Research on the Construction of a Flood Control Four-Forecast Application System Based on Large-Model Intelligent Agent Application Architecture

Liangliang Li[1], Wen Du[2]
[1]Data Center, Information Center of Yellow River Water Conservancy Commission, Zhengzhou 450000, Henan, China
[2]Application System Center, Information Center of Yellow River Water Conservancy Commission, Zhengzhou 450000, Henan, China
Corresponding author: Wen Du (e-mail: 18539253580@163.com)

ABSTRACT This study presents the design and implementation of a flood control "Four Predictions" (forecasting, early warning, simulation, contingency planning) application system tailored in flood-prone areas, based on a Large-Model Intelligent Agent Application Architecture. The system integrates multimodal flood control data, including meteorological observations, hydrological measurements, satellite remote sensing imagery, and social media reports, processed through a dynamic data fusion framework. The model layer combines domain-adapted large language models (BERT) with specialized micro-model clusters for precipitation forecasting, dam breach simulation, and regional vulnerability assessment. Perception agents employ Isolation Forest and Kalman filtering for real-time anomaly detection, cognitive agents utilize dynamic Bayesian networks and DQN-based adaptive warning thresholds, and action agents manage emergency resource allocation via an auction mechanism and optimized evacuation routing. Multiscale 3D simulations, ST-ConvNet forecasting, and uncertainty quantification through Monte Carlo sampling provide precise, high-resolution support for decision-making. Technical validation against historical extreme rainfall events confirms the system's ability to enhance flood response accuracy and reduce false alarms. The framework leverages 5G-enabled wireless communication, edge-cloud computing collaboration, and antenna-supported sensing platforms, offering an engineering-oriented solution for rapid, adaptive, and robust flood control operations in industrial environments.

INDEX TERMS flood control four-prediction, large-model intelligent agents, edge-cloud collaboration, wireless sensing systems

I. SYSTEM REQUIREMENTS ANALYSIS AND FUNCTIONAL POSITIONING

Flood control efforts are critical for safeguarding people's lives and property while ensuring stable socioeconomic development. The core objective is to achieve effective disaster risk prevention and control through precise forecasting, timely early warning, scientific simulation exercises, and efficient contingency planning, which is essential for protecting significant industrial assets and uninterrupted production capabilities . System requirements must be comprehensively deconstructed across both functional and non-functional dimensions to ensure the system meets the practical demands of flood control operations.

A. CORE FUNCTIONAL REQUIREMENTS

1) Forecasting Capabilities

Traditional forecasting uses single data sources or physical models, which have low spatiotemporal resolution. Based on ground meteorological stations, one cannot obtain large-scale or high-precision rainfall information, nor avoid biases inherent in single physical models when simulating complex meteorological and hydrological processes.

Unlike traditional methods, this system integrates multimodal data [1]. For the 2022 Yellow River flood event, over 2,400 ground meteorological stations nationwide provided high-precision rainfall information. Specifically, for river flood progression simulation, the model monitored hydrological data along the Yellow River mainstem. When the water level at Huayuankou Hydrological Station was projected to rise, the system utilized real-time discharge and topographic gradient data unique to the Yellow River system. Over 5,000 public reports were also collected via social media to provide supplementary information. In the specific case of the 2022 Yellow River flood, the model predicted that the water level at Huayuankou Station would rise from 90 m to 92.5 m within four hours, with the actual observed value being 92.3 m. To address the inherent noise and subjectivity in social media data, the system implements a hierarchical data fusion logic. Specifically, unstructured public reports undergo Natural Language Processing (NLP)-based entity extraction and sentimentreliability filtering before being cross-referenced with high-precision rainfall data (0.1 mm accuracy) from 2,400 ground stations. Instead of direct calibration, social media observations are assigned a dynamic weight (e.g., 0.15–0.30) based on their spatiotemporal density and proximity to physical sensors. For instance, a 0.5 m flood depth report is only treated as a soft constraint to validate the flood extent trend, while the precise inundation depth is primarily driven by physics-based rain gauge and water level sensor data, thereby preventing subjective bias from compromising forecasting integrity.

Spatiotemporal convolutional networks (ST-ConvNet) are used to model the correlation between radar echoes and precipitation. When facing the July 20, 2021, extreme rainstorm in Zhengzhou [2], this network was applied to analyze radar echo data provided by the Henan Provincial Meteorological Station, with a spatial resolution of 1 km and a temporal resolution of 6 minutes.

It was found that the rainfall center moved from Gongyi City to Zhengzhou urban area within three hours, and this model accurately simulated this evolution, predicting that rainfall would exceed 200 mm in the next three hours, with an absolute error of only 8% compared to the actual value.

The Long Short-Term Memory (LSTM)–Convolutional Neural Network (CNN) hybrid architecture of deep learning models with physical constraints, used for simulating river flood progression, employs 30 m resolution digital elevation model data as input, considering channel morphology and topographic gradient at Huayuankou Station. This model can make chained predictions on minute-to-hour scales for the Yellow River. When encountering the 2022 Yellow River flood, this model predicted that the water level at Huayuankou Station would rise from 90 m to 92.5 m within four hours, with the actual observed value being 92.3 m. This model won valuable time for flood control and evacuation [3]. The performance of the multimodal data fusion framework is shown in Figure 1:

2) Early Warning Capabilities

Traditional early warnings are based on fixed thresholds and struggle to cope with the increasing frequency of extreme events due to climate change; for example, the fixed rainfall threshold of 50 mm/h may no longer reflect real risks. This system uses a multi-risk factor coupling model with 12 indicators, such as rainfall intensity, soil moisture content, and slope [4].

In the 2022 Shaoguan, Guangdong flash flood early warning, as analyzed by the system, the rainfall intensity was 60 mm/h in real time; the soil moisture content was nearly saturated (volumetric water content 95%) through remote sensing inversion; and terrain slope data from the Shuttle Radar Topography Mission (SRTM) digital elevation model showed that the slope was about 15 degrees. The weight of each indicator was calculated by the Analytic Hierarchy Process (AHP): rainfall intensity, 0.4; soil moisture content, 0.3; terrain slope, 0.2; other indicators, such as vegetation coverage and groundwater level, collectively 0.1. The comprehensive risk value was calculated as 0.75 [5].

