A Deep CNN-Prophet based Drought Early Warning System using Dam Images and Climate Data

Musila Cynthia Katole

134653

Supervisor name: Deperias Kerre

Submitted in Partial Fulfillment of the Requirements of the Bachelor of Science in Informatics and Computer Science at Strathmore University

School of Computing and Engineering Sciences

Strathmore University, Nairobi, Kenya

January 2024

Declaration

I, Musila Cynthia Katole, hereby declare that this concept submitted for the award of a degree in Bachelor of Science in Informatics and Computer Science, is my own original work and has not been submitted to any other institution of higher learning. I further declare that all sources cited or quoted are indicated and acknowledged by means of a comprehensive list of references.

Supervisor

Name: Deperias Kerre

Signature:

Date: 07/01/2024

Student

Name: Cynthia Musila

Signature: Cynthia Katole Musila

Date: 07/01/2024

Abstract

Droughts pose a significant global challenge, impacting society and the environment. Low precipitation, crop damage, and water scarcity are among the consequences of drought. Both natural factors, such as reduced rainfall and climate variability, and human activities, including ineffective water management and climate change, contribute to the occurrence and severity of droughts. Current methods for detecting and monitoring droughts, such as satellite remote sensing, hydrological monitoring, rainfall monitoring, and drought indices, have limitations that hinder their effectiveness. Spatial and temporal resolution issues, restricted coverage, and delays in data availability are some of the challenges these approaches face. To address these limitations, this research proposes an innovative IT solution for drought detection using images of natural dams and climate data specifically temperature and humidity integrated with neural networks and a time series analysis algorithm. By harnessing advanced Artificial intelligence technologies and using high-resolution images, the proposed system aims to improve the accuracy early hydrological drought detection. The application will develop a monitoring system integrating AI algorithms and time series analysis integrated with the CRISP-DM development methodology based on some drought instances. By analyzing real-time data and utilizing neural networks, the system can identify patterns and indicators of drought more accurately and promptly. This will enable proactive measures and prompt response to mitigate the impacts of droughts. In conclusion, the proposed IT solution offered a novel approach to address the challenges of drought detection and monitoring. By leveraging AI and imagery, the system has the potential to revolutionize drought monitoring, providing more accurate and timely information for proactive decision-making and effective mitigation strategies.

Keywords: Drought detection, Localized monitoring, Dams, Neural networks, AI, Early warning system.

Table of Contents

List of Figures

List of Tables

List of abbreviations and acronyms

AI - Artificial Intelligence

BP – Back Propagation

ELM- Extreme Learning Machine

FIA-ML- Feature Importance Analysis with Machine Learning

GRNN- Generalized Regression Neural Network

GRU- Gated Recurrent Unit

HKSVM- Hierarchical Kernel Support Vector Machine

MLR- Multiple Linear Regression

WNN- Wavelet Neural Network

PDSI- Palmer Drought Severity Index

SPI -Standardized Precipitation Index

Chapter 1 Introduction

1.1 Background of the study

A drought is a period of time during which a region or area receives less precipitation than usual. Lack of sufficient precipitation can result in lessened stream flow, crop damsage, decreased soil moisture, groundwater, and general water scarcity. (Society, N. G., n.d.) A combination of natural and man-made forces brings on droughts. Reduced rainfall and higher evaporation rates are two critical meteorological contributors. Drought conditions can be caused by insufficient precipitation or extended periods of low rainfall, and droughts can become worsened by high temperatures and increased evapotranspiration, which enhance water loss from the soil and vegetation. El Nio and La Nia occurrences, examples of climate variability, can further skew rainfall patterns and exacerbate drought conditions. Additionally, droughts are becoming more common and severe in some areas due to changes in weather patterns due to climate change brought on by human activity. (committee, l.r., n.d.)

Hydrological factors also influence the occurrence of drought. Insufficient rainfall and limited groundwater recharge can lead to reduced streamflow and lower river levels, exacerbating drought conditions. Excessive use and unsustainable extraction methods further deplete groundwater supplies, decreasing water availability during dry periods. Poor soil moisture retention and insufficient water-holding capacity worsen the adverse effects of reduced rainfall and dry spells, intensifying the impact of drought on soil and land use. Deforestation, land degradation, and unsustainable land management practices disrupt the water cycle, decrease soil moisture, and increase drought susceptibility. (committee, l.r., n.d.)

Human activity can dramatically impact the frequency and severity of droughts. Ineffective water management and allocation techniques, such as excessive water usage and unsustainable irrigation techniques, can deplete water supplies and worsen drought. Urbanization and rapid population increase strain water supplies, making droughts more likely. Furthermore, drought risks are made worse by human-caused climate change, which alters precipitation patterns and increases the frequency and severity of droughts across a wide range of places. (committee, l.r., n.d.)

In 2023, recurrent drought severely impacted Kenya. Experts project that approximately 5.4 million people will encounter high levels of acute food insecurity between March and June, with an estimated 1.2 million entering the emergency phase. This represents a 43% increase compared to the same period in 2022. The weather forecasts indicated a reasonable chance of another poor rainy season from March to May, which would mark an unprecedented sixth consecutive season of underperformance. Approximately 970,000 children aged 6 to 59 months and 142,000 pregnant or lactating mothers are estimated to experience acute malnutrition throughout 2023, necessitating immediate treatment. Moreover, more than 2.4 million livestock, crucial for pastoralist families' nourishment and livelihoods, have perished. Families are resorting to desperate and unhealthy coping mechanisms to deal with the dire consequences of the drought. (Communications, n.d.)

Drought monitoring and early warning systems typically aim to track, assess and deliver relevant information concerning climatic, hydrologic and water supply conditions and trends.
(dessertification, s.d.) Current ways of detecting drought are Rainfall Monitoring, Satellite Remote Sensing, Hydrological Monitoring and Drought Indices (Portal, s.d.). These methods have been efficient throughout the years, however they have common limitations like real time data availability, Spatial and temporal resolution and localized monitoring (Center). Due to these limitations, there is need for a localized monitoring system that factors real time data availability with high spatial and temporal resolution using images of in land dams captured by drones coupled up with advanced technology for AI i.e. neural networks. This will form the rationale of this research.

1.2 Problem statement

The existing approaches used for detecting and monitoring drought conditions are constrained by various limitations that impede their effectiveness. Satellite remote sensing, which utilizes satellite imagery to monitor drought indicators like vegetation health, soil moisture, and surface temperature, is limited by spatial and temporal resolution. The varying spatial resolutions of satellite sensors restrict the ability to capture detailed information at a fine scale, hindering the monitoring small-scale features and localized areas. Additionally, the cost and accessibility of satellite data pose challenges, as it can be expensive to acquire, process, and interpret, limiting its widespread use, especially in resource-constrained regions. (Hayes, n.d.)

1.3 Research objectives

1.3.1 General objective

To develop a system to detect early stages of drought using images of dams at different water levels and neural networks.

1.3.2 Specific objectives

i. To analyze the challenges posed by droughts and the drought detection.
ii. To review the current techniques used in drought detection.
iii. To develop a model for drought detection.
iv. To evaluate the developed model.

1.4 Research questions

i. What are the challenges posed by drought and drought detection?
ii. Which techniques currently being used for drought detection?
iii. How can a drought detection system be developed and what methodology and framework will be used in the implementation?
iv. How will the system's functionality be tested?

1.5 Justification of the research

Early and localized drought detection is of prime importance, especially to people in areas prone to drought. Drought has significant effects on various aspects of society and the environment. Crop failures and reduced yields are among the agricultural losses caused by severe drought. Projected data indicates that approximately 5.4 million people in Kenya will experience insufficient access to food and water between March and June 2023 due to the severity of the drought. (Program, s.d.) According to the UN environment program, there has been an upward trend of using AI to help deal with some environmental issues. Deep learning, a general area within AI, facilitates the creation of end-to-end models that produce desired results by utilizing input data without requiring manual feature extraction (Bengio, n.d.) Due to the recent drought that Kenya and other countries have been experiencing, there is a pressing need for expertise in this domain, driving a growing interest in developing AI-based detection and early warning systems for drought. AI approaches can potentially address limitations associated with temporal and spatial resolution and localized monitoring.

1.6 Research scope and limitations

The scope of this research was limited to the use of online data to develop the drought detection system, images of dams at different levels and climate data specifically temperature and humidity will be applied.

Chapter 2 Literature Review

2.1 Introduction

This chapter intends to examine various types and origins of drought, alongside the phases it undergoes. Furthermore, we explore the stages involved in detecting drought and its connection with different dams. Several techniques utilized for drought detection will be expounded upon within this chapter. Lastly, we assess the present systems and models employed in the field of satellite remote sensing for drought detection, while emphasizing the limitations of this research.

2.2 Types of Droughts

A drought is a time during which a region or area receives less precipitation than usual. Lack of sufficient precipitation can result in lessened stream flow, crop damsage, decreased soil moisture, groundwater, and general water scarcity (Society, N. G., n.d.). Drought can be categorized into different types based on the affected sector, duration, and severity of the water deficit.

Hydrological Drought centers on water scarcity in rivers, lakes, reservoirs, and groundwater systems. It considers the effects of reduced precipitation on the overall water supply and availability. Hydrological drought impacts water resources, ecosystems, water-dependent industries, and hydropower generation. (Society, N. G., n.d.)

Meteorological drought refers to sustained periods of significantly reduced precipitation compared to the long-term average for a specific region. It primarily focuses on rainfall deficiency, serving as a fundamental precursor for other types of droughts. (Society, N. G., n.d.)

Agricultural Drought emphasizes insufficient soil moisture and water availability, directly impacting agricultural activities. Consequences include diminished crop yields, hindered plant growth, and limited water resources for livestock. It affects farming practices, food production, and the overall agricultural sector. (Society, N. G., n.d.)

Socioeconomic drought evaluates the broader implications of water scarcity on human populations and economic activities. It encompasses drought's social, economic, and environmental consequences, such as water shortages in communities, increased food prices, reduced industrial production, and overall socioeconomic stress. (Society, N. G., n.d.)

Ecological drought focuses on the repercussions of water scarcity on ecosystems and natural environments. It examines disturbances to the balance of ecological systems, affecting plant and animal populations, wetlands, rivers, and overall biodiversity. Ecological drought can lead to habitat degradation, loss of biodiversity, and increased vulnerability to wildfires.

(Society, N. G., n.d.)

The scope of this research focuses on hydrological drought as it explores the relationship between drought and dams.

2.3 Causes of hydrological drought

The causes of drought can stem from various factors, encompassing both natural occurrences and human activities. Within this subtopic, we explore the underlying reasons for drought occurrence.

Firstly, droughts may arise from insufficient precipitation, wherein a specific region experiences prolonged periods of inadequate rainfall or snowfall. This can be influenced by alterations in weather patterns, such as shifts in atmospheric circulation or the impact of global climate phenomena like El Niño or La Niña. (Loon, n.d.)

Additionally, climate change plays a significant role in exacerbating drought conditions. Changes in global climate patterns, including rising temperatures, altered precipitation patterns, and shifts in atmospheric circulation, contribute to more frequent and severe droughts in certain areas. (Loon, n.d.)

