

Research on Fault Warning System for Hydroelectric Facilities Based on Cloud Computing and GAN

Qiuyue Zhong^{1,a} and Jiangkuan Zhao^{2,b*}

¹YCIH (Yunnan Construction and Investment Holding Group) No.1 Water Resources and Hydropower Construction Co., Ltd., Kunming 650501, Yunnan, China

²School of Water Conservancy and Hydroelectric Power, Hebei University of Engineering, Handan 056038, Hebei, China

Because hydropower facilities are one of the country's important energy infrastructures, the safety of their operations is crucial to social and economic development, the security of lives and property, and the protection of the ecological environment. Traditional safety management methods usually rely on manual monitoring and a single system, resulting in long response times, difficulty in information integration, and difficulty in timely capturing potential risks. This approach often has information islands and data fragmentation problems, and cannot achieve comprehensive, real-time monitoring and early warning of facilities, increasing the probability of safety hazards. This study designed and implemented a hydropower facility safety management platform based on cloud computing. A distributed storage structure was used to store collected data and separately build a Kafka transmission architecture to connect the various departments of the platform and speed up the information transmission within the platform. It used generative adversarial network (GAN) technology to identify faults and issue early warnings through the transmission architecture. The experimental results show that during the data transmission process, the average transmission accuracy is 94.6%, the average transmission completeness is 98.2%, and the average transmission synchronization is 97.8%. At the same time, when the platform load is low, the response time can be as short as 0.89 seconds. The evaluation of the platform's early warning capability showed that the AUC value of fault identification is 0.92. This study analyzed key indicators such as throughput, latency, and CPU usage of the platform, and demonstrated that the proposed system can still maintain stable operation in high load and complex environments. Through the application of cloud computing and distributed technology, the proposed hydropower facility safety management platform significantly improves warning capabilities and system stability, ensuring efficient operation of the platform in complex environments and providing intelligent and real-time solutions for the safety management of hydropower facilities

Keywords: Hydropower facility; cloud computing; security management platform; generative adversarial networks; early warning capability

1. INTRODUCTION

Hydropower facilities[1–3] play a key role in ensuring national energy security and promoting social and economic development. With the continuous expansion of the scale of hydropower projects and the increase in the complexity of their operations, the safety management of hydropower facilities has become increasingly important. The safety of facilities is directly related to the stability of power supply,

the safety of people's lives and property, and the protection of the ecological environment. At present, the traditional hydropower safety management method relies on manual operation, single monitoring and local risk assessment, facing the challenges of data silos, information delays and decision lags, and it is difficult to meet the requirements of modern management for real-time and intelligent management. In view of these problems, the research and design of the hydropower facility safety management platform based on cloud computing technology [4–6] has important practical significance and can provide technical support for improving

*Corresponding author. Email: ^a357385572@qq.com, ^bshuisushanxing@163.com

the safety of hydropower facilities, reducing operation risks and improving management efficiency.

In past studies, many scholars have proposed various solutions for the safety management of hydropower facilities. Shaktawat and Shelly believed that the construction phase of a hydropower project was the most critical phase associated with uncertainty, and proposed the use of fuzzy hybrid artificial neural networks and genetic algorithms to analyze the risks of hydropower projects [7]. Vucetic et al. analyzed the risk management of hydropower plants from the perspective of fracture science, using a risk matrix to assess the risk level in a descriptive manner based on probability [8]. Obialor et al. studied the process of reservoir siltation, determined the impact of siltation on reservoir facilities and introduced methods to solve the siltation problem [9]. Nasrat et al. analyzed dam safety and dam accidents from a "human" perspective and made relevant recommendations [10]. Rubin et al. evaluated whether sufficient high-quality engineering maintenance work can be carried out on hydropower stations and put forward recommendations to improve the efficiency of newly built and modernized hydropower facilities [11]. Xiang et al. believed that dam hydraulic structures are facing more and more problems, and further reviewed the technology and application of safety inspection, repair and reinforcement equipment at a 100-meter level in the deep water environment of reservoir dams [12]. Zhu et al. proposed a comprehensive dam safety assessment method based on digital twins, aiming to improve the safety of dams [13]. Kuban and Pretelli studied hydroelectric facilities in the last half century from the perspective of industrial heritage, aiming to provide further research models for the safety management of hydroelectric power plants [14]. Based on existing research and environmental and climate data, Zhang and Shang conducted an in-depth analysis of the way that environmental and climate changes affect dam safety, and proposed relevant measures [15]. Alvi et al. analyzed the causes of dam failure and believed that the cause was "human factors" [16]. Stojkovic et al. proposed a new framework that uses dynamic elasticity to assess the impact of dangerous events on hydropower systems [17]. Overall, although previous research has improved the safety management level of hydropower facilities to a certain extent, it still has shortcomings in terms of data integration, intelligent analysis, real-time response, etc., and cannot effectively respond to complex and changing safety management needs.

