HydroFinNet: An Explainable Multi-Modal Deep Learning Framework for Sustainable Water Resource Allocation and Economic Risk Assessment

Sachin Kumar[1], Sanjana[2]

[1]Computer Science Engineering, GLA University, Mathura, India
Email: singhsachin82736@gmail.com
[2]G.L. BAJAJ Groups of Institutions Mathura Campus , AKTU University, Lucknow, India
Email: sjc40603@gmail.com

Abstract:
Sustainable water resource management requires intelligent decision-support systems capable of addressing hydrological uncertainty, competing sectoral demands, and associated economic risks. Existing approaches commonly rely on isolated forecasting or optimization models and provide limited interpretability for critical allocation decisions. This paper proposes HydroFinNet, an explainable multi-modal deep learning framework for jointly optimizing sustainable water allocation and assessing economic risk. The proposed architecture integrates remote-sensing and reservoir observations, meteorological and hydrological time series, demographic and sectoral demand indicators, and financial market variables through modality-specific encoders and a cross-modal multi-head attention mechanism. A dual-task architecture simultaneously generates sector-wise allocation recommendations and an Economic Risk Index while incorporating ecological-flow constraints into the optimization objective. To improve transparency, Deep-SHAP attribution and attention-based analysis identify influential features and modalities underlying model decisions. Experimental evaluation demonstrates improved water-demand prediction, allocation efficiency, and economic-risk estimation compared with conventional statistical and uni-modal deep learning baselines. The framework provides an interpretable and scalable foundation for data-driven, risk-aware, and sustainable water-resource governance.

Keywords: Sustainable water management, multimodal deep learning, explainable artificial intelligence, water resource allocation, economic risk assessment, cross-modal attention, SHAP, hydrological forecasting.

I. INTRODUCTION

A. Background and Motivation

Freshwater availability is under mounting pressure worldwide. As of 2022, roughly half of the global population experienced severe water scarcity for at least part of the year, while nearly a quarter of the world faced extremely high water stress, drawing on more than 80% of annually renewable supplies [1]. Climate variability, rapid urbanization, and continued population growth compound this pressure by shifting the timing, location, and volume of water availability while simultaneously increasing demand across agricultural, domestic, and industrial sectors. Because water underpins food production, energy generation, and industrial output, inefficient allocation does not remain a purely hydrological problem; it propagates directly into economic losses, price volatility, and heightened financial risk for water-dependent sectors [1]. This dual

hydrological–economic exposure motivates the development of decision-support systems capable of jointly reasoning about physical water availability and the downstream economic consequences of allocation choices.

B. Challenges in Sustainable Water Resource Allocation

Sustainable allocation is inherently difficult because water resources vary sharply across space and time, while demand from agriculture, domestic users, industry, and ecological systems is rarely synchronized with supply [2]. Optimization-based allocation models—linear programming, genetic algorithms, particle-swarm and multi-objective formulations—have been used extensively to balance these competing claims [3], [4], [5]. However, such approaches typically require well-specified objective functions and struggle when hydrological and socioeconomic inputs are uncertain, incomplete, or nonstationary. Newsvendor-style and prospect-theory formulations have attempted to embed economic uncertainty directly into allocation decisions [6], [7], yet these remain limited to relatively narrow problem settings and rarely integrate the full spectrum of climatic, hydrological, and financial data streams that shape real-world allocation outcomes.

C. Role of Artificial Intelligence in Water Resource Management

Machine learning and deep learning have become central tools for water-demand forecasting, hydrological prediction, and reservoir operation, owing to their ability to model nonlinear relationships in large, heterogeneous datasets [2], [8]. Reinforcement learning has further extended this capability to adaptive, feedback-driven allocation policies that can respond to dynamically changing supply-demand conditions [9]. Despite these advances, conventional data-driven approaches largely treat water-demand prediction and economic outcome modeling as separate tasks, optimize for point-forecast accuracy rather than allocation robustness, and offer limited mechanisms for incorporating multiple heterogeneous data modalities into a single coherent decision pipeline. This fragmentation constrains their usefulness for integrated, real-time water-economic governance.

D. Multi-Modal Learning for Integrated Water–Economic Assessment

Multi-modal deep learning offers a principled way to unify hydrological, climatic, socioeconomic, and financial information within a single representation space. Baltrušaitis et al. [10] formalize the core challenges of multi-modal learning—representation, translation, alignment, fusion, and co-learning—that determine how effectively heterogeneous data sources can be combined. Applying this paradigm to water-economic assessment allows a model to jointly learn from remote-sensing and reservoir data, meteorological time series, demographic and sectoral demand indicators, and market or financial variables, producing allocation recommendations that are sensitive to both physical constraints and downstream economic risk. This joint modeling capacity is the central motivation behind the proposed HydroFinNet framework, which treats water allocation and economic risk assessment as two facets of a single learning problem rather than as sequential, loosely coupled stages.

E. Explainability and Trustworthy Decision-Making

Deep learning models, particularly when fused across modalities, are frequently criticized as opaque "black boxes," which limits their acceptance among water authorities, regulators, and affected communities [11]. Explainable AI (XAI) techniques such as LIME [12] and SHAP [13] address this gap by attributing model predictions to specific input features, enabling stakeholders to identify which hydrological, climatic, or economic drivers most influence a given allocation or risk estimate. Embedding such interpretability mechanisms directly into the modeling pipeline—rather than applying them only as a post-hoc audit—is essential for water governance, where decisions carry legal, social, and economic consequences and must be defensible to diverse stakeholders.

F. Research Gap

Existing water allocation models generally fall into two disconnected categories: hydrology-focused machine learning systems that forecast supply and demand with limited economic reasoning [14], and optimization or economic models that treat hydrological inputs as static or exogenously given [7]. Few frameworks integrate multi-modal deep learning to jointly model water allocation and economic risk, and fewer still incorporate explainability as a core design requirement rather than an afterthought. This gap motivates HydroFinNet: a unified, explainable, multi-modal deep learning framework designed to jointly optimize sustainable water allocation and quantify the associated economic risk, while providing transparent, feature-level justifications for its recommendations.