Furthermore, soil and water conditions significantly impact the severity of droughts. Inadequate soil moisture levels and limited water storage capacity in reservoirs, lakes, and rivers intensify the effects of drought. When water sources are already depleted or unable to meet demand, even slight decreases in precipitation can amplify drought impacts.

Human activities, such as deforestation and land use changes, can disrupt the natural water cycle and exacerbate drought vulnerability. Forest clearance and land conversion for agricultural or urban purposes diminish rainfall and increase evaporation rates, heightening the susceptibility to drought in affected regions. (Loon, n.d.)

Excessive extraction of groundwater is another human-induced cause of drought. Over pumping of groundwater for irrigation, industrial use, or domestic consumption depletes underground water reserves. If the extraction rate surpasses the natural recharge rate, it leads to declining groundwater levels and eventual drought conditions.

Lastly, human activities such as excessive water consumption and improper water allocation policies exacerbate drought risk.

2.4 Challenges posed by drought

Challenges posed by drought can be categorized into three main areas: economic impacts, environmental impacts, and social impacts.

Economically, droughts have significant consequences. Farmers face financial losses as their crops are destroyed, leading to food scarcity and requiring increased expenses for irrigation. Businesses dependent on agriculture suffer when drought damages crops or livestock, and the timber industry is adversely affected by wildfires. Power companies relying on hydroelectric power incur higher costs, while water companies must invest in additional water supplies. Reduced water levels impede navigation and impact businesses reliant on water transportation. Moreover, drought leads to increased food prices. (Center, How Does Drought Affect Our Lives?, s.d.)

Environmentally, drought results in the loss or destruction of habitats, causing a lack of food and drinking water for wildlife. Disease occurrence among people and animals rises, migration patterns shift, and endangered species face heightened stress and possible extinction. Water scarcity affects reservoirs, lakes, and ponds, while wetlands are lost. More frequent wildfires and soil erosion further degrade ecosystems and soil quality. (Center, How Does Drought Affect Our Lives?, s.d.)

Socially, drought affects mental health due to economic losses, while reduced water flows and poor quality lead to health problems. Dust poses additional health risks. Loss of human life occurs, and public safety is compromised by increased forest and range fires. Reduced incomes and population movements from rural to urban areas exacerbate social challenges. Recreational activities are limited due to water scarcity. (Center, How Does Drought Affect Our Lives?, s.d.)

2.5 Challenges in drought detection

Satellite remote sensing, which utilizes satellite imagery to monitor drought indicators like vegetation health, soil moisture, and surface temperature, is limited by spatial and temporal resolution. The varying spatial resolutions of satellite sensors restrict the ability to capture detailed information at a fine scale, hindering the monitoring small-scale features and localized areas. Additionally, the cost and accessibility of satellite data pose challenges, as it can be expensive to acquire, process, and interpret, limiting its widespread use, especially in resource-constrained regions. (Hayes, n.d.)

Hydrological monitoring, which assesses streamflow, river levels, and groundwater levels, provides valuable information but faces limitations. The limited distribution of monitoring stations results in restricted coverage and difficulties in capturing drought conditions comprehensively. Furthermore, there may be a lag time in the response of streamflow and groundwater to drought, leading to delays in reflecting drought impacts in hydrological data. (Hayes, n.d.)

Rainfall monitoring, based on collecting and analyzing precipitation data from weather stations, also has limitations. The sparse distribution of weather stations, particularly in remote areas, leads to data gaps and limited accuracy in detecting localized drought conditions. Real-time availability of rainfall data may be lacking, causing delays in promptly identifying and responding to drought conditions. Moreover, the specific locations covered by weather stations make it challenging to capture rainfall variations across larger areas, especially in regions with complex topography.

(Hayes, n.d.)

Drought indices, such as the Standardized Precipitation Index (SPI) and the Palmer Drought Severity Index (PDSI), offer quantifying drought conditions based on meteorological and hydrological data. However, they face limitations such as data requirements and simplified representation. Long-term datasets necessary for calculating drought indices may only be available or reliable in some regions, limiting their applicability. Additionally, while drought indices provide a generalized measure of drought, they may not adequately capture localized or sector-specific impacts, such as those affecting agriculture or water resources. (Hayes, n.d.)

2.6 Methods used in hydrological drought detection

These approaches and methodologies are employed to collect data, examine drought conditions, construct models, and derive valuable insights to enhance the monitoring, prediction, and alleviation of drought:

Remote Sensing: The utilization of satellite-based remote sensing technology allows for the monitoring and analysis of drought conditions. This includes evaluating the health of vegetation, land surface temperature, soil moisture levels, and water storage on the land.

(Adisa, n.d.)

Rainfall Analysis: This technique involves the study of rainfall patterns, encompassing the identification of seasonal variations, interannual and decadal changes, as well as establishing connections to climate phenomena such as El Niño–Southern Oscillation (ENSO) and oceanic fluctuations. (Adisa, n.d.)

Drought Indices: Various indices are employed to quantify and assess drought conditions by integrating data on precipitation, rainfall, and streamflow. Commonly used indices include the Palmer Drought Severity Index (PDSI), the Crop Moisture Index (CMI), the Soil Moisture Drought Index (SMDI), the Standardized Precipitation Index (SPI), the Standardized Precipitation Evapotranspiration Index (SPEI), the Effective Drought Index(EDI), among others. (Adisa, n.d.)

2.7 Stages of hydrological drought

Detecting different stages of drought is crucial for a comprehensive understanding of hydrological droughts. Traditional methods that compare daily water flow to a set threshold often fail to distinguish between short drought episodes, resulting in ambiguity. To address this issue, previous studies have disregarded drought events lasting less than 15 days and attempted to merge separate events occurring within 10 days. However, this approach falls short in identifying droughts that extend across multiple seasons. (Behzad, n.d.)

The method below overcomes these limitations and improves the assessment of hydrological droughts. It categorizes a hydrological drought episode into three distinct stages: growth, persistence, and retreat. The primary criterion for assessing hydrological drought is the persistence period, which spans at least 30 consecutive days during which the streamflow remains below the established threshold level. Once we have identified the persistence period, we can subsequently examine the growth and retreat phases of the drought. (Behzad, n.d.)

Persistence: During the persistence stage, the streamflow remains consistently below the threshold level for at least 30 consecutive days. If multiple periods meet this criterion within a drought episode, the longest period is prioritized as the primary drought persistence stage.

Growth: The growth stage, which marks the onset of drought, is a critical phase for effective drought detection and assessment. When retracing the analysis from the onset of the drought persistence stage, it is observed that the streamflow falls below the threshold level for fewer than 15 days within the defined T-day window. Detecting drought at this stage enables us to intervene early, implement proactive water management strategies, and raise awareness to mitigate potential impacts. It also provides valuable insights into the temporal progression of drought, helping us understand the severity, duration, and underlying factors driving water scarcity.

Retreat: Moving forward from the conclusion of the drought persistence stage, we determine the end of the drought, known as the retreat phase. This occurs when the streamflow falls below the threshold level for less than 15 days within the T-day window. The drought retreat stage spans the period following the conclusion of drought persistence until the termination of the drought.

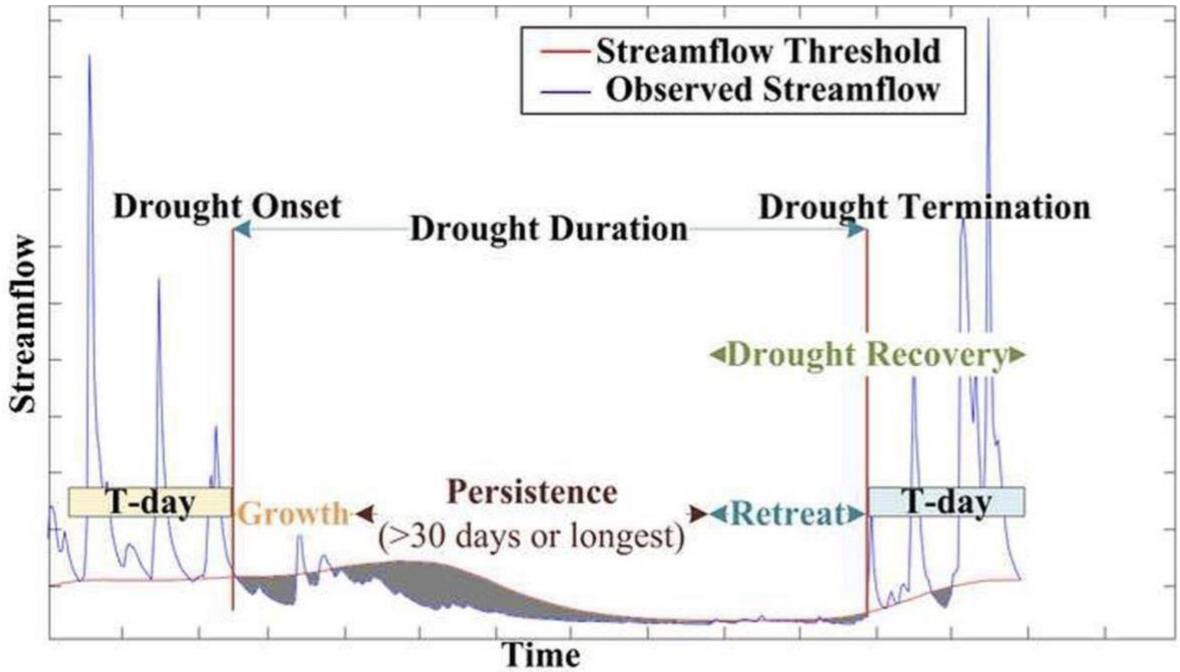


Figure 2-1: Hydrological Threshold determination

(Behzad, n.d.)

2.8 Relationship between hydrological drought and dams

Hydrological droughts have a significant impact on water quantity in dams. During drought episodes, water temperature increases, leading to rapid water loss through evaporation. Considering that hydrological droughts are divided into stages of growth, persistence, and retreat, significant water evaporation occurs during the growth phase. Therefore it is essential to detect drought at that stage. An inverse relationship exists between drought duration and quantity of water at a given water body, indicating that longer drought events are associated with more intense annual streamflow deficits. (Behzad, n.d.)

2.9 Hydrological drought threshold determination

Hydrological drought threshold determination is critical in understanding and characterizing drought events. Various factors, such as the region's characteristics, data availability, and study objectives, influence the method chosen for threshold calculation.

In this study, data on the quantiles of water level based on the long time series are considered to determine the optimum value for streamflow threshold. The analysis includes calculations of monthly, weekly, and yearly quantiles to capture the various time scales of streamflow variability. Through the examination of water level quantiles at different time scales, the study aims to identify the threshold that best captures the characteristics of hydrological droughts. This approach is favored because it effectively captures the low flow regime of a basin. (Behzad, n.d.)

2.10 Relationship between Drought and Temperature

Temperature and drought share a complex and interconnected relationship. Warmer temperatures play a significant role in driving and intensifying drought conditions. Increased evaporation, reduced precipitation, changes in snowmelt timing, and increased plant water demand are all consequences of higher temperatures that contribute to drought. Furthermore, droughts can amplify temperature extremes by reducing water vapor in the atmosphere and altering land surface reflectivity.

