**LITERATURE REVIEW, DATA REVIEW AND TECHNOLOGY REVIEW FOR PREDICTIVE FLOOD MODELING FOR RESILIENT COMMUNITIES IN SIERRA LEONE.**

**Literature Review:**

1. **Introduction:**

Key issues pertaining to community resilience, environmental sustainability, and disaster preparedness are interconnected with the development of a robust flood prediction model for Sierra Leone during the rainy season. This research project is important because it could lessen the devastating effects of flooding on infrastructure, agriculture, and communities, especially in areas that are susceptible to climate change and extreme weather. Floods are the most destructive natural disasters, wreaking havoc on human life, infrastructure, agriculture, and the socioeconomic system. Governments are therefore under pressure to create reliable and accurate maps of flood risk areas, as well as to plan for long-term flood risk management that focuses on prevention, protection, and preparedness. Flood prediction models are crucial for hazard assessment and extreme event management. Robust and accurate prediction are extremely beneficial to water resource management strategies, policy recommendations and analysis, and future evacuation modeling [1].

The main research questions concern how well predictive modeling works to predict and control flood events. It is critical to address these issues as floods during the rainy season become more frequent and intense in Sierra Leone [2], as they do in many other areas. To develop proactive and successful strategies for disaster risk reduction, it is essential to comprehend the underlying patterns and dynamics governing these events.

A thorough review of existing literature is essential for contextualizing and situating our research within the larger scientific landscape. Because of the dynamic nature of climate-related challenges, previous studies, methodologies, and findings in flood prediction and climate adaptation must be thoroughly investigated. We can gain valuable insights, identify gaps in knowledge, and inform the methodology and approach of our own predictive modeling efforts by examining the successes and limitations of previous research.

Furthermore, the review of literature serves as the foundation for developing a theoretical framework that will guide the understanding of the complex interplay between environmental variables, climatic trends, and flood occurrences. I intend to leverage the collective knowledge gathered in the field to improve the accuracy and applicability of our flood prediction model for the specific context of Sierra Leone by drawing on the experiences of similar regions and studies. In essence, this literature review is more than just a survey of previous works; it is an important preliminary step in the journey to meaningfully contribute to sustainable development goals, climate action, and the resilience of communities in flood-prone areas. I strive to establish a solid foundation for my research by meticulously exploring the existing body of knowledge, ensuring that my flood prediction model not only addresses immediate concerns but also aligns with the broader global agenda for a more sustainable and resilient future.

1. **Organization:**

***2.1 Flood Prediction Using Machine Learning Models: Literature Review***

The paper evaluated physically based and numerical models, highlighting their limitations in short-term prediction as well as the difficulties associated with their complexity and data requirements. Despite advances in hybridization and advanced flow simulations, these models may still be lacking in accuracy and robustness.

The shortcomings of physically based and statistical models pave the way for data-driven models, particularly machine learning (ML), to be adopted. ML is a promising tool capable of generating flood nonlinearity solely from historical data, without the need for in-depth knowledge of underlying physical processes. When compared to traditional models, ML methods are praised for their quick development, low computation cost, fast training, and high performance [1].

The performance of various ML algorithms, such as artificial neural networks (ANNs), neuro-fuzzy, support vector machine (SVM), and support vector regression (SVR), is thoroughly evaluated in this paper. It demonstrates the utility of ML in both short-term and long-term flood forecasting. The combination of ML methods and other techniques, such as soft computing, numerical simulations, and physical models, has been identified as a strategy for developing more robust and efficient flood prediction models.

There are some issues with ML models, such as their reliance on training data and the generalization problem. The review emphasizes the importance of robust data enrichment as well as the importance of carefully considering the capabilities of each ML algorithm in relation to the type and amount of available training data as well as the prediction task. Based on flood resource variables, such as water level, river flood, soil moisture, rainfall–discharge, precipitation, river inflow, peak flow, river flow, and numerous others, the review categorizes flood prediction applications. The process generates a long list of search queries by combining ML techniques with flood resource variables.

