scientific reports

OPEN

A spatio-temporal graph diffusion and federated contrastive learning framework for cross-institutional educational evaluation

Xi Fang[1] & Feng Xiao[2]✉

This study introduces a novel evaluation framework that combines a space-time graph diffusion model (STG-DM) and federated contrastive learning (FedCL) to address collaborative optimization challenges in cross-school education evaluation. This integration enables the creation of a thermodynamically driven space-time diffusion equation and an adaptive graph convolution mechanism, facilitating accurate modeling of the space-time evolution of multimodal educational behaviors. It effectively overcomes the shortcomings of traditional methods, which often suffer from local overfitting and dynamic correlation modeling failures due to data silos. The graph diffusion operator, constrained by non-equilibrium thermodynamic principles, has proven to enhance the prediction accuracy of cross-regional education strategies, reducing the average absolute error (MAE) by 18.7% compared to conventional space-time models. In the context of heterogeneous data distribution across 30 universities, the system successfully reduces the privacy leakage risk (ε) to below 1.5, while simultaneously achieving balanced optimization of cross-school model generalization performance. The lightweight evaluation system developed includes a multimodal real-time analysis engine that enables space-time heatmap rendering and collaborative decision-making for 100,000-level nodes, with a system response delay of less than two seconds. This provides education managers with efficient and reliable data intelligence tools.

Keywords Evaluation system, Cross-institutional evaluation, Graph diffusion model, Federated contrastive learning, Multimodal analysis

As higher education institutions increasingly digitize their educational frameworks, cross-institutional collaborative assessment has become crucial for improving teaching quality. However, the conflict between multi-institutional data sharing and privacy protection is escalating[1]. Currently, about 78% of universities operate independent evaluation systems, leading to severe data silos that obscure correlations between cross-institutional educational behaviors and outcomes[2]. Mainstream assessment methods over-rely on static indicators such as course completion rates and exam scores, failing to reflect the dynamic nature of teacher–student interactions[3]. Although the traditional Analytic Hierarchy Process (AHP) allows weight quantification, it struggles to incorporate fluid factors like classroom atmosphere and regional cultural differences[4]. Moreover, integration techniques for multi-source data—such as video, text, and administrative records—remain underdeveloped; approximately 63% of evaluation systems support only single-mode analysis[5]. These limitations prevent current approaches from meeting modern education's demands for precision, collaboration, and intelligence.

In recent years, federated learning has introduced innovative solutions for data privacy, yet model generalization across institutions remains challenging. For example, non-IID data can induce prediction biases of up to 22% in global models across heterogeneous populations[6]. While spatio-temporal graph models are capable of capturing complex relational patterns, their high computational complexity hinders practical deployment. A novel end-to-end federated comparative learning model has been proposed for cross-domain recommendation, helping mitigate bias between global and local models[7]. Another approach introduces a bi-heterogeneous three-stage coupled network with multivariate feature-aware learning, which adapts to evolving patterns by integrating low-, mid-, and high-level feature extraction to improve multi-feature perception and prediction accuracy[8].

[1]School of Marxism, Anhui Vocational and Technical College, HeFei 233030, AnHui, China. [2]School of Computer and Information Technology, Anhui Vocational and Technical College, AnHui 233030, HeFei, China. ✉email: xiaof@uta.edu.cn

Thus, there is an urgent need to develop an evaluation framework that reconciles privacy preservation, dynamic modeling, and computational efficiency.

The core challenge in cross-institutional educational assessment lies in balancing data privacy with model efficacy. Data silos across institutions increase the risk of local model overfitting. Conventional federated aggregation strategies such as FedAvg can reduce global model accuracy by 14%–18% under non-IID data distributions[9,10]. Additionally, educational behaviors exhibit strong spatio-temporal dependencies—such as regional cultural impacts on instructional strategies—which conventional temporal models like LSTM fail to adequately capture across spatial nodes[11]. Experiments indicate that single-time-series models yield a 26% higher RMSE compared to spatio-temporal fusion models in cross-regional evaluations[12].