The impact of temperature on drought varies depending on geographical location, soil type, vegetation cover, and human activities. Soils with lower water holding capacity and areas with sparse vegetation are more susceptible to drought under higher temperatures. Deforestation and improper land management practices can worsen drought impacts by contributing to soil degradation.

Understanding the interplay between temperature and drought is crucial for predicting and mitigating the impacts of drought in a changing climate. By addressing the factors that intensify drought conditions and implementing strategies to conserve water resources, we can strive to reduce the negative consequences of droughts and promote resilience in vulnerable regions (California Department of Water Resources, [Kelly M. Grow], n.d.).

2.11 Relationship between Drought and Humidity

Humidity and drought have a intricate connection, mutually impacting each other and molding the water equilibrium in ecosystems.Higher humidity slows evaporation, conserving soil moisture and potentially mitigating drought. Increased humidity often correlates with augmented precipitation, replenishing water resources. However, these effects vary regionally, influenced by climate. Conversely, droughts reduce evapotranspiration, lowering atmospheric humidity and worsening dry conditions. Altered air circulation patterns during droughts also impact humidity. Additional factors include temperature, where high temperatures intensify drought, wind speed accelerates evaporation, and vegetation cover helps maintain humidity, acting as a buffer against drought impacts. Understanding these dynamics is crucial for managing water resources in diverse ecosystems (California Department of Water Resources,[Kelly M. Grow], n.d.).

2.12 Current drought detection systems

2.12.1 DroughtCast: A Machine Learning Forecast of the United States Drought Monitor

DroughtCast focuses on developing and testing robust drought forecast methods using machine learning techniques. The study covers the entire CONUS (Continental United States) from June 2003 to January 2020. Precipitation plays a crucial role in the DroughtCast model, which combines satellite-observed and modeled meteorological variables as predictors of drought. Other variables include drought status, soil moisture estimates, vegetation gross primary production, land cover data, meteorology data, and long-term climatic indicator. (Burst, n.d.)

Normalized anomalies between -1 and 1 are used to balance the input features, and a Seq2Seq model with gated recurrent units (GRUs) is employed for forecasting. The model is trained using mean squared error (MSE) loss and backpropagation algorithm, validated with spatial and temporal generalization tests, and evaluated using performance metrics. Results demonstrate the model's accuracy in forecasting drought up to 12 weeks in advance, capturing spatial and temporal progression, severity, and historical events. (Burst, n.d.)

2.12.2 Runoff Forecasting Using Machine-Learning Methods: Case Study in the Middle Reaches of Xijiang River

The study proposes a reliable runoff forecasting model by combining hydrological and meteorological data. Four machine learning models (BP, GRNN, ELM, and WNN) are used to forecast mean streamflow and water level up to 7 days in advance. The research also considers the flood propagation mechanism to improve accuracy. Objectives include enhancing runoff forecasting accuracy and efficiency and exploring the relationship between flood propagation and runoff. The study focuses on the Xijiang River basin in China and utilizes data from hydrographic stations and the meteorological center covering 2009 to 2019. (Xiao, n.d.)

The methodology employs artificial neural networks to capture non-linear relationships between input and output variables. The models utilized like GRNN, ELM, and WNN are briefly described, emphasizing their suitability for non-linear problems and their performance in runoff forecasting. In conclusion, the research emphasizes the significance of accurate runoff forecasting and the benefits of machine learning techniques. It presents a specific case study in the Xijiang River basin that utilizes machine learning models for runoff forecasting, contributing to early warning systems for floods and droughts. (Xiao, n.d.)

2.12.3 Climate-informed monthly runoff prediction model using machine learning and feature importance analysis

The study combines machine learning models with feature importance analysis to enhance runoff prediction accuracy. Mutual information and random forest algorithms are utilized to identify key physical factors that impact runoff, improving the understanding of the complex relationship between hydrometeorological factors and runoff. Various machine learning models, including HKSVM, GRNN, and MLR, are optimized using the improved particle swarm optimization algorithm to accurately predict monthly runoff. These techniques provide valuable insights for drought management and water resource planning. (Yan, n.d.)

Focused on the Yingluoxia basin in Northwest China, the research incorporates local meteorological data and large-scale climate factors to improve runoff prediction accuracy. Comparisons with traditional time series analysis models demonstrate the superiority of the proposed feature importance analysis with machine learning (FIA-ML) models in capturing monthly runoff variations, offering valuable insights for drought forecasting and decision-making. Overall, the study enhances runoff prediction accuracy, advances the understanding of the relationship between hydrometeorological factors and runoff, and provides valuable insights for the development of flood and drought early warning systems. (Yan, n.d.)

2.13 Conceptual Framework

The envisioned system processes images of natural dams as its input. It feeds these images into a deep CNN image classification machine learning model. Time Series Data is fed into the Prophet model for time series analysis. The OpenWeather API collects data on Temparature and humidity, this along with the classification data is passed to the Prophet model for forecasting.

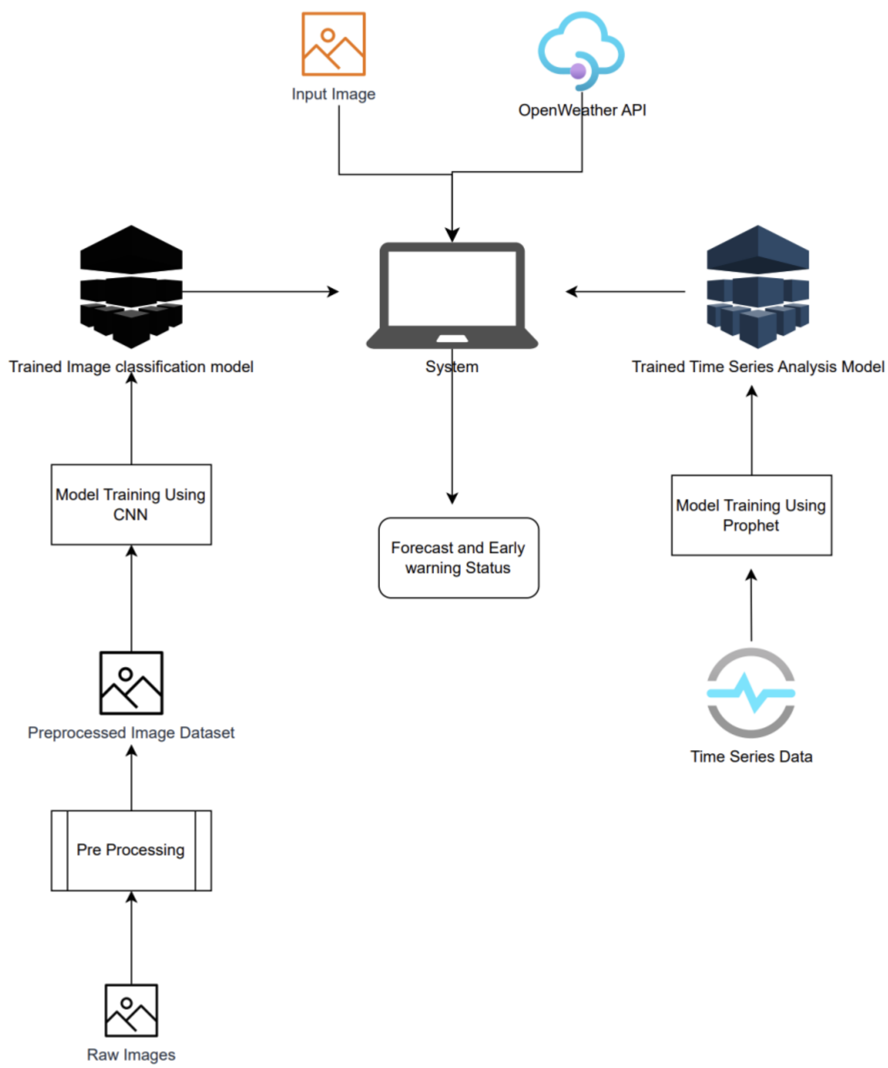


Figure 2-2: Conceptual Framework

Chapter 3 Research Methodology

3.1 Introduction

In this chapter, the research design and software development methodology utilized in creating the deep learning model will be discussed. Furthermore, the chapter will delve into the tools and techniques employed during development. Lastly, it will outline the expected deliverables of the developed model.

3.2 Research Design

Research design encompassed the systematic arrangement of conditions for data collection and analysis, combining relevance to the research purpose with procedural efficiency (Kothari). It provided a blueprint for organizing the research process, including data collection, measurement, and analysis. This study followed an experimental approach, involving the selection, preprocessing, and augmentation of a suitable dataset, training the ensemble model as a proof of concept, and testing and validating the model's performance.

3.3 Model Development

To develop the ensemble the steps that will be followed are as follows:

i. Data collection
ii. Development of the model
iii. Validation of the model

3.3.1 Data Collection

This analysis collected satellite imagery dams from Kaggle. The study's data selection followed specific criteria, including assessing the severity of drought and the period it lasted, affected regions regarding water body land coverage, and prioritizing the highest available resolution. The website Timeanddate and OpenWeather API was used to collect data on temparature and humidity.

3.3.2 Development of the model

The approach used is to build a deep CNN and prophet. The deep Convolution Neural Network used pretrained ResNet50 for better performance in image classification. Prophet by facebook was used for time series analysis for forecasting.

3.3.3 Validation of the model

The proposed model's performance was assessed using metrics like accuracy, F1 score and mean squared error.

3.4 System Development Methodology

The approach used was the CRISP-DM methodology. The CRISP DM methodology presents a cyclical method for project development. It consists of six primary phases: business comprehension, data comprehension, data arrangement, modeling, assessment, and implementation. (Hotz)

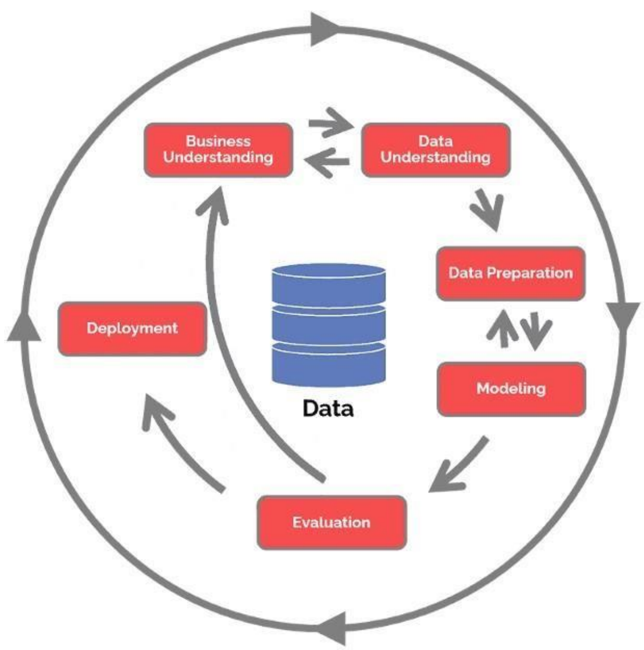


Figure 3-1: CRISP-DM methodology

3.5 System Development Tools and Techniques

3.5.1 Visual Studio Code

Visual Studio Code is a highly efficient and robust source code editor that operates on your computer and is accessible on Windows, macOS, and Linux. It is the preferred choice for developers, especially those working with flutter, due to its comprehensive support for flutter development. (Microsoft, s.d.)

