Flood Forecasting and Alert System using Multivariate Time Series Models with Mobile Application

CUIZON, John Christian A.
Electronics Engineering Department
Technological University of the Philippines

DELA CRUZ, John Miguel F.
Electronics Engineering Department
Technological University of the Philippines

FRANCISCO, Justine V.
Electronics Engineering Department
Technological University of the Philippines

RIVERA, Kevin Gabriel G.
Electronics Engineering Department
Technological University of the Philippines

SOLA Jr., Ronnel S.
Electronics Engineering Department
Technological University of the Philippines

MELEGRITO, Mark P.
Faculty
Electronics Engineering Department
Technological University of the Philippines

Abstract – Philippines' susceptibility to natural disasters, particularly in flood-prone areas like Bulacan, causes frequent disruptions to livelihoods, infrastructure damage, and significant economic setbacks. This study presents a real-time flood forecasting and alert system in Barangay Frances, Calumpit, Bulacan, to mitigate flood damage in the area. It utilizes a wireless sensor network to measure hydrological parameters. To be specific, water level (Ultrasonic Sensor), rainfall (tip bucket), and flow rate (Flow Meter) operate through the Wireless Sensor Networks (WSN) and the Internet of Things (IoT). Moreover, to provide a much more accurate forecast, researchers implement multivariate time-series models, namely LSTM, ARIMA, and Random Forest. The results indicate that the LSTM model achieved a MAPE of 2.84%, demonstrating superior accuracy compared to the Random Forest model, which had a MAPE of 14.67%.

Keywords – flood forecasting, time series, LSTM, Random Forest, ARIMA

I. INTRODUCTION

The Philippines, located in the Northwestern Pacific Basin, the world's most active tropical cyclone basin, is highly vulnerable to flooding, with approximately 60% of its land area and 74% of its population at risk of natural disasters triggered by floods [1]. Despite being able to forecast rainfall and track storm paths using satellite imagery, it is necessary to have real-time monitored data, including flow rates, precipitation, and water levels. This data is essential for making informed decisions and determining necessary actions to minimize the impact of flooding. Bulacan, a province in the Philippines, faces recurrent flooding due to its geographical location and proximity to water bodies. One of the recent catastrophic flooding events in Bulacan occurred in July 2021, exacerbated by southwest monsoon rains intensified by Typhoon Fabian, causing the destruction of agricultural products and infrastructure valued at approximately PhP 155 million and evacuating more than 500 families [2]. With this in mind, good flood management by the local authorities is essential to lessen the damage dealt by the flood. A flood monitoring and forecasting system would benefit the community, allowing more emergency preparation and response time. Damage caused by flooding can be mitigated through monitoring, predicting, simulating, assessing, and analyzing [3].

Over the years, flood monitoring and forecasting have been developed to minimize flood danger in inhabited areas. Since 1973, the Government has piloted a Flood Forecasting and Warning System (FFWS) for the Pampanga River Basin to mitigate losses of life and property brought about by annual flooding occurrences [4]. PAG-ASA project developed a system for real-time monitoring of rainfall and water levels at gauging stations, along with a computer system for processing this data to predict inflow into reservoirs during floods. Recently, interest in utilizing machine learning (ML) techniques for flood prediction has surged. Rajan et al. [5] evaluated the performance of ML models using metrics such as the R2 score, root mean squared error, and validation loss. Their findings suggest that polynomial regression, random forest regression, and long short-term memory (LSTM) exhibited the highest performance levels. Adopting machine learning techniques in flood monitoring and forecasting presents a promising avenue for improving the accuracy of predictions. This integration, alongside traditional methods, can enhance emergency preparedness and response efforts, ultimately reducing the adverse effects of flooding on communities like Bulacan.

II. BACKGROUND OF THE STUDY

The local government of Calumpit, Bulacan, faces challenges evacuating residents during floods due to overcrowding and limited infrastructure. Flood markers are used for monitoring, but evacuation and relief efforts are hindered by poor road conditions, strong currents, and high water levels, worsened by high tide [6]. Most of the time, the only means of communicating updates, dam releases, and flood warnings rely solely on news, television, radio, and manual methods, leading to insufficient awareness and significant damage to properties and lives [7]. As a result, this situation underscores the need for a flood warning system capable of monitoring and forecasting current and potential flood conditions and promptly informing nearby residents. Therefore, this study introduces a project to assist flood-affected communities in Bulacan province, particularly in Barangay Frances, Calumpit.