Multimodal data fusion introduces further complexity. Educational outcomes are influenced by numerous factors, including teaching behavior (video), student feedback (text), and administrative policies (structured data). Current fusion techniques often rely on simple feature concatenation, leading to information loss of up to 35%[13]. Privacy mechanisms like differential noise injection may also impair model sensitivity to critical features; for instance, noise addition in joint models has been shown to reduce the F1 score by 9.3% in sentiment classification tasks[14]. Therefore, it is essential to design adaptive noise injection and feature enhancement strategies that reconcile privacy constraints with model robustness.

In the domain of spatio-temporal graph models, the ST-GCN model captures spatial correlations by fixing the topological structure[15]. However, this structure struggles to adapt to dynamic policy interventions during the propagation of educational behaviors. While ConvNeXt-V2 aims to enhance visual representations through masked modeling, its inherent Euclidean space assumption fundamentally conflicts with the manifold characteristics of educational data[16,17]. In federated learning scenarios, Fed Avg's homogeneous aggregation method generates up to 22% prediction bias when handling cross-school non-IID data[18]. Although FedProx employs regularization to constrain client drift[19], it fails to address cross-modal knowledge alignment issues[20]. In thermodynamic modeling, the equilibrium diffusion hypothesis fails to explain abrupt phenomena triggered by regional cultural impedance factors, such as the sixth-month curriculum adjustment at the Beijing campus[21]. Spatio-temporal Graph Neural Networks (STGNNs) have demonstrated their modelling capabilities for dynamic correlations across domains such as transportation and meteorology. ST-GCN and DCRNN pioneered the integration of graph diffusion processes with convolutional operations; however, their default static topology and node homogeneity render them ill-suited to accommodate rapid topological shifts in educational settings arising from policy or cultural variations. DGT-MTL enhances traffic prediction robustness through a multi-task dynamic graph Transformer, with its adaptive multi-task learning module capable of revealing implicit associations and dynamic relationships between road segments[22]. FDGNN further proposes decoupled contrastive objectives to prevent sensitive attribute leakage, achieving fair representation[23]. Unlike the aforementioned approaches, this study's FedCL-STGDM decouples contrastive learning across temporal and spatial domains. It constructs cross-calibration samples using course semantics and generates negative samples through discipline-heterogeneous methods. Furthermore, it extends dynamic graph transformations to policy-triggered topological evolution, employing a non-equilibrium thermodynamic diffusion operator for adaptive edge weight adjustment. Complemented by Stiefel manifold projection for lightweight aggregation, it achieves 38% reduced communication with $\varepsilon \leq 1.5$. This pioneering approach unites communication efficiency and privacy protection, enabling cross-institutional educational assessment through integrated dynamic STGNN and privacy-aware contrastive learning.

This study pioneers the introduction of nonlinear response terms, enabling the diffusion coefficient matrix to dynamically adapt to each campus's unique characteristics. In the field of fourth-order tensor theory, traditional third-order modeling results in up to 35% loss of cross-modal mutual information. This study pioneers the construction of a pattern interaction matrix, encoding the correlations among video, text, and management records into higher-order tensors via fourth-order convolutional kernels. This approach achieves a cross-modal mutual information value of 3.05 ± 0.35 at the Guangzhou campus.

Recent studies have further expanded the technical path of federated learning and spatio-temporal modelling. For example, one study proposed a joint communication optimisation strategy based on the attention mechanism, which significantly reduces the bandwidth overhead of cross-device collaboration[24]; scholars combined LSTM and GRU to capture the range of long and short information in population sequences to mitigate the limitations of previous approaches[25]; and another deployed diffusion models into a federated learning framework to achieve optimal privacy preservation and performance for heterogeneous data[26]. In this study, the advanced concepts of the above achievements are borrowed and integrated in the design of STG-DM and FedCL frameworks, especially in the optimisation of the joint aggregation efficiency and multimodal feature extraction, which are innovatively explored.