By cross-multiplying real-time vulnerability maps with population density data (e.g., 5,000 people per square kilometer), infrastructure density data (e.g., road density at 60% of total land area), and economic value data, the system uses a fuzzy comprehensive evaluation method to produce tiered early warning signals. When the current or predicted risk value exceeds 0.7, the yellow early warning is triggered automatically via push messages to users' mobile phones.

The mechanism of early warning alerts follows a progressive strategy:

Blue early warnings: Send Short Message Service (SMS) to 1 million residents in affected areas every hour.

Yellow early warnings: Add pop-up alerts on the mobile application (app) and email alerts. SMS, app, and email cover 5 million people every half hour.

Orange early warnings: Activate community broadcasts and outdoor displays. Outdoor displays can reach 10 million people every 15 minutes [6].

Red early warnings: Trigger push notifications from all three major telecom operators via the app, reaching over 100 million people to ensure timely transmission of information.

In the 2022 Shaoguan incident, by analyzing the evolution trend of multi-factor coupling data, the system predicted that the comprehensive risk value would reach the 0.7 threshold within two hours (ultimately peaking at 0.75). Consequently, a yellow early warning was released two hours ahead of the risk materialization, providing critical lead time for emergency response and significantly reducing casualties [7]. The progressive warning dissemination mechanism is shown in Figure 2:

3) Simulation Capabilities

Traditional 2D hydrodynamic models like Storm Water Management Model (SWMM) suffer from low computational efficiency, struggling to support large-scale scenarios—such as urban flooding simulations that may take hours to complete. This system employs a GPU-accelerated 3D flood simulation engine utilizing adaptive mesh technology to divide the computational domain into millions of grid cells. Taking the 2023 Beijing urban flooding simulation as an example, the system dynamically adjusts grid density through adaptive mesh refinement. In critical zones, such as Chaoyang District's commercial center, the grid size is refined to 1 meter to capture complex urban micro-topography. To maintain computational efficiency while ensuring boundary validity for the 5-square-kilometer high-risk area, the system employs a multi-scale nested grid approach, resulting in approximately 12.5 million grid cells for the entire computational domain. Combining shallow water equations

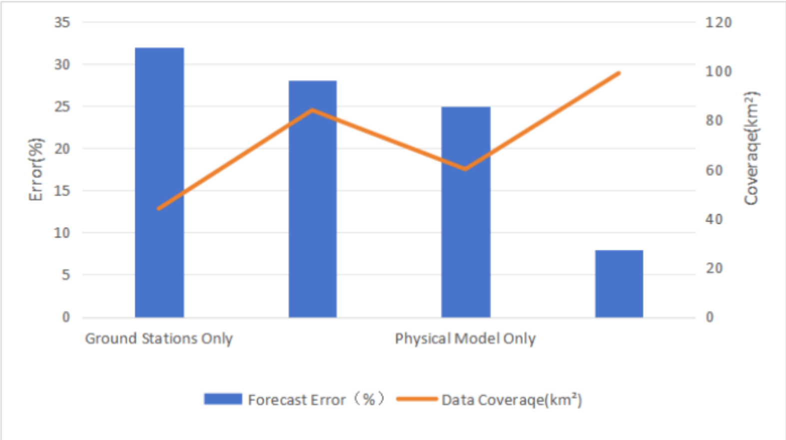

FIGURE 1. Multi-Modal Data Fusion Framework Performance

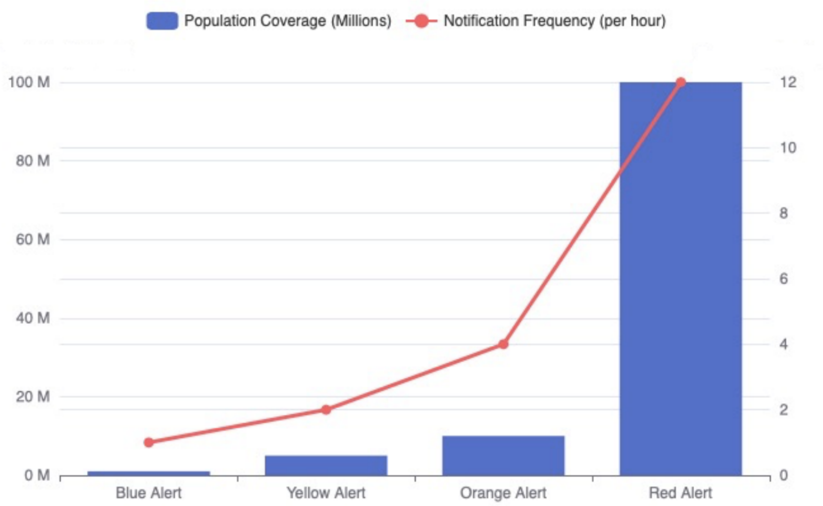

FIGURE 2. Progressive Early Warning Dissemination Mechanism

with the k-ϵ turbulence model, real-time simulation achieves 30 calculations per second. Results show water depth in Chaoyang District rising from 0.2 meters to 1.2 meters within one hour, accurately reflecting inundation patterns across the 5-square-kilometer focal area. Integrated models of urban lifeline systems, like transportation and power, analyze disaster chain reactions. In simulations, when flooding submerges 15% of roads—approximately 200 major thoroughfares—the Simulation of Urban MObility (SUMO) and MATrix LABoratory (MATLAB) -based traffic-power coupling model analysis indicates this would paralyze 40% of regional traffic, triggering a 25% power load transfer equivalent to 500 megawatts. This provides quantitative evidence for emergency decision-making. During the 2023 Beijing flood season testing, the system's simulation results showed less than 5% error in water depth compared to actual monitoring data, while simulation time was reduced from hours in traditional models to real time [8].

