RESEARCH ARTICLE OPEN ACCESS

An Uncertainty-Aware Temporal Fusion Transformer Framework for Reservoir Water-Level Forecasting and Dam Alert Generation

Dr. Nisha. C.M
Assistant Professor
Department of Computer Science
Govt. Victoria College, Palakkad,
Kerala, India.

ABSTRACT

Effective water resource management, flood risk reduction, and dam safety all depend on accurate reservoir water-level forecasts. Using the Temporal Fusion Transformer (TFT) for multi-horizon water-level forecasting, this paper suggests an uncertainty-aware dam alarm production methodology. The model incorporates meteorological and hydrological factors, such as rainfall, reservoir storage, inflow, and discharge, to forecast reservoir water levels of the Malampuzha dam for the next seven days using a 30-day historical observation window. Water-level records that were physically unreasonable were found and corrected through data-quality validation. On a held-out validation period that included the final 365 records and 359 prediction windows, the median forecast's mean absolute error (MAE) was 0.29 m, as opposed to 2.99 m prior to data cleaning. Using a cautious 90th-percentile (P90) decision criteria and predetermined water-level thresholds, probabilistic forecasts were transformed into blue, orange, and red warning phases. Although there were false alarms, the alert-generation framework had an accuracy of 85.5% and no missed alerts. The results show how probabilistic forecasting and threshold-based alert production can be combined to enable early-warning decisions and dam monitoring.

Keywords :— Temporal Fusion Transformer, water-level forecasting, uncertainty quantification, explainable AI, early warning system.

I. INTRODUCTION

Forecasting reservoir water levels is crucial for effective water resource planning, flood risk management, and dam safety. Dam officials can anticipate possible hazards, promptly issue warnings, and make well-informed judgments about reservoir operation and water discharge when they have accurate projections of future reservoir levels. Complex nonlinear interactions between hydrological variables, weather, rainfall patterns, and reservoir operational conditions may be difficult for conventional forecasting techniques to capture. By identifying temporal correlations and interactions between various input variables, recent developments in deep learning have made it possible to enhance hydrological forecasting.

Integrating artificial intelligence into dam and water resource management enhances predictive forecasting, flood risk analysis, and infrastructure resilience against climate challenges[1]. However, widespread adoption requires overcoming data quality limitations, model opacity, ethical concerns, and the need for interdisciplinary collaboration and transparent governance. The study done by [2] developed a seasonality-aware Transformer model for daily inflow forecasting at Oued Makhazine Dam, integrating hydrometeorological variables, seasonal features, and historical inflow patterns. Compared with GRU, LSTM, RNN, ANN, XGBoost, Random Forest, and Linear Regression, the Transformer achieved the best performance, with test R^2 and NSE values, demonstrating strong capabilities in capturing seasonal variations and high-flow events. Future research should focus on testing the model across different basins, incorporating uncertainty estimation and explainable AI, and integrating physics-informed constraints to improve reliability in reservoir management.

This work uses the last 30 days of observations to project reservoir water levels over a seven-day prediction horizon using the Temporal Fusion Transformer (TFT), a deep learning architecture intended for interpretable multi-horizon time-series forecasting. Temperature, humidity, rainfall, reservoir storage, inflow, and outflow via the left-bank canal, right-bank canal, and spillway are among the pertinent meteorological and hydrological variables that are included in the model. To generate both median estimations and upper-

bound projections for risk-sensitive decision-making, quantile-based forecasting is used to describe predictive uncertainty. Using established reservoir-level criteria, the anticipated water levels are then linked to predetermined blue, orange, and red alert phases.