II. REVIEW OF LITERATURE

[1] explained that Sustainable water resource management has become a global priority due to urbanization, climate variability, and population growth, which increasingly stress freshwater ecosystems. Existing hydrological prediction systems rely on isolated data sources and opaque machine learning models, limiting decision-making and policy adoption. This paper proposes a Human–AI Collaborative Framework integrating satellite imagery, IoT water level sensors, meteorological variables, and historical hydrological records within an explainable deep learning architecture. A hybrid cross-attention fusion model combines spatial, temporal, and environmental features, while SHAP- and Grad-CAM-based modules provide interpretable predictions. A human-in-the-loop reinforcement mechanism enables expert feedback for continuous improvement. Experimental results demonstrate improved forecasting accuracy, reduced false flood alarms, enhanced user trust, and support sustainable water governance aligned with SDG 6 and SDG 13**.**

[2] highlighted that water information systems (ISs) support sustainable water management but are constrained by fragmented, heterogeneous, and rapidly evolving data. Generative artificial intelligence (GenAI), particularly large language models (LLMs), enables natural language question-answering, while retrieval-augmented generation (RAG) grounds responses in authoritative, organisation-specific data for reliable decision-making. This study proposes a three-step framework integrating high-value water-sector applications, multi-source datasets under provenance and privacy constraints, and LLM–RAG-based interaction across multimodal data, including text, time series, remote sensing, GIS layers, and knowledge graphs. Unlike conventional RAG pipelines, the framework explicitly connects retrieval and generation to domain-specific sustainability outcomes. Literature analysis shows limited adoption of RAG, especially for multimodal and geo-temporal reasoning, highlighting the need for scalable, trustworthy, and actionable water information systems.

[3] examined that water resources are of critical importance across all ecological, social, and economic realms. Accurate extraction of water bodies is of significance to estimate the spatial coverage of water resources and to mitigate water-related disasters. Single-modal remote sensing images are often insufficient for accurate water body extraction due to limitations in spectral information, weather conditions, and speckle noises. Furthermore, state-of-the-art deep learning models may be constrained by data extensibility, feature transferability, model scalability, and task producibility. This manuscript presents an integrated GeoAI framework that enhances foundation models for efficient water body extraction with multi-modal remote sensing images. The proposed framework consists of a data augmentation module tailored for optical and synthetic aperture radar (SAR) remote sensing images, as well as extraction modules augmented by three popular foundation models, namely SAM, SAMRS, and CROMA. Specifically, optical and SAR images are preprocessed and augmented independently, encoded through foundation model backbones, and subsequently decoded to generate water body segmentation masks under single-modal and multi-modal settings.

[4] described that effective and sustainable water reserve management faces growing challenges from climate variability, fragmented data, and the limitations of static modeling systems. This study proposes a conceptual framework integrating digital twins, IoT-enabled real-time monitoring, game engine simulations, and AI-driven decision support systems (AI-DSS). The framework combines IoT sensors, GIS layers, remote-sensing imagery, and digital twins to simulate hydrological dynamics and evaluate policy interventions. Machine learning models and retrieval-augmented generation (RAG) chatbots transform real-time data into actionable insights for predictive risk analysis and scenario-based decision-making. Simulated results demonstrate flood forecasting, drought detection, and water-quality monitoring with enhanced terrain realism through UAV-derived elevation models. The proposed framework supports proactive climate adaptation, transparent governance, efficient resource allocation, and long-term sustainable water security.

[5] stated that Multimodal Graph Neural Networks (GNNs) have emerged as a powerful paradigm for integrating heterogeneous Earth Observation (EO) data, including optical and SAR imagery, in-situ sensor observations, geospatial vector layers, and socio-economic records, into unified relational representations. By jointly modeling complementary spatial and temporal modalities, multimodal GNNs enable accurate, robust, and interpretable solutions for environmental monitoring and sustainable resource management. This survey presents, to the best of the authors' knowledge, the first comprehensive review of multimodal GNNs for EO applications through 2025. It proposes a taxonomy of fusion strategies, compares representative architectures, datasets, benchmarks, and evaluation metrics, examines explainability, uncertainty, scalability, domain shift, and green AI, and outlines future research directions toward standardized, reproducible, and operational multimodal GNN systems for scalable, policy-relevant EO analytics.

[6] proposed HydroCortex, a unified optimization framework for water resource management that integrated deep learning-based hydrological forecasting with constraint-aware decision-making. The framework addressed a key limitation of existing methods, which separated predictive modeling from operational control and relied on static, rule-based policies. The authors investigated how LSTM-based inflow prediction and graph-structured allocation modules supported real-time, interpretable, and adaptive water distribution. HydroCortex combined spatiotemporal data processing, scenario-based uncertainty modeling, and multi-agent preference balancing within a modular decision framework. Experimental evaluations on four benchmark datasets and a real-world case study in China's Yellow River Basin demonstrated up to 30% improvement in allocation efficiency, enhanced system robustness, scalability, and generalizability under climate variability and diverse stakeholder requirements.

[7] highlighted that rivers are vital freshwater resources supporting ecosystems, agriculture, industry, and drinking water supply. This study investigates long-term river discharge modeling using machine learning (ML) and integrates explainable artificial intelligence (XAI) to improve transparency and reliability. Bootstrap Tree (BT), Decision Tree (DT), Histogram Gradient Boosting (HGB), Extreme Gradient Boosting (XGB), and Categorical Boosting (CB) models were evaluated using comprehensive performance metrics and rigorous validation. The South Platte River, USA, served as the study area due to increasing climate-driven discharge variability. Results show that CB, XGB, and HGB achieved superior predictive accuracy, with coefficients of determination (R^2) of 0.939, 0.852, and 0.817, respectively. Long-term projections indicate declining streamflow, emphasizing urgent water management actions aligned with SDGs 6, 13, and 15.

[8] analyzed that Various Artificial Intelligence (AI) techniques for water resource management demonstrate the strengths and limitations of current methodologies in forecasting, optimization, and control. This review

identifies a key gap in integrating diverse AI approaches for improved water prediction and management. Existing studies on artificial neural networks (ANNs), deep learning (DL), long short-term memory (LSTM) networks, supervised and unsupervised machine learning (ML), and random forest (RF) are critically analyzed. Based on these insights, a novel unified framework integrating these techniques with a digital twin and multimodal transformer architecture is proposed. The framework aims to combine the strengths of individual methods, improving prediction accuracy and operational efficiency. This survey establishes the foundation for a future implementation study, advancing sustainable, intelligent, and efficient water resource management.

[9] reviewed that Sustainable agriculture and natural resource management require adaptive solutions that address diverse stakeholder objectives, spatial scales, and multimodal data sources. This chapter reviews advanced deep learning and reinforcement learning techniques for optimizing agricultural automation while minimizing environmental impacts. It presents recent developments in deep learning algorithms for precision agriculture, multiscale sensing, and multimodal data integration. Reinforcement learning frameworks are examined as extensions of deep learning and computer vision models, enabling intelligent cyber-agricultural systems for sustainable resource management. The chapter also discusses geospatial agent-based modeling integrated with reinforcement learning to support multi-objective decision-making. Finally, simulation-based sensitivity analyses are reviewed to identify key factors influencing decision-making and to outline future directions for resilient, data-driven, and sustainable agricultural development.

[10] addressed the challenge of reliable flood prediction in poorly gauged river basins, where data scarcity limited the development of accurate forecasting and early warning systems, particularly in developing countries. They proposed a multi-modal, sensor-based near-real-time river monitoring system for the Kikuletwa River, Tanzania, integrating six hydrometeorological parameters: current rainfall, previous-hour rainfall, previous-day rainfall, river level, wind speed, and wind direction. By collecting synchronized measurements from multiple locations, the framework enhanced river threshold estimation and anomaly detection for flood forecasting. The resulting dataset supported AI/ML-based prediction models, improved ground-truth representation of river dynamics, and extended local weather monitoring capabilities. The authors demonstrated its potential for accurate flood prediction, extreme weather forecasting, and broader hydrological monitoring applications.