3.5.2 Google Colab

Google Colab is a free, cloud-based platform for collaborative coding in Python through Jupyter Notebooks. It provides access to GPU and TPU resources, supports real-time collaboration, and integrates with Google Drive. It was to implement both models.

3.5.3 Streamlit

This python library was used for the creation of web app for visualizing the time series analysis data.

3.5.4 Flask

This python web framework was used for aiding on the process of uploading images to the classification model.

3.5.5 Ethical Considerations

In the study the researcher utilized publicly available data and attribution for the use of this data will be given to the owners. Any information that was gathered was treated with utmost confidentiality and integrity.

Chapter 4 System Analysis and Design

4.1 Introduction

This chapter discusses the various components of the system identifying both the functional and non-functional requirements of the system. It also discusses system analysis and design diagrams designed using the design paradigm CRISP-DM methodology which is a data-driven approach. Some of the diagrams in this chapter include a use case diagram, system sequence diagram, and class diagram.

4.2 System Requirements

Several system requirements were examined as part of the project analysis, including:

4.2.1 Functional Requirements

Some of the services that the system has to offer are described below.

i. Authentication

This requirement enables environmentalists to be registered and logged into the system. The environmentalists provide their emails and passwords during registration. The details are stored in the Firebase database with the passwords hashed using the scrypt hashing algorithm. When logging in, they provide these details, which are then verified to be in the right format. Once the details are verified and a secure connection is established using SSL, they are granted access to the system and open their session. Eventually, once they are done using the system, they can be logged out.

ii. Upload Dam Images

A flask-based file transfer module for uploading images to the classification model was implemented. Users interact with an HTML form to select and submit image files. The Flask server, configured to handle file uploads, receives and securely saves the uploaded files to a designated server folder. Once stored, the module integrates image data with the classification model, triggering predictions.

iii. Water level classification

This is the main functionality of the system. Uploaded dam images are processed and using the deep CNN model, they are classified and output is generated indicating the water level as either low, normal, or critical.

iv. Real-time climate data collection

The system implemented an API for real-time collection of Temperature and humidity which are essential for forecasting drought and

v. Updating time series data in MySQL database

This is a critical functionality of the system as it enables the collection of time series data which is important for time series analysis and forecasting.

vi. Forecasting

The system then feeds the compiled time series data into a Prophet model for forecasting, using classification data as the main variable and temperature and humidity as regressors.

vii. Early warning

The environmentalist is notified in case the forecast is above the set threshold thus enabling them to prepare against critical drought situations.

4.2.2 Non-Functional Requirements

Some of the system attributes are described below.

i. Response Time

The system was quick to give the user the required response. The mobile app and web app load in less than a second for the users, this was tested by timing how long it takes to load the pages.

ii. Accuracy

The system produced accurate results, the classifications were precise to ensure that accurate predictions of drought are made.

iii. Usability

The system was easy use, memorable and easy to learn. Therefore, the system had a familiar look, actions were easily identifiable and easy to execute.

iv. System security

The user passwords were hashed with the scypt hashing algorithm, the system also implemented a criterion to make sure the passwords set were strong using regular expressions. The web app pages were secured with SSL.

4.3 System Narrative

The primary goal of this project was to establish a robust system utilizing climate data to issue early warnings. Current drought early warning systems rely on normalized anomalies of climate data within seq2se2 models for drought forecasting. However, this approach poses several challenges, including labor intensity, time consumption, and the waste of useful metadata. To address these issues, the project introduces an automated method of inputting climate data through an API. The devised solution incorporates an image classification model using a Deep CNN, which categorizes dam images based on varying water levels: low, normal, and critical. The training set images include NIR, R, and G bands to accentuate dam edges. Subsequently, the system integrates the classification data with additional climate data from the API, specifically temperature and humidity, into a time series forecasting model for predictive analysis.

If the forecast surpasses the predetermined threshold, the system issues an early warning to the environmentalist, enabling proactive measures against critical drought situations. This innovative approach not only streamlines the input of climate data but also enhances the accuracy and efficiency of drought predictions, contributing to more effective environmental risk management.

4.4 System Analysis Diagrams

The diagrams in this section illustrate the system's anticipated functionality, including the use case diagram, system sequence diagram, and class diagram. These design artifacts adhere to the CRISP-DM methodology (Cross-Industry Standard Process for Data Mining). The CRISP-DM process involves understanding project goals and requirements, exploring and preparing the data, selecting and refining modeling techniques, evaluating model performance, deploying the system, and iterating as needed. Aligning with CRISP-DM ensures a systematic and effective approach to system design within the broader context of data mining best practices.

4.4.1 Use Case Diagram

Figure 4 provides a visual representation of the interactions between various actors and distinct use cases within the system. The system involves four key actors: the environmentalist, the Open Weather API, the image classification model, and the time series analysis model. The environmentalist actor is endowed with the capabilities of registration, login, and logout. Authentication is a mandatory step during registration and login processes. Additionally, the environmentalist has the privilege to upload dam images to the system and access visual representations of both the image classification and the forecast through the user interface. These functionalities collectively contribute to a user-friendly and comprehensive experience for the environmentalist within the system.

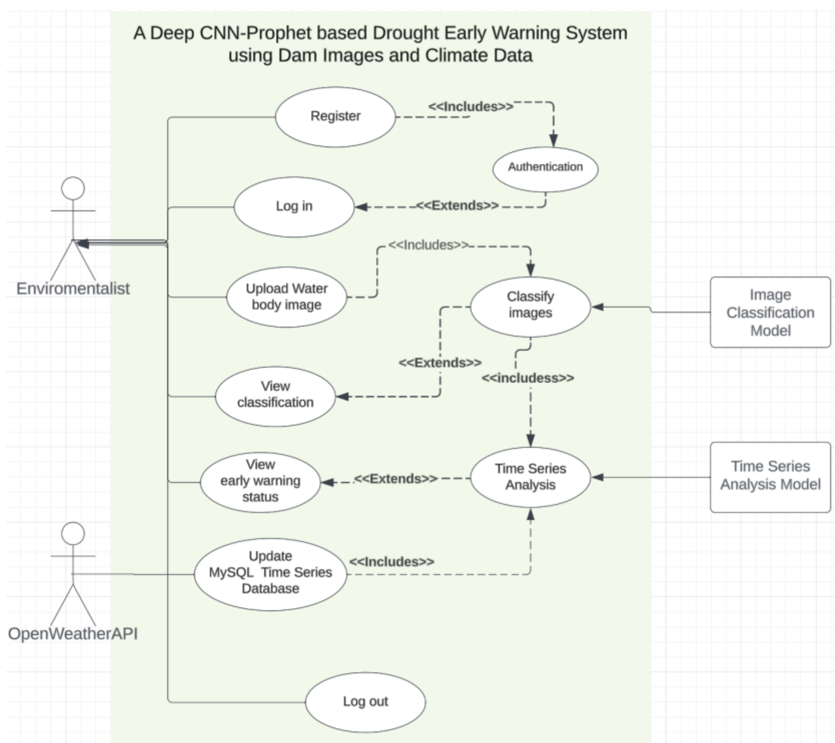


Figure 4-1: Use case diagram

4.4.2 System Sequence Diagram

Figure 5 visually outlines the system's sequential tasks. Starting with log-in, the environmentalist inputs their email and password for verification. Upon successful authentication, they access the landing page, unlocking various functionalities. The diagram details the classification and forecasting process. The environmentalist uploads dam images, triggering classification and updating of classified and climate data in the Time Series Data in the MySQL database. This merged dataset seamlessly feeds into the forecasting model, generating predictions and early warnings. The visual guide succinctly captures user interactions and their significant contributions to essential system processes.

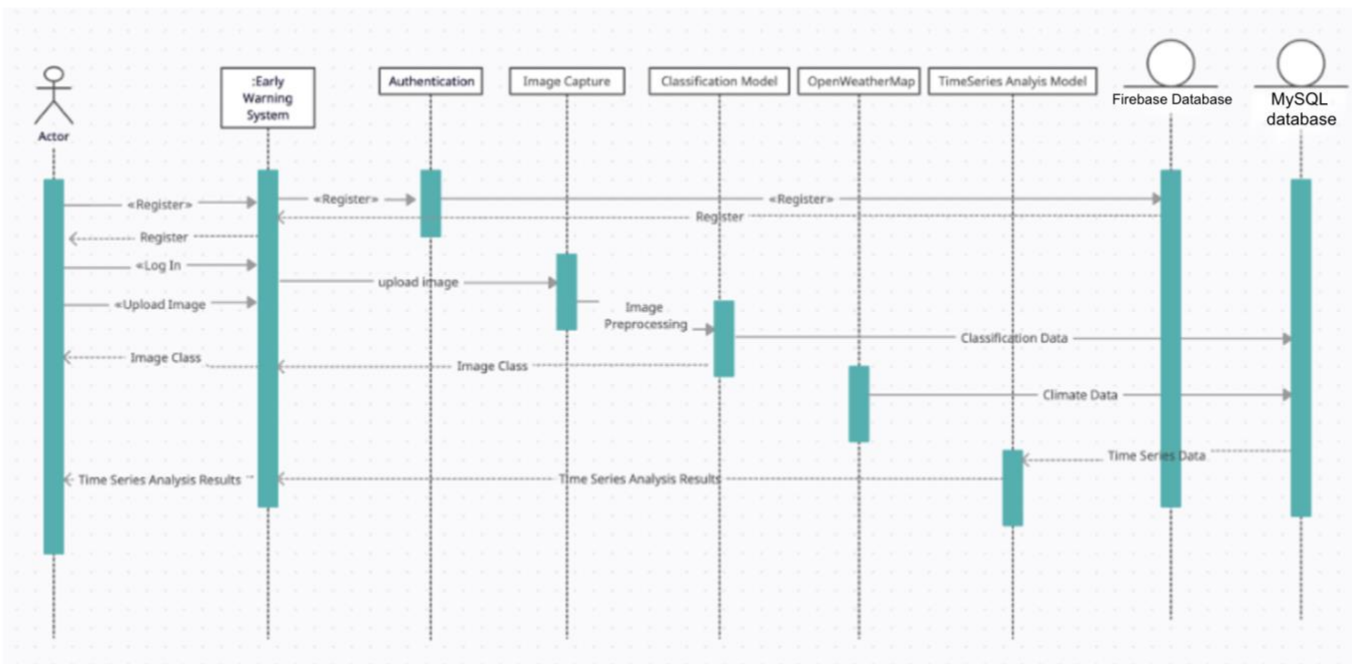


Figure 4-2: System sequence diagram

4.5 System Design Diagrams

In this section, we present models, including user interfaces, that fulfill all documented system requirements while showcasing the intricacies of outputs, inputs, and processes. These models are delineated through the utilization of wireframes, database schema depictions, and system architecture representations.

