



Research Paper

CONNECTING FLOODWATER, RAINFALL AND RESERVOIR DATA TO INTELLIGENT WATER REDISTRIBUTION FOR DROUGHT-RESILIENT RESOURCE MANAGEMENT - (AQUA RESCUE)

¹ Mr. R. Ramu, ² Golla Indu Prasanna, ³ Mallavalap Divya Reddy, ⁴ Boddapati Manjunadh, ⁵ Pembarthi Rohith

¹Assistant Professor, ²³⁴⁵Students

Department of CSE (Data Science)

Siddhartha Institute of Technology & Sciences, Narapally

ramuramavath@siddhartha.org.in, 24tq1a6716@siddhartha.co.in, 24tq1a6730@siddhartha.co.in,
24tq1a6748@siddhartha.co.in, 25tq5a6703@siddhartha.org.in

ABSTRACT

Many regions of India face floods and droughts in the same year, sometimes within a few hundred kilometres of each other. During a heavy monsoon spell, reservoirs in one basin release large volumes of water to protect their dams, and much of this water flows to the sea, while districts in a neighbouring basin run short of drinking and irrigation water a few months later. The data that could help manage this imbalance, including rainfall records, reservoir storage, river gauge levels and flood extents, is collected by different agencies and is rarely analysed together. This paper presents Aqua Rescue, a data pipeline and decision support system that connects these sources to plan the redistribution of surplus water to drought-prone areas.

The pipeline ingests daily rainfall from rain gauge networks and gridded products, reservoir levels and releases from water resources department bulletins, river discharge from gauging stations, groundwater levels from observation wells and flood extent maps derived from satellite radar imagery. Each source is cleaned, converted to common units and mapped to a shared spatial framework of basins, sub-basins, reservoirs and districts. A curated data warehouse keeps both the daily history and the current state of every reservoir and district, and data quality checks catch missing bulletins, unit errors and sudden unrealistic jumps in storage values.

Machine learning models use this curated data in three ways. A Long Short-Term Memory network forecasts reservoir inflows for the next fourteen days from recent rainfall, upstream discharge and catchment soil moisture. A gradient boosting model estimates the water demand of each district for drinking, irrigation and industry using population, cropping pattern and temperature data. A drought risk score is computed from the Standardised Precipitation Index, groundwater decline and storage deficit. The forecast surplus and the estimated deficits are then passed to a linear programming optimiser that recommends how much water to move through existing canals, link channels and pumping schemes while respecting capacity, flood cushion and minimum environmental flow constraints.

I. INTRODUCTION

Water availability in India depends heavily on the south-west monsoon, which delivers most of the annual rainfall in about four months. The distribution of this rainfall is uneven across space and time. A single week of intense rain can fill reservoirs to their safe limits and force large releases downstream, causing floods in low-lying villages, while other basins receive below-normal rainfall and enter the summer with low storage. The same state may therefore declare flood relief in some districts and drought relief in others within the same year, and drinking water is transported by tanker to villages that lie within reach of canals that carried surplus water only months earlier.

The decisions that govern reservoir releases and inter-basin transfers are usually made using rule curves, experience and daily bulletins. Rainfall is recorded by the meteorological department, reservoir levels by the irrigation department, river flows by central and state water agencies, groundwater levels by another board, and flood extents by disaster management and remote sensing centres. Each agency publishes its own reports in its own format and at its own frequency. When an engineer wants to know whether the surplus in one reservoir could be used to fill tanks in a drought-hit district, the information must be gathered manually from several sources, and by the time it is assembled the opportunity has often passed.

Advances in data engineering and machine learning make a more coordinated approach possible. Automated pipelines can collect bulletins, gauge readings and satellite products as soon as they are published, clean and align them to a common hydrological framework, and store them in a warehouse suitable for analysis. Forecasting models can predict inflows days in advance, demand models can estimate how much water each district will need, and optimisation methods can work out feasible transfers within the physical limits of canals and pumps. Together these steps turn scattered observations into a timely redistribution plan.

