DengueWatch: A System for Real-Time Dengue Monitoring and Forecasting in Iloilo City

A Special Problem Proposal
Presented to
the Faculty of the Division of Physical Sciences and Mathematics
College of Arts and Sciences
University of the Philippines Visayas
Miag-ao, Iloilo

In Partial Fulfillment
of the Requirements for the Degree of
Bachelor of Science in Computer Science by

AMODIA, Kurt Matthew A.
BULAONG, Glen Andrew C.
ELIPAN, Carl Benedict L.

Francis D. DIMZON
Adviser

November 6, 2024

Abstract

In response to a marked rise in dengue cases, Iloilo City and Province are enhancing control measures. As of August 10, 2023, the Iloilo Provincial Health Office reported 4,585 cases and 10 fatalities, reflecting a 319% increase from last year's 1,095 cases and one death. This research includes the development of a centralized system for monitoring and forecasting dengue trends in the Iloilo region. This study explores the application of artificial intelligence (AI) for dengue prediction, using a deep learning approach with Long Short-Term Memory (LSTM) networks. The LSTM model is compared with traditional statistical methods, including non-seasonal and seasonal Autoregressive Integrated Moving Average (ARIMA) models and the Kalman Filter for state estimation algorithm in noisy data conditions. Forecasting was based on climate variables such as temperature, rainfall, relative humidity, and previous monthly case counts, with performance evaluated using Root Mean Square Error (RMSE). This research, aimed at supporting public health agencies like the Department of Health (DOH), advocates for AI-driven solutions that improve outbreak response beyond traditional reporting systems.

Keywords: ARIMA, artificial intelligence, dengue prediction, LSTM, Kalman Filter, deep learning, climate variables, public health, outbreak mitigation

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

From 2020 to 2022, dengue cases declined due to reduced surveillance during the COVID-19 pandemic, but cases surged in 2023 as restrictions were lifted. This year saw an increase in dengue outbreaks worldwide, with over five million cases and more than 5,000 deaths reported in over 80 countries. (?, ?) Dengue is endemic in the Philippines, leading to longer and more widespread seasonal outbreaks. Globally, dengue infections have increased significantly, posing a major public health challenge. The World Health Organization (?, ?) reported a ten-fold rise in cases between 2000 and 2019, with a peak in 2019 when the disease spread across 129 countries.

Iloilo City and Province are intensifying efforts to curb the rising dengue cases. As of August 10, 2023, the Iloilo Provincial Health Office recorded 4,585 cases and 10 deaths, a 319% increase from last year's 1,095 cases and one death. Governor Arthur Defensor Jr. confirmed that the province has reached the dengue outbreak threshold based on Department of Health (DOH) criteria, and a formal declaration is pending. Local government units (LGUs) have been informed, and the province's disaster management office is on blue alert, indicating disaster mode. (?, ?)

In Iloilo City, 649 dengue cases were recorded during the same period, with two deaths. Cases cluster in 40 out of 180 barangays, meaning multiple cases are being reported in these areas over several weeks. The city's health officer, Dr. Roland Jay Fortuna, reported high utilization of non-COVID-19 hospital beds, reaching over 76%, prompting concerns about hospital capacity.

This study explores the monitoring and forecasting of dengue outbreaks by analyzing key factors such as temperature, relative humidity, and historical dengue cases, using different models. The findings aim to provide an advanced, AI-driven alternative for dengue prevention and control, targeting agencies like the Department of Health (DOH). By aligning with the national AI Roadmap, particularly in Iloilo City, this research aspires to improve outbreak responses through cutting-edge technology rather than traditional reporting methods.

1.2 Problem Statement

The problem being addressed here is that dengue cases remain a critical public health issue worldwide, with rising cases attributed to the easing of COVID-19 restrictions and increased global mobility. From 2020 to 2022, dengue cases saw a temporary decline due to reduced surveillance efforts amidst the pandemic. However, 2023 witnessed a significant resurgence, with over five million cases and more than 5,000 deaths reported across 80 countries, indicating the continued vulnerability of dengue-endemic regions like the Philippines. In Iloilo City and Province, dengue cases surged dramatically by 319% as of August 2023, with local health systems struggling to manage the influx. High hospitalization rates due to dengue, with over 76% of non-COVID-19 hospital beds occupied, have raised concerns about healthcare capacity and the need for enhanced predictive measures.

1.3 Research Objectives

1.3.1 General Objective

This study aims to develop an AI-based dengue forecasting and monitoring system for Iloilo City and Province. The system will use Long Short-Term Memory (LSTM) to predict dengue case trends based on climate data and historical dengue cases to help public health officials in possible dengue case outbreaks.