In order to address the shortcomings of traditional safety management methods, the safety management platform of hydropower facilities based on cloud computing has gradually attracted attention in recent years. Cloud computing technology has powerful data storage, processing and computing capabilities, and can integrate massive amounts of information from various monitoring devices, sensors and historical data to achieve real-time data collection, storage and analysis [18]. Sang et al. proposed a method combining extended cloud model with extended hierarchical method to solve the uncertainty problem of overall safety trend evaluation of dams and selection of safety trend warning indicators [19]. Fera and Spandonidis developed a special real-time framework to identify the damage of hydropower

station infrastructure, especially the corrosion of generator conductors [20]. Krishna and Saini used an adaptive neuro-fuzzy interface system to perform online monitoring of turbines and other equipment [21]. Guo explored the use of Huawei Cloud Platform to optimize the reservoir scheduling system through cloud services to cope with complex water resource management needs and extreme climate events [22]. Solanki et al. studied multi-criteria decision-making methods to identify and evaluate factors for implementing IoT and cloud computing in hydropower facilities, but these methods have not yet been used [23]. Markovic-Petrovic et al. proposed a new safety risk assessment method for monitoring and data acquisition and evaluated it in a case study of an actual run-of-the-river hydropower station [24]. Afridi et al. introduced artificial intelligence and cloud computing to develop a predictive maintenance system to identify faults before they occur, thereby minimizing equipment downtime and improving work efficiency [25]. Zhang et al. developed a sensor detection system that can evaluate the structural state of the sensor to detect whether there is an abnormality and help with early warning [26]. Li et al. studied and summarized whether new information technologies such as big data cloud computing can help the intelligent construction of dams, and analyzed the relevant technical characteristics and related management models [27]. However, although these studies have achieved the improvement of data integration and analysis capabilities, there are still problems such as low platform integration and imperfect real-time response mechanisms, making it difficult to achieve seamless collaboration and instant processing among multiple systems.

In order to achieve all-round and real-time monitoring and early warning of hydropower facilities, this study designed a transmission architecture using Kafka [28–30] in addition to the basic architecture of the platform. The function of this architecture is to transmit data between the various departments of the platform to ensure data security and efficiency. At the same time, the generative neural network technology [31–33] was used to generate fault data and perform cosine similarity [34–35] analysis on real data, and early warning is issued based on similar results. The analysis of experimental data showed that the platform designed in this study achieved an average transmission accuracy, average transmission completeness, and average transmission synchronization of 94.6%, 98.2%, and 97.8% respectively for information transmission. When the platform load is small, it can achieve a minimum response time of 0.89 seconds. Even if the load is high, it can respond within 7 seconds. At the same time, the platform's early warning capability was analyzed. The precision-recall curve showed that the AUC value reached 0.92, indicating that the platform had a good ability to identify faults and early warning. The relationship between the platform's throughput, latency and CPU usage was analyzed, and the results showed that the platform had good stability. In summary, the platform researched and designed in this paper not only performed well in terms of data transmission and response time, but also was able to cope effectively with challenges under high load through accurate fault identification and early warning mechanisms. This provides a reliable guarantee for the safe operation of

Table 1 Detection equipment and sensor information table.

Device Type	Measurement parameters	Application Scenario	Accuracy requirements	Communication Method	Environmental adaptability
Water level sensor	Water level	Monitoring of water bodies such as reservoirs, dams, and spillways	$\pm 1\text{cm}$	RS485, Modbus, Wi-Fi	Waterproof, anti-corrosion, high temperature resistant
Flow Meter	Water flow, flow rate	Water pump, water pipe, generator set water inlet flow monitoring	$\pm 0.5\%$	Modbus, LoRa, NB-IoT	Waterproof, anti-corrosion, vibration-resistant, high-pressure-resistant
Pressure Sensors	Water pressure, pipe pressure	Pressure monitoring of pipelines, dams, and equipment at hydropower facilities	$\pm 0.2\%$	Modbus, RS485, Zigbee	High temperature, high pressure and corrosion resistance
Temperature and humidity sensor	Temperature, humidity	Equipment room and generator set environmental monitoring	$\pm 1^\circ\text{C}$, $\pm 2\%\text{RH}$	Wi-Fi, Zigbee, LoRa	Waterproof, anti-corrosion, strong adaptability to high and low temperatures
Vibration Sensor	Vibration acceleration, frequency	Generators, pump stations, rotors and other mechanical equipment monitoring	$\pm 0.01\text{g}$	RS485, Modbus, Wi-Fi	Dustproof, waterproof and vibration-resistant
Current/voltage sensor	Current, voltage, power, power factor	Power equipment, generator sets, power transmission line monitoring	$\pm 0.5\%$	Modbus, CAN, RS485	Waterproof, dustproof, anti-electromagnetic interference

hydropower facilities, significantly improves the intelligence level and system stability of the platform, and provides valuable reference for future applications in similar fields.

2. PLATFORM DESIGN AND IMPLEMENTATION

2.1 Data Acquisition and Transmission Module

The data acquisition and transmission module is one of the core components of the security management platform designed in this study. It is responsible for obtaining real-time data from various sensors and monitoring equipment in hydropower facilities, and transmitting the data to the cloud for processing and analysis through efficient communication technology. This module ensures the real-time monitoring of various parameters of hydropower facilities (such as water level, flow, pressure, equipment operating status, etc.), providing a data basis for subsequent data processing, analysis and early warning. The detection equipment and sensor information for data collection are shown in Table 1.