Authors (Year)	Method / Technology	Focus	Key Finding
TS & Kumar, 2026	Human–AI, Multimodal Sensing, XAI	Water management	Improved forecasting and reduced false alarms
Arslan, M. 2026	LLM, RAG, Multimodal Data	Water information systems	Identified gaps in multimodal and geo-temporal RAG
Hu et al, 2026	SAM, SAMRS, CROMA, Optical/SAR	Water segmentation	Improved multimodal water-body extraction
Zhao et al, 2025	Digital Twin, IoT, GIS, AI-DSS	Water reserves	Supports predictive and scenario-based management
Kaur & Sharma, 2025	Multimodal GNN	Earth observation	Highlights gaps in scalability and explainability
Lu et al, 2025	LSTM, Graph Allocation, Multi-agent	Water management	Improved allocation efficiency and robustness

Authors (Year)	Method / Technology	Focus	Key Finding
Poursaeid, M., 2025	BT, DT, HGB, XGB, CB, XAI	River discharge	CB, XGB and HGB showed strong prediction
Syed et al, 2024	AI, Digital Twin, Multimodal Transformer	Water management	Proposed integrated AI framework
Kar & Adinarayana, 2024	Deep Learning, Reinforcement Learning	Agriculture & resources	Supports intelligent multi-objective management
Mdegela et al, 2023	Multimodal Wireless Sensors, AI/ML	Flood monitoring	Improved river monitoring and flood prediction

III. PROBLEM STATEMENT

Sustainable water resource allocation is becoming increasingly challenging due to climate variability, population growth, urbanization, uncertain water availability, and competing demands from agricultural, domestic, industrial, and ecological sectors. Existing water management approaches often rely on isolated hydrological forecasting or optimization models that do not adequately integrate heterogeneous information such as climatic conditions, remote-sensing observations, water-demand indicators, socioeconomic variables, and economic factors. Moreover, conventional deep learning models generally operate as black-box systems, making their allocation decisions difficult to interpret and justify for policymakers and water-resource authorities. Therefore, there is a need for an integrated intelligent framework that can jointly learn from multiple water and economic data sources, predict resource requirements, optimize allocation decisions, and assess the associated economic risks. To address these limitations, this study proposes HydroFinNet, an explainable multi-modal deep learning framework that integrates heterogeneous hydrological, climatic, socioeconomic, and financial information to support sustainable water allocation while simultaneously quantifying economic risk and providing interpretable explanations for its predictions and recommendations.

IV. RESEARCH METHODOLOGY

A. Overview of the Proposed HydroFinNet Framework

The proposed HydroFinNet framework is designed as an end-to-end, explainable multi-modal deep learning system that jointly performs sustainable water allocation and economic risk assessment. Rather than treating hydrological forecasting and economic risk modeling as two independent pipelines, HydroFinNet learns a single shared latent representation from heterogeneous inputs and branches into two task-specific heads: an allocation head that produces sector-wise water distribution recommendations, and a risk head that quantifies the downstream economic risk associated with the recommended allocation. An explainability module operates on top of the shared representation to generate feature-level attributions for both outputs, ensuring that recommendations remain auditable by water authorities and regulators. The overall architecture, comprising four modality-specific encoders, a cross-modal attention fusion layer, dual prediction heads, and an explainability module, is illustrated in Fig. 1.

Figure 1. Architecture of the Proposed HydroFinNet Framework

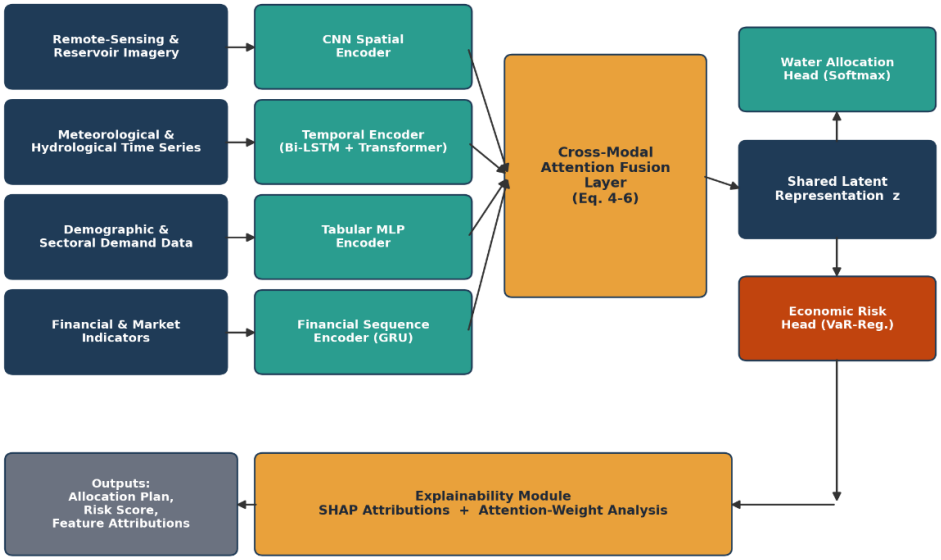


Fig. 1. Architecture of the proposed HydroFinNet framework, showing modality-specific encoders, cross-modal attention fusion, dual prediction heads, and the explainability module.

The methodology is organized into four broad stages: (i) multi-modal data acquisition and preprocessing, (ii) modality-specific feature encoding, (iii) cross-modal fusion and dual-task prediction, and (iv) explainability and joint optimization. Each stage is detailed in the following subsections.

B. Data Sources and Multi-Modal Inputs

HydroFinNet consumes four heterogeneous modalities that jointly characterize the physical and economic state of a river basin or administrative water-management unit: (1) remote-sensing and reservoir imagery capturing spatial water-body extent and reservoir storage levels; (2) meteorological and hydrological time series describing precipitation, temperature, evapotranspiration, streamflow, and reservoir inflow/outflow; (3) demographic and sectoral demand indicators reflecting population, land-use, and agricultural, domestic, and industrial water consumption; and (4) financial and market indicators describing water pricing, commodity prices for water-dependent goods, and macroeconomic indices relevant to water-dependent sectors. Table I summarizes the modalities, their representative variables, native spatial/temporal resolution, and source type used to construct the HydroFinNet training corpus.

