Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 21s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Forecasting Analysis of Dengue Cases with Geospatial Mapping Using SARIMAX Algorithm

John Michael Turija[1], Angele E. Gonzales[1], Juswyn C. Espinosa[1], Glaiza R. Presilda[1], John B. Cabilin[1], Jr., Darlene Genesis P. Arcilla[1], Reymon M. Santiañez[*1]

[1]School of Technology and Computer Studies, Biliran Province State University, Naval, Biliran, Philippines.

***Corresponding Author:**

Reymon M. Santiañez

Email ID: reymonsantianez@bipsu.edu.ph

Cite this paper as: John Michael Turija, Angele E. Gonzales, Juswyn C. Espinosa, Glaiza R. Presilda, John B. Cabilin, Jr., Darlene Genesis P. Arcilla, Reymon M. Santiañez, (2025). Forecasting Analysis of Dengue Cases with Geospatial Mapping Using SARIMAX Algorithm. *Journal of Neonatal Surgery*, 14 (21s), 256-268.

ABSTRACT

Dengue fever remains a significant public health concern in many parts of the world and a considerable challenge to health authorities regarding timely intervention and resource allocation. Forecasting of dengue cases can be crucial in effective disease management and control. The study proposed an innovative method to predict the spread of dengue cases by combining geospatial mapping with the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) algorithm. The study utilizes geospatial data, demographic factors, and historical dengue case records to build a comprehensive forecasting model. The SARIMAX algorithm integrates the temporal patterns of dengue incidence with relevant exogenous variables. The SARIMAX algorithm with geospatial mapping outperforms the traditional methods in forecasting dengue cases. Integrating spatial information allows for better identification of disease hotspots and potential risk areas, facilitating targeted intervention strategies. Incorporating exogenous variables enhances the model's accuracy and provides a more comprehensive understanding of the factors influencing dengue transmission dynamics. The developed application presents a promising approach to forecasting dengue cases, contributing to improved disease surveillance and public health management. The developed web application can assist health authorities in making informed decisions, allocating resources effectively, and implementing timely preventive measures to combat the spread of dengue fever.

Keyword: *Dengue Cases, Forecasting, Geospatial Mapping, Sarimax Algorithm*

1. INTRODUCTION

Technology is changing at a rapid pace in unpredictable ways. The scale of their impact is far-reaching. ICT's rapid development affected all branches, led to evolution, and influenced people's worldviews and desires. The importance of ICT extends beyond identifying, tracing, understanding, managing, treating, and perceiving pandemics [1].

Dengue is a mosquito-borne viral infection caused by the Dengue Virus (DENV), which can result in serious and potentially life-threatening complications. It has steadily grown for decades, putting almost half of humans at risk. It was labeled the most significant global health problem in many tropical countries, including the Philippines. Dengue fever (DF) is caused by four closely related serotypes: DENV1, DENV2, DENV3, and DEN44. The virus is transmitted to humans by the bites of infected female Aedes aegypti and Aedes albopictus mosquitoes [2]. They are among the most significant vectors since they are the leading indirect source of sickness and mortality in humans, more than any other group of species, and there is no specific treatment for the illness.

Indeed, because no specific treatment or vaccine is available, the only solution to prevent the disease is a control strategy, which requires the identification of risk areas and risk periods. According to Jain and Sharma [3] several campaigns were conducted to lessen the incidence and mortality of dengue, but there is a lack of local surveillance and monitoring programs.

According to Herbuela et al. [4], early detection and effective control of epidemics depend on appropriate surveillance methods. In the Philippines, the approach to surveillance is largely passive, relying on notifications from local health centers, municipal or city health offices, hospitals, clinics, and quarantine units. This method often restricts reporting cases diagnosed clinically but lacks laboratory confirmation [5]. Madewell et al. [6] stated that passive surveillance is a cost-effective approach and serves as the foundation of dengue monitoring, it often leads to underreporting, particularly for cases that do not require hospitalization.

Dengue surveillance is vital to strategies for prevention and control, and strengthening surveillance is significant to the global response to emerging infectious diseases [7]. Salem [8] a well-functioning surveillance of dengue cases is significant in assessing risks and responding to outbreaks. It helps track disease patterns over time and evaluates the effectiveness of health programs, ensuring timely and informed decision-making. Disease surveillance is a vital aspect of health geography, helping to analyze how diseases spread across different areas and identify emerging health concerns. A deep understanding of this process is essential for developing effective strategies to address public health challenges [9].

Although there is an existing study about forecasting analysis of dengue cases, this study forecasts future dengue cases weekly to monitor the track of dengue cases in Biliran Province easily. Considering the lack of surveillance and monitoring systems here in the Philippines, we propose the use of geospatial data analysis with forecasting analysis using SARIMAX algorithm based on the confirmed number of cases by medical institutions in the province in every municipality. It utilizes Geographic Information Systems (GIS) to identify, assess and visualize potential risk factors involved in disease transmission. It also uses forecasting analysis which is important in the health sector due to its capability to anticipate the spread of dengue.

With the alarming incidence of dengue in Biliran Province, the ability of the developed system will help forecast and visualize dengue cases in the province is necessary. This approach could help communities and policymakers by providing strategic information without requiring extensive capacity or resources. Hence, this study aims to forecast possible dengue outbreaks in Biliran using a web-based geospatial analysis system with SARIMAX forecasting that generates and allows the input of a newly reported number of cases daily to update the forecasting patterns.

1.1 Objective of the Study

This study aimed to design and develop a forecasting analysis of dengue cases using SARIMAX algorithm.

Specifically seeks to:

1. Collect data on the number of dengue cases in Biliran Province.
2. Create a model using SARIMAX algorithm to forecast dengue cases in the province.
3. Develop a website application to forecast dengue cases and visualize the risk of dengue in Biliran Province.

2. METHODOLOGY

The proponents use the standard software development life cycle (SDLC), specifically the agile method, to describe and present solutions to the problems identified in this study. The project was divided into phases and required constant collaboration with the target users. The agile method's core principles are defining target users and identifying the scope of the problem, opportunities, and value to be addressed [10].