Table 1 lists the detection equipment and sensors used in this study: water level sensor, flow meter, pressure sensor, temperature and humidity sensor, etc., and their corresponding information such as parameter information, accuracy, connection method, etc. Understanding this information can help the research proceed more smoothly, and at the same time help ensure that various important data are accurate and collected in real-time in the cloud computing platform

to ensure the safe operation and management of hydropower facilities [37].

Due to the large number of sensors and the large data flow, this study used Apache Kafka as the data transmission architecture. As a distributed stream processing platform, Kafka can provide a high-throughput, low-latency and high-reliability data transmission mechanism. The core components of Kafka include producers that are responsible for sending data to the Kafka cluster. Topics are logical channels for storing data streams. Data in the Kafka architecture is stored by partitions to ensure high concurrency and high throughput. The Kafka cluster consists of multiple proxy servers, each of which is responsible for processing and storing data from producers. Consumers are responsible for subscribing to data from Kafka, that is, extracting data.

Kafka supports high-throughput data transmission. The throughput is calculated with:

$$T_{total} = \frac{M_{total}}{t_{total}} \quad (1)$$

where M_{total} represents the total amount of data transmitted, t_{total} represents the total transmission time, and T_{total} represents the throughput.

In order to ensure the security and efficiency of data transmission, this study used the AES symmetric encryption algorithm to encrypt the transmission process. AES uses a shared key for encryption and decryption to ensure the efficiency of the process. The encryption and decryption process is as follows.

Encryption process, for a plaintext block M , the symmetric key K is used for encryption:

$$C = AES(M, K) \quad (2)$$

Decryption process, using the same key K for decryption:

$$M = AES^{-1}(C, K) \quad (3)$$

2.2 Cloud Data Storage and Processing Architecture

In order to cope with large-scale and high-concurrency storage requirements, the platform proposed in this paper adopted a distributed storage structure to store data. In distributed storage, data can be divided into multiple data blocks, and these databases can be stored on different nodes respectively. The system can automatically perform redundant backup of the data to ensure data reliability.

Assuming there are N storage nodes, the allocation algorithm when data blocks are allocated to these nodes is:

$$Node_i = (\text{hash}(\text{DataBlock}) \bmod N) \quad (4)$$

where $\text{hash}(\text{DataBlock})$ is the value obtained by hashing the data block content, and N is the total number of storage nodes. This algorithm ensures data load balancing and can prevent the problem of node failure or malfunction.

Assuming that each data block has k backups, the total amount of data R after redundant backup is:

$$R = k \cdot \text{Node} \quad (5)$$

In order to make the subsequent intelligent analysis model more effective, the cloud data processing architecture of the platform studied in this paper preprocesses the data on the basis of data storage, and performs missing value and outlier processing and feature extraction on the collected and stored data.

This study applied the mean filling method to fill missing values. Assuming that the missing value of a sensor data is $x_i = \text{null}$, the formula for filling missing values is:

$$x_i = \frac{1}{N} \sum_{i=1}^N x_i \quad (6)$$

where N is the total number of historical data of the sensor, and $\sum_{i=1}^N x_i$ is the sum of historical data.

This study uses the Z-score method to detect outliers. If the Z-score value of a point exceeds the set threshold of 3, the point is considered to be an outlier. The Z-score calculation formula is:

$$Z_i = \frac{x_i - \mu}{\sigma} \quad (7)$$

where x_i is the sample value of the current data point, μ is the mean of all samples, and σ is the standard deviation of the samples. If $|Z_i| > 3$, the data is considered abnormal and the abnormal value is marked for subsequent processing.

In the feature extraction stage, in addition to extracting some basic features such as maximum value, minimum value,

average value, and standard deviation, this study designed a health index based on this range because the data of a healthy system is within a certain range. In this stage, the health index is calculated to prepare for subsequent warnings.

Mean calculation:

$$\mu = \frac{1}{n} \sum_{i=1}^N x_i \quad (8)$$

Standard deviation calculation:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (9)$$

Health Index Calculation:

$$HI = \sum_{i=1}^n w_i \cdot f_i \quad (10)$$

where f_i is the value of the i -th feature, and w_i is the weight of the feature. By comparing the obtained health index with the set health index, it can be preliminarily determined that the system may have a fault.

2.3 Platform Management Interface

Figure 1 depicts the platform management interface designed in this study. The left and right sides of the interface display the flow, water level, power and today's warning information of the left and right banks of the hydropower building. The middle of the interface is the status display of related settings, with a total of three status displays. Green means that the facility is operating normally, and yellow means that the system has detected that the facility may have a fault and is further confirming it. Red means that the facility has been confirmed by the system to be faulty and needs further confirmation and maintenance by staff. Through this management interface, staff can clearly view the information for related facilities and relevant information of hydropower buildings in order to carry out further work.

3. INTELLIGENT ANALYSIS AND EARLY WARNING

3.1 Generative Adversarial Network Model

In order to carry as a powerful deep learning technology, generative adversarial networks have shown significant advantages in many fields due to their unique structure and training methods.

Unlike traditional supervised learning methods, generative adversarial networks do not require labeled data for training. Generative adversarial networks consist of two parts: a generator and a discriminator. The basic idea is to use the game training of the generator and the discriminator to enable the generator to generate sufficiently realistic "fake" data, while the discriminator can identify real data and generated data.