II. LITERATURE SURVEY

Reservoir inflow forecasting has a long history in hydrology. Conceptual rainfall-runoff models represent a catchment as a set of connected storages for soil moisture, groundwater and channel flow, and they are calibrated using observed rainfall and discharge. These models are physically interpretable but require careful calibration and detailed catchment information that is often missing for Indian basins. Data-driven approaches have therefore gained attention. Studies using artificial neural networks, support vector regression and, more recently, Long Short-Term Memory networks report that learned models can match or exceed conceptual models when enough history is available, especially for forecasts a few days ahead. Researchers note that performance depends strongly on the input set, and that upstream discharge, antecedent rainfall over several days and a soil moisture indicator are among the most informative predictors. Several works also emphasise evaluating forecasts with hydrological measures such as the Nash-Sutcliffe efficiency rather than simple error values.

Drought monitoring research has produced a family of indices that describe different aspects of water shortage. The Standardised Precipitation Index measures rainfall deficit over chosen time scales, the Standardised Precipitation Evapotranspiration Index adds the effect of temperature, and vegetation indices from satellite imagery

reflect agricultural stress. Hydrological drought is commonly measured through reservoir storage and streamflow deficits, and groundwater drought through the decline of well levels relative to seasonal norms. Studies agree that no single index describes drought completely and that composite indicators combining meteorological, agricultural and hydrological measures give a more reliable picture for administrative decisions. Machine learning has been used to predict these indices ahead of time, allowing agencies to prepare relief measures before shortages become severe.

Flood mapping from satellite data is another well-developed area. Synthetic aperture radar sensors can observe the ground through cloud cover, which is a major advantage during monsoon storms, and open water appears dark in radar images because it reflects the signal away from the sensor. Threshold-based and machine learning classifiers have been applied to these images to delineate inundated areas, and freely available radar data now allows flood maps to be produced within a day or two of an event. Research has combined such maps with elevation models to estimate flood volumes and with population data to estimate exposure. Only a few studies, however, have used flood extent and volume information as an indicator of surplus water that could be captured and redirected, rather than purely as a hazard to be reported.

Water allocation and reservoir operation have been studied extensively with optimisation methods. Linear programming, dynamic programming and evolutionary algorithms have been used to determine releases that balance irrigation, drinking water, hydropower and flood control objectives. Multi-reservoir studies show that coordinated operation across a basin can deliver more water to users than independent operation of each dam. Research on inter-basin transfer schemes discusses the potential of link canals to move surplus water to deficit regions, while also highlighting the engineering, environmental and social constraints involved. A recurring observation in this literature is that the quality of allocation decisions is limited by the timeliness and accuracy of the input data, particularly inflow forecasts and demand estimates, which are often based on outdated assumptions.

Recent work on water information systems describes platforms that gather hydrological and meteorological data into shared databases and web portals. National and state water resource information systems publish reservoir levels, river flows and groundwater data, and some research prototypes add dashboards and alerting. Studies of these platforms note that they are mostly descriptive: they show the current state but do not forecast, do not integrate flood and drought information in one view and do not recommend actions. Data engineering challenges such as inconsistent station identifiers, delayed bulletins and mixed units are repeatedly mentioned as barriers to automated analysis. Aqua Rescue builds on these observations by combining an automated ingestion and cleaning pipeline with forecasting, drought scoring and optimisation so that the output is a practical redistribution plan.

III. SYSTEM ANALYSIS

The system is analysed as an end-to-end pipeline that begins with hydrological and meteorological observations and ends with a recommended redistribution plan. Rainfall arrives daily from rain gauge stations and as gridded products, reservoir storage and release figures are extracted from department bulletins published as tables or PDF files, and river discharge is read from gauging station feeds. Groundwater

levels come from observation wells at monthly or seasonal intervals, while flood extent maps are produced from satellite radar images after major rain events. Static layers describe basin and district boundaries, reservoir capacities, canal and pipeline routes with their carrying capacities, and district population and cropping patterns. The analysis showed that the main difficulties are inconsistent identifiers, mixed units such as thousand million cubic feet and million cubic metres, delayed or missing bulletins, and different time resolutions. A common spatial key and a daily time step were therefore chosen as the basis for integration, with forecasting, demand estimation, drought scoring and optimisation built on the curated daily tables.