1.3.2 Specific Objectives

Specifically, this study aims to develop a system that can:

1. Gather dengue data from the Iloilo Provincial Health Office and climate data from online sources. Combine these data into a unified dataset to facilitate comprehensive dengue case forecasting.

2. Develop and evaluate deep learning models, including LSTM, ARIMA, Seasonal ARIMA, and Kalman Filter, for predicting dengue cases. Compare the performance of these models to determine the most accurate forecasting approach.

3. Integrate the predictive model into a web-based data analytics dashboard. This dashboard will include features such as data visualizations and data entry, offering public health stakeholders an interactive tool for analyzing dengue trends and making informed decisions.

1.4 Scope and Limitations of the Research

This study aimed to develop an AI-based dengue forecasting and monitoring system specifically designed for Iloilo City. The system focuses on two major features: dengue case prediction and risk area identification. The dengue case prediction feature utilizes climate variables—such as temperature, rainfall, and relative humidity—along with historical dengue case data to forecast monthly dengue cases. The results will be displayed in a user-friendly interface, providing public health officials with actionable insights to enhance outbreak management and resource allocation. However, this study has several limitations. The accuracy of the dengue case predictions heavily relies on the quality and completeness of the input data. Inconsistent or incomplete historical data may lead to reduced prediction accuracy. Additionally, the model's performance may fluctuate based on variations in climate patterns, which are not always predictable. The model utilizes advanced machine learning techniques, but it cannot account for all factors influencing dengue transmissions, such as socio-economic conditions or public health interventions, which may further impact case dynamics. Finally, the dataset used for training the predictive models has not undergone peer review but has been validated by local public health experts to ensure its relevance and accuracy for the study's context. As a result, the findings should be interpreted with caution, and ongoing validation and adjustments may be necessary to enhance the model's robustness and applicability in real-world settings.

1.5 Significance of the Research

This study's development of an AI-based dengue forecasting and monitoring system has wide-reaching significance for various stakeholders in Iloilo City:

- Public Health Agencies: Organizations like the Department of Health (DOH) and local health units in Iloilo City and Province stand to benefit greatly from the system. With dengue predictions, we can help these agencies optimize their response strategies and implement targeted prevention measures in high-risk areas before cases escalate.

- Local Government Units (LGUs): LGUs can use the system to support their disaster management and health initiatives by proactively addressing dengue outbreaks. The predictive insights allow for more efficient planning and resource deployment in barangays and communities most vulnerable to outbreaks, improving overall public health outcomes.

- Healthcare Facilities: Hospitals and clinics, which currently face high bed occupancy rates during dengue season will benefit from early outbreak forecasts that can help in managing patient inflow and ensuring adequate hospital capacity.

- Researchers and Policymakers: This AI-driven approach contributes valuable insights for researchers studying infectious disease patterns and policymakers focused on strengthening the national AI Roadmap. The system's data can support broader initiatives for sustainable health infrastructure and inform policy decisions on resource allocation for dengue control.

- Community Members: By reducing the frequency and severity of outbreaks, this study ultimately benefits the community at large. This allows for timely awareness campaigns and community engagement initiatives, empowering residents with knowledge and preventative measures to protect themselves and reduce the spread of dengue.

Chapter 2

Review of Related Literature

2.1 Existing System: RabDash DC

RabDash, developed by the University of the Philippines Mindanao, is a web-based dashboard for rabies data analytics. It combines predictive modeling with genomic data, enabling local health authorities to optimize interventions and allocate resources more effectively. RabDash's modules include trend visualization, geographic hotspot mapping, and predictive forecasting, utilizing Long Short-Term Memory (LSTM) models for time-series forecasting (?, ?).

For DengueDash, RabDash serves as a strong inspiration, particularly in its monitoring, historical trend visualization, and forecasting capabilities. These features align well with the needs of dengue control efforts, providing real-time insights into outbreak trends and enabling more effective, data-driven decision-making. RabDash's architecture is relevant to the DengueDash, as dengue outbreaks similarly require time-series forecasting models. By using LSTM, RabDash effectively models trends in outbreak data, which provides a framework for adapting LSTM to dengue forecasting. Research indicates that LSTM models outperform traditional methods, such as ARIMA and MLP, in handling the complexities of time-dependent epidemiological data (?, ?).

2.2 Deep Learning

The study of Kim Dianne Ligue and Kristine Joy Ligue highlights how data-driven models can help predict dengue outbreaks. The authors compared traditional statistical methods, such as non-seasonal and seasonal autoregressive integrated moving average (ARIMA), and traditional feed-forward network approach using a multilayer perceptron (MLP) model with a deep learning approach using the long short-term memory (LSTM) architecture in their prediction model. They find that the LSTM model performs better in terms of accuracy. The LSTM model achieved a much lower root mean square error (RMSE) compared to both MLP and ARIMA models, proving its ability to capture complex patterns in time-series data (?, ?). This superior performance is attributed to LSTM's capacity to capture complex, time-dependent relationships within the data, such as those between temperature, rainfall, humidity, and mosquito populations, all of which contribute to dengue incidence (?, ?).

