2023 IEEE 15th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) | 979-8-3503-8117-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/HNICEM60674.2023.10589082

Project Apaw: Spatiotemporal Forecasting of River Flood using Deep Learning

Joseph Jessie S. Oñate*, Challiz D. Omorog†, Rosel O. Onesa‡, Kaela Marie N. Fortuno§, Brenda D. Benosa¶, Keith Marlon R. Tabal, Ian P. Benitez, Philip Alger M. Serrano, Freddie B. Prianes and Ichelle F. Baluis
AI Research Center for Community Development, Camarines Sur Polytechnic Colleges
Nabua, Camarines Sur, Philippines
Email: *josephjessie97@cspc.edu.ph, †challizomorog@cspc.edu.ph, ‡roselonesa@cspc.edu.ph,
§kfortuno@cspc.edu.ph, §ninbenosa@cspc.edu.ph

***Abstract*—Flooding presents a growing global threat, particularly in vulnerable areas like the municipality of Nabua, Philippines, a low-lying municipality that serves as a catch basin for floodwaters from neighboring regions. This study, Project Apaw: Spatiotemporal Forecasting of River Flood using Deep Learning, aims to explore the appropriateness of Spatiotemporal Forecasting of River Flood using Deep Learning in the Municipality of Nabua. Hence, a localized weather monitoring system was implemented, consisting of an Indoor Control Unit (ICU) and an Outdoor Control Unit (OCU), which collect data on temperature, humidity, rainfall, water level, and wind speed. These units communicate via RS485 modbus, exchanging and uploading data to a webserver. Long-Short Term Memory model was used for flood prediction based on the collected data. The trained model demonstrated its ability to detect patterns and relationships in the data. Moreover, the findings demonstrate the potential of deep learning-based flood prediction models to accurately forecast river flood risks, with the constructed Spatiotemporal model achieving proficiency in capturing complex spatial and temporal correlations, as indicated by an RMSE value of 0.0101 and MAE of 0.0097. However, challenges related to data availability, quality, interpretability, and computational resources need to be addressed for widespread implementation. By incorporating social and economic data, considering spatial variability and climate dynamics, and collecting more data over an extended period, deep learning-based flood prediction models can become valuable tools for proactive flood mitigation and preparedness in vulnerable regions like Nabua, thus contributing to improved flood risk management practices.**

***Index Terms*—Deep Learning, Flood Prediction, Long-Short Term Memory, Spatiotemporal Forecasting**

I. Introduction

Flooding, caused by extreme precipitation and rainfall intensity, has become a global threat in recent years [3]. Flash floods, characterized by their rapid and unpredictable nature, pose the most significant danger [9]. The year 2021 witnessed a surge in natural disasters, with floods being the most prevalent, leading to thousands of deaths and significant economic losses. Asia, including the Philippines, suffered the most severe consequences, with the country being hit by an average of 20 typhoons annually [1], [2], [8]. These extreme weather conditions and heavy rainfall make flooding a recurring and potentially catastrophic phenomenon in the Philippines. The impacts of floods extend beyond property and infrastructure damage, affecting public services, production, and the overall well-being of the population [5].

The Philippines, located along the Pacific typhoon belt, experiences a high flood risk index and bears the brunt of typhoons and monsoon rains. Rural areas are particularly vulnerable to the effects of typhoons, with drowning being a leading cause of mortality. Assessing the cost and extent of flood-related damage is challenging, making it difficult to measure the true impact [10]. However, the immeasurable consequences of floods, including service disruptions and loss of production, highlight the urgent need for effective flood risk management and mitigation strategies [11].