An overview of the state of ML modeling for flood prediction is provided by the literature review, which divides ML techniques into short- and long-term prediction models. Technical explanations of machine learning techniques are presented in the paper along with a synopsis of their flood prediction applications.

This paper discusses the classification of prediction models, performance analysis of over 6000 articles, and identifies 180 influential articles for detailed review. Four major trends for improving prediction quality were identified in this paper including novel hybridization, data decomposition, ensemble methods, and add-on optimizer algorithms.

* 1. ***Urban flood forecasting using a hybrid modeling approach based on a deep learning technique.***

With global flood risks increasing due to climate change, accurate forecasting has become critical for effective disaster prevention and management. To improve urban flood forecasting, this paper proposes a hybrid modeling approach that combines a rainfall-runoff model and Long Short-Term Memory (LSTM) deep learning techniques [3]. The study, which focuses on small urban river basins, employs virtual rainfall scenarios that combine rainfall amounts, durations, and time distributions. The proposed method yields promising results, with a high accuracy correlation coefficient of 0.9 and a Nash-Sutcliffe efficiency of 0.8, providing a practical solution for flood occurrence and timing predictions based solely on forecasted rainfall data [3].

This study emphasizes the importance of flood prediction, particularly in the context of climate change. Despite their accuracy, physical-based models face data availability and computational efficiency challenges. Because of their lower computational costs and ability to capture complex relationships between input and output data, data-driven models, particularly ANN and LSTM, emerge as promising alternatives. The research contributes to the growing interest in LSTM models by emphasizing their successful applications in a variety of domains, including rainfall-runoff simulations.

***2.3 Flood Prediction Model using Artificial Neural Network***

The focus of this paper is to apply optimized ANN for next day’s river water level forecasting by determination of suitable input parameters and designing the best network architecture. The study reported in this article led to the conclusion that MLP type network consistently performed better compared to other network. Among the water level prediction after 24 hours, 48 hours and 72 hours; prediction after 24 hours performs well. Therefore, the researchers’ ANN model with MLP was used only for predicting next day (24 hours) water level.

1. ***Summary and Synthesis:***
2. The first literature provides a thorough and insightful overview of ML models in flood prediction. It effectively communicates the importance of ML in addressing traditional model challenges and presents a systematic approach to reviewing a large body of literature. The paper successfully identified key trends, challenges, and future directions in the field of flood prediction using ML, providing valuable insights for hydrology and machine learning researchers and practitioners.
3. The second study used a hybrid modeling approach to show a significant improvement in flood forecasting, particularly for small urban river basins. The method overcomes data limitations and accurately predicts flooding time under various rainfall conditions by combining a rainfall-runoff model with LSTM. The study adds valuable insights into improving flood prediction accuracy and emphasizes the proposed approach's practical applicability for real-time flood forecasting and warnings.
4. The third body of literature employed Artificial Neural Networks (ANNs) for the critical task of forecasting next-day river water levels. The study unequivocally demonstrates the superior performance of the Multi-Layer Perceptron (MLP) network over other network types by meticulous consideration of input parameters and network architecture design. Notably, an analysis of water level predictions at various time horizons revealed that the 24-hour prediction consistently outperformed predictions for the following time intervals, namely 48 and 72 hours. As a result, the ANN model, specifically the MLP architecture, was used selectively for precise and effective prediction of next-day (24 hour) water levels. This focused approach not only improves the forecasting model's accuracy but also emphasizes the importance of tailoring ANN architectures to the specific situation.
5. **Conclusion:**

The review of literature provides a thorough overview of flood prediction models, emphasizing the critical role of Machine Learning (ML) in overcoming the challenges posed by traditional models. These findings highlight the importance of accurate flood prediction models, especially considering climate change. The goal of the research project is to significantly contribute to flood prediction models in Sierra Leone by using ML techniques to improve accuracy and align with global agendas for sustainable development and community resilience in flood-prone areas.

