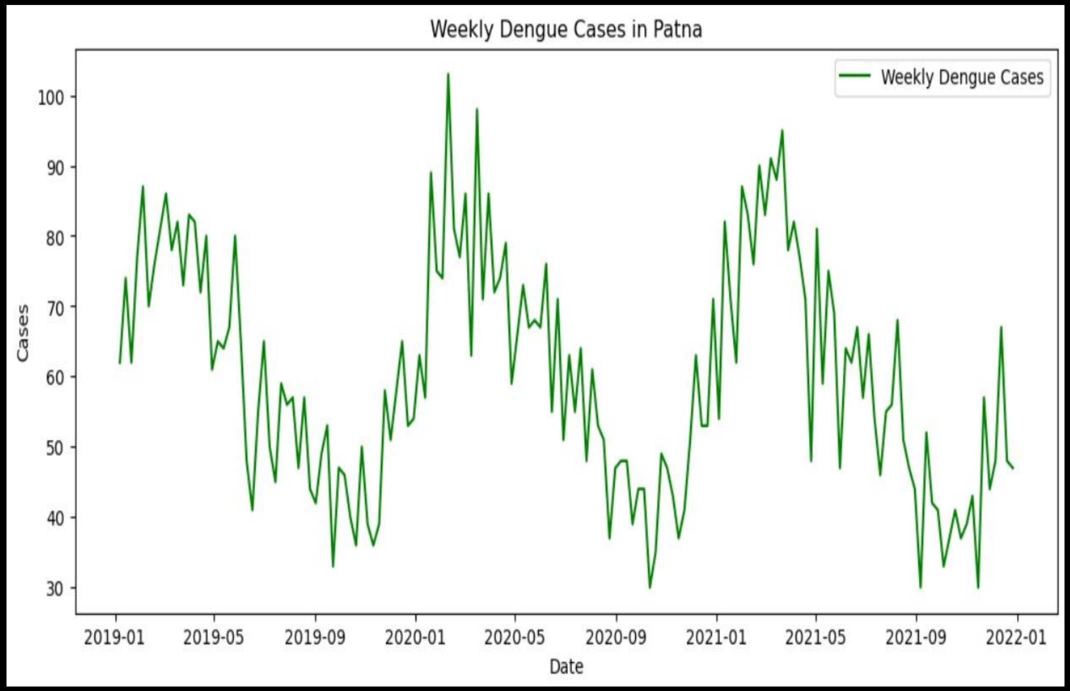
7.96

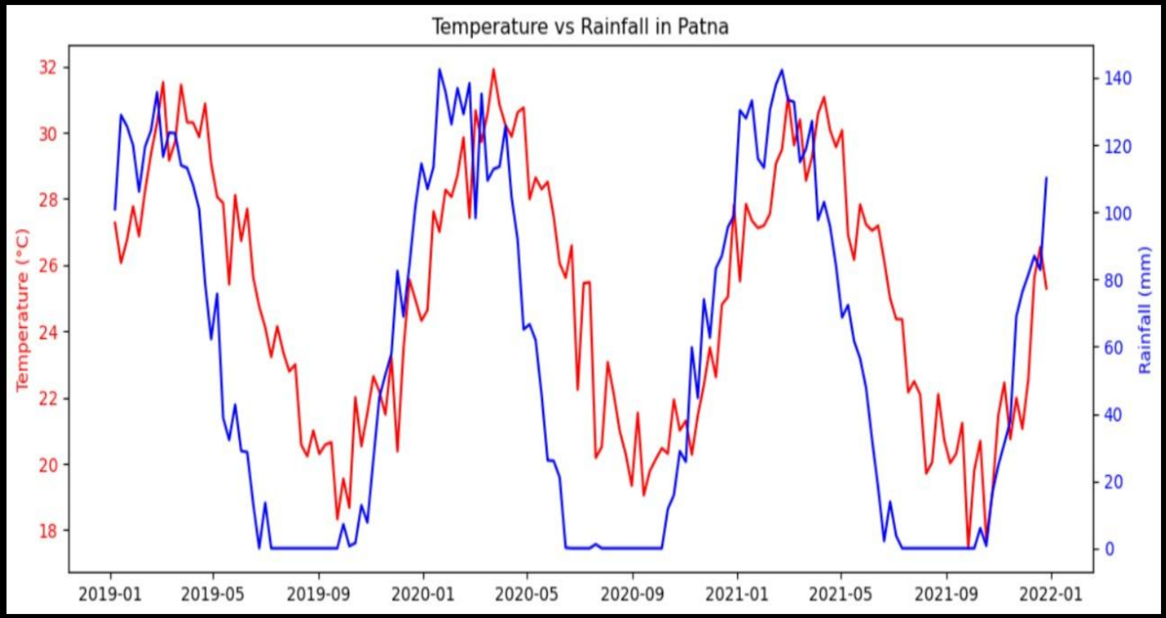
MAE Score

The reduced MAE reflects the model's closer alignment with actual case counts in the region.

The clearer seasonal cycles observed in Patna contributed directly to this