

¹ DENGUEWATCH: A SYSTEM FOR REAL-TIME
² DENGUE MONITORING AND FORECASTING IN ILOILO
³ PROVINCE

⁴ A Special Problem Proposal
⁵ Presented to
⁶ the Faculty of the Division of Physical Sciences and Mathematics
⁷ College of Arts and Sciences
⁸ University of the Philippines Visayas
⁹ Miag-ao, Iloilo

¹⁰ In Partial Fulfillment
¹¹ of the Requirements for the Degree of
¹² Bachelor of Science in Computer Science by

¹³ AMODIA, Kurt Matthew A.
¹⁴ BULAONG, Glen Andrew C.
¹⁵ ELIPAN, Carl Benedict L.

¹⁶ Francis D. DIMZON, Ph.D.
¹⁷ Adviser

¹⁸ May 31, 2025

Abstract

Dengue fever remains a significant public health concern in the Philippines, with cases rising dramatically in recent years. Nationwide outbreaks have placed immense strain on healthcare systems, underscoring the need for innovative approaches to surveillance and response. In Iloilo City, this national trend was reflected in a significant surge, with the Iloilo Provincial Health Office reporting 4,585 cases and 10 fatalities as of August 10, 2023—a 319% increase from the previous year's 1,095 cases and one death. This rise overwhelmed local healthcare systems, with over 76% of non-COVID-19 hospital beds occupied by dengue patients. The absence of a reliable system to monitor and forecast dengue outbreaks contributed to delayed interventions, exacerbating public health risks and the burden on medical resources. To address this gap, this study developed a centralized system for monitoring and modernizing data management of dengue cases in public health institutions, making it more efficient and acceptable. Using data gathered from the Iloilo Provincial Health Office and online sources, several deep learning models were trained to predict dengue cases, utilizing weather variables and historical case data as inputs. These models included Long Short-Term Memory (LSTM), ARIMA, Seasonal ARIMA, Kalman Filter (KF), and a hybrid KF-LSTM model. The models underwent time series cross-validation strategies to mimic real-world conditions as closely as possible and were evaluated using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The LSTM model demonstrated the best performance with the lowest RMSE of 16.90, followed by the hybrid KF-LSTM model at 25.56. The LSTM model was then integrated into the system to provide forecasting features that could support health institutions by offering actionable insights for proactive intervention strategies.

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

⁴⁶ **Contents**

⁴⁷ 1 Introduction	¹
⁴⁸ 1.1 Overview	¹
⁴⁹ 1.2 Problem Statement	²
⁵⁰ 1.3 Research Objectives	²
⁵¹ 1.3.1 General Objective	²
⁵² 1.3.2 Specific Objectives	²
⁵³ 1.4 Scope and Limitations of the Research	³
⁵⁴ 1.5 Significance of the Research	⁴
⁵⁵ 2 Review of Related Literature	⁵
⁵⁶ 2.1 Dengue	⁵
⁵⁷ 2.2 Outbreak Definition	⁵
⁵⁸ 2.3 Existing System: RabDash DC	⁶
⁵⁹ 2.4 Deep Learning	⁷
⁶⁰ 2.5 Kalman Filter	⁷
⁶¹ 2.6 Weather Data	⁸
⁶² 2.7 Chapter Summary	⁸

63	3 Research Methodology	9
64	3.1 Research Activities	10
65	3.1.1 Gather Dengue Data and Climate Data to Create a Com-	
66	plete Dataset for Forecasting	10
67	3.1.2 Develop and Evaluate Deep Learning Models for Dengue	
68	Case Forecasting	12
69	3.1.3 Integrate the Predictive Model into a Web-Based Data An-	
70	alytics Dashboard	16
71	3.1.4 System Development Framework	16
72	3.2 Development Tools	18
73	3.2.1 Software	18
74	3.2.2 Hardware	19
75	3.2.3 Packages	20
76	3.3 Application Requirements	21
77	3.3.1 Backend Requirements	21
78	3.3.2 User Interface Requirements	23
79	3.3.3 Security and Validation Requirements	25
80	4 Results and Discussion/System Prototype	26
81	4.1 Data Gathering	26
82	4.2 Exploratory Data Analysis	26
83	4.3 Outbreak Detection	32
84	4.4 Model Training Results	33
85	4.4.1 LSTM Model	33
86	4.4.2 ARIMA Model	35

87	4.4.3 Seasonal ARIMA (SARIMA) Model	36
88	4.4.4 Kalman Filter Model	38
89	4.5 Model Simulation	40
90	4.6 System Prototype	40
91	4.6.1 Home Page	40
92	4.6.2 User Registration, Login, and Authentication	41
93	4.6.3 Encoder Interface	43
94	4.6.4 Admin Interface	51
95	4.7 User Testing	56
96	5 Conclusion	58
97	References	60
98	A Data Collection Snippets	63
99	B Resource Persons	67

¹⁰⁰ List of Figures

¹⁰¹	3.1 Workflow for forecasting the number of weekly dengue cases	9
¹⁰²	3.2 Testing Process for DengueWatch	17
¹⁰³	3.3 Entity-Relationship Database Schema Hybrid Diagram for DengueDash Database Structure	22
¹⁰⁵	3.4 Use Case Diagram for Admins	23
¹⁰⁶	3.5 Use Case Diagram for Encoder	24
¹⁰⁷	4.1 Snippet of the Combined Dataset	27
¹⁰⁸	4.2 Data Contents	27
¹⁰⁹	4.3 Dataset Statistics	28
¹¹⁰	4.4 Trend of Dengue Cases	28
¹¹¹	4.5 Ranking of Correlations	29
¹¹²	4.6 Pre-Transform Feature Distributions	30
¹¹³	4.7 Scatterplots	30
¹¹⁴	4.8 Post-Transform Feature Distributions	31
¹¹⁵	4.9 Transformed Distributions: Scatterplots	31
¹¹⁶	4.10 Ranking of Correlations with New Distributions	32
¹¹⁷	4.11 Training Folds - Window Size 5	34

118	4.12 Testing Folds - Window Size 5	35
119	4.13 ARIMA Prediction Results for Test Set	36
120	4.14 Seasonal ARIMA Prediction Results for Test Set	37
121	4.15 Kalman Filter Prediction Results for Test Set	39
122	4.16 Predicted vs Actual Dengue Cases 2025	40
123	4.17 Home Page	41
124	4.18 Sign Up Page	42
125	4.19 Login Page	42
126	4.20 First Part of Case Report Form	43
127	4.21 Second Part of Case Report Form	44
128	4.22 Bulk Upload of Cases using CSV	45
129	4.23 Dengue Reports	46
130	4.24 Detailed Case Report	47
131	4.25 Update Report Dialog	48
132	4.26 Delete Report Alert Dialog	49
133	4.27 Forecasting Page	50
134	4.28 Retraining Configurations	51
135	4.29 Start Retraining	52
136	4.30 Retraining Results	52
137	4.31 List of Verified Accounts	53
138	4.32 List of Pending Accounts	54
139	4.33 Account Details	54
140	4.34 Disease Reporting Unit Registration	55

¹⁴¹	4.35 List of Disease Reporting Units	55
¹⁴²	4.36 Disease Reporting Unit details	56
¹⁴³	A.1 Snippet of Consolidated Data	63
¹⁴⁴	A.2 Snippet of Weather Data Collection	64
¹⁴⁵	A.3 Letter of Approval for User Testing in UPV HSU	65
¹⁴⁶	A.4 System Usability Score Questionnaire	66

¹⁴⁷ List of Tables

¹⁴⁸	4.1 Comparison of different models for dengue prediction	33
¹⁴⁹	4.2 Comparison of Window Sizes	33
¹⁵⁰	4.3 Time-Series Cross Validation Results: Comparison of Window Sizes	34
¹⁵¹	4.4 Comparison of SARIMA performance for each fold	38
¹⁵²	4.5 Comparison of KF-LSTM performance for each fold	39
¹⁵³	4.6 Computed System Usability Scores per Participant	56

¹⁵⁴ **Chapter 1**

¹⁵⁵ **Introduction**

¹⁵⁶ **1.1 Overview**

¹⁵⁷ From 2020 to 2022, dengue cases declined due to reduced surveillance during
¹⁵⁸ the COVID-19 pandemic (WHO, 2023), but cases surged in 2023 as restrictions
¹⁵⁹ were lifted. In 2025, there was an increase in dengue outbreaks worldwide, with
¹⁶⁰ over five million cases and more than 5,000 deaths reported in over 80 countries
¹⁶¹ (Bosano, 2023). Dengue is endemic in the Philippines, leading to longer and more
¹⁶² widespread seasonal outbreaks. Globally, dengue infections have increased signif-
¹⁶³ icantly, posing a major public health challenge. The World Health Organization
¹⁶⁴ reported a ten-fold rise in cases between 2000 and 2019, with a peak in 2019 when
¹⁶⁵ the disease spread across 129 countries (WHO, 2024).

¹⁶⁶ Iloilo City and Province are intensifying efforts to curb the rising dengue cases
¹⁶⁷ (Lena, 2024). As of August 10, 2023, the Iloilo Provincial Health Office recorded
¹⁶⁸ 4,585 cases and 10 deaths, a 319% increase from last year's 1,095 cases and one
¹⁶⁹ death. Governor Arthur Defensor Jr. confirmed that the province has reached the
¹⁷⁰ dengue outbreak threshold based on Department of Health (DOH). Local govern-
¹⁷¹ ment units (LGUs) have been informed, and the province's disaster management
¹⁷² office is on blue alert, indicating disaster mode (Perla, 2024).

¹⁷³ In Iloilo City, 649 dengue cases were recorded this year 2024, with two deaths.
¹⁷⁴ Cases cluster in 40 out of 180 barangays, meaning multiple cases are being reported
¹⁷⁵ in these areas over several weeks. The city's health officer, Dr. Roland Jay
¹⁷⁶ Fortuna, reported high utilization of non-COVID-19 hospital beds, reaching over
¹⁷⁷ 76%, prompting concerns about hospital capacity.

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

185 **1.2 Problem Statement**

186 Dengue remains a critical public health challenge worldwide, with cases increasing
187 due to the easing of COVID-19 restrictions and heightened global mobility. While
188 a temporary decline in cases was observed during the pandemic (2020–2022) due
189 to reduced surveillance efforts, 2023 marked a resurgence, with over five million
190 cases and more than 5,000 deaths reported across 80 countries. In dengue-endemic
191 regions like the Philippines, the threat is particularly severe. In Iloilo City and
192 Province, dengue cases rose by 319% as of August 2023, overwhelming local health-
193 care systems. This surge strained resources, with over 76% of non-COVID-19 hos-
194 pital beds occupied by dengue patients, highlighting the urgent need for effective
195 predictive tools. The lack of a reliable system to monitor and forecast dengue
196 outbreaks contributes to delayed interventions, exacerbating public health risks
197 and healthcare burdens in the region.

198 **1.3 Research Objectives**

199 **1.3.1 General Objective**

200 This study aims to develop a centralized monitoring and analytics system for
201 dengue cases in Iloilo City and Province with data management and forecasting
202 capabilities. The researchers will train and compare multiple deep learning models
203 to predict dengue case trends based on climate data and historical dengue cases
204 to help public health officials in possible dengue case outbreaks.

205 **1.3.2 Specific Objectives**

206 Specifically, this study aims to:

- 207 1. Gather dengue data from the Iloilo Provincial Health Office and climate data
208 (including temperature, rainfall, wind, and humidity) from online sources,
209 and combine and aggregate these into a unified dataset to facilitate compre-
210 hensive dengue case forecasting;
- 211 2. Train and evaluate deep learning models for predicting dengue cases using
212 metrics such as Mean Absolute Error (MAE), Root Mean Squared Error
213 (RMSE), and Mean Squared Error (MSE), and determine the most accurate
214 forecasting approach; and
- 215 3. Develop a web-based analytics dashboard that integrates the predictive
216 model, provides a data management system for dengue cases in Iloilo City
217 and the Province, and assess its usability and effectiveness through struc-
218 tured feedback from health professionals and policymakers.

219 1.4 Scope and Limitations of the Research

220 This study aims to gather dengue data from the Iloilo Provincial Health Office
221 and climate data from online sources such as PAGASA or weatherandclimate.com.
222 These data will be preprocessed, cleaned, and combined into a unified dataset to
223 facilitate comprehensive dengue case forecasting. However, the study is limited by
224 the availability and completeness of historical data. Inconsistent or missing data
225 points may introduce biases and reduce the quality of predictions. Furthermore,
226 the granularity of the data will be in a weekly format.

227 To evaluate deep learning models for predicting dengue cases, the study will
228 train and compare the performance of various models, using metrics like Mean
229 Absolute Error (MAE) and Root Mean Square Error (RMSE). While these models
230 aim to provide accurate forecasts, their performance is heavily influenced by the
231 quality and size of the dataset. Limited or low-quality data may lead to suboptimal
232 predictions. Additionally, the models cannot fully account for external factors
233 such as public health interventions or socio-economic conditions which may impact
234 dengue transmission dynamics.

235 The study also involves developing a web-based analytics dashboard that in-
236 tegrates predictive models and provides a data management system for dengue
237 cases in Iloilo City and the Province. This dashboard will offer public health
238 officials an interactive interface to visualize dengue trends, input new data, and
239 identify risk areas. However, its usability depends on feedback from stakeholders,
240 which may vary based on their familiarity with analytics tools. Moreover, exter-
241 nal factors such as limited internet connectivity or device availability in remote

²⁴² areas may affect the system's adoption and effectiveness. While the dashboard
²⁴³ provides valuable insights, it cannot incorporate all factors influencing dengue
²⁴⁴ transmission, emphasizing the need for ongoing validation and refinement.

²⁴⁵ 1.5 Significance of the Research

²⁴⁶ This study's development of an AI-based dengue forecasting and monitoring sys-
²⁴⁷ tem has wide-reaching significance for various stakeholders in Iloilo City:

²⁴⁸ • Public Health Agencies: Organizations like the Department of Health (DOH)
²⁴⁹ and local health units in Iloilo City and Province stand to benefit greatly
²⁵⁰ from the system. With dengue predictions, we can help these agencies opti-
²⁵¹ mize their response strategies and implement targeted prevention measures
²⁵² in high-risk areas before cases escalate.

²⁵³ • Local Government Units (LGUs): LGUs can use the system to support
²⁵⁴ their disaster management and health initiatives by proactively addressing
²⁵⁵ dengue outbreaks. The predictive insights allow for more efficient planning
²⁵⁶ and resource deployment in barangays and communities most vulnerable to
²⁵⁷ outbreaks, improving overall public health outcomes.

²⁵⁸ • Healthcare Facilities: Hospitals and clinics, which currently face high bed
²⁵⁹ occupancy rates during dengue season will benefit from early outbreak fore-
²⁶⁰ casts that can help in managing patient inflow and ensuring adequate hos-
²⁶¹ pital capacity.

²⁶² • Researchers and Policymakers: This AI-driven approach contributes val-
²⁶³ uable insights for researchers studying infectious disease patterns and policy-
²⁶⁴ makers focused on strengthening the national AI Roadmap. The system's
²⁶⁵ data can support broader initiatives for sustainable health infrastructure
²⁶⁶ and inform policy decisions on resource allocation for dengue control.

²⁶⁷ • Community Members: By reducing the frequency and severity of outbreaks,
²⁶⁸ this study ultimately benefits the community at large. This allows for timely
²⁶⁹ awareness campaigns and community engagement initiatives, empowering
²⁷⁰ residents with knowledge and preventative measures to protect themselves
²⁷¹ and reduce the spread of dengue.

²⁷² **Chapter 2**

²⁷³ **Review of Related Literature**

²⁷⁴ **2.1 Dengue**

²⁷⁵ Dengue disease is a tropical and subtropical mosquito-borne viral illness and is a
²⁷⁶ major health concern in the Philippines (Bravo, Roque, Brett, Dizon, & L'Azou,
²⁷⁷ 2014). The majority of individuals with dengue experience no symptoms. Fever is
²⁷⁸ the most common symptom, typically 4 to 7 days after being bitten by an infected
²⁷⁹ mosquito (Zhou & Malani, 2024). In recent years, the trend of dengue cases in
²⁸⁰ the Philippines has shown notable fluctuations, with periodic outbreaks occur-
²⁸¹ ring every 3 to 5 years, often influenced by climatic and environmental changes.
²⁸² According to the Department of Health (DOH), the number of reported cases
²⁸³ has steadily increased over the past decades, attributed to urbanization, popula-
²⁸⁴ tion growth, and inadequate vector control measures (World Health Organization
²⁸⁵ (WHO), 2018). Moreover, studies suggest that El Niño and La Niña events have
²⁸⁶ significant effects on dengue incidence, with warmer temperatures and increased
²⁸⁷ rainfall providing favorable breeding conditions for mosquitoes (Watts, Burke,
²⁸⁸ Harrison, Whitmire, & Nisalak, 2020). The study of Carvajal et. al. highlights
²⁸⁹ the temporal pattern of dengue cases in Metropolitan Manila and emphasizes the
²⁹⁰ significance of relative humidity as a key meteorological factor, alongside rainfall
²⁹¹ and temperature, in influencing this pattern (Carvajal et al., 2018).

²⁹² **2.2 Outbreak Definition**

²⁹³ The definition of an outbreak is a critical factor in disease surveillance, as it
²⁹⁴ determines the threshold at which an unusual increase in cases is considered a

295 public health concern. Studies suggest that outbreak thresholds should be context-
296 specific, given the variability in transmission dynamics across different locations
297 (Runge-Ranzinger et al., 2016). Outbreak definitions defined using the Endemic
298 Channel often base thresholds on 2 standard deviations (SD) above the mean
299 number of historic dengue cases. Other studies (Hemisphere, 2015) also used an
300 alert threshold of a 5-year mean plus 2 standard deviations. A study by (Brady,
301 Smith, Scott, & Hay, 2015) discussed that definitions of dengue outbreaks differ
302 significantly across regions and time, making them inconsistent and incomparable.

303 From the studies above, this research implements an outbreak definition basing
304 the threshold on 2 standard deviations (SD) above the mean number of historic
305 dengue cases. It is important to note that for future outbreak definitions, addi-
306 tional local context like available hospital space etc., must be taken into account
307 for a more effective outbreak definition.

308 **2.3 Existing System: RabDash DC**

309 RabDash, developed by the University of the Philippines Mindanao, is a web-
310 based dashboard for rabies data analytics. It combines predictive modeling with
311 genomic data, enabling local health authorities to optimize interventions and al-
312 locate resources more effectively. RabDash's modules include trend visualization,
313 geographic hotspot mapping, and predictive forecasting, utilizing Long Short-
314 Term Memory (LSTM) models for time-series forecasting (RabDashDC, 2024).