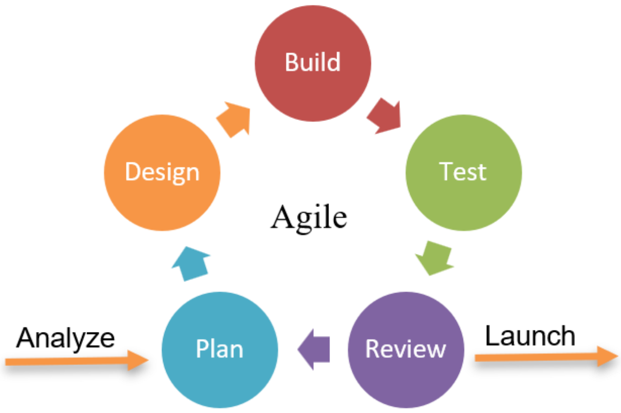

Fig. 1: Agile Model (SDLC)

Figure 1 illustrates the Agile Model (Software Development and Life Cycle (SDLC)). Where Planning, Requirements Analysis, Design, Coding, and Testing are present in the Agile process.

Analyze. The proponents define the objectives and scope of the software system to be developed. For this study, this phase plays an important role in ensuring that the system meets the intended research goals and provides accurate, data-driven insights into dengue case forecasting. Align to the objectives of the study to forecast dengue cases and visualize dengue case distributions based on historical data. The system aims to assist public health agencies, researchers, and policymakers in identifying high-risk areas and implementing timely interventions.

Plan. The proponents gather and document all necessary information to ensure the software system aligns with the needs and expectations of the intended users. In this study it involves understanding the specific requirements of public health agencies, and policymakers who will utilize the forecasting model. The proponents conduct consultations, interviews, and surveys to determine the key features that stakeholders require for the development of real-time geospatial mapping, data visualization, and predictive analytics for dengue case trends. The collected information is then used to create detailed documentation, including functional requirements, system architecture, and user interface specifications. It also involves defining the system's performance criteria, ensuring that the model delivers accurate predictions based on historical dengue data. By establishing clear objectives and a structured development plan, the proponents ensure that the system remains adaptive, scalable, and capable of meeting the client's expectations, ultimately contributing to public health interventions and strategic decision-making.

Design. In this section, the requirements from the surveillance agency are referred to tore build the software structure. The proponents determine approaches and present the chosen design for the system and its development.

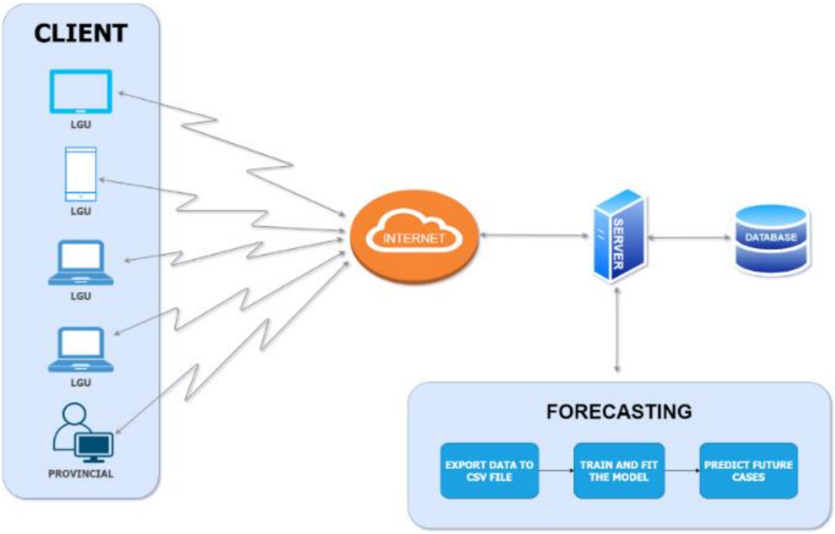

Fig. 2: System Architecture

This figure represents the overall architecture and workflow of the dengue forecasting system using a forecasting algorithm. It highlights how users, primarily Local Government Units (LGUs) and provincial authorities, interact with the system to access and analyze dengue-related data. These users can connect through various devices, such as computers, laptops, and mobile phones, ensuring accessibility and convenience. Their requests and data submissions are transmitted through the Internet, which serves as the communication bridge between the clients and the system's backend. The server is at the core of the system, which processes incoming data and manages computational tasks. The server is linked to a database, where historical dengue records, and other important datasets are stored. This data is essential for training the forecasting model and generating predictions. The forecasting process involves three key steps: exporting data to a CSV file, which prepares the dataset for analysis; training and fitting the model, where historical trends are analyzed to recognize patterns; and predicting future dengue cases. By leveraging this system, government agencies can efficiently monitor dengue trends, anticipate outbreaks, and implement timely preventive measures, ultimately reducing the impact of dengue in affected areas in the province.

2.1. Auto-regressive Integrated Moving Average with eXogenous factors (SARIMAX)

S. Stands for seasonality. In case you identify that the data patterns are repeating every month /year, then yes, it is seasonality.

AR. Auto-Regressive means that the current value depends on all lagged or past values.

I. Stand for several differences created from the previous existing values. If a dataset is not stationary, make it stationary, as in the case of the ARIMA model.

MA. Moving Average and to extent we need to roll it up.

X stands for other exogeneous variables that cause the variable to change. It is used when we are doing multivariate time series analysis [11].