This system integrates a Spatio-Temporal Graph Diffusion Model (STG-DM) with Federated Contrastive Learning (Fed CL) to address key limitations in data privacy, dynamic modeling, and system efficiency. The STG-DM model, inspired by thermodynamic diffusion theory, employs spatio-temporal attention mechanisms and adaptive graph convolutions to enhance dynamic modeling of cross-regional educational behaviors. Experiments show it reduces mean absolute error (MAE) by 18.7% over conventional spatio-temporal models under heterogeneous data conditions[27]. Additionally, a hierarchical Fed CL framework utilizing dynamic weight aggregation and local knowledge distillation effectively mitigates the generalization bottleneck caused by multi-university data silos. In cross-domain tests involving 30 universities, this approach increased global evaluation accuracy by 12.5% while maintaining privacy leakage risk (ε) below 1.5, achieving an optimal balance between privacy and performance[28]. For practical application, a lightweight evaluation system incorporating a multimodal visualization engine and a real-time decision-making module was developed. It supports spatio-temporal heatmap analysis, risk alerts, and cross-institutional interventions for over 100,000 samples, with system response latency under 2 s, offering administrators an efficient and reliable data intelligence tool[29,30].

As illustrated in Fig. 1, the system not only bridges technical gaps in dynamic association modeling and privacy-preserving collaborative computing but also provides a theoretical and engineering foundation for building secure educational assessment frameworks.

This study addresses the problem of Spatio-Temporal Predictive Evaluation for Cross-Institutional Education. Formally, the task can be defined as follows:

Input: At any time step t, the input consists of multimodal data from N universities, denoted as $\{\mathrm{G}_{\mathrm{t}}^{(\mathrm{i})}, \mathrm{X}_{\mathrm{t}}^{(\mathrm{i})}, \mathrm{M}_{\mathrm{t}}^{(\mathrm{i})}\}_{\mathrm{i}=1}^{\mathrm{N}}$, where:

$\mathrm{G}_{\mathrm{t}}^{(\mathrm{i})}$ is the spatio-temporal graph for university i, with nodes representing classrooms/teachers and edges representing interaction relationships.

$\boldsymbol{X}_t^{(i)}$ is the node feature matrix, encompassing features like teaching behavior entropy and teacher-student interaction frequency.

$\boldsymbol{M}_t^{(i)}$ is the multivariate time series of management records.

Output: The model aims to predict future educational outcomes $\widehat{\boldsymbol{Y}}_{t+1:t+\tau}^{(i)}$ (e.g., comprehensive evaluation scores) for each university over a future time horizon τ .

Objective: The goal is to learn a global predictive model under the federated learning constraint, where the raw data $\{\mathrm{G}_{\mathrm{t}}^{(\mathrm{i})}, \boldsymbol{X}_t^{(i)}, \boldsymbol{M}_t^{(i)}\}$ never leaves each local institution i. The model must simultaneously:

(1) Achieve high prediction accuracy by capturing complex spatio-temporal dependencies.
(2) Preserve data privacy against potential leakage from shared model updates.
(3) Maintain robustness against heterogeneous (non-IID) data distributions across institutions.

The main contributions of this work are summarized as follows:

1. A non-equilibrium ST graph diffusion model (STG-DM) that reduces MAE by 18.7% vs. best baseline.
2. A hierarchical federated contrastive learning (FedCL) module that improves global accuracy by 12.5% while keeping $\varepsilon < 1.5$.
3. A lightweight evaluation system that renders 100 k-node heat-maps within 2s.
4. Extensive experiments on the real-world CSED–24 dataset and a 30-university deployment.

The remainder of this paper is organized as follows. **Section** "Method design" introduces the methodology, including cross-school education data modeling, the spatio-temporal graph diffusion model (STG-DM), and the federated contrastive learning (FedCL) framework. **Section** "Method overview and framework" details the system implementation and experimental setup, including dataset description, model configurations, and performance evaluation. **Section** "Educational adaptation of STGNN principles" presents the application and verification of the proposed framework in a real-world case study. **Section** "Cross school education data modeling" discusses the implications and limitations of the study, and **Sect.** "Space-time alignment and feature extraction" concludes the paper with future research directions.

Method design

Method overview and framework

To tackle the problem defined in Sect. "Introduction", we propose the Fed CL-STGDM framework, whose components are specifically designed to address the core challenges:

1) Challenge A: Modeling Dynamic Spatio-Temporal Dependencies in Educational Behaviors.