Rainfall forecasting has received significant scientific importance recently due to its intricacies and long-term uses, such as flood forecasts and monitoring. Existing approaches employ complicated statistical models that are frequently excessively expensive, both computationally and financially [8]. Therefore, Machine learning has been considered a powerful tool in forecasting hydrological phenomena such as floods, utilizing time-series historical data to overcome these drawbacks [9]. Additionally, hydrological forecasting, including flood prediction, is typically categorized based on lead time duration. Short-term forecasts usually have a lead time of up to 2 days, while medium-term forecasts extend from around 2 to 10 days. Long-term forecasts cover periods exceeding 10 days, while seasonal forecasts span several months ahead [10]. Although the forecast lead time is usually chosen to fit operational requirements, whether daily work or long-term planning, selecting the right time-series scale pattern to feed into the forecast model is challenging because of data variability and availability issues. Therefore, comparing multivariate time series models offers a viable way to identify the dominant model to provide an accurate forecast [9]. In recent years, prior studies focused on flood monitoring, forecasting, and warning systems; however, they have overlooked the forecasting of receding water levels. Thus, the researchers aim to create a more comprehensive system with a predictive flood-receding feature. By exploring the integration of mobile application with selected multivariate time series models, this study aims to contribute to the technology in disaster management to alert authorities and affected individuals of an impending flood, promoting efficient disaster preparedness and improving emergency response.

III. OBJECTIVES

The objective of this journal is to develop flood forecasting – including rising and receding water levels, specifically, it aims to:

- Develop a Mobile application for real-time notification and alert.
- Develop a predictive model using LSTM, Random Forest, and ARIMA

Along with the forecasting capabilities of the system, the researchers aim to develop a mobile application for real-time alerts and notifications. Through the hydrological sensors, the system would be able to provide sensor database and flood warning levels, which will help the residents check the current and potential flood situation.

This research journal aims to develop a predictive model using three multivariate time series algorithms to forecast rising and receding water levels. Using the necessary data, such as rainfall, water level, and flow rate, the researchers aim to choose the most suitable model that will generate an accurate forecast.

IV. RELATED LITERATURE

The study's proponent seeks to develop a predictive model and mobile application for forecasting the rising and receding of floods, as well as distributing real time notification and alerts to inform local authorities and nearby residents in Barangay Frances, Calumpit, Bulacan. The goal is to provide an accurate forecast and to send real time water level status to help local authorities and residents take preventive measures about the current and upcoming situation of floods, tides, and dam water releases. With real-time data analysis, the system will provide early warnings, allowing for timely evacuation and efficient resource allocation. This initiative seeks to protect lives and property in Barangay Frances, Calumpit, Bulacan, by increasing community resilience and disaster preparedness.

A study by S. Ibañez et al. [11] on forecasting reservoir water levels using deep neural networks implemented several statistical and machine learning-based methods for forecasting water levels. The study evaluated six different forecasting methods, namely: naïve/persistence, seasonal mean, autoregressive integrated moving average (ARIMA), gradient boosting machines (GBM), and two deep neural networks (DNN) using a long short-term memory-based (LSTM) encoder-decoder architecture, specifically a univariate model (DNN-U) and a multivariate model (DNN-M). The findings indicate that in the 1-day-ahead scenario, the DNN-U model achieved the highest accuracy, with a mean absolute error (MAE) and root mean square error (RMSE) of 0.2 m. However, in the 30-day, 90-day, and 180-day-ahead scenarios, the DNN-M model performed better than the other models, with MAE (RMSE) scores of 2.9 (3.3), 5.1 (6.0), and 6.7 (8.1) meters, respectively.