315 For DengueWatch, RabDash serves as a strong inspiration, particularly in
316 its monitoring, historical trend visualization, and forecasting capabilities. These
317 features align well with the needs of dengue control efforts, providing real-time
318 insights into outbreak trends and enabling more effective, data-driven decision-
319 making. RabDash's architecture is relevant to the DengueDash, as dengue out-
320 breaks similarly require time-series forecasting models. By using LSTM, RabDash
321 effectively models trends in outbreak data, which provides a framework for adapt-
322 ing LSTM to dengue forecasting. Research indicates that LSTM models outper-
323 form traditional methods, such as ARIMA and MLP, in handling the complexities
324 of time-dependent epidemiological data (Ligue & Ligue, 2022).

325 2.4 Deep Learning

326 The study of (Ligue & Ligue, 2022) highlights how data-driven models can help
327 predict dengue outbreaks. The authors compared traditional statistical meth-
328 ods, such as non-seasonal and seasonal autoregressive integrated moving average
329 (ARIMA), and traditional feed-forward network approach using a multilayer per-
330 ceptron (MLP) model with a deep learning approach using the long short-term
331 memory (LSTM) architecture in their prediction model. They found that the
332 LSTM model performs better in terms of accuracy. The LSTM model achieved a
333 much lower root mean square error (RMSE) compared to both MLP and ARIMA
334 models, proving its ability to capture complex patterns in time-series data (Ligue
335 & Ligue, 2022). This superior performance is attributed to LSTM's capacity
336 to capture complex, time-dependent relationships within the data, such as those
337 between temperature, rainfall, humidity, and mosquito populations, all of which
338 contribute to dengue incidence (Ligue & Ligue, 2022).

339 2.5 Kalman Filter

340 The Kalman Filter is another powerful tool for time-series forecasting that can be
341 integrated into our analysis. It provides a recursive solution to estimating the state
342 of a linear dynamic system from a series of noisy measurements. Its application
343 in epidemiological modeling can enhance prediction accuracy by accounting for
344 uncertainties in the data(Li et al., 2022). Studies have shown that Kalman filters
345 are effective in predicting infectious disease outbreaks by refining estimates based
346 on observed data. A study published in Frontiers in Physics utilized the Kalman
347 filter to predict COVID-19 deaths in Ceará, Brazil. They found that the Kalman
348 filter effectively tracked the progression of deaths and cases, providing critical in-
349 sights for public health decision-making (Ahmadini et al., 2021). Another research
350 article in PLOS ONE focused on tracking the effective reproduction number (R_t)
351 of COVID-19 using a Kalman filter. This method estimated the growth rate of
352 new infections from noisy data, demonstrating that the Kalman filter could main-
353 tain accurate estimates even when case reporting was inconsistent(Arroyo-Marioli,
354 Bullano, Kucinskas, & Rondón-Moreno, 2021).

355 This study compares ARIMA, seasonal ARIMA, Kalman Filter, and LSTM
356 models using collected dengue case data along with weather data to identify the
357 most effective model for real-time forecasting.

358 2.6 Weather Data

359 The relationship between weather patterns and mosquito-borne diseases is inher-
360 ently nonlinear, meaning that fluctuations in disease cases do not respond propor-
361 tionally to changes in climate variables(Colón-González, Fezzi, Lake, & Hunter,
362 2013) Weather data, such as minimum temperature and accumulated rainfall, are
363 strongly linked to dengue case fluctuations, with effects observed after several
364 weeks due to mosquito breeding and virus incubation cycles. Integrating these
365 lagged weather effects into predictive models can improve early warning systems
366 for dengue control(Cheong, Burkart, Leitão, & Lakes, 2013). A study also suggests
367 that weather-based forecasting models using variables like mean temperature and
368 cumulative rainfall can provide early warnings of dengue outbreaks with high sen-
369 sitivity and specificity, enabling predictions up to 16 weeks in advance(Hii, Zhu,
370 Ng, Ng, & Rocklöv, 2012).

371 This study utilizes weather data, including variables such as temperature,
372 rainfall, and humidity, as inputs for our dengue forecasting model. Given the
373 strong, nonlinear relationship between climate patterns and dengue incidence,
374 these weather variables, along with their lagged effects, are essential for enhancing
375 prediction accuracy and providing timely early warnings for dengue outbreaks.

376 2.7 Chapter Summary

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

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

³⁹¹

Chapter 3

³⁹²

Research Methodology

³⁹³ 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.

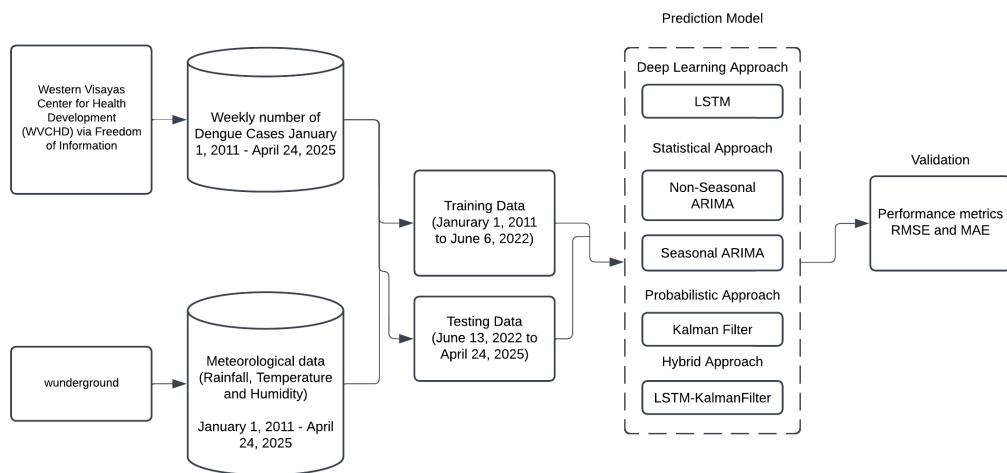


Figure 3.1: Workflow for forecasting the number of weekly dengue cases

³⁹⁶ This summarizes the workflow for forecasting the number of weekly dengue cases. This workflow focuses on using statistical, deep learning, and probabilistic ³⁹⁷ models to forecast the number of reported dengue cases. The approach involves ³⁹⁸ deploying several models for prediction, including ARIMA and Seasonal ARIMA ³⁹⁹ as statistical approaches, LSTM as a deep learning approach, and the Kalman ⁴⁰⁰ Filter as a probabilistic approach. These methods are compared with each other ⁴⁰¹

402 to determine the most accurate model.

403 3.1 Research Activities

404 3.1.1 Gather Dengue Data and Climate Data to Create a 405 Complete Dataset for Forecasting

406 Acquisition of Dengue Case Data

407 The historical dengue case dataset used in this study was obtained from the Hu-
408 manitarian Data Exchange and the Western Visayas Center for Health Develop-
409 ment (WVCHD) via Freedom of Information (FOI) requests. The decision to use
410 weekly intervals was driven by the need for precision and timeliness in captur-
411 ing fluctuations in dengue cases and weather conditions. Dengue transmission is
412 influenced by short-term changes in weather variables such as rainfall and temper-
413 ature, which impact mosquito breeding and virus transmission cycles. A weekly
414 granularity allowed the model to better capture these short-term trends, enabling
415 more accurate predictions and responsive public health interventions.

416 Moreover, using a weekly interval provided more data points for training the
417 models compared to a monthly format. This is particularly critical in time series
418 modeling, where larger datasets help improve the robustness of the model and its
419 ability to generalize to new data. Also, the collection of weather data was done
420 by utilizing web scraping techniques to extract weekly weather data (e.g., rainfall,
421 temperature, and humidity) from Weather Underground (wunderground.com).

422 Data Fields

424 • **Time.** Represents the specific year and week corresponding to each entry
425 in the dataset.

426 • **Rainfall.** Denotes the observed average rainfall, measured in millimeters,
427 for a specific week.

428 • **Humidity.** Refers to the observed average relative humidity, expressed as
429 a percentage, for a specific week.

430 • **Max Temperature.** Represents the observed maximum temperature, mea-
431 sured in degrees Celsius, for a specific week.

- **Average Temperature.** Represents the observed average temperature, measured in degrees Celsius, for a specific week.
- **Min Temperature.** Represents the observed minimum temperature, measured in degrees Celsius, for a specific week.
- **Wind.** Represents the observed wind speed, measured in miles per hour (mph), for a specific week.
- **Cases.** Refers to the number of reported dengue cases during a specific week.

440 Data Integration and Preprocessing

441 The dengue case data was integrated with the weather data to create a com
 442 prehensive dataset, aligning the data based on corresponding timeframes. The
 443 dataset undergoed a cleaning process to address any missing values, outliers, and
 444 inconsistencies to ensure its accuracy and reliability. To ensure that all features
 445 and the target variable were on the same scale, a MinMaxScaler was applied to
 446 normalize both the input features (climate data) and the target variable (dengue
 447 cases).

448 Exploratory Data Analysis (EDA)

- Analyzed trends, seasonality, and correlations between dengue cases and weather factors.
- Created visualizations like time series plots and scatterplots to highlight relationships and patterns in the data.

453 Outbreak Detection

454 To detect outbreaks, we computed the outbreak threshold value of dengue cases
 455 using the formula,

$$\text{Outbreak Threshold Value} = \mu + 2\sigma \quad (3.1)$$

456 where μ is the historical mean and σ is the standard deviation.

457 **3.1.2 Develop and Evaluate Deep Learning Models for**
458 **Dengue Case Forecasting**

459 The deep learning models were developed and trained to forecast weekly dengue
460 cases using historical weather data (rainfall, temperature, wind, and humidity)
461 and dengue case counts. The dataset was normalized and divided into training and
462 testing sets, ensuring temporal continuity to avoid data leakage. The methodology
463 for preparing and training the model are outlined below.

464 **Data Preprocessing**

465 The raw dataset included weekly aggregated weather variables (rainfall, tempera-
466 ture, wind, humidity) and dengue case counts. The "Time" column was converted
467 to a datetime format to ensure proper temporal indexing. To standardize the data
468 for training, MinMaxScaler was employed, normalizing the feature values and tar-
469 get variable to a range of 0 to 1. This step ensured that the models could efficiently
470 process the data without being biased by feature scaling differences.

471 **LSTM Model**

472 The dataset was split into training and test sets to evaluate the model's perfor-
473 mance and generalizability:

- 474 • **Training Set:** 80% of the data (572 sequences) was used for model training,
475 enabling the LSTM to learn underlying patterns in historical dengue case
476 trends and their relationship with weather variables.
- 477 • **Test Set:** The remaining 20% of the data (148 sequences) was reserved for
478 testing

479 To prepare the data for LSTM, a sliding window approach was utilized. Se-
480 quences of weeks of normalized features were constructed as input, while the
481 dengue case count for the subsequent week was set as the target variable. This
482 approach ensured that the model leveraged temporal dependencies in the data for
483 forecasting. To enhance the performance of the LSTM model in predicting dengue
484 cases, Bayesian Optimization was employed using the Keras Tuner library. The
485 tuning process aimed to minimize the validation loss (mean squared error) by
486 adjusting key model hyper-parameters. The search space is summarized below:

487 **LSTM units:**

- 488 • min value: 32
- 489 • max value: 128
- 490 • step: 16
- 491 • sampling: linear

492 **Learning Rate:**

- 493 • min value: 0.0001
- 494 • max value: 0.01
- 495 • step: None
- 496 • sampling: log

497 The tuner was instanstiated with:

- 498 • **max trials = 10:** Limiting the search to 10 different configurations
- 499 • **executions per trial = 3:** Running each configuration thrice to reduce
500 variance
- 501 • **validation split = 0.2:** Reserving 20% of the training data for validation

502 The hyperparameter tuning was conducted for three different window sizes of
503 data: 5, 10, and 20. This allows the model to have the optimal hyperparameters
504 used for each window size. Training was conducted over 100 epochs with early
505 stopping to prevent overfitting while maintaining computational efficiency. A
506 batch size of 1 was used, enabling the model to process individual sequences,
507 which is suitable for smaller datasets but results in longer training times. The
508 Adam optimizer, known for its adaptive learning capabilities and stability was
509 employed.

510 To validate the effectiveness of the model, cross-validation was implemented.
511 However, standard k-fold cross-validation randomly shuffles the data, which isn't
512 suitable for time series since the order of observations is important. To address
513 this, a time series-specific cross-validation strategy was used with TimeSeriesS-
514 plit from the scikit-learn library. This method creates multiple train-test splits

515 where each training set expands over time and each test set follows sequentially.
516 This approach preserves the temporal structure of the data while helping reduce
517 overfitting by validating the model across different time segments.

518 After training, predictions on both the training and test datasets were rescaled
519 to their original scale using the inverse transformation of MinMaxScaler. Model
520 performance was evaluated using the mean squared error (MSE), root mean
521 squared error (RMSE) and mean absolute error (MAE).

522 **ARIMA**

523 The ARIMA model was utilized to forecast weekly dengue cases, leveraging histori-
524 cal weather data—including rainfall, maximum temperature, and humidity—as
525 exogenous variables alongside historical dengue case counts as the primary depen-
526 dent variable. The dataset was partitioned into training (80%) and testing (20%)
527 sets while maintaining temporal consistency.

528 To identify the optimal ARIMA configuration, a comprehensive grid search
529 was performed across the following parameter ranges:

- 530 • Autoregressive order (p): 0 to 3
- 531 • Differencing order (d): 0 to 2
- 532 • Moving average order (q): 0 to 3

533 Each combination of (p,d,q) was used to fit an ARIMA model, and perfor-
534 mance was evaluated based on the mean squared error (MSE) between the pre-
535 dicted and actual dengue cases on the test set. The parameter set that achieved
536 the lowest MSE was selected as the final model configuration.

537 Following model selection, the best-fit ARIMA model was retrained on the
538 training set and subsequently used to forecast dengue cases for the test period.
539 The predictions were assigned to the **PredictedCases** column in the test dataset.
540 Model performance was further assessed using key evaluation metrics, including
541 MSE, root mean squared error (RMSE), and mean absolute error (MAE). Visual
542 comparisons between actual and predicted dengue cases were produced through
543 line plots to better illustrate the model's forecasting accuracy.

544 **Seasonal ARIMA (SARIMA)**

545 The SARIMA modeling process began with data preprocessing, which included
546 handling missing values through interpolation or imputation, and standardizing
547 features to ensure stable model training. The dataset was then split into training
548 and testing sets in an 80:20 ratio, preserving the temporal order of observations.
549 Seasonality analysis was conducted using time series decomposition and autocor-
550 relation plots, which revealed a periodicity of 52 weeks—justifying the adoption
551 of a seasonal model. To fine-tune the model, a grid search was performed over a
552 range of SARIMA parameters $(p,d,q)(P,D,Q)[S]$, while stationarity was validated
553 using the Augmented Dickey-Fuller (ADF) test. The model was then trained
554 on the dataset using rainfall, temperature, and humidity as exogenous variables,
555 with convergence ensured by setting a maximum number of iterations. Residual
556 diagnostics were used to confirm that residuals were uncorrelated, indicating a
557 good model fit. For evaluation, forecasts were compared against actual values,
558 and results were visualized with line plots. Finally, to validate the model's gener-
559 alizability across different time periods, Time Series Cross-Validation with three
560 folds was applied. This allowed assessment of the model's performance on multi-
561 ple time segments, providing insights into its robustness in real-world forecasting
562 scenarios.

563 **Kalman Filter:**

- 564 • Input Variables: The target variable (Cases) was modeled using three re-
565 gressors: rainfall, max temperature, and humidity.
- 566 • Training and Testing Split: The dataset was split into 80% training and
567 20% testing to evaluate model performance.
- 568 • Observation Matrix: The Kalman Filter requires an observation matrix,
569 which was constructed by adding an intercept (column of ones) to the re-
570 gressors.

571 The Kalman Filter's Expectation-Maximization (EM) method was employed
572 for training, iteratively estimating model parameters over 10 iterations. After
573 training, the smoothing method was used to compute the refined state estimates
574 across the training data. Observation matrices for the test data were constructed
575 in the same manner as for the training set, ensuring compatibility with the learned
576 model parameters. On the test data, the Kalman Filter applied these parameters
577 to predict and correct the estimated dengue cases, providing more stable and
578 accurate forecasts compared to direct regression models. Additionally, a hybrid

579 Kalman Filter–LSTM (KF-LSTM) model was developed to combine the strengths
580 of both approaches. In this setup, the LSTM model was first used to predict
581 dengue cases based on historical data and weather features. The Kalman Filter
582 was then applied as a post-processing step to the LSTM predictions, smoothing
583 out noise and correcting potential errors.

584 **3.1.3 Integrate the Predictive Model into a Web-Based**
585 **Data Analytics Dashboard**

586 **Dashboard Design and Development**

- 587 • Design an intuitive, user-friendly web-based dashboard incorporating:
- 588 – Interactive visualizations of yearly dengue case trends.
589 – Data input and update forms for dengue and weather data.
590 – Map display of dengue cases in each district in Iloilo City

591 **Model Integration and Deployment**

- 592 • Deploy the best-performing model within the dashboard as a backend service
593 to enable real-time or periodic forecasting.

594 **3.1.4 System Development Framework**

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

601 **Design and Development**

602 After brainstorming and researching the most appropriate type of application to
603 accommodate both the prospected users and the proposed solutions, the team
604 has decided to proceed with a web application. Given the time constraints and
605 available resources, it has been decided that the said means is the most pragmatic

and practical move. The next step is to select modern and stable frameworks that align with the fundamental ideas learned by the researchers in the university. The template obtained from WVCHD and Iloilo Provincial Epidemiology and Surveillance Unit was meticulously analyzed to create use cases and develop a preliminary well-structured database that adheres to the requirements needed to produce a quality application. The said use cases serve as the basis of general features. Part by part, these are converted into code, and with the help of selected libraries and packages, it resulted in the desired outcome that may still modified and extended to achieve scalability.