Existing System

In the existing practice, each agency manages its own data and decisions are made largely through manual coordination. Reservoir operators follow rule curves that specify target levels for each month and release water when storage exceeds the flood cushion, with limited foresight about inflows expected in the coming days. Rainfall and river flow information is shared through daily reports and phone calls, and groundwater data is published months later. Drought declarations are based on rainfall deficit and crop surveys carried out at district level, and relief usually takes the form of tanker supply and emergency borewells. Transfers between reservoirs or basins are planned on a case-by-case basis using spreadsheets prepared by engineers. There is no single platform in which surplus and deficit can be seen together on a map, no forecast of how much surplus will arrive, and no systematic method to decide which transfers are feasible within canal capacities and environmental flow limits.

Disadvantages of Existing System

- Rainfall, reservoir, river and groundwater data are held by separate agencies.
- Bulletins use different formats, units and station names, making integration slow.
- Reservoir releases rely on rule curves with little forecast of future inflows.
- Large volumes of flood water are released downstream without being captured.
- Drought is recognised late, usually after crops and drinking supplies are affected.
- Flood and drought conditions are never viewed together on one map.
- Transfer decisions are prepared manually in spreadsheets under time pressure.
- Canal capacities and environmental flows are not checked systematically.
- District water demand is estimated from outdated norms.
- Missing or delayed bulletins go unnoticed and distort daily reports.

Proposed System

The proposed system, Aqua Rescue, is an integrated data pipeline and decision support platform that links rainfall, floodwater, reservoir and groundwater data to redistribution planning. Automated connectors collect gauge readings, gridded rainfall, reservoir bulletins, discharge feeds, well levels and satellite radar scenes, and an extraction step parses tabular and PDF bulletins into structured records. All data is validated, converted to common units and keyed to basins, reservoirs and districts before loading into a curated warehouse with daily history. On this foundation an LSTM network forecasts reservoir inflows for fourteen days, a gradient boosting

model estimates district water demand, and a composite drought risk score combines the Standardised Precipitation Index, groundwater decline and storage deficit. A linear programming optimiser uses forecast surplus, demand deficits, canal capacities, flood cushion requirements and minimum environmental flows to recommend feasible transfers. A basin dashboard presents surplus and deficit maps, storage trends, forecasts, drought alerts and the recommended plan, and it records the decisions finally taken by engineers.

Advantages of Proposed System

- Brings rainfall, reservoir, river, groundwater and flood data into one warehouse.
- Parses department bulletins automatically and standardises units and identifiers.
- Forecasts reservoir inflows up to fourteen days ahead for better release planning.
- Identifies surplus flood water that can be stored or diverted instead of lost.
- Detects drought risk early using a composite of rainfall, groundwater and storage.
- Estimates district water demand from population, crops and temperature.
- Recommends transfers that respect canal capacities and environmental flows.
- Shows flood and drought conditions together on a single basin map.
- Keeps a record of plans and decisions for later review and learning.
- Flags missing or suspicious data before it affects forecasts and plans.

IV. METHODOLOGY

The methodology starts with data acquisition, in which scheduled Apache Airflow tasks download rainfall grids, gauge readings, discharge feeds and well levels, and a parser built with table extraction libraries converts reservoir bulletins into rows of storage, inflow and release. Satellite radar scenes over the basins are classified with a threshold and random forest method to produce flood extent polygons, which are intersected with elevation data to estimate floodwater volume. During preprocessing, units are converted to million cubic metres, station names are matched to a master list, gaps shorter than three days are interpolated and values that violate physical limits are flagged. Daily features such as cumulative rainfall over three, seven and fifteen days, upstream discharge and storage change are computed. The LSTM inflow model and the demand model are trained on ten years of history with time-ordered validation. The drought score is computed weekly, and the optimiser, written with the PuLP library, runs each day to produce a transfer plan that is published to the dashboard.