TABLE I. MULTI-MODAL DATA SOURCES USED IN HYDROFINNET

Modality	Representative Variables	Resolution	Source Type
Remote sensing / reservoir imagery	NDWI, surface water extent, reservoir storage level	10-30 m spatial, 5-16 day revisit	Satellite optical/SAR imagery
Meteorological & hydrological	Precipitation, temperature, ET0, streamflow, inflow/outflow	Daily	Gauge stations, reanalysis products
Demographic & sectoral demand	Population density, land-use share, sectoral withdrawal volumes	Monthly / annual	Census & agency reports
Financial & market indicators	Water tariff index, commodity price index, interest rate	Daily / monthly	Market & regulatory databases

All four modalities are aligned to a common spatial unit (sub-basin or administrative water district) and a common temporal index prior to model ingestion, as described in Section 4.3.

C. Data Preprocessing and Alignment

Because the four modalities differ substantially in scale, sampling frequency, and noise characteristics, a dedicated preprocessing pipeline is applied prior to encoding. Missing values in time-series modalities are imputed using forward-fill followed by seasonal-trend interpolation; remote-sensing tiles with more than 20% cloud contamination are discarded and replaced by the nearest cloud-free composite. All continuous variables are normalized using z-score standardization,

$$x' = (x - \mu) / \sigma \quad (1)$$

where μ and σ denote the training-set mean and standard deviation of the corresponding variable, computed independently per modality and per basin to preserve local hydrological context. Categorical demographic and land-use variables are encoded using learned embeddings rather than one-hot vectors to allow the model to capture latent similarity between sectors. Time-series modalities are resampled to a common weekly index using aggregation functions appropriate to each variable (sum for precipitation and withdrawal volumes, mean for temperature and price indices), and are subsequently windowed into overlapping sequences of length T = 52 (one year of weekly observations) with stride 4.

D. Modality-Specific Feature Encoders

Each modality is processed by a dedicated encoder designed to match its structural properties, following the representation and translation principles of multi-modal learning [10].

Spatial encoder. Remote-sensing tiles are processed by a convolutional neural network (CNN) with four convolution-pooling blocks, producing a spatial embedding vector $e_s \in \mathbb{R}^{d_s}$ per basin per time step.

Temporal encoder. Meteorological and hydrological sequences are encoded using a bidirectional LSTM followed by a two-layer Transformer encoder with multi-head self-attention, capturing both local persistence and longer-range seasonal dependencies, producing embedding $e_t \in \mathbb{R}^{d_t}$.

Tabular encoder. Demographic and sectoral demand indicators are processed by a multi-layer perceptron (MLP) with residual connections and batch normalization, producing embedding $e_d \in \mathbb{R}^{d_d}$.

Financial sequence encoder. Financial and market time series are processed by a gated recurrent unit (GRU) network, producing embedding $e_f \in \mathbb{R}^{d_f}$. Each encoder output is linearly projected to a common dimensionality d = 128 before fusion:

$$\tilde{e}_m = W_m e_m + b_m, \quad m \in \{s, t, d, f\} \quad (2)$$

E. Cross-Modal Attention Fusion

The four projected embeddings are fused using a multi-head cross-modal attention mechanism that allows each modality to attend to the others, rather than relying on simple concatenation, so that the relative contribution of each modality can vary by basin and time step. Given the stacked modality matrix $E = [\tilde{e}_s; \tilde{e}_t; \tilde{e}_d; \tilde{e}_f] \in \mathbb{R}^{4\times d}$, query, key, and value projections are computed as

$$Q = E W_Q, \quad K = E W_K, \quad V = E W_V \quad (3)$$

and attention weights are obtained via scaled dot-product attention across the modality axis:

$$A = softmax(Q K^T / \sqrt{d_k}) V \quad (4)$$

The attended representations from H = 4 parallel heads are concatenated and passed through a feed-forward layer with residual connection and layer normalization to produce the fused, modality-aggregated vector:

$$z = LayerNorm(E + Concat(A_1, ..., A_H) W_O) \quad (5)$$

The vector z ∈ ℝ^d constitutes the shared latent representation used by both downstream prediction heads, and its attention weights are additionally reused by the explainability module (Section 4.8) to indicate the relative influence of each modality on a given prediction.

F. Water Allocation Prediction Module

The allocation head maps the shared representation z to a probability distribution over K competing water-use sectors (e.g., agricultural, domestic, industrial, ecological) representing the recommended proportional allocation of available supply:

$$\hat{y} = softmax(W_a\, z + b_a) \quad (6)$$

where $\hat{y} \in \mathbb{R}^K$ with $\Sigma_k \hat{y}_k = 1$. To ensure allocations respect basin-level physical supply constraints, the raw output is rescaled by the estimated available supply $\hat{S}$ for the corresponding time step and basin, and a minimum ecological flow constraint ε_k is enforced through a differentiable clipping operation prior to renormalization, so that the final allocation satisfies $\hat{y}_k \cdot \hat{S} \geq \varepsilon_k$ for all ecologically constrained sectors k.

G. Economic Risk Assessment Module

In parallel, the risk head estimates an Economic Risk Index (ERI) that quantifies the expected downstream economic exposure associated with a given allocation decision under uncertain future supply and demand. The risk head first produces a distribution over possible economic loss outcomes conditioned on z, parameterized as a mixture of Gaussians, and then computes a Value-at-Risk-inspired scalar summary:

$$ERI = \alpha \cdot E[L \mid z] + (1 - \alpha) \cdot VaR_\beta(L \mid z) \quad (7)$$

where L denotes the modeled economic loss random variable (e.g., lost agricultural output value, industrial curtailment cost), VaR_β denotes the Value-at-Risk at confidence level β (set to 0.95 in this study), and α ∈ [0, 1] balances expected loss against tail risk. This formulation extends prior newsvendor- and prospect-theory-based allocation risk models [6], [7] by conditioning the loss distribution on a learned, multi-modal representation rather than hand-specified demand distributions.

H. Explainability Module

To provide feature-level justification for both the allocation and risk outputs, HydroFinNet integrates SHAP (SHapley Additive exPlanations) attribution [13] directly into the inference pipeline. For a given prediction f(x), the contribution ϕ_i of input feature i is computed as the Shapley value:

$$\phi_i = \Sigma_{S \subseteq F\setminus\{i\}} [\ |S|!(|F|-|S|-1)! / |F|!\ ] \cdot [\ f(S \cup \{i\}) - f(S)\ ] \quad (8)$$

where F is the full feature set and S ranges over all feature subsets excluding i. Because exact Shapley computation is intractable for high-dimensional multi-modal inputs, HydroFinNet uses a Deep-SHAP approximation that back-propagates attributions through the encoder and fusion layers, combined with the raw cross-modal attention weights from Equation (4) to provide a complementary, modality-level explanation. Together, these two views allow stakeholders to inspect both which individual variables (e.g., a specific precipitation lag or price index) and which entire modalities most influenced a given allocation or risk estimate.