2.2. SARIMAX Model Formula

The SARIMAX model is represented as:

$$Y_t=\mu+\phi_1 Y_{t-1}+\phi_2 Y_{t-2}+ \cdots+\phi_p Y_{t-p}+\theta_1 \epsilon_{t-1}+\theta_2 \epsilon_{t-2}+ \cdots+\theta_q \epsilon_{t-q}+X_t\beta+S_t+\epsilon$$

where:

- Y_t = The dependent variable (time series value) at time t
- μ = Constant term (mean)
- ϕ_i = Autoregressive (AR) coefficients (impact of past values)
- p = Number of autoregressive terms (AR order)
- θ_i = Moving average (MA) coefficients (impact of past errors)
- q = Number of moving average terms (MA order)
- ϵ_t = White noise error term at time t
- X_t = Exogenous variables (optional, for SARIMAX only)
- β = Coefficients for exogenous variables
- S_t = Seasonal component (captures periodic fluctuations)
- d= Order of differencing (to make the series stationary)

2.3. SARIMAX Notation

The SARIMAX model is often written as[12]:

SARIMA(p,d,q)×(P,D,Q,s)

where:

- (p,d,q) = Non-seasonal ARIMA terms:

p = Number of autoregressive terms (AR)

d = Number of differencing steps

q = Number of moving average terms (MA)

- (P,D,Q,s) = Seasonal ARIMA terms:

P = Seasonal autoregressive terms

D = Seasonal differencing

Q = Seasonal moving average terms

s = Seasonal period

Build. This phase is where the conceptual design is transformed into a fully functional system. After finalizing the design documentation, developers begin writing the source code to implement the system's features and functionalities. This stage involves developing the predictive model, integrating geospatial mapping tools, and ensuring seamless data processing. The algorithm model is coded and optimized to analyze historical dengue cases. Developers also work on creating a user-friendly interface that allows Local Government Units (LGUs) and health agencies to input, retrieve, and visualize data. The system is built to handle real-time data updates while maintaining accuracy in forecasting future dengue trends. Moreover, database management is implemented to store and organize datasets efficiently. Throughout this phase, proponents conduct initial testing to identify and resolve issues, ensuring that the system operates as expected. By carefully translating design specifications into a working software solution, this stage plays an important role in building a reliable tool for health authorities to predict and mitigate dengue outbreaks.

Test and Review. This phase plays an important role in the study where the developed system undergoes constant testing to ensure its functionality, accuracy, and reliability. At this point, the dengue forecasting system, built using the forecasting algorithm, is evaluated to confirm that it meets the intended requirements and operates efficiently. Testing methods using the ISO 25010 software quality standards, were conducted to identify any potential issues or inconsistencies of the system. Additionally, data validation is performed to verify that the model accurately processes historical dengue case data and generates reliable predictions. The geospatial mapping component is also reviewed to ensure that dengue hotspots are correctly visualized on the interface. The proponents collaborate with stakeholders, including Local Government Units (LGUs) and public health agencies, to gather feedback and make necessary improvements. This phase is essential in refining the system's performance, ensuring it delivers accurate forecasting results, and making it a dependable tool for disease monitoring and outbreak prevention efforts.

Launch. After testing and reviewing, the developed software will be deployed and can now be accessible to the target users. The Geospatial Data Analysis of Dengue Cases with Forecasting Analysis using the SARIMAX Algorithm System is needed in Hospitals, LGUs, and other agencies that monitor and supervise dengue cases.

3. RESULTS AND DISCUSSIONS

The study's results are presented and discussed with reference to the aim of the study, which was stated in the study's objectives.

3.1. Datasets on the number of dengue cases in Biliran Province from every barangay

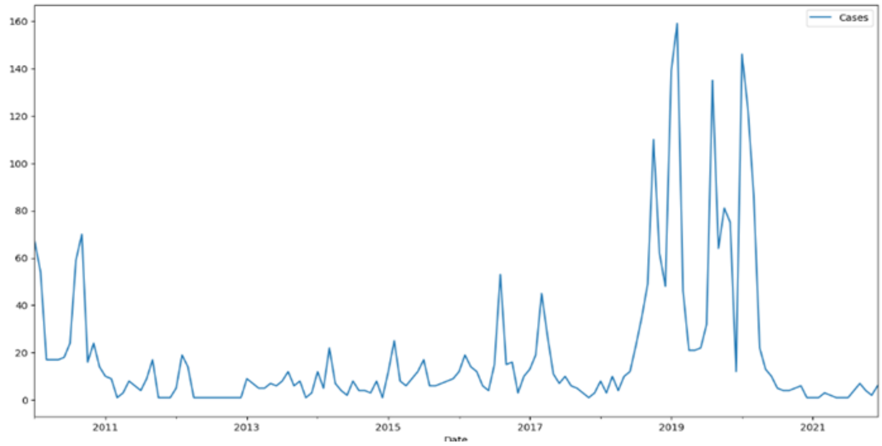

Fig. 3: Provincial Monthly Dengue Cases

Figure 3 shows the monthly number of dengue cases in Biliran Province from 2010 to 2021. Observing the graph, the proponents note a peak in dengue incidence between 2019 and 2020.

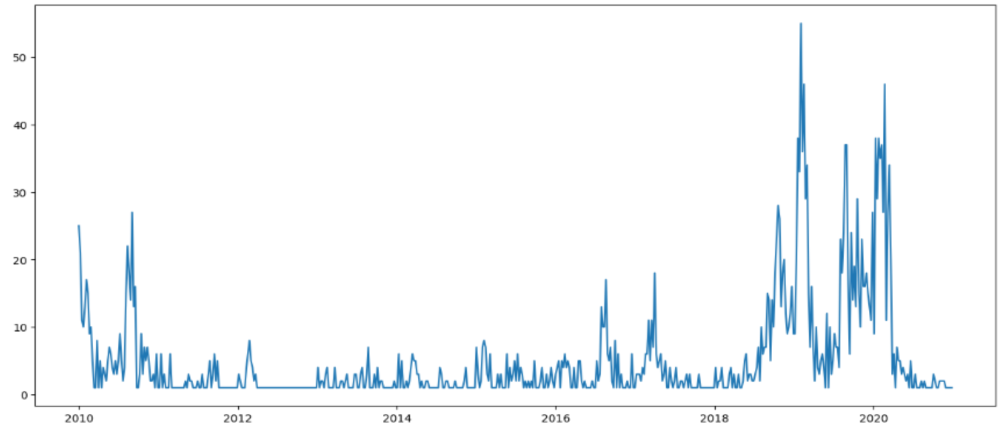

Fig. 4: Provincial Weekly Dengue Cases

Figure 4 shows the weekly number of dengue cases in Biliran province from 2010-2020. Observing the graph, the proponents noticed a peak in dengue incidence between 2019-2020.

