Collaborative Analysis of Multi-Regional Cultural Heritage Protection Enabled by Federated Learning

Xiaona Xie[1], Wenliang Zhang[2]
[1]Department of Teaching Management, Henan College of Transportation, Zhengzhou 451460, Henan, China
[2]Academy of Science & Technology, Zhengzhou University, Zhengzhou 450001, Henan, China
Corresponding author: Xiaona Xie (e-mail: xiexiaona026@163.com)

ABSTRACT Secure and distributed information sharing has become increasingly important for intelligent monitoring systems and large-scale collaborative sensing applications associated with electromagnetic communication networks. To address data fragmentation and privacy constraints in multi-regional cultural heritage protection, this study proposes a collaborative risk analysis and strategy evaluation framework based on federated learning. A spatiotemporal graph attention network (ST-GAT) is deployed at the client side, where a two-layer long short-term memory network captures temporal dependencies and a graph attention mechanism models interregional spatial correlations under privacy-preserving constraints. FedProx regularization and differential privacy are incorporated into federated optimization to improve robustness against non-IID data while protecting sensitive information. Furthermore, a federated Bayesian decision layer evaluates conservation strategies through posterior inference, multi-objective optimization, and Monte Carlo sampling. Experimental results demonstrate that the proposed framework achieves an RMSE of 0.113 and a risk mitigation rate of 95.45%, while maintaining stable cross-regional prediction performance and substantially reducing the success rates of membership inference and statistical attacks. The proposed approach provides an effective solution for privacy-preserving collaborative conservation and offers practical insights into secure distributed perception and intelligent information transmission for future electromagnetic communication and sensing environments.

INDEX TERMS federated learning, cultural heritage protection, spatiotemporal graph attention network, differential privacy

I. INTRODUCTION

CULTURAL heritage carries regional memory and social identity. Its conservation involves not only the physical restoration and long-term monitoring of cultural relics, but also local governance, resource allocation, and cultural policymaking. In this context, specialized heritage techniques are particularly salient, as their preservation directly connects to issues of local craftsmanship, the sustainability of artisanal industries, and the very embodiment of regional identity through material culture [1-3]. Cultural heritage encompasses a variety of forms, including physical artifacts, historical buildings and sites, as well as their corresponding archives and digital 3D models. It faces multiple risks, including natural disasters, environmental degradation, structural safety, human damage, management and maintenance, and technological and information risks. With the widespread application of digital technology in cultural heritage archives, 3D scanning, environmental monitoring, and public imagery, data-driven approaches have become a key path to improving conservation efficiency and scientific decision-making [4,5]. In reality, data is highly fragmented across regions and institutions, with significant data silos. These silos are caused by factors including closed-loop management of archives and images, a lack of unified standards and interoperability specifications, disparities in financial and technical capabilities, and fragmented archives and property rights disputes. These factors hinder sample aggregation, exacerbate bias and unfairness, delay emergency response and decision-making, lead to duplicate data collection and information loss, and weaken the resilience of cultural heritage preservation. Furthermore, legal, ethical, and institutional governance constraints hinder the cross-domain flow of raw data, severely hindering cross-regional collaborative analysis and unified decision-making. Differences in natural environments, social utilization, and conservation capabilities across regions lead to significant regional variations in risk patterns and conservation priorities. Traditional centralized modeling struggles to balance global generalization with local customization. Achieving cross-regional knowledge sharing, risk prediction, and strategic collaboration while ensuring data sovereignty and privacy

has become a pressing research challenge.

In the field of cultural heritage protection analysis, a large number of studies have conducted risk assessment and prediction for a single region or specific heritage type. Lei Y et al. proposed a fire risk prediction method for historical wooden structures based on entropy-weighted TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) and XGBoost (Extreme Gradient Boosting), which achieved high accuracy in the Palace Museum, but is difficult to generalize to multi-region collaborative analysis [6]. Moreno M et al. developed a platform to formulate protection strategies through fuzzy logic, GIS, and expert evaluation, but it relied on limited cases and manual input [7]. Mekonnen H conducted a survey in the Northern Shea region of Ethiopia and found that insufficient management, funding, professional resources, and community participation restricted protection efficiency [8]. Existing research has achieved certain results in cultural heritage risk prediction and protection strategy formulation, but there are generally problems such as data fragmentation, focusing on a single region or a single type of heritage, lack of cross-regional collaborative analysis and protection strategy integration, and data privacy and sharing restrictions.

To address data silos and poor data privacy in regional cultural heritage conservation, a challenge especially pronounced in sensitive and proprietary documentation, and to achieve accurate cross-regional risk assessment and actionable collaborative conservation decisions for cultural heritage while respecting data sovereignty and privacy, this paper applies federated learning on data from 21 provinces and cities. First, a multi-task ST-GAT (Spatio-Temporal Graph Attention Network) model is constructed and federated trained, using LSTM (Long Short Term Memory) as the temporal encoder and GAT (Graph Attention Networks) as the spatial attention backbone. The prediction head employs a dual-channel regression/classification model trained with an uncertainty-weighted joint loss. FedProx regularization and DP (Differential Privacy) noise are applied into the federated aggregation to adapt to non-independent and identically distributed (IID) distributions and preserve privacy. Subsequently, a Bayesian-based posterior update and Monte Carlo strategy sampling are performed on the server side on the aggregated prediction distribution to achieve collaborative strategy selection under resource constraints. Results show that this method achieves significant performance while preserving privacy. On the prediction side, the FL(DP)-ST-GAT algorithm achieves an RMSE (Root Mean Square Error) of 0.113, a 0.015 reduction compared to ST-GAT, a MAE (Mean Absolute Error) of 0.085, and a classification accuracy of 96.72%. Using Beijing as a benchmark for cross-regional generalization, the average RMSE is 0.118, demonstrating stable performance. On the decision side, the FL(DP)-Bayes algorithm achieves an expected utility of 0.91, a Conditional Value-at-Risk (CVaR) of 6.12%, a risk mitigation rate of 95.45%, and a resource allocation efficiency of 91.2%. The results demonstrate that coupling federated spatiotemporal graph modeling with Bayesian decision making can simultaneously enhance cross-regional generalization capabilities and the quality of collaborative decision-making, providing an operational path for cultural heritage protection under restricted data sharing conditions. This research aligns with the National Cultural Digitalization Strategy and the United Nations Sustainable Development Goals. By promoting cross-regional data collaboration within a framework of legal compliance and privacy protection, it can promote the digital transformation of cultural heritage, enhance the accessibility and equitable sharing of cultural resources, strengthen cultural governance at the local and national levels, and enhance cultural identity and confidence among the public, providing technical and policy references for achieving the long-term goal of integrated "protection, utilization, and inheritance."

II. MODEL BUILDING AND FEDERATED LEARNING

A. FEDERATED MULTI-TASK ST-GAT RISK PREDICTION MODEL

1) Temporal Encoding

This paper combines federated learning with ST-GAT to construct a federated multi-task ST-GAT risk prediction model [9,10]. A unidirectional LSTM is used for each node to map the original time series into a temporally compressed representation sequence and a final hidden state. The LSTM uses a standard gating formula, as shown in (1) [11-13]. The final hidden state h_T of the sequence is taken as the node time series representation z_i.

$$\begin{cases} f_t = \sigma(W_f x_t^{(i)} + U_f h_{t-1} + b_f) \\ i_t = \sigma(W_i x_t^{(i)} + U_i h_{t-1} + b_i) \\ o_t = \sigma(W_o x_t^{(i)} + U_o h_{t-1} + b_o) \\ \tilde{c}_t = \tanh(W_c x_t^{(i)} + U_c h_{t-1} + b_c) \\ c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ h_t = o_t \odot \tanh(c_t) \end{cases} \tag{1}$$

In formula (1), $x_t^{(i)}$ represents the input at time t; h_t represents the hidden state; and c_t represents the cell state. W_f, U_f, and bias b_f represent the science parameters. σ represents sigmoid, and $\odot$ represents element-wise multiplication.

2) Spatial Graph Construction and Normalization

The construction and normalization of spatial maps are shown in formula (2).

$$w_{ij} = \exp\left(-\frac{\|p_i - p_j\|^2}{2\sigma_d^2}\right) \cdot \frac{\langle m_i, m_j \rangle}{\|m_i\|\|m_j\|} \tag{2}$$

In formula (2), m_i represents the node attribute vector, including material, age, etc., p_i represents the geographic coordinates between regions, and w_{ij} represents the adjacency weight.

The k-th largest weight of each node is retained to form a sparse adjacency matrix, resulting in the adjacency matrix A.

This matrix is then symmetrized and normalized, as shown in formula (3).

$$\tilde{A} = D^{-1/2} A D^{-1/2}, \quad D_{ii} = \sum_j A_{ij} \tag{3}$$

In formula (3), $\tilde{A}$ represents the adjacency matrix after symmetry and normalization.

3) Spatial Attention Calculation

At each federated client, a multi-head GAT is applied to the temporal embedding zi of the node to obtain a spatial representation. The attention coefficient of the k-th head is first calculated as the unnormalized energy by linear transformation, as shown in (4) [14,15].

$$e_{ij}^{(k)} = \text{LeakyReLU}\left(a^{(k)^\top}\left[W^{(k)} z_i \| W^{(k)} z_j\right]\right) \tag{4}$$

In formula (4), $W^{(k)}$ represents the projection matrix, $a^{(k)}$ represents the attention vector, and $\|$ represents the vector connection.