The Municipality of Nabua, situated in a coastal province, serves as a minor central business district, predominantly consisting of rural settlements. Being a low-lying community, Nabua is highly susceptible to flooding, earning its reputation as a ”catch basin of floodwaters.” Even nearby municipalities' heavy rainfall can result in widespread flooding across the town. In 2020, during Typhoon Rolly, more than half of the village areas, including downtown Nabua, were submerged in floodwaters, posing a severe threat to its residents [6]. Given the local conditions, the increasing unpredictability of weather patterns, and the ongoing global issue of flooding, it is imperative to implement mitigation measures and anticipate the potential effects of such disasters.

Nabua's geographic location makes it prone to floods when there is consistent and uniform rainfall distribution. It acts as a convergence point for floodwaters from neighboring areas, particularly the Albay province's third district and the Camarines Sur's Rinconada areas. Floods significantly impact the livelihoods and economy of the Bicol region, affecting 40% of the area regularly, including Nabua. By establishing reliable and accurate guidance systems, flood-prone zones and waterways can better prepare for potential flood hazards through prevention, protection, and preparedness measures. Early warning systems can be crucial in mitigating the consequences of floods, saving lives, and minimizing the implications on people's livelihoods and the financial system [1].

Real-time flood forecasting is crucial in emergency response efforts in areas prone to flooding. Various factors, such as soil water content and the location of rainfall centers, impact flood predictions in small and medium-sized watersheds. Over the past decade, researchers have developed different strategies to improve the reliability of flood forecasting and minimize

979-8-3503-8117-7/23/$31.00 ©2023 IEEE

the impacts of floods. Two main approaches in flood forecasting research are physical or hydrological conceptual models and data-driven models [7], [12]. Physical models utilize mathematical equations to describe hydrological processes like rain, evaporation, and flow concentration [15]. On the other hand, data-driven models directly establish mathematical relationships between different hydrological parameters and runoff levels, based on historical data. Unlike physical models, data-driven models focus on mapping flood indicators to flow rates without considering detailed physical processes [16].

Spatiotemporal events are occurrences that happen at specific times and places, holding significance for various stakeholders. Detecting and understanding these events involves capturing the ”what” (theme), ”where” (spatial information), and ”when” (temporal information) aspects, as well as identifying participants, thematic attributes, causes, and effects [4]. The unique nature of spatiotemporal datasets requires a shift in data mining methodologies to leverage their rich geographical and temporal linkages. Challenges include visualizing patterns, adapting data mining approaches, and effectively handling spatial and temporal relationships. Considering neighboring instances and their impact is crucial for accurately analyzing spatiotemporal data [13], [14].

Deep learning techniques, particularly LSTM-based architectures like STA-LSTM, are proving to be crucial for spatiotemporal forecasting of river floods. Real-time flood prediction in seconds allows for timely dissemination of information to the public, aiding in the prevention of injuries and fatalities [1]. These deep learning models leverage attention mechanisms to extract dynamic features from the data, mimicking human visual attention. In a study conducted by Noor, different neural network models, including ANN, LSTM, SALSTM, TALSTM, and STALSTM, were employed for multivariate water level prediction. The dataset was divided into training and testing sets, with gaps in the data filled using an imputation approach. The models' performance was evaluated using various assessment metrics [2], [15], [17], [18]. These advancements in deep learning-based flood prediction techniques hold promise for effective flood prevention and mitigation efforts.

Considering these implications, the study seeks to contribute in the context of technology as applied to disaster preparedness using Deep Learning Spatiotemporal Forecasting, to reduce the impact of flooding in the Municipality of Nabua. By focusing on the Municipality of Nabua, the study aims to answer several research questions. Firstly, it seeks to determine the specific data requirements necessary for accurate river flood forecasting in the area. This understanding will contribute to the development of a comprehensive forecasting model that can provide reliable predictions. Secondly, the research aims to identify the potential benefits that can be derived from implementing river flood forecasting using deep learning techniques in the Municipality of Nabua. By assessing the outcomes and advantages of such a system, the study can evaluate the practical value and potential improvements that can be achieved through this approach. Lastly, the study intends to explore how the utilization of deep learning in flood forecasting fits within the observation and understanding of river floods. This investigation will provide insights into the compatibility and integration of deep learning techniques within the existing framework of river flood observation, contributing to a more comprehensive understanding of the applicability and effectiveness of this approach.