SYSTEM ARCHITECTURE

The architecture of Aqua Rescue is organised as a set of stages connected around a central hydrological data warehouse. The Data Sources stage includes rain gauges and gridded rainfall, reservoir bulletins, river gauging stations, groundwater observation wells, satellite radar imagery and static layers for basins, districts and canals. The Ingestion and Extraction stage uses Airflow schedules, bulletin parsers and a radar flood mapping job. The Validation and Harmonisation stage performs unit conversion, identifier matching, gap filling and quality flagging. The Hydrological Data

Warehouse, built on PostgreSQL with PostGIS, holds daily curated tables and spatial layers. The Analytics and Models stage contains the LSTM inflow forecaster, the demand estimator and the drought risk scorer. The Optimisation stage runs the linear programming allocator with its capacity and environmental flow constraints. The Decision Support stage provides the basin dashboard, drought and surplus alerts for district officials and a log of approved transfers that feeds back into later analysis.

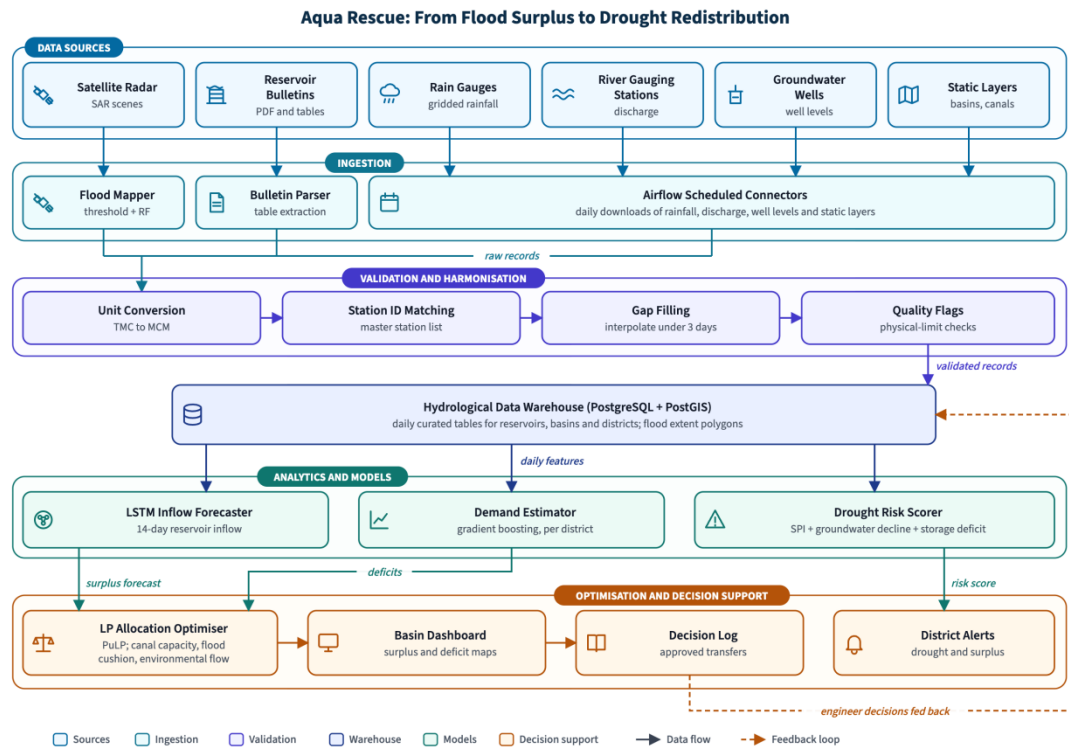


Fig. 1 System Architecture

V. RESULT AND OUTPUT

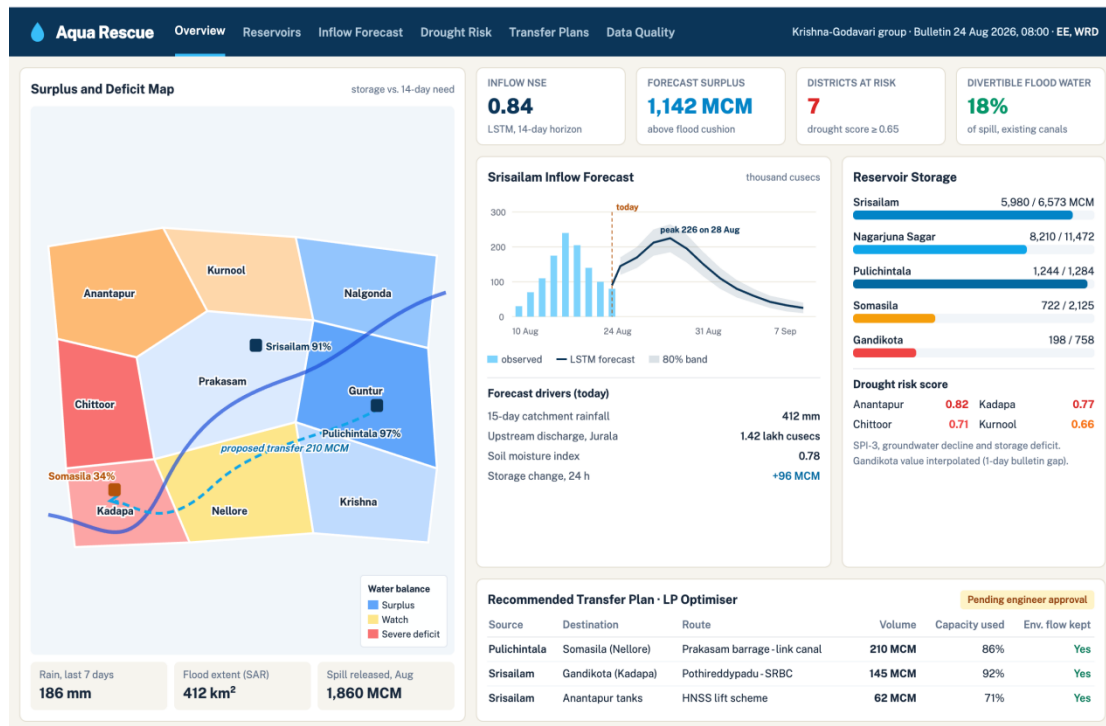


Fig. 2 Result and Output Screen

VI. CONCLUSION

This paper presented Aqua Rescue, a data pipeline and decision support system that connects rainfall, floodwater, reservoir, river and groundwater data to plan the redistribution of surplus water towards drought-prone areas. By automating the collection and harmonisation of data from several agencies, the system replaces slow manual compilation with a curated daily view in which surplus and deficit can be examined side by side.