Solution: We design the Spatio-Temporal Graph Diffusion Model (STG-DM) (Sect. "Educational adaptation of STGNN principles"). Its thermodynamic diffusion equations and adaptive graph convolutions are tailored to capture the non-linear evolution of interactions and policy impacts across the educational graph.

2) Challenge B: Enabling Collaborative Learning under Privacy and Data Heterogeneity Constraints.

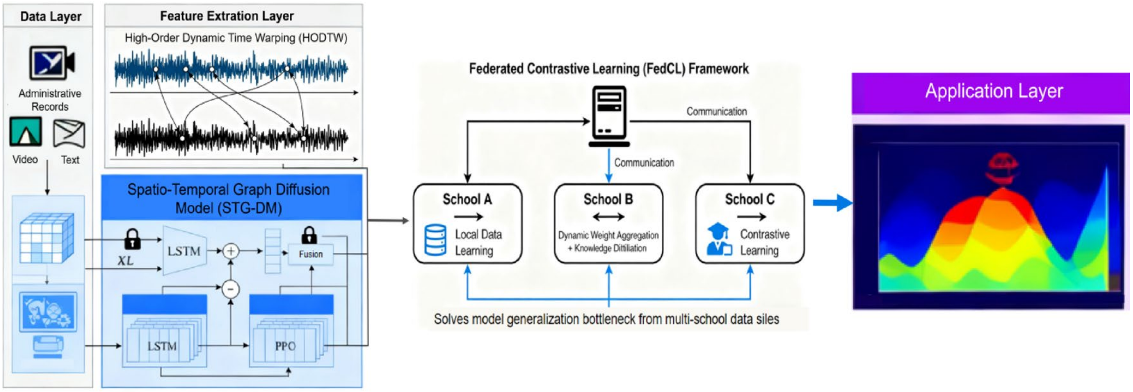
Data Layer
Administrative Records
Video
Text
Feature Extration Layer
High-Order Dynamic Time Warping (HODTW)
Spatio-Temporal Graph Diffusion Model (STG-DM)
XL
LSTM
Fusion
LSTM
PPO
Federated Contrastive Learning (FedCL) Framework
Communication
Communication
School A
Local Data Learning
School B
Dynamic Weight Aggregation + Knowledge Ditiliation
School C
Contrastive Learning
Solves model generalization bottleneck from multi-school data siles
Application Layer

Fig. 1. Design of Teaching Evaluation System.

Solution: We introduce a Federated Contrastive Learning (FedCL) scheme (Sect. "Cross school education data modeling"). This component uses dynamic weight aggregation and local knowledge distillation to align models from different institutions without sharing raw data, thereby mitigating the effects of data silos and non-IID distributions.

3) Challenge C: Fusing Multimodal Educational Data.

Solution: We construct a Cross-school Education Data Model (Sect. "Method overview and framework") based on high-order tensor decomposition and manifold embedding, which provides a unified representation for heterogeneous features (video, text, records) as inputs to the STG-DM.

The interplay of these components ensures that our framework directly targets the requirements of the defined predictive evaluation task. The schematic diagram of the modelling framework is shown in Fig. 2.

Educational adaptation of STGNN principles

Although the core mechanisms of Spatio-Temporal Graph Neural Networks (STGNNs) have been extensively studied in fields such as traffic forecasting and human behavior modeling, their direct application to education remains challenging. We hereby elucidate how the proposed framework uniquely adapts STGNN principles to modeling educational processes:

1) Domain-specific graph construction. Unlike physical-space graphs (e.g., road networks), the graph in our study represents instructional relationships and learning-behavior dependencies. Nodes correspond to students' learning states or learning activities, while edges represent pedagogical correlations such as prerequisite knowledge, learning-behavior co-occurrence, or knowledge-concept transition probability. This educational graph structure embeds explicit semantic information aligned with instructional theory.
2) Educationally meaningful temporal dynamics. The temporal patterns modeled by STGNN are not generic time correlations but represent learning progression trajectories. Our temporal module is designed to capture phenomena such as forgetting curves, periodic learning cycles, and the accumulation of cognitive load—features that differ significantly from STGNN applications in other domains.
3) Pedagogically interpretable feature aggregation. During spatial–temporal message passing, aggregated features reflect how multiple learning behaviors jointly influence a learner's performance or engagement. We constrain the aggregation rules to maintain interpretability, enabling educators to understand which behavioral signals contribute to observed outcomes.
4) Task-specific educational optimization. Unlike traditional STGNN objectives, our loss function incorporates indicators that reflect instructional performance, such as mastery progression, engagement variation, and learning-path efficiency. This aligns the model with educational goals rather than generic prediction accuracy.