This study examines the interpretability of the TFT model using temporal attention weights and variable-selection importance scores in addition to assessing predicting accuracy. These methods aid in determining which input variables and past observations have the most influence on the model's predictions. The efficacy of a conservative alert-generation technique based on the 90th-percentile forecast, which prioritizes avoiding missed warnings while accepting the chance of false alarms, is also examined in this study. The findings highlight the significance of interpretable forecasts, uncertainty-aware forecasting, and data-quality validation in creating trustworthy decision-support systems for dam safety. The results, however, are limited by single-reservoir validation, the lack of future rainfall projections, and the requirement for additional assessment under other hydrological conditions.

II. LITERATURE REVIEW

Regionally-trained LSTM networks can accurately model rainfall-runoff dynamics across diverse catchments using solely meteorological inputs[3]. A triple-confirmation explainability framework revealed that model performance is significantly influenced by catchment characteristics and hyperparameters like input sequence length. These insights support the development of more interpretable and reliable AI models for operational flood forecasting in complex regions.

The study[4] proposes a Combined Optimizer (CO) and Dual-AI framework—integrating deep learning with eXplainable AI (XAI)—to enhance dam inflow prediction and water resource management. Compared to traditional optimizers, the Dual-AI model significantly reduces prediction error (RMSE) and overestimation tendencies while improving accuracy (R^2). Although current limitations involve complex parameter calibration, these advancements provide a reliable foundation for flood forecasting and drought mitigation.

The study [5] combined an LSTM model with eXplainable AI methods (SHAP and EG) to evaluate the impact of the Three Gorges Dam (TGD) on Yangtze River flood dynamics. The results showed that primary flood drivers shifted from upper-basin runoff before the dam's operation to local precipitation afterward, highlighting both the dam's regulatory effectiveness and the value of XAI in uncovering complex hydrological mechanisms.

Among multiple machine learning models evaluated for flood susceptibility prediction by [6], XGBoost demonstrated superior accuracy with the lowest error rate and highest predictive power. Using SHAP explainability analyses, the study revealed that distance to streams, topographic wetness index (TWI), and elevation are the primary drivers of flood risk, while factors like soil type play a minimal role.

Evaluating runoff forecasting methods in Nepal's steep and data-limited catchments, the study done by [7] compared the Temporal Fusion Transformer (TFT) against Random Forest and Long Short-Term Memory models using 40 years of hydrometeorological data. The TFT achieved superior accuracy (RMSE = 22.5, R^2 = 0.88), with its attention mechanism successfully pinpointing precipitation and prior runoff as the principal drivers of flood dynamics.

Reviewing 83 publications on explainable AI (XAI) in hydrological forecasting, the paper [8] synthesizes current methods, model architectures, and key research gaps while emphasizing post-hoc approaches like SHAP and Grad-CAM. The review highlights that XAI effectively aligns model outputs with physical hydrological reasoning and diagnoses failure modes, provided it is embedded directly into the modeling workflow rather than applied as a late visualization step.

The study [9] evaluates five machine learning models—Bootstrap Tree, Decision Tree, Histogram Gradient Boosting, Extreme Gradient Boosting, and Categorical Boosting—for long-term river discharge prediction in the South Platte River, USA, incorporating explainable AI (XAI) to improve model transparency and reliability. The results show that Categorical Boosting, XGBoost, and Histogram Gradient Boosting achieved the best predictive performance, with R^2 values of 0.939, 0.852, and 0.817, respectively. Long-term projections indicate a continuous decline in streamflow, potentially leading to river drying within two decades, highlighting the urgent need for sustainable water management and climate action.

The study [10] proposed a hybrid CNN-LSTM-Attention model for flood and danger-level prediction in the Dinajpur district of Bangladesh, outperforming CNN, LSTM, SVM, MNB, KNN, and Logistic Regression models. Validation using the Kerala Rainfall dataset also demonstrated better predictive performance, highlighting the model's potential for early flood warnings, disaster risk reduction, and improved flood management.