The forecasting and optimisation components show that this unified data can support better decisions. The LSTM inflow model achieved a Nash-Sutcliffe efficiency of 0.84 for fourteen-day forecasts, which gives operators time to plan releases and transfers before flood peaks arrive. In the retrospective study, the linear programming allocator indicated that about 18 percent of the water released during flood spells could have been stored or diverted to deficit districts through existing canals and pumping schemes without violating capacity or environmental flow constraints.

The work also highlights the importance of data engineering in water management. Parsing bulletins, matching station names, converting units and flagging suspicious values consumed a large part of the development effort, but these steps were what made reliable forecasts and plans possible. The composite drought score, built from rainfall, groundwater and storage indicators, identified districts at risk several weeks earlier than declarations based on rainfall deficit alone.

Future work will add real-time telemetry from reservoir gates and canal flow meters, estimate demand at the level of villages and irrigation commands, and include water quality parameters so that transfers do not move polluted water into drinking supply sources. Extending the optimiser to multiple objectives, including energy used for

pumping and fair sharing among districts, and conducting a pilot with a state water resources department are the next planned steps.

REFERENCES

- [1] Kumar, R. D., Prudhviraaj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
- [2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.
- [3] Sv satyakrishna, shirisha rang, bhargavi nalacheruve. (2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
- [4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
- [5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
- [6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.
- [7] R. D. Kumar, V. N. S. Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
- [8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
- [9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
- [10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.
- [11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.