2025 13th International Conference on Intelligent Embedded, MicroElectronics, Communication and Optical Networks (IEMECON) | 979-8-3315-7201-3/25/$31.00 ©2025 IEEE | DOI: 10.1109/IEMECON69302.2025.11365823

Integrated Climate and Water Analysis: Forecasting Floods and Droughts with Machine Learning

Subhadra Bhattacharyya
Dept. of ECE
Institute of Engineering & Management
University of Engineering & Management
Kolkata, India
Subhadra.Bhattacharyya2023@iem.edu.in

Sudip Bhattacharya
Dept. of ECE
Institute of Engineering & Management
University of Engineering & Management
Kolkata, India
Sudip.Bhattacharya2023@iem.edu.in

Amit Kumar Das
Dept. of ECE
Institute of Engineering & Management
University of Engineering & Management
Kolkata, India
amit.das@iem.edu.in

Abstract – Water is the most important thing in our life, and this paper focuses on monitoring and predicting water resources. The prediction is for flood and drought using rainfall forecasting. This forecasting was done by using monthly Normalized Difference Water Index (NDWI) and temperature, rainfall datasets. It is helpful for construction and agricultural activity. The prediction can be visualized in a graph of rainfall vs NDWI monthly. And for predicting this, we have used a Long Short-Term Memory (LSTM) model. For accomplishing such a thing, we have organized the data set accordingly for LSTM model and we have use prophet for advance climate forecasting and by using various plot like heat map, box plot, plotting Histogram to understand the Distribution the different Features NDWI + Rainfall Distribution Together to understand points with high NDWI and high rainfall as well as identify the flood thresholds.

Keyword – LSTM, NDWI, Prophet, Forecasting, Rainfall distribution.

I. INTRODUCTION

To manage water resources effectively, a system is needed to monitor water level changes, as sudden rises can cause floods and declines can lead to droughts. Preserving the water resources is not feasible for all the time because there are many types of water resources like rivers, lakes, streams which cannot be controlled easily .For doing such thing we need con tenuous surveillance on those water bodies. We all know that rainfall effect greatly to these kind of water resources but we can control the water flow of river by creating dams and maintaining the water level of pond and lake by draining the excess water or if the water level is low because of excessive use of water despite of insufficient rainfall then we can divert rainwater from rooftops, roads, and open land into the lake or pond or create recharge trenches and percolation pits which can help to maintain water level in small water resources like lakes and ponds .

This system uses forecasting as data and some ml models like LSTM and algorithm like random forest and Prophet. The dataset is used in this model includes monthly NDWI and monthly temperature and rainfall. NDWI is an index which monitors changes in water content leaves using near infrared wave length. Water reflects green light well, but absorbs near-infrared and it can detect areas where NDWI values increased, indicating new water resources like lakes and ponds, that is very helpful for sorting out the resources.

The system uses forecasting as data and some ml models like LSTM and algorithm like random forest and prophet. The datasets used for accomplishing the work are monthly NDWI and monthly temperature and rainfall. NDWI is an index which monitors changes in water content of leaves using near infrared and short wave infrared wave length. Water reflects green light well but absorbs near-infrared and it can detect areas where NDWI values increased, indicating new water resources, and also it can monitor how much flood evolves. After forecasting future NDWI, we can use Prophet for forecasting monthly rainfall, which can lead to the final prediction. Changes in temperature, along with rains, help provide a more comprehensive understanding of the environmental conditions related to drought and flood events.

II. METHODOLOGY

To build this system, we first collected and organized the required data — a process known as data cleaning. The cleaned data were then visualized and used to develop models such as LSTM and Prophet, along with algorithms like Random Forest. Several visualizations were also generated, which are discussed in the following sections.

A. Data processing and visualization

This process is crucial for building a successful system, as any error can lead to significant resource loss. In this model, monthly NDWI, rainfall, and temperature data were collected, dates were standardized, and datasets were merged based on date. Missing values were removed, and the data were visualized using correlation heatmaps. Heatmaps efficiently show relationships between variables, where correlation values range from -1 to 1 — brighter colors indicate strong correlations, while faded ones represent weaker relationships.