This review of [11] over 230 studies highlights the role of Explainable Artificial Intelligence (XAI) methods, such as SHAP, LIME, PDP, and ALE, in improving model interpretability and understanding hydrological processes, with SHAP being the most widely used technique. However, research remains limited in areas such as evapotranspiration, groundwater, and forecasting. Future studies should focus on integrating physical knowledge with XAI, improving computational scalability, and applying explainable models to practical decision-making to promote transparent, reliable, and sustainable water resource management.

Using the Temporal Fusion Transformer (TFT), this study provides an integrated framework for seven-day reservoir water-level forecasting and uncertainty-aware dam alarm generating. This study uses predefined thresholds and a conservative P90-based decision rule to translate probabilistic water-level forecasts into actionable blue, orange, and red alert stages, in contrast to earlier research that mostly focused on rainfall–runoff modelling, dam inflow prediction, river discharge forecasting, or flood susceptibility assessment [3–10]. Analysis of the key hydrological variables and past observations is made possible by the model's integration of multi-horizon forecasting with inherent interpretability through variable-selection weights and temporal attention.

Additionally, the median forecast MAE was lowered from 2.99 m to 0.29 m by data-quality validation, highlighting the significance of fixing physically implausible records. False alarms are still a problem, although the alert system achieved 85.5% accuracy with no missed alerts in the assessed validation intervals. The integration of forecasting accuracy, uncertainty estimate, interpretability, and safety-oriented alert production into a single framework for reservoir management and early warning is hence the main benefit.

III. METHODOLOGY

This methodology utilizes a Temporal Fusion Transformer (TFT) to forecast multi-horizon reservoir water levels and generate uncertainty-aware dam alerts. The dataset integrates historical meteorological and hydrological observations of ten years encompassing temperature, dew point, humidity, wind speed, atmospheric pressure, water level, reservoir storage, daily and accumulated rainfall, and various discharge metrics (canal, spillway, total, and inflow) of the Malampuzha dam situated in Kerala, a southernmost state of India. Data preprocessing involved cleaning and correcting invalid entries—specifically addressing physically impossible water-level values to prevent model degradation—through removal or appropriate interpolation before preparing the features for training. A sliding-window technique was employed, utilizing the previous 30 days of observations to forecast reservoir water levels for the following seven days. The dataset was split chronologically, reserving the final 365 records as a validation set to assess the model's predictive performance on an unseen subsequent period.

The TFT architecture was designed to identify temporal dependencies and correlations among input variables while providing probabilistic forecasts. With approximately 107,000 trainable parameters, the model generated several forecast quantiles, including the median (P50) and upper bounds (P90 and P98). The median forecast served to evaluate central prediction accuracy, while the upper quantiles captured higher plausible water-level scenarios to facilitate risk-aware alert generation.

To generate dam alerts, forecasted water levels were compared against predefined thresholds: 113.0 m (Blue), 114.0 m (Orange), and 114.46 m (Red). A conservative decision rule was applied, triggering an alert if the P90 forecast reached or exceeded any threshold at any point during the seven-day forecast horizon. Model performance was evaluated by comparing predicted alert levels against those derived from observed water levels. Key performance metrics included alert accuracy, missed alerts, and false alarms, with a strong emphasis on minimizing missed warnings due to their implications for dam safety. Finally, to evaluate the model's utility as a decision-support tool for reservoir monitoring and early warning, the study analyzed forecast uncertainty through quantiles, alongside variable-importance and attention patterns. The methodology also acknowledges limitations stemming from data-quality concerns, the lack of future rainfall forecasts, and the use of only a single reservoir and a single hold-out period for validation.

IV RESULTS AND DISCUSSION

A. Forecast accuracy

The Temporal Fusion Transformer (TFT, 107 K trainable parameters) was trained on a 30-day history to forecast reservoir water level for the next 7 days. On the held-out validation period (the last 365 records, 359 forecast windows), the median (P50) forecast achieved a mean absolute error (MAE) of 0.29 m. For context, the blue, orange and red alert levels (113.0, 114.0 and 114.46 m) are separated by 0.46 to 1.0 m, so the error is smaller than the gap between adjacent alert stages.