In the study of S. Puttinaovarat and P. Horkaew [12], different machine learning techniques for flood forecasting systems, Decision Trees, Random Forest, Naïve Bayes, ANN, Support Vector Machine, and Fuzzy Logic based on fusing meteorological, hydrological, geospatial, and crowdsource big data. The flood forecasting outcomes were divided into four classes, they were as follows: 1) no flood is expected, 2) the flood level is less than 20 cm, 3) the flood level is between 20 and 49 cm, and 4) the flood level is 50 cm or higher. Corrected Classified Instances (CCI), Kappa, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), True Positive (TP), False Positive (FP), Precision, Recall, F-Measure, and Area under ROC were the standard accuracy metrics used. The result shows that MLP ANN, SVM, and Random Forest were more accurate as machine learning algorithms in this study in comparison to others. (i.e., 97.83%, 96.67%, and 96.67%) and Kappa coefficients (i.e., 0.89, 0.84, and 0.84).

The study by Mohd Sabre et al. [13] focuses on flood warning and monitoring system integrated with the Blynk application. It mainly concentrates on the current status of flood and rain by viewing the interface and receiving a push notification that is available in the Blynk application via IOS or Android smartphones. It developed a flood monitoring system application for monitoring data from NodeMCU, which is connected to ultrasonic and rain sensors over the internet using a smartphone. The flood detection system is represented by an LCD display, which provides information on the water level status, indicating whether it is at a Safety level, Warning level, or Critical level. Subsequently, three LEDs are exhibited as indicators, activating in response to the flood level state. A rain intensity level widget is included to visually represent the degree of rainfall, serving as the initial monitoring system for rain before a significant flood event occurs. The system continuously sends alert notifications and emails to the victim when the flood water level and rain intensity reach a specific hazard threshold. This study will serve as a basis for notification of the system. Similar to this study, push notification will be utilized in order to send specified alerts that includes, Alert, Alarm, and Critical levels. However, the use of mobile application's dashboard that provides information for the three parameters of the flood monitoring will be used in this study instead of an LCD display.

According to Kavitha et al. [14], combining standard SMS with WhatsApp messages increases the reach and efficiency of the early flood warning system. Their study highlights the characteristics of data processing and data transmission to be used. The ATmega8 microcontroller utilizes a serial communication interface to facilitate the efficient transmission of flood-related alarms to authorized personnel. Incorporating the WhatsApp module, which facilitates the transmission of warnings and notifications through the widely utilized messaging application WhatsApp, represents a unique characteristic of their system. As a modification, the study's proponent will incorporate an android application with several features that inform the residents about the flood-related alarms within the area. Lastly, Singh et al. [15] served as an inspiration for this study about creating the live feed of all the data obtained by the sensors. The study introduced a very efficient approach for an Internet of Things (IoT) enabled flood monitoring and alerting system. This system fetches real-time data for flood monitoring and alerting utilizing Java, XML, and Android Studio. In line with this, the proponents will utilize web hosting to create a domain that fetches real-time data using hostinger and later on will be displayed to the android application.

V. METHODOLOGY

The primary goal of the research is to forecast rising and receding flood levels in Barangay Frances, Calumpit, Bulacan, using the data from its upstream area in Barangay Sulipan, Apalit, Pampanga. The forecasted flood levels will be then display on the mobile application for real-time alerts and warnings.

A. Front-End Design and Development

The researchers developed a mobile application to facilitate the residents of Barangay Frances in Calumpit, Bulacan, to access information more conveniently. The platform was created by employing various programming languages to construct the interface and web servers to host the mobile application on the Internet. In addition to displaying monitored and predicted data, the website includes additional features such as flood level monitoring, water dam release information, routing options, flood safety guidelines, emergency contact information, and details about the individuals or organizations responsible for the mobile app.