improved accuracy, highlighting the potential for highly effective, targeted public health interventions.

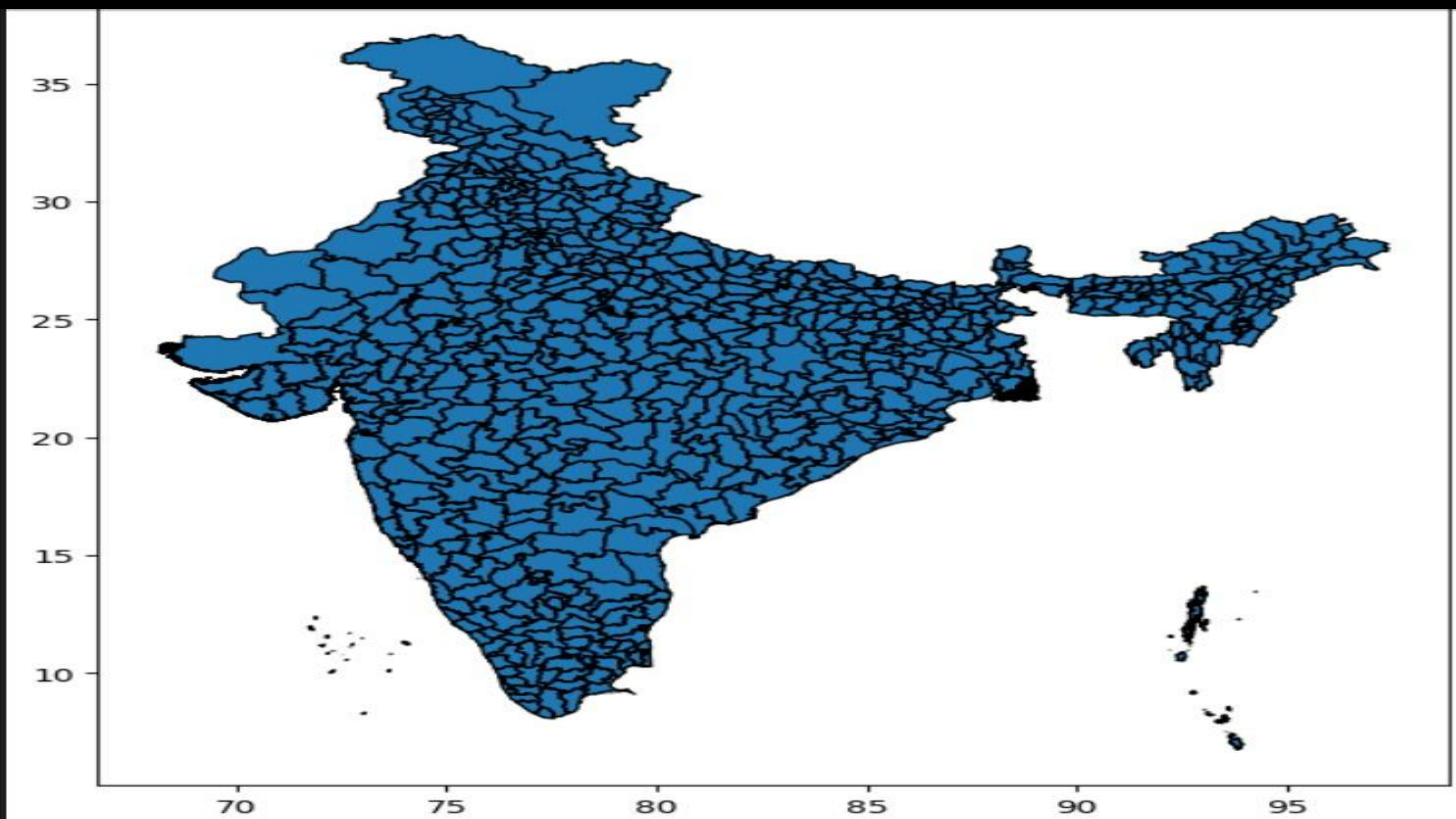
This focused approach allows for more efficient resource allocation and proactive measures in high-risk urban areas.