Data quality had a large effect. Before cleaning, physically impossible water-level records (up to 1402 m) were present, and the MAE was 2.99 m. After removing and interpolating these values, the MAE fell to 0.29 m. This shows that data validation is a necessary step in operational dam forecasting, and it should be reported as part of the method.

B. Alert prediction

Forecast quantiles were converted into alert stages using the official thresholds. An alert was raised when the 90th-percentile (P90) forecast reached a threshold at any point in the 7-day horizon, an early-warning rule that favours caution. Against alert stages derived from observed levels:

All misclassifications were on the safe side. In flood-control applications a missed warning is far more costly than a false alarm, so this error pattern is favourable. The false alarms arise from the conservative P90 rule, and the threshold quantile gives operators a tunable trade-off between sensitivity and false-alarm rate.

TABLE 1
ALERT VALUES

Metric	Value
Alert-Accuracy	85.5%
Missed alerts (under-warning)	0
False alarms (over-warning)	52 (about 14.5% of windows)

C. Interpretability

The TFT provides two built-in explanation mechanisms: variable-selection weights, which rank the input variables within the encoder, decoder and static groups, and attention weights, which show which past time steps influenced a forecast. Figures 1 to 5 show these for the validation windows, together with an example forecast. All importance values below are read from the figures and are approximate.

D. Temporal attention

Averaged over the validation windows, attention rises steadily from about 0.023 at lags of 25 to 28 days to about 0.046 at the most recent day, a roughly twofold difference (Figure 1). The model therefore relies mainly on the latest observations and gives progressively less weight to older ones, while still drawing on the whole 30-day window, since attention never approaches zero. This is physically plausible: the level seven days ahead is largely set by the current reservoir state, with recent inflow and release history adjusting it. The small rise at the oldest lag is minor and should not be over-interpreted.

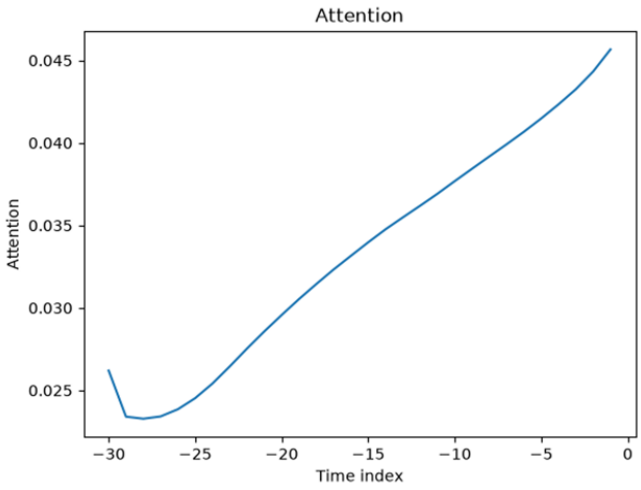


Fig 1. Mean attention weight by lag (days before the forecast origin).

E. Encoder variable importance

Reservoir storage dominates the encoder, with about 43% of the total weight, followed by humidity at about 10%. No other variable exceeds about 5% (Figure 2). Inflow (about 4.5%), water level (about 4.3%) and the canal and spillway discharges (about 3 to 4% each) form the next tier. Grouped, the reservoir-state variables (storage and level) account for roughly 47% of the weight, the weather variables (humidity, dew point, pressure, wind, temperature) for about 19%, the release variables (LBC, RBC, spillway and total discharge) for about 13%, and the rainfall variables for about 12%.

Among the rainfall features, the longer accumulation windows (15-day, about 3.3%; 7-day, about 2.7%) received more

weight than the 3-day sum (about 0.9%). This is consistent with a catchment that responds to accumulated rainfall with a lag, although the differences are small. Spillway discharge carries about 3.4%, like the canal discharges, which suggests the model uses release history as part of the reservoir's water balance.