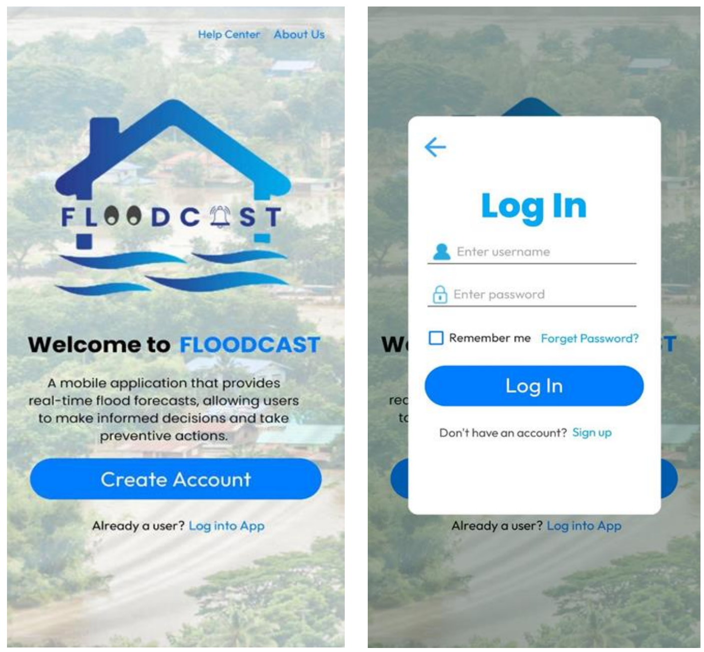

Fig. 1. *Launching and Login Screen*

The provided figure illustrates the front-end design of the launch and login screens within the Floodcast app. The launch screen prominently displays the app's logo and a navigation menu with subcategories detailing app information and its stakeholders. It offers two distinct buttons for user login and registration, facilitating access to flood incident information. The login screen features designated input fields for Username and Password, allowing users to enter their credentials. Upon selecting the Login button, positioned at the screen's bottom, users are seamlessly directed to the Floodcast application's Dashboard page, granting access to their accounts.

B. Back-End Development

For monitoring sensor readings from the NodeMCU v3, the researchers utilized XAMPP to create and manage the database. XAMPP combines the Apache web server, MySQL database, and PHP, providing a versatile environment for database operations. The NodeMCU v3 records sensor data, and XAMPP facilitates the storage of this data in the MySQL database. To enable real-time access and visualization on a mobile application, an API (Application Programming Interface) is implemented. The API acts as a link between the database and the mobile application, facilitating secure and efficient data transfer. Through well-defined endpoints, the mobile application retrieves sensor readings from the MySQL database using HTTP requests, ensuring smooth integration and timely updates on the mobile interface.

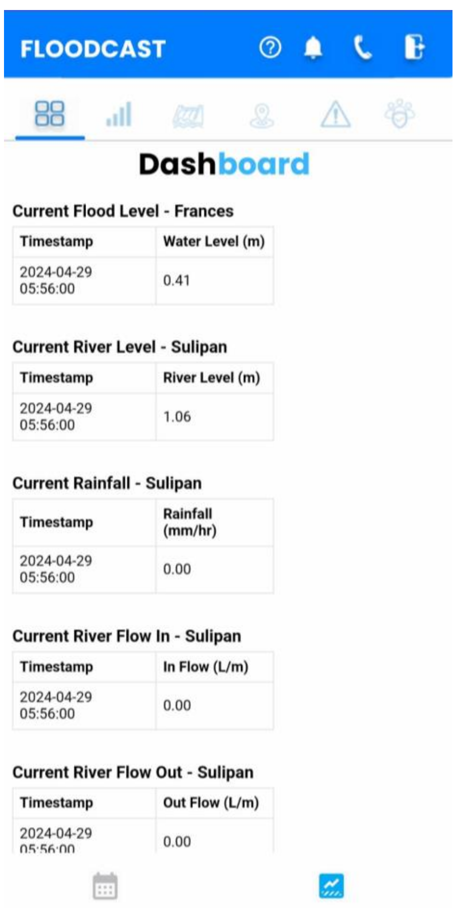

Fig. 2. *Live Feed of the Sensor*

Figure 2 illustrates the live feed data of every sensor connected to each monitoring station of the device. All of the data were labeled to their corresponding parameter together with their monitoring location. Furthermore, the timestamp was also included for real time purposes.