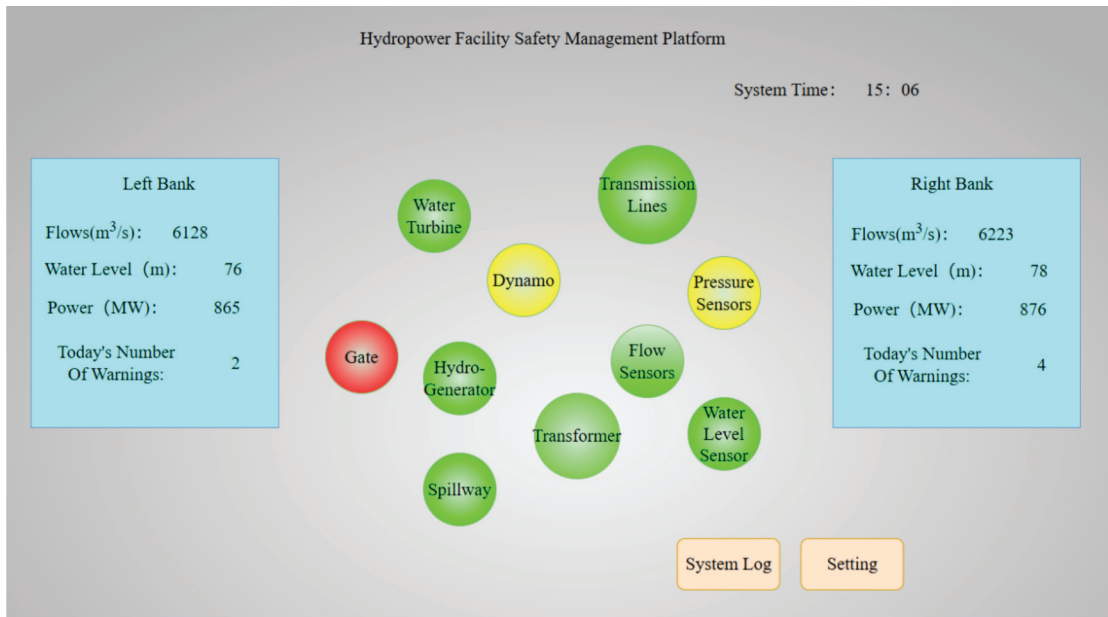


Figure 1 Platform management interface.

In this study, the task of the generator is to generate equipment operation data that simulates a real fault or a field from noise data, with the purpose of “deceiving” the discriminator so that it cannot distinguish between the generated data and the real data. The task of the discriminator is to determine whether the input data is real equipment data or generated data, and the output format is a probability value, which indicates the probability that the data is real data.

The generator hopes that the generated data can be considered real by the discriminator, so its goal is to minimize the discriminant error of its generated data, that is, the loss function of the generator:

$$\Gamma_G = -E_{z \sim p_z(z)}[\log D(G(z))] \quad (11)$$

where $p_z(z)$ is the noise distribution, $G(z)$ is the real data generated by the generator, and $D(G(z))$ is the judgment of the discriminator on the generated data, indicating the probability that the generated data is the real data.

The goal of the discriminator is to maximize its ability to correctly classify the real data and misclassify the generated data, so the loss function of the discriminator is:

$$\Gamma_D = -E_{x \sim p_{data}(x)}[\log D(x)] - E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (12)$$

where $D(x)$ represents the discriminator’s prediction of the real data, and $D(G(z))$ represents the discriminator’s prediction of the generated data. The goal is to make the discriminator judge the real data as 1 (true) and the generated data as 0 (false).

In addition to the loss function, the evaluation of the generative adversarial network also requires some criteria to measure the quality and diversity of the generated samples. This study used Fréchet Inception Distance (FID) to measure the distribution difference between the generated data and the real data.

The FID formula is:

$$FID(G) = \|\mu_r - \mu_g\| + \text{Tr}\left(\sum_r + \sum_g - 2\left(\sum_r \sum_g\right)^{1/2}\right) \quad (13)$$

where μ_r and $\sum_r$ are the mean and covariance matrix of the real data, μ_g and $\sum_g$ are the mean and covariance matrix of the generated data, and Tr is the trace of the covariance matrix.

After calculating the mean of the real data set X and the generated data set G using the mean formula, their covariance matrix can be calculated.

The covariance matrix $\sum_r$ of the real data is calculated with:

$$\sum_r = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_r)(x_i - \mu_r)^T \quad (14)$$

The covariance matrix $\sum_g$ of the generated data is calculated with:

$$\sum_g = \frac{1}{M-1} \sum_{j=1}^M (g_j - \mu_g)(g_j - \mu_g)^T \quad (15)$$

In adversarial training, the generator and the discriminator optimize each other alternately. The ultimate goal is to enable the generator to generate realistic data and the discriminator to be unable to distinguish the generated samples. For this overall goal, this study expresses it as a max-min problem, and the formula for which is expressed as:

$$\min_G \max_D E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (16)$$

where D tries to maximize the probability that the real data is judged as real while minimizing the probability that the generated sample is judged as false. G tries to make the discriminator “think” that the generated sample is real by minimizing the loss function.



Figure 2 Data transmission quality analysis diagram.

