AquaAlert: An AI-Powered Early Water Scarcity Alarming System

Project Report Submitted in Partial Fulfilment of the Requirements for the Degree of

Bachelor of Technology

in

Computer Science and Engineering

Submitted by

Harshit Yadav(Roll No. 2210110298)
Shreeya Arora (Roll No. 2210110569)
Umang Dwivedi(Roll No. 2210110633)

Under the Supervision of

Dr. Suchi Kumari

Department of Computer Science and Engineering

SHIV NADAR
INSTITUTION OF EMINENCE DEEMED TO BE
UNIVERSITY
DELHI NCR

Shiv Nadar University

October 2025

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented, fabricated, or falsified any idea, data, fact, or source in our submission. We understand that any violation of the above will be cause for disciplinary action by the University and may also invoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Name of the Student	**Signature and Date**
Harshit Yadav	15-10-2025
Shreeya Arora	15-10-2025
Umang Dwivedi	15-10-2025

Abstract

Water scarcity poses a critical threat to sustainable development, particularly in data-scarce and rapidly urbanizing regions such as India. *AquaAlert* proposes an AI-powered early warning system designed to forecast and visualize potential water stress conditions across Indian states. The system integrates multi-source datasets, including groundwater levels, rainfall, soil moisture, and demographic demand into a unified analytical pipeline. Machine learning models such as Random Forest, Gradient Boosting, and Neural Networks are employed to predict regional water scarcity patterns, which are then quantified through a Composite Water Stress Index (WSI) aligned with NITI Aayog's Composite Water Management Index (CWMI). Results are visualized via an interactive heatmap interface to support proactive policy and community level decision making. The study further outlines ongoing work on WSI optimization and hybrid learning models to enhance predictive reliability and operational scalability.

Name of Supervisor: Dr. Suchi Kumari

Signature and Date: ______________________________

AquaAlert: An AI-Powered Early Water Scarcity Alarming System

I. Introduction

Water scarcity has emerged as one of the most pressing environmental and socio-economic challenges of the 21st century, particularly in developing nations such as India. Rapid population growth, urbanization, climate variability, and unsustainable water consumption patterns have significantly widened the gap between water demand and supply. According to NITI Aayog's *Composite Water Management Index*, nearly 600 million Indians currently face high to extreme levels of water stress, highlighting the urgent need for proa

AquaAlert addresses this challenge by proposing an AI-powered Early Water Scarcity Warning System that leverages data-driven insights for sustainable water management. The system integrates environmental parameters, including rainfall, groundwater levels, population density, soil moisture, consumption rates, and reservoir capacities. By applying advanced machine learning models, AquaAlert predicts potential water scarcity trends across Indian states.

The predicted results are visualized through an interactive heatmap interface, enabling policymakers, researchers, farmers, and citizens to effectively monitor regional water stress levels. Through this intelligent, data-centric approach, AquaAlert aims to bridge the gap between prediction and prevention—empowering early intervention and fostering sustainable resource utilization across India

II. Literature Survey

The literature survey for AquaAlert encompasses a robust spectrum of contemporary research and operational frameworks central to AI-powered water scarcity prediction and management. Beginning with Bahri and Dasarathy (IJISAE, 2024), the survey highlights the effectiveness of statistical trend tests and machine learning models—Random Forest, Gradient Boosting, and Neural Networks,in modelling groundwater level fluctuations, establishing the operational potential of ML for identifying intervention zones and supporting region-level scarcity forecasting. The NITI Aayog's Composite Water Management Index (CWMI, 2023) brings policy relevance by providing standardized evaluation metrics and reference indicators that calibrate and validate AquaAlert's model outputs to national water management benchmarks. Data from the Central Ground Water Board's annual reports serve as foundational operational inputs, informing groundwater-related features and guiding data validation for AquaAlert's predictive analytics, while MOSPI's environmental and social statistics enable context-sensitive demand estimation by integrating socio-demographic drivers like population density and urbanization. Urban system inefficiencies and governance issues are explored through IWA's case study on Bengaluru's water challenges (Water Policy, 2024), supporting AquaAlert's expanded inclusion of operational and policy dimensions in its predictors.