4.5.1 Class Diagram

Fig 6 shows the main classes including Environmentalist, representing users who can register, log in, and perform other essential actions. OpenWeather API and Image Classification Model classes depict external systems that provide climate data and classify dam images, respectively. The Time Series Analysis Model and Update_Database classes represent the core components responsible for processing and analyzing data. Associations between these classes illustrate how they collaborate and exchange information, forming a cohesive structure that underpins the functionalities of the system. The class diagram serves as a blueprint for understanding the system's architecture.

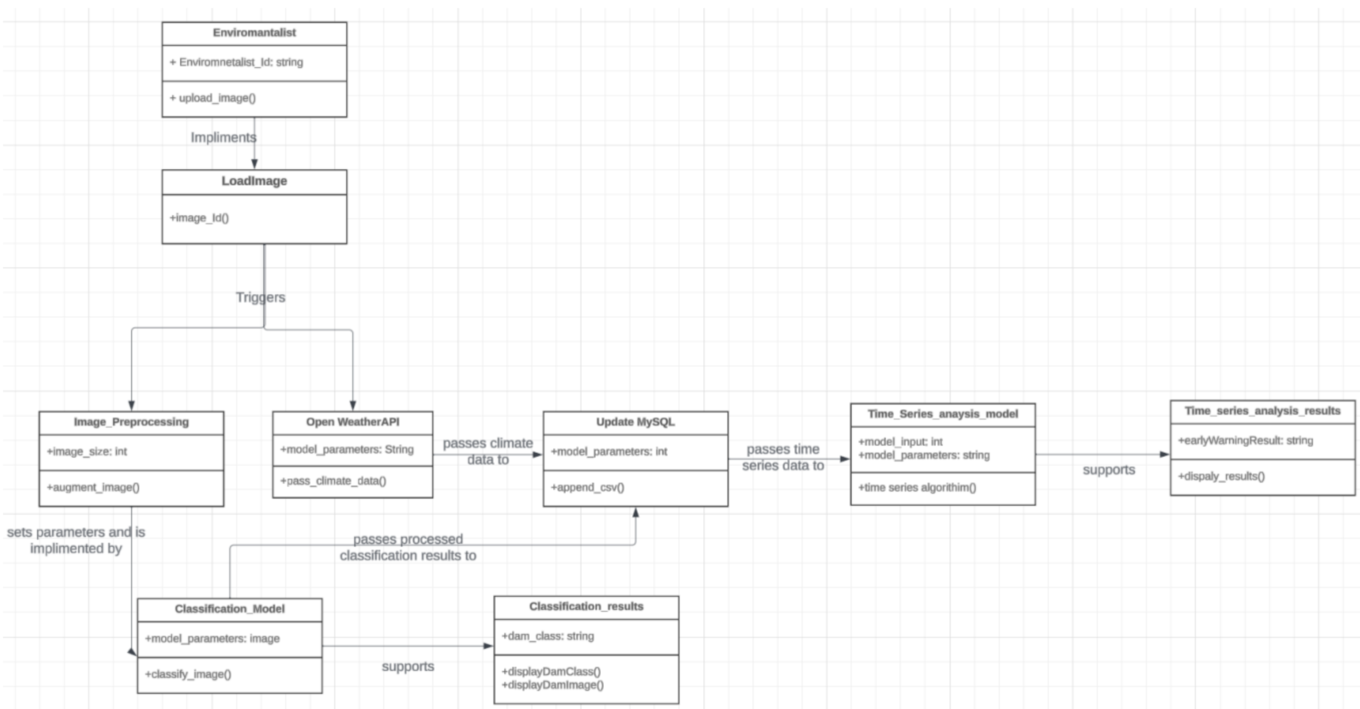


Figure 4-3: Class diagram

4.5.2 System Architecture

Figure 4-4 illustrates how the different components of the system, users, web application interface, image classification model, OpenWeather API, and the Time Series Analysis model interact.

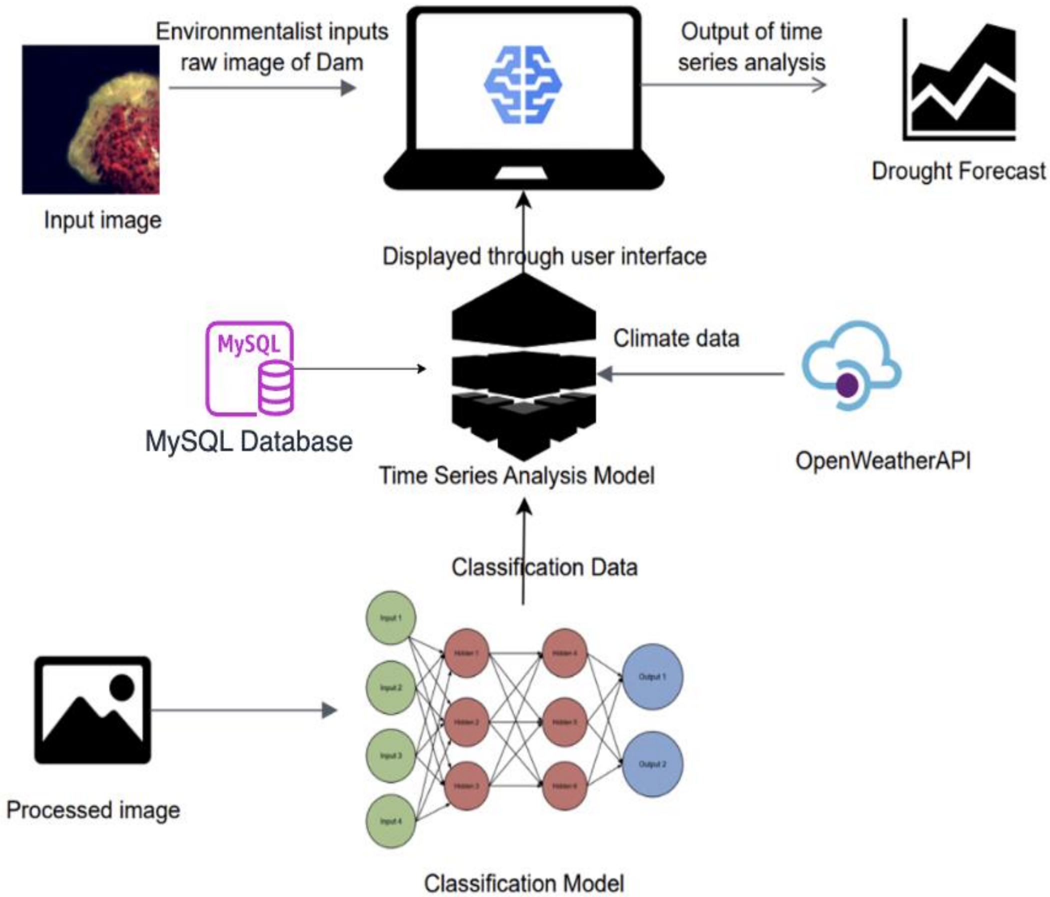


Figure 4-4: System Architecture

Chapter 5 System Implementation and Testing

5.1 Introduction

This chapter reviews the hardware and software specifications required to train and perform inference on the model. Additionally, this chapter reviews the dataset building, preprocessing training, testing, and inference processes of the model and how it is used by the web application.

5.2 Description of the Implementation Environment

The flutter mobile app was implemented from VS code, which required Xcode and Android studio for emulators. The web app was implemented form VS code using streamlit and flask. MySql was used for storing the time series data and Firebase was used to store user credentials.

5.2.1 Hardware Specifications

These are the hardware's technical minimum requirements and configurations for the system. Table 5-1: Hardware Specifications

Table 5-1: Hardware requirements

Hardware Name	**Specifications**	**Justification**
Processor	Apple M1 chip	A powerful processor is required to execute the functions of the system efficiently.
RAM	Minimum RAM of 8GB	Enough storage for data being processed and intermediate outputs should be provided to enable the system to operate well.
GPU	Minimum of 8GB	A GPU is necessary to accelerate the computations of the model to ensure that the results are displayed promptly.
Hard Disk storage	Minimum of 40GB storage space	Sufficient storage is necessary to store the model, its weights, and the data it needs to access.

5.2.2 Software Specifications

These are the software's technical minimum requirements and configurations for the system.

Table 5-2: Software requirements

Software Name	Specifications	Justification
Operating System	macOS with a minimum version of 13.6.1	It manages the computing resources between the processes of the developed project when it is executing
Streamlit	Minimum version of 2.29.0	It allows for fast and simple development and management of web applications through developed APIs.
Flask	Minimum version of 3.0.0	It allows for fast and simple development and management of web applications through developed APIs.
Python	Python 3.12.1	It is supported by the Flask framework used in the development of the web application. It also supports machine learning libraries.

5.3 Description of the Dataset

5.3.1 Classification model

The data was obtained from Kaggle, this dataset focuses on remote sensing image patches depicting water supply dams in the State of São Paulo, Brazil, spanning the years 2015 to 2021. The selected area encompasses nine dams in the drought-prone state of São Paulo, as illustrated below. The data was partitioned into a 70:20:10 ratio for training, testing, and validation.

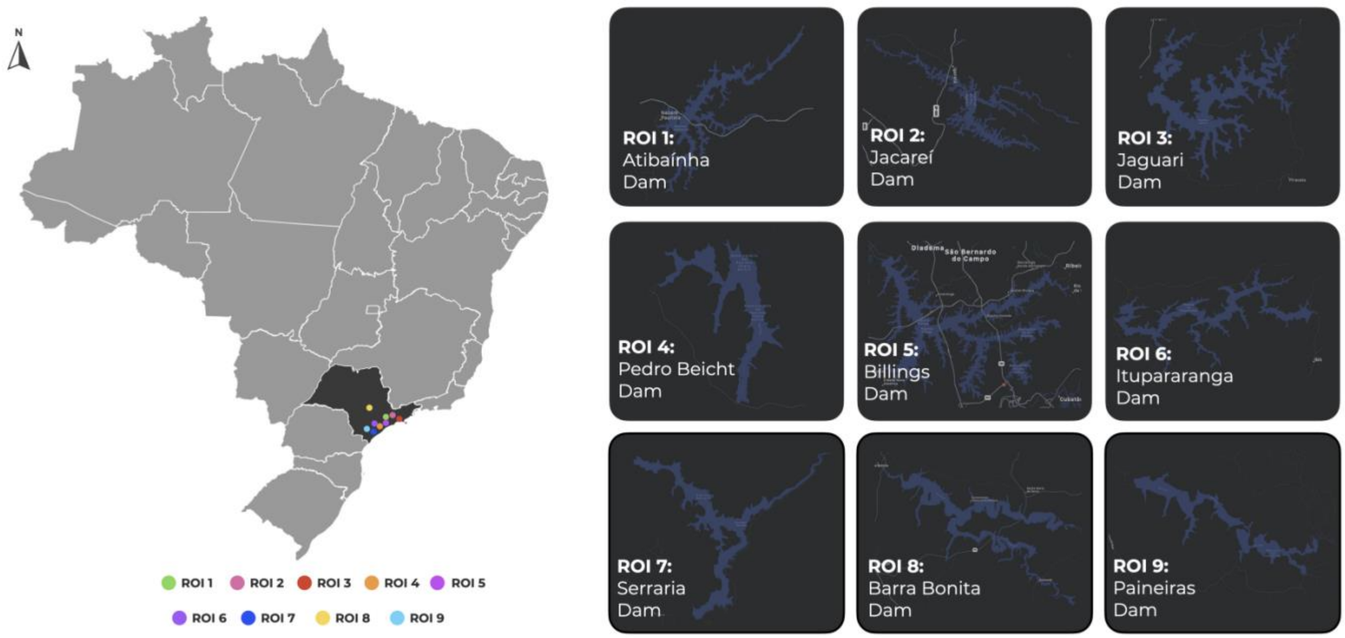


Figure 5-1: Sao paulo study region

The training set comprises 3884 images, each measuring 8.562 × 12.736 pixels and captured by the CBERS-4's PAN10M sensor, boasting a 10-meter spatial resolution. To emphasize dam edges, the NIR, R, and G bands were employed, represented by pink, red, and green colors in Figure 9. The NIR (Near-Infrared), R (Red), and G (Green) bands are specific wavelength ranges used in remote sensing. NIR is sensitive to vegetation health, R is within the visible spectrum and useful for land cover analysis, and G provides information on chlorophyll absorption in vegetation. Utilizing hydrological data from [SABESP 2021], images in the training and testing datasets were classified based on dam volume into:

i) Normal: volume exceeding 60% of total capacity;

ii) Low: volume ranging between 40% and 60% of full capacity;

iii) Critical: volume falling below 40% of total capacity.