II. Methodology

A. Study Area

Nabua is considered the catch basin of floodwaters from the neighboring third district of Albay province and the Rinconada areas in Camarines Sur. The Bicol River crosses the central part of Nabua, because majority of Nabua's barangays, including the Poblacion, are located in low-lying locations, the municipality is inevitable to become a catch basin for water runoffs during typhoons and severe rainfall.

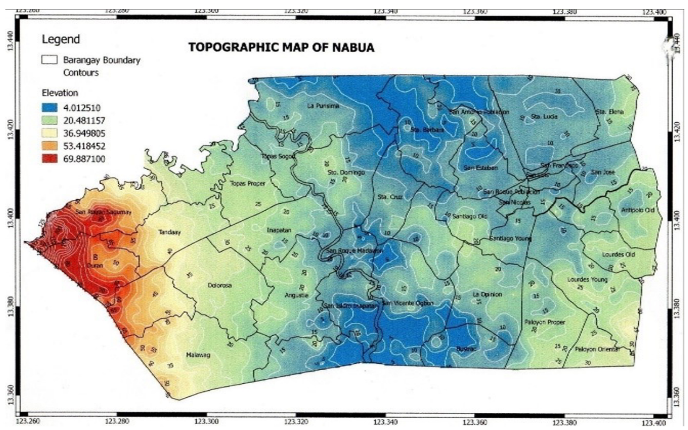


Fig. 1. Topographic Map of Municipality of Nabua [19]

According to the Nabua-MDRRMO record, the historical timeline of the devastating typhoon that affected Nabua shows that flooding is the municipality's main problem, involving the agricultural sector, transportation sector, market/vendor sector, religious sector, academic sector, barangay sector, local folk sector, senior citizen sector, and youth sector.

B. Data Collection

The localized weather monitoring system is composed of two (2) sub-units, the Indoor Control Unit (ICU) and the Outdoor Control Unit (OCU). The ICU is composed of the Arduino Mega2560 which serves as the main controller of the unit, the Anemometer that measures and determines the windspeed and wind direction, and two peripheral devices, the OLED display and SD card module. On the other hand, the OCU comprises the NodeMCU as the main controller and establishes the connection of the whole system to the Access Point, the rain gauge, the water level sensor, the heat and humidity sensor and the real-time clock.

In the current setup, the ICU is deployed at the 4th level of the College of Computer Studies (CCS) building to facilitate the installation of the Anemometer at least 20 feet above the ground. This setup aims to ensure that the anemometer is

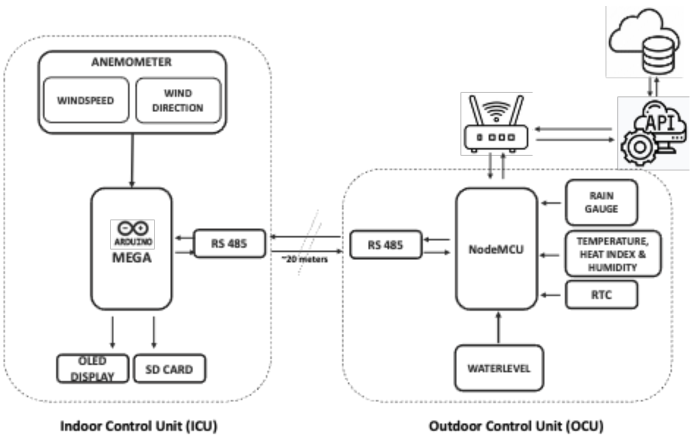


Fig. 2. Localized Weather Monitoring System Architecture

exposed to the wind and is not blocked by the surrounding building and other barriers that would prohibit the optimal sensor data acquisition. Meanwhile, the OCU is deployed at the river canal beside the CCS building. The said river canal is directly connected to the main rivers in the town of Nabua which makes it an ideal and practical location to install the water level sensor which is made up of a waterproof ultrasonic sensor. The rain gauge is also installed in the OCU along with the real-time clock (RTC) that provides the rate at which the rain is accumulated in an hourly and daily basis. Also, a DHT11 sensor was also used to measure the temperature, heat-index and humidity in the area.