3.2 Abnormal Pattern Generation and Comparison

The generator can generate simulated equipment failure data based on random noise. In order to detect whether the equipment is faulty, the generated abnormal data model can be used as a benchmark for comparison. If it is similar or identical to the generated abnormal data benchmark, it may be faulty. For example, if a pump fault data generator is trained, when the vibration signal of the pump exceeds the normal range, the generator can simulate the corresponding fault data, which can reflect the classic performance of equipment failure, such as excessive vibration frequency, excessive temperature, abnormal pressure, etc.

This study compares the real-time monitoring transmission data in the Kafka transmission architecture with the abnormal pattern produced by the generator. The comparison method uses the cosine similarity calculation of the two data:

$$\text{cosine_similarity}(x_t, G(z)) = \frac{x_t \cdot G(z)}{\|x_t\| \cdot \|G(z)\|} \quad (17)$$

where x_t is the monitoring data at the current time t , and $G(z)$ is the fault data generated by the generator. If the calculated cosine similarity is high, it means that the real-time data may be similar to the fault data, indicating that a fault may occur.

3.3 Warning Trigger Mechanism

There are two ways to trigger the platform warning mechanism proposed in this paper. The first is the health coefficient calculated based on the extraction of data features mentioned above, and the second is to set the cosine similarity threshold ℓ :

$$\text{if } \text{cosine_similarity}(x_t, G(z)) > \ell \quad (18)$$

If the formula (18) is satisfied, the device is considered to be faulty.

When the system detects a device fault, the message is transmitted to the staff through the Kafka transmission architecture. In this way, the fault message can be transmitted in real time, thus breaking the information island and responding to the fault information at a high speed.

The evaluation indicator used in this study for the early warning mechanism of the platform shows the early warning accuracy, which indicates the ability of the platform to correctly identify faults or security threats. In order to

determine the accuracy of the early warning system, the following key concepts need to be taken into account: True Positives (TP), that is, the system correctly issues an alarm and there is actually a problem; False Positives (FP), that is, the system incorrectly issues an alarm, but there is actually no problem; True Negatives (TN), that is, the system correctly determines that there is no problem and there is actually no problem; False Negatives (FN), that is, the system fails to issue an alarm, but there is actually a problem. By calculating the number of times that each of these occur, the accuracy, precision, and recall of the early warning system are obtained to measure its overall performance.

Accuracy formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (19)$$

The accuracy rate can reflect the overall accuracy of the warning.

Precision formula:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (20)$$

Precision can indicate the number of events that are actually faulty among those that are judged as faulty by the platform.

Recall formula:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (21)$$

The recall rate indicates the proportion of all events with faults that the platform can correctly identify.

4. RESULT ANALYSIS

4.1 Data Transmission Quality Analysis

For the hydropower facility safety management platform proposed in this paper, the quality of data transmission directly affects the normal operation of the platform and the reliability of the final results. It is important to ensure that the data remains accurate, complete and consistent during the transmission process. The accuracy, completeness and consistency of multiple experimental data are presented below.

Figure 2 is a dashboard analysis diagram showing the data transmission quality. There are three dashboards representing

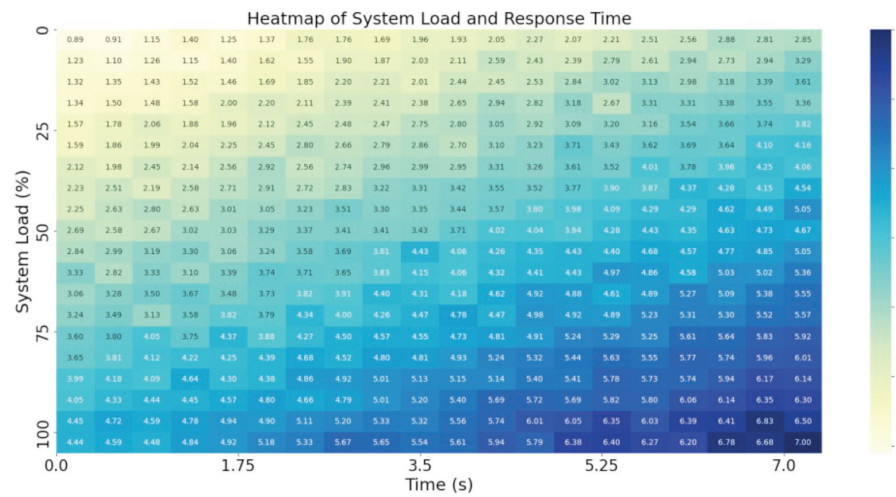


Figure 3 Response time heat map.