2.3 Kalman Filter

The Kalman Filter is another powerful tool for time-series forecasting that can be integrated into our analysis. It provides a recursive solution to estimating the state of a linear dynamic system from a series of noisy measurements. Its application in epidemiological modeling can enhance prediction accuracy by accounting for uncertainties in the data(?, ?). Studies have shown that Kalman filters are effective in predicting infectious disease outbreaks by refining estimates based on observed data. A study published in Frontiers in Physics utilized the Kalman filter to predict COVID-19 deaths in Ceará, Brazil. They found that the Kalman filter effectively tracked the progression of deaths and cases, providing critical insights for public health decision-making (?, ?). Another research article in PLOS ONE focused on tracking the effective reproduction number (R_t) of COVID-19 using a Kalman filter. This method estimated the growth rate of new infections from noisy data, demonstrating that the Kalman filter could maintain accurate estimates even when case reporting was inconsistent(?, ?).

Our study will compare ARIMA, seasonal ARIMA, Kalman Filter, and LSTM models using our own collected dengue case data along with weather data to identify the most effective model for real-time forecasting.

2.4 Weather Data

The relationship between weather patterns and mosquito-borne diseases is inherently nonlinear, meaning that fluctuations in disease cases do not respond proportionally to changes in climate variables(?, ?) Weather data, such as minimum

Study	Method	Weather Variables	Forecasting Model	Region	Key Findings	Gaps Identified
RabDash DC	Centralized System with Visualizations and Forecasting	N/A	LSTM	Davao City, Mindanao, Philippines	Real-time tracking and prediction for rabies outbreaks	Lack of dengue-specific application
Ligue et al. (2022)	Comparative Study for forecasting dengue cases	Temperature, Rainfall, Humidity	LSTM, ARIMA, MLP	Mindanao, Philippines	LSTM outperforms traditional models for dengue forecasting	Limited focus on weather effects and lagged impacts
Arroyo et al. (2021)	Kalman Filter	N/A	Kalman Filter	COVID-19 in Brazil	Kalman filter accurately tracked infection rates despite data inconsistencies	Application limited to COVID-19; requires adaptation for dengue
Ahmadini et al. (2021)	Kalman Filter	N/A	Kalman Filter	COVID-19 worldwide	Accurately tracked R_t values in fluctuating infection data	Lack of application for vector-borne diseases like dengue
Hii et al. (2012)	Multivariate model	Mean Temperature, Cumulative Rainfall	Poisson regression, ROC curve	Singapore	16-week forecast potential with high sensitivity and specificity	Regional limitations for forecast applicability

temperature and accumulated rainfall, are strongly linked to dengue case fluctuations, with effects observed after several weeks due to mosquito breeding and virus incubation cycles. Integrating these lagged weather effects into predictive models can improve early warning systems for dengue control(?, ?). A study also suggests that weather-based forecasting models using variables like mean temperature and cumulative rainfall can provide early warnings of dengue outbreaks with high sensitivity and specificity, enabling predictions up to 16 weeks in advance(?, ?).

We will utilize weather data, including variables such as temperature, rainfall, and humidity, as inputs for our dengue forecasting model. Given the strong, nonlinear relationship between climate patterns and dengue incidence, these weather variables, along with their lagged effects, are essential for enhancing prediction accuracy and providing timely early warnings for dengue outbreaks.

2.5 Chapter Summary

This chapter reviewed key literature relevant to our study, focusing on existing systems, predictive modeling techniques and the role of weather data in forecasting dengue outbreaks. We examined systems like RabDash DC, which integrates predictive modeling with real-time data to inform public health decisions, providing a foundational structure for our Dengue Watch System. Additionally, deep learning approaches, particularly Long Short-Term Memory (LSTM) networks, were highlighted for their effectiveness in time-series forecasting, while alternative methods such as ARIMA and Kalman Filters were considered for their ability to model complex temporal patterns and handle noisy data.

The literature further underscores the significance of weather variables—such as temperature and rainfall—in forecasting dengue cases. Studies demonstrate that these variables contribute to accurate outbreak prediction models. Leveraging these insights, our study will incorporate both weather data and historical dengue case counts to build a reliable forecasting model.

Chapter 3

Research Methodology

This chapter lists and discusses the specific steps and activities that will be performed to accomplish the project. The discussion covers the activities from pre-proposal to Final SP Writing.