Testing and Integration

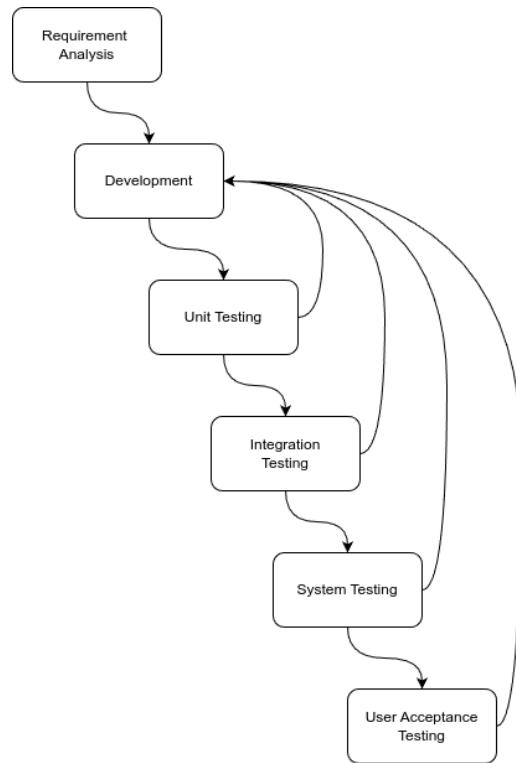


Figure 3.2: Testing Process for DengueWatch

Implementing testing is important to validate the system's performance and efficacy. Thus a series of tests were conducted to identify and resolve bugs during the developmental phase. Each feature was rigorously tested to ensure quality assurance, with particular emphasis on prerequisite features, as development cannot progress properly if these fail. Because of this, integration between each feature

621 serves as a pillar for a cohesive user experience. Since dengue reports include
622 confidential information, anonymized historical dengue reports were used to train
623 the model and create the foundational architecture of the system. By using func-
624 tional tests, data validation and visualization can be ensured for further continual
625 improvements. Security testing is also important as it is needed to safeguard
626 confidential information when the system is deployed. It includes proper authen-
627 tication, permission views, and mitigating common injection attacks. Finally, a
628 user acceptance test from the prospected users, in this case, doctors, nurses, and
629 other health workers is crucial to assess its performance and user experience. It
630 enables the developers to confirm if the system meets the needs of the problem.

631 **3.2 Development Tools**

632 **3.2.1 Software**

633 **Github**

634 GitHub is a cloud-based platform that tracks file changes using Git, an open-
635 source version control system (*About GitHub and Git - GitHub Docs*, n.d.). It is
636 used in the project to store the application's source code, manage the system's
637 source version control, and serve as a repository for the Latex files used in the
638 actual research.

639 **Visual Studio Code**

640 Visual Studio Code is a free, lightweight, and cross-platform source code editor
641 developed by Microsoft (*Why Visual Studio Code?*, 2021). As VS Code supports
642 this project's programming and scripting languages, it was chosen as the primary
643 source code editor.

644 **Django**

645 Django is a free and open-sourced Python-based web framework that offers an
646 abstraction to develop and maintain a secure web application. As this research
647 aims to create a well-developed and maintainable application, it is in the best
648 interest to follow an architectural pattern that developers and contributors in the

649 future can understand. Since Django adheres to Model-View-Template (MVT)
650 that promotes a clean codebase by separating data models, business logic, and
651 presentation layers, it became the primary candidate for the application's back-
652 bone.

653 **Next.js**

654 A report by Statista (2024) claims that React is the most popular front-end frame-
655 work among web developers. However, React has limitations that can be a nui-
656 sance in rapid software development, which includes routing and performance op-
657 timizations. This is where Next.js comes in—a framework built on top of React.
658 It offers solutions for React's deficiency, making it a rising star in the framework
659 race.

660 **Postman**

661 As the application heavily relies on the Application Programming Interface (API)
662 being thrown by the backend, it is a must to use a development tool that facilitates
663 the development and testing of the API. Postman is a freemium API platform
664 that offers a user-friendly interface to create and manage API requests (*What is*
665 *Postman? Postman API Platform*, n.d.).

666 **3.2.2 Hardware**

667 The web application was developed on laptop computers with minimum specifica-
668 tions of an 11th-generation Intel i5 CPU and 16 gigabytes of RAM. Furthermore,
669 an RTX 3060 mobile GPU was also utilized to retrain the integrated model locally.

670 **3.2.3 Packages**

671 **Django REST Framework**

672 Django Rest Framework (DRF) is a third-party package for Django that provides a
673 comprehensive suite of features to simplify the development of robust and scalable
674 Web APIs (Christie, n.d.). These services include Serialization, Authentication
675 and Permissions, Viewsets and Routers, and a browsable API .

676 **Leaflet**

677 One of the features of the web application is the ability to map the number
678 of cases using a Choropleth Map. Leaflet is the only free, open-sourced, and
679 most importantly, stable JavaScript package that can do the job. With its ultra-
680 lightweight size, it offers a comprehensive set of features that does not trade
681 off performance and usability (*Leaflet — an open-source JavaScript library for*
682 *interactive maps*, n.d.).

683 **Chart.js**

684 Another feature of the application is to provide users with informative, approach-
685 able data storytelling that is easy for everyone to understand. The transformation
686 of pure data points and statistics into figures such as charts is a big factor. Thus,
687 there is a need for a package that can handle this feature without compromising
688 the performance of the application. Chart.js is a free and open-source JavaScript
689 package that is made to meet this criteria as it supports various types of charts
690 (*Chart.js*, n.d.).

691 **Tailwind CSS**

692 Using plain CSS in production-quality applications can be counterproductive.
693 Therefore, CSS frameworks were developed to promote consistency and accelerate
694 the rapid development of web applications (Joel, 2021). One of these is Tailwind,
695 which offers low-level utility classes that can be applied directly to each HTML
696 element to create a custom design (*Tailwind CSS - Rapidly build modern web-*
697 *sites without ever leaving your HTML.*, n.d.). Given the limited timeline for this
698 project, using this framework is a wise choice due to its stability and popularity
699 among developers.

700 **Shadcn**

701 Shadcn offers a collection of open-source UI boilerplate components that can be
702 directly copied and pasted into one's project. With the flexibility of the provided
703 components, Shadcn allows developers to have full control over customization and
704 styling. Since this is built on top of Tailwind CSS and Radix UI, it is supported
705 by most modern frontend frameworks, including Next.js (Shadcn, n.d.).

706 **Zod**

707 Data validation is integral in this web application since it will handle crucial data
708 that will be used for analytical inferences and observations. Since Zod is primarily
709 used for validating and parsing data, it ensures proper communication between
710 the client and the server (Zod, n.d.).

711 **3.3 Application Requirements**

712 **3.3.1 Backend Requirements**

713 **Database Structure Design**

714 Determining how data flows and how it would be structured is crucial in creating
715 the system as it defines how extendible and flexible it would be for future features
716 and updates. Thus, creating a comprehensive map of data ensures proper nor-
717 malization that eliminates data redundancy and improves data integrity. Figure
718 3.3 depicts the designed database schema that showcases the relationship between
719 the application's entities.

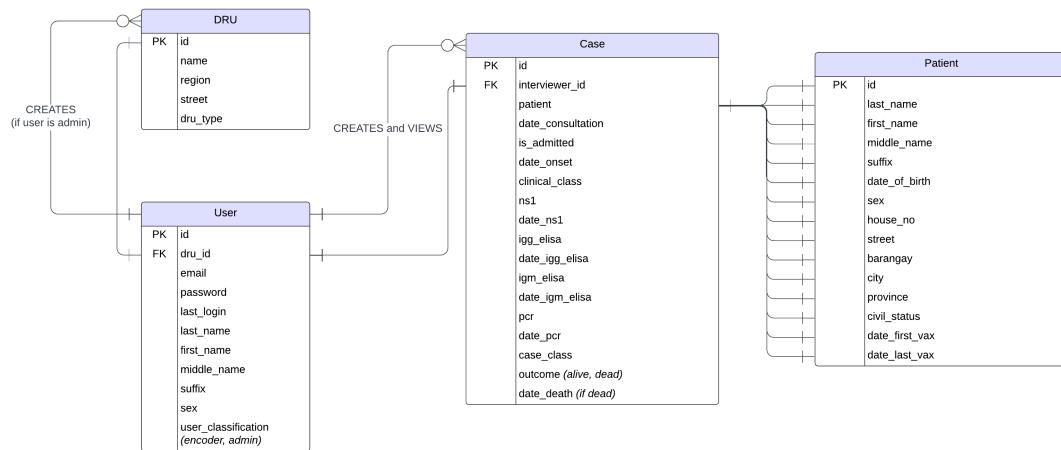


Figure 3.3: Entity-Relationship Database Schema Hybrid Diagram for DengueDash Database Structure

⁷²⁰ **3.3.2 User Interface Requirements**

⁷²¹ **Admin Interface**

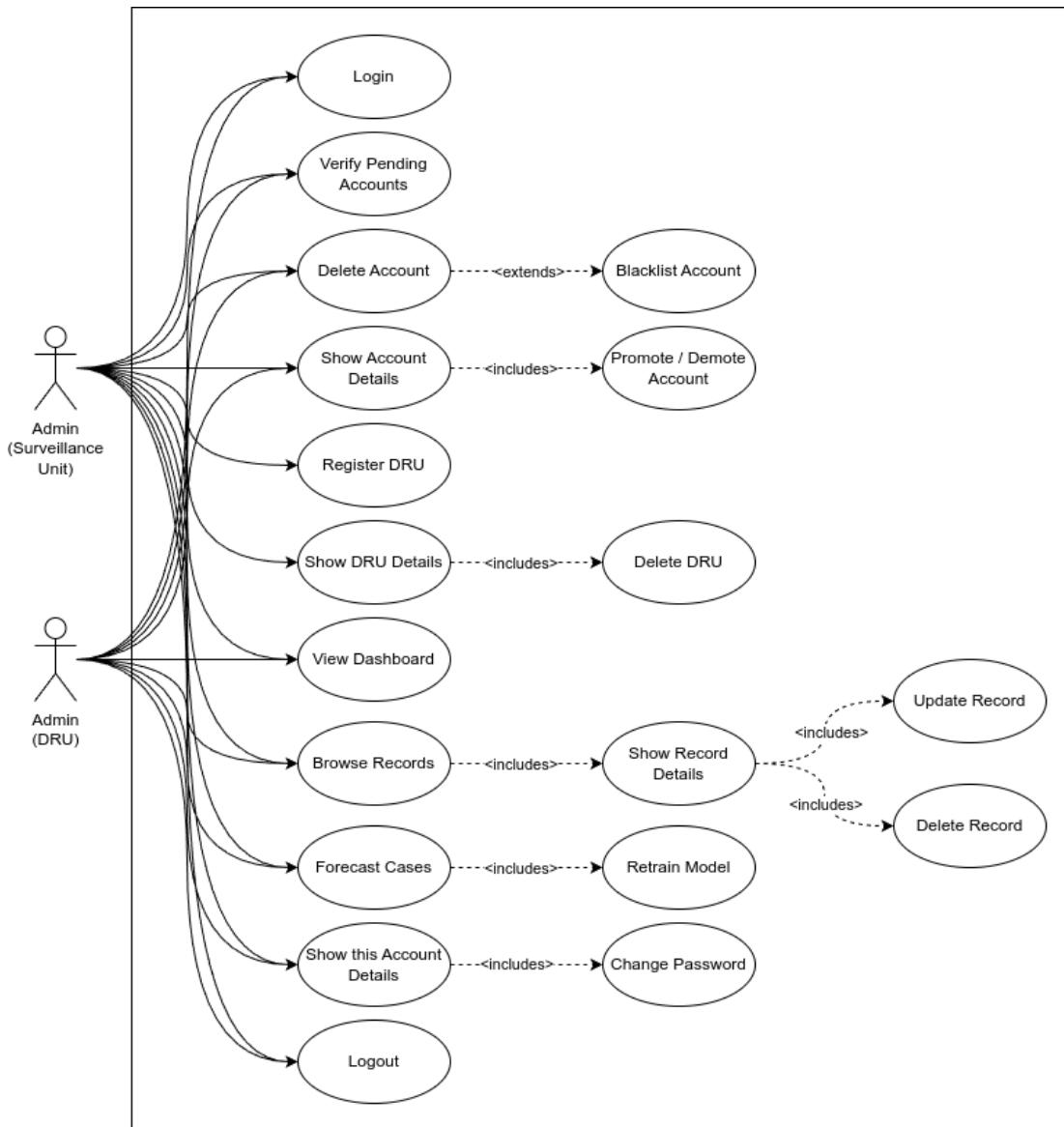


Figure 3.4: Use Case Diagram for Admins

⁷²² Figure 3.4 shows the actions of an admin for a specific Disease Reporting Unit (DRU) and an admin for a specific Surveillance Unit can take in the application.
⁷²³ Both of them include the management of accounts, browsing records, and forecasting and retraining all the consolidated data under their supervision. Most
⁷²⁴
⁷²⁵

726 importantly, these users must verify the encoders who register under their ju-
727 risdiction before allowing their account to access the application in the name of
728 safeguarding the integrity of the data. The only advantage of the latter type of ad-
729 ministrator is that it has a one-step higher authorization as it manages the DRUs.
730 In addition, only the authorized surveillance unit administrator can register and
731 create a DRU to uphold transparency and accountability.

732 **Encoder Interface**

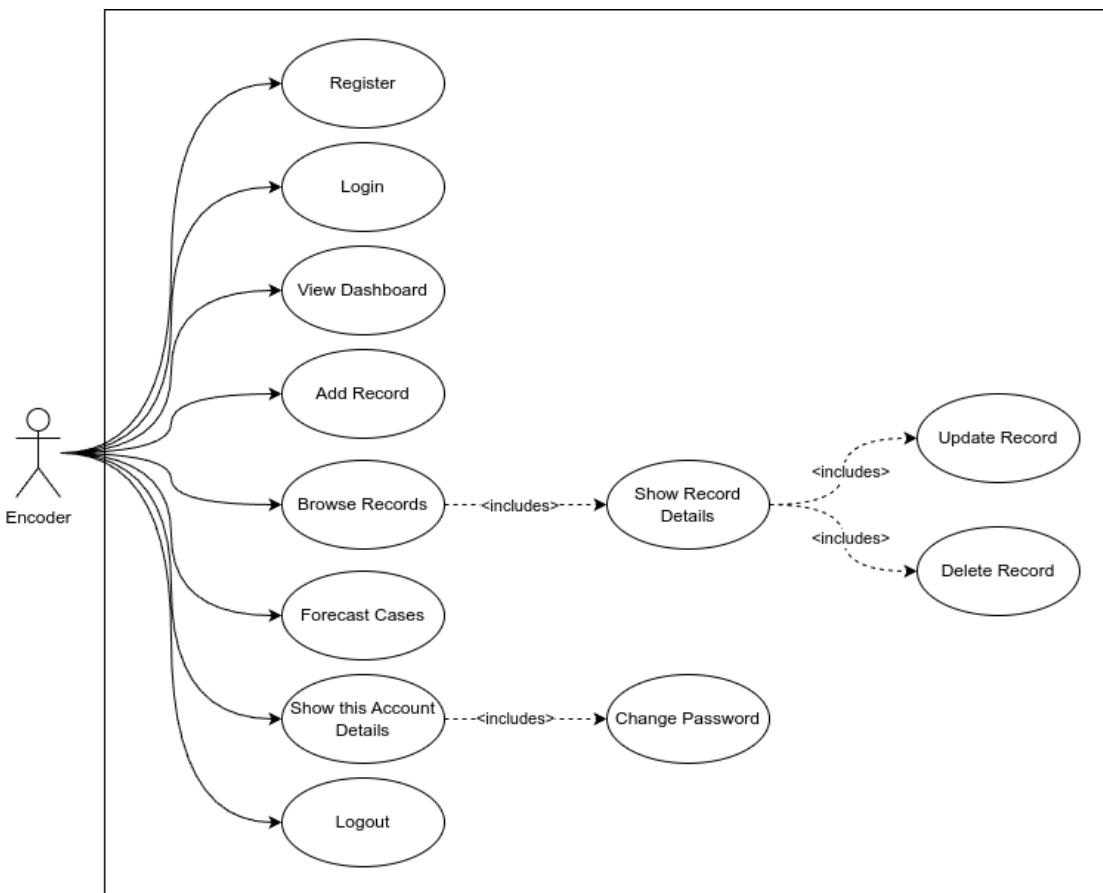


Figure 3.5: Use Case Diagram for Encoder

733 Figure 3.5, on the other hand, illustrates the use cases for the system's primary
734 users. These users can register but must wait for further verification to access the
735 application. Similar to the previous interfaces, encoders can browse and manage
736 records, as well as forecast the consolidated cases under a specific surveillance or
737 disease reporting unit, but they are not allowed to retrain the model. Lastly, they

738 are the only type of user that can file and create dengue cases by filling out a form
739 with the required details.

740 3.3.3 Security and Validation Requirements

741 Password Encryption

742 Storing passwords as plain text in the database is a disgrace and a mortal sin in
743 production. It is important to implement precautionary methods such as hashing
744 and salting, followed by encryption with a strong algorithm, to prevent bad actors
745 from using the accounts for malicious transactions. By default, Django generates
746 a unique random salt for each password and encrypts it with Password-Based Key
747 Derivation Function 2 (PBKDF2) with a SHA256 hash function. Utilizing these
748 techniques ensures that in the event of a data breach, cracking these passwords
749 would be time-consuming and useless for the attackers.

750 Authentication

751 DengueWatch utilizes JSON Web Tokens (JWT) to authenticate the user. Since
752 the mechanism operates in a stateless manner, tokens are served only after a
753 successful login, eliminating the need for the server to keep a record of the token,
754 which is vulnerable to session hijacking. In addition, these tokens are signed with
755 a secret key, ensuring they have not been tampered with.

756 Data Validation

757 Both the backend and frontend should validate the input from the user to preserve
758 data integrity. Thus, Zod is implemented in the latter to help catch invalid inputs
759 from the user. By doing this, the user can only send proper requests to the server
760 which streamlines the total workflow. On the other hand, Django has also a built-
761 in validator that checks the data type and ensures that the input matches the
762 expected format on the server side. These validation processes ensure that only
763 valid and properly formatted data is accepted, which reduces the risk of errors
764 and ensures consistency across the web application.

⁷⁶⁵ Chapter 4

⁷⁶⁶ Results and Discussion/System ⁷⁶⁷ Prototype

⁷⁶⁸ 4.1 Data Gathering

⁷⁶⁹ The data for dengue case prediction was gathered from a variety of reliable sources,
⁷⁷⁰ enabling a comprehensive dataset spanning from January 2011 to October 2024.
⁷⁷¹ This dataset includes 720 rows of data, each containing weekly records of dengue
⁷⁷² cases along with corresponding meteorological variables, such as rainfall, temper-
⁷⁷³ ature, and humidity.