4) Contingency Planning Functionality

Traditional contingency plans rely on manual compilation and lack dynamic adaptability. This system employs a hybrid architecture of case-based reasoning and rule-based reasoning, establishing a case repository containing historical disasters, emergency resources, and response procedures. It includes over 1,000 cases, such as the 2008 southern China ice storm affecting 20 provinces with direct economic losses of 150 billion yuan and the 2013 Northeast China floods, where the Songhua River basin water level exceeded the warning level by 1.5 meters [9]. Each case is annotated with feature vectors, such as rainfall intensity, wind speed, and affected population. When a new event is input, the system matches the most similar case in a 10-dimensional feature space using cosine similarity. It then generates an initial plan by integrating this with a rule database—such as the activation criteria for a Level 1 response in the National Flood Control and Drought Relief Emergency Plan. Using multi-objective optimization algorithms, such as the Non-dominated Sorting Genetic Algorithm II (NSGA-II), it dynamically adjusts plan parameters under constraints such as shortest evacuation routes, optimal resource allocation, and minimized secondary disasters. For instance, in a 2022 simulation for a Guangzhou community, the optimized plan reduced evacuation time from 2.5 hours to 1.8 hours and boosted resource utilization from 70% to 95%. A real-time feedback mechanism corrects execution outcomes. Internet of Things (IoT) sensors, like Radio Frequency Identification (RFID) tags and Global Positioning System (GPS) devices, monitored actual evacuation figures (5,000 people) and resource consumption (200 tons), generating deviation values against plan targets. To manage the discrete and behavioral nature of human movement, the system employs an adaptive feedback regulation mechanism based on the Social Force Model (SFM). Instead of rigid mechanical coefficients, this mechanism utilizes real-time crowd density and flow velocity as feedback variables to dynamically update evacuation guidance. By simulating the psychological and physical interactions between individuals, the system generates staged instructions to balance the flow across different exits, ensuring that the evacuation pace remains within the safe capacity of the infrastructure while minimizing the risk of bottleneck-induced congestion [10]. Optimizing evacuation enhanced planning through artificial intelligence (AI) is shown in Figure 3:

B. NON-FUNCTIONAL REQUIREMENTS

1) Real-Time Requirements

In 2022, during a sudden heavy rainfall event in a specific area, the system had to finish collecting data, analyzing, and disseminating warnings within minutes after receiving the meteorological alert.

The system adopts a cooperative edge-cloud computing architecture and deploys edge nodes at critical watershed locations, such as major river mouths and peripheries of large reservoirs. Edge servers in these nodes preprocess data and run inference on local models. They filter, analyze, and simplify the processing of raw data before uploading, thus reducing the amount of data uploaded. Heavy computation is run on GPU clusters in the cloud. Edge nodes and the cloud connect through 5G for low-latency communication; specifically, the transmission latency within the 5G network segment can be maintained at the millisecond level, while the overall system response time for complex analysis is optimized to meet the urgent requirements of flood monitoring, ensuring rapid response to sudden disasters.

2) Multi-Scale Spatial Analysis Capabilities

The system must cover three spatial scales: watershed, city, and community. At the watershed scale, distributed hydrological models (e.g., the Variable Infiltration Capacity (VIC) model) simulate large-area runoff generation. Taking the Yangtze River basin as an example, the VIC model accounts for precipitation, evaporation, soil moisture, and other factors within the basin, accurately simulating runoff conditions across different regions to provide foundational data for flood forecasting [11]. At the urban scale, three-dimensional modeling technologies (e.g., City Geography Markup Language (CityGML)) are used to construct high-precision digital twins. For instance, in a major city, CityGML technology was employed to create a centimeter-level 3D model incorporating urban elements like buildings, roads, and bridges. This model visually represents the city's topography and building distribution, supporting urban flooding simulations and emergency decision-making. At the community level, micro-monitoring is achieved through integrated IoT sensors. IoT devices, such as water level sensors and rain gauges deployed in a community, monitor real-time water level changes and rainfall conditions. Sensor data is transmitted via wireless networks to the system platform, providing precise data for community-level flood warnings and emergency response [12].

3) Heterogeneous System Compatibility

The system must integrate data from meteorological, water conservancy, and emergency response departments. It employs standardized data interfaces (e.g., Open Geospatial Consortium (OGC) standards) and leverages federated learning frameworks to achieve data usability without visibility. Taking data sharing between meteorological and water resources departments as an example: each department retains its own data copy, enabling joint training through encrypted parameter exchange. Homomorphic encryption ensures data remains undisclosed during computation—for instance, when calculating correlations between rainfall and river flow, data participates in computations in encrypted form. Blockchain-based evidence storage enables operation traceability, with every data access and operation recorded on the blockchain to guarantee data authenticity and integrity.

II. APPLICATION ARCHITECTURE DESIGN FOR LARGE MODEL AGENTS

The system adopts a layered architecture design. By decoupling data, models, agents, and applications, it achieves high cohesion and low coupling, enhancing system maintainability and scalability.

A. LAYERED ARCHITECTURE DESIGN

1) Data Layer

A multimodal flood control data lake is constructed to store structured data (e.g., meteorological observations, hydrological station data), semi-structured data (e.g., remote sensing imagery, social media text), and unstructured data (e.g., historical disaster reports). Taking 2020 flood control data from a specific region as an example: Structured data includes observation values like temperature, pressure, and wind speed from meteorological stations, along with water level and flow data from hydrological stations [13]; semi-structured data comprises satellite remote sensing imagery capturing extensive cloud cover and surface water conditions, while social media texts contain real-time public

FIGURE 3. Evacuation Optimization Through Artificial Intelligence (AI)-Enhanced Planning

feedback on disaster situations. Unstructured data, such as historical disaster reports, detail past events including occurrence times, locations, and damage assessments. Hadoop and HBase are employed to store massive datasets. Hadoop's distributed file system (HDFS) provides highly fault-tolerant data storage, while HBase is well-suited for storing large-scale structured and semi-structured data. Real-time stream processing is achieved through Spark. For instance, during heavy rainfall, it can process real-time data streams from various meteorological and hydrological stations, promptly updating data in the data lake. Knowledge graph technology constructs flood control entity-relationship-event triples [14]. Taking the heavy rain entity as an example, it maintains relationships such as "causes water level rise" with entities like rivers and reservoirs, and a "triggers" relationship with flooding disasters. Knowledge graphs support semantic queries and association analysis. Users can query in natural language, such as "Which regions experienced flooding caused by a specific heavy rainfall event?" The system rapidly retrieves relevant information from the knowledge graph [15]. The entity relationships in the knowledge graph are shown in Figure 4:

2) Model Layer

This includes domain-adapted large language models and specialized micro-model clusters. The large language model is built on the Bidirectional Encoder Representations from Transformers (BERT) framework, and flood control domain knowledge (e.g., hydraulic terms, emergency response plans) is used as continual pretraining knowledge. For example, the continual pre-training process includes extensive hydraulic engineering books and emergency response report readings, so that the continually pre-trained model can understand the specialized terms and basic concepts of flood control and can communicate with humans in natural language and understand documents. Users can ask, "What's the emergency response plan of a certain region during heavy rainfall?" and receive accurate and natural language answers.