**Data Research:**

1. **Introduction:**

The goal of the data review is to investigate data to create a predictive flood model for Sierra Leone. A thorough investigation serves as the foundation for developing a reliable flood prediction model. For this project, data will be sourced from weather observations stations to create a comprehensive dataset. We will use historical weather data for Sierra Leone for the past forty (40) years and will be sourced from Open Weather (<https://openweathermap.org/> ) due to data constraints from the Sierra Leone Meteorological Agency. This extensive dataset, available in CSV format, contains critical information on rainfall patterns and different weather conditions of Sierra Leone measured hourly since January 1, 1979. By addressing missing values and optimizing the dataset for machine learning model training, data research ensures the accuracy of predictions.

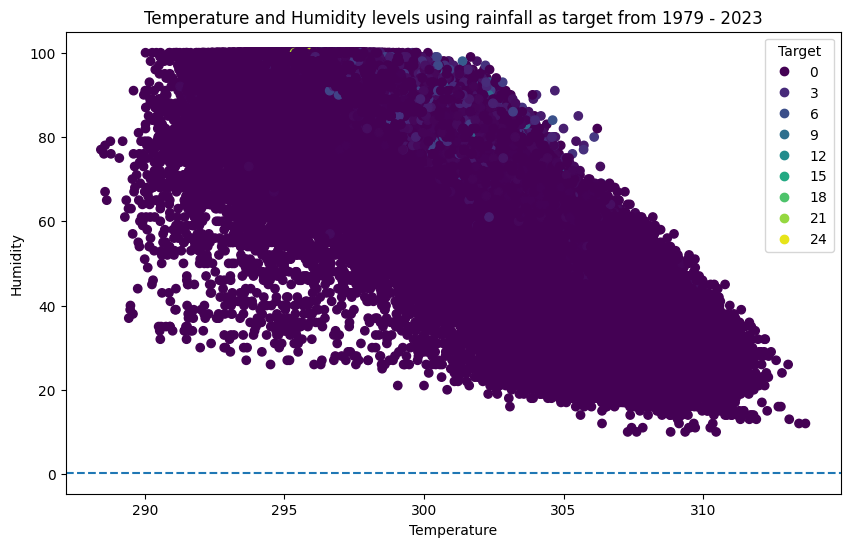
1. **Organization:**

The data I will be using for the model training of the predictive flood modelling for Sierra Leone records different weather parameters such as date, time zone, longitudes and latitudes, temperature, dew point, amount of rain in one hour, amount of rain in 3 hours, visibility, windspeed wind gust clouds, main weather and weather description observed on current timestamp, hourly and daily basis from January 1, 1979 to December 2, 2023 [4].