- ⁷⁷⁴ 1. Dengue Case Data: The primary source of historical dengue cases came
⁷⁷⁵ from the Humanitarian Data Exchange and the Western Visayas Center for
⁷⁷⁶ Health Development (WVCHD). The dataset, accessed through Freedom of
⁷⁷⁷ Information (FOI) requests, provided robust case numbers for the Western
⁷⁷⁸ Visayas region. The systematic collection of these data points was essential
⁷⁷⁹ for establishing a reliable baseline for model training and evaluation.
- ⁷⁸⁰ 2. Weather Data: Weekly weather data was obtained by web scraping from
⁷⁸¹ Weather Underground, allowing access to rainfall, temperature, wind, and
⁷⁸² humidity levels that correlate with dengue prevalence.

⁷⁸³ 4.2 Exploratory Data Analysis

⁷⁸⁴ From Figure 4.2, the dataset consists of 720 weekly records with 8 columns:

	Time	Rainfall	MaxTemperature	AverageTemperature	MinTemperature	Wind	Humidity	Cases
0	2011-01-03	9.938571	29.444400	25.888890	23.888900	11.39	86.242857	5
1	2011-01-10	8.587143	30.000000	26.705556	24.444444	7.32	88.028571	4
2	2011-01-17	5.338571	30.000000	26.616667	25.000000	7.55	84.028571	2
3	2011-01-24	5.410000	30.555556	26.483333	20.555556	10.67	80.971429	7
4	2011-01-31	2.914286	28.333333	25.283333	18.650000	11.01	74.885714	2

Figure 4.1: Snippet of the Combined Dataset

- 785 • **Time.** Weekly timestamps (e.g. "2011-w1")
- 786 • **Rainfall.** Weekly average rainfall (mm)
- 787 • **MaxTemperature, AverageTemperature, MinTemperature.** Weekly
788 temperature data (C)
- 789 • **Wind.** Wind speed (m/s)
- 790 • **Humidity.** Weekly average humidity (%)
- 791 • **Cases.** Reported dengue cases

#	Column	Non-Null Count	Dtype
0	Time	720 non-null	datetime64[ns]
1	Rainfall	720 non-null	float64
2	MaxTemperature	720 non-null	float64
3	AverageTemperature	720 non-null	float64
4	MinTemperature	720 non-null	float64
5	Wind	720 non-null	float64
6	Humidity	720 non-null	float64
7	Cases	720 non-null	int64
dtypes: datetime64[ns](1), float64(6), int64(1)			
memory usage: 45.1 KB			

Figure 4.2: Data Contents

- 792 From the statistics in figure 4.3, the number of cases ranges from 0 to 319.
 793 The average number of dengue cases per week is 23.74, with a median of 12 cases
 794 and a standard deviation of 37.14. The distribution is highly skewed, with some
 795 weeks experiencing significant number of cases (up to 319 cases). Rainfall shows
 796 a wide variation (0 to 445mm), while temperature remains relatively stable, with

	Time	Rainfall	MaxTemperature	AverageTemperature	MinTemperature	Wind	Humidity	Cases
count	720	720.000000	720.000000	720.000000	720.000000	720.000000	720.000000	720.000000
mean	2017-12-02 11:22:00	13.957499	32.191142	28.110319	25.038472	6.172417	81.609442	23.744444
min	2011-01-03 00:00:00	0.000000	-14.600000	24.494444	12.222222	1.910000	73.185714	0.000000
25%	2014-06-21 06:00:00	1.270000	31.666667	27.504167	25.000000	4.117500	79.885713	5.000000
50%	2017-12-07 12:00:00	4.318000	32.222222	28.161111	25.000000	5.725000	81.771429	12.000000
75%	2021-05-11 18:00:00	10.414000	32.777778	28.751389	25.555556	7.860000	83.503571	26.000000
max	2024-10-28 00:00:00	445.008000	58.333333	30.916667	32.222222	19.200000	89.571429	319.000000
std		NaN	35.448846	2.616379	0.999800	1.291659	2.446703	2.831674

Figure 4.3: Dataset Statistics

797 an average of 28.1 degree celsius. Humidity levels ranges from 73% to 89% with
 798 a mean of 81.6%.

799 Figure 4.4 illustrates the trend of weekly dengue cases over time. The data
 800 reveals periodic spikes in the number of cases, suggesting a seasonal pattern in
 801 dengue cases. Notably, peak cases are observed during certain periods approx-
 802 imately 3 years, potentially aligning with specific climatic conditions such as
 803 increased rainfall or temperature changes. This underscores the importance of
 804 incorporating climate variables into the forecasting model.

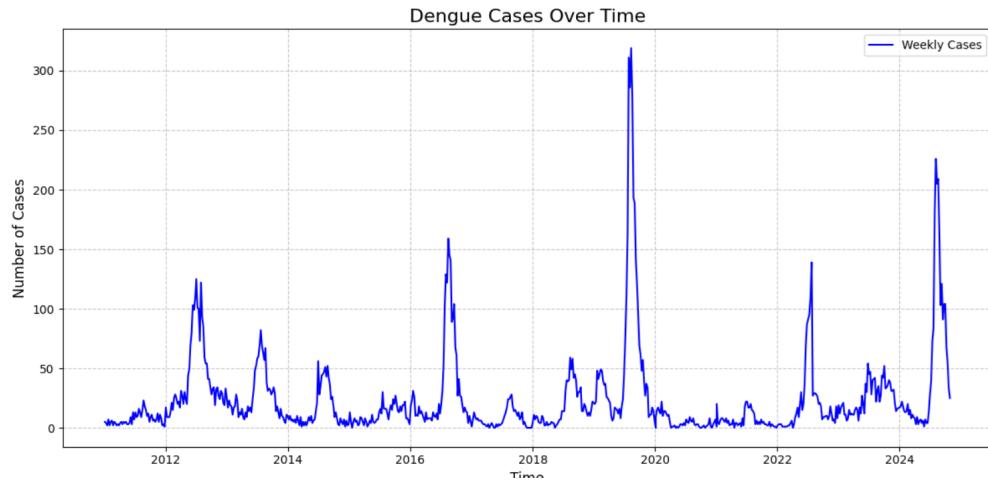


Figure 4.4: Trend of Dengue Cases

805 Figure 4.5 shows the ranking of correlation coefficients between dengue cases
 806 and selected features, including rainfall, humidity, maximum temperature, aver-
 807 age temperature, minimum temperature, and wind speed. Among these, rainfall
 808 exhibits the highest positive correlation with dengue cases (correlation coefficient
 809 0.13), indicating that increased rainfall may contribute to higher cases counts.
 810 This aligns with existing studies suggesting that stagnant water from heavy rain-

fall creates breeding grounds for mosquitos. It is followed by humidity (0.10), suggesting that higher humidity levels may enhance mosquito reproduction, leading to more dengue cases. Temperature has a weak to moderate positive correlation with dengue cases, with maximum temperature (0.09) showing a stronger relationship than average and minimum temperature.

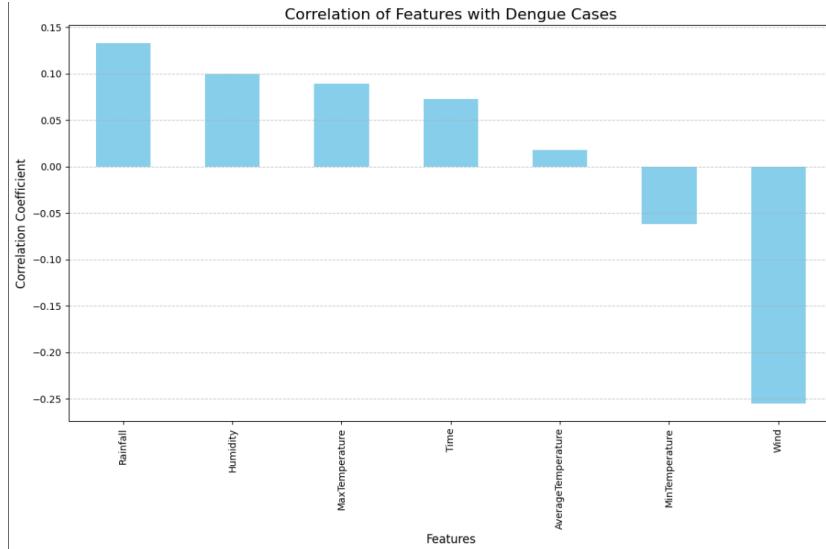


Figure 4.5: Ranking of Correlations

Figure 4.6 shows the distributions of each feature while Figure 4.7 shows scatterplots of each feature against the number of cases. The distributions of Rainfall, Max Temperature, Min Temperature, and Wind appear skewed, which is common for many real-world variables. This skewness can distort correlation estimates, as Pearson correlation assume linear relationships and are more reliable when variables follow a symmetric or approximately normal distribution (Bobbitt, 2021). Applying a log transformation can help normalize these distributions, improve linearity, and thus lead to more meaningful and accurate correlation analysis.

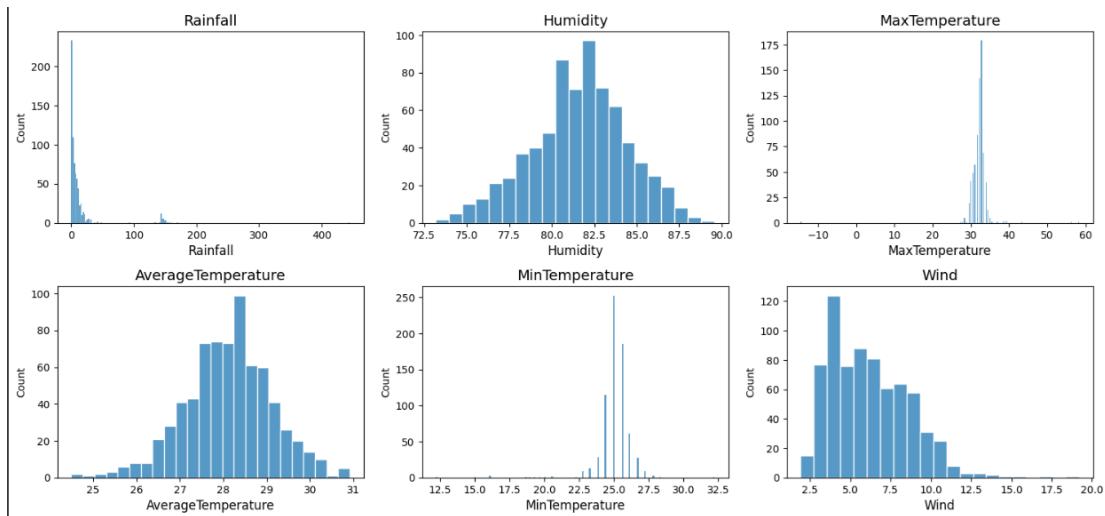


Figure 4.6: Pre-Transform Feature Distributions

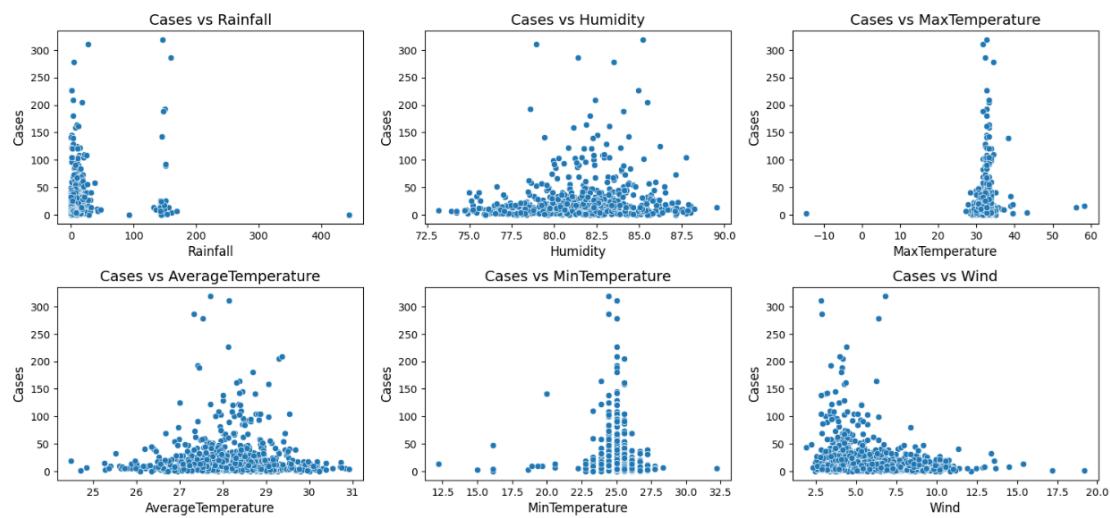


Figure 4.7: Scatterplots

824 After applying a log transformation, Figure 4.8 shows the new distributions for
 825 the previously skewed distributions, while Figure 4.9 shows the new scatterplots
 826 of each feature against the number of cases. Now, all distributions exhibit a
 827 somewhat normal distribution which is ideal for computing linear computations
 828 such as Pearson's correlation.

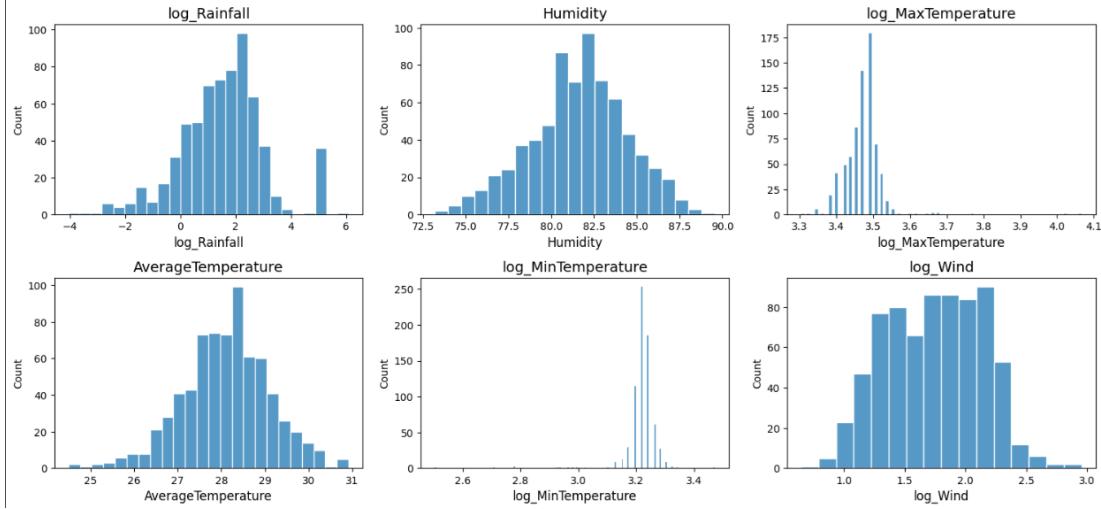


Figure 4.8: Post-Transform Feature Distributions

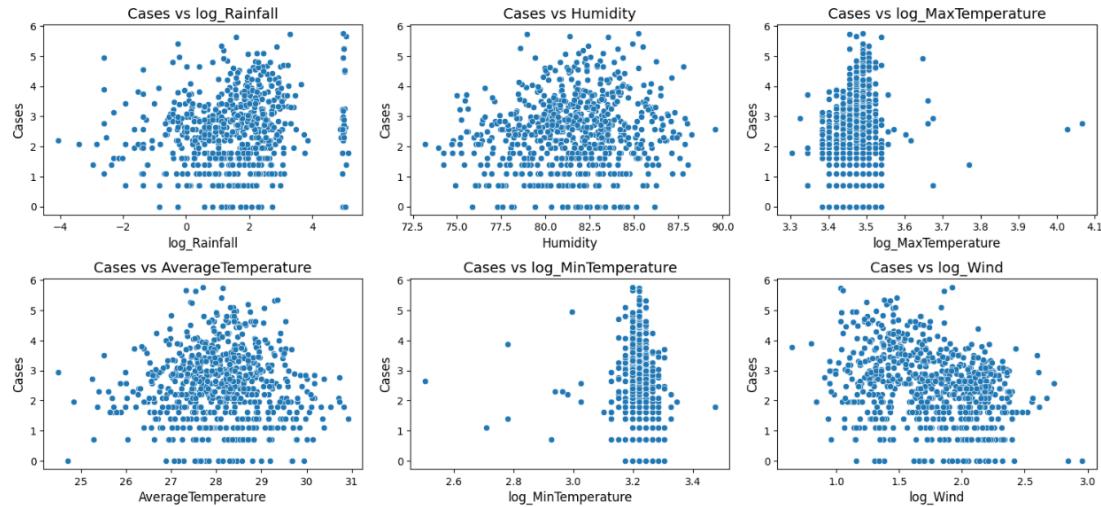


Figure 4.9: Transformed Distributions: Scatterplots

829 Figure 4.10 presents the recomputed correlation coefficients between dengue
 830 cases and the log-transformed weather features. Rainfall shows the strongest cor-
 831 relation at 0.16, followed by Max Temperature at 0.12, and Humidity at 0.10.
 832 While other features are included, their correlation values are very small and not

833 considered meaningful. Although the individual correlations are weak, they provide
 834 valuable signals that, when combined in a multivariate model, may contribute
 835 meaningfully to predictive performance., As a result, Rainfall, Max Temperature,
 836 and Humidity are selected as the key features for model training.

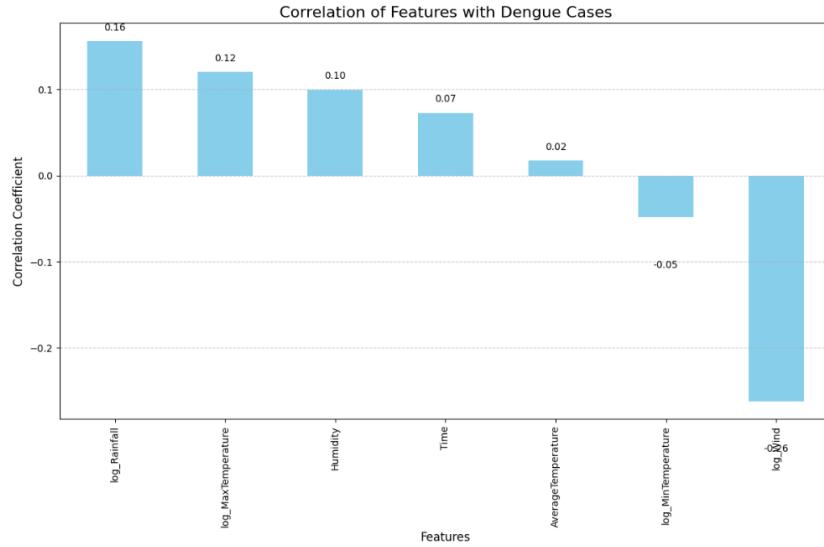


Figure 4.10: Ranking of Correlations with New Distributions

837 4.3 Outbreak Detection

838 To identify outbreaks, we calculated the outbreak threshold value using the historical
 839 mean as the endemic channel. The threshold is determined using the formula:

$$\text{Outbreak Threshold Value} = \mu + 2\sigma \quad (4.1)$$

$$= 23.744444 + 2(37.144813) \quad (4.2)$$

$$= 23.744444 + 74.289626 \quad (4.3)$$

$$= 98.03407 \quad (4.4)$$

840 where μ is the historical mean and σ is the standard deviation.