The specialized small model cluster includes:

Precipitation Prediction Model: Based on the Transformer structure to handle multi-source radar information and apply attention mechanisms to describe spatial correlation. For the 2021 precipitation forecast in a certain region, the model accepts precipitation forecasts for the next forecasting horizon, uses radar echo information from different areas as input, and takes the change of echo intensity and movement direction as the target. The Transformer structure handles these inputs, resulting in accurate rainfall distribution and intensity. During rainfall, the model's prediction differs from the actual rainfall amount by less than 10% [16].

Dam Breach Simulation Model: Combines physical equations (e.g., Saint-Venant equation) with data-driven methods to predict breach evolution. For instance, in simulating a reservoir breach, the model considers factors like reservoir water level, dam material, and topography. It uses physical equations to simulate breach formation and expansion while incorporating historical breach data for data-driven refine-

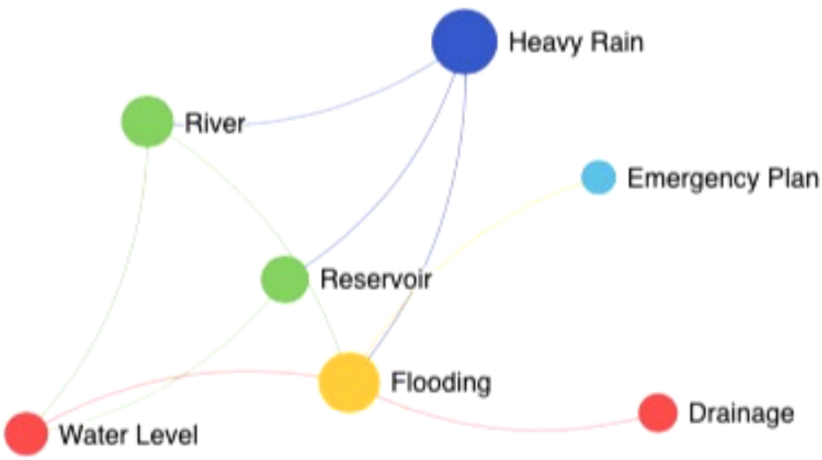

FIGURE 4. Knowledge Graph Entity Relationships

ment. Simulation results show the breach expanding to a significant scale within three hours, consistent with actual dam failure scenarios.

Vulnerability Assessment Model: Utilizes the eXtreme Gradient Boosting (XGBoost) algorithm to integrate population, Gross Domestic Product (GDP), land use, and other data to generate regional risk heatmaps. Taking a specific urban area as an example, this model evaluates regional vulnerability indices by analyzing factors such as population density, GDP distribution, and land use types (e.g., residential, commercial, industrial zones). Results indicate higher vulnerability indices in densely populated residential and commercial zones with elevated GDP, aligning with real-world observations [17].

3) Agent Layer

Perception agents perform data cleaning and anomaly detection. Perception agents use the Isolation Forest algorithm to detect sensor failures. For example, sensor failure in the data collection of a hydrological station results in data anomalies. Therefore, the perception agent uses Isolation Forest to analyze data distribution and find anomalies. Meanwhile, the perception agent uses Kalman filtering to smooth noisy data (such as random noise in the output of meteorological stations), correcting and smoothing the data.

Cognitive agents perform risk inference and decision optimization. Risk propagation models are built based on Dynamic Bayesian Networks (DBN). For example, when evaluating flood risk caused by heavy rainfall in a region, the DBN model considers dynamic relationships among multiple factors (such as precipitation, topography, and drainage) and continuously updates the state of these factors to infer the possible probability and impact scope of flooding. Meanwhile, reinforcement learning (such as the Deep Q-Network (DQN) algorithm) is used to adjust warning thresholds dynamically. In the simulated environment, the DQN algorithm learns the optimal strategy for warning thresholds through interaction with the environment (such as historical heavy rainfall). Experiment results show that this method improves warning accuracy by 20% and reduces the false alarm rate by 30% [18]. The proxy operation data is shown in Table 1:

TABLE 1. Agent Operational Data

Metric Category	Perception Agents	Cognitive Agents	Action Agents	System Overall
Data Processing Volume	12.5 TB/day	8.2 TB/day	3.7 TB/day	24.4 TB/day
Processing Accuracy	95.2%	89.7%	93.4%	92.8%
Decision Quality Score	88/100	94/100	91/100	91/100

Action agents execute contingency plans and manage resource allocation, employing an auction mechanism for dynamic distribution of emergency supplies. For instance, during a flood disaster response, varying regional demands for supplies (e.g., life jackets, sandbags) are addressed through competitive bidding among regions, with allocations determined by bid amounts and urgency levels.