3.2. Created a model using the SARIMAX algorithm to forecast the number of dengue cases in Biliran province.

```python
from statsmodels.tsa.stattools import adfuller

def adf_test(series):
    result = adfuller(series)
    print('ADF statistics: {}'.format(result[0]))
    print('p-value {}\n'.format(result[1]))

    if result[1] <=  0.05:
        print('Dataset is Stationary')
    else:
        print('Dataset is Non-stationary')
```

```
ADF statistics: -5.336424450444182
p-value 4.600360864168638e-06

Dataset is Stationary
```

Fig. 5: Checking of Stationary and Non-Stationary

This figure shows the checking if the series is stationary or non-stationary using the Augmented Dickey Fuller Test (adfuller()), from the statsmodel package. If the autocorrelation is positive for many numbers of lags (10 or more), then the series needs further differencing, as seen in the figure above, the p- value is negative or is less than 0.05. Therefore, time series is already stationary and there is no need for differencing. Therefore, our d is 0. Now we have (?, 0, ?).

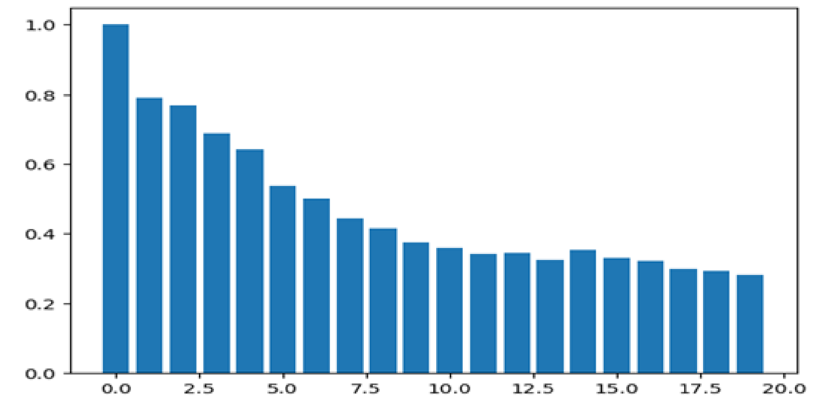

Fig. 6: Autocorrelation Function (ACF)

Figure 6 shows the Autocorrelation Function, which tells how many MA (Moving Average) terms are required to remove any autocorrelation in the stationarized series. The figure above shows no significant correlation since there is no spike or sudden rise of the number of dengue cases in Biliran; it just gradually decreases, therefore time series is not fit for MA (q), and we will leave it as 0. We now have (?, 0, 0).

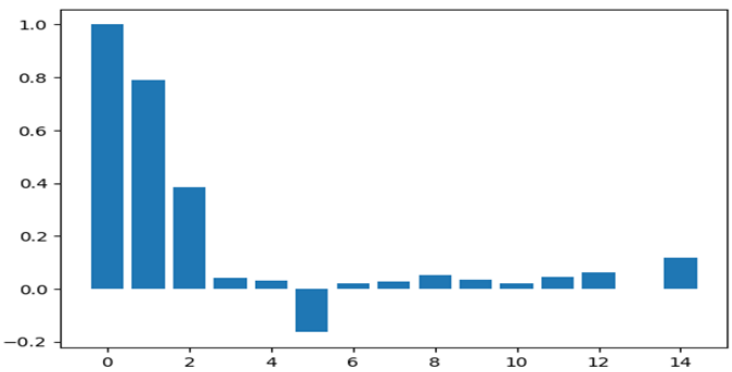

Fig. 7: Partial Correlation Function (PCF)

This figure shows the Partial Correlation Function (PCF), which can be imagined as the correlation between the series and its lag, it conveys the pure correlation between a lag and the series[13]. In the figure shown above, we can observe that 1 is quite significant. Therefore, our AR (p) is 1. Now we have (1, 0, 0). Now already have an order of (1, 0, 0).

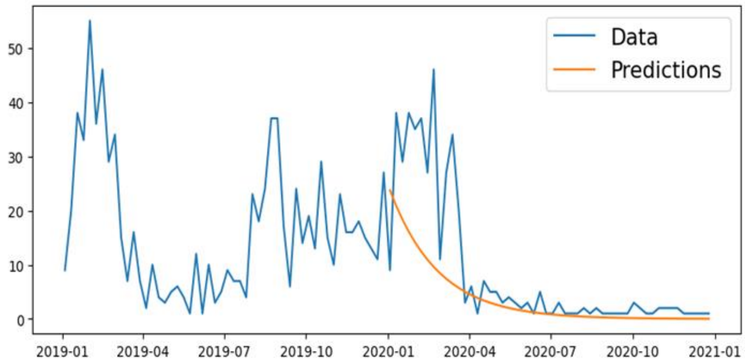

Fig. 8: Training and testing set

In this figure, the proponents have split the data into a training set and a testing set wherein the set will contain the last 52 data points (52 weeks or 1 year). After training a model that would predict and fit for order (1, 0, 0), results show a weak and not strong prediction since it predicted 52 periods at once instead of one at a time.

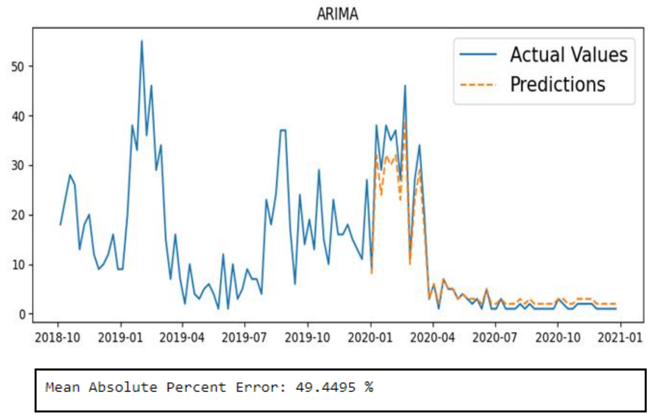

Fig. 9: Rolling Forecasting ARIMA

Figure 9 shows the other technique in forecasting dengue cases which is the Rolling Forecasting ARIMA in which it forecasts based on a month it passes, forecasting the next month automatically.

Rolling Forecasting ARIMA showed a better result than the first technique being used. Its Mean Absolute Percent Error is 49.4495%, meaning its prediction is 49% far from its actual values.