841 This result indicates that dengue cases exceeding 98 in Iloilo City can be
 842 considered an outbreak. However, it is important to note that this threshold
 843 serves only as a baseline. Additional parameters, such as the number of hospital

844 beds available in the city, must be considered to compute a more effective threshold
845 and develop an appropriate response strategy.

846 4.4 Model Training Results

847 The models were evaluated using three metrics: MSE, RMSE, and MAE. The
848 table below provides a summary and comparative analysis of each model's results
849 across these metrics, offering insights into the strengths and limitations of each
850 forecasting technique for dengue case prediction in Iloilo City. The lower values
851 of the three metrics indicate better forecasting performance. Table 4.1 shows that
852 the models performed differently on testing data. LSTM outperformed the other
853 models with the lowest RMSE, MSE, and MAE while the other three models had
854 relatively higher values for the three metrics.

Method	LSTM	Seasonal ARIMA	ARIMA	Kalman Filter	KF-LSTM
Testing MSE	285.54	1261.20	1521.48	1474.82	785.35
Testing RMSE	16.90	34.45	39.00	38.40	25.56
Testing MAE	9.45	18.73	25.80	22.33	14.55
Best Parameters	Window Size: 5 Learning Rate: 0.01 Units: 64	(2,0,2)(0,1,1)	(1,2,2)	Observation Covariance: 10.0 Transition Covariance: 0.1 × Identity	Same as LSTM

Table 4.1: Comparison of different models for dengue prediction

855 4.4.1 LSTM Model

856 The LSTM model was tuned for the following parameters: learning rate and units.
857 The hyperparameter tuning was conducted for each window size, finding the best
858 parameters for each window size. Further evaluating which window size is most
859 suitable for the prediction model, Table 4.2 shows the evaluation metrics for each
window size used in the LSTM model training.

Window Size	MSE	RMSE	MAE	R ²
5	285.54	16.90	9.45	0.83
10	334.63	18.29	9.85	0.80
20	294.85	17.17	9.35	0.83

Table 4.2: Comparison of Window Sizes

860
861 The results indicate that a window size of 5 weeks provides the most accurate
862 predictions, as evidenced by the lowest MSE and RMSE values. Furthermore, the

863 R^2 score of 0.83 indicates that 83% of the variability in the target variable (cases)
 864 is explained by the independent variables (the inputs) in the model, making it a
 865 reliable configuration overall.

866 As shown in Table 4.3, the results from time series cross-validation indicate
 867 consistent performance trends, with a window size of 5 yielding the highest average
 868 RMSE across all folds compared to the other window sizes.

Window Size	Average RMSE	Average MAE	Average R^2
5	16.69	9.06	0.79
10	17.08	10.40	0.75
20	16.93	8.75	0.81

Table 4.3: Time-Series Cross Validation Results: Comparison of Window Sizes

869 Figure 4.11 illustrates the model's performance in predicting dengue cases
 870 for each fold using a window size of 5. As shown in the plot, the training set
 871 progressively increases with each fold, mimicking a real-world scenario where more
 872 data becomes available over time for dengue prediction. Figure 4.12 demonstrates
 873 that the predicted cases closely follow the trend of the actual cases, indicating
 874 that the LSTM model successfully captures the underlying patterns in the data.
 875 It is also evident that as the fold number increases and the training set grows, the
 876 accuracy of the predictions on the test set improves. Despite the test data being
 877 unseen, the model exhibits a strong ability to generalize, suggesting it effectively
 878 leverages past observations to predict future trends.

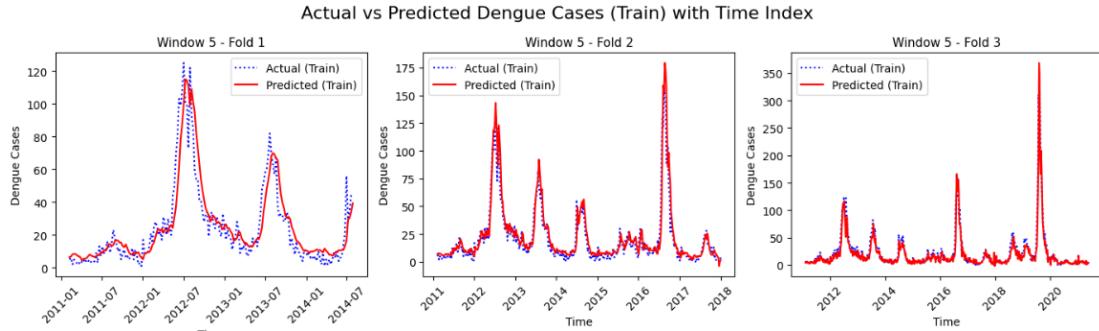


Figure 4.11: Training Folds - Window Size 5

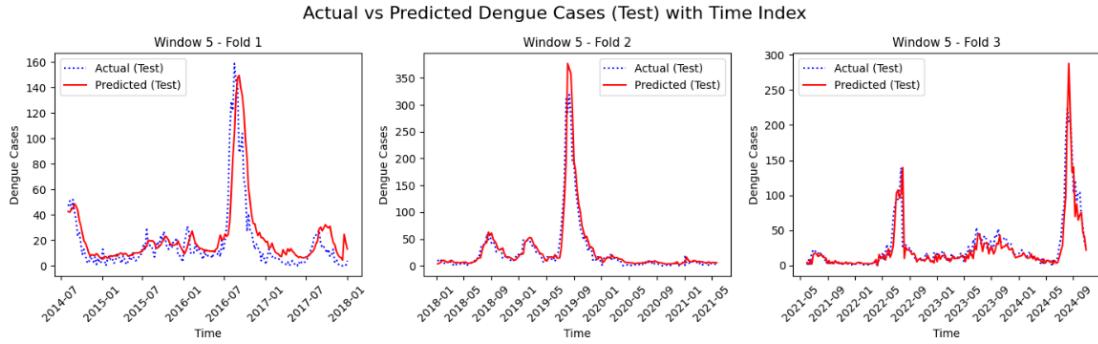


Figure 4.12: Testing Folds - Window Size 5

879 4.4.2 ARIMA Model

880 The ARIMA model was developed to capture non-seasonal trends in the data.
 881 To determine the best model configuration, grid search was used to explore vari-
 882 ous combinations of ARIMA parameters, ultimately selecting **ARIMA(1, 2, 2)**.
 883 The model was iteratively refined over **400 iterations** to ensure convergence to
 884 an optimal solution. Figure 4.13 illustrates the comparison between actual and
 885 predicted dengue cases in the test set. As shown in the plot, the ARIMA model
 886 struggled to capture the non-linear characteristics and abrupt spikes in the data.
 887 Consequently, it failed to accurately reflect the fluctuations and outbreak patterns
 888 seen in the actual case counts.

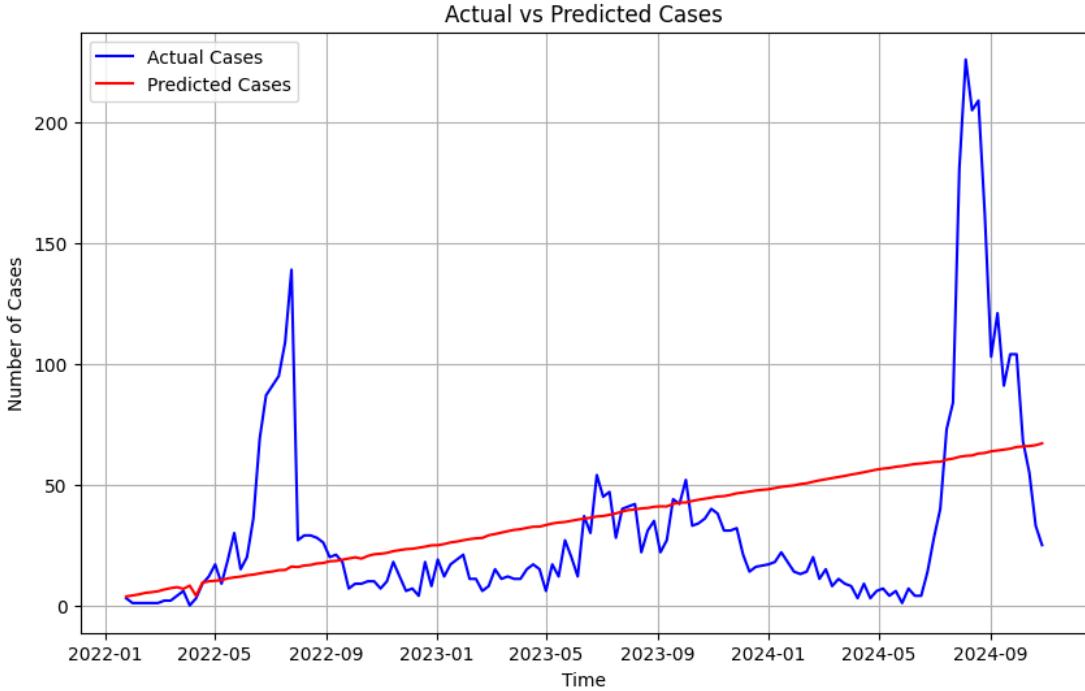


Figure 4.13: ARIMA Prediction Results for Test Set

889 The model's performance was assessed using regression metrics to evaluate its
 890 forecasting capability. The ARIMA model yielded the following error metrics:

- 891 • **MSE (Mean Squared Error):** 1521.48
- 892 • **RMSE (Root Mean Squared Error):** 39.01
- 893 • **MAE (Mean Absolute Error):** 25.80

894 4.4.3 Seasonal ARIMA (SARIMA) Model

895 To address the limitations of the ARIMA model, a Seasonal ARIMA (SARIMA)
 896 model was developed to capture both non-seasonal and seasonal variations in the
 897 data.

898 This model incorporates seasonal parameters, which were tuned using grid
 899 search to find the best configuration: **SARIMA(2, 0, 2)(0, 1, 1)[52]**. As with
 900 ARIMA, **400 iterations** were applied to ensure a robust fit. As shown in Figure
 901 4.14, the SARIMA model demonstrates a notable improvement in performance.

902 Unlike its non-seasonal counterpart, it effectively captures the general trend and
903 aligns more closely with the peaks observed in the actual dengue cases, indicating
904 its ability to model seasonal dynamics.

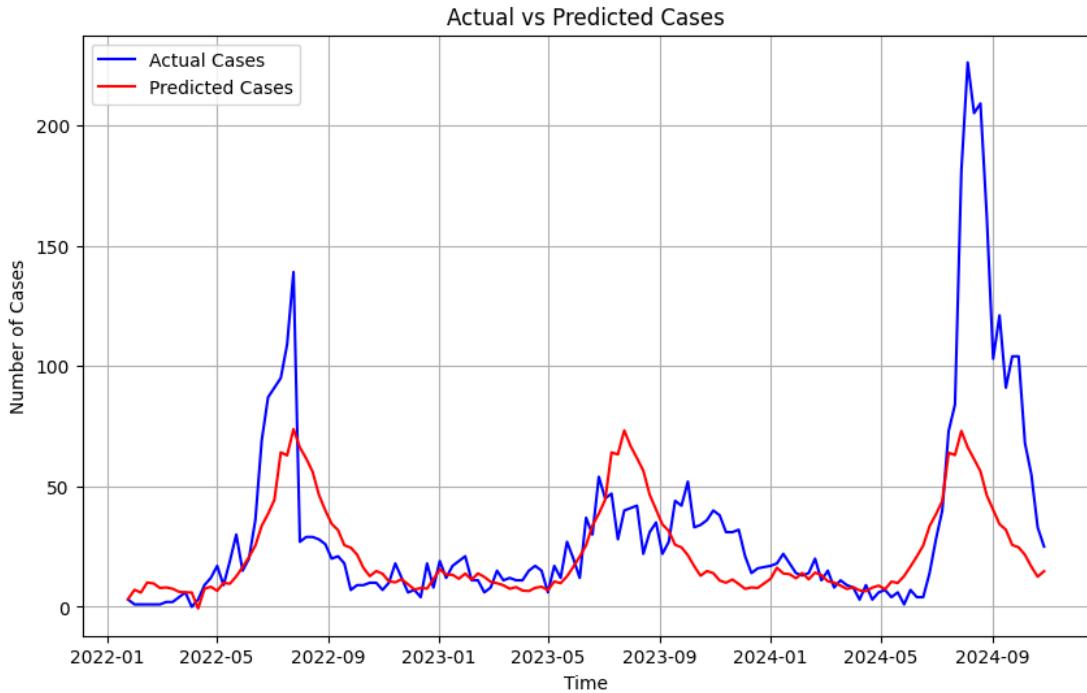


Figure 4.14: Seasonal ARIMA Prediction Results for Test Set

905 The model's performance was assessed using regression metrics to evaluate its
906 forecasting capability. The SARIMA model yielded the following error metrics:

- 907 • **MSE:** 1109.69
908 • **RMSE:** 33.31
909 • **MAE:** 18.09

910 The lower error values, when compared to the ARIMA model, highlight the
911 SARIMA model's superior capability in forecasting dengue cases. Its effective-
912 ness in capturing seasonal patterns contributed to a more accurate representation
913 of the actual cases.

914 After training the model, the SARIMA model was validated using the same
915 Time Series Cross-Validation strategy employed in the LSTM model. Table 4.4

916 presents the performance metrics for each fold, as well as the average metrics
917 across all folds. The average RMSE and MAE values were close to those obtained
918 during the initial training phase, indicating that the SARIMA model performed
919 consistently across different time segments.

Fold	MSE	RMSE	MAE
1	659.68	25.68	16.00
2	2127.49	46.12	21.30
3	996.43	31.56	18.89
Average	1261.20	34.45	18.73

Table 4.4: Comparison of SARIMA performance for each fold

920 4.4.4 Kalman Filter Model

921 Figure 4.15 shows the comparison between the actual dengue cases and the pre-
922 dicted values on the test set. As illustrated in the plot, the Kalman Filter model
923 demonstrates a moderate ability to follow the general trend of the actual data.
924 While it effectively captures some rising and falling patterns, it still struggles to
925 accurately replicate the sharp peaks and extreme values found in the real case
926 counts. This limitation is particularly noticeable during the large spikes in 2022
927 and 2024. The model's performance was evaluated using standard regression met-
928 rics. The results are as follows:

$$\text{MSE} = 1474.82, \quad \text{RMSE} = 38.40, \quad \text{MAE} = 22.34$$

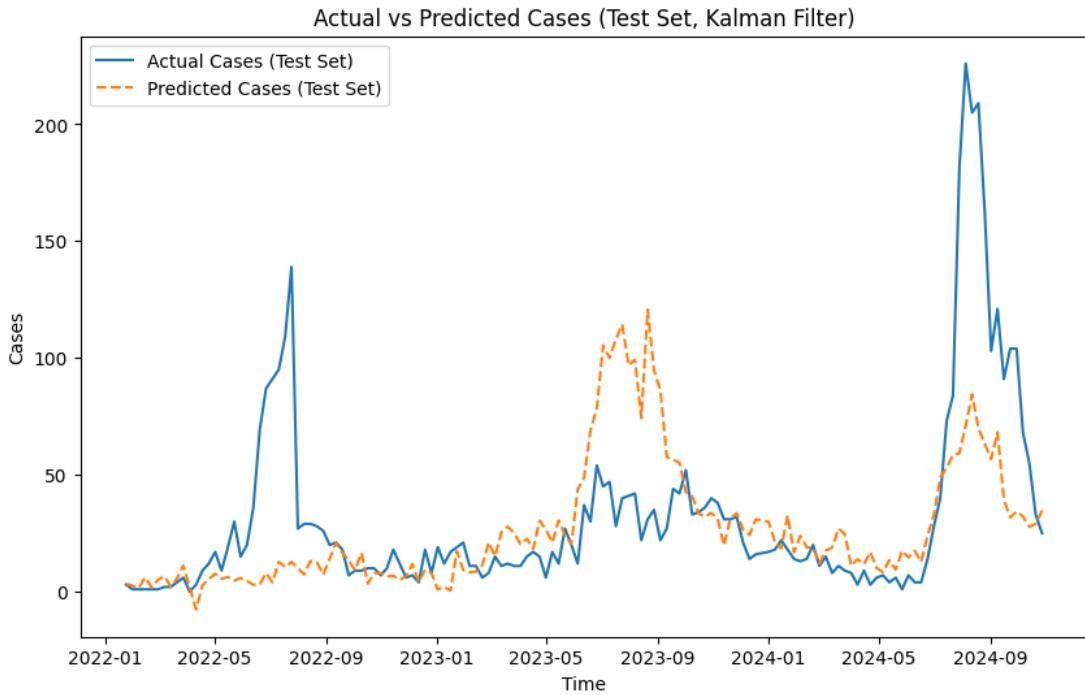


Figure 4.15: Kalman Filter Prediction Results for Test Set

929 The Kalman Filter was then combined with the LSTM model in order to see
 930 improvements in its predictions. Table 4.5 shows the metrics across three folds
 931 using the same Time Series Cross Validation Strategy employed in the previous
 932 models to see how it performed on different time segments.

Fold	MSE	RMSE	MAE
1	113.59	10.66	6.42
2	752.51	27.43	12.11
3	1489.95	38.60	25.13
Average	785.35	25.56	14.55

Table 4.5: Comparison of KF-LSTM performance for each fold

933 As can be seen in the table above, the performance of the hybrid model demon-
 934 strated improvements in all metrics as compared to just using the Kalman Filter
 935 alone.