PRECISION IN PREDICTION: ADVANCING DENGUE FORECASTING

This presentation outlines a robust framework for predicting dengue outbreaks, empowering public health officials and data scientists with timely, actionable insights for proactive disease management.



ACTIONABLE INSIGHTS: FORECASTING DENGUE TRENDS

Our advanced forecasting model generates critical predictions, providing public health agencies with a clear vision of upcoming dengue activity.

12-Week Predictive Outlook

Detailed forecasts extend three months into the future, offering ample time for strategic planning and resource deployment.

Anticipating Post-Monsoon Surges

The model specifically highlights expected rises in dengue cases following the crucial monsoon period, a peak transmission season.

Empowering Proactive Strategies

These timely predictions enable health departments to shift from reactive responses to proactive, preventative public health interventions.

TRANSLATING FORECASTS INTO PUBLIC HEALTH ACTION

Accurate predictions serve as the foundation for targeted interventions, optimizing the use of vital resources and protecting communities.

Early Warning Systems

Implementing automated alerts for local health departments, ensuring rapid notification of impending risk.

Targeted Interventions

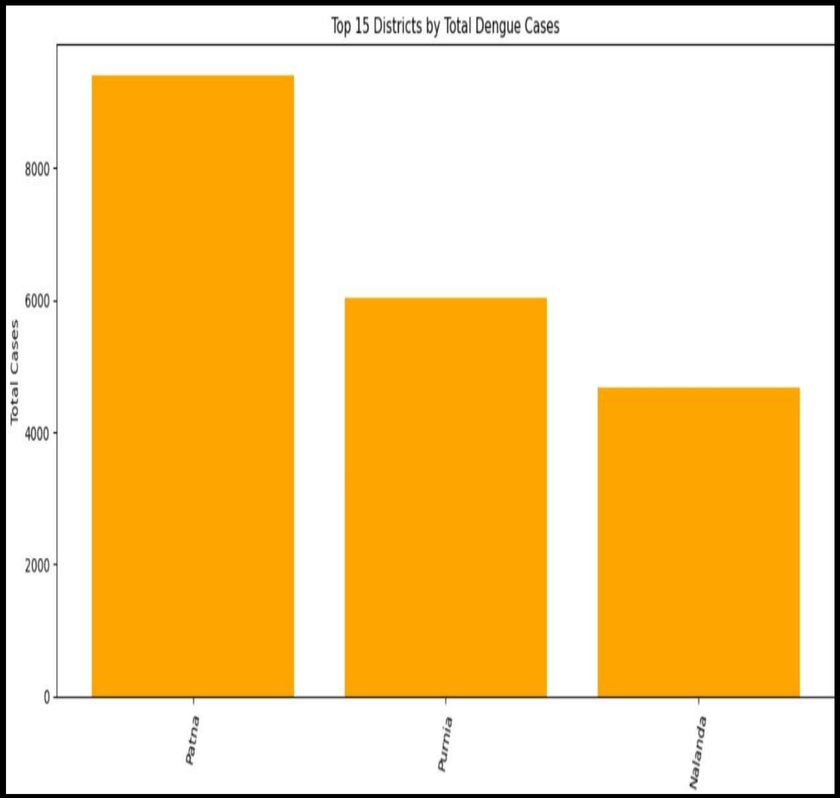
Directing fogging operations and intensive community awareness campaigns to high-risk zones identified by the forecast.