1. **Data Description:**

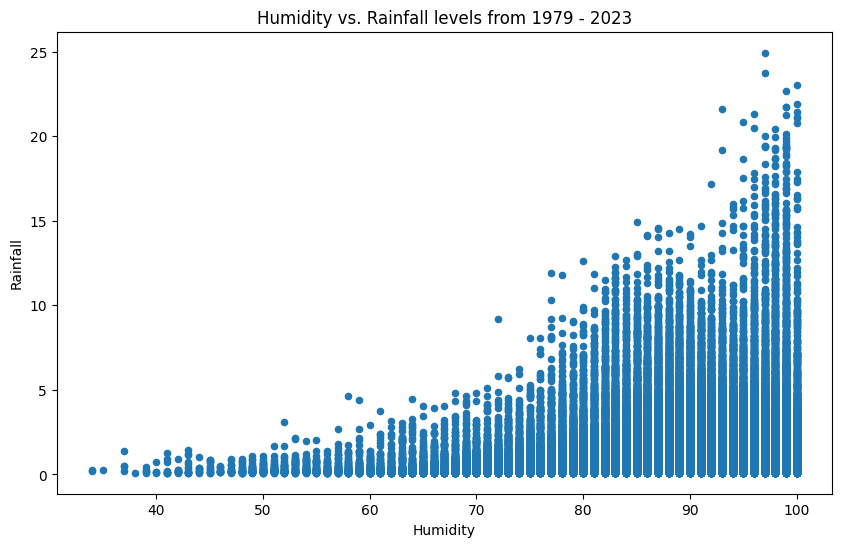
The data for my flood prediction model is sourced from OpenWeather (<https://openweathermap.org/> ), a team of IT experts and data scientists that has been practicing deep weather data science since 2014. For each point on the globe, OpenWeather provides historical, current, and forecasted weather data via light-speed APIs. Headquarters in London, UK, specifically historical weather forecast on current, hourly, and daily basis dataset for all coordinates in Sierra Leone encompassing temperature, minimum temperature, maximum temperature, feels like, Pressure, Humidity, Clouds, Weather conditions, Rain, Snow, Dew point, Visibility, Wind (speed, direction, gust) records spanning 1st January 1979 to 2nd December 2023. It includes information on the date and time the weather was recorded, time zone, longitudes and latitudes, city name, temperature, minimum temperature, maximum temperature, feels like, Pressure, Humidity, Clouds, Weather conditions, Rain, Snow, Dew point, Visibility, Wind (speed, direction, gust). The data is accessible in CSV format and serves as a valuable resource due to its comprehensive coverage of historical weather patterns of Sierra Leone as a whole. This dataset's relevance to my project lies in its alignment with the geographical and climatic conditions, offering a basis for training and validating the flood prediction model. As earlier mentioned, the data is sourced from Open Weather accessible at <https://home.openweathermap.org/marketplace/my_orders> [4], CSV format with a size of 63.2 megabytes.

1. **Data Analysis and Insights:**
2. The data contains twenty-eight (28) columns and has three hundred and ninety-three thousand seven hundred and seventy-three (393,773) records (rows) describing the weather and climatic condition of Sierra Leone. It can be observed from the data that there are weather descriptions for every forecast and the descriptions include, scattered clouds, broken clouds, overcast clouds, light rain, moderate rain, heavy intensity rain. These descriptions help to find relevant features in the data that are pertinent to predicting flood events.
3. Rainfall is recorded in two stages, firstly a record of rainfall within one hour is recorded and if it rains over one hour, it is recorded as rainfall in 3 hours, minimum and maximum temperatures are read and averaged to give a temperature of a particular place. Pressure, humidity, sea level, wind speed and wind gust cloud condition, the degree of wind and the condition of snow are all recorded.
4. It can be observed from the data set that rains are frequent from the months of May to October each year with records of high rainfall in June, July and August and moderate rain in September and October. During these distinct wet seasons, May to October, the region experiences a significant portion of the annual rainfall falls. During the remaining months, particularly December to March, dry spells are observed.
5. Some years, particularly in the late 2010s, exhibit exceptionally high rainfall. Examples include 2016, 2017, 2018, and 2019 etc.
6. In recent decades (2000-2018), there seems to be an increasing trend in annual rainfall, with several years exceeding 3500 mm.
7. The year 2017 stands out with exceptionally high rainfall, contributing to severe floods in the region.



A screenshot of a computer screen

Description automatically generated



**Conclusion:**

1. The dataset, comprising 28 columns and 393,773 records, offers a detailed depiction of the weather and climatic conditions, enabling valuable insights for flood prediction.
2. One significant observation from the data is the presence of distinct wet seasons from May to October, marked by frequent rains, with June, July, and August experiencing particularly high rainfall and September and October witnessing moderate rain. Conversely, dry spells are evident during the remaining months, notably from December to March. This temporal analysis is crucial for anticipating and preparing for flood events during the wet seasons.
3. The dataset highlights specific years, notably in the late 2010s, such as 2016, 2017, 2018, and 2019, exhibiting exceptionally high rainfall. The year 2017 stands out as it contributed to severe floods in the region. This emphasizes the need for the predictive flood model to be sensitive to such anomalous years and capable of providing timely warnings to mitigate the impact of extreme weather events.
4. The dataset, with its rich coverage of weather parameters and alignment with Sierra Leone's geographical and climatic conditions, serves as a valuable resource for training and validating the flood prediction model.
5. The insights derived from the data analysis not only contribute to the development of a reliable flood prediction model but also facilitate informed decision-making and preparedness measures for the community in Sierra Leone. This project, rooted in thorough data research and analysis, holds the potential to significantly enhance the region's resilience to flood events and minimize their adverse effects.