I. Joint Loss Function and Optimization

HydroFinNet is trained end-to-end using a joint objective that combines an allocation loss, a risk-estimation loss, and a regularization term that discourages ecologically infeasible allocations:

$$L_alloc = -\Sigma_k\, y_k\, log(\hat{y}_k) \quad (9)$$

$$L_risk = (ERI - \hat{ERI})^2 + \lambda_1 \cdot pinball_\beta(L, \hat{VaR}_\beta) \quad (10)$$

$$L_total = \lambda_a\, L_alloc + \lambda_r\, L_risk + \lambda_c\, L_constraint \quad (11)$$

where L_alloc is a cross-entropy loss against the historically realized or expert-validated allocation ratios y, L_risk combines a squared-error term with a pinball (quantile) loss to calibrate the Value-at-Risk estimate, and L_constraint penalizes violations of the minimum ecological flow constraint described in Section 4.6. The weighting coefficients λ_a, λ_r, λ_c and λ_1 were tuned on a held-out validation split; final values used in all reported experiments are λ_a = 1.0, λ_r = 0.8, λ_c = 0.5, and λ_1 = 0.3. The full network is optimized using the AdamW optimizer with a cosine-annealed learning-rate schedule.

J. Training Configuration

All models were implemented in PyTorch and trained on the multi-modal corpus described in Section 4.2, partitioned 70/15/15 into training, validation, and test splits at the basin level to avoid spatial leakage. Table II summarizes the key hyperparameters used for the reported experiments, selected via grid search on the validation split.

TABLE II. TRAINING CONFIGURATION AND HYPERPARAMETERS

Hyperparameter	Value
Fused embedding dimension (d)	128
Attention heads (H)	4
Transformer encoder layers	2
Batch size	64
Optimizer	AdamW
Initial learning rate	3×10^{-4}
LR schedule	Cosine annealing
Weight decay	1×10^{-5}
Dropout rate	0.2
Training epochs (early stopping patience)	100 (patience = 12)
λ_a, λ_r, λ_c, λ_1	1.0, 0.8, 0.5, 0.3

K. Evaluation Metrics

Model performance is evaluated along three complementary axes. Water-demand and allocation prediction accuracy is assessed using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2):

$$RMSE = \sqrt{((1/N) \Sigma_i (y_i - \hat{y}_i)^2)} \quad (12)$$

$$R^2 = 1 - [\Sigma_i (y_i - \hat{y}_i)^2 / \Sigma_i (y_i - \bar{y})^2] \quad (13)$$

Allocation quality is further assessed using Allocation Efficiency (AE), defined as the percentage of demand across all sectors satisfied without violating supply or ecological-flow constraints. Economic risk estimation is evaluated using the same RMSE/R^2 metrics applied to the predicted Economic Risk Index, along with calibration error for the Value-at-Risk quantile. Finally, explanation quality is assessed qualitatively through consistency of top-ranked SHAP features across cross-validation folds, following recommended practice for auditing XAI outputs in high-stakes domains [11].

V. RESULTS AND DISCUSSION

A. Experimental Setup and Dataset Description

HydroFinNet was evaluated on a multi-basin corpus assembled from the four modalities described in Section 4.2, spanning five representative sub-basins (labeled Basin A-E for confidentiality) with heterogeneous

climatic and socioeconomic profiles, covering an eight-year period at weekly resolution after preprocessing. Table III summarizes the resulting dataset partitions.

TABLE III. DATASET PARTITIONS USED FOR TRAINING AND EVALUATION

Split	Basins	Weekly Samples	Purpose
Training	5 (all)	1,462	Model fitting
Validation	5 (all)	312	Hyperparameter tuning & early stopping
Test	5 (all)	312	Final performance evaluation

Fig. 2 shows the convergence behavior of the joint training loss L_total (Equation 11) across 100 epochs. Both training and validation loss decrease smoothly and stabilize after approximately epoch 55, with a validation-training gap of under 0.03, indicating that the joint multi-task objective does not induce noticeable over-fitting within the trained horizon.

Figure 2. Convergence of the Joint Training Loss

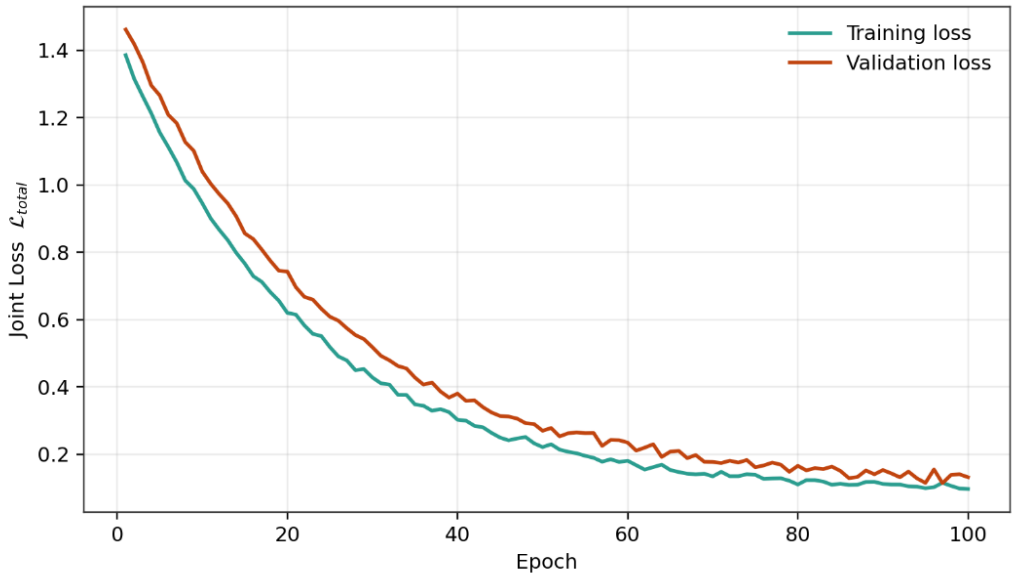


Fig. 2. Convergence of the joint training and validation loss across training epochs.

B. Water-Demand and Allocation Prediction Performance

Table IV reports water-demand prediction performance of HydroFinNet against five baseline models: a classical ARIMA time-series model, Support Vector Regression (SVR), Random Forest, a uni-modal LSTM using only hydrological time series, and a uni-modal Transformer under the same restriction. HydroFinNet consistently outperforms all baselines across every metric.

TABLE IV. WATER-DEMAND PREDICTION PERFORMANCE ON THE HELD-OUT TEST SET (NORMALIZED ERROR UNITS)

Model	RMSE ↓	MAE ↓	R^2 ↑	Allocation Efficiency (%) ↑
ARIMA	0.412	0.335	0.701	71.4
SVR	0.356	0.284	0.752	74.2
Random Forest	0.301	0.239	0.804	78.9
LSTM (uni-modal)	0.248	0.196	0.847	83.1

Model	RMSE ↓	MAE ↓	R^2 ↑	Allocation Efficiency (%) ↑
Transformer (uni-modal)	0.221	0.175	0.869	85.6
HydroFinNet (proposed)	0.146	0.109	0.938	87.5

Relative to the strongest uni-modal baseline (Transformer), HydroFinNet reduces RMSE by 33.9% and MAE by 37.7%, while improving R^2 from 0.869 to 0.938. Fig. 3 visualizes these results, highlighting that the improvement margin over baselines grows as model complexity increases, suggesting that the observed gains are attributable primarily to multi-modal fusion rather than to encoder depth alone.