3.1 Research Activities

A total of 452 rows of dengue cases and the associated average weekly weather data from January 2016 to September 2024 were used in training and forecasting the proposed models. These meteorological data points include rainfall, temperature, and humidity. The cases were gathered from both the Humanitarian Data Exchange and the Western Visayas Center for Health Development (WVCHD) via Freedom of Information (FOI). On the other hand, the weather variables were scraped from the Weather Underground. However, to produce more accurate and reliable results, the researchers are planning to request the same dataset from the Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAG-ASA). Moreover, the forms and templates that inspire the system came from WVCHD. The said document was meticulously analyzed to create a well-structured database that adheres to the requirements needed to produce a quality application.

3.1.1 Development Framework

The Agile Model is the birthchild of both iterative and incremental approaches in Software Engineering. It aims to be flexible and effective at the same time by being adaptable to change. It's also important to note that small teams looking to construct and develop projects quickly can benefit from this kind of methodology. As the Agile Method focuses on continuous testing, quality assurance is a guarantee since bugs and errors are quickly identified and patched.

3.1.2 Development Tools

Software

GitHub
GitHub is a cloud-based platform that uses Git, an open-source version control system, to track file changes (GitHub Docs, n.d.). According to Sewers (2023), it is the most popular collaboration tool with over 100 million users and 420 million repositories. It stores source code, manages system versions, and serves as a repository for LaTeX files used in this research.

Visual Studio Code
Visual Studio Code (VS Code) is a lightweight, cross-platform source code editor developed by Microsoft (2021). Although not a full IDE, it supports features such as syntax highlighting, code completion, and debugging. Due to its support for various programming languages, it is chosen as the primary code editor.

Django
Django is a free, open-source Python-based web framework that facilitates secure web application development by abstracting complexities (MDN Web Docs, 2024). It adheres to the Model-View-Template (MVT) architecture, providing a clean codebase by separating data models, business logic, and presentation layers.

Next.js
Statista (2024) claims React is the most popular front-end framework, but it has limitations like routing and performance optimizations. Next.js, built on React, addresses these limitations, enhancing the development experience through features such as hot module replacement, TypeScript support, and a rich plugin ecosystem.

Postman
Postman is a freemium API platform that provides a user-friendly interface to

create and manage API requests (Postman API Platform, n.d.).

3.1.3 Hardware

The application development is carried out on laptops with at least an 11th-gen Intel i5 CPU and 16GB RAM.

Packages

Django Rest Framework
Django Rest Framework (DRF) is a third-party package for Django that simplifies the development of robust Web APIs by providing features such as Serialization, Authentication, and a Browsable API (Christie, n.d.).

Leaflet
The application includes a Choropleth Map to display case counts, utilizing Leaflet—an open-source JavaScript package known for its performance and usability.

Chart.js
Chart.js provides data storytelling through various chart types. It is open-source and highly customizable, leveraging HTML5's canvas for compatibility across modern browsers.

Tailwind CSS
Tailwind CSS offers utility classes for creating custom designs directly in HTML (Tailwind CSS Docs, n.d.), making it an efficient choice for this project.

Shadcn
Shadcn offers open-source UI boilerplate components compatible with modern frameworks like Next.js, enhancing flexibility and customizability.

3.2 Model and Algorithm

Data Preprocessing and Feature Selection:

- Clean and preprocess the dataset, handling missing values and outliers.
- Select important features such as temperature, rainfall, and humidity.

Model Selection and Justification:

- Train an LSTM model due to its efficiency in handling long-term data relationships.
- Compare performance with ARIMA, SARIMA, and Kalman Filter.

Seasonal ARIMA (SARIMA):

- Set seasonal parameters based on observed seasonality in cases.
- Apply SARIMA to capture short-term and seasonal trends.

Kalman Filter:

- Use Kalman Filter for real-time updates and handle missing values dynamically.

LSTM (Long Short-Term Memory) Neural Network:

- Build a sequential LSTM model with tuned hyperparameters.

Model Evaluation:

- Evaluate performance with Mean Absolute Error (MAE), RMSE, and R-squared metrics.

3.3 Calendar of Activities

A Gantt chart showing the schedule of the activities is included below. Each bullet represents approximately one week of activity.

Table 3.1: Timetable of Activities

Activities (2024)	Aug	Sept	Oct	Nov	Dec
Project Initiation and Team Formation	●●				
Literature Review and Data Gathering	●●	●●●●			
Data Cleaning and Feature Selection		●●		●	
Creating System Dashboard		●	●●	●	
Analysis and Interpretation of Results			●		
Documentation		●	●●●●	●	