Recent advances in remote sensing and integrated modelling (ScienceDirect/SAGE, 2023–2024) further underscore the value of satellite derived hydrological indicators—particularly GRACE/GRACE-FO data—when combined with ground-based observations, enabling enhanced spatial-temporal forecasting accuracy for large-scale prediction. Building on these foundations, additional literature including Wang et al. (2024) demonstrates the utility of Capsule Neural Networks for improved urban water stress forecasting; Ibrahim et al. (2024) reviews state of the art remote sensing technologies for groundwater monitoring, reinforcing AquaAlert's technical pipeline; Gautam et al. (2024) validates GRACE datasets for subsurface groundwater estimation; and Zheng et al. (2024) advances ensemble methodologies for accurate water parameter prediction. Jointly, these sources affirm the importance of integrating cutting-edge AI/ML, national-level benchmarking, authoritative operational data, and multi-sensor modelling to create a scalable, policy aligned, and technically advanced warning system for water scarcity.

Despite these advancements, notable research and implementation gaps persist. Integration between emerging neural architectures (e.g., Capsule Networks, GRUs, hybrid ensembles) and conventional ML models remains limited, reducing robustness across spatial and multi-sensor data contexts. Existing frameworks often lack real time multi-source fusion, combining satellite, ground, and socio-economic inputs only partially, which affects contextual accuracy. Moreover, validation standards rarely align with national benchmarks like CWMI, weakening policy relevance. Finally, few studies emphasize visualization or actionable early warning outputs for stakeholders. AquaAlert addresses these deficiencies through its hybrid AI design, integrated data fusion, CWMI-aligned Water Stress Index, and interactive visualization interface—bridging methodological, infrastructural, and policy gaps in existing research. .

III. Problem Statement

Water scarcity has become a major socio-environmental challenge in India, affecting agriculture, domestic consumption, and economic stability. Declining groundwater levels, erratic monsoon patterns, and unsustainable extraction practices have led to frequent shortages, reduced crop productivity, and unequal access to safe water, particularly in rural regions. These conditions have intensified rural distress, migration, and regional disparities in water availability. Although vast hydrological and environmental datasets exist, current monitoring frameworks remain fragmented and largely descriptive, lacking the predictive intelligence required for proactive management. They provide historical assessments but fail to forecast emerging scarcity zones or quantify future risks. AquaAlert seeks to bridge this gap by leveraging historical, environmental, and socio-economic data to forecast potential water scarcity across Indian states. Using machine learning-based modelling and spatial visualisation through an interactive heatmap, the system aims to support policymakers, planners, and communities in anticipating shortages and implementing timely interventions, transforming water management from reactive response to data-driven prevention.

IV. Objectives

The Primary objective of AquaAlert is to design and develop an AI-driven predictive intelligence system that forecasts future water scarcity patterns across Indian states and districts. The system aims to transform traditional water monitoring into a strategic foresight model that empowers policymakers, researchers, and citizens to make proactive decisions.

1) Develop a machine learning–based forecasting framework that predicts future water scarcity levels using multi-year hydrological, climatic, and socio-economic datasets.
2) Integrate multi-source data pipelines into a unified, high-quality analytical model, including:
   - Groundwater levels from India-WRIS GW Time Series Data,
   - Rainfall and soil moisture data from data.gov.in,
   - Population and water demand estimates extracted from Aadhaar-based demographic datasets, with per capita consumption calculated using LPCD formulas.
3) Construct a Water Stress Index (WSI) that quantifies predicted scarcity severity and aligns with the NITI Aayog Composite Water Management Index (CWMI 2.0) benchmarks to ensure policy relevance and comparability
4) Visualise predicted scarcity trends through an interactive, colour-coded heatmap interface, providing intuitive access to insights for experts, planners, and the general public.
5) Generate actionable early-warning insights and recommendations to support stakeholders in planning recharge programs, optimising water demand management, and implementing sustainable allocation strategies ahead of critical shortages.

V. Methodology

The proposed methodology for the AquaAlert system is structured into five key stages: data collection, preprocessing, machine learning-based forecasting, water stress index calculation, and visualization with interpretation.