936 4.5 Model Simulation

937 To evaluate the LSTM model's real-world forecasting ability, a simulation was
938 conducted to predict dengue cases for the year 2025. The model was trained
939 exclusively on data from 2011 to 2024, using both dengue cases and weather vari-
940 ables. Importantly, the actual dengue case values for 2025 were never included
941 during training. Instead, only the weather variables collected for 2025 were input
942 into the model to generate predictions for that year. After prediction, the fore-
943 casted dengue cases for 2025 were compared against the true observed cases to
944 assess the model's accuracy. Figure 4.16 shows that the predicted values closely
945 follow the trend, although it may overestimate the dengue cases in some weeks.

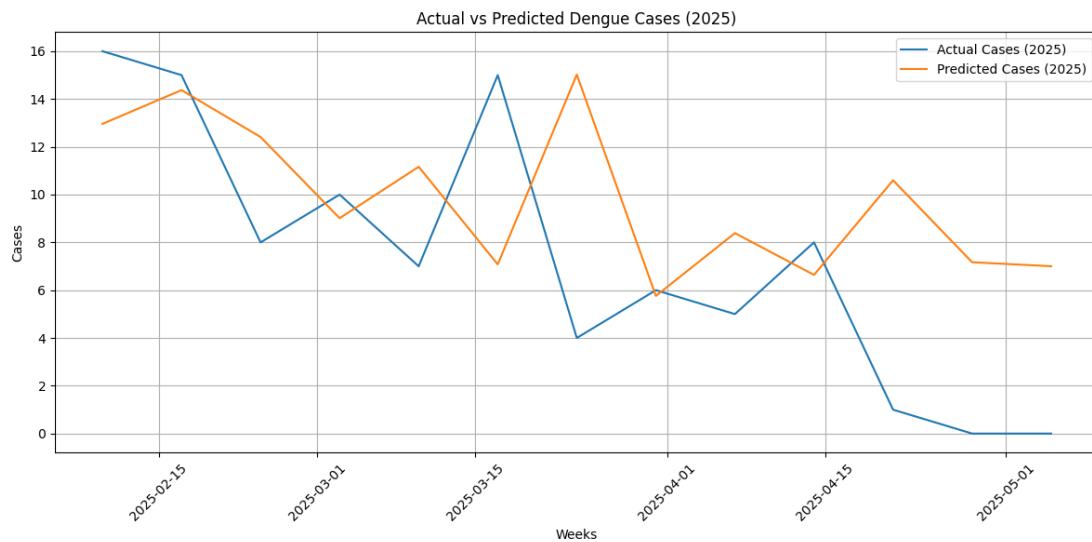


Figure 4.16: Predicted vs Actual Dengue Cases 2025

946 4.6 System Prototype

947 4.6.1 Home Page

948 The Home Page is intended for all visitors of the web application. The Analytics
949 Dashboard, which displays relevant statistics for dengue cases at a certain year
950 and location, is the primary component highlighted, as seen in Figure 4.17. This
951 component includes a combo chart that graphs the number of dengue cases and
952 deaths per week in a specific year, a choropleth map that tracks the number of

953 dengue cases per location, and various bar charts that indicate the top locations
 954 affected by dengue.

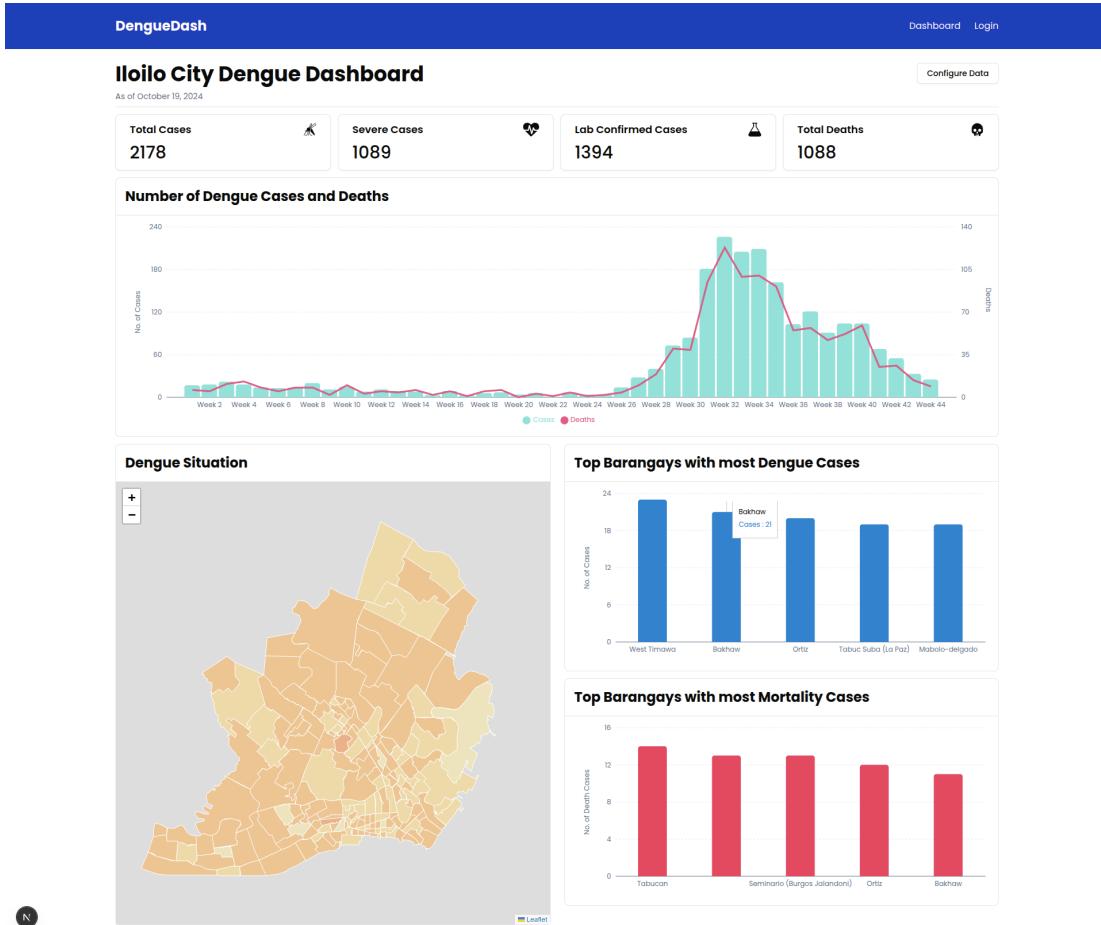


Figure 4.17: Home Page

955 4.6.2 User Registration, Login, and Authentication

956 The registration page, as shown in Figure 4.18, serves as a gateway to access the
 957 authenticated pages of the web application. Only prospected encoders can create
 958 an account since administrator accounts are only made by existing administrator
 959 accounts to protect the data's integrity in production. After registering, the
 960 "encoder account" cannot access the authorized pages yet as it needs to be veri-
 961 fied first by an administrator managing the unit the user entered. Once verified,
 962 the user can log in to the system through the page shown in Figure 4.19. Af-
 963 ter entering the correct credentials, which consist of an email and password, the

964 system uses HTTP-only cookies containing JSON Web Tokens (JWT) to prevent
965 vulnerability to Cross-site Scripting (XSS) attacks. It will then proceed to the
966 appropriate page the type of user belongs to.

The screenshot shows the 'Sign Up' page of the DengueDash application. At the top, there is a blue header bar with the text 'DengueDash' on the left and 'Dashboard Login' on the right. Below the header, the page title 'Sign Up' is centered, with the sub-instruction 'Create your account to get started' underneath it. The form consists of several input fields: 'First Name' (John), 'Middle Name (Optional)' (David), 'Last Name' (Doe), 'Sex' (Select gender), 'Email' (john@example.com), 'Region' (Select region), 'Surveillance Unit' (Select surveillance unit), 'DRU' (Select DRU), 'Password' (a field with placeholder text 'Must be at least 8 characters long'), and 'Confirm Password'. At the bottom of the form is a large black button labeled 'Create Account'. Below this button, there is a link 'Already have an account? Sign in'.

Figure 4.18: Sign Up Page

The screenshot shows the 'Welcome back!' page of the DengueDash application. At the top, there is a blue header bar with the text 'DengueDash' on the left and 'Dashboard Login' on the right. The main content area has a white background with a central box titled 'Welcome back!'. Inside this box, there are two input fields: 'Email' (placeholder 'Enter your email') and 'Password' (placeholder 'Enter your password'). Below these fields are two small checkboxes: 'Remember me' and 'Forgot password?'. At the bottom of the box is a large green button labeled 'Continue'.

Figure 4.19: Login Page

967 4.6.3 Encoder Interface

968 Case Report Form

969 Figures 4.20 and 4.21 show the digitized counterpart of the form obtained from the
970 Iloilo Provincial Epidemiology and Surveillance Unit. As the system aims to sup-
971 port expandability for future features, some fields were modified to accommodate
972 more detailed input. It is worth noting that all of the included fields adhere to the
973 latest Philippine Integrated Disease and Surveillance Response (PIDS) Dengue
974 Forms, which the referenced form was based on. By doing this, if implemented
975 on a national scale, the transition between targeted users will be easier. More-
976 over, the case form includes the patient's basic information, dengue vaccination
977 status, consultation details, laboratory results, and the outcome. On the other
978 hand, encoders can also create case records using a "bulk upload" feature that
979 makes use of a formatted CSV file template. As shown in Figure 4.22, an encoder
980 can download the template using the "Download Template" button, and insert
981 multiple records inside the file, then upload it by clicking the "Click to upload"
982 button. The web application automatically checks the file for data inconsis-
983 tencies and validation.

The screenshot shows the 'Case Report Form' page within the 'DengueDash' application. The left sidebar contains a navigation menu with 'Analytics', 'Forms' (selected), 'Data Tables', and 'Settings'. A user profile at the bottom left shows 'CN Elizabeth Thomas Ra...' and an email address. The main content area has a header 'Case Report Form' with a 'Bulk Upload' button. Below is a 'Personal Information' section divided into 'Personal Detail' and 'Clinical Status'. 'Personal Detail' includes fields for First Name, Middle Name, Last Name, Suffix, Sex (dropdown), Date of Birth (date picker), and Civil Status (dropdown). 'Clinical Status' is currently empty. The 'Address' section includes Region, Province, City, Barangay, Street, and House No. dropdowns. The 'Vaccination' section includes Date of First Vaccination and Date of Last Vaccination date pickers. A 'Next' button is at the bottom right.

Figure 4.20: First Part of Case Report Form

DengueDash

Building Your Application > Data Fetching

Case Report Form

Bulk Upload

Personal Information Clinical Status

Consultation

Date Admitted/Consulted/Seen
Pick a date

Is Admitted?
Select

Date Onset of Illness
Pick a date

Clinical Classification
Select

Laboratory Results

NS1
Pending Result Date done (NS1)
Pick a date

IgG ELISA
Pending Result Date done (IgG ELISA)
Pick a date

IgM ELISA
Pending Result Date done (IgM ELISA)
Pick a date

PCR
Pending Result Date done (PCR)
Pick a date

Outcome

Case Classification
Select Outcome
Select

Date of Death
Pick a date

Previous Submit

CN Elizabeth Thomas Ra... zlewis@example.com

N

Figure 4.21: Second Part of Case Report Form

The screenshot shows the DengueDash application interface. On the left, there's a sidebar with 'Case Report Form' selected under 'Forms'. The main area is titled 'Case Report Form' and has tabs for 'Personal Information' and 'Clinical Status'. A central modal window is open, titled 'Bulk Upload Patient Cases', with the sub-instruction 'Upload a CSV file to create multiple patient cases at once'. It features a large dashed box for file upload, a 'Click to upload' button, and a note 'CSV files only (max 5MB)'. Below this are 'Need a template?' (with a download link), 'CSV Format Requirements' (with a bulleted list of 9 items), and 'Reset' and 'Upload CSV' buttons. The background of the main form shows fields for 'First Name', 'Last Name', 'Sex', 'Date of Birth', 'Address', 'Region', 'City', and 'Street', along with a vaccination section for 'Date of First Vaccination' and 'Date of Last Vaccination'.

Figure 4.22: Bulk Upload of Cases using CSV

984 Browsing, Update, and Deletion of Records

985 Once the data generated from the case report form or the bulk upload is vali-
 986 dated, it will be assigned as a new case and can be accessed through the Dengue
 987 Reports page, as shown in Figure 4.23. The said page displays basic information
 988 about the patient related to a specific case, including their name, address, date
 989 of consultation, and clinical and case classifications. It is also worth noting that
 990 it only shows cases the user is permitted to view. For example, in a local Disease
 991 Reporting Unit (DRU) setting, the user can only access records that belong to
 992 the same DRU. On the other hand, in a consolidated surveillance unit such as a
 993 regional, provincial, or city quarter, its users can view all the records from all the
 994 DRUs that report to them. Moving forward, Figure 4.24 shows the detailed case
 995 report of the patient on a particular consultation date.

DengueDash

Building Your Application > Data Fetching

Dengue Reports

Case ID	Name	Barangay	City	Date Consulted	Clinical Classification	Case Classification	Action
25017095	Hodges, Destiny Michelle	Pale Benedicto Rizal (Manduriao)	ILOILo CITY (Capital)	2024-11-03	Severe dengue	Probable	<button>Open</button>
25017077	Cuevas, Robert Rebecca	Democracia	ILOILo CITY (Capital)	2024-11-03	With warning signs	Confirmed	<button>Open</button>
25017090	Middleton, Joseph Michael	Tanza-Esperanza	ILOILo CITY (Capital)	2024-11-01	With warning signs	Suspect	<button>Open</button>
25017089	Medina, Michael Paige	Tocas	ILOILo CITY (Capital)	2024-11-01	With warning signs	Probable	<button>Open</button>
25017081	Love, Paula Kimberly	Magsaysay	ILOILo CITY (Capital)	2024-11-01	With warning signs	Suspect	<button>Open</button>
25017073	Smith, Anna Andrea	Desamparados	ILOILo CITY (Capital)	2024-11-01	Severe dengue	Confirmed	<button>Open</button>
25017094	Morrison, Michael Sarah	El 98 Castillo (Claudio Lopez)	ILOILo CITY (Capital)	2024-10-31	Severe dengue	Probable	<button>Open</button>
25017093	Barnes, Charles Robert	Rima-Rizal	ILOILo CITY (Capital)	2024-10-31	With warning signs	Suspect	<button>Open</button>

< Previous 1 2 ... 2137 Next >

CN Iloilo City Epidemiol... ilococeu@gmail.com

Figure 4.23: Dengue Reports

The screenshot shows the DengueDash application interface. On the left is a sidebar with a navigation menu:

- Modules
 - Accounts
 - DRU
 - Analytics
 - Data Tables
 - Dengue Reports
 - Settings

Below the sidebar, the user's information is displayed: CN: Iloilo City Epidemiology, Email: iloiloepi@gmail.com.

The main content area is titled "Building Your Application > Data Fetching". It displays a "Personal Information" section with fields for Full Name (Medina, Michael Paige), Date of Birth (October 11, 1935), Sex (Male), and Civil Status (Widowed). It also shows a "Vaccination Status" section with First Dose (April 26, 2023) and Last Dose (May 31, 2020).

A "Case Record #25017089" section is shown, with "Update Case" and "Delete Case" buttons. This section includes fields for Date of Consultation (November 1, 2024), Patient Admitted? (No), Date Onset of Illness (October 23, 2024), Clinical Classification (With warning signs), and a "Case Record" section with details like NS1 (Negative, Date Done: October 27, 2024), IgG Elisa (Equivocal, Date Done: October 30, 2024), IgM Elisa (Pending Result, Date Done: N/A), PCR (Pending Result, Date Done: N/A), Case Classification (Probable, Outcome: Dead), and Date of Death (October 31, 2024).

The "Interviewer" section lists Interviewer (Daniels, Lisa Long) and DRU (Molo District Health Center).

Figure 4.24: Detailed Case Report

996 To update the case, the user can click the "Update Case" button, where a
 997 dialog will appear, and the updateable fields will be shown. It is worth noting
 998 that in this case, only fields under Laboratory Results and Outcome are included
 999 since they are the only ones that are time-based, where the result may change in
 1000 the future. After updating, a prompt will show confirming the action of the user.
 1001 Moving forward, to delete a case record, the user must click the "Delete Case"
 1002 button, and a prompt verifying the action will appear. After confirming, the case
 1003 will be deleted permanently.