Softmax normalization is performed on the neighbors, as shown in formula (5).

$$\alpha_{ij}^{(k)} = \frac{\exp(e_{ij}^{(k)})}{\sum_{j' \in N(i)} \exp(e_{ij'}^{(k)})} \tag{5}$$

The output of head k is a weighted sum and is activated, and multiple heads are concatenated to obtain a spatial representation, as shown in formula (6).

$$\begin{cases} u_i^{(k)} = \sigma\left(\sum_{j \in N(i)} \alpha_{ij}^{(k)} W^{(k)} z_j\right) \\ s_i = \|_{k=1}^{K} u_i^{(k)} \end{cases} \tag{6}$$

In formula (6), s_i represents the spatial representation obtained by multi-head concatenation, and $u_i^{(k)}$ represents the weighted sum and the result of the activation operation.

The LSTM temporal representation zi, the attribute vector mi, and the GAT spatial output si are linearly mapped and fused into a shared representation, as shown in formula (7).

$$h_i = \text{ReLU}(W_z z_i \| W_m m_i \| W_s s_t + b_h) \tag{7}$$

In formula (7), W_z, W_m, W_s, and b_h represent learnable parameters, ReLU represents the rectified linear unit, and h_i represents the shared encoding in this paper.

This paper explicitly defines spatial dependence in the context of multi-regional cultural heritage as a cross-regional risk association structure induced by geographical proximity, similarity of environmental exposure, and historical-craft associations, rather than a simple physical distance relationship. The edges in the graph structure do not represent direct interactions between heritage objects, but rather the transferability of risk evolution mechanisms and conservation experiences across different regions. On the one hand, geographically adjacent regions or those in the same climate zone (such as monsoon, arid, or plateau regions) often share similar temperature and humidity fluctuations, pollution exposure, and natural disaster patterns; their textiles exhibit highly correlated temporal dynamics in terms of material aging, dye fading, and fiber embrittlement. On the other hand, regions with similar historical periods, craft systems, or restoration traditions also show statistical similarities in risk type distribution and intervention responses. Based on this understanding, this paper views the spatial graph as a cross-regional risk similarity graph, with its adjacency weights composed of a weighted combination of geographical distance, similarity of environmental statistical distribution, and similarity of cultural/craft attributes, used to characterize "which regions' risk patterns and evolutionary experiences are of reference value to each other." The role of GAT is to adaptively learn and adjust the importance of these potential relationships, thereby enabling the selective transfer and aggregation of cross-regional risk knowledge without sharing the original data. This provides a clear and interpretable logical basis for establishing graph structures between multi-regional cultural heritage data centers.

The constructed multi-region graph structure is shown in Figure 1.

In Figure 1, nodes represent provinces, municipalities, and autonomous regions, and edge weights reflect the potential strength of interregional connections in cultural heritage protection. Geographically adjacent provinces, such as Hebei-Beijing, Sichuan-Chongqing, and Henan-Anhui, possess higher edge weights, indicating a stronger coupling between them in cultural heritage risk prediction and collaborative protection. Cross-regional connections, such as Liaoning-Jilin and Guangdong-Fujian, reflect connections driven by cultural and geographical similarities. Node size is determined by the strength of the connection.

Larger nodes, such as Henan, Sichuan, and Hebei, indicate a core position in the multi-regional collaborative network, responsible for greater information exchange and strategic coordination.

4) Local Training Objectives and Federated Updates and Communications

The client trains a local multi-task model on the shared code hi. The training of the shared code is achieved through a joint loss function as shown in (8).

$$\mathcal{L}_{local} = \lambda_r \mathcal{L}_{reg} + \lambda_c \mathcal{L}_{cls} + \frac{\mu}{2} \|\theta_\beta - \theta_g\|^2 \tag{8}$$

In formula (8), θ_β represents the local model parameter set, θ_g represents the current global parameter copy, $\mathcal{L}_{reg}$ represents the regression loss, $\mathcal{L}_{cls}$ represents the classification loss, and λ_r, λ_c, and μ represent hyperparameters. The last term is a FedProx-style regularization term used to suppress local update bias caused by non-IID.

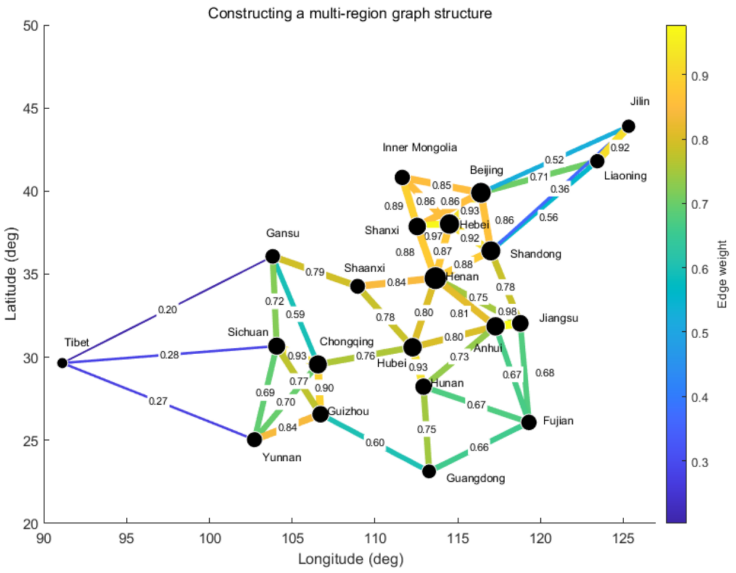

FIGURE 1. Constructed multi-region graph structure

After the client completes several local epochs of optimization, using Adam [16-18], it calculates the model differential $\Delta\theta_\beta = \theta_\beta - \theta_g$ and adds noise $\Delta\tilde{\theta}_\beta = \Delta\theta_\beta + \xi_\beta$, $\xi_\beta \sim N(0, \sigma^2 I)$ to the differential when applying differential privacy. The server performs weighted average aggregation, as shown in (9).

$$\theta_g^{(t+1)} = \sum_{\beta=1}^{\beta} \frac{n_\beta}{N} \Delta\tilde{\theta}_\beta \tag{9}$$

In formula (9), n_β represents the number of samples from client β.

If differential updates are used, the noised differences can be equivalently aggregated to achieve differential privacy. After aggregation, the latest $\theta_g^{(t+1)}$ is sent to the selected client to continue the next round of training.

B. MULTI-TASK PREDICTION HEAD DESIGN

In the federated multi-task ST-GAT model, the prediction head is used to map the shared spatiotemporal representation h_i to risk score regression values and risk level classification probabilities, thereby achieving multi-dimensional prediction of cultural heritage.

1) Regression Prediction Head

For each node, the regression head consists of two fully connected layers, as shown in (10).

$$r_i = W_2 \,\mathrm{ReLU}(W_1 h_i + b_1) + b_2 \tag{10}$$

In formula (10), W_1 and W_2 represent the learnable weight matrices in the fully connected layer, and b_1 and b_2 represent the corresponding biases. (Rectified Linear Unit) represents a continuous risk score, and r_i represents the risk level of a node due to potential damage, environmental impact, or material degradation.

The loss function uses Mean Squared Error (MSE), as shown in formula (11).

$$\mathcal{L}_{reg} = \frac{1}{N} \sum_{i=1}^{N} (r_i - y_i^{reg})^2 \tag{11}$$

In formula (11), y_i^{reg} represents the expert annotation or historical risk measurement value.

2) Classification Prediction Head

The structure of the classification head is similar to that of the regression head, but the output is a class probability vector, as shown in (12).

$$p_i = \mathrm{Softmax}(W_4 \,\mathrm{ReLU}(W_3 h_i + b_3) + b_4) \tag{12}$$

In formula (12), W_3 and W_4 represent the learnable weight matrices corresponding to the classification, and b_3 and b_4 represent the bias.

The loss function uses cross-entropy (CE), as shown in formula (13).

$$\mathcal{L}_{cls} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{ri=1}^{RI} y_{i,c}^{cls} \log p_{i,c} \tag{13}$$

Formula (13) represents the number of risk categories, and $y_{i,c}^{cls}$ represents the true category of the node. The structure of the federated multi-task ST-GAT risk prediction model is shown in Figure 2.

Figure 2 shows the overall structure of the federated multi-task ST-GAT risk prediction model. Each client first inputs multi-source information, including environmental

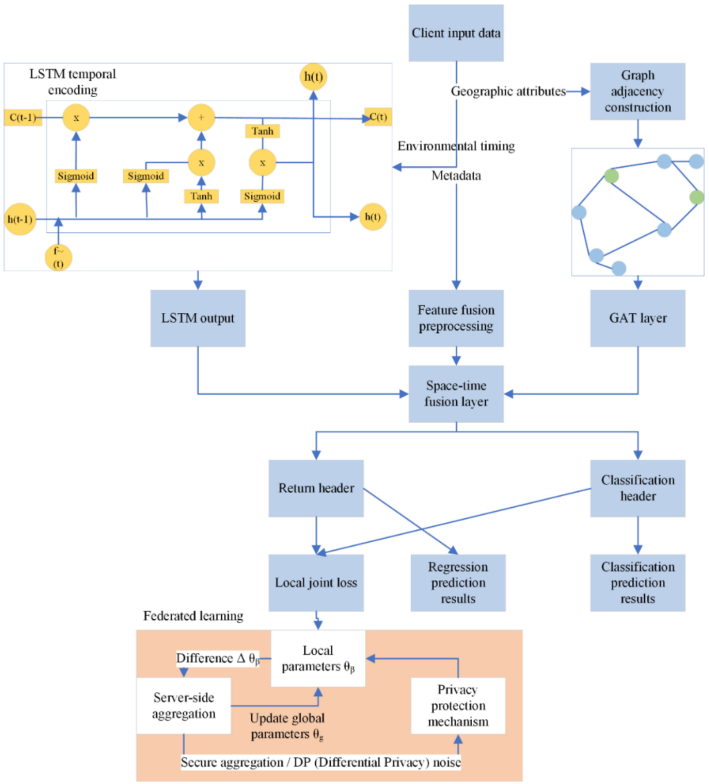

FIGURE 2. Structure of the federated multi-task ST-GAT risk prediction model

time series, geographic attributes, and metadata. The time series features are encoded by LSTM to obtain hidden representations, and the spatial relationships are input into the GAT layer based on the constructed adjacency matrix to extract spatial dependencies. Subsequently, they are integrated with the metadata in the spatiotemporal fusion layer to generate a unified node representation.