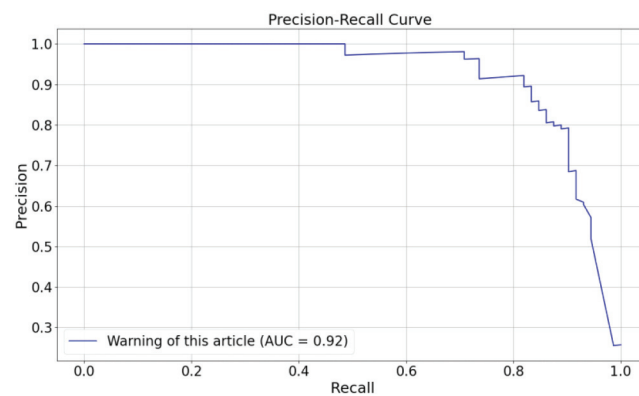


Figure 4 Precision-recall curve.

the probability analysis of three quality indicators in data transmission. The probability analysis is calculated by averaging the data collected from multiple experiments. The blue progress bar dashboard for analyzing the data accuracy during data transmission. The orange progress bar dashboard that analyzes the data integrity during data transmission shows that the average accuracy is 94.6%. The red progress bar dashboard that analyzes the data synchronization during data transmission shows that the average completeness is 98.2%, and the average synchronization is 97.8%. It can be seen from the numerical values that the hydropower facility safety management platform proposed in this paper has a high accuracy rate for data transmission, and the data is transmitted completely and maintains a high consistency with the original data.

4.2 Response Time Analysis

The response time of the platform is an important indicator of the platform's performance, especially the hydropower facility safety management platform proposed in this paper. The response time of the platform directly affects the monitoring and early warning mechanism, and also affects the decision-making of the staff. The response time of the platform is analyzed below.

Figure 3 shows the response time heat map of the hydropower facility safety management platform proposed in

this paper. The horizontal axis represents time in seconds, and the vertical axis represents the load of the platform, from 0–100%. The data in the heat map shows that when the platform load is small, the platform can respond in the shortest time: 0.89 seconds. However, when the platform is under heavy load, the response time is longer, but it can still respond within 7 seconds. This time lag occurs because the platform may have resource bottlenecks or delays under high load, which can directly affect the platform's response time. Therefore, when the platform has a slow response time, it is also a reminder to staff that certain load balancing measures may be needed to reduce delays and reduce the platform's response time.

4.3 Identification and Early Warning Capabilities

The hydropower facility safety management platform proposed in this paper provides early warnings of faulty hydropower facilities. Since the amount of normally operating equipment is far greater than the amount of faulty equipment, only a small number of facilities need to be warned. Therefore, the precision-recall curve is drawn to determine the identification and early warning capabilities of the platform.

Figure 4 shows the precision-recall curve based on the platform warning data. The horizontal axis represents the recall rate, which is the proportion of facilities that are judged

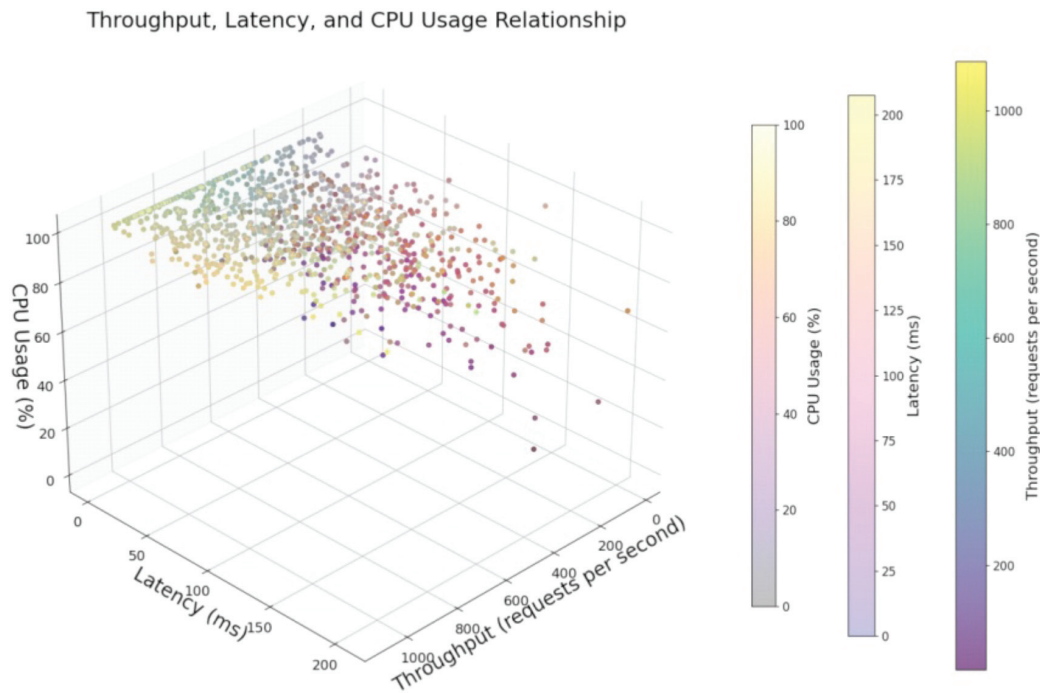


Figure 5 Stability Analysis Graph.

as faulty by the platform among all the facilities that are actually faulty. The vertical axis represents the precision, which is the proportion of total facilities that are actually faulty as determined by the platform. The area enclosed by the blue line and the coordinate axis is the AUC value, which directly indicates the strength of the platform's early warning capability. The closer the AUC value is to 1, the stronger the platform's ability to identify faults and issue early warnings. The AUC value of the early warning in this study is 0.92, which means that the proposed platform can effectively identify faults and issue early warnings.