C. Development of Predictive Models

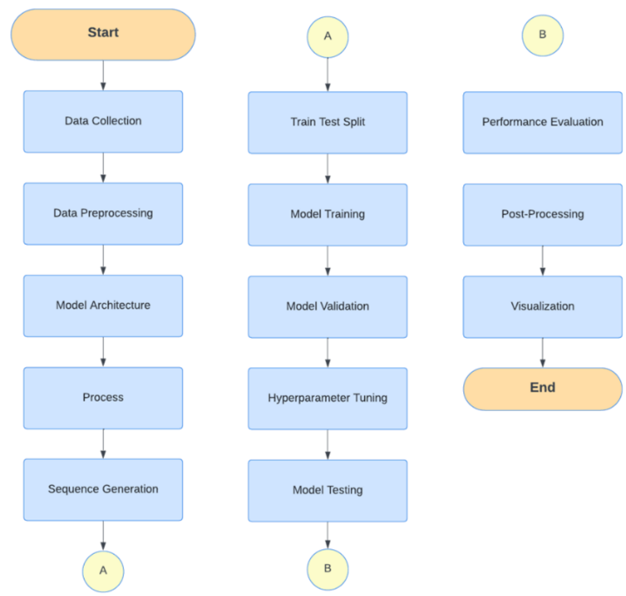

Fig. 3. Flowchart for the Development of Predictive Algorithms

This research employs a systematic process depicted in the flowchart. It started with data collection, gathering river levels and rainfall patterns using hydrological sensors. This data then undergoes preprocessing, involving cleaning and transformation, to ensure consistency and quality of the data. After the initial data preprocessing, sequences are generated for time series models, including LSTM, ARIMA, and Random Forest. This involves creating pairs of input and output to predict future events. Subsequently, the data is divided into a training set for model development and a testing set for assessing the model's effectiveness. The models are then trained using the training data to learn patterns and relationships within the data. Validation follows, ensuring the models perform well and adjusting parameters to prevent overfitting. Hyperparameter tuning optimizes model performance using grid or random search techniques. The models are then tested on unseen test data to evaluate their predictive accuracy and generalization capabilities.

Performance evaluation is conducted using metrics such as RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and accuracy to ensure the models meet the required standards for flood prediction. Post-processing of the model predictions is done to make them more interpretable, which involves smoothing predictions or translating them into alerts. Finally, the results are visualized using charts and graphs to understand the projections, aiding in interpretation and decision-making. The Process concludes with the system being ready to deliver reliable flood predictions based on the trained models, reassuring the audience about the thoroughness of the system's performance evaluation and instilling confidence in its reliability and effectiveness.

D. Accuracy Testing and Validation of Predictive Algorithm

In the Accuracy Testing and Validation of the Predictive Algorithm, Mean Absolute Error is used to evaluate how well the program works. It calculates its Mean Absolute Percentage Error regarding date and time by comparing the actual and forecasted flood levels for one week's hourly data.

Table 1. Parameters for Accuracy Testing and Validation

MAPE	**Interpretation**
<10	Highly accurate forecasting
10 – 20	Good forecasting
20 – 50	Reasonable forecasting
> 50	Inaccurate forecasting

The output values for actual and forecasted flood levels are then computed and compared. The performance of the model in each instance and its average are determined using the formula of Mean Absolute Percentage Error and Mean.

$$MAPE = \frac{Forecasted\ Value - Actual\ Value}{Actual\ Value}$$

$$Mean\ Value = \frac{Sum\ of\ Values}{Number\ of\ Values}$$

VI. RESULTS AND DISCUSSION

This section provides a thorough overview of the project's mobile application, as well as the results of the forecasted data using the three predictive algorithms.