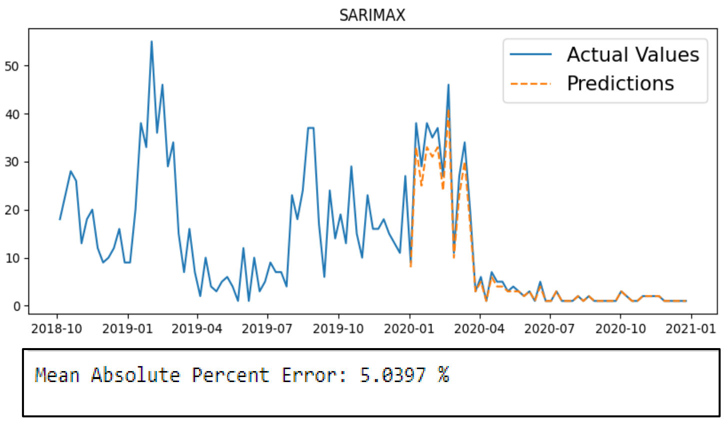

Fig. 10: SARIMAX Algorithm

This figure shows that the SARIMAX algorithm forecasts the number of dengue cases. Here, the proponents will not use the Seasonal and Exogenous exponent; instead, the proponents used the same order (1, 0, 0) using SARIMAX algorithm.

Results showed that the Mean Absolute Percent Error of the rolling forecasting SARIMAX is 5.0397% which means that its prediction is 5% near the actual values.

A good Mean Absolute Percent Error (MAPE) Score should be less than 10%; among all the techniques and algorithms being used, the Rolling Forecasting SARIMAX is the best fit in predicting the number of dengue cases in Biliran Province since it is 5% near the actual values[14,15].

3.3. An application to forecast dengue cases and a map to visualize the risk of dengue in Biliran province.

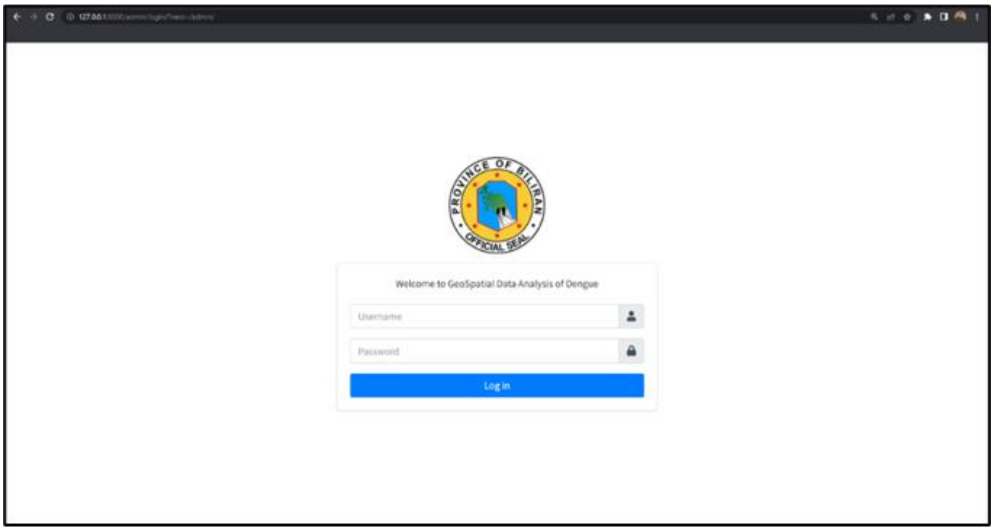

Fig. 11: Login Page

Figure 11, it shows the login page of the application. It allows superusers and staff users to access their accounts in the application with their usernames and password.

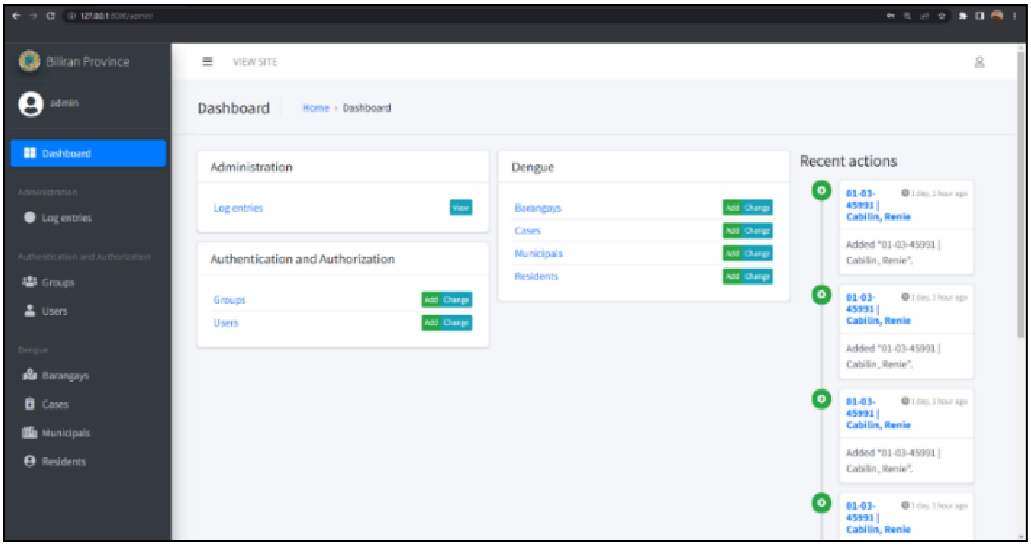

Fig. 12: *Dashboard*

Figure 12 shows the application's dashboard. It gives users, both superusers and staff, an overview of the whole feature set and functionalities of the application; this includes record transactions, where users can add, edit, delete, view, and even print reports of the records.