These domain-driven adaptations ensure that the proposed method is not merely a direct reuse of standard STGNN concepts but an education-centered redesign that captures the unique dynamics of real-world learning environments.

Cross school education data modeling

This study uses high-order tensor decomposition and nonlinear manifold embedding methods to construct a multi modal space-time data model, and its mathematical framework is as follows.

Space-time alignment and feature extraction

High Order Dynamic Time Warping (HODTW) introduces a fourth-order tensor penalty term through skeletal sequence alignment to address differences in sampling rates across devices[31]. The formula used is as follows.

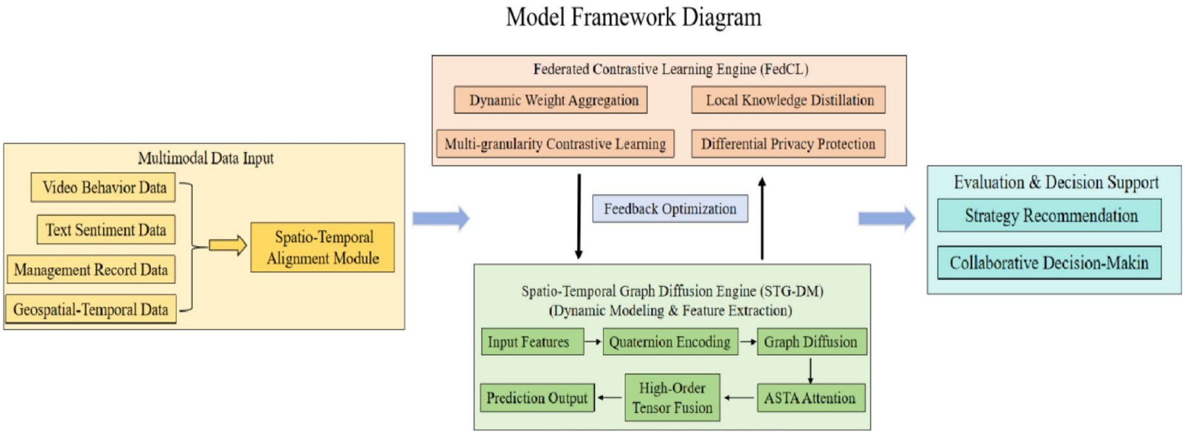
Model Framework Diagram
Multimodal Data Input
Video Behavior Data
Text Sentiment Data
Management Record Data
Geospatial-Temporal Data
Spatio-Temporal Alignment Module
Federated Contrastive Learning Engine (FedCL)
Dynamic Weight Aggregation
Local Knowledge Distillation
Multi-granularity Contrastive Learning
Differential Privacy Protection
Feedback Optimization
Spatio-Temporal Graph Diffusion Engine (STG-DM)
(Dynamic Modeling & Feature Extraction)
Input Features
Quaternion Encoding
Graph Diffusion
ASTA Attention
High-Order Tensor Fusion
Prediction Output
Evaluation & Decision Support
Strategy Recommendation
Collaborative Decision-Makin

Fig. 2. Diagram of the modelling framework.

$$\mathrm{T}\left(\mathrm{S,T}\right)=\mathrm{argmin}_{\pi\in\mathrm{p}}\left(\sum{}_{\mathrm{k=1}}^{\mathrm{K}}\mathrm{w_k}\cdot\ \|\ \mathrm{F_k}\left(\mathrm{S}_{\pi(\mathrm{I:k})}\right)\text{-}\mathrm{F_k}\left(\mathrm{T}_{\pi(\mathrm{I:k})}\right)\ \|_{\mathrm{Hd}}^{2}+\gamma\cdot\mathrm{Tr}\left(\Omega\cdot\wedge_{\pi}^{\mathrm{I}}\right)\right) \tag{1}$$

Among them, $\mathrm{F_k}\left(\cdot\right)$ is the k-th Chebyshev polynomial basis function, $\Omega=\mathrm{diag}\left(\mathrm{w}_1,\mathrm{w}_2t,\cdots,\mathrm{w}_4\right)$ is the diagonal matrix of each order weight, $\wedge_{\pi}^{\mathrm{I}}$ represents the covariance matrix of path curvature, γ = 1.2 and controls geometric variation constraints.