The two sub-units can communicate with each other and exchange data via the RS485 modbus. The RS485 modbus allows the 2 units to communicate with each over a long distance up to approximately 1 kilometer via a CAT6 copper cable. The figure below shows how the ICU and OCU share their data and how these data are saved and uploaded to the webserver via the API.

In the figure, the 2 items in blue are sensor data collected by the OCU while the 7 items in red are the collected data from the ICU. The ICU first sends its collected data comprising the windspeed and wind direction to the OCU, the OCU then combines these data with its own collected data namely the timestamp, temperature, heat-index, humidity, rain(hourly), rain(daily) and water level before uploading it to the database. The database therefore receives all the sensor data collected by both the ICU and OCU in a predetermined sequence. After the uploading all these sensor data to the database, the OCU also sends the 7 data it collected to the ICU. The ICU then parses these data and combine it with the wind speed and wind direction in a sequence similar to the ones uploaded in the database. The complete data are then saved into the SD card for backup purposes just in case a network failure occurs. Aside from saving the data to the SD card, the sensor data are also displayed on the OLED. This enables the monitoring of the sensor detection without opening the portal. Also, the display would help us determine if there are anomalies or problems in the system, thereby providing immediate actions to fix the problem.

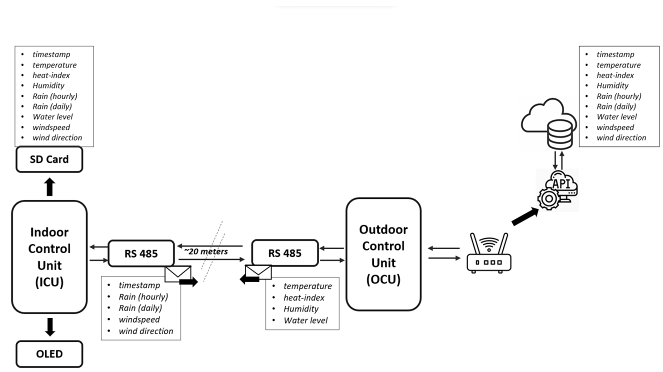


Fig. 3. Data Collection System

In the system, the decision on what sensors to include in each control unit relied on the power requirements by the microcontrollers and each electronic component. The NodeMCU-32s in the OCU is not 5-volt tolerant, therefore, only electronic components that can work on 3.6 volts may be connected in it. Meanwhile, the anemometer requires a 12-volt power supply for both its windspeed and wind direction sensor to operate normally and it produces a maximum output voltage of 5 volts which would burn the NodeMCU-32s instantly but would work perfectly on a 5-volt tolerant Arduino Mega2560.

C. Deep Learning Model

For this project, the researchers employed a Deep Learning method which is the Long-Short Term Memory (LSTM) method, similar to the Recurrent Neural Network. LSTM is designed to create a robust many-to-one model for hydrological time series, comparable to the structure of RNN memory cells' input, self-recurrent connection, forget, and output gates.