A. Data Collection

The system integrates multiple data sources to build a comprehensive dataset:

- **Groundwater Levels:** Collected from the India-WRIS Groundwater Time Series Data for monitoring seasonal and long-term groundwater trends.
- **Rainfall and Soil Moisture:** Retrieved from `data.gov.in`, providing historical climatic and hydrological context.
- **Population Density:** Extracted from Aadhaar-based demographic datasets for accurate state- and district-level estimates. Monthly per capita water consumption (*LPCD*) is calculated as:

$$LPCD = \frac{\text{Monthly Water Consumption}}{\text{Population} \times \text{Days in Month}}$$

B. Data Preprocessing

Collected datasets undergo preprocessing to ensure quality and consistency before modeling:

- **Cleaning:** Handling missing or inconsistent entries across groundwater, rainfall, and socio-economic datasets.
- **Normalization:** Standardizing features to ensure uniform input into machine learning algorithms.
- **Temporal Aggregation:** Aggregating raw time-series data into monthly or seasonal intervals to capture temporal trends.
- **Feature Engineering:** Integrating climatic, hydrological, and socio-economic variables into combined predictive features.

C. Machine Learning-Based Forecasting

Machine learning models are employed to predict water scarcity levels across regions.

- **Model Selection:** Supervised learning algorithms such as Random Forest (RF), Gradient Boosting Machines (GBM), and Neural Networks (NN) are considered for time-series forecasting.
- **Training and Validation:** Historical data is split into training and validation sets. Model performance is evaluated using metrics such as Root Mean Squared Error (RMSE) and the coefficient of determination (R^2).
- **Water Stress Index (WSI) Computation:** Predicted water scarcity levels are transformed into a standardized Water Stress Index aligned with NITI Aayog's CWMI benchmarks.

D. Water Stress Index (WSI) Calculation

A composite WSI is constructed by integrating multiple hydrological and socio-economic indicators:

- **Standardized Precipitation Index (SPI):**

$$SPI = \Phi^{-1}\big(F_{\text{rain},s,m}(\text{rainfall})\big)$$

- **Standardized Soil Moisture Index (SSI):**

$$SSI = \Phi^{-1}\big(F_{\text{sm},s,m}(\text{soil moisture})\big)$$

Alternatively, a z-score may be used and classified similarly to SPI.

- **Standardized Groundwater Index (SGI):**

$$SGI = \Phi^{-1}\big(F_{\text{gw},s,m}(\pm\text{groundwater level})\big)$$

- **Per Capita Water Consumption (LPCD):**

$$LPCD = \frac{\text{Monthly Water Consumption}}{\text{Population} \times \text{Days in Month}}$$

The composite Water Stress Index is calculated as:

$$\begin{aligned} \text{WSI} =&0.25 \cdot \min\{1, \max(0, -SPI/2)\}+ \\ &0.25 \cdot \min\{1, \max(0, -SSI/2)\}+ \\ &0.25 \cdot \min\{1, \max(0, -SGI/2)\}+ \\ &0.25 \cdot \min\{1, \max(0, (LPCD - 135)/65)\} \end{aligned}$$

The resulting WSI is then classified into four categories—*Low*, *Moderate*, *High*, and *Extreme*—to enable clear interpretation and decision-making.

E. Visualization and Interpretation

The final stage of the AquaAlert system focuses on presenting predictive outcomes in an intuitive and actionable format. The visualization framework transforms analytical results into comprehensible spatial and interactive representations.

- **Interactive Heatmap:** A color-coded, district- and state-level map visualizes predicted scarcity zones based on the computed Water Stress Index (WSI). The heatmap dynamically represents spatial variations in water stress, aiding regional comparison and prioritization.
- **User Interface:** A web-based dashboard is designed for policymakers, planners, and citizens to intuitively interpret predicted risks. The interface presents region-wise WSI values, trend charts, and filter controls for user-defined analysis.
- **Scenario Analysis:** The system enables users to explore multiple water stress scenarios by adjusting key input variables such as rainfall deficits, groundwater levels, or population growth. This functionality allows decision-makers to simulate potential outcomes under different climatic or demographic conditions.

F. Early-Warning Insights

The predictive intelligence generated by AquaAlert translates into actionable guidance for both policymakers and communities. These early-warning insights support proactive water management and equitable resource distribution.

- **Policy Recommendations:** Based on predicted high-stress zones, stakeholders can implement targeted interventions such as groundwater recharge programs, optimized irrigation schedules, and sustainable water allocation strategies.
- **Community Awareness:** Visual reports and dashboards enable local communities to anticipate water scarcity and adopt preventive measures. The participatory visualization component encourages collective action toward water conservation and sustainable usage.