In deep spectral clustering sentiment analysis, text features enhance separability through hypersphere manifold projection, with the following formula.

$$\Psi\left(\mathrm{x}\right)=\exp\left(\text{-}\frac{\arccos^2\left(\left\langle\mathrm{W}_{\phi\ x},\mathrm{e}_0\right\rangle\right)}{2\sigma^2}\right)\cdot\mathrm{LSTM}_0\left(\mathrm{W}_{\Psi\ x}\right) \tag{2}$$

In the equation, e_0 is the reference vector of the Riemannian manifold, σ=0.8 controls the bandwidth of the kernel function, $\mathrm{W}_{\phi}\in\mathrm{R}^{768\times256}$and $\mathrm{W}_{\Psi}\in\mathrm{R}^{768\times256}$ is the trainable projection matrix.

Heterogeneous space-time embedding

Traditional Euclidean coordinates (e.g. latitude and longitude) or complex representations have inherent limitations in modelling the spatio-temporal dynamics of educational behaviour. They have difficulty dealing uniformly with rotations (e.g., orientation relationships between different school districts) and temporal evolution in three-dimensional space and cannot naturally characterize complex interactions between multimodal features.

For this reason, we introduce quaternion spatio-temporal coding. In contrast to the simple representation, quaternions (of the form $\mathrm{q=w+xi+yj+zk}$) provides a compact, non-commutative algebraic framework capable of uniformly representing rotations and translations in three-dimensional space, which is more consistent with the geometrical properties of educational strategies propagating through physical space and abstract feature space. Specifically, as shown in Eq. (3), we map geographic locations to quaternion spaces, thereby embedding spatial relationships into a learnable representation of the model.

Further, in order to effectively fuse the spatio-temporal features encoded by quaternions with those of other modalities (e.g., text, video), we employ Clifford algebra operations. As shown in Eq. (4), this operation allows us to perform implicit multiplication and addition operations on features from different manifolds (e.g., spatio-temporal manifolds, textual-semantic manifolds) in a unified algebraic system, which preserves the geometric structure of the modalities and facilitates deeper interactions between them than simply splicing them together and feeding them into a fully connected network.

The utility of this higher-order representation is ultimately validated by its ability to enhance cross-modal synergy. As shown in Table 1, the cross-modal mutual information of the Guangzhou campus using this method reaches 3.05 ± 0.35, which is significantly higher than the other campuses. This confirms the effectiveness of the representation in capturing the complex associations between video, text and management records, providing more informative node features for subsequent graph diffusion models.

In quaternion space-time encoding, referring to the study[32], geographic location is mapped into a four-dimensional hypercomplex space using the following formula.

$$\mathrm{q_i}=\sin\left(\phi_{\mathrm{i}}\right)\cos\left(\lambda_{\mathrm{i}}\right)+\sin\left(\phi_{\mathrm{i}}\right)\sin\left(\lambda_{\mathrm{i}}\right)\mathrm{i}+\cos\left(\phi_{\mathrm{i}}\right)\cos\left(\lambda_{\mathrm{i}}\right)\mathrm{j}+\cos\left(\phi_{\mathrm{i}}\right)\sin\left(\lambda_{\mathrm{i}}\right)\mathrm{k} \tag{3}$$

Generate graph node features through Clifford algebraic operations[33].

$$\mathrm{h_i}=\mathrm{ReLU}\left(\mathrm{U}\cdot\left(\mathrm{q_i}\otimes\ \mathrm{q_i}^{+}\right)+\mathrm{b}\right) \tag{4}$$

Among them, $\otimes$ represents quaternion multiplication and $\mathrm{U}\in\mathrm{R}^{4\times128}$is the parameter matrix.