In each round of federated communication, the client first encodes its local environmental time series using a two-layer LSTM to generate temporally compressed embeddings (temporal representations). These embeddings only reflect the local raw observations and do not contain the raw data of neighbors. Spatial aggregation (GAT) then follows, executed locally on the client. GAT takes the local temporal embeddings and local static attribute vectors as input and uses an adjacency matrix broadcast by the server in advance without sensitive raw data to define the neighbor structure. To obtain cross-regional neighbor information, the client uses a snapshot of the aggregated and differentially privacy-processed neighbor embeddings from the server in the previous round as the neighbor features of GAT in each round, thereby completing message passing and attention weighting without sharing the raw time series. The output of GAT and the temporal embeddings of the local LSTM are linearly mapped and merged in the local fusion layer to form the final spatiotemporal representation, which is then fed into the regression/classification head for prediction. The client only reports model parameters or embedding differences to the server (protected by FedProx regularization and DP noise), ensuring that the cross-domain information source of spatial aggregation is clear, traceable, and complies with collective privacy constraints.

C. FEDERATED AGGREGATION AND PERSONALIZATION STRATEGY

During the federated multi-task ST-GAT model training process, each client node maintains local model parameters, and the server maintains global parameters. Each client performs optimization for a local epoch on the local sample set. A joint loss is computed for the shared representation h_i and the multi-task prediction head, with the gradients updated using the Adam optimizer. After the federated aggregation described above, and after the global model parameters are distributed, each client performs personalized fine-tuning to account for the non-IID characteristics of each region. Let the local personalized parameter be θ_k^{per}, and minimize the personalized loss on the local sample set, as shown in (14).

$$\mathcal{L}_k^{per} = \lambda_r \mathcal{L}_{reg,k}^{per} + \lambda_c \mathcal{L}_{cls,k}^{per} + \gamma \|\theta_k^{per} - \theta_g\|^2 \qquad (14)$$

In formula (14), $\mathcal{L}_k^{per}$ represents the personalized loss minimized on the local sample set, $\mathcal{L}_{reg,k}^{per}$ and $\mathcal{L}_{cls,k}^{per}$ represent the corresponding regression and classification losses, and γ represents the personalized regularization coefficient.

To further alleviate the differences in data distribution between different regions, the client-side adaptive coefficient

δ_β is applied in the aggregation stage, as shown in formula (15) [19-21].

$$\theta_g^{(t+1)} = \theta_g^{(t)} + \sum_{\beta=1}^{\beta} \delta_\beta \Delta\tilde{\theta}_k, \quad \delta_\beta = \frac{n_\beta}{N} \cdot \frac{1}{1 + KL(pro_\beta \| pro_g)} \tag{15}$$

In formula (15), pro_β represents the local label distribution of the client, pro_g represents the global distribution estimate, and KL represents the Kullback-Leibler divergence.

D. FEDERATED BAYESIAN COLLABORATIVE DECISION-MAKING FRAMEWORK

In multi-regional cultural heritage conservation, the decision-making process involves integrating risk prediction results from different regions and selecting strategies. To achieve cross-regional collaborative decision-making while protecting data privacy, this paper designs a federated Bayesian collaborative decision-making framework. This framework constructs a Bayesian inference model on the global server side. It uses the risk prediction distribution and classification probability uploaded by the client as input and selects conservation measures through posterior updating, strategy sampling, and collaborative optimization.

At the model level, FL-ST-GAT outputs not only a point estimate (mean) for each object at each node, but also an uncertainty measure (prediction variance or confidence level). The input to the decision layer is constructed based on this pair (mean + uncertainty). The specific process is as follows: The client calculates the regression mean and uncertainty for each object in its local samples, and locally performs "soft allocation" of this continuous output according to a predefined risk level range (i.e., assigning an object to low/medium/high levels based on probability according to the mean and uncertainty) to avoid coarse hard thresholds. These soft-allocated probabilities are then weighted by confidence level and converted into pseudo-counts, with a small smoothing term added to prevent zero counts, resulting in a class count vector for each client. The server receives the aggregated pseudo-count snapshots processed with differential privacy and accumulates them globally to form a Dirichlet parameter vector (i.e., the alpha value of each class). This Dirichlet posterior reflects both the average risk propensity of each client and retains uncertainty information. Based on this, the server uses Dirichlet-multinomial posterior sampling (Monte Carlo) assesses the success probability and risk distribution of different protection strategies, thus providing robust Bayesian decision inputs for protection strategy selection while taking uncertainty into account.

In this paper, textile collections are particularly well-suited to the proposed federated learning + Bayesian optimization architecture. Textile data is typically multimodal and highly dispersed—including high-resolution fabric images, dyeing/dye formulation records, microfiber and point cloud scans, restoration logs, and expert annotations—most of which are privately held by different museums or restoration institutions, with limited sample sizes within individual institutions, making centralized aggregation both impractical and potentially infringing on design intellectual property rights. Furthermore, textiles are highly sensitive to local environments (temperature, humidity, contaminants) and treatment history, resulting in significant non-IID distributions across different regions. Therefore, it is necessary to enhance predictive stability by preserving local characteristics while leveraging cross-domain information.

The federated ST-GAT architecture precisely meets this need by preserving time-series data and fragment/sample-level metadata locally and exchanging them through controlled, differentially privacy-preserving embeddings. Many textiles are preserved in a "fragmented" form (fragments, fragments, localized dyeing differences), and their restoration and risk assessment rely on fusing fragment-level features and rare process knowledge from multiple institutions. The federated framework can aggregate this information without exposing the original fragment data or process formulations, thereby improving the determination of dating/restoration priorities. Ultimately, the Bayesian decision-making layer can systematically integrate the forecast mean and uncertainty from various sources into the probability of strategy applicability, supporting resource allocation and intervention choices while protecting intellectual property rights and sensitive expert annotations. Therefore, for textile collections that need to balance data sovereignty, scarce samples and process confidentiality, this method has clear advantages in practical usability and compliance.

1) Risk Integration and Prior Construction

The regression results $r_i^{(\beta)}$ predicted by each client are aggregated into a Gaussian distribution hypothesis, as shown in (16).

$$r_i \sim \widehat{\mathcal{N}}(\mu_i, \sigma_i^2), \quad \mu_i = \frac{1}{\beta} \sum_{\beta=1}^{\beta} r_i^{(\beta)}, \quad \sigma_i^2 = \frac{1}{\beta} \sum_{\beta=1}^{\beta} \left(r_i^{(\beta)} - \mu_i \right)^2 \tag{16}$$

The classification probability is modeled using Dirichlet distribution, as shown in (17).

$$p_i \sim \mathrm{Dir}(\nu_i), \quad \nu_i = \sum_{\beta=1}^{\beta} p_i^{(k)} \cdot \iota \tag{17}$$

In formula (17), ν_i represents the parameter vector of the Dirichlet distribution, and ι represents the smoothing coefficient, which ensures that information about the minority class is not lost. Through the above aggregation, the server obtains the prior distribution of the global risk representation r_i and p_i for Bayesian inference.

2) Policy Posterior Update

For each strategy, a suitability scoring function $q_m(r_i, p_i; \theta_m)$ is defined, and the output is the strategy success probability.

The Bayesian update formula is used, as shown in (18) [22,23].

$$P(\theta_m \mid o) \propto P(o \mid \theta_m)P(\theta_m) \tag{18}$$

In formula (18), o represents the integrated global risk prediction, and $P(o \mid \theta_m)$ represents the likelihood function. The definition of the likelihood function is shown in formula (19) [24].

$$P(o \mid \theta_m) = \prod_{i=1}^{N} \widehat{\mathcal{N}}\left(r_i \mid \mu_i(\theta_m), \sigma_i^2(\theta_m)\right) \cdot \mathrm{Mult}\left(y_i \mid p_i(\theta_m)\right) \tag{19}$$

In formula (19), $\widehat{\mathcal{N}}$ represents the Gaussian likelihood, and Mult represents the polynomial likelihood. $\mu_i(\theta_m)$, $\sigma_i^2(\theta_m)$, and $p_i(\theta_m)$ are the predicted mean, variance, and probability distribution under the policy parameters, respectively. By optimizing the posterior $P(o \mid \theta_m)$ of θ_m through gradient descent, the applicability probability of each policy under the current risk conditions is obtained.

3) Collaborative Strategy Sampling and Combination

Monte Carlo sampling of strategies is performed using the posterior distribution, and maximum a posteriori selection is performed on strategy samples from multiple clients, so as to select the optimal strategy at the global level that takes into account the risk characteristics and resource constraints of multiple regions.

$$st^* = \arg\max_{st_m} \mathbb{E}_{m \sim P(\theta_m \mid o)} \left[q_m(r_i, p_i; \theta_m)\right] \tag{20}$$

In formula (20), st^* represents the final selected strategy.