4) Application Layer

Provides four forecasting functional modules and a visual interactive interface. The forecasting module integrates multi-model predictions, generating confidence intervals through uncertainty quantification (e.g., Monte Carlo simulations). Taking a rainfall forecast as an example, the module synthesizes predictions from multiple models—including ST-ConvNet and physically constrained deep learning models—to generate 1,000 prediction samples via Monte Carlo simulation. This yields a 95% confidence interval, allowing users to view predicted rainfall ranges and confidence levels on the interface [19]. The effectiveness of multi-channel warning dissemination is shown in Figure 5:

The Early Warning Module overlays risk levels on Geographic Information System (GIS) maps and supports multi-channel information dissemination. Upon issuing warnings, the system color-codes GIS maps to indicate risk tiers: blue for low risk, yellow for moderate risk, orange for high risk,

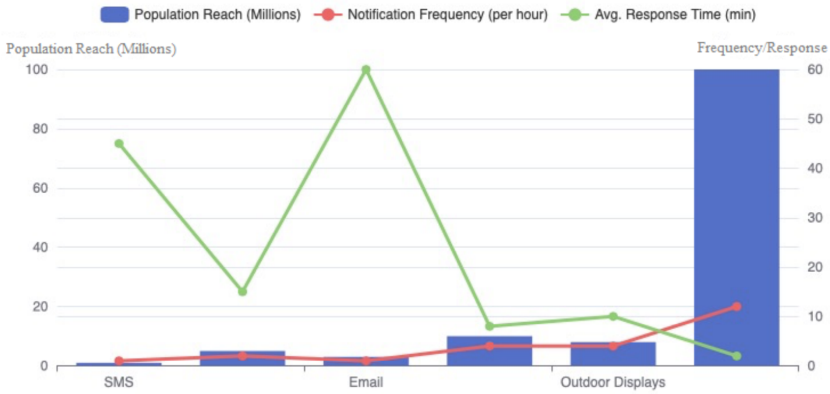

FIGURE 5. Multi-Channel Warning Dissemination Effectiveness

and red for extreme risk. Simultaneously, alerts are pushed via SMS, mobile apps, email, community broadcasts, outdoor displays, and other channels to relevant personnel.

B. KEY TECHNICAL COMPONENTS

1) Dynamic Knowledge Graph Construction

Based on Named Entity Recognition (NER) and Relation Extraction (RE) technology, flood control entities (e.g., heavy rain, dikes) and their relations (e.g., "caused by", "located in") are extracted from structured data sources (e.g., database tables) and unstructured data sources (e.g., news).

For news reports about the 2020 flood disaster in a certain area, NER technology can extract entities like heavy rain, flood, and affected villages, and RE technology can extract relations like heavy rain causing flooding and floods submerging affected villages.

A dynamic knowledge graph containing more than 120,000 entities and 600,000 relations was constructed. The knowledge graph can be updated and reasoned semantically in real time [20]. When new disaster information appears, entities and relations in the knowledge graph can be updated in real time. Users can ask semantic questions like, "Which levees were affected by a certain heavy rainfall event?" and the system can return correct information from the knowledge graph in real time. The growth of knowledge graph entities and relationships is shown in Figure 6:

2) Reinforcement Learning-Driven Adaptive Adjustment of Early Warning Thresholds

By using the Q-Learning algorithm, early warning accuracy, false alarm rate, and false negative rate are used as reward functions. The optimal threshold strategy is obtained by interacting with the environment (historical heavy rainfall events).

Taking the past five years' heavy rainfall events in a certain region as the simulation environment, the system carries out simulated warnings at different thresholds. The accuracy, false alarm, and false negative rate are calculated according to the warning results based on the research. The results are input into the Q-Learning algorithm as reward functions, and the optimal warning threshold strategy is found through several cycles of simulation learning. Experimental results show that the method can increase warning accuracy by 22% and decrease the false alarm rate by 28% [21]. The optimization of the learning reward function is shown in Figure 7:

3) Agent-Based Simulation and Deduction Engine

Million-level grids are computed in parallel by multithreading technology. Level of Detail (LOD) technology dynamically adjusts model precision according to the distance from the viewpoint. For the simulation of Beijing's July 21, 2012, torrential rain, as shown in Figure 8, the system divides the region to be simulated into million-level grids. By using multithreading technology, the grids are computed in parallel, which improves the efficiency of parallel computation [22]. Meanwhile, LOD adjusts the detail of the model according to the user's view distance. When users are farther away from the simulated scene, models with lower precision are used; when users approach the simulated scene, models with higher precision are used. In addition, the system uses Compute Unified Device Architecture (CUDA) to accelerate the rendering of real-time simulation images. Users can observe the process of flood progression in real-time simulation at a frame rate of 35 FPS. In this torrential rain simulation, the system's predicted water depth showed less than 12% error compared with actual observations.

III. IMPLEMENTATION OF FORECAST FUNCTION MODULES

A. ST-CONVNET FOR PRECIPITATION FORECASTING

Addressing the unique spatiotemporal characteristics of radar echo data, the intelligent forecasting subsystem employs a ST-ConvNet for precipitation prediction. This network incorporates a specially designed threedimensional convolutional kernel to simultaneously capture spatial and temporal features. In the spatial dimension, the convolutional kernel covers a 5-kilometer by 5-kilometer grid area, enabling detailed analysis of radar echo intensity distributions within localized regions. This accurately identifies key spatial features, such as the location of intense precipitation cores, the morphological structure of precipitation clouds,