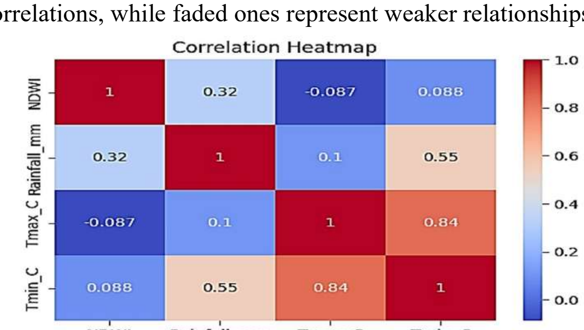

Fig. 1. Correlation heatmap

Then we plot the histogram to understand the distribution of the different features.

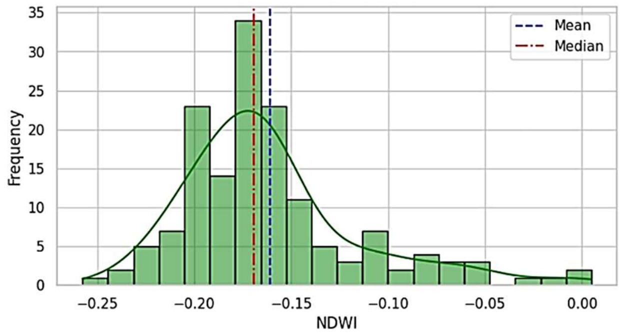

Fig. 2.1 NDWI distribution with KDE

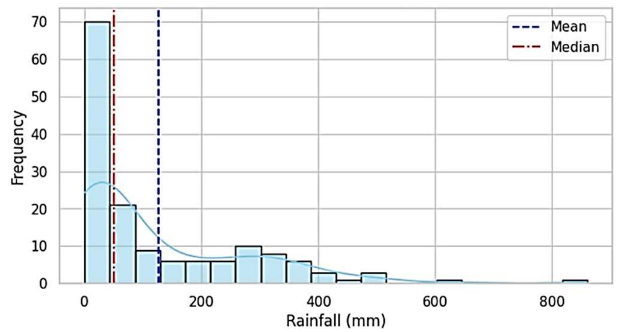

Fig. 2.2 Rainfall distribution with KDE

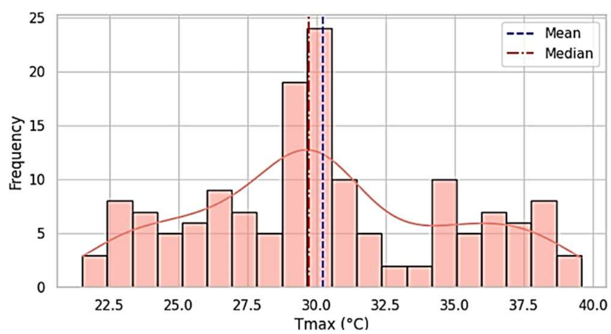

Fig. 2.3 Max Temperature distribution with KDE

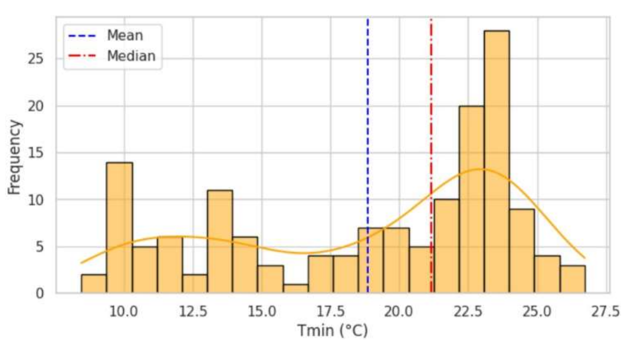

Fig. 2.4 Min Temperature distribution with KDE

After that we plot NDWI + Rainfall distribution together to understand points to identify the flood threshold. We know that if NDWI is less than 0.1 and rain fall is less than 50 mm then it can be called drought and if the rainfall is more than 223.97 mm and NDWI is more than 0.15 then it can be called flood by theses information we have plot condition class distribution which is number of records vs class label.