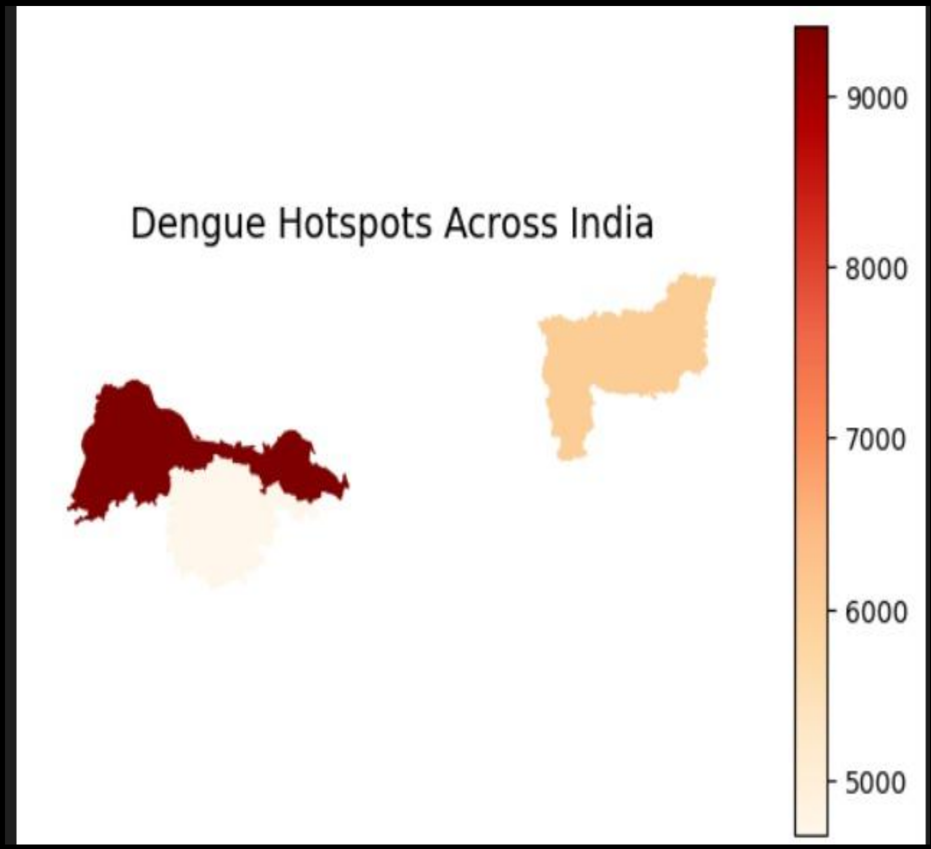
Optimized Resource Allocation

Pre-positioning essential medical supplies, hospital beds, and healthcare personnel where they are most needed.

Scalable Framework

The methodology is designed for scalability, allowing for replication and adaptation to other urban centers and regions facing similar challenges.