4.4 Platform Stability Analysis

Figure 5 is a 3D relationship graph of the throughput, latency, and CPU usage of the platform studied in this paper. The stability of the platform is indirectly evaluated by analyzing the relationship between the three. From the image, the greater the platform throughput, the lower the latency, and the higher the platform CPU usage. As shown in the figure, most of the points are yellow. This indicates that the platform's CPU usage is very high most of the time, but under this high CPU usage, the latency of most points is relatively low, and only a few points have a large latency. This shows that the security management platform implemented in this study has good stability and can still maintain relatively low latency even under high CPU usage and high throughput.

5. CONCLUSION

In order to achieve all-round and real-time monitoring and early warning of hydropower facilities, this paper designs a

Kafka architecture specifically for transmitting data within the platform based on the original architecture, and uses a generative neural network to generate fault data and compare it with real data in real time to achieve rapid data transmission, real-time monitoring, and early warning. The research results show that the platform designed in this paper has high data transmission quality, strong early warning capability and good system stability, which provides a useful reference for the safety management of hydropower facilities. However, there are still some shortcomings in the research, such as the long response time of the platform under high load, and when the CPU usage is high, the delay may cause a slower response. Future research could focus on optimizing the response speed of the platform and reducing delays to achieve faster real-time response.

REFERENCES

1. Ocko, Ilissa B., and Steven P. Hamburg. "Climate impacts of hydropower: enormous differences among facilities and over time". *Environmental Science & Technology* 53.23 (2019): 14070–14082.
2. Mahmut, Tayeb. "Hydropower plant and its environmental effects". *Eurasian Journal of Chemical, Medicinal and Petroleum Research* 1.4 (2022): 130–137.
3. Leander, Johan, Klaminder, Jonatan, Jonsson, Micael, Brodin, Tomas, Leonardsson, Kjell, Hellstrom, Gustav. "The old and the new: evaluating performance of acoustic telemetry systems in tracking migrating Atlantic salmon (*Salmo salar*) smolt and European eel (*Anguilla anguilla*) around hydropower facilities". *Canadian Journal of Fisheries and Aquatic Sciences* 77.1 (2020): 177–187.
4. Attaran, Mohsen, and Jeremy Woods. "Cloud computing technology: improving small business performance using the