A. Mobile Application

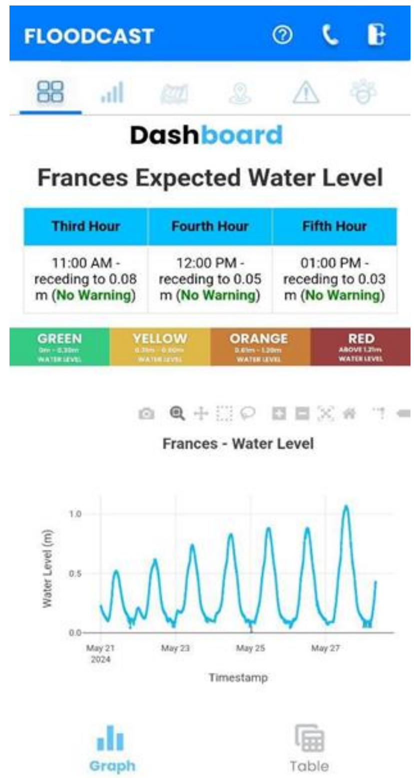

Fig. 4. Inside Page of the Mobile App

Figure 4 illustrates the Mobile App's dashboard page. This page functions as the focal point for accessing flood-related information and updates. Several navigation icons are positioned atop the Dashboard Page, facilitating access to different sections of the mobile application. Users can utilize this page to access predictive and real-time flood data presented in a graphical form derived from sensor readings.

Fig. 5. Routing Page Implementing the Google Maps JavaScript API

Figure 5 illustrates the Routing Page, which integrates the Google Maps API. Users can interact with the map interface by zooming in or out, panning across different geographical areas, and selecting specific markers to retrieve location-specific information. Notably, the depiction includes two green pins representing the positions of four nodes, with three nodes situated in Barangay Sulipan, Pampanga, and one node in Barangay Frances, Bulacan. Additionally, two red pins indicate the locations of critical government facilities within Barangay Frances, Bulacan, including the barangay hall and the evacuation center of Barangay Frances in Calumpit, Bulacan.

Fig. 6. Flood Forecasting and Warning Display

Figure 6 shows the forecasted flood and corresponding warning level integrated into the mobile app. This section describes the predicted time whether the flood rises or recedes in the next 3 to 5 hours lead time.

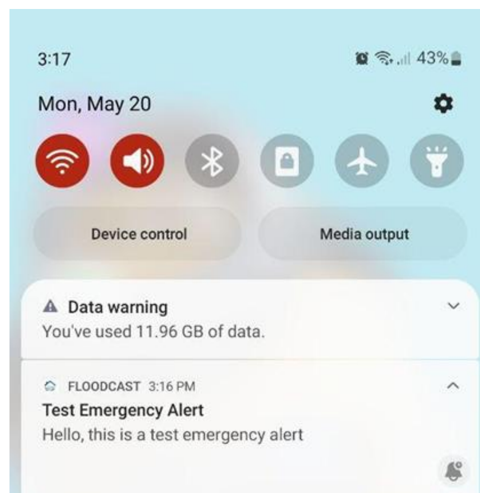

Fig. 7. Push Notification in Mobile Phone

Figure 7 illustrates the mobile application's push notification feature. This notification will automatically appear once the user has installed the application. It contains information and data from the mobile app, and upon clicking, users will be redirected to the app's dashboard.

B. Accuracy of the Predictive Models

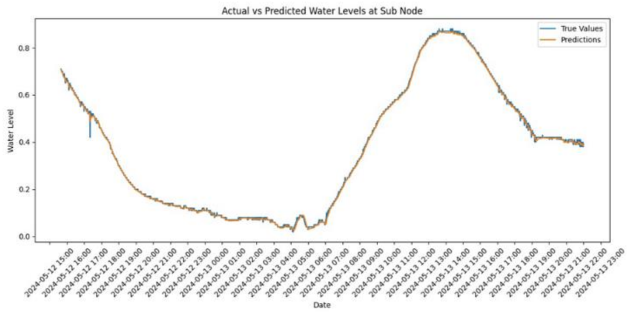

Fig. 8. Actual and Predicted Values of LSTM Model

Figure 8 shows the actual and predicted value of the LSTM model achieves an average Mean Absolute Percentage Error (MAPE) of 2.84%. This indicates that the LSTM model's predictions are, on average, only 2.84% different from the actual values gathered from the sensors, demonstrating a high level of accuracy in forecasting.