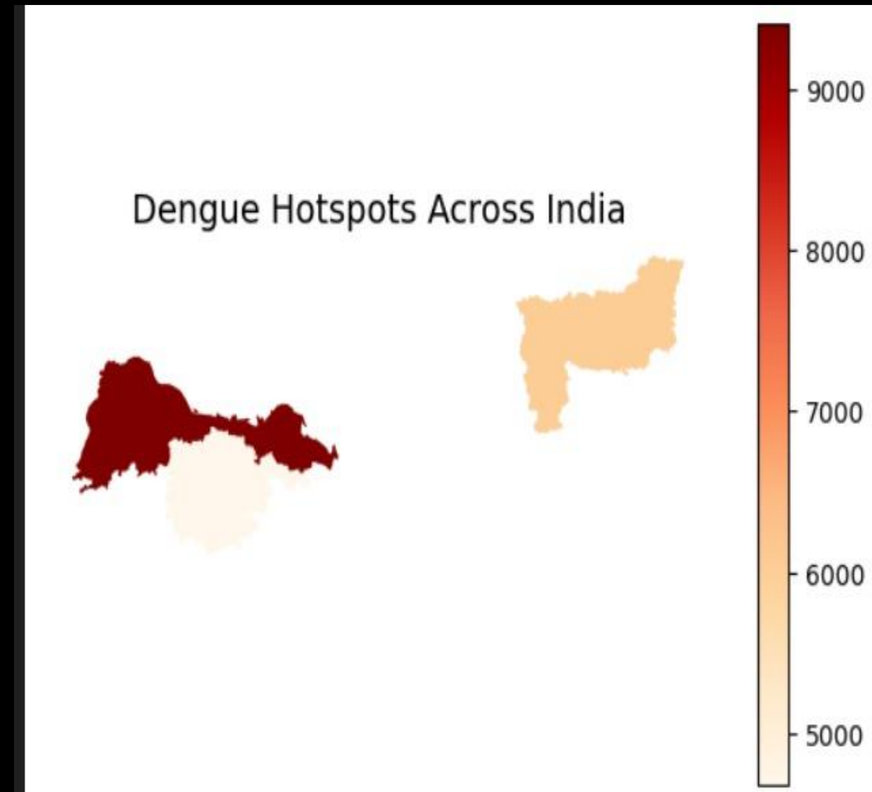




ROBUST FORECASTS FOR RESILIENT COMMUNITIES

Our project successfully demonstrates the power of data science in enhancing public health preparedness and response capabilities against dengue

`Climate Factors: Key Drivers` Our research confirms the strong influence of climatic variables on dengue transmission patterns, making them crucial inputs for accurate



`ARIMAX Model: Proven Accuracy`

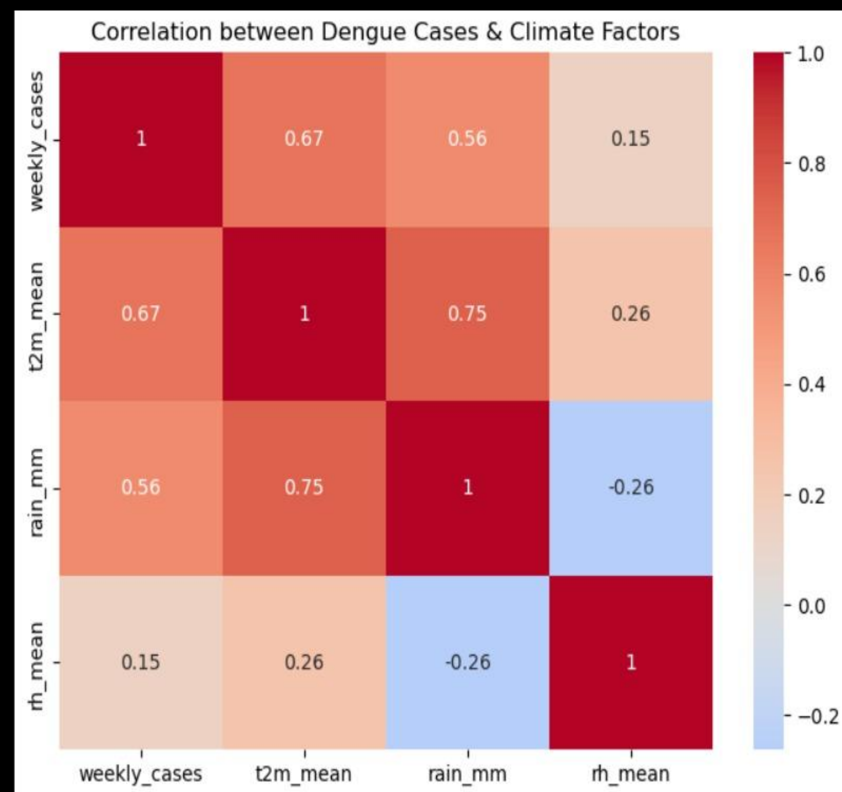
The AutoRegressive Integrated Moving Average with eXogenous inputs (ARIMAX) model achieved high reliability, consistently producing accurate predictions.