Concerning the test and validation dataset, a total of 884 images were created using the NIR, R, and G bands. Notably, 100 images were sourced from the CBERS-4A's WPM camera, featuring multi-spectral and panchromatic lenses with 8 and 2 meters of spatial resolution, respectively. Each image measured 56.842 × 58.344 pixels. The remaining images in the test and validation sets were obtained from the same camera used in the training set. Including these diverse images, especially those from the CBERS-4A's WPM camera, enhances the dataset's robustness and ensures a more comprehensive evaluation of the model's performance across different sensors and resolutions.

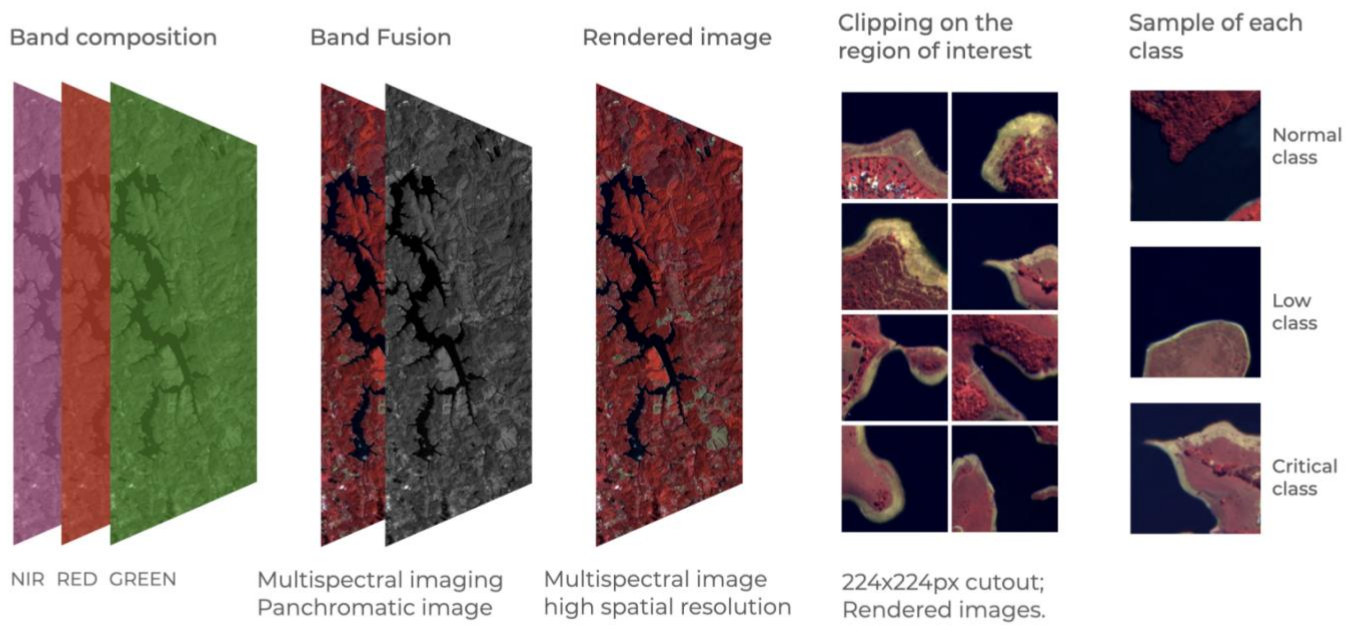


Figure 5-2: Classification image description

5.3.2 Time Series Analysis Model

We focused on the drought that occurred in Sao Paulo from from January to July 2021. Therefore the simulation covered 365 days from October 1st, 2020 to September 30, 2021, where in October 2020 there were more normal conditions which progressively deteriorated as the drought was approaching. The first step was to collect date timestamps encompassing the specified period. After classification, values are assigned to the respective classes as; critical as 15, low as 10, and normal as 5. This is then multiplied by the probability of classification to produce the variable Forecast_value. Respective historic values for corresponding temperature and Humidity between October 1st, 2020, and September 30, 2021, were obtained from Timeanddate.com (timeanddate, s.d.). This compiled formed the Time series data as shown below

	DATE	forecast_value	Temparature	Humidity
0	2020-10-01	3.3665	36	25
1	2020-10-02	3.3665	36	23
2	2020-10-03	3.3665	25	76
3	2020-10-04	3.3665	20	87
4	2020-10-05	3.3665	35	58
...	...	...	...	...
360	2021-09-26	3.3560	29	49
361	2021-09-27	3.3560	29	47
362	2021-09-28	3.3560	32	31
363	2021-09-29	3.3560	26	67
364	2021-09-30	3.3560	22	72

365 rows × 4 columns

Figure 5-3: Time Series Data Snippet

5.4 Data Preprocessing

5.4.1 Classification model

The image data preprocessing is facilitated through the utilization of `ImageDataGenerator`. This tool orchestrates a series of transformations to augment the training dataset, enhancing the model's generalization capabilities. These transformations include rescaling pixel values to a normalized range between 0 and 1. Additionally, diverse data augmentation techniques are employed, such as random rotations, shifts, shearing, zooming, and horizontal flipping. These augmentations artificially expand the training dataset, improving the model's ability to handle variations in real-world data. The specified data directory and batch size, along with the target size for resizing images, contribute to the creation of data generators for both training and testing datasets. The `flow_from_directory` function automatically infers class labels from the directory structure, streamlining the data loading process. Overall, this preprocessing pipeline prepares the input data for the convolutional neural network (CNN) model, facilitating robust training and evaluation on image classification tasks. The code also mounts Google Drive as the data was stored there.

```
# data transformations for data augmentation and normali
train_datagen = ImageDataGenerator(
    rescale=1./255,
    rotation_range=20,
    width_shift_range=0.2,
    height_shift_range=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode='nearest'
)

test_datagen = ImageDataGenerator(rescale=1./255)
```

```
data_dir = '/content/drive/MyDrive/Dataset1'

#data generators
batch_size = 4
train_generator = train_datagen.flow_from_directory(
    os.path.join(data_dir, 'Train'),
    target_size=(224, 224),
    batch_size=batch_size,
    class_mode='categorical'
)

test_generator = test_datagen.flow_from_directory(
    os.path.join(data_dir, 'Test'),
    target_size=(224, 224),
    batch_size=batch_size,
    class_mode='categorical'
)
```

Figure 5-4: Classification preprocessing

5.4.2 Time Series Analysis model

Assigned the class name a value and getting the product of the value and accuracy. Got the timestamp.

```python
# Convert the predicted class index to the class name
class_name = image_datasets['Train'].classes[predicted_class.item()]

# Print the predicted class
print(f"Predicted Class: {class_name}")

print(f'1. Class: {class_name}')
print(f'2. Accuracy: {epoch_acc:.4f}')

# Assign numerical values to classes
numerical_values = {'critical': 15, 'low': 10, 'normal': 5}
numerical_value = numerical_values.get(class_name, 0)

print(f'3. Numerical Value: {numerical_value}')

# Get the timestamp with only the date
timestamp = datetime.strptime(user_date, "%Y-%m-%d").date()

print(f'4. Timestamp (Date): {timestamp}')

# Calculate the product of accuracy with the numerical class value
product = epoch_acc * numerical_value

print(f'5. Product of Accuracy and Numerical Value: {product:.4f}')

# Return timestamp and product
return timestamp, class_name, product
```

Figure 5-5: Product(forecast_value) function

Function used to get the data on current temperature and humidity.

```python
def get_weather_info(api_key, city):
    base_url = "http://api.openweathermap.org/data/2.5/weather"
    params = {
        'q': city,
        'appid': api_key,
        'units': 'metric'
    }

    try:
        response = requests.get(base_url, params=params)
        response.raise_for_status()  # Check if the request was successful
        data = response.json()

        temperature = data['main']['temp']
        humidity = data['main']['humidity']

        return temperature, humidity

    except requests.exceptions.RequestException as e:
        print(f"Error: Unable to fetch data from OpenWeatherMap API. {e}")
        return None, None
```

Figure 5-6: API function

Converted date to acceptable date format:

```
[ ]  print(DFS['DATE'].dtype)

     datetime64[ns]
```

Figure 5-7: Date format conversion

Renamed the columns:

```
df_train = df_train.rename(columns={'forecast_value': 'y', 'DATE':'ds'})
df_train['y_orig'] = df_train['y']
df_train['y'] = np.log(df_train['y'])
```

Figure 5-8: Rename columns function.