**Technology Review:**

1. **Introduction:**

The technology review is critical for informed model development decision-making because it aims to investigate the importance of machine learning algorithms in building flood prediction models while ensuring precision and scalability. The review serves as a guideline for the selection and integration of cutting-edge tools, allowing for the adoption of methodologies required for a cutting-edge flood prediction model. Its importance lies in bridging the gap between project objectives and technological feasibility, allowing for the identification of optimal tools that improve adaptability, accuracy, and applicability to Sierra Leone's environmental dynamics.

1. **Technology Overview:**

For this project I will use the Machine Learning algorithms for predictions:

1. ***Logistic Regression***

**Purpose**: Predicts the probability of an event occurring.

**Key Features**: Binary outcome prediction, suitable for classification tasks.

**Common Usage in Relevant Fields:** Widely used in fields such as finance for credit scoring, healthcare for disease prediction, and, in this context, predicting flood occurrences.

1. ***K-Nearest Neighbors (KNN):***

**Purpose**: Classifies data based on similarity to its neighbors.

**Key Features**: Non-parametric, simple implementation, effective for pattern recognition.

**Common Usage in Relevant Fields:** Applied in various domains, including recommendation systems, image recognition, and environmental modeling like predicting floods based on similar historical patterns.

1. ***Support Vector Machine Algorithm (SVM):***

**Purpose:** Classifies data by finding the hyperplane that best separates classes.

**Key Features:** Effective in high-dimensional spaces, versatile for classification tasks.

**Common Usage in Relevant Fields**: Commonly used in image classification, text categorization, and, in the context of flood prediction, for classifying regions prone to flooding based on historical data.

1. ***Decision Tree Algorithm***

**Purpose:** Makes decisions by mapping possible outcomes of a decision in a tree-like structure.

**Key Features:** Intuitive, interpretable, and capable of handling both numerical and categorical data.

**Common Usage in Relevant Fields**: Employed in finance for decision-making, in healthcare for diagnosis, and, in this project, for creating a decision structure to predict the likelihood of floods based on various factors.

1. ***Random Forest Classifier***

**Purpose:** Utilizes a tree-like structure to map potential decision outcomes, facilitating decision-making processes.

**Key Features:** Characterized by intuitiveness, interpretability, and the ability to handle both numerical and categorical data seamlessly.

**Common Usage in Relevant Fields:** Frequently employed in finance for decision-making scenarios, in healthcare for diagnostic purposes, and, notably in this project, utilized to construct a decision framework predicting the likelihood of floods based on diverse factors.

1. **Relevance to the Project:**
2. Machine learning algorithms excel in handling complex datasets. Logistic Regression and Decision Tree provide simplicity and interpretability, crucial for understanding complex relationships in rainfall patterns and flood occurrences.
3. K-Nearest Neighbors and Support Vector Machine, with their ability to consider spatial proximity and patterns, are instrumental in spatial analysis—essential for predicting floods in specific regions based on historical data.
4. Decision Tree algorithm is adept at capturing interactions between various environmental variables, enabling an understanding of the factors influencing flood probabilities.
5. Logistic Regression's adaptability to binary outcomes, KNN's flexibility in handling various data types, SVM's effectiveness in high-dimensional spaces, and Decision Tree's versatility make them suitable for the diverse and dynamic environmental conditions of Sierra Leone.
6. Random Forest's ensemble learning, adaptability to complex climatic data, and feature importance analysis make it a pertinent choice for enhancing the accuracy and interpretability of the predictive flood model, particularly in addressing missing data and capturing temporal patterns in Sierra Leone.