4) Cost-Benefit Constraints and Multi-Objective Optimization

In actual deployment, the strategy cost and benefit function are applied to form a multi-objective optimization problem, as shown in (21).

$$st^* = \arg\max_{st_m} \left[\rho_1 \mathbb{E}\left[q_m(r_i, p_i; \theta_m)\right] - \rho_2 \pi_m - \rho_3 \varrho_m\right] \tag{21}$$

In formula (21), ρ_1, ρ_2, and ρ_3 represent weight hyperparameters, π_m represents the policy cost, and ϱ_m represents the benefit function.

Based on the client's regional characteristics and historical risk patterns, the server generates personalized parameters and sends them to the client for local policy adjustment. The client performs policy simulation and fine-tuning locally, using local data to provide feedback on the policy's effectiveness and update the local posteriors. Ultimately, through iterative updates, the local policy gradually converges while preserving cross-regional knowledge sharing.

In this study, the candidate protection strategy set includes a set of parameterizable and practically feasible measures, such as microclimate control (temperature and humidity stabilization, graded by intensity), physical isolation/protective shield installation, local repair/reinforcement (billed by repair intensity and labor hours), priority digitization and high-precision recording (billed by sample size), emergency relocation/temporary storage, and community/institutional capacity building. Each strategy is represented by a finite-dimensional vector (binary/intensity variable) to facilitate combination and scheduling. The cost function consists of three parts: direct costs (materials, labor, operation and maintenance), time costs (construction period, exhibition/usage interruption losses), and opportunity costs (delays in other projects due to resource occupation). The benefit function mainly quantifies the expected reduction in damage, the number of grade downgrades (number of objects from high to medium/low), and the number of years of extended safe life. To balance fairness and robustness, the objectives also include minimizing inter-regional performance differences and minimizing risk tail metrics (such as CVaR). In its optimization implementation, the Dirichlet (classification) and Gaussian (regression) posteriors are first obtained from the server. Monte Carlo sampling is then used to propagate each sample to a policy cost-benefit simulator to estimate the distribution of benefits, costs, and risks of the policy combination under that sample. Expectations and quantiles (for CVaR) are calculated for a large number of samples to compute expected utility, cost, and risk tail indicators for each candidate combination. Subsequently, a weighted scaling method is used to search for the Pareto front in a multi-objective space of "maximizing expected utility / reducing CVaR / minimizing cost / improving cross-regional fairness." Under budget and resource constraints, the final policy combination is selected using the maximum posterior probability or the decision-maker preference rule.

This process preserves the uncertainty information of the Bayesian posterior and directly couples policy evaluation with actual cost-benefit analysis, thereby concretizing the abstract "collaborative solution" into an executable, comparable, and traceable decision output.

The experimental parameter settings are shown in Table 1.

TABLE 1. Experimental parameters

Parameters	Value	Parameters	Value
Number of clients	21	Fusion representation dimension	256
Number of global communication rounds	200	Multi-task fully connected hidden layer dimension	Regression: 128; Classification: 128
Number of local training rounds	5	Multi-task weights	Initial values are 1.0
Local batch size	32	Differential privacy noise	0.5
Local learning rate	1×10^{-3}	Clipping norm	1
Optimizer	Adam (0.9, 0.999)	Personalized fine-tuning	3 epochs
Number of LSTM layers	2	Communication compression	8-bit quantization
LSTM hidden dimensions	128	Dropout probability	0.3
Number of GAT heads	8	Weight decay (L2)	1×10^{-5}
GAT output dimensions per head	64	Early stopping	Global validation: no improvement for 10 rounds

III. EVALUATION PLAN

A. EXPERIMENTAL DATA AND PREPROCESSING

The experimental data in this paper covers 21 regions: Beijing, Hebei, Shanxi, Shandong, Gansu, Chongqing, Shaanxi, Jiangsu, Hubei, Tibet Autonomous Region, Yunnan, Henan, Sichuan, Anhui, Liaoning, Jilin, Guangdong, Fujian, Inner Mongolia Autonomous Region, Hunan, and Guizhou. The data comes from three sources: (1) Archives of local cultural institutions and partner museums: video materials, restoration records, and expert evaluation forms obtained through data use agreements signed with provincial and municipal cultural relics bureaus, museums, and site management units. This section is the main source of the annotation, including high-definition images of the real objects, before-and-after comparison photos, and manual evaluation records. (2) Environmental and time-series monitoring data: Data from national/provincial public meteorological and environmental monitoring platforms, including hourly temperature, relative humidity, precipitation, pollutant concentrations, etc., as well as long-term observation series from nearby meteorological stations and environmental monitoring points. Remote sensing images using Sentinel-2 and Landsat series satellite images are used for large-scale environmental change and vegetation cover analysis. (3) 3D reconstruction and point cloud data: 3D scanning data (structured light/laser scanning) of cultural relics/structures provided by local cultural institutions, as well as point cloud summaries and deformation records obtained based on multi-view photogrammetry. Supplementary metadata includes geographic coordinates, source material, age, known restoration records, and inspection logs.

During preprocessing, all images and 3D data are unified in resolution and color space, and automated quality screening is performed to remove blurred or heavily exposed samples. Geometric correction, histogram equalization, and texture-based local contrast enhancement are performed on the images to improve detail visibility. Manual annotations are cleaned through a double-blind review process, with damage classification and grading standards standardized. The original labels from different institutions are mapped to a unified labeling body. Point cloud data is downsampled and feature summaries (surface normals, curvature, etc.) are performed to reduce computational overhead. 3D models are then normalized and locally registered, with the output of a unified geometric descriptor format for subsequent use in GAT node attributes.

B. COLLABORATIVE ANALYSIS SOLUTION

The experiment shards the data in each region into local shards as federated clients, keeping the data local and local. Two types of experiments are conducted to compare prediction and classification models: a centralized/local baseline and FL(DP). The centralized baseline merges data on a central server to train the corresponding model; the local baseline trains each client independently and reports site performance.

Federated experiments are conducted using the following schemes: FL(DP)-XGBoost, FL(DP)-LightGBM, FL(DP)-LSTM, FL(DP)-TFT (Temporal Fusion Transformer), and FL(DP)-ST-GAT. Several privacy budgets are compared to evaluate the privacy-performance tradeoff. Training hyperparameters are uniformly controlled, and experiments are repeated under different non-IID scenarios, with each configuration run 10 times to obtain the average. Evaluation metrics include regression (RMSE, MAE), classification (Accuracy, Precision/Recall/F1), and cross-region fairness (performance gap between sites).

IV. RESULTS OF COLLABORATIVE ANALYSIS ON MULTI-REGIONAL CULTURAL HERITAGE PROTECTION

A. OVERALL PREDICTION ACCURACY AND CLASSIFICATION PERFORMANCE

To validate the effectiveness of the proposed method in cultural heritage risk prediction and classification, it compares the performance of various federated and non-federated baseline models across 21 regions, including Beijing, Hebei, and Shanxi provinces. Prediction accuracy corresponds to the regression value of the risk score (ranging from 0 to 1), while classification performance measures the ability to discriminate heritage risk levels. The overall prediction accuracy and classification performance are shown in Figure 3.

In Figure 3(a), taking RMSE as an example, the RMSE of FL(DP)-XGBoost is 0.167, while that of FL(DP)-ST-GAT is 0.113, a decrease of 0.054 compared to FL(DP)-XGBoost, corresponding to a 0.039 decrease in MAE and a 7.19% decrease in MAPE. The tree-based models (XGBoost and LightGBM) perform poorly, while LSTM and TFT improve in turn (LSTM RMSE 0.142, TFT 0.137). ST-GAT further reduces the RMSE to 0.128. Finally, the federated ST-GAT with DP, after integrating multi-source regional

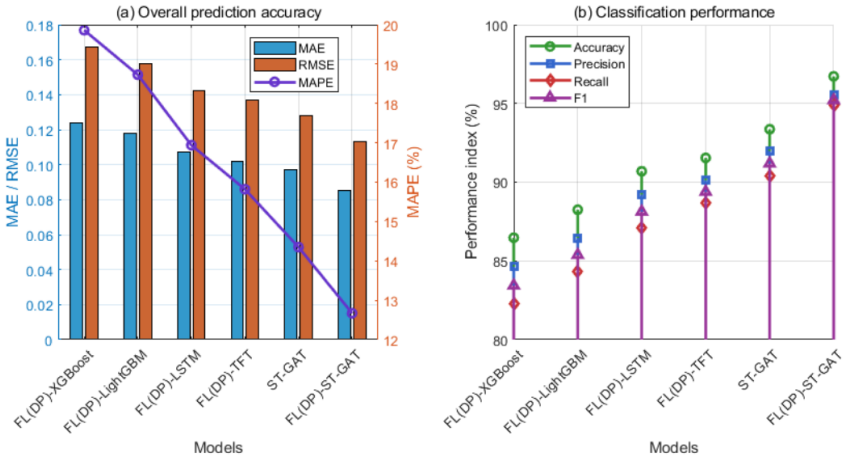

FIGURE 3. (a) Overall prediction accuracy; Figure 3 (b) Classification performance
Overall prediction accuracy and classification performance

information, further reduces the RMSE to 0.113, a further 0.015 decrease compared to the non-federated ST-GAT. The combination of federated aggregation and spatial-temporal graph modeling is most effective in reducing regression error and improving numerical prediction accuracy. Meanwhile, FL(DP)-ST-GAT achieves a significant error reduction while preserving privacy, indicating that the diversity benefits of cross-region aggregation outweigh the noise cost applied by DP.