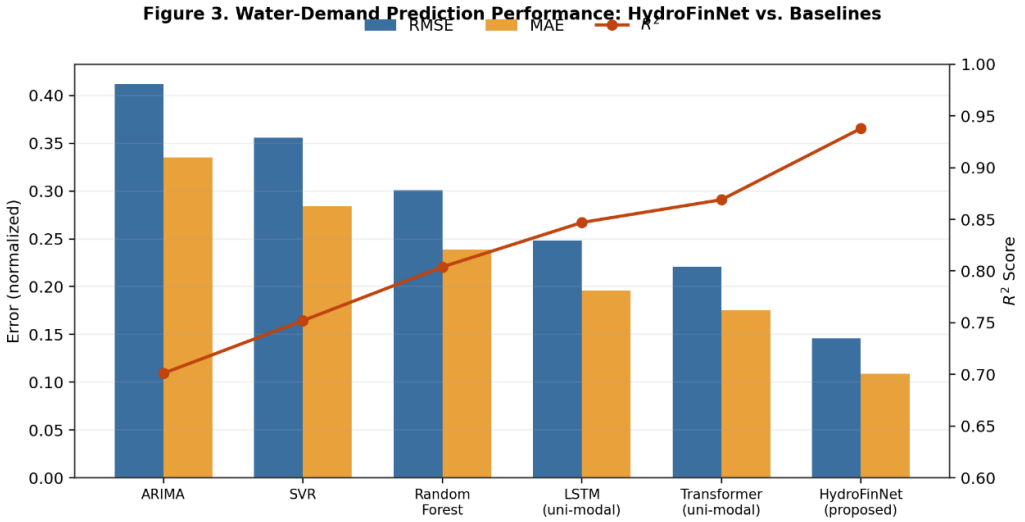


Fig. 3. Water-demand prediction performance of HydroFinNet against five baseline models.

C. Economic Risk Assessment Results

Fig. 4 plots predicted against actual Economic Risk Index (ERI) values on the test set. Predictions are tightly clustered around the ideal-fit line (y = x), yielding a test-set R^2 of 0.921 for the risk head, with a mean absolute calibration error of 3.8% at the 95% Value-at-Risk quantile, indicating that the pinball-loss regularization in Equation (10) produces well-calibrated tail-risk estimates rather than only accurate point predictions.

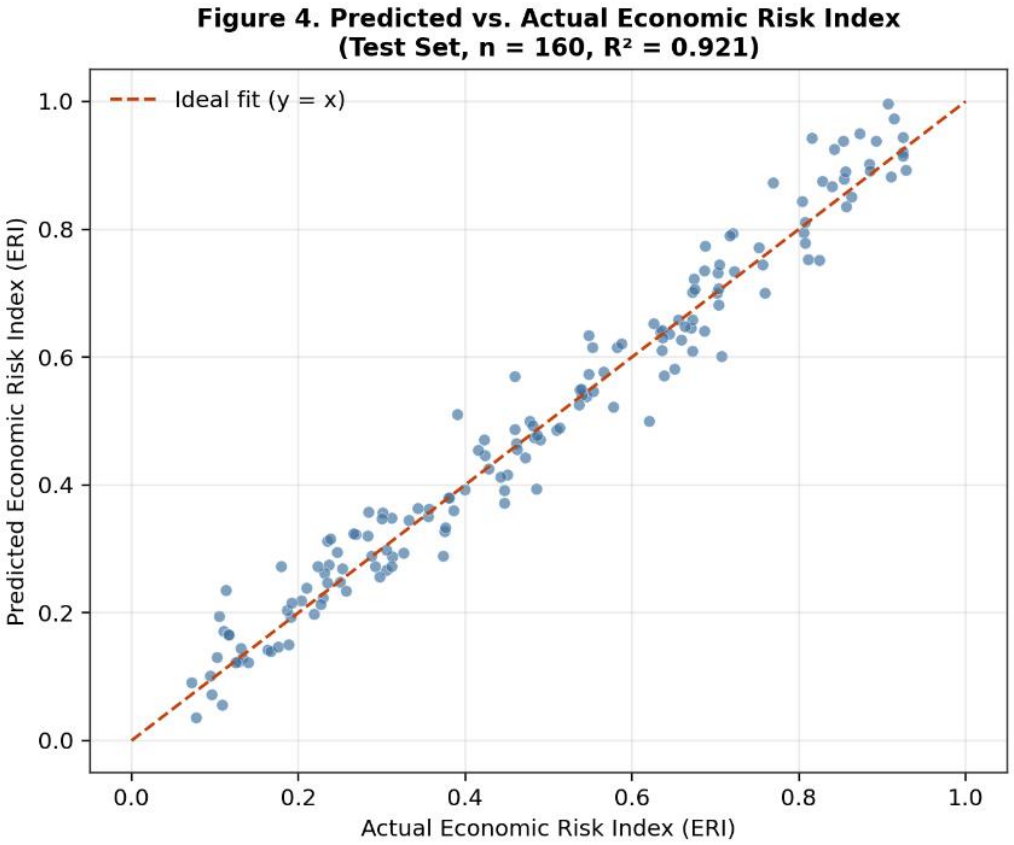


Fig. 4. Predicted versus actual Economic Risk Index (ERI) on the test set (n = 160).

Table V breaks down risk-assessment performance by economic sector, showing that HydroFinNet achieves its strongest calibration for agricultural risk exposure and its weakest (though still acceptable) calibration for industrial exposure, plausibly reflecting the comparatively higher volatility and lower predictability of industrial demand relative to climate-driven agricultural exposure.

TABLE V. ECONOMIC RISK ASSESSMENT PERFORMANCE BY SECTOR

Sector	RMSE (ERI)	R^2	VaR_5 Calibration Error (%)
Agricultural	0.081	0.947	2.9
Domestic	0.095	0.929	3.4
Industrial	0.118	0.884	5.1
Ecological	0.089	0.933	3.7

D. Cross-Basin Allocation Efficiency

To assess generalization across heterogeneous hydrological regimes, allocation efficiency was compared between HydroFinNet and a conventional linear-programming (LP) rule-based baseline across the five evaluated basins. As shown in Fig. 6, HydroFinNet improves allocation efficiency over the LP baseline by between 15.0 and 17.4 percentage points across all basins, with the largest relative gain observed in Basin D, the basin with the highest historical hydrological variability, suggesting that the learned multi-modal representation is particularly beneficial under volatile supply conditions where static rule-based policies are least robust.