Storage and water level carry almost the same physical information, so their importance is split between them, with storage favoured. The combined weight should be read as "current reservoir fullness matters most", not as storage causing an alert. An ablation that removes one of the two would show whether this redundancy affects accuracy.

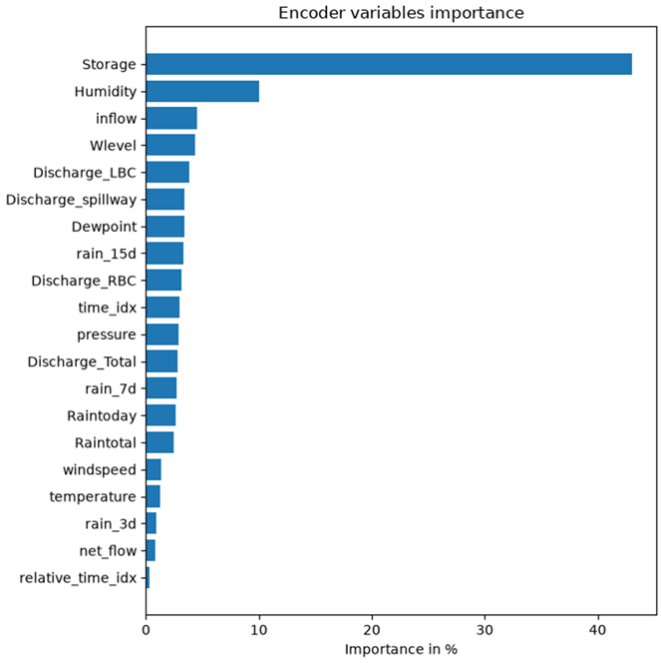


Fig 2. Encoder variable importance (aggregated over validation windows).

Humidity ranking second in the aggregate probably reflects the monsoon season, when high humidity coincides with heavy rain and rising inflow. Because the dataset has no timestamps, the model could not receive calendar information directly, so humidity is likely acting as a seasonal proxy. For the single highest-risk forecast window, humidity (0.414) outranked storage (0.203), so the dominant drivers differ between typical and extreme situations, and the aggregate is the better summary of overall model behaviour. Variable-importance scores describe what the trained model relied on, not physical causal effects.

F. Decoder and static variable importance

The only inputs known in advance were the absolute time index and the position within the forecast horizon, so the decoder panel has limited physical content (Figure 3). The time index accounts for about 81% and the relative position for about 19%. The dominance of the absolute time index shows that the model uses the position within the record as a trend or seasonal cue. This cannot be extrapolated to future dates, so it should be replaced by calendar variables (month, day of year) once the dates of the records are known.

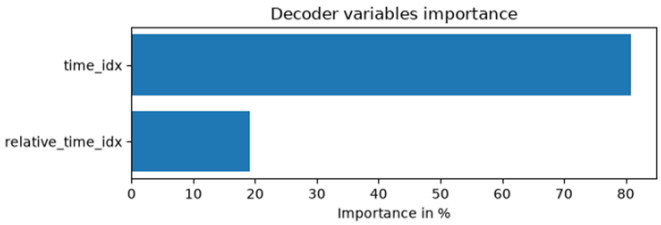


Fig 3. Decoder variable importance.

Among the static variables (Figure 4), the encoder length (about 41%) and the group identifier (about 40%) are constant or nearly constant in a single-reservoir dataset, so their weights carry no interpretable meaning and should not be discussed as findings. The target's normalisation statistics, Wlevel_center (about 16%) and Wlevel_scale (about 4%), indicate that the model uses the typical level of the series to calibrate its output.