In Figure 3(b), FL(DP)-ST-GAT also performs best, with accuracy increasing by 10.25% from 86.47% for FL(DP)-XGBoost to 96.72%. Precision also increased by 10.93% from 84.63% to 95.56%, recall by 12.59% from 82.28% to 94.87%, and F1 by 11.77% from 83.44% to 95.21%. Compared to non-federated ST-GAT, FL(DP)-ST-GAT improves accuracy by 3.36% from 93.36% to 96.72%, demonstrating that federated aggregation can further refine the class discrimination boundaries and significantly improve recall. In summary, the cross-model improvements in classification metrics are particularly significant in recall and F1, indicating that FL(DP)-ST-GAT has the greatest practical value in reducing missed high-risk objects. These results also indicate that graph attention and cross-region information fusion play a key role in improving classification robustness and recall.

B. VERIFICATION OF GENERALIZATION PERFORMANCE IN EACH REGION

To evaluate the model's generalization ability in cross-regional applications, it trains the model using data from Beijing and applies it to 20 other regions (such as Hebei, Shanxi, and Shandong). It compares the performance of local baselines, centralizes training, and FL (DP) in regression and classification tasks. The results are shown in Figure 4.

In Figure 4(a), the FL(DP) model significantly reduces regression error across all regions. Taking Beijing as an example, the RMSE for the local baseline is 0.145, the centralized approach is 0.123, and the FL(DP) approach is 0.113. Compared to the local baseline, FL(DP) reduces the RMSE in Beijing by 0.032, and further by 0.010 compared to the centralized approach. Spatially, the improvement in error occurs in both dense and sparse sample areas. For example, the error in Guangdong Province decreases from 0.152 for the local approach to 0.113 for the FL(DP) approach, while it decreases from 0.179 to 0.125 for the Tibet Autonomous Region. Overall, the RMSE fluctuation range across regions under FL(DP) is relatively small, while the fluctuation range for the centralized and local baselines is larger. This indicates that federated aggregation not only reduces absolute error but also significantly reduces performance differences between regions, improving cross-regional generalization stability.

In the classification task shown in Figure 4(b), FL(DP) also achieves widespread and significant improvements. Taking Beijing as an example, the accuracy increases from 87.56% for local training to 92.41% for centralized training, and then to 96.72% for FL(DP), representing a 9.16% improvement over local training and a 4.31% improvement over centralized training. The improvement is particularly pronounced in regions with sparse samples or large environmental differences: Tibet shows an increase from 78.32% for local training to 91.64% for FL(DP); and Hebei shows an increase from 83.21% to 94.87%. FL(DP) achieves high accuracy across all regions, while the centralized and local baselines are relatively low. This indicates that federated training not only improves overall discriminative ability but also effectively narrows the gap in classification performance between regions. It is particularly effective in recalling high-risk categories (reducing missed classifications).

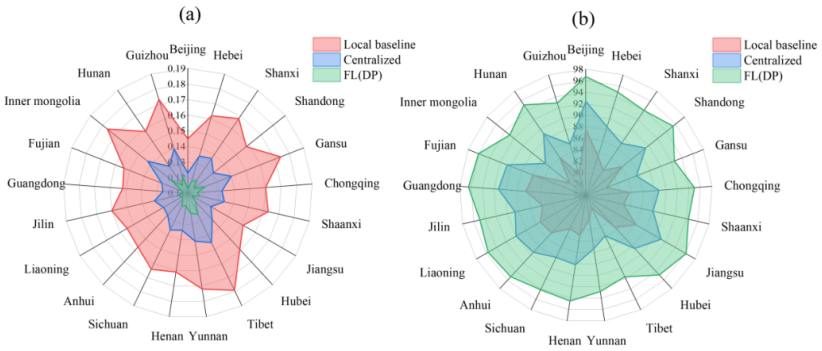

FIGURE 4. (a) RMSE; Figure 4 (b) Accuracy
Verification of generalization performance in each region

C. DIFFERENCES IN CULTURAL HERITAGE RISK SCORE REGRESSION VALUES AND RISK TYPES ACROSS MULTIPLE REGIONS

The regression values of cultural heritage risk scores and risk types in multiple regions are shown in Figure 5.

Figure 5 shows significant regional differences in the multi-regional cultural heritage risk scores. High-risk areas are concentrated in economically developed and highly urbanized regions, such as Guangdong Province (0.74), Beijing (0.73), Jiangsu Province (0.72), Liaoning Province (0.70), and Anhui Province (0.71). Remote or geographically unique regions have relatively low scores, such as the Tibet Autonomous Region (0.56), Gansu Province (0.58), Inner Mongolia Autonomous Region (0.57), Guizhou Province (0.60), and Yunnan Province (0.59). Intermediate-level regions include Hebei Province (0.66), Shanxi Province (0.61), Hubei Province (0.67), Shandong Province (0.69), and Chongqing Municipality (0.65), demonstrating significant regional differences in cultural heritage risk. The average score is approximately 0.65, indicating that most regions face medium-to-high risks and require focused attention on risk management and resource allocation strategies.

The distribution of risk types reveals significant differences in key risk factors across regions. In Beijing, structural safety risks (28%) and environmental degradation risks (22%) are the primary risk factors, while human damage and management and maintenance risks account for a high proportion (18%), indicating that cultural heritage in urban centers is significantly impacted by human traffic and environmental pressures. In the Tibet Autonomous Region, natural disaster risks (18%) and environmental degradation risks (24%) predominate, while technology and information risks account for only 11%, indicating that remote areas are more constrained by natural and environmental factors. In Gansu Province, management and maintenance risks account for a high 46%, while structural safety risks only account for 6%, indicating that resource scarcity and inadequate maintenance are primary issues. In Guizhou Province, human damage risks account for 43%, while natural disaster and structural risks are relatively low, reflecting the significant threat posed by local human factors to cultural heritage. Overall, the distribution of risk types varies across regions, influenced by both geographic environment and socioeconomic development, management capacity, and human traffic.

D. DECISION-MAKING EFFECTS AND ACTUAL BENEFIT EVALUATION OF DIFFERENT METHODS

To compare the benefits and decision quality of different decision-making methods in actual deployment scenarios, the experiment analyzes expected utility, cost/benefit ratio, CVaR and decision transparency scores, as well as risk mitigation rate, resource allocation efficiency, and policy coverage/accuracy indicators. The results are shown in Table 2.

In Table 2, the FL(DP)-Bayes algorithm performs best, achieving an expected utility of 0.91, significantly higher than the traditional Bayesian algorithm's 0.88 and far exceeding FL(DP)-PPO (0.79) and FL(DP)-DQN (0.76). In terms of cost-benefit ratio, the FL(DP)-Bayes algorithm's is only 0.28, lower than all other compared methods, demonstrating its cost-effectiveness while maintaining high utility. The FL(DP)-AHP algorithm, on the other hand, has the highest ratio (0.55), indicating relatively high costs. Regarding risk control, the FL(DP)-Bayes algorithm also has the lowest CVaR (6.12%), demonstrating a more robust performance compared to FL(DP)-DQN (9.11%) and FL(DP)-AHP (11.20%). In terms of decision transparency, FL(DP)-AHP scores the highest (8.95), but FL(DP)-Bayesian algorithm comes in second (8.90), outperforming FL(DP)-PPO (6.05) and FL(DP)-DQN (5.40), achieving both interpretability and stability.

E. MULTI-REGION COLLABORATIVE DECISION-MAKING EFFECT

The effect of multi-region collaborative decision-making is shown in Figure 6.

As shown in Figure 6(a), the risk mitigation rate in most regions remains above 90%, but there is significant variation. The most outstanding regions are Hebei and Henan Provinces (both at 99.90%), Sichuan Province

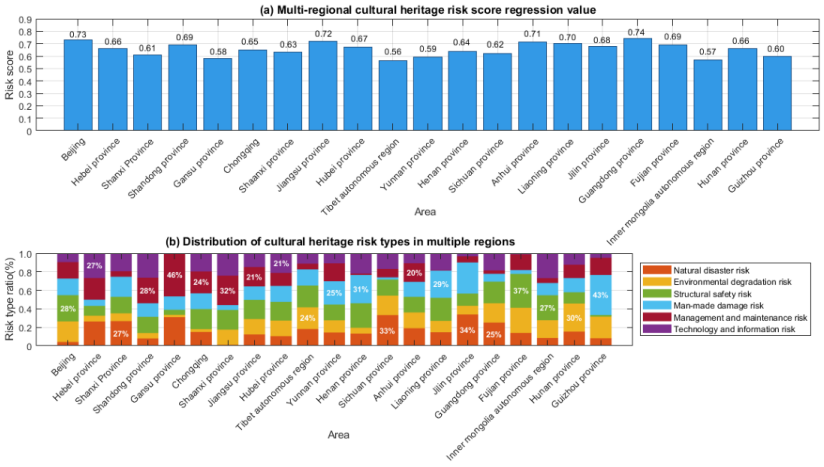

FIGURE 5. (a) Regression values of cultural heritage risk scores for multiple regions; Figure 5 (b) Distribution of cultural heritage risk types for multiple regions; Regression values of cultural heritage risk scores and risk types for multiple regions

TABLE 2. Evaluation of the actual benefits of different decision-making methods

Methods	Expected utility	Cost/Benefit ratio	CVaR (%)	Decision transparency score (0-10)	Risk mitigation rate (%)	Resource allocation efficiency (%)	Strategy selection coverage (%)	Strategy selection accuracy (%)
FL(DP)-bayesian algorithm	0.91	0.28	6.12	8.90	95.45	91.2	98.5	94.3
Bayesian algorithm	0.88	0.32	6.98	8.20	92.12	88.35	95.4	92.1
FL(DP)-PPO	0.79	0.45	8.34	6.05	88.33	85.1	92.3	87.55
FL(DP)-DQN	0.76	0.48	9.11	5.40	85.88	83.85	90.15	85.2
FL(DP)-RF	0.81	0.36	7.50	7.20	81.4	80.12	86.6	81.9
FL(DP)-AHP	0.69	0.55	11.20	8.95	77.75	76.5	79.25	78.45