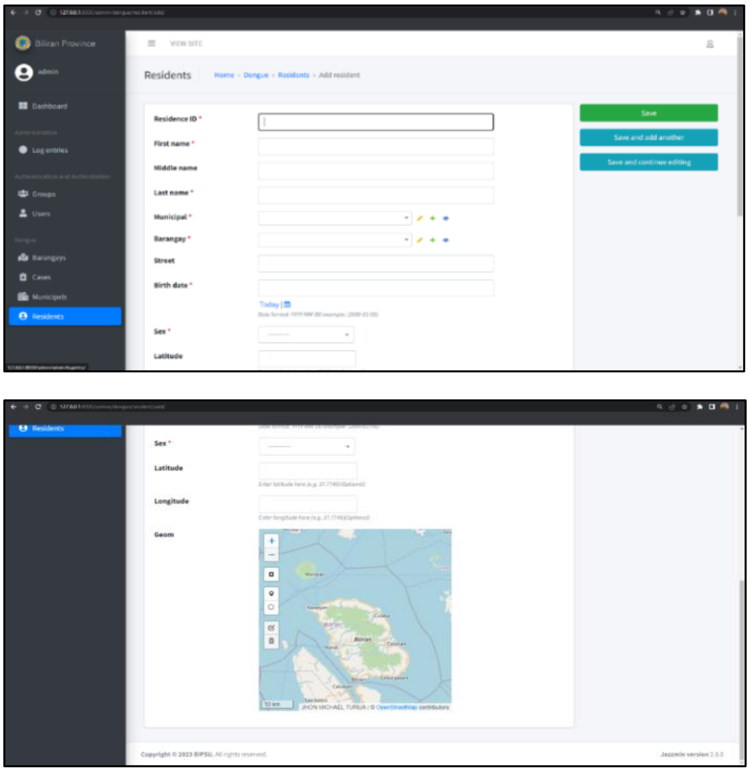

Fig. 13: *Add Residents*

Figure 13 shows that the superusers and staff can add new residents to the application, their details including their name, address, date of birth, and gender

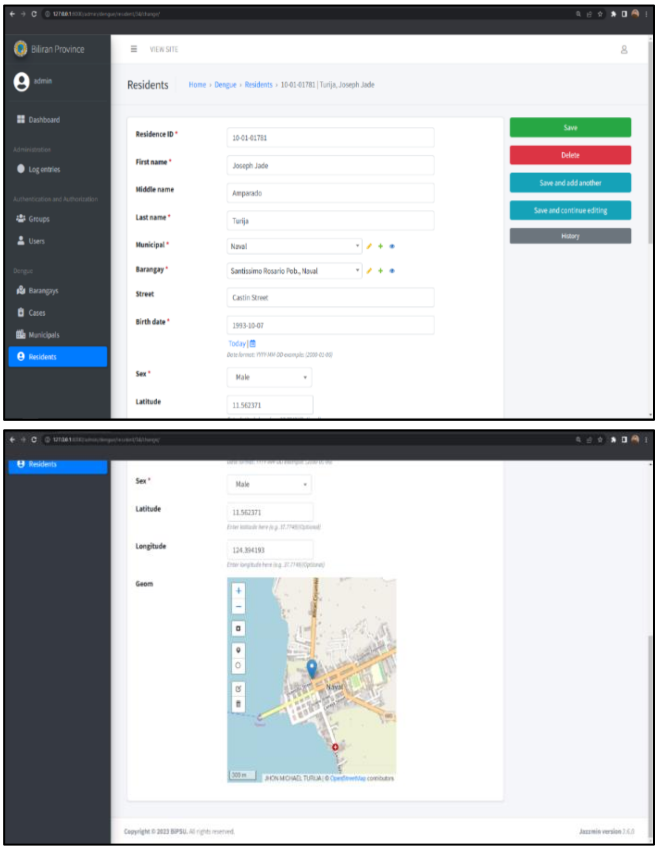

Fig. 14: Delete and Edit Residents

Figure 14 shows that the users manage and modify resident's information. By managing, the user can delete a resident, and by modifying, the user can update information about the resident.

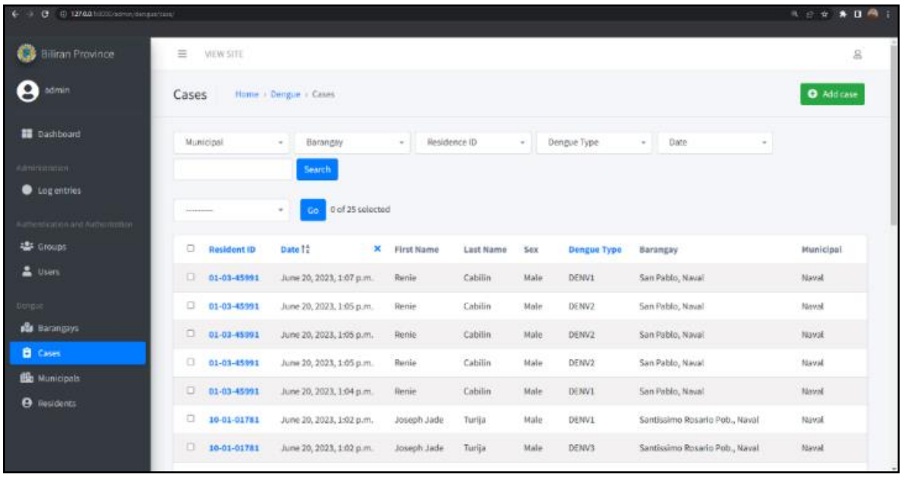

Fig. 15: View Cases

This figure allows the superuser or authorized users to access and view dengue cases-specific information within the application. It provides access to information about the number of dengue cases in Biliran province. This includes patient information such as name, address, dengue type, gender, and when the patient was diagnosed with dengue.

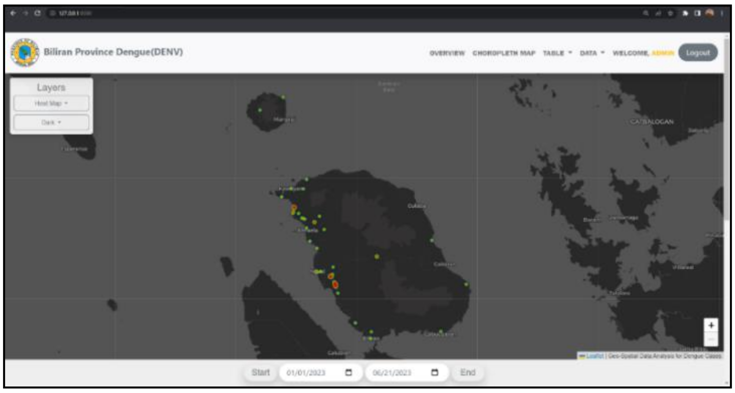

Fig. 16: Heat Map

This figure shows a Heat map that visualizes the intensity of the risk areas of dengue. It uses gradient colors that convey the magnitude or intensity of the dengue cases in a certain area. The darker the color, the higher the intensity; the lighter the color, the lower the intensity, which can help identify high-risk dengue areas.