DengueDash

Building Your Application > Data Fetching

Personal Information

Full Name: Medina, Michael Paige
Date of Birth: October 11, 1935

Sex: Male Civil Status: Widowed

Full Address: 995 Monique Spur, Tacas, ILOIO CITY (Capital), Iloilo

Vaccination S

First Date: April 26, 2023

Case Record #

Date of Consultation: November 1, 2024

Date Onset of illness: October 23, 2024

Laboratory Results

NS1	Date Done: n/a
IgG Elisa	Date Done: November 7th, 2024
IgM Elisa	Date Done: November 7th, 2024
PCR	Date Done: November 5th, 2024

Outcome

Case Classification: Probable	Outcome: Alive
-------------------------------	----------------

Interviewer

Interviewer: Daniels, Lisa Long DRU: Molo District Health Center

Update Case #25017095

Laboratory Results

NS1	Date Done: n/a
IgG Elisa	Date Done: November 7th, 2024
IgM Elisa	Date Done: November 7th, 2024
PCR	Date Done: November 5th, 2024

Outcome

Case Classification: Probable	Outcome: Alive
-------------------------------	----------------

Interviewer

Interviewer: Daniels, Lisa Long DRU: Molo District Health Center

Figure 4.25: Update Report Dialog

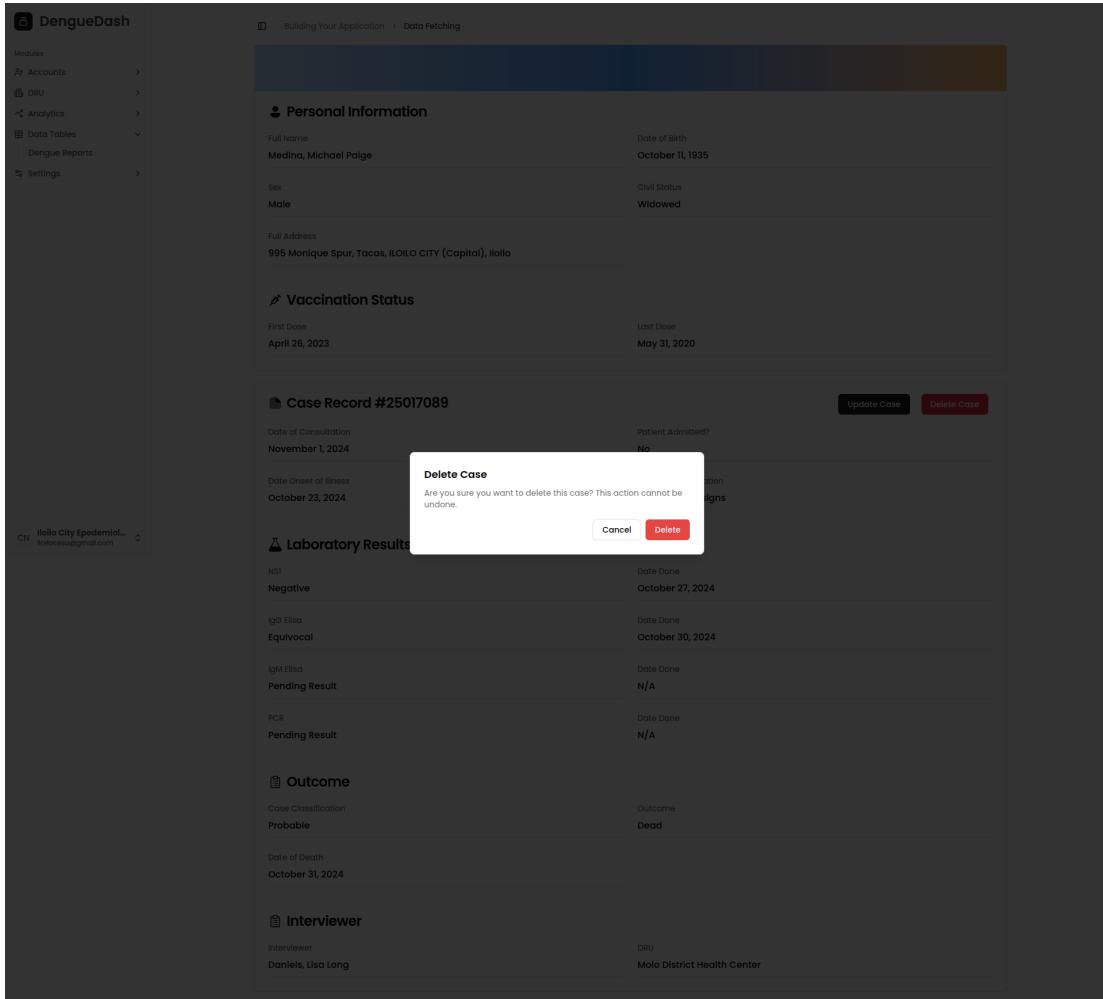


Figure 4.26: Delete Report Alert Dialog

1004 Forecasting

1005 The piece de resistance of the web application's feature is the Forecasting Page.
 1006 This is where users can forecast dengue cases for the next following weeks. To
 1007 predict, the application utilizes the exported LSTM model in a Keras format
 1008 derived from training the consolidated data from the database. It requires the
 1009 recent weekly dengue cases and weather variable data (temperature, humidity, and
 1010 rainfall) based on the window size. This allows the web application to display a line
 1011 chart with the anticipated number of dengue cases over the following four weeks.
 1012 Moving forward, the Forecasting page, as shown in Figure 4.27, introduces a user-
 1013 friendly interface that shows the current cases for the week and the predictions for
 1014 the next two weeks with a range of 90 percent to 110 percent confidence interval

1015 that is presented in a simple but organized manner. There is also a line chart
 1016 that shows the number of cases from the last 5 weeks plus the forecasted weekly
 1017 cases. In addition, the current weather data for a specific week is also shown, as
 1018 well as the forecasted weather data fetched from the OpenWeather API. Lastly,
 1019 locations where dengue cases have been reported for the current week are listed
 1020 in the Location Risk Assessment component.

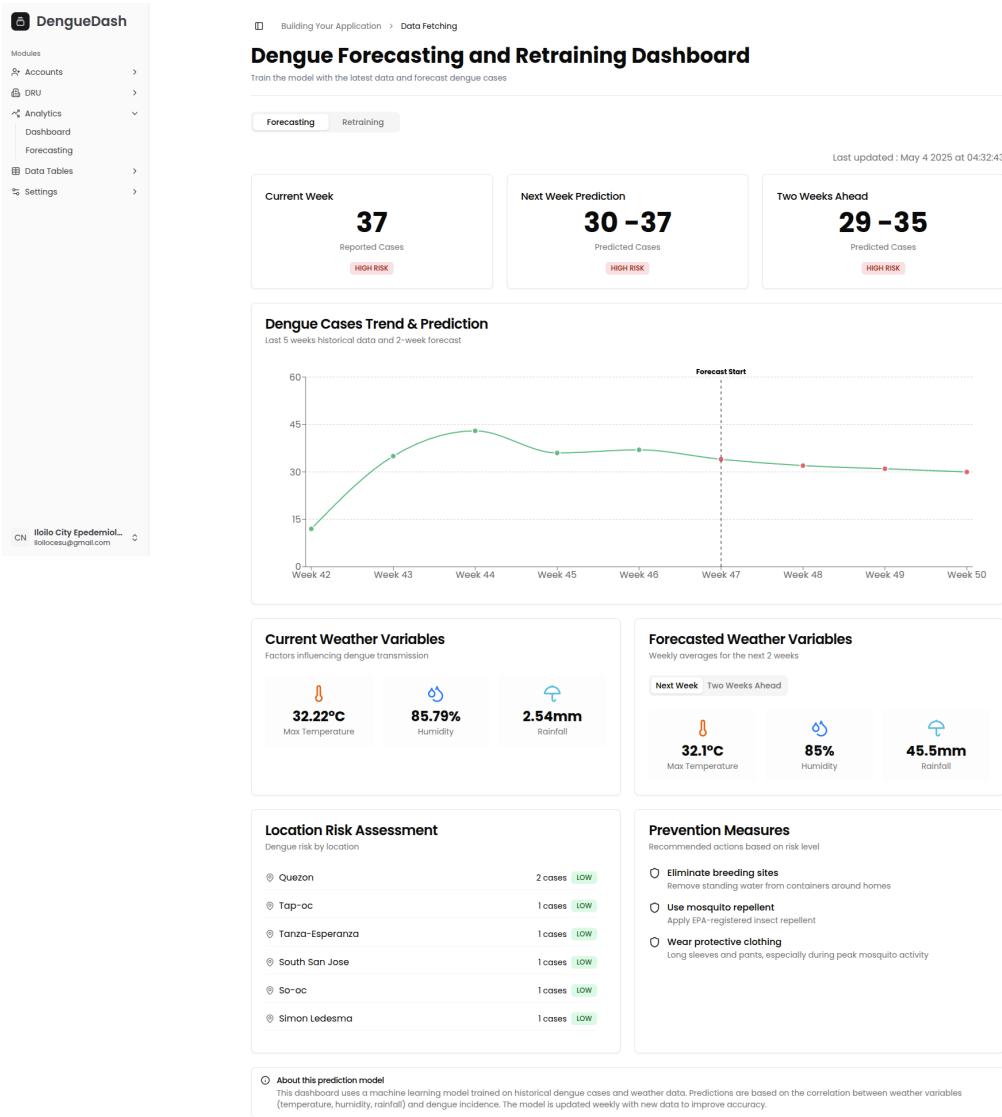


Figure 4.27: Forecasting Page

1021 4.6.4 Admin Interface

1022 Retraining

1023 With LSTM being the best-performing model among the models used in forecast-
1024 ing dengue cases, it is the model chosen to power the prediction and retraining
1025 of the consolidated data within the web application. Since the retraining process
1026 consumes a lot of processing power and requires a more advanced understanding
1027 of how it works, it was decided that the said feature should only be available
1028 to admin users. Furthermore, the retraining component in the Forecasting page
1029 includes three additional components that include the configuration of LSTM pa-
1030 rameters (Figure 4.28), the actual retraining of the consolidated data from the
1031 database (Figure 4.29), and the results of the retraining that shows the current
1032 and previous model metrics depending on the parameters entered (Figure 4.30).
1033 It is also worth noting that when trained, the model used a seeded number to
1034 promote reproducibility.

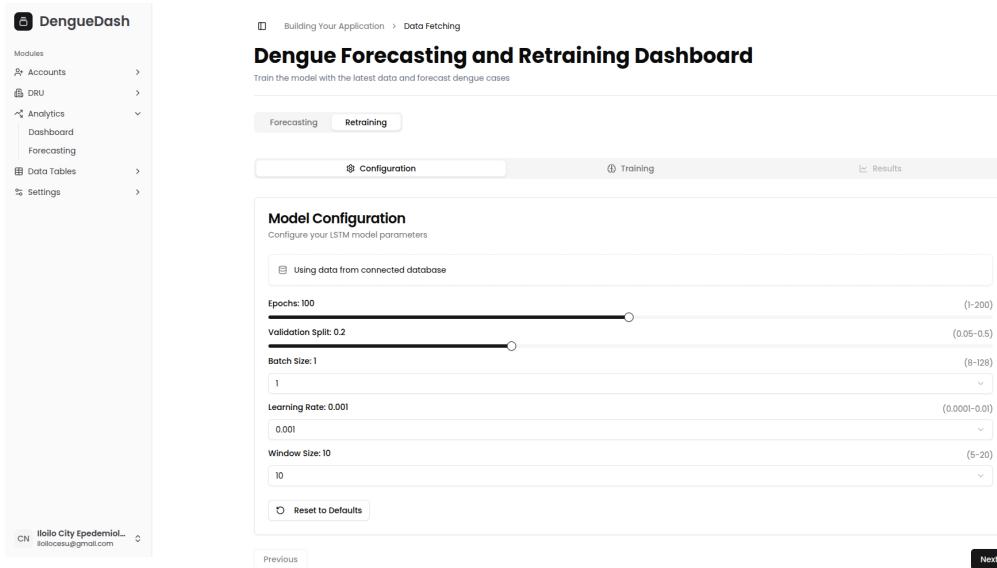


Figure 4.28: Retraining Configurations

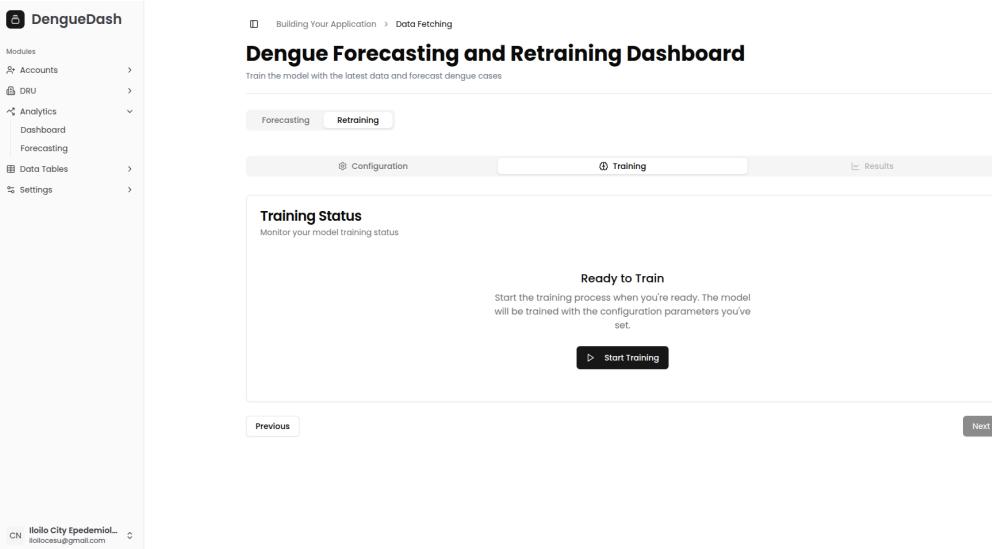


Figure 4.29: Start Retraining

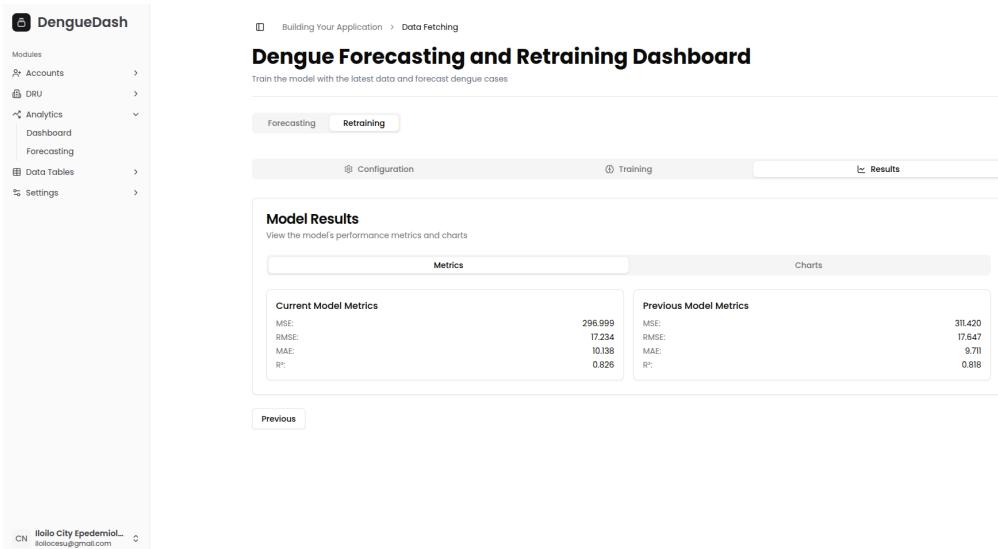


Figure 4.30: Retraining Results

1035 Managing Accounts

1036 Proper management of accounts is important to protect the integrity and confi-
1037 dentiality of data. Thus, it is crucial for administrators to track their users and
1038 control the flow of information. As discussed in the user registration of encoders,
1039 admin users from a specific DRU or surveillance have the power to grant them ac-
1040 cess to the web application. Figure 4.32 illustrates the interface for this scenario,
1041 as the admins can approve or reject their applications. Once approved, these users
1042 can access the features given to encoders and may be promoted to have admin-
1043 istrative access, as shown in Figure 4.33. When deleting an account, the user's
1044 email will be blacklisted and illegible to use when creating another account, and
1045 all the cases reported by this user will be soft-deleted. The same figure also shows
1046 the expanded details of the user, which include personal information and brief
1047 activity details within the system.

The screenshot shows the DengueDash application interface. On the left is a sidebar with a navigation menu:

- Modules
 - Accounts (selected)
 - DRU
 - Analytics
 - Data Tables
 - Settings

The main content area is titled "Manage Accounts". At the top, there are three buttons: "Verified" (highlighted), "Pending", and "Blacklisted". Below this is a table with the following data:

Name	Email	Role	Sex	Actions
Cheryl Hernandez King	omarpatterson@example.net	Encoder	Female	<button>Open</button>

At the bottom left of the main area, there is a small user profile icon with the text "CN illo City Epidemiol..." and "illocessu@gmail.com".

Figure 4.31: List of Verified Accounts

Name	Email	Sex	Created At	Actions
John David Doe	testereee@example.gov.ph	Male	2025-04-26	<button>Approve</button> <button>Delete</button>

Figure 4.32: List of Pending Accounts

Figure 4.33: Account Details

1048 Managing DRUs

1049 Unlike the registration of encoder accounts, the creation of Disease Reporting
 1050 Units can only be done within the web application, and the user performing the
 1051 creation must be an administrator of a surveillance unit. Figure 4.34 presents the

1052 fields the admin user must fill out, and once completed, the new entry will show
 1053 as being managed by that unit, as shown in Figure 4.35. Figure 4.36, on the other
 1054 hand, shows the details provided in the registration form as well as its creation
 1055 details. There is also an option to delete the DRU, and when invoked, all the
 1056 accounts being managed by it, and the cases reported under those accounts will
 1057 be soft-deleted.

The screenshot shows the 'DengueDash' application interface. On the left, there's a sidebar with 'Modules' listed: Accounts, DRU (selected), Analytics, Data Tables, and Settings. The main content area is titled 'Register Disease Reporting Unit' with the sub-instruction 'Add a new Disease Reporting Unit to the surveillance system.' It contains several input fields:

- Name:** Enter DRU name (The official name of the Disease Reporting Unit).
- Address Information:**
 - Region:** Select Region (dropdown menu).
 - Province:** Select Province (dropdown menu).
 - City/Municipality:** Select City/Municipality (dropdown menu).
 - Barangay:** Select Barangay (dropdown menu).
- Street Address:** House/Building No., Street Name.
- Email:** example@health.gov
- Contact Number:** +63 912 345 6789
- DRU Type:** Select DRU type (dropdown menu).
- Description:** The category that best describes this reporting unit.

A large black 'Register DRU' button is at the bottom right of the form.

Figure 4.34: Disease Reporting Unit Registration

The screenshot shows the 'DengueDash' application interface. On the left, there's a sidebar with 'Modules' listed: Accounts, DRU (selected), Analytics, Data Tables, and Settings. The main content area is titled 'Manage Disease Reporting Units' with the sub-instruction 'View and manage Disease Reporting Units.' It displays a table of existing DRUs:

DRU Name	Email	Action
Molo District Health Center	molodistricthealth@gmail.com	<button>Open</button>
Jaro Health Center	jarohealth@gmail.com	<button>Open</button>
Saint Paul's Hospital	saintpaul@gmail.com	<button>Open</button>

A small watermark 'Iloilo City Epidemiol...' is visible at the bottom left of the screen.

Figure 4.35: List of Disease Reporting Units

Figure 4.36: Disease Reporting Unit details

1058 4.7 User Testing

1059 To evaluate the usability of the system, the System Usability Scale (SUS) was
 1060 utilized. SUS is a Likert-scale-based questionnaire comprising 10 items that are
 1061 critical to assessing system usability. A total of five participants completed the sur-
 1062vey. Their responses were processed following the step-by-step calculation method
 1063 adopted from (Babich, 2015). The resulting usability scores for each participant
 1064 are shown in Table 4.6.

Participant	Usability Score
1	95.0
2	90.0
3	85.0
4	87.5
5	85.0
Average	88.5

Table 4.6: Computed System Usability Scores per Participant

1065 The average System Usability Scale (SUS) score across systems is typically
 1066 68 (Babich, 2015). In this testing, the system achieved an average SUS score
 1067 of 88.5, indicating a highly positive user experience. This score suggests that
 1068 participants found the system not only enjoyable to use but also intuitive enough

¹⁰⁶⁹ to recommend to others. Furthermore, it demonstrates that the system is suitable
¹⁰⁷⁰ for real-world applications without presenting significant complexity for first-time
¹⁰⁷¹ users.