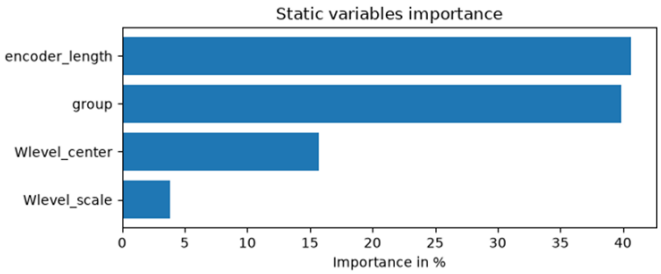


Fig 4. Static variable importance.

G. Example forecast

In the example window (Figure 5), the observed level rose from about 113.5 m to 114.5 m over the preceding 30 days. The attention trace (grey) rises toward the present, with peaks in the last 12 days or so, in agreement with the aggregate profile in Figure 1. The median forecast starts near 114.9 m and drifts down to about 114.85 m, with the outer prediction band spanning roughly 114.6 to 115.2 m. The observed level

continued rising slowly to about 114.6 m, close to the lower edge of the outer band. First, the forecast over-predicts the level by about 0.3 m and misses the continued upward trend, so it is conservative here. Second, both the forecast and the observed levels exceed the red threshold of 114.46 m, so the alert category is correct despite the level error. This is consistent with the reported absence of missed alerts and with the false alarms produced by the upper-quantile rule.

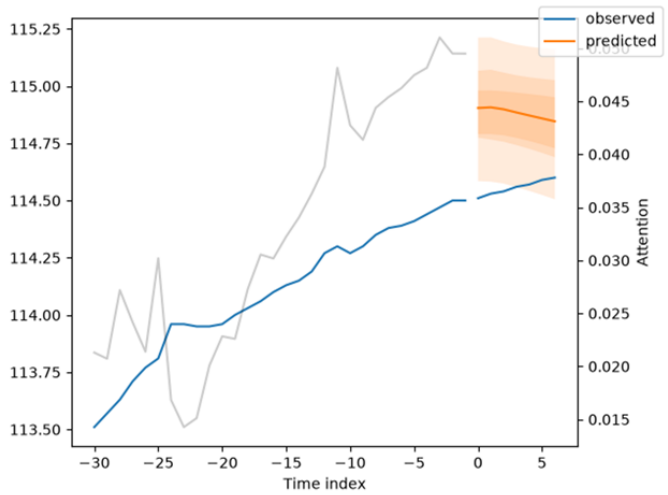


Fig 5. Example 7-day forecast: observed (blue), median forecast with quantile bands (orange), attention (grey, right axis).

H. Operational forecast

For the most recent record, the model projected a gradual decline in median level from 112.2 m to 111.9 m over seven days. Even the P98 forecast stayed below the 113 m blue threshold, so no alert was issued. The widening gap between P50 and P98 shows the model's uncertainty growing with lead time, which is useful information for decisionmakers.

I. Limitations

- Validation covers a single dam and a single hold-out year, so generalization to other seasons, years or reservoirs is untested.
- No future rainfall forecasts were used. Adding numerical weather prediction or IMD rainfall forecasts as known-future inputs would likely improve multi-day skill.
- Data errors were corrected by interpolation, and some may remain in other variables.
- Stage thresholds are crisp, so small errors near a boundary change the alert category. Reporting results near the boundaries separately would be informative.

V Conclusion

A Temporal Fusion Transformer (TFT)-based framework for seven-day reservoir water-level forecasting and uncertainty-aware dam alarm production was provided in this paper. After data cleaning, the model produced a median forecast mean absolute error (MAE) of 0.29 m using a 30-day historical observation window, underscoring the significance of data quality in hydrological forecasting. Despite 52 false alarms, the suggested P90-based alert method produced 85.5% alert accuracy with no missed alerts in the assessed validation windows. Reservoir storage was shown to be the most significant input by variable-selection weights, and recent observations were highlighted by temporal attention. These results show how dam safety and reservoir management can be supported by combining probabilistic forecasting, model interpretability, and threshold-based warning production. However, the results' generalizability is limited by validation with a single reservoir and hold-out period. To increase dependability and operational application, future research should include projected rainfall data, assess alternate alarm quantiles, and validate the system across several years and reservoirs.