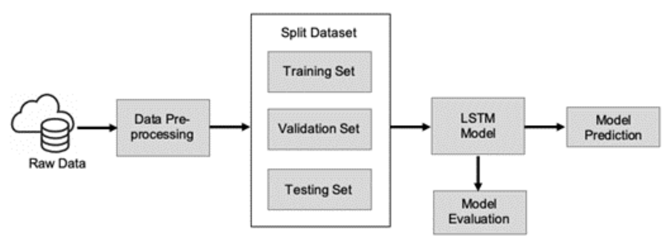


Fig. 4. Water level time series forecasting

Data collection is conducted through a localized weather station, gathering various parameters such as Temperature, Humidity, Heat Index, Hourly Rain, Daily Rain, Water Level, and Wind Speed. Once collected, the data undergoes preprocessing, which involves tasks like file conversion, data type conversion, grouping, outlier detection and removal, handling missing values, category encoding, and normalizing or scaling numerical characteristics. Following preprocessing, a deep learning model is trained and validated to accurately predict outputs based on given inputs. The model learns patterns and relationships from the input features and corresponding target variables (labels), while validation ensures that the model does not overfit the training data. Subsequently, the model is tested and evaluated using a separate dataset that was not utilized during training or validation. This process provides an unbiased assessment of the model's performance in real-life scenarios, determining its effectiveness and reliability.

III. Results and Discussion

A. Localized Weather Monitoring System

Nabua is a municipality in the Philippines' Camarines Sur province. It's in the Rinconada region, which is known for its rugged topography and high rainfall. As a result, Nabua is frequently referred to as a "catch basin" for floodwaters from nearby locations.

Fig. 5. Localized Weather Monitoring System at Balayan Creek

Nabua's topography contributes significantly to its vulnerability to flooding. The municipality is located in a valley, and the neighboring mountains act as a water barrier. When it rains heavily, the water has nowhere to go but to collect in the valley.

Hence, a flood sensor was placed in Balayan Creek in Nabua, Camarines Sur. Similar to the forecasting strategy mentioned in the study of Zhu et. al (2020), which is following the data-driven approach in generating a model, the implementation of a Localized Weather Monitoring System, used to collect and analyze weather data from the said creek.

B. Data Pre-processing

The dataset's characteristics are presented in the table below. Every minute, statistics are collected from the localized weather system from January to March 2023. In total, 5,000 samples were used in the study for analysis. Figure 6 shows the summary of the collated data.

The weather conditions as shown in Figure 6 are noted as Temperature, Humidity, HeatIndex, HourlyRain, DailyRain, WaterLevel, and WindSpeed. The dataset summary depicts the total number of dataset, mean, standard deviation (std), minimum (min), and maximum (max), as well as the lower (25%) and upper (75%) percentile.

Moreover, as part of this phase, the identification of anomalies in the dataset and its possible implications. Data summary on Figure 6, highlights potential outlier values like in the case

	Temperature	Humidity	HeatIndex	HourlyRain	DailyRain	WaterLevel	WindSpeed
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
mean	28.210500	13.831000	26.663132	1.734644	1.734794	3.284398	0.789490
std	3.484517	17.377688	2.747291	35.321506	35.321500	43.090731	1.690719
min	23.700000	5.000000	23.580000	0.000000	0.000000	-23.550000	0.000000
25%	25.700000	6.000000	24.757500	0.000000	0.000000	0.477500	0.060000
50%	26.800000	7.000000	25.690000	0.000000	0.000000	0.720000	0.060000
75%	30.025000	10.000000	27.850000	0.000000	0.000000	0.750000	1.480000
max	39.600000	95.000000	37.120000	770.000000	770.000000	770.000000	62.000000

Fig. 6. Data summary

of WaterLevel where there is irrelevant variance in the readings as shown by the maximum and minimum values vs the mean and percentile values. Figure 7 exhibits the graphical representation of these dataset (WaterLevel), highlighting irregularities in the values.