1072 **Chapter 5**

1073 **Conclusion**

1074 **Revolutionizing Dengue Surveillance: The Rise of AI-Driven Forecasting**

1076 The development of DengueWatch marks a transformative leap forward in public health technology, providing Iloilo City with a centralized system to combat one of the most persistent mosquito-borne diseases. Previously, data was recorded manually on paper, making tracking and analysis slow and error-prone. DengueWatch digitizes this process, enabling faster, more accurate monitoring. More than an academic project, DengueWatch serves as a practical solution aimed at shifting the approach from reactive outbreak response to proactive prevention. By combining deep learning models with real-time climate data integration, the system achieves a level of accuracy and usability that makes it viable for real-world deployment.

1086 At the heart of DengueWatch is a Long Short-Term Memory (LSTM) neural network, which outperformed traditional forecasting models such as ARIMA and Kalman Filter. The LSTM model achieved a Root Mean Square Error (RMSE) of 16.90, compared to 39.00 and 38.40 for ARIMA and Kalman, respectively—demonstrating a substantial improvement in predictive capability. This advantage stems from the LSTM’s ability to capture long-term dependencies and model nonlinear relationships between environmental factors and disease patterns.

1093 The analysis also revealed that climate indicators, particularly rainfall and humidity, play a significant role in dengue outbreaks, typically leading to a surge in cases three to five weeks after anomalies are detected. By incorporating these lagged effects, DengueWatch achieved an explanatory power of 83% ($R^2 = 0.83$), offering a game-changing advantage for early intervention and resource allocation.

1098 Usability testing further underscored DengueWatch's readiness for real-world
1099 deployment. The system achieved an average System Usability Scale (SUS) score
1100 of 88.5, significantly above the industry benchmark of 68. This indicates that
1101 users found the system intuitive, efficient, and suitable for operational use in
1102 public health contexts. Key features such as a user-friendly dashboard, a two-week
1103 forecasting window aligned with mosquito life cycles, and automated outbreak
1104 alerts ensure that the system supports timely, effective responses.

1105 Beyond its immediate application in Iloilo City, the framework behind Dengue-
1106 Watch holds the potential for broader impact. With minor adaptations, it can
1107 be scaled nationally through integration with Department of Health surveillance
1108 systems.

1109 DengueWatch exemplifies how deep learning can bridge the gap between data
1110 science and life-saving interventions. It empowers health workers to act preemp-
1111 tively, policymakers to allocate resources strategically, and communities to en-
1112 gage in early preventive measures. As climate change accelerates the spread of
1113 vector-borne diseases, systems like DengueWatch will become indispensable in
1114 safeguarding public health. This system not only demonstrates the power of AI
1115 in epidemiological forecasting but also lays the foundation for a smarter, more
1116 resilient approach to combating infectious diseases in the years ahead.

¹¹¹⁷ References

- ¹¹¹⁸ About GitHub and Git - GitHub Docs. (n.d.). Retrieved from
¹¹¹⁹ <https://docs.github.com/en/get-started/start-your-journey/about-github-and-git>
- ¹¹²⁰ Ahmadini, A. A. H., Naeem, M., Aamir, M., Dewan, R., Alshqaq, S. S. A., & Mashwani, W. K. (2021). Analysis and forecast of the number of deaths, recovered cases, and confirmed cases from covid-19 for the top four affected countries using kalman filter. *Frontiers in Physics*, *9*, 629320.
- ¹¹²¹ Arroyo-Marioli, F., Bullano, F., Kucinskas, S., & Rondón-Moreno, C. (2021). Tracking r of covid-19: A new real-time estimation using the kalman filter. *PloS one*, *16*(1), e0244474.
- ¹¹²² Babich, N. (2015). *How to use the system usability scale (sus) to evaluate the usability of your website*. Usability Geek. Retrieved from <https://usabilitygeek.com/how-to-use-the-system-usability-scale-sus-to-evaluate-the-usability-of-your-website/> (Accessed: 2025-04-26)
- ¹¹²³ Bobbitt, Z. (2021, November 17). *The five assumptions for pearson correlation*. Retrieved from <https://www.statology.org/pearson-correlation-assumptions/> (Statology)
- ¹¹²⁴ Bosano, R. (2023). *Who: Ph most affected by dengue in western pacific*. Retrieved
¹¹²⁵ Use the date of access, from <https://news.abs-cbn.com/spotlight/12/22/23/who-ph-most-affected-by-dengue-in-western-pacific>
- ¹¹²⁶ Brady, O. J., Smith, D. L., Scott, T. W., & Hay, S. I. (2015). Dengue disease outbreak definitions are implicitly variable. *Epidemics*, *11*, 92–102.
- ¹¹²⁷ Bravo, L., Roque, V. G., Brett, J., Dizon, R., & L'Azou, M. (2014). Epidemiology of dengue disease in the philippines (2000–2011): a systematic literature review. *PLoS neglected tropical diseases*, *8*(11), e3027.
- ¹¹²⁸ Carvajal, T. M., Viacrusis, K. M., Hernandez, L. F. T., Ho, H. T., Amalin, D. M., & Watanabe, K. (2018). Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan manila, philippines. *BMC infectious diseases*, *18*, 1–15.
- ¹¹²⁹ Chart.js. (n.d.). Retrieved from <https://www.chartjs.org/>

- 1149 Cheong, Y. L., Burkart, K., Leitão, P. J., & Lakes, T. (2013). Assessing weather
1150 effects on dengue disease in malaysian. *International journal of environmental
1151 research and public health*, 10(12), 6319–6334.
- 1152 Christie, T. (n.d.). *Home - Django REST framework*. Retrieved from <https://www.djangoproject-rest-framework.org/>
- 1153 Colón-González, F. J., Fezzi, C., Lake, I. R., & Hunter, P. R. (2013). The effects
1154 of weather and climate change on dengue. *PLoS neglected tropical diseases*,
1155 7(11), e2503.
- 1156 Hemisphere, N. (2015). Update on the dengue situation in the western pacific
1157 region. *Update*.
- 1158 Hii, Y. L., Zhu, H., Ng, N., Ng, L. C., & Rocklöv, J. (2012). Forecast of dengue
1159 incidence using temperature and rainfall. *PLoS neglected tropical diseases*,
1160 6(11), e1908.
- 1161 Joel, C. (2021, 10). *6 reasons to use Tailwind over traditional CSS*. Retrieved
1162 from <https://dev.to/charliejoel/6-reasons-to-use-tailwind-over-traditional-css-1nc3>
- 1163 Leaflet — an open-source JavaScript library for interactive maps. (n.d.). Retrieved
1164 from <https://leafletjs.com/>
- 1165 Lena, P. (2024). *Iloilo beefs up efforts amid hike in dengue cases*. Retrieved Use
1166 the date of access, from <https://www.pna.gov.ph/articles/1231208>
- 1167 Li, X., Feng, S., Hou, N., Li, H., Zhang, S., Jian, Z., & Zi, Q. (2022). Applications
1168 of kalman filtering in time series prediction. In *International conference on
1169 intelligent robotics and applications* (pp. 520–531).
- 1170 Ligue, K. D. B., & Ligue, K. J. B. (2022). Deep learning approach to forecasting
1171 dengue cases in davao city using long short-term memory (lstm). *Philippine
1172 Journal of Science*, 151(3).
- 1173 Perla. (2024). *Iloilo beefs up efforts amid hike in dengue cases*. Retrieved Use
1174 the date of access, from <https://www.pna.gov.ph/articles/1231208>
- 1175 RabDashDC. (2024). *Rabdash dc*. Retrieved Use the date of access, from <https://rabadash.com>
- 1176 Runge-Ranzinger, S., Kroeger, A., Olliaro, P., McCall, P. J., Sánchez Tejeda, G.,
1177 Lloyd, L. S., ... Coelho, G. (2016). Dengue contingency planning: from
1178 research to policy and practice. *PLoS neglected tropical diseases*, 10(9),
1179 e0004916.
- 1180 Shadcn. (n.d.). *Introduction*. Retrieved from <https://ui.shadcn.com/docs>
- 1181 *Tailwind CSS - Rapidly build modern websites without ever leaving your HTML*.
1182 (n.d.). Retrieved from <https://tailwindcss.com/>
- 1183 Watts, D. M., Burke, D. S., Harrison, B. A., Whitmire, R. E., & Nisalak, A.
1184 (2020). Effect of temperature on the transmission of dengue virus by aedes
1185 aegypti. *The American Journal of Tropical Medicine and Hygiene*, 36(1),
1186 143–152.
- 1187 1188 1189 1190 *What is Postman? Postman API Platform*. (n.d.). Retrieved from <https://>

- 1191 www.postman.com/product/what-is-postman/
- 1192 WHO. (2023). *Dengue - global situation*. Retrieved Use the date of ac-
1193 cess, from <https://www.who.int/emergencies/diseases-outbreak-news/>
1194 [item/2023-DON498](#)
- 1195 WHO. (2024). *Dengue and severe dengue*. Retrieved Use the date
1196 of access, from <https://www.who.int/news-room/fact-sheets/detail/>
1197 [dengue-and-severe-dengue](#)
- 1198 Why Visual Studio Code? (2021, 11). Retrieved from <https://code.visualstudio.com/docs/editor/whyvscode>
- 1199 World Health Organization (WHO). (2018). Dengue and severe dengue in the
1200 philippines. *WHO Dengue Factsheet*. (Available at: <https://www.who.int>)
- 1201 Zhou, S., & Malani, P. (2024). What is dengue? *Jama*, 332(10), 850–850.
- 1202 Zod. (n.d.). *TypeScript-first schema validation with static type inference*. Re-
1203 trieved from <https://zod.dev/?id=introduction>

¹²⁰⁶ **Appendix A**

¹²⁰⁷ **Data Collection Snippets**

Time	Rainfall	MaxTemperature	AverageTemperature	MinTemperature	Wind	Humidity	Cases
2011-w1	9.938571429	29.4444	25.88889	23.8889	11.39	86.24285714	5
2011-w2	8.587142857	30	26.70555556	24.44444444	7.32	88.02857143	4
2011-w3	5.338571429	30	26.61666667	25	7.55	84.02857143	2
2011-w4	5.41	30.55555556	26.48333333	20.55555556	10.67	80.97142857	7
2011-w5	2.914285714	28.33333333	25.28333333	18.65	11.01	74.88571429	2
2011-w6	3.44	31.11111111	27.53333333	25	3.69	85.27142857	5
2011-w7	1.444285714	30	27.06111111	25	6.69	82.1	6
2011-w8	0.7685714286	30.55555556	27.03333333	25	7.82	76.18571429	2
2011-w9	4.064285714	30.55555556	26.75	24.44444444	6.64	84.1	5
2011-w10	1.252857143	31.11111111	27.7	25.55555556	7.92	80.18571429	4
2011-w11	7.068571429	28.33333333	26.03888889	23.88888889	10.35	85.7	2
2011-w12	3.567142857	30.55555556	26.4	24.44444444	9.57	87.74285714	3
2011-w13	2.745714286	30	26.61666667	25	10.05	86.25714286	2
2011-w14	0.09285714286	30	27.44444444	25	8.25	76.32857143	4
2011-w15	1.008571429	31.11111111	26.89444444	22.77777778	7.93	80.64285714	5
2011-w16	0.8771428571	31.66666667	28.43333333	24.44444444	4.76	80.35714286	3
2011-w17	2.15	32.77777778	28.87222222	25.55555556	5.54	81.77142857	5
2011-w18	3.062857143	32.77777778	28.62777778	25	5.15	81.9	4
2011-w19	9.828571429	32.77777778	27.97222222	25	4.18	85.77142857	5
2011-w20	2.848571429	33.33333333	29.71111111	25.55555556	3.77	79.65714286	3
2011-w21	11.74142857	32.77777778	28.51111111	25	4.28	81.81428571	3
2011-w22	14.52857143	32.77777778	28.47222222	25.55555556	4.04	82.45714286	4
2011-w23	11.51714286	32.22222222	28.26111111	24.44444444	3.91	84.24285714	9
2011-w24	8.784285714	32.22222222	27.71111111	24.44444444	3.87	84.7	3
2011-w25	10.43714286	32.77777778	27.47777778	24.44444444	4.46	86.52857143	13
2011-w26	6.702857143	32.22222222	28.39444444	25	3.94	81.87142857	7
2011-w27	14.48	32.22222222	27.66666667	25	4.19	85.42857143	13
2011-w28	3.32	32.22222222	27.49444444	19.68333333	6.69	80.78571429	9
2011-w29	17.33714286	32.77777778	27.95555556	25	3.57	80.05714286	10
2011-w30	22.77	31.66666667	26.63888889	24.44444444	5.13	88.01428571	16
2011-w31	9.875714286	32.77777778	27.81111111	25	4.66	83.11428571	13

Figure A.1: Snippet of Consolidated Data



Figure A.2: Snippet of Weather Data Collection

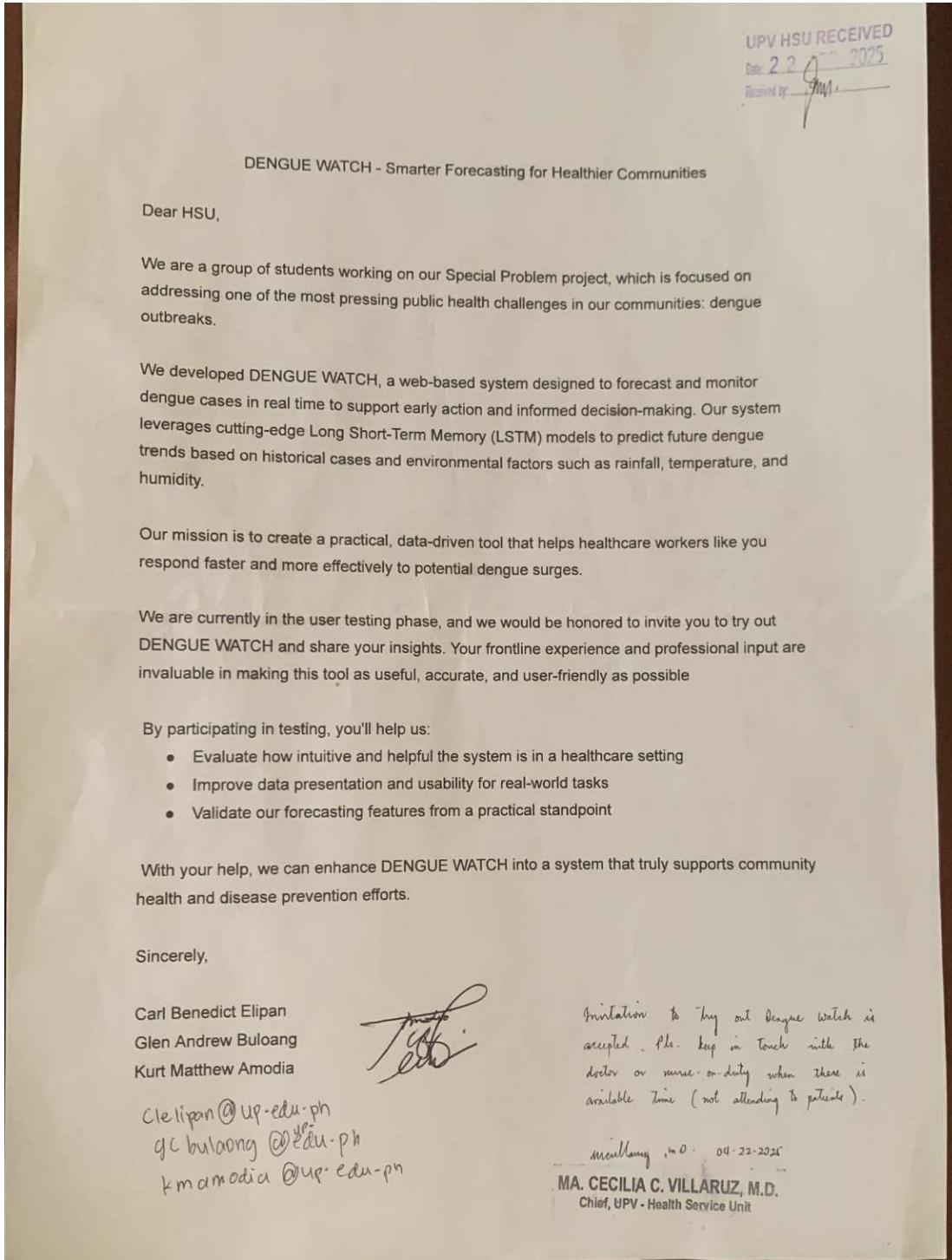


Figure A.3: Letter of Approval for User Testing in UPV HSU

Please enter your participant number: 2

System Usability Scale (SUS)

This is a standard questionnaire that measures the overall usability of a system. Please select the answer that best expresses how you feel about each statement after using the system today.

	Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1. I think I would like to use DengueWatch frequently to monitor and forecast dengue cases.	<input type="checkbox"/>				
2. I found DengueWatch unnecessarily complex when entering or analyzing data.	<input type="checkbox"/>				
3. I thought DengueWatch was easy to use, even for someone unfamiliar with AI forecasting tools.	<input type="checkbox"/>				
4. I think I would need the support of an IT or technical person to use all the features of DengueWatch.	<input type="checkbox"/>				
5. I found the visualizations (charts, maps, graphs) in DengueWatch to be well-integrated and informative.	<input type="checkbox"/>				
6. I thought there was too much inconsistency in how different parts of the DengueWatch system behaved.	<input type="checkbox"/>				
7. I believe most public health workers would learn to use DengueWatch very quickly.	<input type="checkbox"/>				
8. I found DengueWatch very cumbersome to use when entering weekly case reports or viewing forecasts.	<input type="checkbox"/>				
9. I felt confident in using DengueWatch to get the information I needed for decision-making.	<input type="checkbox"/>				
10. I needed to learn a lot of things before I could get going with DengueWatch.	<input type="checkbox"/>				

How likely are you to recommend this website to others? (please circle your answer)

Not at all likely 0 1 2 3 4 5 6 7 8 9 10 Extremely likely

Please provide important comments about the website:

Figure A.4: System Usability Score Questionnaire

₁₂₀₈ **Appendix B**

₁₂₀₉ **Resource Persons**

₁₂₁₀ **Mr. Firstname1 Lastname1**

₁₂₁₁ Role1

₁₂₁₂ Affiliation1

₁₂₁₃ emailaddr1@domain.com

₁₂₁₄ **Ms. Firstname2 Lastname2**

₁₂₁₅ Role2

₁₂₁₆ Affiliation2

₁₂₁₇ emailaddr2@domain.net

₁₂₁₈