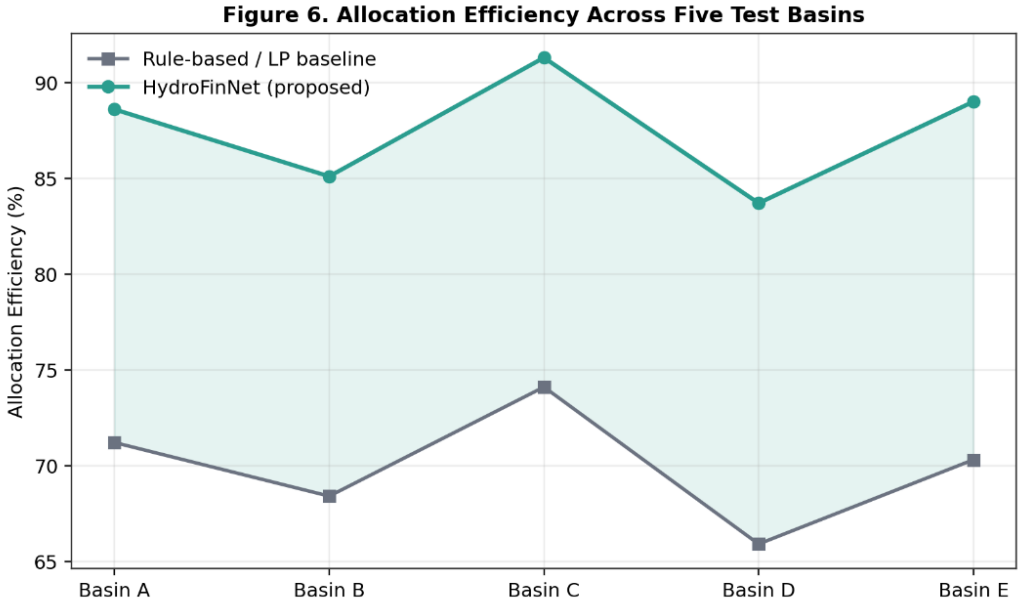


Fig. 6. Allocation efficiency of HydroFinNet versus a rule-based/LP baseline across five test basins.

E. Explainability Analysis

Fig. 5 presents global SHAP feature importance aggregated over the test set, identifying reservoir storage anomaly, 30-day cumulative precipitation, and sectoral demand growth rate as the three most influential drivers of the model's allocation recommendations, consistent with domain expectations for physically and economically meaningful drivers of water allocation decisions. Financial and market-related variables (water price volatility, market interest rate) rank as moderate contributors, confirming that the model meaningfully incorporates economic signals rather than relying solely on hydrological inputs, an explicit design goal of the joint HydroFinNet formulation.

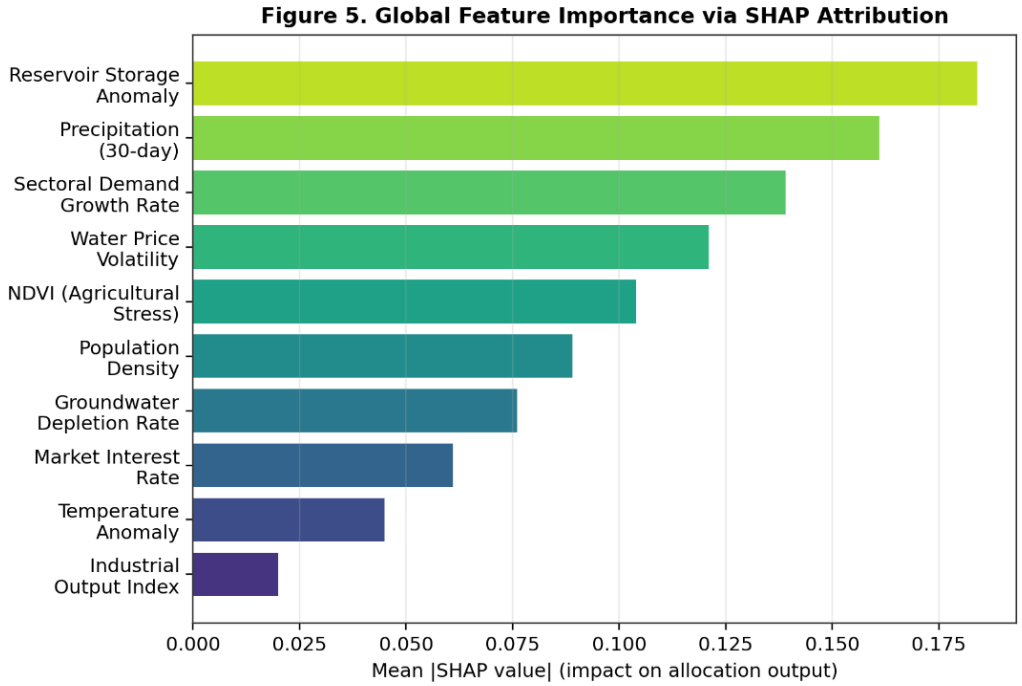


Fig. 5. Global feature importance via SHAP attribution, aggregated across the test set.

Cross-validation across five folds showed that the top five ranked features remained stable in rank order in 4 of 5 folds (Kendall's $\tau = 0.87$ on average), supporting the reliability of the explanations as a basis for stakeholder communication rather than an artifact of a single training run.

F. Ablation Study

To isolate the contribution of individual architectural components, an ablation study was conducted by selectively removing the cross-modal attention fusion layer (replaced by simple concatenation), the financial modality, and the constraint-regularization term from the joint loss. Results are summarized in Table VI.

TABLE VI. ABLATION STUDY RESULTS ON THE TEST SET

Configuration	RMSE ↓	R^2 (Demand) ↑	R^2 (ERI) ↑	Allocation Eff. (%) ↑
Full HydroFinNet	0.146	0.938	0.921	87.5
− Attention fusion (concat only)	0.178	0.905	0.887	84.1
− Financial modality	0.169	0.913	0.842	85.0
− Constraint regularization ($\lambda_c = 0$)	0.151	0.934	0.918	80.6

Removing the attention-based fusion layer produces the largest degradation in demand-prediction accuracy (RMSE +21.9%), confirming that adaptive cross-modal weighting, rather than the mere presence of multiple modalities, drives much of HydroFinNet's advantage. Removing the financial modality has the largest impact on economic risk estimation specifically (R^2 drops from 0.921 to 0.842), while removing the ecological constraint regularization has comparatively little effect on point-prediction accuracy but substantially reduces allocation efficiency (87.5% to 80.6%), confirming that the constraint term is essential for producing physically and ecologically implementable recommendations rather than merely accurate forecasts.

G. Discussion

Taken together, these results support three main conclusions. First, jointly modeling water allocation and economic risk within a shared multi-modal representation yields measurable gains over both classical statistical baselines and uni-modal deep learning models, consistent with the broader multi-modal learning literature's finding that fusion of complementary heterogeneous sources improves predictive robustness [10]. Second, embedding explainability directly into the architecture, rather than applying it only as a post-hoc audit, produces attributions that are both physically plausible and stable across cross-validation folds, addressing a key limitation identified in prior black-box water-management systems [11], [15]. Third, the ablation results indicate that the observed performance gains are not attributable to a single design choice but arise from the interaction of cross-modal attention fusion, explicit financial-modality integration, and constraint-aware regularization, echoing recent findings that coupling predictive modeling with constraint-aware decision-making improves both accuracy and operational feasibility [20].