5.5 Description of Training

5.5.1 Classification model

The code snippet below constructed a convolutional neural network (CNN) for image classification using the ResNet50 architecture with transfer learning. The ResNet50 base model, pretrained on the ImageNet dataset, was employed to extract high-level features. These features were then fed into a sequential model, which included a global average pooling layer for dimensionality reduction, a dense layer with 512 neurons and ReLU activation, and a final dense layer with 3 neurons for classification using softmax activation. The model was compiled with the Adam optimizer, a categorical crossentropy loss function, and accuracy as the metric. The training loop iterated through a specified number of epochs, during which the model was trained and evaluated on both training and testing datasets. The loss and accuracy metrics were printed for each phase, providing insights into the model's performance over the training process.

```
# ResNet50-based model using Keras
base_model = ResNet50(weights='imagenet', include_top=False, input_shape=(224, 224, 3))

model = models.Sequential()
model.add(base_model)
model.add(layers.GlobalAveragePooling2D())
model.add(layers.Dense(512, activation='relu'))
model.add(layers.Dense(3, activation='softmax'))
```

```
Downloading data from https://storage.googleapis.com/tensorflow/keras-applications/resnet/resnet50_weights_tf
94765736/94765736 [==============================] - 0s 0us/step
```

```
# Compile the model using learning_rate instead of lr
model.compile(optimizer=Adam(learning_rate=0.001), loss='categorical_crossentropy', metrics=['accuracy'])
```

```
# Training loop
num_epochs = 10
for epoch in range(num_epochs):
    for phase in ['Train', 'Test']:
        if phase == 'Train':
            generator = train_generator
        else:
            generator = test_generator

        history = model.fit(generator, epochs=1, verbose=1)

        print(f'{phase} Loss: {history.history["loss"][-1]:.4f} Acc: {history.history["accuracy"][-1]:.4f}')

print("Training complete!")
```

Figure 5-9: Training classification model

Results:

```
971/971 [==============================] - 2576s 3s/step - loss: 0.4896 - accuracy: 0.7889
Train Loss: 0.4896 Acc: 0.7889
221/221 [==============================] - 574s 3s/step - loss: 0.4931 - accuracy: 0.7715
Test Loss: 0.4931 Acc: 0.7715
971/971 [==============================] - 2564s 3s/step - loss: 0.4709 - accuracy: 0.7984
Train Loss: 0.4709 Acc: 0.7984
221/221 [==============================] - 580s 3s/step - loss: 0.4212 - accuracy: 0.8088
Test Loss: 0.4212 Acc: 0.8088
971/971 [==============================] - 2574s 3s/step - loss: 0.4580 - accuracy: 0.7958
Train Loss: 0.4580 Acc: 0.7958
221/221 [==============================] - 575s 3s/step - loss: 0.3985 - accuracy: 0.8145
Test Loss: 0.3985 Acc: 0.8145
971/971 [==============================] - 2581s 3s/step - loss: 0.4469 - accuracy: 0.8133
Train Loss: 0.4469 Acc: 0.8133
221/221 [==============================] - 579s 3s/step - loss: 0.3720 - accuracy: 0.8269
Test Loss: 0.3720 Acc: 0.8269
Training complete!
```

Figure 5-10: Training results

5.5.2 Time Series Analysis model

A new instance of the Prophet class was created, and two regressors, 'Temperature' and 'Humidity', were added to the model. The 'add_regressor' method was utilized to incorporate additional external features (regressors) into the model, allowing it to consider these factors when making predictions. In this case, 'Temperature' and 'Humidity' were assumed to be external factors influencing the target variable.

```
model_new = Prophet() #instantiate Prophet
model_new.add_regressor('Temparature')
model_new.add_regressor('Humidity')
```

```
<prophet.forecaster.Prophet at 0x7bdb03cf9de0>
```

Figure 5-11: Prophet instance

5.6 Description of Testing and Evaluation

5.6.1 Classification model

The model was evaluated using the metrics accuracy, precision, recall and F1 score, the results are as follows:

```
Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.74      0.84       328
           1       0.74      0.98      0.84       274
           2       0.99      0.96      0.97       282

    accuracy                           0.88       884
   macro avg       0.90      0.89      0.89       884
weighted avg       0.91      0.88      0.89       884


Evaluation Metrics:
Loss: 0.3959
Accuracy: 0.8846
```

Figure 5-12: Classification Evaluation

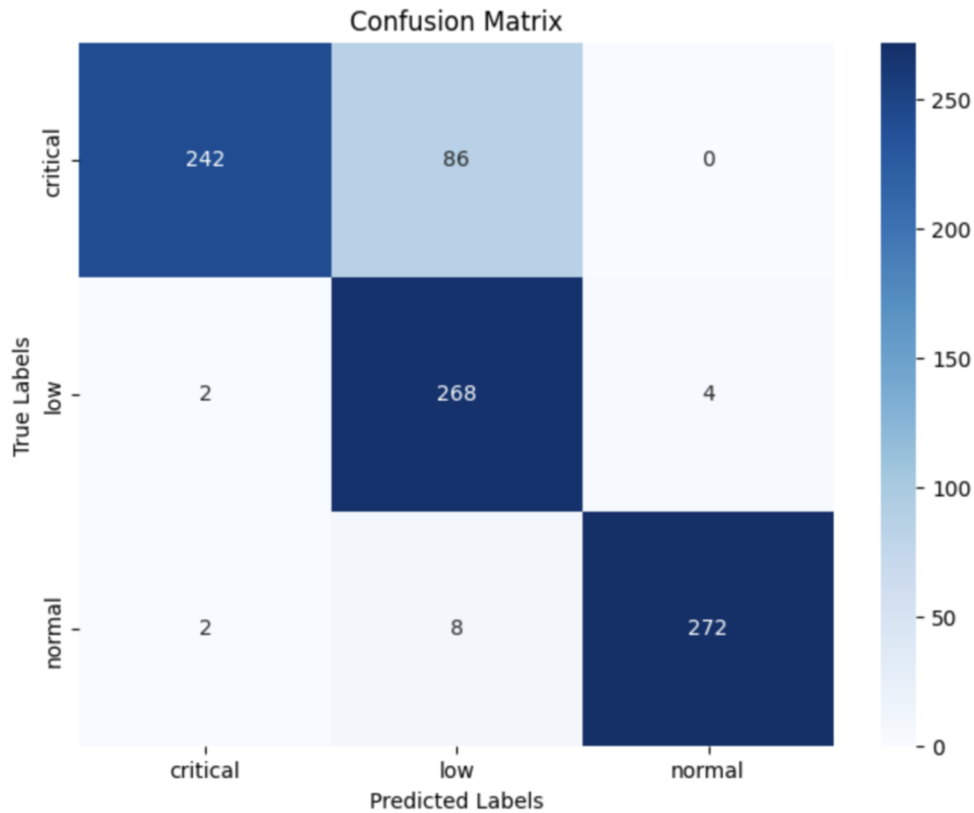


Figure 5-13: Confusion Matrix

5.6.2 Time Series Analysis model

The plot below is a plot of the forecast with regressors in blue and the actual data inferred from the classification in black.

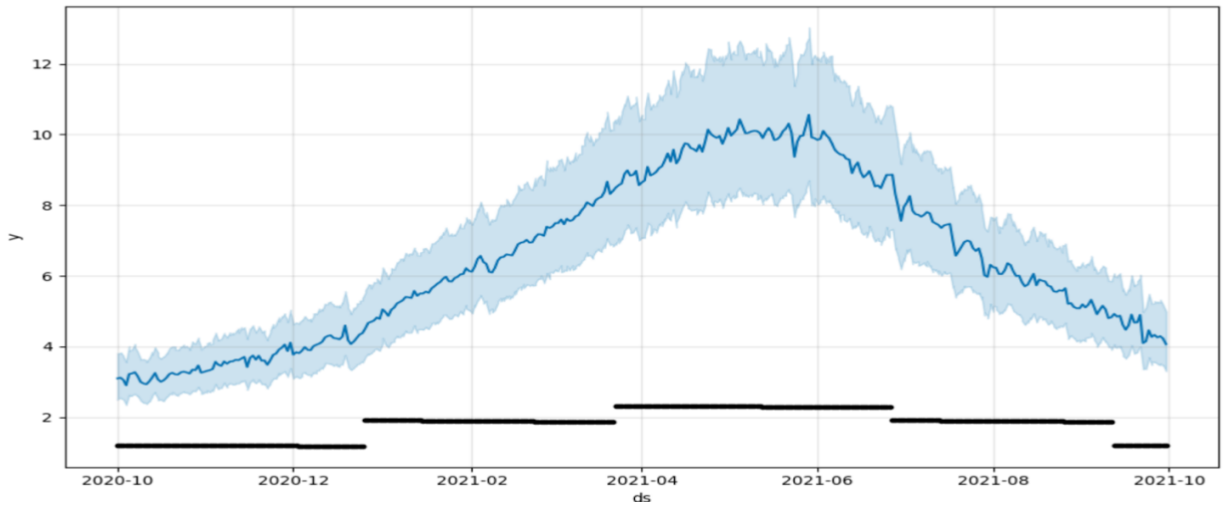


Figure 5-14: Forecast(blue) with classification data(black)

The plot below is a forecast for the next 6 months from the last date with the range highlighted

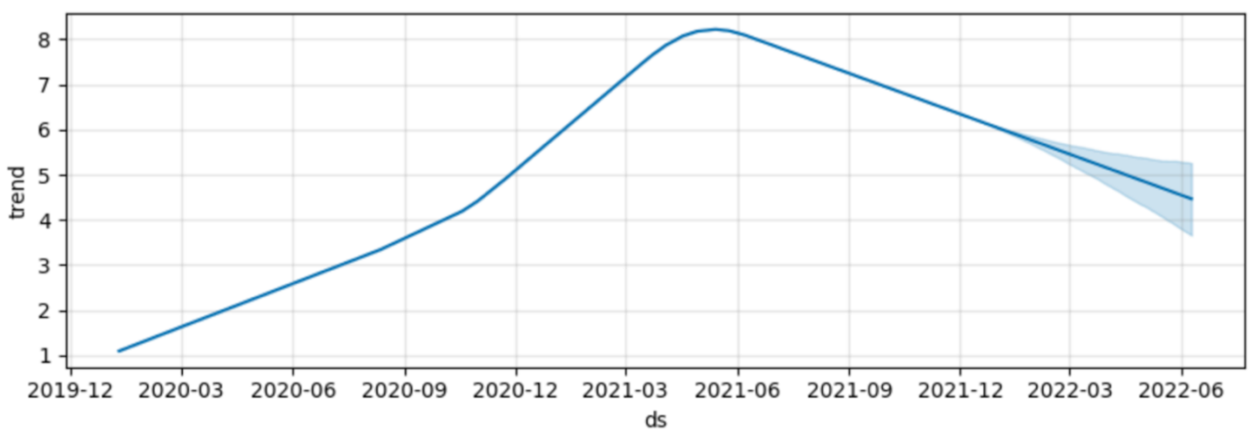


Figure 5-15: Six months forecast.

5.7 System Testing Paradigms

The testing paradigms used in testing the implemented system were black box testing and unit testing.

5.7.1 Black Box Testing

This involved examining the functionality of a system without prior knowledge of the internal structure. This testing was done for all the modules implemented in the system. Test data with predicted known output was used in testing. The expected output was compared with the system's actual output to understand system behavior and identify possible errors.

5.7.2 White Box Testing

White box testing involves evaluating the internal structure of an application. This testing paradigm requires in-depth knowledge of the system and requires that each line of code in the system is evaluated. Due to its exhaustive nature, the testing was done for sensitive modules. The image segmentation module was tested using this paradigm to ensure that there were no mistakes in the code that could lead to incorrect classification of uploaded images.

5.8 Testing Results

The testing results are represented in the table below.