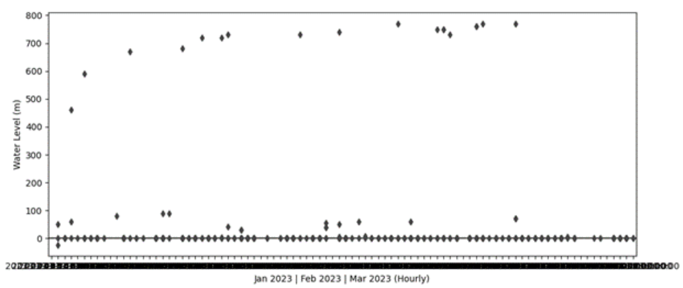


Fig. 7. Outlier Detection (Hourly Water Level)

The value that lies outside most of the other or all other average values are identified. It is first detected within the range of 10 - 100 water levels. Observed is another extremely high value of water level that lies between 450 - 800 water level. The detected outliers signifies an abnormally high level of water that causes floods.

Furthermore, the recorded Daily Water Level, as presented in Fig. 8, has no significant difference during the timeframe of data collection.

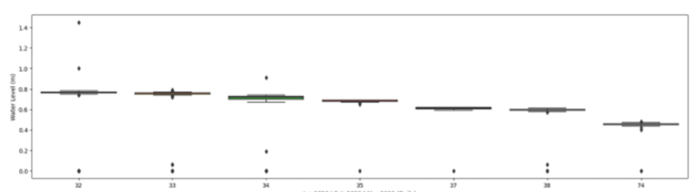
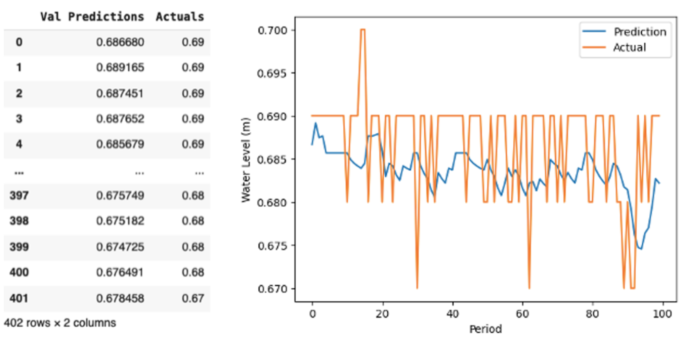


Fig. 8. Outlier Detection (Daily Water Level)

The recorded daily water level was consistent throughout the data collection period, considering the weather conditions during the time of data collection, by which there was very little rainfall.

C. Deep Learning Model

The dataset was divided into three subsets: 80% for training, 10% for validation, and 10% for testing. The specific data split percentages was purposely implemented to establish a balance between having adequate data for training, effective evaluation during the training, and a reliable assessment of the model's performance. The LSTM model was trained on the training set and then used to forecast at a different time on the testing sets. In order to gather insights into how effectively the model conveys the patterns and trends in the dataset, Figures 9 and 10 exhibit the visual comparison of the actual and predicted values during the validation, and testing, respectively.

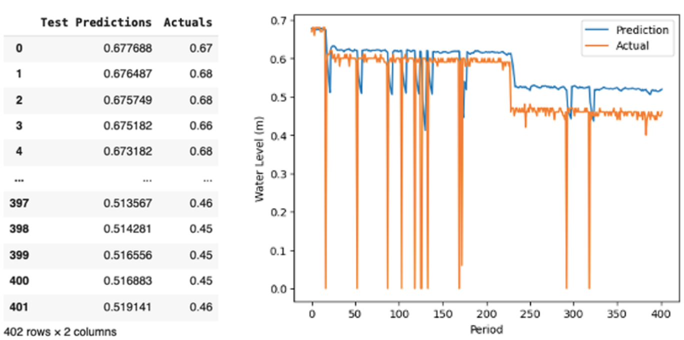


	Val Predictions	Actuals
0	0.686680	0.69
1	0.689165	0.69
2	0.687451	0.69
3	0.687652	0.69
4	0.685679	0.69
...	...	...
397	0.675749	0.68
398	0.675182	0.68
399	0.674725	0.68
400	0.676491	0.68
401	0.678458	0.67