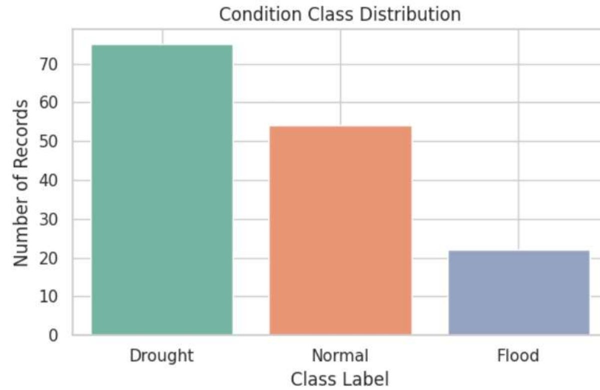

Fig. 3. Condition class distribution

that was possible because we manage to label the data that up to which records it is flood or drought or it is normal. It became possible because of random forest model.

B. Training of Random Forest model

Training the random forest model is essential for predicting water condition like flood, drought based on features like NDWI, rainfall, temperature. Training of random forest models requires initialization, then making prediction after that evaluation. After initializing starts the defining process where are two variables divided according to the requirement of training, one is created by selecting the feature and other one is assigning the target variables like flood, drought, normal. Then after splitting the data parameters are set and 20% of data for testing and the rest of the data for training. In the next step the prediction of water condition is saved and exported entire data frame to a csv file. In the last step a confusion matrix has plotted for evaluating the performance of random forest classifier.

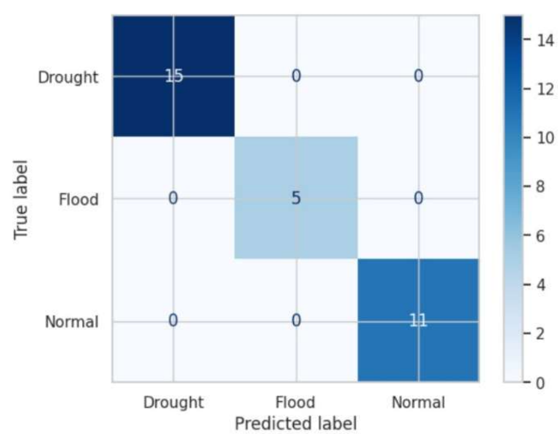

Fig. 4. Confusion matrices of Random Forest model

C. Forecasting NDWI using LSTM

Long Short-Term Memory (LSTM) is a Neural Network which is an upgrade of Recurrent Neural Network (RNN) and also it is an ideal for time series forecasting because it can capture long term dependencies. LSTM works using three logics -

- *Forget Gate* – Forgets irrelevant past information. It decides which memory to keep or which is to forget by using this formula - $f_t = \sigma\ (W_f \cdot [h_{t-1}, x_t] + b_f)$ when f_t is close to 1 it keeps the memory otherwise it forgets it.

- *Input Gate* – Decides what new info to store. It takes decision based on these formulas:
  $it = \sigma\ (Wi \cdot [ht-1, xt] + bi)$
  and
  $C\ \hat{}t = \tanh\ (Wc \cdot [ht-1, xt] + bc)$,
  it decides how much to the new memory to add and C^t is the new memory. Updating the memory is necessary so it is done by using this formula:
  $Ct = ft \odot Ct-1 + it \odot C\hat{}t$,
  Where, ft ⊙ Ct−1 parts of old memory added to the parts of old memory it ⊙ C^t

- *Output Gate* – Decide the output using,

$o_t = \sigma\ (W_o \cdot [h_{t-1}, x_t] + b_o) h_t = o_t \odot \tanh\ (C_t)$, basically there are two type of main functions at work, sigmoid function outputs value between 0 to 1 uses for making how 'much' decision and tanh function outputs value +1 to –1 uses for compressing information.