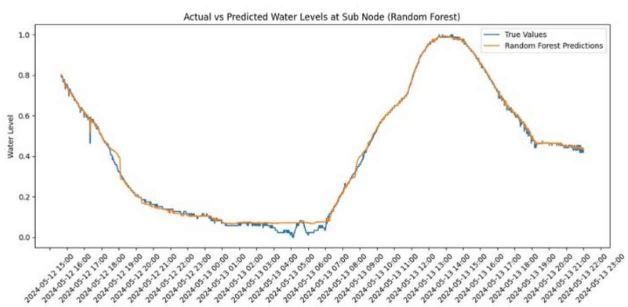

Fig. 9. Actual and Predicted Values of Random Forest Model

Figure 9 demonstrates the actual and predicted values from the Random Forest model, which achieves an average Mean Absolute Percentage Error (MAPE) of 14.67%. This means that the model's predictions deviate from the actual sensor values by an average of 14.67%. Although this error rate is higher than the LSTM model, it still reflects good forecasting accuracy.

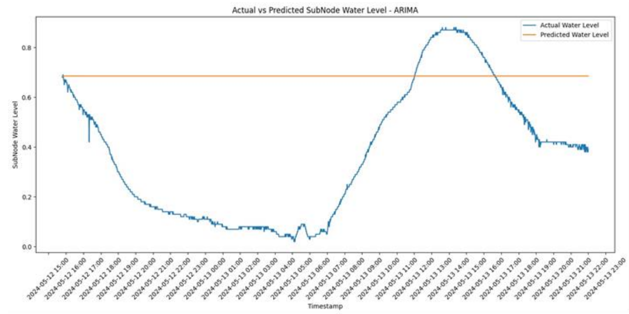

Fig. 10. Actual and Predicted Values of ARIMA Model

Figure 10 shows the actual and predicted values of the ARIMA model. The graph shows the predicted values are only linear due to the gathered data being measured every minute.

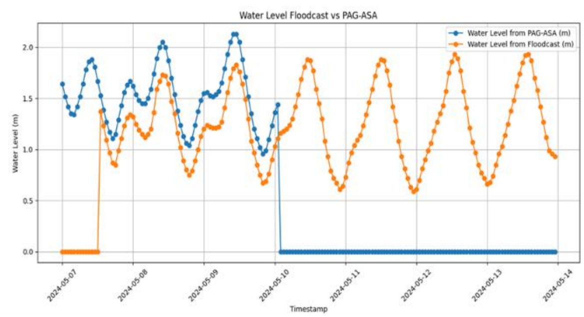

Fig. 11. Data Visualization for Sulipan – Water Level from Floodcast and PAG-ASA

Figure 11 displays the water level measured by Floodcast and PAG-ASA. As seen from the graph, Floodcast's gathered water level only differs by an average of 0.9 compared to PAG-ASA's readings. The recorded water level data from PAG-ASA was 0 from May 10 - May 14 due to the downtime of their system.

VII. CONCLUSION

The research provided a mobile application as a central platform that presents all the essential data gathered from the sensors, including the forecasted data, alerts, and warning systems. The study was developed using three algorithm models: LSTM, Random Forest, and ARIMA. These three algorithm models were trained using the datasets gathered from the sensors in the system. LSTM provides highly accurate forecasting, attaining an average of 2.84% Mean Absolute Percentage Error (MAPE). While Random Forest, despite obtaining an average of 14.67% MAPE, still reflects a relatively good forecasting accuracy. Lastly, ARIMA achieved a high

MAPE average due to the wrong process of gathering data. Although it is less accurate than LSTM and Random Forest, it can still be helpful in specific scenarios where simpler models are preferred. After evaluating the performance of the three models, we decided to integrate LSTM into our system due to its superior accuracy, ensuring more reliable flood predictions and timely alerts.