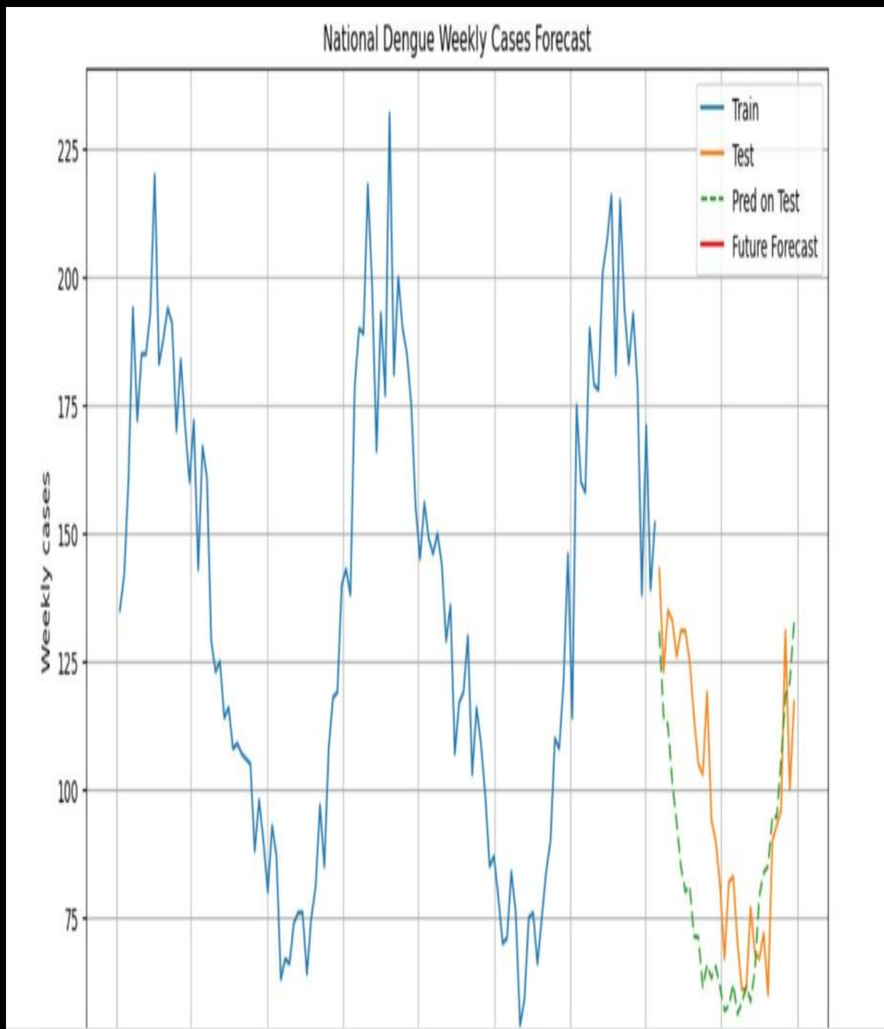
`Practical & Deployable Tool`

The developed forecasting tool is not just theoretically sound but is also highly practical and ready for real-world application by health agencies.

`Next Step: Automated Alerts`

Future development will focus on integrating these forecasts into automated alert dashboards for seamless, real-time dissemination of risk information.





Reflections: A Data Science Learning Curve

- Mastered end-to-end data pipeline development, from collection to deployment.
- Enhanced confidence in handling complex, real-world messy datasets.
- Deepened understanding of the critical public health relevance of data science.
- Prepared for future AI and health sector projects with practical experience.

Key References: Guiding Our Research

- World Health Organization (WHO) Dengue Guidelines
- ERA5 Climate Reanalysis Data (ECMWF)
- Official District Dengue Surveillance Reports
- Statsmodels and Pandas Software Documentation

OUR DEEPEST GRATITUDE

We extend our sincere appreciation to the **GUVI platform** for their invaluable support, providing the resources and environment necessary for this project. Our profound thanks also go to the distinguished faculty of **IIT Patna** for their expert guidance, mentorship, and unwavering encouragement throughout this research journey. Together, we are building healthier communities.