For forecasting NDWI we need LSTM and for using it we have cleaned the data and split the data into training and testing section. The training data includes all data points except the last 12 months, which are reserved for testing the forecast. In the next step we have created sequences for LSTM, it uses Min-Max to scale the training NDWI values to a range between 0 and 1. Then we have defined the scaled data for creating sequence where each input sequences is a window of past data points, and the corresponding output, it has window size of 12 for 12 months. After that we have trained the LSTM model by initializing sequential model and used Adam optimizer algorithm, sets the mean squared error as the loss function. In this training process the model will iterate over the entire training dataset 100 times and it will update its weights after processing every 8 samples. Finally, we have forecast the NDWI.

D. Forecasting climate data using Prophet & final prediction

Prophet is simple and easy to use accurate and robust forecasting library which is built by Meta for business application. Prophet is used for time series forecasting like temperature, daily sales and the job is done by decomposing the time series into components.

Here is a formula for this: $y(t) = g(t) + s(t) + h(t) + \varepsilon_t$, where y(t) is actual observation g(t) is the trend and s(t) is the seasonality representing pattern like daily, weekly and monthly and h(t) represents holidays or events and ε_t noise or errors. Prophet follows these following steps -

i. Input data – For starting the process it needs a data frame with two columns one is time stamp like date (d_s) and another one is measurement values like temperature, sales etc.
ii. Modelling the components – There are five components are used in prophet it models every component.
   - At first it models the overall direction of the data allowing change points where the trend can suddenly shift due to real world changes like lockdown and it finds the most likely change points
   - It models seasonality by handling repeating cycles like daily, weekly, monthly etc. It uses Fourier series for modelling seasonality:
     $S(t) = \sum^{N} (a)n \cos(2\pi nt/P) + b_n \sin(2\pi nt/P)$
     where P is a period and a_n and b_n coefficient prophet learns
   - It lets user manually define holidays or events (like Diwali, Black Friday) that affect the data.
iii. Prophet uses Stan, a Bayesian modeling engine, to estimate the values of all parameters for trend slope, seasonal strength, holiday effect, Changepoint location and it fits the model to the data by minimizing error.
iv. After training, it can predict the future values and it gives predicted value and 95% confidence interval and component breakdown.

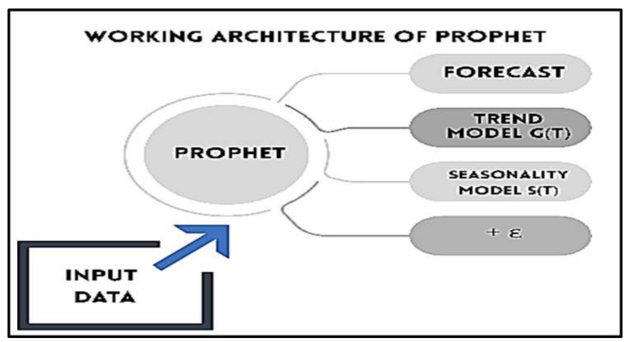

Fig. 5. Prophet working diagram

For forecasting the climate using prophet we have prepare the dataset which contains monthly rainfall and temperature and checked that the data correctly loaded or not. Then we have defined a function with three arguments for input data frame, containing the time series data and the name of the column in the data frame, which is to be forecasted and the number of future periods that means future prediction time line, which is by default 12 months. After that we have created an instance of Prophet and prepares the data in the right format. Then we have trained the model using historical data along with the data frame with future dates for the prediction which is already created. In the next step the trained Prophet model to generate forecasts for the future date created the previous step extracted from the forecast result. Then we have got the data frame containing the forecasted values for the specified column. After that we have utilized the function defined previously to generate future forecasting for rainfall, maximum and minimum temperature. We have combined these forecasts into a single data frame, and it automates the process of generating individual forecasts for multiple climate variables and then conveniently combines them for further analysis or visualizations.

For final prediction we have merged NDWI and climate forecast data frames into single data frame. We have defined a function which takes a row of the data frame as input and returns a classification based on predicted NDWI and rainfall values in that row and by using it we have stored the resulting classification and after processing we have managed to predict and plot the forecasted NDWI, rainfall, and the predicted water condition for the specified number of months. Finally, the system is complete.