402 rows × 2 columns

Fig. 9. Data Validation - Visual Comparison of the actual values vs the predicted values

From the visualization, the orange line represents the actual values, while the blue line denotes the predicted trend. As measured by visual inspection, the proximity of these lines suggests the effectiveness of the LSTM-based model, indicating the model's high-degree of accuracy. Further, the dataset was also resampled to test the LSTM model's ability to forecast water level at different time intervals: fifteen and thirty minutes, respectively. Since the outliers were removed, there were gaps in the resampled dataset resulting to NAN values, which were filled by employing an interpolation technique.



	Test Predictions	Actuals
0	0.677688	0.67
1	0.676487	0.68
2	0.675749	0.68
3	0.675182	0.66
4	0.673182	0.68
...	...	...
397	0.513567	0.46
398	0.514281	0.45
399	0.516556	0.45
400	0.516883	0.45
401	0.519141	0.46

402 rows × 2 columns

Fig. 10. Testing - Visual Comparison of the actual values vs the predicted values

D. Model Evaluation

To supplement the visual observation on the LSTM model's performance, the calculation of the Mean Absolute Error (MAE) and Root Mean Square Error were used. The root mean square error (RMSE) is a measurement of the average difference between projected and actual values. The square root of the mean of the squared errors is used to calculate it. Table 1 summarizes the MAE and RSME results between the predicted and actual values during the training and testing phase.

TABLE I
EVALUATION OF LSTM MODEL IN THREE DIFFERENT TIME INTERVALS

Evaluation Metric		1 Minute	15 Minutes	30 Minutes
MAE	Training	0.0568	0.0095	0.0217
MAE	Testing	0.0612	0.0160	0.0097
RMSE	Training	0.1366	0.0258	0.0384
RMSE	Testing	0.1151	0.0162	0.0101

Table 1 shows the results of the evaluation metrics used in three different time intervals: every minute, every fifteen minutes, and every thirty minutes.

For the MAE, the closer it is to 0, the more accurate the model is, the LSTM testing model with value of 0.0097 is the most accurate at thirty minutes. Moreover, the RSME at thirty minutes has the best value for the LSTM testing model, with the value of 0.0101 for it is closer to 0. As observed in the results, the longer the time intervals, the MAE and RMSE of the LSTM model becomes more accurate.

IV. CONCLUSION

The study's findings indicate that although data collection is utilized to create a model for predicting river floods, the short timeframe of data collection limits the representation of predictors, potentially affecting the accuracy of the model for flood prediction. Despite this limitation, deep learning models show promise in accurately predicting flood risks. However, there are obstacles to overcome, including addressing challenges related to data availability, quality, interpretability, and computational resources. By addressing these challenges, deep learning can become a valuable tool for flood risk management, supporting decision-making processes, and enhancing preparedness and resilience in the face of flooding events. It's important to note that the accuracy of flood prediction varies depending on factors such as flood type, location, and available data. Models tend to be more accurate for large, slow-moving floods compared to small, fast-moving floods, and accuracy is also influenced by the amount of available data in the specific flood-prone areas.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to all individuals and organizations who have contributed to the successful completion of this research especially the Department of Science and Technology - Philippine Council for Industry, Energy and Emerging Technology Research and Development (DOST-PCIEERD) for the funding of this project. We extend our appreciation to the residents and local authorities of Nabua for their support and cooperation throughout the data collection process. We are grateful for the guidance and mentorship provided by our research advisors, as well as the institutions and funding agencies that supported this project. We also acknowledge the contributions of researchers in the field and the unwavering support of our families, friends, and colleagues. We recognize that this research would not have been possible without the collective efforts of everyone involved.

REFERENCES

[1] "Information on disaster risk reduction of the member countries," Asian Disaster Reduction Center (ADRC), https://www.adrc.asia/nationinformation.php?NationCode=608, 2019.