VI. Limitations

The development of AquaAlert faced several limitations primarily related to data availability and quality. Many of the datasets used, such as rainfall, groundwater, and soil moisture records, exhibited temporal inconsistencies, incomplete spatial coverage, and restricted accessibility due to the absence of unified APIs or open data channels. These factors limited the granularity and continuity of the analysis across districts. Additionally, the predictive models may exhibit reduced accuracy in regions with sparse or irregular historical data, underscoring the need for improved data standardization and broader dataset integration in future iterations.

VII. Future Work

The next phase of the AquaAlert project will focus on model optimization, performance benchmarking, and system validation to transition the framework from research design to an operational prototype. Two primary directions will guide this phase: Water Stress Index (WSI) optimization and hybrid machine learning model refinement.

A. Water Stress Index (WSI) Optimization

In the upcoming stage, the static weighted formulation of the Water Stress Index will be extended through a Principal Component Analysis (PCA)-based weighting mechanism. This approach aims to determine the contribution of each key variable rainfall, soil moisture, groundwater level, and population driven demand by quantifying their explained variance within the first principal component. The resulting weights (w_i) will dynamically adjust based on regional and temporal variability, making the index empirically grounded and adaptive to diverse hydroclimatic contexts. Comparative evaluation between the equal-weight and PCA-weighted formulations will be conducted to identify the most robust and generalizable index representation.

B. Hybrid Model Development and Benchmarking

AquaAlert's predictive layer will be enhanced using hybrid learning architectures integrating Random Forest (RF), XGBoost, and Long Short-Term Memory (LSTM) models. Two alternative forecasting paradigms will be evaluated: (1) direct WSI prediction as a categorical classification problem (*Low*, *Moderate*, *High*, *Extreme*) and (2) input-based forecasting using time series models such as Prophet or ARIMA for each hydrological component, followed by index aggregation. Each model will undergo hyperparameter tuning and cross-validation, with performance assessed using regression metrics (RMSE, R^2) and classification metrics (Accuracy, F1-score). The optimal hybrid framework will be selected based on predictive accuracy, robustness, and computational efficiency.

C. Validation and Deployment

Subsequent validation will compare predicted WSI outcomes against historical drought reports and official assessments, such as the *First Census of Water Bodies*. This retrospective benchmarking will ensure real-world reliability and alignment with ground-truth scarcity conditions. The final optimized pipeline including data integration, ML forecasting, and the advisory engine. The complete system, documentation, and final analytical results will be compiled and presented in the CSD493 Project-1 Monsoon 2025 final report.

References

[1] Central Ground Water Board (CGWB), *Ground Water Year Book – India 2022-2023*. Government of India, 2023. [Online]. Available: https://cgwb.gov.in/

[2] Y. Wang, J. Li, and M. Zhao, "Capsule Neural Network Approach for Water Stress Prediction," *Environmental Modelling & Software*, vol. 175, pp. 101–118, 2024.

[3] NITI Aayog, *Composite Water Management Index 2.0*. Government of India, 2023. [Online]. Available: https://www.niti.gov.in/

[4] R. Gautam, P. Singh, and S. Das, "Assessing Groundwater Variations Using GRACE Satellite Data: A Machine Learning Perspective," *Remote Sensing of Environment*, vol. 295, pp. 112–123, 2024.

[5] C. Bahri and B. Dasarathy, "AI/ML for Groundwater-Level Forecasting — Arkavathi Watershed," *International Journal of Intelligent Systems and Applications in Engineering*, 2024.

[6] IWA Publishing, "Building Bengaluru's water future: identifying challenges and a strategic blueprint for sustainable management," *Water Policy Journal*, 2024, doi: 10.2166/wp.2025.086.

[7] A. Ibrahim, L. Chen, and M. Rahman, "Remote Sensing Technologies for Groundwater Monitoring: An Integrated Review," *Hydrology and Earth System Sciences*, vol. 28, no. 3, pp. 2021–2038, 2024.

[8] Ministry of Statistics and Programme Implementation (MOSPI), *Social Statistics in India: Section 6 - Water Supply and Sanitation*. Government of India, 2016. [Online]. Available: https://mospi.gov.in/

[9] Q. Zheng, Y. Liu, and R. Kumar, "Ensemble Modelling for Water Parameter Prediction Using Hybrid ML Frameworks," *Journal of Hydrology*, vol. 635, pp. 129–144, 2024.