- Internet". *Journal of Small Business & Entrepreneurship* 31.6 (2019): 495–519.
5. Thavi, Riddhi, Jhaveri, Rujuta, Narwane, Vaibhav, Gardas, Bhaskar, Navimipour, Nima Jafari. "Role of cloud computing technology in the education sector". *Journal of Engineering, Design and Technology* 22.1 (2024): 182–213.
6. Ali, Omar, Shrestha, Anup, Osmanaj, Valmira, Muhammed, Shahnawaz. "Cloud computing technology adoption: an evaluation of key factors in local governments". *Information Technology & People* 34.2 (2021): 666–703.
7. Shaktawat, Anuja, and Shelly Vadhera. "Risk management of hydropower projects for sustainable development: a review". *Environment, Development and Sustainability* 23.1 (2021): 45–76.
8. Vucetic, Ivana, Kirin, Snezana, Sedmak, Aleksandar, Golubovic, Tamara, Lazic, Mirjana. "Risk management of a hydro power plant–fracture mechanics approach". *Tehnički vjesnik* 26.2 (2019): 428–432.
9. Obialor, C. A., Okeke, O.C, Onunkwo, A.A., Fagorite, V.I., Ehujuo, N.N. "Reservoir sedimentation: causes, effects and mitigation". *International Journal of Advanced Academic Research* 5.10 (2019): 92–109.
10. Adamo, Nasrat, Al-Ansari, Nadhir, Sissakian, Varoujan, Laue, Jan, Knutsson, Sven. "Dam safety: hazards created by human failings and actions". *Journal of Earth Sciences and Geotechnical Engineering* 11.1 (2021): 65–107.
11. Rubin, Oleg D., Antonov, Anton.S, Bellendit, Evgeny.N, Kobochkina, Ekaterina.M, Kotlow, Oleg.N. "Development of the design module of the software and hardware complex to ensure the safety of mutually influencing HPS". *Structural Mechanics of Engineering Constructions and Buildings* 15.2 (2019): 96–105.
12. Xiang, Yan, Sheng, JinBao, Wang, Lei, Cai, YueBo, Meng, Ying, Cai, Wei". Research progresses on equipment technologies used in safety inspection, repair, and reinforcement for deepwater dams". *Science China Technological Sciences* 65.5 (2022): 1059–1071.
13. Zhu, Xi, Bao, TengFei, Yeoh, Justin K.W., Jia, NingXiao, Li, Hui. "Enhancing dam safety evaluation using dam digital twins". *Structure and Infrastructure Engineering* 19.7 (2023): 904–920.
14. Kuban, Nurdan, and Marco Pretelli. "Industrial heritage assessment and guidelines for the architectural conservation of hydroelectric plants". *International Journal of Architectural Heritage* 15.10 (2021): 1526–1546.
15. Zhang, J., and Y. Shang. "Nexus of dams, reservoirs, climate, and the environment: a systematic perspective". *International Journal of Environmental Science and Technology* 20.11 (2023): 12707–12716.
16. Alvi, Irfan A., and Isabelle S. Alvi. "Why dams fail: A systems perspective and case study". *Civil Engineering and Environmental Systems* 40.3 (2023): 150–175.
17. Stojkovic, Milan, Marjanovic, Dusan, Rakic, Dusan, Ivetic, Damjan, Simic, Visnja, Milivojevic, Nikola, et al. "Assessment of water resources system resilience under hazardous events using system dynamic approach and artificial neural networks". *Journal of Hydroinformatics* 25.2 (2023): 208–225.
18. Wen Jiang. "The role of data fusion internet cloud computing in enterprise economic data management". *Engineering Intelligent Systems* 31(1): 2023.
19. Sang, Liansheng, Wang, Jun, Sui, JueYi, Dziedzic, Mauricio. "A new approach for dam safety assessment using the extended cloud model". *Water Resources Management* 36.15 (2022): 5785–5798.
20. Fera, Fation T., and Christos Spandonidis. "An Artificial Intelligence and Industrial Internet of Things-Based Framework for Sustainable Hydropower Plant Operations". *Smart Cities* 7.1 (2024): 496–517.
21. Kumar, Krishna, and R. P. Saini. "Adaptive neuro-fuzzy interface system based performance monitoring technique for hydropower plants". *ISH Journal of Hydraulic Engineering* 29.5 (2023): 611–621.
22. Guo, Jianhong. "Optimization and exploration of reservoir dispatching system based on cloud services". *Advances in Engineering Technology Research* 10.1 (2024): 7–7.
23. Solanki, Arpit, Debasis Sarkar, and Dhruv Shah. "Evaluation of factors affecting the effective implementation of Internet of Things and cloud computing in the construction industry through WASPAS and TOPSIS methods". *International Journal of Construction Management* 24.2 (2024): 226–239.
24. Markovic-Petrovic, Jasna D., Mirjana D. Stojanovic, and Slavica V. Bostjancic Rakas. "A fuzzy AHP approach for security risk assessment in SCADA networks". *Advances in Electrical and Computer Engineering* 19.3 (2019): 69–74.
25. Afridi, Yasir Saleem, Kashif Ahmad, and Laiq Hassan. "Artificial intelligence based prognostic maintenance of renewable energy systems: A review of techniques, challenges, and future research directions". *International Journal of Energy Research* 46.15 (2022): 21619–21642.
26. Zhang, ShaoQuan, Tan, YanKe, Ge, HanBin, Zhang, QiLin. "An ultra-thin bolt tension sensor and online monitoring system: For application in hydropower plant unit". *Frontiers of Structural and Civil Engineering* 18.9 (2024): 1388–1400.
27. Qingbin, Li, Ma, Rui, Hu, Yu, HuangFu, Zehua, Shen, YiYuan, Zhou, ShaoWu, et al. "A review of intelligent dam construction techniques". *Journal of Tsinghua University (Science and Technology)* 62.8 (2022): 1252–1269.
28. Peddireddy, Kiran. "Kafka-based architecture in building data lakes for real-time data streams". *International Journal of Computer Applications* 185.9 (2023): 1–3.
29. Vyas, Bhuman. "Optimizing Data Ingestion and Streaming for AI Workloads: A Kafka-Centric Approach". *International Journal of Multidisciplinary Innovation and Research Methodology*, ISSN: 2960–2068 1.1 (2022): 66–70.
30. Sethi, S., S. Panda, and S. Hooda. "Design and Optimization of a Zookeeper and Kafka-Based Messaging Broker for Asynchronous Processing in High Availability and Low Latency Applications". *J Curr Trends Comp Sci Res* 3.2 (2024): 01–07.
31. Goodfellow, Ian, Pouget-Abadie, Jean, Mehdi, Mirza, Xu, Bing, Warde-Farley, David, Ozair, Sherjil et al. "Generative adversarial networks". *Communications of the ACM* 63.11 (2020): 139–144.
32. Gui, Jie, Sun, ZheNan, Wen, YongGang, Tao, DaCheng, Ye, JiePing. "A review on generative adversarial networks: Algorithms, theory, and applications". *IEEE transactions on knowledge and data engineering* 35.4 (2021): 3313–3332.
33. Wang, Chaoyue, Xu, Chang, Yao, Xin, Tao, DaCheng. "Evolutionary generative adversarial networks". *IEEE Transactions on Evolutionary Computation* 23.6 (2019): 921–934.
34. Park, Kwangil, June Seok Hong, and Wooju Kim. "A methodology combining cosine similarity with classifier for text classification". *Applied Artificial Intelligence* 34.5 (2020): 396–411.
35. Rafiq, Muhammad, Ashraf, Shahzaib, Abdullah, Saleem, Mahmood, Tahir, Muhammad, Shakoor. "The cosine similarity measures of spherical fuzzy sets and their applications in decision making". *Journal of Intelligent & Fuzzy Systems* 36.6 (2019): 6059–6073.

36. Singh, RamniHarbir, Maurya, Sargam, Tripathi, Tanisha, Narula, Tushar, Srivastav, Gaurav. "Movie recommendation system using cosine similarity and KNN". *International Journal of Engineering and Advanced Technology* 9.5 (2020): 556–559.
37. Xi Chen and Yongbin Tan. Data center energy management based on cloud computing and artificial intelligence". *Engineering Intelligent Systems* 32(3):2024.