**METHODOLOGY**

**1. Data Collection**

Primary Meteorological Data:

The source of the main data in this study is the Trans-African Hydro-Meteorological Observatory-TAHMO, operating a network of three stations across Kampala. This dataset provided high-resolution records of relative humidity, precipitation, maximum temperature, minimum temperature, and mean temperature, which we used for the flood prediction. These are the meteorological and hydrological variables that precede flood events and have been deemed essential for an accurate urban flood forecast by Musoke et al. (2022) and the International Rescue Committee (2018).

Additional Data Sources:

Historical flood data from the Global Disaster Alert and Coordination System For insight into patterns of rainfall intensities, critical in flood prediction, the data uses Intensity-Duration-Frequency curves, tailored for urban environments (Madsen et al., 2009). Also, the degradation of natural flood management systems, such as Kampala's Lubigi wetland system, as a result of human activities, such as wetland encroachment, has been included since such activities have contributed to increasing the risks of urban flooding (Gideon & Bernard, 2018).

**2. Data Preprocessing**

The preprocessing is also another important aspect of this project. We merged  the satellite meta data and station data. Missing data values were imputed, and outliers were removed to ensure dataset quality. Temporal alignment across variables captures time-sensitive relationships critical for flood prediction. Standardization of variables via normalization ensures uniform scales, enhancing the model's performance. Furthermore, categorical variables were converted into numeric forms, and detailed correlation analyses were performed to uncover hidden relationships between variables.

**3. Feature Engineering**

Key Predictive Variables:

Feature engineering focuses on the identification of key variables that are indicative of flood risk, including cumulative precipitation, temperature fluctuations, changes in relative humidity, and wind speed. These are very important variables that will capture real conditions at any given time that lead to flooding, as they drive the pattern of rainfall and flood potential. According to Babar et al. (2022, p. 197), "The relationship found is that by increasing humidity and cloudiness, the possibility of rain increases, whereas temperature and radiation from the sun have an inverse relationship". This highlights the importance of analyzing these interconnected factors to enhance the accuracy of flood prediction models.

Temporal and Seasonal Indicators:

From the existing Kampala researchers, the group assured the collection and aims of the data. Seasonal features are incorporated to capture Kampala’s bimodal rainfall pattern, which aligns with its two main rainy seasons. These features improve the model's adaptability to seasonal variations (Mukwaya et al., 2012; Zhong et al., 2019).

**4. Model Selection**

For flood prediction in Kampala, we utilized Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models. RNN’s main distribution lies in its ability to model and process sequential data, unlike traditional neural networks that handle inputs independently. Its recurrent nature and hidden state give it a significant edge in tasks requiring temporal understanding, though it comes with challenges in training and efficiency for long sequences. LSTMs are superior to RNNs because they can capture long-term dependencies and mitigate the vanishing gradient problem. This makes LSTM more suitable for handling hourly data, which involves extensive temporal dependencies. For this flood prediction study, LSTMs are optimal due to their robust gating mechanisms, which enable them to learn from detailed, sequential data. LSTM is particularly effective for managing non-linear relationships and avoiding vanishing gradient issues, making it ideal for rainfall prediction tasks (Haidar & Verma, 2018; Babar et al., 2022).

This study differs from previous ones in that the models used are recurrent neural networks (RNN) and long short-term memory (LSTM), which are derived from hourly meteorological and hydrological data. Using hourly meteorological and hydrological data enhances the precision of flood predictions. With such a high temporal resolution, models such as RNN and LSTM are able to identify short-term trends that are crucial to accurately predicting sudden flood events. As a result, it improves the ability to provide timely and reliable warnings, reducing the risk of flood-related damage.

**5. Model training**

The dataset was divided into subsets for testing (30%) and training (70%), each serving a distinct purpose in model development (Hastie et al., 2009). The training set is used to adjust the model's internal parameters by minimizing the error during learning. The validation set guides the tuning of hyperparameters, such as learning rates, dropout rates, and hidden layer configurations, helping to prevent overfitting and improve generalization. The testing set, at the end, verifies that the model generalizes well when exposed to previously unseen data and works reliably in real-world situations. This formalized approach balances the fitting of data with predictive accuracy.

**6. Model Evaluation and Validation**

The performance was evaluated on several key metrics to ensure accuracy and reliability such as the Root Mean Square Error for the continuous rainfall estimations by comparing the predicted and actual values, and the Area Under the Receiver Operating Characteristic Curve for the flood classification problem, which measures the ability of the model to discern between the two classes-a flood and a no-flood event. Additionally, the F1 Score balances precision (correct flood predictions) and recall (detecting most flood events), ensuring the model provides dependable flood event predictions (Chollet, 2017).

To improve robustness and minimize overfitting, K-fold cross-validation was employed, splitting the dataset into multiple parts and testing the model on each one. This ensures the model performs consistently across different data subsets. For real-world validation, recent flood events recorded in the GDACS database were used as external datasets, allowing the model to demonstrate its effectiveness in actual flood scenarios (Rentschler & Salhab, 2020).

References

Babar, M., Akhtar, S., & Khan, T. (2022). A deep learning-based rainfall prediction for flood management. *Proceedings of the 17th International Conference on Emerging Technologies (ICET 2022)*, 196–199. IEEE. <https://doi.org/10.1109/ICET56601.2022.10004663>

Chollet, F. (2017). *Deep learning with Python*. Manning Publications. <https://www.manning.com/books/deep-learning-with-python>

Haidar, A., & Verma, B. (2018). Monthly rainfall forecasting using one-dimensional deep convolutional neural network. *IEEE Access, 6*, 69053–69063. <https://doi.org/10.1109/ACCESS.2018.2880113>

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>

Nevo, I., & Fishman, E. (2021). Addressing data gaps in hydrological modeling for urban flood management. *Natural Hazards Review, 22*(1), 05020015. <https://doi.org/10.1061/(ASCE)NH.1527-6996.0000455>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.[**https://www.deeplearningbook.org**](https://www.deeplearningbook.org/)

Madsen, H., et al. (2009). Update of regional intensity–duration–frequency curves in Denmark: Tendency towards increased storm intensities. *Atmospheric Research, 92*(3), 343–349. <https://doi.org/10.1016/j.atmosres.2009.01.013>

Mukwaya, P., Sengendo, H., & Lwasa, S. (2012). Enhancing security and resilience of low-income communities to climate change in growing cities: An assessment of flood management and planning regimes in Kampala City, Uganda. In *Climate change adaptation in practice: From strategy development to implementation* (pp. 371–386). Springer**.** [**https://doi.org/10.1007/978-3-642-28626-1\_26**](https://doi.org/10.1007/978-3-642-28626-1_26)

Rentschler, J., & Salhab, M. (2020). People in harm’s way: Flood exposure and poverty in 189 countries. *World Bank Group*. Retrieved from[**https://www.worldbank.org/**](https://www.worldbank.org/)

Zhong, R., & Wang, J. (2019). Seasonal flood prediction using machine learning techniques. *Environmental Modelling & Software, 123,* 104520.[**https://doi.org/10.1016/j.envsoft.2019.104520**](https://doi.org/10.1016/j.envsoft.2019.104520)