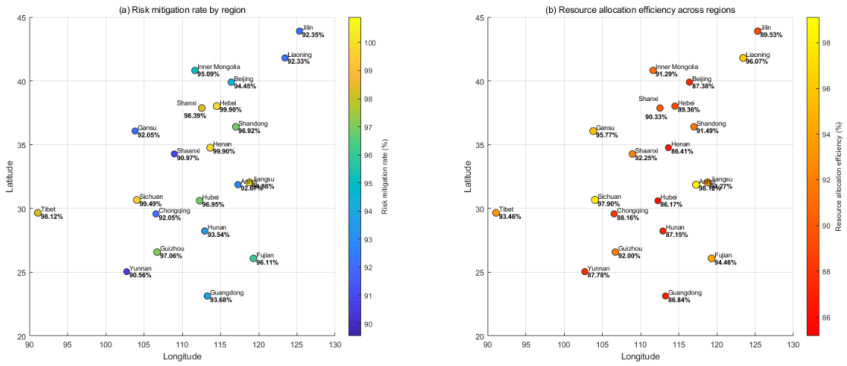

FIGURE 6. (a) Risk mitigation rate by region; Figure 6 (b) Resource allocation efficiency by region; Multi-region collaborative decision-making effectiveness

(99.49%), and Jiangsu Province (99.86%), all approaching 100%, demonstrating that these regions are extremely robust in risk prevention and emergency response. Tibet Autonomous Region and Shanxi Province also perform well, reaching 98.12% and 98.39%, respectively. In contrast, the risk mitigation rates in some regions are relatively low, such as Yunnan Province (90.56%), Shaanxi Province (90.97%), Chongqing Municipality, and Gansu Province (both at 92.05%), placing them at the bottom of the national rankings. Overall, provinces with outstanding risk mitigation rates are concentrated in East China (Jiangsu and Shandong) and key nodes in central and western China (Sichuan and Tibet), demonstrating that the collaborative model is able to address risk control in both developed eastern regions and key western regions.

In this paper, the risk mitigation rate does not refer to the direct percentage decrease in the risk regression value in a continuous sense, nor is it equivalent to the percentage reduction in risk score. Instead, it is a discrete utility indicator oriented towards decision-making outcomes, used to measure the actual intervention effectiveness of collaborative decision-making strategies on high-risk objects. This paper first classifies cultural heritage objects into three risk levels—low, medium, and high—based on the risk scores output by the risk prediction model. Then, after generating and implementing protection/intervention strategies in the federated Bayesian decision-making module, based on subsequent inspection records, expert reviews, or simulation evolution results, it statistically analyzes the percentage of objects initially classified as "high-risk" that

were successfully downgraded to "medium-risk or low-risk" after intervention. The risk mitigation rate is defined as the percentage of high-risk objects successfully downgraded out of the initial total number of high-risk objects. 95.45% indicates that, under this experimental setting, approximately 95.45% of high-risk cultural heritage objects achieved effective risk mitigation after strategy implementation. This indicator supplements prediction accuracy indicators such as RMSE from a decision-making and management perspective, reflecting the operational value of the model output in actual protection actions, rather than replacing or obscuring the risk regression error itself.

Judging from the resource allocation efficiency shown in Figure 6(b), some regions demonstrate exceptionally high levels of resource coordination. Sichuan Province reached 97.90%, followed by Anhui Province (98.12%), Liaoning Province (96.07%), and Gansu Province (95.77%), demonstrating a high degree of optimization in resource allocation and distribution through multi-regional collaboration. Fujian Province and the Tibet Autonomous Region in the southeastern coastal areas also performed exceptionally well, reaching 94.46% and 93.46%, respectively. Regions with relatively low efficiency include Hubei Province (86.17%), Henan Province (86.41%), Guangdong Province (86.84%), and Hunan Province (87.15%). Overall, the eastern coastal areas and some central and western provinces excel in resource coordination, while some provinces in central China still have certain shortcomings in cross-regional collaboration and need to further improve their collaborative allocation mechanisms.

F. NON-IID SENSITIVITY ANALYSIS

To evaluate the robustness and generalization capabilities of different methods to data distribution heterogeneity, this paper systematically compares several federated and centralized models under non-IID strengths (1, 0.7, 0.5, 0.3, and 0.1). The non-IID strength is set to 1, indicating near-IID distribution, and 0.1, indicating high heterogeneity. The results of the non-IID sensitivity analysis are shown in Figure 7.

In Figure 7, as the non-IID degree decreases from 1.00 to 0.10, the regression errors of all models increase significantly, but the magnitude of the increase varies. Starting from the near-IID (1.00) case, the initial RMSE for FL(DP)-ST-GAT is the smallest at 0.111, while for FL(DP)-XGBoost it is 0.171, ST-GAT is 0.130, and both FL(DP)-LSTM and FL(DP)-TFT are 0.142. When heterogeneity increases to 0.10, the RMSE increases to 0.172 for FL(DP)-ST-GAT, 0.281 for FL(DP)-XGBoost, 0.202 for ST-GAT, 0.232 for FL(DP)-LSTM, and 0.212 for FL(DP)-TFT. In the extreme non-IID case (0.10), the error range is 0.172–0.281, and the RMSE of FL(DP)-ST-GAT (0.172) is 0.109 lower than the worst FL(DP)-XGBoost (0.281). This shows that under strong heterogeneity conditions, the graph-time series joint representation and federated aggregation enable FL(DP)-ST-GAT to maintain a clear absolute advantage in numerical prediction and be more robust to heterogeneity.

Classification performance also declines with increasing non-IID, but the relative ranking between models remains stable. When the IID is approximately 1.00, the highest accuracy is achieved by FL(DP)-ST-GAT at 96.72%, while ST-GAT achieves 93.36% and FL(DP)-TFT achieves 91.54%. When the IID is extremely high (0.10), the accuracy of the models is 85.34% for FL(DP)-ST-GAT, 79.61% for ST-GAT, 74.38% for FL(DP)-TFT, and 66.88% for FL(DP)-XGBoost. At a heterogeneity of 0.10, FL(DP)-ST-GAT outperforms ST-GAT by 5.73% and significantly outperforms tree models and most time series models, demonstrating that federated spatiotemporal graph models are more resistant to distribution shift in class discrimination while maintaining high recall and precision. When the differences in client data distribution increase, the directional drift caused by local updates makes it difficult for the global model of simple weighted aggregation to take into account the characteristics of each domain, especially affecting tree models that rely on boundaries/segments (XGBoost/LightGBM). Therefore, the error and accuracy drop the most under non-IID.

G. PRIVACY AND SECURITY ASSESSMENT

Figure 8(a) shows the performance (RMSE/accuracy) curves for different privacy budgets. To verify the privacy protection effectiveness of Federated Learning-DP, typical attacks, including membership inference attacks and statistical attacks, are applied to centralized, centralized-DP, FL, and FL-DP models to explore the attack success rate. The results are shown in Figure 8(b). The privacy and security evaluation is also shown in Figure 8.

Under different privacy budget conditions in Figure 8, the prediction accuracy and classification performance of the model show a "privacy-performance trade-off". The smaller the privacy budget, the higher the noise injection intensity, the higher the RMSE and the lower the accuracy. When the budget is 0.1, the RMSE is as high as 0.225, and the accuracy is only 85.12%; when the budget is increased to 0.5, the RMSE drops significantly to 0.170, and the accuracy rises to 90.56%; when the budget is further increased to 2, the RMSE is controlled at 0.125, and the accuracy reaches 95.02%. Under a higher budget (5), the RMSE is as low as 0.113 and the accuracy is as high as 96.72%, which is almost the same as the ideal performance without differential privacy. This shows that differential privacy does have an impact on model performance, but when the budget is greater than or equal to 1, the performance loss is relatively small, and it can balance privacy protection and prediction accuracy.