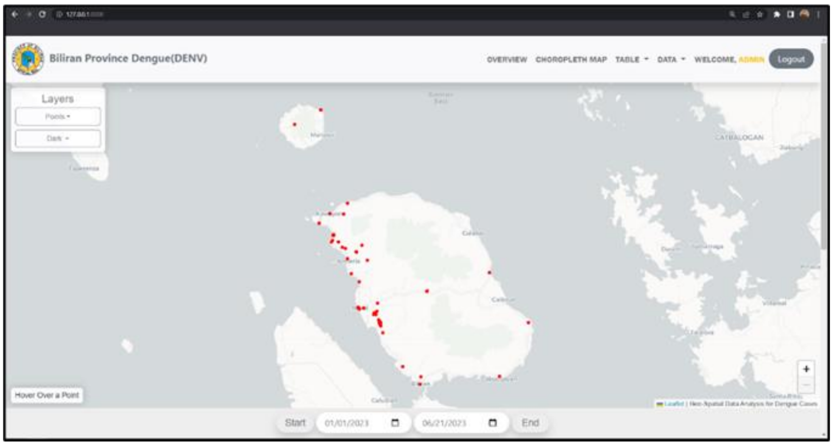

Fig. 17: Point Map

In this figure, it shows the point map, which presents data points allowing users to visually identify a specific location of a dengue patient based on the geographic coordinates.

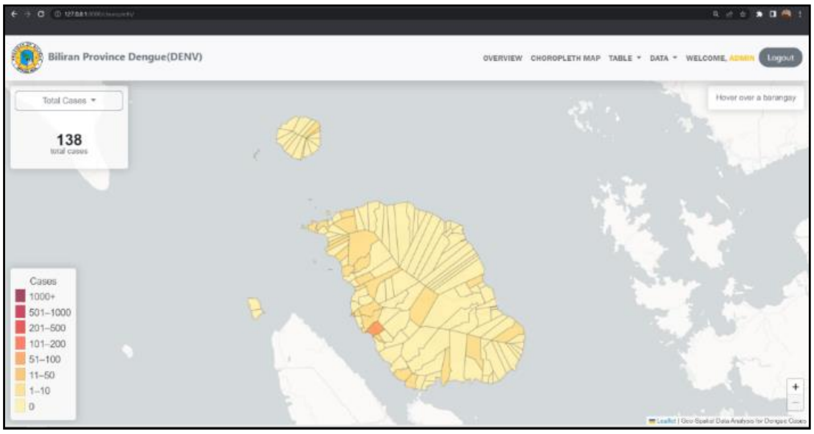

Fig. 18: Choropleth Map

This figure shows the Choropleth map, which visualizes geographical areas divided with colors, shades, or patterns and presents the density of the number of dengue cases in a certain region.

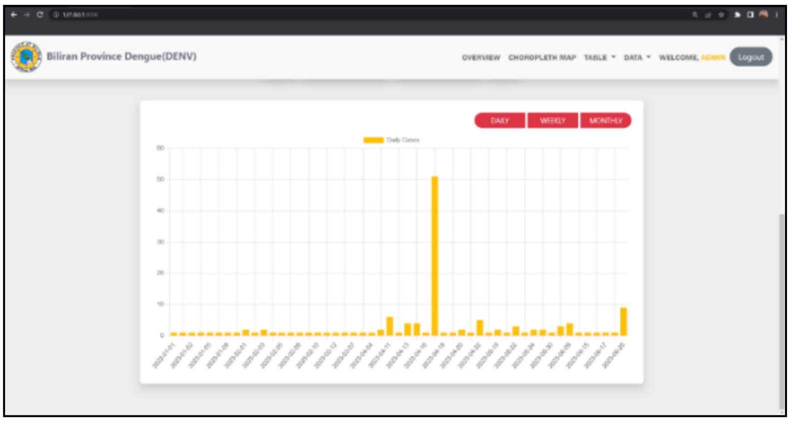

Fig. 19: Dengue Cases Chart

In this figure, it shows the chart that presents and illustrates the data on the number of dengue cases in Biliran province, which helps to give users clear and precise values without referring to a separate table.

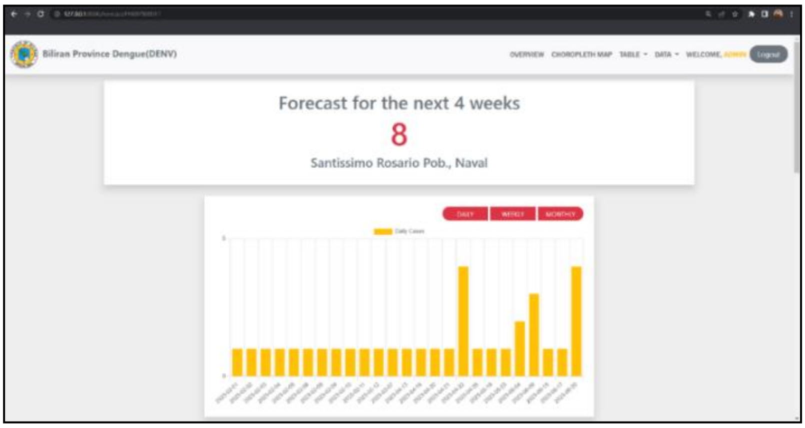

Fig. 20: Barangay Forecasting

This figure allows users to see the forecast or the possible number of dengue cases in the next four weeks of a specific barangay, which helps to take immediate action to prevent the spread of dengue.