**Use Cases and Examples:**

1. Prediction of Flood in Bangladesh using k-Nearest Neighbors Algorithm [5].
2. Hybrid model for movie recommendation system using content K-nearest neighbors and restricted Boltzmann machine [6].
3. Spatial prediction of flood in Kuala Lumpur city of Malaysia using logistic regression [7].
4. Prediction of Flood in Barak River using Hybrid Machine Learning Approaches: A Case Study [8].
5. An Intelligent Early Flood Forecasting and Prediction Leveraging Machine and Deep Learning Algorithms with Advanced Alert System [9].

**Conclusion:**

The technology review emphasizes the importance of machine learning algorithms in developing a reliable flood prediction model for Sierra Leone. The algorithms chosen—Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Decision Tree—offer distinct advantages critical for addressing the region's complex flood prediction challenges.

Logistic Regression and Decision Tree bring simplicity and interpretability, vital for comprehending intricate relationships in rainfall patterns and flood occurrences.

K-Nearest Neighbors and Support Vector Machine excel in spatial analysis, essential for predicting floods in specific regions based on historical data.

Logistic Regression's adaptability, KNN's flexibility, SVM's effectiveness in high-dimensional spaces, and Decision Tree's versatility make them well-suited for Sierra Leone's diverse and dynamic environmental conditions.

Random Forest's ensemble learning, adaptability to complex climatic data, and feature importance analysis make it a pertinent choice for enhancing the accuracy and interpretability of the predictive flood model, particularly in addressing missing data and capturing temporal patterns in Sierra Leone.

**References:**

[1] A. Mosavi, P. Ozturk, and K.-W. Chau, “Flood Prediction Using Machine Learning Models: Literature Review.”

[2] “Reflections on Sierra Leone’s mudslide disaster - BBC News.” Accessed: Nov. 20, 2023. [Online]. Available: https://www.bbc.com/news/world-africa-40973539

[3] H. Moon, S. Yoon, and Y. Moon, “Urban flood forecasting using a hybrid modeling approach based on a deep learning technique,” *Journal of Hydroinformatics*, vol. 25, no. 2, pp. 593–610, Mar. 2023, doi: 10.2166/hydro.2023.203.

[4] “OpenWeather Marketplace.” Accessed: Dec. 09, 2023. [Online]. Available: https://home.openweathermap.org/marketplace/my\_orders

[5] N. Gauhar, S. Das, and K. S. Moury, “Prediction of Flood in Bangladesh using k-Nearest Neighbors Algorithm,” *International Conference on Robotics, Electrical and Signal Processing Techniques*, pp. 357–361, 2021, doi: 10.1109/ICREST51555.2021.9331199.

[6] D. K. Behera, M. Das, S. Swetanisha, and P. K. Sethy, “Hybrid model for movie recommendation system using content K-nearest neighbors and restricted Boltzmann machine,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 1, pp. 445–452, Jul. 2021, doi: 10.11591/IJEECS.V23.I1.PP445-452.

[7] A. Tella, M. R. Ul Mustafa, A. O. Balogun, C. J. Okolie, I. B. Yamusa, and M. B. Ibrahim, “SPATIAL PREDICTION OF FLOOD IN KUALA LUMPUR CITY OF MALAYSIA USING LOGISTIC REGRESSION,” in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, International Society for Photogrammetry and Remote Sensing, Feb. 2023, pp. 363–369. doi: 10.5194/isprs-archives-XLVIII-4-W6-2022-363-2023.

[8] A. Sahoo, S. Samantaray, and D. K. Ghose, “Prediction of Flood in Barak River using Hybrid Machine Learning Approaches: A Case Study,” *Journal of the Geological Society of India*, vol. 97, no. 2, pp. 186–198, Feb. 2021, doi: 10.1007/S12594-021-1650-1/METRICS.

[9] I. M. Hayder *et al.*, “An Intelligent Early Flood Forecasting and Prediction Leveraging Machine and Deep Learning Algorithms with Advanced Alert System,” *Processes 2023, Vol. 11, Page 481*, vol. 11, no. 2, p. 481, Feb. 2023, doi: 10.3390/PR11020481.