In the membership inference and statistical attack scenarios, different deployment modes exhibit significant differences in their ability to resist attacks. The centralized model presents the highest risk, with a membership inference attack success rate of 72.45% and a statistical attack success rate of 60.1%, demonstrating that centralized data aggregation is highly susceptible to leaking sensitive information. Af-

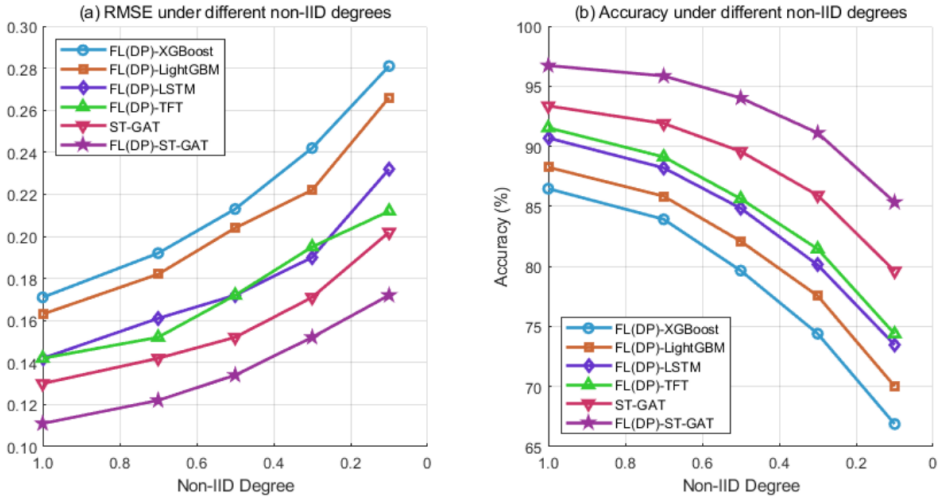

FIGURE 7. (a) RMSE at different levels; Figure 7 (b) Accuracy at different levels; Non-IID Sensitivity Analysis

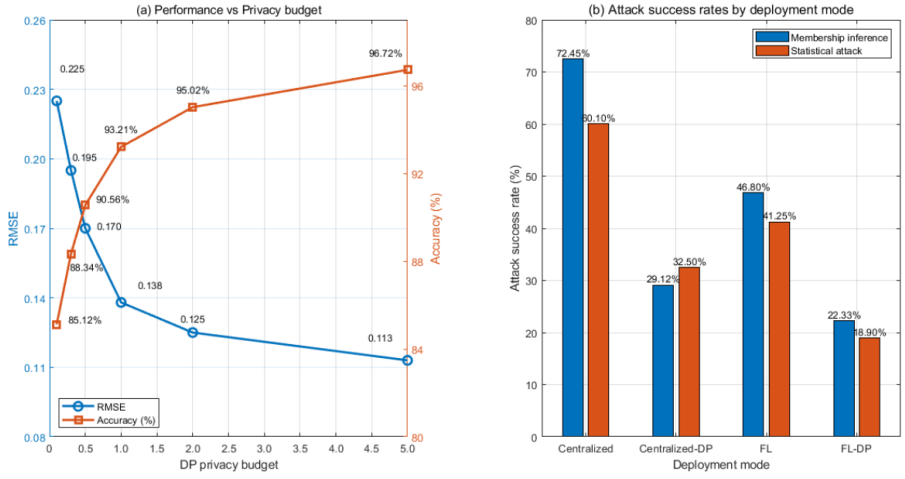

FIGURE 8. (a) Performance (RMSE/Accuracy) curves under different privacy budgets; Figure 8 (b) Attack success rates for different deployment modes; Privacy and Security Evaluation

ter applying differential privacy, the centralized-DP attack success rate is significantly reduced to 29.12% and 32.5%. In contrast, the FL model inherently improves privacy but still presents risks, with an attack success rate of 46.8% and 41.25%. The FL-DP model, while enhanced by differential privacy, performs best, yet has the lowest success rate, at only 22.33% (membership inference) and 18.9% (statistical attack). This demonstrates that federated learning combined with differential privacy not only maintains excellent performance on a high budget but also significantly reduces attack risk, balancing practicality and security in cross-regional collaborative analysis for cultural heritage protection.

H. DIFFERENTIAL PRIVACY (DP) AND PERFORMANCE TRADE-OFF EXPERIMENTS

This paper designs and executes comparative experiments, running (A) federated FL-ST-GAT (Non-DP baseline) without differential privacy and (B) FL(DP)-ST-GAT with Gaussian DP noise injected into global aggregation, under the same data splitting, non-IID strength, model structure and training hyperparameters.

The experimental steps were as follows: (1) Fix the random seed and repeat each configuration 10 times on 21 clients with the number of communication rounds and local training settings given in Table 1 to reduce random fluctuations; (2) Test the privacy budget $\epsilon = 5, 2, 0.5, 0.1$ (and $\epsilon \to \infty$ corresponding to Non-DP) one by one in the DP experimental group, injecting Gaussian noise of corresponding intensity into the model difference reported by the client each time; (3) Record and average the regression metrics (RMSE, MAE), classification metrics (accuracy), actual decision utility (approximated by risk mitigation rate), and privacy adversarial metrics (member inference attack success rate and statistical attack success rate) for each run; the results are shown in Table 3.

In Table 3, FL-ST-GAT exhibits a typical and controllable privacy-performance trade-off after introducing differential privacy. Compared to the federated baseline without dif-

TABLE 3. Performance and Privacy Adversarial Evaluation of FL-ST-GAT under Different Differential Privacy Budgets

Configuration	Privacy Budget ϵ	RMSE	Accuracy (%)	Risk Mitigation Rate (%)	Member Inference Attack Success Rate (%)	Statistical Attack Success Rate (%)
Non-DP (Baseline)	∞	0.108	97.10	96.8	46.8	41.3
FL(DP)-ST-GAT	5	0.113	96.72	95.45	22.3	18.9
FL(DP)-ST-GAT	2	0.125	95.02	94.1	18.0	15.0
FL(DP)-ST-GAT	0.5	0.170	90.56	89.3	10.5	8.6
FL(DP)-ST-GAT	0.1	0.225	85.12	75.0	4.2	3.5

ferential privacy (RMSE = 0.108, accuracy = 97.10%, risk mitigation rate = 96.8%), under a more lenient privacy budget $\epsilon = 5$, the model performance only slightly decreases. RMSE increases to 0.113 (+0.005), accuracy decreases to 96.72% (-0.38%), and the risk mitigation rate remains at 95.45%. However, the success rate of member inference attacks and statistical attacks significantly decreases from 46.8% and 41.3% to 22.3% and 18.9%, respectively, demonstrating significant privacy gains. As privacy budgets tightened further ($\epsilon = 2$ and 0.5), RMSE increased to 0.125 and 0.170, respectively, while accuracy decreased to 95.02% and 90.56%.

Simultaneously, attack success rates continued to decline to 18.0%/15.0% and 10.5%/8.6%, indicating a monotonic relationship between performance loss and privacy enhancement. When $\epsilon = 0.1$, the model achieved the strongest privacy protection (attack success rate below 5%), but RMSE significantly increased to 0.225, and risk mitigation rate decreased to 75.0%, showing that excessive noise significantly weakens prediction and decision-making capabilities. Overall, these results demonstrate that within the range of $\epsilon \approx 2$–5, FL(DP)-ST-GAT can significantly improve privacy security at the cost of limited accuracy, thus achieving a more practical performance-privacy balance for collaborative analysis of multi-regional cultural heritage.

I. NON-INDEPENDENT IDENTICALLY DISTRIBUTED (NON-IID) QUANTIFICATION AND FEDPROX GAIN ANALYSIS

To quantitatively characterize the heterogeneity of multi-regional cultural heritage data and evaluate the actual contribution of FedProx, this paper compares two federated optimization strategies, standard FedAvg and FedProx, while maintaining consistency in model structure (ST-GAT), privacy budget, and training hyperparameters. The specific experimental steps are as follows: First, for 21 regional clients, non-independent identically distributed (IID) indices were calculated on three dimensions: risk level label distribution, statistical distribution of key environmental characteristics (temperature, humidity/pollutants), and the proportion of textile material and degradation type. These indices included Jensen–Shannon Divergence (JSD) and Earth Mover's Distance (EMD) to quantify the distribution differences between regions. Then, under the same non-IID intensity setting (controlled by the Dirichlet parameter α), FedAvg and FedProx were used for federated training, with each configuration repeated 10 times and the average performance reported. Finally, the differences between the two methods in regression, classification, and cross-regional fairness indices were compared to evaluate the actual stabilization effect of FedProx on this multi-regional cultural heritage dataset. The performance comparison of FedAvg and FedProx under the same non-IID level is shown in Table 4.

Table 4 quantifies the degree of non-independent and identically distributed nature of multi-regional cultural heritage data and compares the performance differences between FedAvg and FedProx under different levels of heterogeneity. It can be seen that as the Dirichlet parameter α decreases from 1.0 to 0.1, the average JSD increases from 0.021 to 0.186, and the average EMD increases from 0.034 to 0.243, indicating a significant increase in the distribution differences among regions in risk level, environmental conditions, and textile material and degradation type, posing a greater challenge to federated training.

Under near-IID conditions (α=1.0), FedAvg and FedProx perform similarly, with RMSEs of 0.114 and 0.113 respectively, and accuracies of 96.55% and 96.72%, respectively. The difference in RMSE variance between regions is very small (0.0021 vs. 0.0019). However, under moderate to strong non-IID conditions, FedProx's advantages gradually become apparent: when α=0.5, FedProx reduces RMSE from 0.129 to 0.121 (a reduction of 0.008), improves accuracy by 0.92%, and compresses the inter-region RMSE variance from 0.0068 to 0.0049; under stronger heterogeneity α=0.3, FedProx's RMSE advantage over FedAvg expands to 0.016, with an accuracy improvement of 1.53%. In the extreme non-IID scenario (α=0.1), FedProx still maintains a significant stability advantage, with RMSE decreasing from 0.198 to 0.172 (a reduction of 0.026) and accuracy improving by 2.38%. At the same time, it reduces the cross-regional error variance from 0.0213 to 0.0137, significantly alleviating the client update drift problem caused by regional differences in textile types and uneven environmental conditions. This verifies the effectiveness of FedProx in handling highly heterogeneous data in multi-regional cultural heritage protection scenarios.