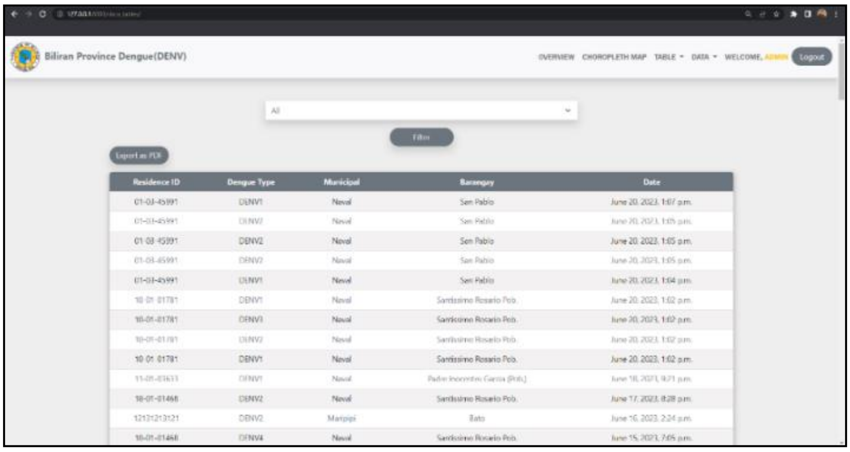

Fig. 21: All Records

This figure shows all the records of dengue patients in every barangay, including the dengue type and the date on which the patient was diagnosed and identified with dengue. The records can be exported and printed, which will be easily distributed to the respective LGU-RHU in every municipality of Biliran province.

3. CONCLUSION

This study successfully takes on the significant challenge of predicting dengue cases by integrating geospatial mapping with the SARIMAX algorithm. By incorporating exogenous variables, the model achieves enhanced prediction accuracy and gains a better understanding of how the disease spreads. The ability to forecast dengue outbreaks with greater precision strengthens disease surveillance and enables more proactive response strategies, which are essential for mitigating the impact of this global health threat. While dengue continues to pose significant challenges worldwide, combining geospatial mapping and the SARIMAX algorithm is a promising tool for improving public health outcomes. This study highlights the practical application of advanced analytical techniques in addressing pressing public health issues, demonstrating the potential to transform disease prediction and management. This study contributes meaningfully to the ongoing efforts to combat dengue by bridging the gap between data-driven innovation and real-world implementation. It underscores the importance of integrating technology and public health strategies for a healthier future

REFERENCES

[1] Nayak, M.; and Narayan, K.A. (2019). Forecasting Dengue Fever Incidence Using ARIMA Analysis. International Journal of Collaborative Research on Internal Medicine and Public Health. 11. 924-932.

[2] Wilson, C.; and Jumbert, M.G. (2018) The new informatics of pandemic response: humanitarian technology, efficiency, and the subtle retreat of national agency, Journal of International Humanitarian Action 3 (8): 1–13.

[3] Jain, S.; and Sharma, S.K. (2017). Challenges & options in dengue prevention & control: A perspective from the 2015 outbreak," The Indian Journal of Medical Research, vol. 145, no. 6, p. 718, doi: 10.4103/ijmr.ijmr_1325_16.

[4] Herbuela, V. R. D.; Karita, T.; Carvajal, T.; Ho, H.; Lorena, J. M.; Regalado, R.; Sobrepeña, G.; and Watanabe, K. (2021). Correction: Early detection of dengue fever outbreaks using a surveillance app (Mozzify): Cross-sectional Mixed Methods Usability study. JMIR Public Health and Surveillance, vol. 7, no. 4, p. e29795, Apr. 2021, doi: 10.2196/29795.

[5] Undurraga, E. A. (2017). Disease burden of dengue in the Philippines: Adjusting for underreporting by comparing active and passive dengue surveillance in Punta Princesa, Cebu City. American Journal of Tropical Medicine and Hygiene, vol. 96, no. 4, pp. 887–898, doi: 10.4269/ajtmh.16-0488.

[6] Madewell Z.J.; Hernandez-Romieu A.C.; and Wong, J.M. (2024). Sentinel Enhanced Dengue Surveillance System — Puerto Rico, 2012–2022. MMWR Surveill Summ 2024;73(No. SS-3):1–29. DOI: http://dx.doi.org/10.15585/mmwr.ss7303a1

[7] WHO. (2024). Dengue - Global situation. Retrieved July 7, 2024, from https://www.who.int/emergencies/disease-outbreak-news/item/2024-DON518.

[8] Salim, M. F.,; Satoto, T. B. T.; Danardono, D.; and Daniel, D. (2024). Digital Health Interventions in Dengue Surveillance to Detect and Predict Outbreak: A scoping review. The Open Public Health Journal, vol. 17, no. 1, doi: 10.2174/0118749445283264240116070726.

[9] Murray, J. and Cohen, A.L.(2016). Infectious disease surveillance. in Elsevier eBooks, pp. 222–229. doi:

10.1016/b978-0-12-803678-5.00517-8.

[10] GeeksforGeeks. (2024). Software Development Life Cycle (SDLC). GeeksforGeeks, Retrieved Dec 10, 2024, from https://www.geeksforgeeks.org/software-development-life-cycle-sdlc/

[11] Artley, B. (2025). Time Series Forecasting with ARIMA , SARIMA and SARIMAX. Towards Data Science, from https://towardsdatascience.com/time-series-forecasting-with-arima-sarima-and-sarimax-ee61099e78f6/

[12] J. A. R. and Ortiz, J. E. Forecasting SARIMAX and ARIMA models - Skforecast Docs, from https://skforecast.org/0.7.0/user_guides/forecasting-sarimax-arima

[13] Kis, A. (2024). Understanding autocorrelation and partial autocorrelation functions (ACF and PACF), Medium, Retrieved Nov. 19, 2024, from https://medium.com/@kis.andras.nandor/understanding-autocorrelation-and-partial-autocorrelation-functions-acf-and-pacf-2998e7e1bcb5

[14] Moreno, J. M.; Pol, A. P.; Abad, A. S.; and Blasco, B. C. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy," Psicothema, vol. 4, no. 25, pp. 500–506, Nov. 2013, doi: 10.7334/psicothema2013.23.

[15] Roberts, A. (2023). Mean Absolute percentage error (MAPE): what you need to know," Arize AI, Retrieved Feb. 10, 2023, from https://arize.com/blog-course/mean-absolute-percentage-error-mape-what-you-need-to-know/

...