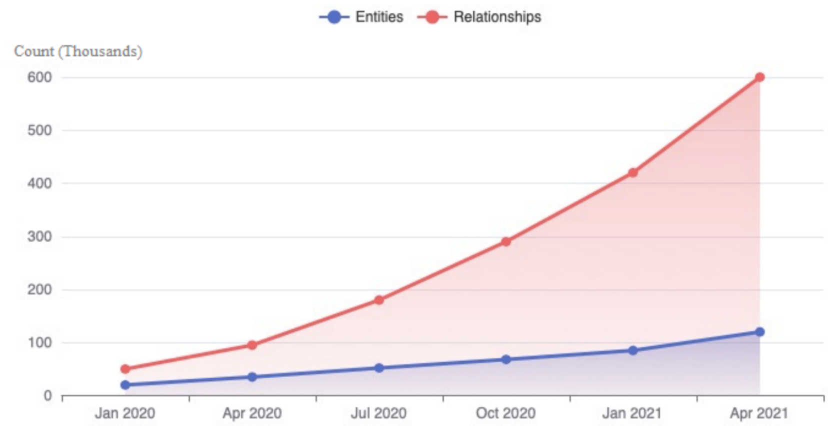

FIGURE 6. Knowledge Graph Entity & Relationship Growth

FIGURE 7. Q-Learning Reward Function Optimization

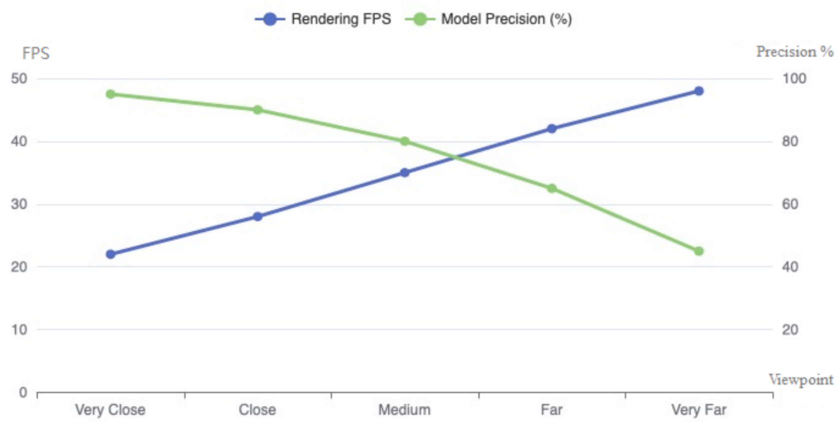

FIGURE 8. LOD Model Precision vs. Rendering Performance

and their coverage extent. In the temporal dimension, the network correlates six consecutive radar echo sequences to track rainfall's dynamic evolution over time, including precipitation cloud movement paths, velocity, and intensity trends. To overcome the common vanishing gradient problem in deep network training, the model incorporates a residual connection mechanism [23]. This mechanism directly passes outputs from preceding layers to subsequent ones, effectively ensuring gradient stability and efficiency during backpropagation. This guarantees model training convergence and performance. The performance comparison between ST convolutional neural network and traditional Z-R method is shown in Figure 9:

In the practical application scenario of rainfall forecasting in the Pearl River Basin, the model achieved a critical success index of 0.75 in one-hour rainfall prediction applica-

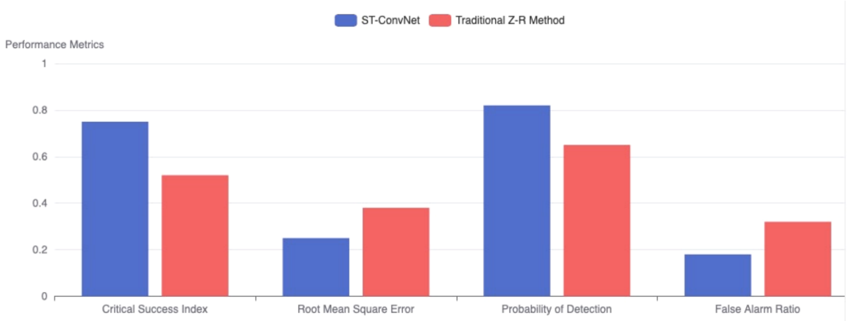

FIGURE 9. ST-ConvNet vs. Traditional Z-R Method Performance

tions. The heavy precipitation estimation based on the deep learning method can achieve root mean square error (RMSE) less than 2.5 mm/h in the validation results of the AI identification and forecasting benchmark dataset for short-term heavy precipitation developed by the National Meteorological Information Center. Compared with the traditional physicalstatistical method based on the Z-R relationship, the performance of this method is improved by more than 30%, and the accuracy is significantly improved.

The performance of the 3DuA-Net, a radar echo extrapolation model for short-term precipitation proposed by the research team at Lanzhou University of Technology, which adopts 3D convolutional and dual-end attention methods, has been improved in multiple performance indicators [24]. The critical success index is improved by an average of 7.74 percentage points, and the structural similarity index is improved by 8.86 percentage points. These technologies improve the accuracy and reliability of technical support for regional rainfall forecasting.

B. PHYSICALLY CONSTRAINED DEEP LEARNING FLOOD EVOLUTION MODELS

As a core component of the system, the intelligent forecasting subsystem incorporates physically meaningful constraints, such as Manning's formula and the continuity equation, into the neural network framework. The Lagrange multiplier method effectively calibrates the relationship between the data-driven method and the physical mechanism. In a typical application scenario of simulating the failure of the Xiaolangdi Reservoir on the Yellow River, physical equations were embedded as hard constraints in the loss function during model training. The Lagrange multiplier method adjusted the network parameters effectively by changing the multiplier, which promoted the model to strictly follow the physical mechanism of fluid motion and data patterns. Simulation results show that the model can predict the arrival time of the flood with an error within six minutes compared with observed data, and the coincidence rate of inundation area prediction results is 93% [25].

As shown in the flood forecasting method implemented by researchers from Wuhan University, coupling spatiotemporal feature contribution and deep learning, this method has obvious advantages in the application of the Jianxi River Basin. Combining CNN, bidirectional long short-term memory (BiLSTM), and attention mechanisms, the method builds a CNN-BiLSTM-Attention hybrid model. The accuracy rate for peak flood, total flood volume, and peak occurrence time all reach Class A. Compared with the traditional LSTM model, the method can effectively extend the lead time of the forecast and improve the accuracy of flood prediction. Especially in forecasting flood processes under complex spatial and temporal conditions, the method has clear advantages. Graph neural network-based basin flood forecasting technology improves the accuracy of prediction by building a graph-guided spatiotemporal correlation model. Multiple LSTMs encode the temporal correlation feature of historical attributes in historical monitoring points, while graph convolutional neural networks (GCNs) extract geospatial dependency between monitoring points. The simulation accuracy of Xiaolangdi Reservoir dam break is shown in Figure 10:

C. UNCERTAINTY QUANTIFICATION METHODS

The intelligent forecasting subsystem uses the Monte Carlo dropout method to quantify uncertainty. When training the model, fifty percent of neurons are randomly dropped. Ultimately, several groups of prediction samples are obtained by randomization; one thousand groups of prediction results are generated when applying the model, covering the prediction distribution range. Then, the kernel density estimation method is used to construct the probability distribution and intuitively display the range and characteristics of the prediction results' probability distribution.