Table 5-3: Testing results

Test Case	Description	Test Data	Expected Outcome	Test Result	Verdict
Registration and Login	Registration with all fields provided	Email-'test@gmail.com' Password-test@123	User details added to the database and user is redirected to landing page	As expected	Pass

	Registration with some fields provided not	One of the fields missing data	Error message is displayed asking user to enter all fields	As expected	Pass
	Registration with existing email	Email -‘test@ gmail.com’ Password - test@123	Error message is displayed informing user that account exists	As expected	Pass
	Login with valid details	Email -‘test@ gmail.com’ Password - test@123	Login is successful and user is redirected to landing page	As expected	Pass
	Login with incorrect details	Email -‘test@ gmail.com’ Password - 12345677	Error message is displayed informing that the details are incorrect	As expected	Pass
Upload Image	Upload Dam image for classification	A sample dam image	Image uploaded successfully	As expected	Pass

Image Classification	Image Classification on uploaded Dam Image	Uploaded image	Image classified correctly	As expected	Pass
Update MySQL database	Add relevant data to database	Classification and climate data	Database updated well	As expected	Pass
Time Series Analysis	Analyze the time series data and forecast 6 months	Time Series Data	Forecast the next 6 months	As expected	Pass
Early warning	Give early warning above the set threshold of 6(y value)	Time series data	Give early warning if above 6	As expected	Pass

Chapter 6 Conclusions, Recommendations and Future Works

6.1 Conclusion

The aim of this research was to develop a system utilizing climate data for the purpose of issuing drought early warnings using a deep CNN and Prophet model. This was aimed to alleviate the challenges brought about by current systems including manual forecasting which involves the collection of data manually, to utilize unused data because the current drought and flood forecasting systems do not use images of water bodies which is a waste of metadata and to implement a more up to date forecasting algorithm as the current and reviewed drought forecasting study uses a seq2seq model and we implemented the more recent prophet algorithm. To address these issues, the project introduces an automated method of inputting climate data through an API. The devised solution incorporates an image classification model using a Deep CNN, which categorizes dam images based on varying water levels: low, normal, and critical. The training set images include NIR, R, and G bands to accentuate dam edges. Subsequently, the system integrates the classification data with additional climate data from the API, specifically temperature and humidity, into a time series forecasting model for predictive analysis. A function is added to the Prophet algorithm to issue a warning in case the forecast is above the set threshold. The resulting model, upon testing and validation, was able to classify the dam images, forecast for drought, and give early warnings.

6.2 Recommendations

It is recommended that the project undergoes a hardware upgrade to enhance computational efficiency and speed during model training and forecasting. The implementation of a system with increased RAM, transitioning from 8GB to at least 16 GB, is suggested to accommodate the complexity of the deep convolutional neural network (CNN) and the forecasting algorithm, particularly when processing large datasets. The model was trained using hosted GPUs. Therefore, due to the size of the model, it is recommended that TPUs be used in the training of the model. This will help accelerate the training of the model and allow for more experiments to be carried out in hyperparameter tuning. It is also recommended that class weights be applied in the training of the model to counter the issue of class imbalances and improve classification performance for each class. In addition, sourcing a larger dataset will affect the model performance positively.

6.3 Future works

Training the model using a different algorithm like transformers could lead to better classification results. Including more climate factors as regressors could lead to better drought prediction. Lastly, collaboration with relevant stakeholders, such as meteorological agencies, can enhance the integration of real-time climate data such as the vegetation indices which are obtained by directly reading the values of the near-infrared (NIR) and red (R) bands from satellites, ensuring the model remains up-to-date and capable of providing timely and accurate drought early warnings.

Bibliography

Adisa, n.d., O. M. (n.d.). *Bibliometric Analysis of Methods and Tools for Drought Monitoring and Prediction in Africa.* Retrieved from MDPI: https://www.mdpi.com/2073-4441/13/24/3498

Behzad, n.d., A. (n.d.). *Hydrological drought persistence and recovery over the CONUS: A multistage framework considering water quantity and quality.* Retrieved from ScienceDirect: https://www.sciencedirect.com/science/article/abs/pii/S0043135418309813

Bengio, Y. (n.d.). *Deep learning*. Retrieved from Nature: https://www.nature.com/articles/nature14539

Burst, n.d., C. (n.d.). *DroughtCast: A Machine Learning Forecast of the United States Drought Monitor.* Retrieved from Frontiers: https://www.frontiersin.org/articles/10.3389/fdata.2021.773478/full

California Department of Water Resources, [Kelly M. Grow]. (n.d.). Temperature and Drought A SCIENCE ASSESSMENT BY A SUBGROUP OF THE DROUGHT TASK FORCE. *Temperature and Drought*. https://repository.library.noaa.gov/view/noaa/17720/noaa_17720_DS1.pdf?

Center, N. D. (n.d.). *Future Opportunities and Challenges in Remote Sensing of Drought.* Drought Mitigation Center Faculty Publications.

Center, N. D. (n.d.). *How Does Drought Affect Our Lives?* Retrieved from National Drought Mitigation Center:

https://drought.unl.edu/Education/DroughtforKids/DroughtEffects.aspx committee, I. r. (n.d.). *Drought Explainer*. Retrieved from International rescue committee: https://www.rescue.org/uk/article/what-drought-causes-impact-and-how-we-canhelp

Communications, I. G. (n.d.). *International Rescue Committee*. Retrieved from International Rescue Committee: https://www.rescue.org/press-release/irc-severedroughtprojected-leave-about-54-million-people-kenya-withoutadequate#:~:text=%ED%95%9C%EA%B5%AD,IRC%3A%20Severe%20drought%20is%20projected%20to%20leave%20about%2 05.4%20million,between%20March%20and%20June

dessertification, U. N. (n.d.). *Monitoring and early warning*. Retrieved from United Nations Convention to combact dessertification: https://www.unccd.int/landandlife/drought/toolbox/monitoring-and-early-warning

Hayes, n.d., M. J. (n.d.). *Drought Monitoring: Historical and Current Perspectives.* Retrieved from Drought Monitoring: Historical and Current Perspectives: https://core.ac.uk/download/pdf/188138949.pdf

Intellij. (n.d.). Retrieved from Phpstorm: https://www.jetbrains.com/phpstorm/

Kothari. (n.d.). *RESEARCH DESIGNS AND METHODOLOGY.* Retrieved from RESEARCH DESIGNS AND METHODOLOGY: http://www.kenpro.org/researchdesign-and-methodology/

Loon, n.d., A. F. (n.d.). *Hydrological drought explained.* Retrieved from WIREsWater: https://wires.onlinelibrary.wiley.com/doi/full/10.1002/wat2.1085

Microsoft. (n.d.). *Visual Studio Code*. Retrieved from https://code.visualstudio.com/

Nimble. (n.d.). *What Is Scrum Methodology? & Scrum Project Management.* Retrieved from

Nimble: https://www.nimblework.com/agile/scrum-methodology/

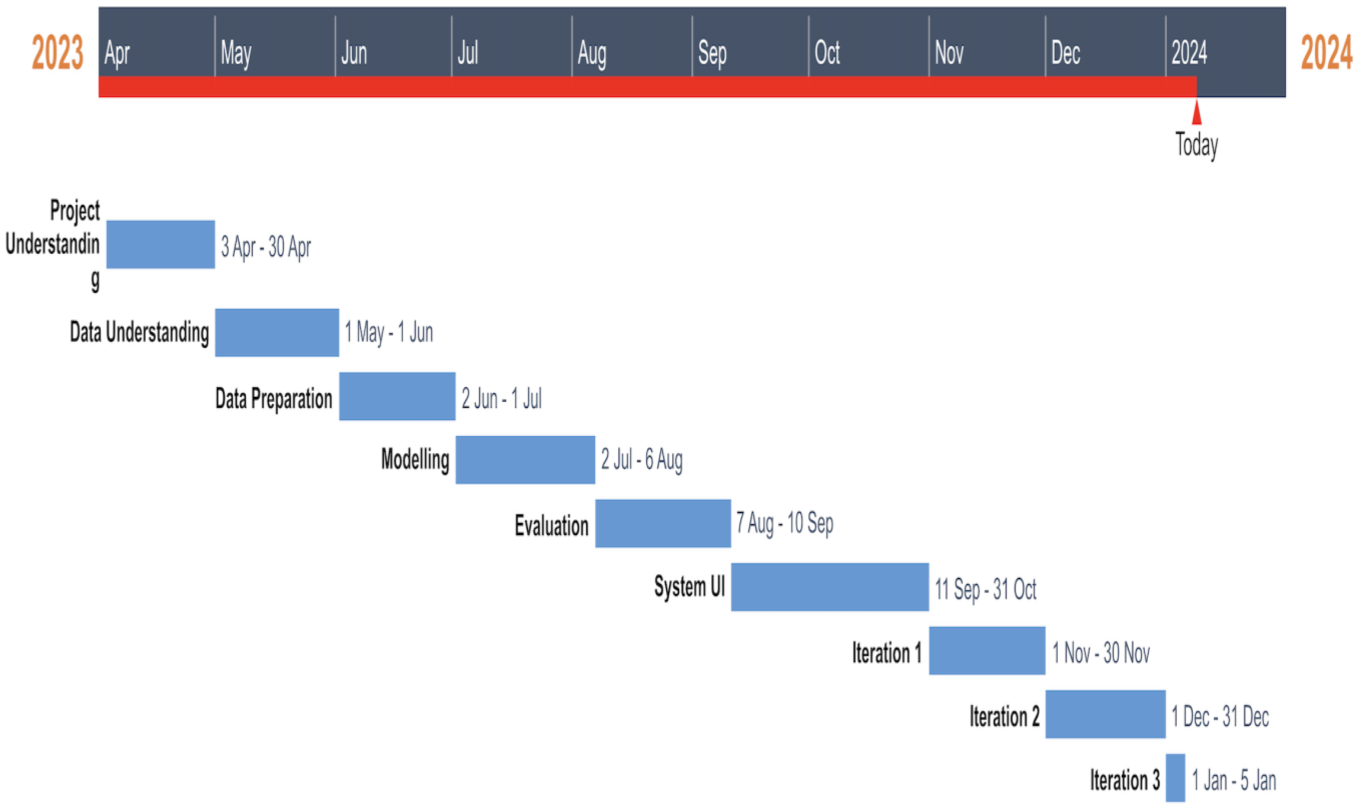
Portal, S. (n.d.). *Space technologies for drought monitoring and management*. Retrieved from Space4Water Portal: https://www.space4water.org/news/spacetechnologiesdrought-monitoring-andmanagement#:~:text=Thus%2C%20indicators%20such%20as%20Standardized ,dro ught%20occurrences%20(Aghakouchak%20et%20al.

Society, N. G. (n.d.). *Drought*. Retrieved from National Geography: https://education.nationalgeographic.org/resource/drought/

Xiao, n.d., L. (n.d.). *Runoff Forecasting Using Machine-Learning Methods: Case Study in the Middle Reaches of Xijiang River.* Retrieved from Frontiers: https://www.frontiersin.org/articles/10.3389/fdata.2021.752406/full

Yan, n.d., L. (n.d.). *Climate-informed monthly runoff prediction model using machine learning and feature importance analysis.* Retrieved from Frontiers: https://www.frontiersin.org/articles/10.3389/fenvs.2022.1049840/full#:~:text=We% 20proposed%20the%20model%20by,factors%20into%20the%20ML%20models.

Gantt Chart



2023
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2024
2024
Today
Project Understanding
3 Apr - 30 Apr
Data Understanding
1 May - 1 Jun
Data Preparation
2 Jun - 1 Jul
Modelling
2 Jul - 6 Aug
Evaluation
7 Aug - 10 Sep
System UI
11 Sep - 31 Oct
Iteration 1
1 Nov - 30 Nov
Iteration 2
1 Dec - 31 Dec
Iteration 3
1 Jan - 5 Jan