TABLE 4. Performance Comparison of FedAvg and FedProx under Different Non-IID Levels

Non-IID Intensity (Dirichlet α)	Mean JSD	Mean EMD	Method	RMSE	Accuracy (%)	Inter-region RMSE Variance
1.0 (Near IID)	0.021	0.034	FedAvg	0.114	96.55	0.0021
1.0 (Near IID)	0.021	0.034	FedProx	0.113	96.72	0.0019
0.5 (Moderate Heterogeneity)	0.067	0.091	FedAvg	0.129	94.10	0.0068
0.5 (Moderate Heterogeneity)	0.067	0.091	FedProx	0.121	95.02	0.0049
0.3 (Strong Heterogeneity)	0.112	0.156	FedAvg	0.158	90.84	0.0125
0.3 (Strong Heterogeneity)	0.112	0.156	FedProx	0.142	92.37	0.0081
0.1 (Strong Non-IID)	0.186	0.243	FedAvg	0.198	82.96	0.0213
0.1 (Strong Non-IID)	0.186	0.243	FedProx	0.172	85.34	0.0137

V. CONCLUSION

This paper explores the common problems of data silos and privacy constraints in multi-regional cultural heritage protection, particularly the practical needs of historical textiles in environmental risk assessment and technological knowledge protection. A federated learning-driven risk modeling and strategy generation framework is proposed. This method models environmental time series data in each regional client and integrates spatial correlation information under a privacy-preserving federated training mechanism to achieve stable predictions of textile degradation risks. Multi-objective protection strategies are generated and evaluated through federated Bayesian inference and Monte Carlo sampling to guide resource allocation and intervention priority determination. Experimental results show that the framework achieves an RMSE of 0.113 and a classification accuracy of 96.72% in the risk regression task. In the cross-regional setting, it maintains an average RMSE of 0.118 and an accuracy of 94.47%. Furthermore, under differential privacy constraints, it achieves a risk mitigation rate of 95.45% and a resource allocation efficiency of 91.2%, indicating that the model possesses good generalization performance and decision-making reference value without sharing original data. It should be noted that current experiments mainly rely on structured indicators, which still have limited ability to characterize the long-term evolution of textile aging, dyeing fading, and structural fragility under complex conditions. Future research will introduce protective record texts, historical video materials, and expert knowledge on traditional dyeing and weaving processes, and combine them with actual regional management and feedback from the textile industry to further improve the model and strategy generation mechanism.

AUTHOR CONTRIBUTIONS

Conceptualization –Xiaona Xie and Wenliang Zhang; methodology – Xiaona Xie and Wenliang Zhang; investigation – Xiaona Xie and Wenliang Zhang; writing-original draft preparation – Xiaona Xie and Wenliang Zhang. 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] A. J. Folarin, "Management of cultural monuments site: Protection, preservation and rehabilitation of art work and monuments," UNIZIK Journal of Educational Research and Policy Studies, vol. 18, no. 2, pp. 426-434, 2024, [Online]. Available: https://www.unijerps.org/index.php/unijerps/article/view/787.

[2] M. Rossi and D. Bournas, "Structural health monitoring and management of cultural heritage structures: a state-of-the-art review," Applied Sciences, vol. 13, no. 11, pp. 6450-6486, 2023, doi: 10.3390/app13116450.

[3] N. Laohaviraphap and T. Waroonkun, "Integrating Artificial Intelligence and the Internet of Things in Cultural Heritage Preservation: A Systematic Review of Risk Management and Environmental Monitoring Strategies," Buildings, vol. 14, no. 12, pp. 3979-4006, 2024, doi: 10.3390/buildings14123979.

[4] I. Siliutina, O. Tytar, M. Barbash, N. Petrenko, and L. Yepyk, "Cultural preservation and digital heritage: Challenges and opportunities," Amazonia Investiga, vol. 13, no. 75, pp. 262-273, 2024, doi: 10.34069/AI/2024.75.03.22.

[5] J. Liu, "Digitally Protecting and Disseminating the Intangible Cultural Heritage in Information Technology Era," Mobile Information Systems, vol. 2022, no. 1, pp. 1-10, 2022, doi: 10.1155/2022/1115655.

[6] Y. Lei, Z. Shen, F. Tian, X. Yang, F. Wang, R. Pan, et al., "Fire risk level prediction of timber heritage buildings based on entropy and XGBoost," Journal of Cultural Heritage, vol. 63, no. 1, pp. 11-22, 2023, doi: 10.1016/j.culher.2023.06.024.

[7] M. Moreno, R. Ortiz, D. Cagigas-Muniz, J. Becerra, J. M. Martin, A. J. Prieto, et al., "ART-RISK 3.0 a fuzzy—based platform that combine GIS and expert assessments for conservation strategies in cultural heritage," Journal of Cultural Heritage, vol. 55, no. 1, pp. 263-276, 2022, doi: 10.1016/j.culher.2022.03.012.

[8] H. Mekonnen, Z. Bires, and K. Berhanu, "Practices and challenges of cultural heritage conservation in historical and religious heritage sites: Evidence from North Shoa Zone, Amhara Region, Ethiopia," Heritage Science, vol. 10, no. 1, pp. 1-22, 2022, doi: 10.1186/s40494-022-00802-6.

[9] L. Che, J. Wang, Y. Zhou, and F. Ma, "Multimodal federated learning: A survey," Sensors, vol. 23, no. 15, pp. 6986-7006, 2023, doi: 10.3390/s23156986.

[10] H. Sui, X. Sun, J. Zhang, B. Chen, and W. Li, "Multi-level membership inference attacks in federated learning based on active GAN," Neural Computing and Applications, vol. 35, no. 23, pp. 17013-17027, 2023, doi: 10.1007/s00521-023-08593-y.

[11] C. Wang, H. Shi, B. Song, L. Cai, and L. Wu, "Hierarchical Weighted LSTM with One-class Classifier for Preventive Protection of Cultural Heritage in Museums," ACM Journal on Computing and Cultural Heritage, vol. 18, no. 1, pp. 1-20, 2025, doi: 10.1145/3703633.

[12] Y. Wu, Y. Dong, Z. Shan, X. Meng, Y. He, P. Jia, et al., "Enhancing anomaly detection for cultural heritage via long short-term memory with attention mechanism," Electronics, vol. 13, no. 7, pp. 1254-1274, 2024, doi: 10.3390/electronics13071254.

[13] A. Ramirez-Arellano, E. M. Munoz-Silva, M. Antonio-Cruz, and J. Irving Vasquez-Gomez, "Deng entropy and LSTM neural network to classify built cultural heritage with severe damage," Journal of Cultural Heritage, vol. 73, no. 1, pp. 286-294, 2025, doi: 10.1016/j.culher.2025.04.003.

[14] Y. Wang, J. Liu, W. Wang, J. Chen, X. Yang, L. Sang, et al., "Construction of cultural heritage knowledge graph based on graph attention neural network," Applied Sciences, vol. 14, no. 18, pp. 8231-8258, 2024, doi: 10.3390/app14188231.

[15] A. G. Vrahatis, K. Lazaros, and S. Kotsiantis, "Graph attention networks: A comprehensive review of methods and applications," Future Internet, vol. 16, no. 9, pp. 318-351, 2024, doi: 10.3390/fi16090318.
[16] M. Reyad, A. M. Sarhan, and M. Arafa, "A modified Adam algorithm for deep neural network optimization," Neural Computing and Applications, vol. 35, no. 23, pp. 17095-17112, 2023, doi: 10.1007/s00521-023-08568-z.
[17] Y. Zhang, C. Chen, N. Shi, R. Sun, and Z. Q. Luo, "Adam can converge without any modification on update rules," Advances in Neural Information Processing Systems, vol. 35, no. 1, pp. 28386-28399, 2022, doi: 10.48550/arXiv.2208.09632.
[18] R. Abdulkadirov, P. Lyakhov, and N. Nagornov, "Survey of optimization algorithms in modern neural networks," Mathematics, vol. 11, no. 11, pp. 2466-2503, 2023, doi: 10.3390/math11112466.
[19] Y. Zheng, S. Lai, Y. Liu, X. Yuan, X. Yi, and C. Wang, "Aggregation service for federated learning: An efficient, secure, and more resilient realization," IEEE Transactions on Dependable and Secure Computing, vol. 20, no. 2, pp. 988-1001, 2022, doi: 10.1109/TDSC.2022.3146448.
[20] M. Moshawrab, M. Adda, A. Bouzouane, H. Ibrahim, and A. Raad, "Reviewing federated learning aggregation algorithms; strategies, contributions, limitations and future perspectives," Electronics, vol. 12, no. 10, pp. 2287-2321, 2023, doi: 10.3390/electronics12102287.
[21] L. Luo, C. Zhang, H. Yu, G. Sun, S. Luo, and S. Dustdar, "Communication-efficient federated learning with adaptive aggregation for heterogeneous client-edge-cloud network," IEEE Transactions on Services Computing, vol. 17, no. 6, pp. 3241-3255, 2024, doi: 10.1109/TSC.2024.3399649.
[22] S. Pourdoustmohammadi and R. Ansari, "Developing a Scenario-Based Optimization Model for Planning Risks in Construction Projects by Integrating a Decision Support System with Bayesian Belief Network Analysis Approach: A Case Study in High-Rise Buildings," Iranian Journal of Science and Technology, Transactions of Civil Engineering, vol. 49, no. 3, pp. 2779-2801, 2025, doi: 10.1007/s40996-024-01590-8.
[23] R. Norouzi Isfahani, A. Talaee Malmiri, A. BahooToroody, and M. M. Abaei, "A Bayesian-based framework for advanced nature-based tourism model," Journal of Asian Business and Economic Studies, vol. 30, no. 2, pp. 86-104, 2023, doi: 10.1108/JABES-11-2020-0119.
[24] A. V. D. Sano, A. A. Stefanus, E. D. Madyatmadja, H. Nindito, A. Purnomo, and C. P. M. Sianipar, "Proposing a visualized comparative review analysis model on tourism domain using Naïve Bayes classifier," Procedia Computer Science, vol. 227, no. 1, pp. 482-489, 2023, doi: 10.1016/j.procs.2023.10.549.