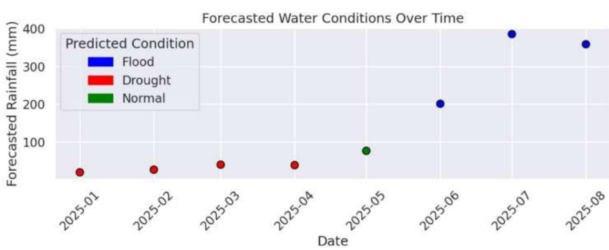

Fig. 6. forecasted water condition over time

III. *RESULT & ANALYSIS*

We have predicted the state of the water resources by utilizing LSTM and Prophet and the prediction is for eight months. We have classified the data into 'Drought', 'Normal', and 'Flood' using the data set based on NDWI and temperature, rainfall within a certain time. We have plotted the cleaned and process data for clear visualization. After that we have forecasted the climate changes and also NDWI. Forecasting is the part of this system based on forecasting we narrowed down the data frame for previously classified thresholds for changing the state of water bodies. Then we have processed the data accordingly for predicting the future water condition. In this whole process we have thoroughly visualized and checked the newly modified data frames and tried to understand the pattern for preparing that dataset accordingly for specific models or specific operation. Then we have defined function which can ensure the data sets have any missing value or not. After ensuring the result forecasting and predicting are correct, we have to create a final dataset containing all the predicted values for every feature with future dates for eight months. Then we have plotted as per the data frame for visualization.

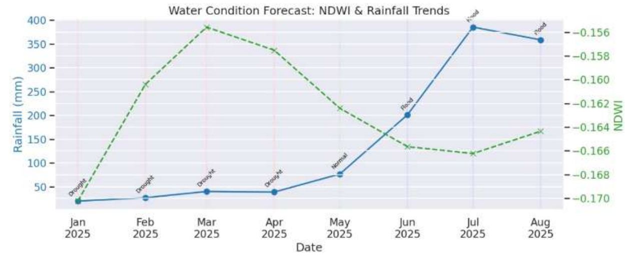

Fig. 7. Water condition forecast: NDWI & Rainfall trends

IV. *CONCLUSION*

In this paper, we have developed an integrated system for water resource monitoring and prediction using advanced machine learning techniques like LSTM and forecasting models like Prophet. By leveraging historical and current climate data, including NDWI, rainfall, and temperature, we successfully forecasted future water conditions to identify potential flood and drought scenarios. Our method provides an automated approach to classify regions into ‘Flood’, ‘Drought’, or ‘Normal’ categories, which is crucial for proactive resource management, agricultural planning, and disaster mitigation.

Overall, this system demonstrates a practical and scalable solution for environmental monitoring, which can be further enhanced by incorporating more granular data, real-time sensor inputs, or satellite imagery. With continuous improvements, such models can serve as powerful tools for government agencies, urban planners, and farmers, aiding in sustainable water management and disaster prevention strategies.

REFERENCES

[1] Gonzalo Mateo-García, Jordi Muñoz-Marí, Luis Gómez-Chova, “Cloud Detection on the Google Earth Engine Platform,” *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 10.1109/IGARSS.2017.8127359, 2017.

[2] Djamel Mansour, Bouchra Touil, “Cereal Crop Monitoring Using Sentinel Time Series Through Google Earth Engine,” *2024 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*, 10.1109/M2GARSS57310.2024.10537411, 2024.

[3] Jiang Wei, Kong Fanping, Li Xin, Ding Xiaohui, Cui Shiai, Feng Tianshi, “Total Phosphorus Mapping Using Sentinel-2 Images on the GEE Platform,” *2023 30th International Conference on Geoinformatics*, 10.1109/Geoinformatics60313.2023.10247737, 2023.

[4] Rustam Khamdamov, Khoshim Rakhmanov, Dilafruz Fattaeva, “Algorithm for Calculating Vegetation Indices NDVI for Time Series of Landsat 8 Multispectral Satellite Images on Google Earth Engine Platform,” *2021 International Conference on Information Science and Communications Technologies (ICISCT)*, 10.1109/ICISCT52966.2021.9670365, 2021.