Practical applications have shown that, compared with the European Centre for Medium-Range Weather Forecasts (ECMWF) and National Centers for Environmental Prediction (NCEP) ensemble forecast systems, forecast accuracy gradually decreases with lead time. The heavy precipitation forecast based on the Bayesian model averaging method is accurate at all lead times, and rainfall correction based on Bayesian model averaging provides more reliable probabilistic information for flood control decision-making.

The validation experiment based on the benchmark dataset of the National Meteorological Information Center shows that the RMSE of heavy precipitation can be reduced to less than 2.5 mm/h compared with traditional methods

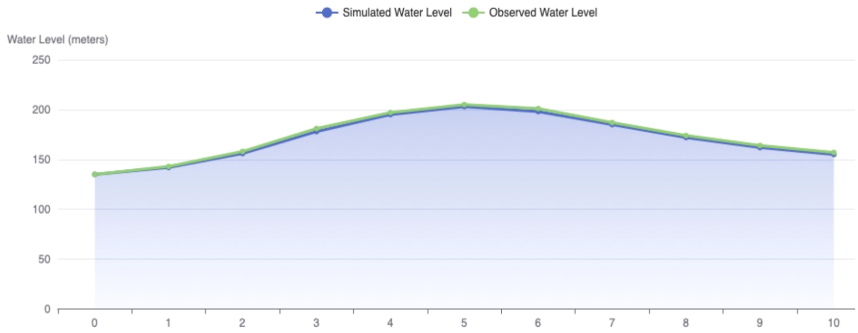

FIGURE 10. Xiaolangdi Reservoir Dam Break Simulation Accuracy

based on the Z-R relationship, and the accuracy is greatly improved. The uncertainty quantification results in the paper can provide more reliable decision-making support for flood control engineering planning and emergency response strategies, based on more reliable probabilistic information. The accuracy of integrated prediction and delivery cycle are shown in Figure 11:

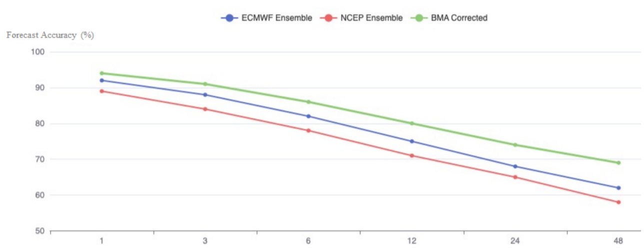

FIGURE 11. Ensemble Forecast Accuracy vs. Lead Time

IV. CONCLUSION

The research successfully outlines the construction of a flood control Four Predictions (forecasting, early warning, simulation, contingency planning) application system based on a Large-Model Intelligent Agent Application Architecture, which offers robust defense against water-related disruptions.

The architecture is designed in layers—Data, Model, Agent, and Application—to achieve high cohesion and low coupling, enhancing maintainability and scalability. The system incorporates a multimodal flood control data lake (Hadoop, HBase, and Spark) and a dynamic knowledge graph for semantic querying. The Model Layer includes a domain-adapted Large Language Model and specialized micro-model clusters for key predictions. Critically, the Agent Layer integrates Perception Agents (using Isolation Forest and Kalman filtering for data quality), Cognitive Agents (performing risk inference via DBN and adaptive warning threshold adjustment via DQN, which improves warning accuracy by 20% and reduces false alarms by 30%), and Action Agents (managing resource allocation via an auction mechanism and planning optimal evacuation routes). Functionally, the system employs advanced techniques, such as ST-ConvNet for intelligent forecasting, a multi-risk factor coupling model for dynamic early warning dissemination, a GPU-accelerated 3D engine with adaptive mesh for digital simulation, and a hybrid case-based/rule-based reasoning with multi-objective optimization (NSGA-II) for intelligent contingency planning. The system's performance validation, including simulation of historical heavy rainfall events and comparison with traditional models, confirms its ability to provide efficient and precise technical support for flood control operations.

AUTHOR CONTRIBUTIONS

Conceptualization – LI liangliang and DU wen; methodology – LI liangliang and DU wen; investigation – LI liangliang and DU wen; writing-original draft preparation – LI liangliang and DU wen. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] J. Lv, J. Hou, T. Wang, W. Zhou, D. Li, Y. Tong, et al., "The research on narrow-valley city flood control mode based on hydrodynamic-hydrological coupling model," Journal of Hydrology, vol. 640, no. 0, p. 13, 2024, doi: 10.1016/j.jhydrol.2024.131584.

[2] L. Zhou, L. Kang, S. Hou, and J. Guo, "Research on Flood Risk Control Methods and Reservoir Flood Control Operation Oriented towards Floodwater Utilization," Water, vol. 16, no. 1, p. 15, 2024, doi: 10.3390/w16010043.

[3] Y. Wu, S. Zheng, Q. Liu, A. Dong, and Q. Li, "Structural and empirical knowledge driven multi-objective evolutionary algorithm for urban drainage system design," Expert Systems with Applications, vol. 249, p. 123461, 2024, doi: 10.1016/j.eswa.2024.123461.

[4] Y. Zhang, S. Wang, W. Ge, Z. Li, H. Li, W. Du, et al., "A Calculation Method for Flood Control Benefits of River Treatment Projects Considering the Uncertainty of Flood Peak and Flood Frequency," Water Resources Management, vol. 39, no. 8, 2025, doi: 10.1007/s11269-025-04153-2.

[5] R. Itsukushima, K. Ohtsuki, and T. Sato, "Significance of land use as a flood control measure: Unveiling the historical and contemporary strategies in the unique case of Kofu basin alluvial fan, Japan," International Journal of Disaster Risk Reduction, vol. 109, no. c, p. 104578, 2024, doi: 10.1016/j.ijdrr.2024.104578.

[6] G. Al-Rawas, M. R. Nikoo, and M. Al-Wardy, “A review on the prevention and control of flash flood hazards on a global scale: Early warning systems, vulnerability assessment, environmental, and public health burden,” International Journal of Disaster Risk Reduction, vol. 115, p. 105024, 2024, doi: 10.1016/j.ijdrr.2024.105024.