ACKNOWLEDGMENT

The author would like to thank the Irrigation department of the Malampuzha division for providing access to data.

REFERENCES

[1] L. Abdulameer, N. M. L. Maimuri, A. H. Nama, F. L. Rashid, H. I. Mohammed, and A. N. G. Al-Dujaili, “Review of Artificial Intelligence Applications in Dams and Water Resources: Current Trends and Future Directions,” *J. Adv. Res. Fluid Mech. Therm. Sc.*, vol. 128, no. 2, pp. 205–225, Mar. 2025, doi: 10.37934/arfmts.128.2.205225.

[2] C. Achahboun, M. Chikhaoui, M. Naimi, M. Bellafkih, and O. Bassairate, “Comparative Assessment of Transformer and Conventional Deep Learning Models for Dam Inflow Prediction,” *Water Resour Manage*, vol.

40, no. 9, p. 429, Jul. 2026, doi: 10.1007/s11269-026-04791-0.

[3] F. Hosseini, C. Prieto, and C. Álvarez, "An explainable AI approach for interpreting regionally optimized deep neural networks in hydrological prediction," *Journal of Hydrology*, vol. 661, p. 133689, Nov. 2025, doi: 10.1016/j.jhydrol.2025.133689.

[4] Y. M. Ryu and E. H. Lee, "Development of dam inflow prediction technique based on explainable artificial intelligence (XAI) and combined optimizer for efficient use of water resources," *Environmental Modelling & Software*, vol. 187, p. 106380, Apr. 2025, doi: 10.1016/j.envsoft.2025.106380.

[5] X. Wei, G. Wang, P. F. Grosser, and B. Schmalz, "Using explainable artificial intelligence (XAI) methods to understand the nonlinear relationship between the Three Gorges Dam and downstream flood," *Journal of Hydrology: Regional Studies*, vol. 53, p. 101776, Jun. 2024, doi: 10.1016/j.ejrh.2024.101776.

[6] B. Choubin, A. Jaafari, J. Henareh, O. Karimi, and F. Sajedi Hosseini, "Explainable artificial intelligence (XAI) for interpreting predictive models and key variables in flood susceptibility," *Results in Engineering*, vol. 27, p. 105976, Sep. 2025, doi: 10.1016/j.rineng.2025.105976.

[7] M. G. K. Mishra, "Attention-Based Deep Learning for Runo! Forecasting: Evaluating the Temporal Fusion Transformer Against Traditional Machine Learning Models".

[8] A. M. Bramm, P. V. Matrenin, and A. I. Khalyasmaa, "A Review of XAI Methods Applications in Forecasting Runoff and Water Level Hydrological Tasks," *Mathematics*, vol. 13, no. 17, p. 2830, Sep. 2025, doi: 10.3390/math13172830.

[9] M. Poursaeid, "Long-term Study of Water Resource by Machine Learning: A New Application of Explainable Artificial Intelligence Model Towards Sustainable Development Goals," *Water Resour Manage*, vol. 39, no. 15, pp. 7995–8016, Dec. 2025, doi: 10.1007/s11269-025-04326-z.

[10] M. Hasan, Md. F. Rabbi, M. A. Hamja, K. Fatema, and M. M. Hassan, "Enhancing flood forecasting performance using effective and transparent explainable hybrid deep learning model," *Earth Sci Inform*, vol. 18, no. 2, p. 417, Jun. 2025, doi: 10.1007/s12145-025-01930-w.

[11] M. Zounemat-Kermani and M. Kheimi, "Explainable Artificial Intelligence in Hydrology: A Review," *Water Resour Manage*, vol. 40, no. 3, p. 106, Feb. 2026, doi: 10.1007/s11269-025-04435-9.

.