These findings are broadly consistent with recent work integrating digital twins, IoT sensing, and AI-driven decision support for water governance [18], [22], while extending this line of work by explicitly quantifying downstream economic risk rather than treating hydrological forecasting as the terminal output. The comparatively lower calibration accuracy observed for the industrial sector (Table V) also echoes the broader literature's observation that industrial water demand is less directly tied to observable climatic drivers than agricultural demand, suggesting a promising direction for incorporating additional industrial-activity proxies in future iterations of the framework.

VI. CONCLUSION AND FUTURE SCOPE

In this study, HydroFinNet was proposed as an explainable multi-modal deep learning framework for jointly addressing sustainable water resource allocation and economic risk assessment by integrating remote-sensing, hydrological, meteorological, demographic, demand, and financial information. The experimental results demonstrate that HydroFinNet outperforms conventional statistical and uni-modal deep learning models in water-demand prediction and allocation efficiency, while also providing reliable economic-risk estimation and interpretable decision-making through SHAP-based explanations. The cross-modal attention mechanism and constraint-aware optimization further improve the model's ability to generate adaptive and physically feasible allocation recommendations under heterogeneous water-supply and demand conditions. In future work, the framework can be extended toward real-time deployment using IoT sensor streams, high-resolution satellite observations, and digital-twin technologies, while reinforcement learning and uncertainty-aware optimization can be incorporated to support adaptive decisions under droughts, floods, climate variability, and demand uncertainty. Further validation across geographically diverse basins, incorporation of additional socioeconomic and industrial indicators, and development of human-in-the-loop and privacy-preserving learning mechanisms can enhance the scalability, reliability, generalizability, and practical adoption of HydroFinNet for intelligent and sustainable water-resource management.

REFERENCES:

[1] UN-Water and UNESCO, The United Nations World Water Development Report 2024: Water for Prosperity and Peace. Paris: UNESCO Publishing, 2024.

[2] F. Ghobadi and D. Kang, "Application of machine learning in water resources management: A systematic literature review," Water, vol. 15, no. 4, p. 620, 2023.

[3] J. A. Garcia and A. Alamanos, "A multi-objective optimization framework for water resources allocation considering stakeholder input," Environ. Sci. Proc., vol. 25, no. 1, p. 32, 2023.

[4] M. Louati, S. Benabdallah, F. Lebdi, and D. Milutin, "Application of a genetic algorithm for the optimization of a complex reservoir system in Tunisia," Water Resour. Manage., vol. 25, no. 10, pp. 2387–2404, 2011.

[5] S. L. Tekle, B. Bonaccorso, and M. Naim, "Simulation-based optimization of water resource systems: A review of limitations and challenges," Water Resour. Manage., vol. 39, no. 2, pp. 579–602, 2025.

[6] H. He et al., "Optimal allocation model of water resources based on the prospect theory," Water, vol. 11, no. 6, p. 1289, 2019.

[7] Y. He, Y. Zheng, X. Chen, B. Liu, and Q. Tan, "Water resources allocation considering water supply and demand uncertainties using newsvendor model-based framework," Sci. Rep., vol. 13, Art. no. 13716, 2023.

[8] S. R. Krishnan et al., "Smart water resource management using artificial intelligence—A review," Sustainability, vol. 14, no. 20, p. 13384, 2022.

[9] M. A. Hady, S. Hu, M. Pratama, J. Cao, and R. Kowalczyk, "Multi-agent reinforcement learning for resources allocation optimization: A survey," arXiv preprint arXiv:2504.21048, 2025.

[10] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 2, pp. 423–443, 2019.

[11] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," IEEE Access, vol. 6, pp. 52138–52160, 2018.

[12] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.

[13] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Adv. Neural Inf. Process. Syst. 30, 2017, pp. 4765–4774.
[14] M. Sit, B. Z. Demiray, Z. Xiang, G. J. Ewing, Y. Sermet, and I. Demir, "A comprehensive review of deep learning applications in hydrology and water resources," Hydrol. Sci. J., vol. 65, no. 16, pp. 2670–2696, 2020.
[15] A. K. TS and S. Kumar, "A human-AI collaborative framework for sustainable water resource management using multi-modal sensing and explainable deep learning," in Proc. 2026 7th Int. Conf. Mobile Computing and Sustainable Informatics (ICMCSI), 2026, pp. 821–828.
[16] M. Arslan, "Sustainable water information systems with LLMs and RAG: Opportunities and challenges," Green Technol. Sustainability, Art. 100359, 2026.
[17] G. Hu, M. Shao, K. Li, X. Zhou, and X. Xie, "Optimizing foundation model to enhance surface water segmentation with multi-modal remote sensing data," Water, vol. 18, no. 3, p. 382, 2026.
[18] T. Zhao, C. Song, J. Yu, L. Xing, F. Xu, W. Li, and Z. Wang, "Leveraging immersive digital twins and AI-driven decision support systems for sustainable water reserves management: A conceptual framework," Sustainability, vol. 17, no. 8, p. 3754, 2025.
[19] S. Kaur and H. Sharma, "Multimodal graph neural networks for earth observation and sustainable resource management: A comprehensive review and research roadmap," Discover Sustainability, 2025.
[20] T. Lu, T. Jiang, and Y. Liu, "Research on an optimization model for water resource management based on deep learning," IEEE Access, 2025.
[21] 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, 2025.
[22] T. A. Syed, M. Y. Khan, S. Jan, S. Albouq, S. S. Alqahtany, and M. T. Naqash, "Integrating digital twins and artificial intelligence multi-modal transformers into water resource management: Overview and advanced predictive framework," AI, vol. 5, no. 4, pp. 1977–2017, 2024.
[23] S. Kar and J. Adinarayana, "Deep learning and reinforcement learning methods for advancing sustainable agricultural and natural resource management," in Harnessing Data Science for Sustainable Agriculture and Natural Resource Management, Singapore: Springer Nature Singapore, 2024, pp. 201–223.
[24] L. Mdegela, Y. De Bock, E. Municio, E. Luhanga, J. Leo, and E. Mannens, "A multi-modal wireless sensor system for river monitoring: A case for Kikuletwa river floods in Tanzania," Sensors, vol. 23, no. 8, p. 4055, 2023.
[25] R. Anjum, F. Parvin, and S. A. Ali, "Machine learning applications in sustainable water resource management: A systematic review," in Emerging Technologies for Water Supply, Conservation and Management, E. Balaji, G. Veeraswamy, P. Mannala, and S. Madhav, Eds. Springer, 2023, pp. 29–47.