[5] Egidio Arai, Yosio E. Shimabukuro, Andeise C. Dutra, Valdete Duarte, "Detection and Analysis of Forest Degradation by Fire Using LANDSAT/OLI Images in Google Earth Engine," *IGARSS 2019 – 2019 IEEE International Geoscience and Remote Sensing Symposium*, 10.1109/IGARSS.2019.8899250, 2019.

[6] Francesca Cecinati, Donato Amitrano, Lemia Benevides Leoncio, Elvis Walugendo, Raffaella Guida, Pasquale Iervolino, Sukumar Natarajan, "Exploitation of ESA and NASA Heritage Remote Sensing Data for Monitoring the Heat Island Evolution in Chennai with the Google Earth Engine," *IGARSS 2019 – 2019 IEEE International Geoscience and Remote Sensing Symposium*, 10.1109/IGARSS.2019.8898040, 2019.

[7] Mohammad Kakooei, Yasser Baleghi, "Spectral Unmixing of Time Series Data to Provide Initial Object Seeds for Change Detection on Google Earth Engine," *2019 27th Iranian Conference on Electrical Engineering (ICEE)*, 10.1109/IranianCEE.2019.8786494, 2019.

[8] Zhiyu Jiang, Liang Hong, "Monitoring of Surface Water Area and Water Level Changes in Nine Plateau Lakes in Yunnan and Analysis of Influencing Factors," *2024 IEEE 6th Advanced Information Management, Communications, Electronic and Automation Control Conference (IMCEC)*, 2024.

[9] Xiaozhou Yang, Liang Hong, "Spatiotemporal Variation and Attribution Analysis of Surface Water in Yunnan–Guizhou Plateau Based on Google Earth Engine," *2024 IEEE 6th Advanced Information Management, Communications, Electronic and Automation Control Conference (IMCEC)*, 10.1109/IMCEC59810.2024.10575081, 2024.

[10] Yong Du, Xiaoyu Zhang, Binbin Jiang, "Monitoring Water Quality of Surface Drinking Water Sources on Typical Islands with Remote Sensing Time Series Dataset," *2024 12th International Conference on Agro-Geoinformatics*, 10.1109/AgroGeoinformatics262780.2024.10660810, 2024.

[11] D. R. Manjunath, P. Jagadeesh, "Surface Water Dynamics Study Using Google Earth Engine: A Review," *Soft Computing and Geospatial Techniques in Water Resources Engineering*, Conference Paper, First Online: 02 December 2024.

[12] Sumona Akter Shimu, Mouluda Aktar, Masud Ibn Afjal, Adiba Mahjabin Nitu, Md. Palash Uddin, Md. Al Mamun, "NDVI-Based Change Detection in Sundarban Mangrove Forest Using Remote Sensing Data," *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, 10.1109/EICT48899.2019.9068819, 22 December 2019.

[13] Qiang Zhang, Fujun Sun, Qiangqiang Yuan, Liangpei Zhang, "Thick Cloud Removal for Sentinel-2 Time-Series Images via Combining Deep Prior and Low-Rank Tensor Completion," *2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 10.1109/IGARSS47720.2021.9553230, 11–16 July 2021.

[14] Linna Chai, Haiying Jiang, Wade T. Crow, Shaomin Liu, Shaojie Zhao, Jin Liu, "Estimating Corn Canopy Water Content from Normalized Difference Water Index (NDWI): An Optimized NDWI-Based Scheme and Its Feasibility for Retrieving Corn VWC," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 59, Issue 10, 10.1109/TGRS.2020.3041039, 11 December 2020.

[15] Yuyu Liu, Hao Wang, Ce Han, Yongfei Fu, "Analysis of Normalized Difference Wetness Index (NDWI) Using Landsat Imagery for Confluent Area of Multi Water Resources in Western Jinan," *2024 6th International Conference on Communications, Information System and Computer Engineering (CISCE)*, 10.1109/CISCE62493.2024.10653415, 2024.