[2] Centre for Research on the Epidemiology of Disasters, "2021 Disasters in numbers", DOI:https://doi.org/10.1787/eee82e6e-en, 2021.

[3] Earth Networks, "What is a flood? - earth networks," Earth Networks, https://www.earthnetworks.com/flooding/, 2023.

[4] R. Felipe and N. De Aguiar, "Spatiotemporal Forecasting At Scale", August 2019.

[5] J. Gray, S. Lloyd, S. Healey, and A. Opdyke, "Urban and rural patterns of typhoon mortality in the Philippines," Progress in Disaster Science, https://doi.org/10.1016/j.pdisas.2022.100234, March 2022.

[6] Ma. A. Mier-Manjares, "Camarines Sur villages still submerged in Flood," INQUIRER.net, https://newsinfo.inquirer.net/1358068/, November 2020.

[7] A. Tuovila, "Forecasting: What it is, how it's used in business and investing," Investopedia, https://www.investopedia.com/terms/f/forecasting.asp, May 2023.

[8] Statista Research Department, "Philippines: Risk index for natural disasters 2022," Statista, https://www.statista.com/statistics/921036/philippines-risk-index-for-natural-disasters/, May 2023.

[9] N. M. Roberts Westra, S., H. J. Fowler, J.P. Evans, L. V. Alexander, P. Berg, F. Johnson, E. J. Kendon, G. Lenderink, "Future changes to the intensity and frequency of short-duration extreme rainfall", in Review of Geophysics, Volume 52, Issue 3, September 2014, pp. 522-555.

[10] "Floods," PAGASA, https://www.pagasa.dost.gov.ph/learning-tools/floods.

[11] M. Allaire, "Socio-economic impacts of flooding: A review of the empirical literature," Water Security, vol. 3, pp. 18–26, 2018. doi:10.1016/j.wasec.2018.09.002

[12] Y. Zhu, J. Feng, L. Yan, T. Guo, and X. Li, "Flood prediction using rainfall-flow pattern in data-sparse watersheds," IEEE Access, vol. 8, pp. 39713–39724, 2020. doi:10.1109/access.2020.2971264

[13] M. Yu et al., "Spatiotemporal Event Detection: A Review," International Journal of Digital Earth, vol. 13, no. 12, pp. 1339–1365, 2020. doi:10.1080/17538947.2020.1738569

[14] S. Dhamodaran and M. Lakshmi, "Design and analysis of spatial-temporal model using hydrological techniques," 2017 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC), 2017. doi:10.1109/iccpeic.2017.8290333

[15] Y. Wu, Y. Ding, Y. Zhu, J. Feng, and S. Wang, "Complexity to forecast flood: Problem definition and spatiotemporal attention LSTM solution," Complexity, vol. 2020, pp. 1–13, 2020. doi:10.1155/2020/7670382

[16] H. Wang, C. Wang, Y. Wang, X. Gao, and C. Yu, "Bayesian forecasting and uncertainty quantifying of stream flows using Metropolis–Hastings Markov chain Monte Carlo algorithm," Journal of Hydrology, vol. 549, pp. 476–483, 2017. doi:10.1016/j.jhydrol.2017.03.073

[17] Y. Ding, Y. Zhu, Y. Wu, F. Jun, and Z. Cheng, "Spatio-temporal attention LSTM model for flood forecasting," 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2019. doi:10.1109/ithings/greencom/cpscom/smartdata.2019.00095

[18] F. Noor et al., "Water level forecasting using spatiotemporal attention-based long short-term Memory Network," Water, vol. 14, no. 4, p. 612, 2022. doi:10.3390/w14040612

[19] R. B. Sotto, "Flood Control Measures in One Municipality in Camarines Sur, Philippines: Bases for Community-Based Flood ControlInterventions," Journal of Education, Management and Development Studies, pp. 30–39, May 2022. doi:0.52631/jemds.v2i2.115