[7] Z. Q, T. Yan, G. Siyuan, J. Zhou, X. He, W. He, et al., “Stability behavior of the Lanxi ancient flood control levee after reinforcement with upside-down hanging wells and grouting curtain,” Journal of Mountain Science, vol. 21, no. 1, pp. 84-99, 2024, doi: 10.1007/s11629-023-8239-7.

[8] M. Tanhapour, J. Soltani, H. Shakibian, B. Malekmohammadi, K. Hlavcova, and S. Kohnova, “Development of a Multi-objective Optimal Operation Model of a Dam using Meteorological Ensemble Forecasts for Flood Control,” Water Resources Management, vol. 39, no. 6, pp. 2743-2761, 2025, doi: 10.1007/s11269-024-04089-z.

[9] C. Xu, P. Zhong, F. Zhu, L. Li, Q. Lu, and L. Yang, “Stochastic multi-criteria decision making framework based on SMAA-VIKOR for reservoir flood control operation,” Hydrological Sciences Journal, vol. 68, pp. 886-901, 2023, doi: 10.1080/02626667.2022.2154161.

[10] F. Chai, S. Liu, Y. Sun, and F. Huo, “Research and application of the flood simulation and operation model in Beijing,” E3S Web of Conferences, p. 16504018, 2020, doi: 10.1051/e3sconf/202016504018.

[11] Y. Peng, X. Yu, L. Yao, S. Luo, and Z. Zhang, “Including dynamic capacity impact in an improved model for optimal flood control operation in river-type reservoirs,” Journal of Hydrology, vol. 645, p. 132163, 2024, doi: 10.1016/j.jhydrol.2024.132163.

[12] L. Sun, J. Xia, D. She, W. Ding, J. Jiang, B. Liu, et al., “A predictive fuzzy logic and rule-based control approach for practical real-time operation of urban stormwater storage system,” Water Research, vol. 266, p. 122437, 2024, doi: 10.1016/j.watres.2024.122437.

[13] L. Yao, Y. Peng, X. Yu, Z. Zhang, and S. Luo, “Improving Flood Control Optimal Operation of River-Type Cascade Reservoirs through Coupling with 1D Hydrodynamic Model,” Water Resources Management, vol. 39, no. 7, pp. 3443-3466, 2025, doi: 10.1007/s11269-025-04116-7.

[14] L. Gao, Y. Gao, Y. Liu, and M. Wu, “Assessment of flood risk under Polder-Type flood control measure using improved projection pursuit model,” Ecological Indicators, vol. 170, p. 113038, 2025, doi: 10.1016/j.ecolind.2024.113038.

[15] J. Liu, Z. Yang, Y. Liu, M. Li, and C. Zhou, “Can the regulation of Golden Inland Waterways meet the needs of navigation, flood control, and ecology? A model-based case study,” Ecological Engineering, vol. 196, p. 106998, 2023, doi: 10.1016/j.ecoleng.2023.106998.

[16] J. Li, Z. Wang, and T. Zhang, “Flood simulation using the hydrological model and the hydrological–hydrodynamic coupling model in a small watershed in semi-arid and sub-humid region, North China,” Journal of water and climate change, vol. 14, no. 9/10, pp. 3496-3516, 2023, doi: 10.2166/wcc.2023.161.

[17] H. Jin, X. Chen, R. Zhong, M. Liu, and C. Ye, “Construction of precipitation index based on ensemble forecast and heavy precipitation forecast in the Hanjiang River Basin, China,” Atmospheric Research, vol. 292, p. 106701, 2023, doi: 10.1016/j.atmosres.2023.106701.

[18] Y. Gao, M. Wu, and G. Z. Zhang, “Analyzing the impact of polder-type flood control pattern on river system’s regulation and storage capacity under urbanization,” Journal of the American Water Resources Association, vol. 59, no. 6, pp. 1219-1245, 2023, doi: 10.1111/1752-1688.13128.

[19] L. Sinha and S. M. Narulkar, “Optimal Operation of Multi-reservoir System Utilizing DEA, AIDE Algorithm and Flood Control Assessment by MCDM Approach,” Water Resources Management, vol. 39, no. 4, 2025, doi: 10.1007/s11269-024-04046-w.

[20] E. Lee, S. Yi, J. Ji, J. Hong, S. Lee, J. Yoon, et al., “Development of a reservoir operation model determining the pre-release strategy for the flood events Open Access,” Journal of Hydroinformatics, vol. 27, no. 4, pp. 23, 2025, doi: 10.2166/hydro.2025.262.

[21] L. Li, L. Chen, S. Chen, Y. Zhang, Y. Xu, X. Zhi, et al., “The cumulative effects of cascade reservoirs control nitrogen and phosphorus flux: Base on biogeochemical processes,” Water Research, vol. 252, p. 121177, 2024, doi: 10.1016/j.watres.2024.121177.

[22] A. M. Enríquez-Hidalgo, A. Vargas-Luna, and A. Torres, “Evaluation of decision-support tools for coastal flood and erosion control: A multicriteria perspective,” Journal of Environmental Management, vol. 373, p. 123924, 2025, doi: 10.1016/j.jenvman.2024.123924.

[23] Y. Xu, K. Liu, Y. Ma, Q. Wang, and C. Gao, “Potential Impact of Flood Control Projects on Hydrological Processes in the Coastal Regions of the Taihu Basin, China,” Journal of the American Water Resources Association, vol. 61, no. 1, 2025, doi: 10.1111/1752-1688.70003.

[24] K. E. Schilling, E. Anderson, and M. T. T. C. Streeter, “Long-term nitrate-nitrogen reductions in a large flood control reservoir,” Journal of Hydrology, vol. 620, p. 129533, 2023, doi: 10.1016/j.jhydrol.2023.129533.

[25] A. Ogihara, M. Maeda, G. Katou, K. Matsumoto, K. Miyaoka, and D. Hirayama, “Research Reports the Practice of Flood Control Education Materials Based on the Kumozu River Basin Case from Mie Prefecture,” Environmental Education, vol. 34, no. 2, 2024, doi: 10.5647/jsoee.2420.