REFERENCES

[1] "World Bank Climate Change Knowledge Portal." https://climateknowledgeportal.worldbank.org/country/philippines/vulnerability

[2] Lagos, Devralin & Rebadolla, Justin & Anne, Feliza & Farin, Feliza Anne Bethina & Eco, Rodrigo. "Flooding and Subsidence Research: Bulacan Province, Northern Manila Bay Area, Philippines." https://www.researchgate.net/publication/369742698_Flooding_and_Subsidence_Research_Bulacan_Province_Northern_Manila_Bay_Area_Philippines

[3] Sunkpho, Jirapon & Oottamakorn, Chaiwat. "Real-time flood monitoring and warning system." Songklanakarin Journal of Science and Technology. 33. 227-235. https://www.researchgate.net/publication/263922229_Real-time_flood_monitoring_and_warning_system

[4] L. Mercado, “The Flood Forecasting and Warning System for Dam Operation (FFWSDO),” Dost.gov.ph, 2024. https://prsd.pagasa.dost.gov.ph/index.php/79-hydromet/hydromet/348-the-flood-forecasting-and-warning-system-for-dam-operation-ffwsdo#the-project (accessed Mar. 10, 2024).

[5] Rajab A, Farman H, Islam N, Syed D, Elmagzoub MA, Shaikh A, Akram M, Alrizq M. "Flood Forecasting by Using Machine Learning: A Study Leveraging Historic Climatic Records of Bangladesh. "Water. 2023; 15(22):3970. https://doi.org/10.3390/w15223970

[6] L. K. S. Tolentino et al., “Real Time Flood Detection, Alarm and Monitoring System Using Image Processing and Multiple Linear Regression,” SSRN Electronic Journal, 2023, doi: https://doi.org/10.2139/ssrn.4319789.

[7] E. R. Yabut et al., “e-wasBaha: A mobile application framework for flood monitoring in Metro Manila using crowdsourcing,” 2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Dec. 2017, doi: https://doi.org/10.1109/hnicem.2017.8269533.

[8] A. Y. Barrera-Animas, L. O. Oyedele, M. Bilal, T. D. Akinosho, J. M. D. Delgado, and L. A. Akanbi, “Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting,” Machine Learning with Applications, vol. 7, p. 100204, Nov. 2021, doi: https://doi.org/10.1016/j.mlwa.2021.100204.

[9] N. Khairudin, N. Mustapha, T. Noranis, M. Aris, and M. Zolkepli, “A STUDY TO INVESTIGATE THE EFFECT OF DIFFERENT TIME-SERIES SCALES TOWARDS FLOOD FORECASTING USING MACHINE LEARNING,” Journal of Theoretical and Applied Information Technology, vol. 99, no. 23, 2021, Accessed: Mar. 09, 2024. [Online]. Available: http://www.jatit.org/volumes/Vol99No23/9Vol99No23.pdf

[10] World Meteorological Organization, TECHNICAL REGULATIONS: Volume III - Hydrology, no. 2. 2006.

[11] S. C. Ibañez, C. V. G. Dajac, M. P. Liponhay, E. F. T. Legara, J. M. H. Esteban, and C. P. Monterola, “Forecasting Reservoir Water Levels Using Deep Neural Networks: A Case Study of Angat Dam in the Philippines,” Water, vol. 14, no. 1, p. 34, Dec. 2021, doi: https://doi.org/10.3390/w14010034.

[12] S. Puttinaovarat and P. Horkaew, “Flood Forecasting System Based on Integrated Big and Crowdsource Data by Using Machine Learning Techniques,” IEEE Access, vol. 8, pp. 5885–5905, 2020, doi: https://doi.org/10.1109/access.2019.2963819.

[13] M. S. Mohd Sabre, S. S. Abdullah, and A. Faruq, “Flood Warning and Monitoring System Utilizing Internet of Things Technology,” Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, vol. 4, no. 4, p. 287, Oct. 2019, doi: https://doi.org/10.22219/kinetik.v4i4.898.

[14] P. Kavitha et al., “The Development of Early Flood Monitoring and a WhatsApp-Based Alert System for Timely Disaster Preparedness and Response in Vulnerable Communities,” Engineering Proceedings, vol. 62, no. 1, p. 18, 2024, doi: https://doi.org/10.3390/engproc2024062018.

[15] G. Singh, N. Bisht, P. Bisht, and P. Singh, “Iot Based Flood Monitoring and Alerting System with Weather Forecasting,” International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 6, pp. 559–563, Apr. 2020, doi: https://doi.org/10.35940/ijitee.f3854.049620.