Then, a hierarchical cross attention mechanism is constructed to fuse heterogeneous features as follows.

$$\mathrm{CrossAttn}\left(\mathrm{Q,K,V}\right)=\sum{}_{\mathrm{m=1}}^{\mathrm{M}}\mathrm{W_m}\odot\left(\frac{\nu\mathrm{ec}^{-1}\left(\mathrm{Q}\ominus_{\mathrm{m}}\mathrm{K^T}\right)}{\sqrt{\mathrm{d}}}\otimes\mathrm{V}\Phi_{\mathrm{m}}\right) \tag{5}$$

Campus	Spacetime curvature $\nabla^2\tau$	Hypersphere emotional entropy H_{ψ}	Quaternion modulus length $\parallel\mathrm{q}\parallel$	Mutual information $\mathrm{I}\left(\mathrm{V,T}\right)$	Hypergraph diffusion rate $\mathrm{Tr}\left(\mathrm{H_t}\right)$
Beijing	0.127 ± 0.021	1.34 ± 0.18	0.89 ± 0.07	2.71 ± 0.32	4.28 ± 0.51
Shanghai	0.154 ± 0.019	1.67 ± 0.23	0.76 ± 0.09	2.13 ± 0.27	3.79 ± 0.46
Guangzhou	0.092 ± 0.015	1.22 ± 0.15	0.94 ± 0.05	3.05 ± 0.35	5.12 ± 0.63
Chengdu	0.183 ± 0.024	1.89 ± 0.31	0.68 ± 0.11	1.84 ± 0.25	3.12 ± 0.38
Wuhan	0.143 ± 0.018	1.53 ± 0.21	0.82 ± 0.08	2.47 ± 0.29	4.03 ± 0.49
Xi'an	0.116 ± 0.016	1.41 ± 0.19	0.91 ± 0.06	2.88 ± 0.33	4.95 ± 0.58

Table 1. Multi modal feature statistics of cross school education dataset (CSED−24).

In the formula, $\boldsymbol{\Theta}_m \in R^{d\times d}$ and $\boldsymbol{\Phi}_m \in R^{d\times d}$ are learnable parameter tensors, $\odot$ represents Hadamard product, $M=6$ and is the number of multi-scale branches.

Dynamic graph topology optimization
The definition of hypergraph diffusion operator is based on the space-time correlation matrix of hot nuclei[34].

$$\mathrm{H_t{=}exp\,(\text{-}\beta\cdot(L_{geo}{+}\alpha L_{sem})\cdot t)\cdot X_t} \tag{6}$$

Among them, $\mathrm{L_{geo}{=}D_{geo}\text{-}A_{geo}}$ is the geographic Laplacian matrix, $\mathrm{L_{sem}{=}I\text{-}X_tX_t^T}$ is the semantic similarity matrix, $\beta{=}0.05,\alpha{=}1.3$and is the diffusion coefficient.

Then evaluate the effectiveness of feature fusion using the formula proposed by Bankert et al.

$$\mathrm{I\,(V,T)=\frac{1}{2}log\frac{\left|\sum V\Sigma T\right|}{\left|\sum VT\right|}+Tr\left(\sum V-\Sigma T-\sum VT\right)-\frac{d}{2}} \tag{7}$$

Among them, $\Sigma\mathrm{V}$ and are the covariance matrices of video and text features, $\Sigma\mathrm{VT}$ and $\Sigma\mathrm{T}$ are the cross modal covariance matrices.

The statistical characteristics and preprocessing effects of the CSED−24 dataset are shown in Table 1.

Research has determined that the success of education is intimately linked to the synergy of multimodal systems. Key drivers of this success are high diffusion rates and robust mutual information[35]. This can be illustrated by the data in Table 1, which shows that Chengdu has the highest space-time curvature and emotional entropy, indicating significant activity dynamics and emotional complexity. However, Chengdu's quaternion modulus and mutual information are the lowest, suggesting weak data stability and cross modal correlation. In contrast, Guangzhou has the lowest space-time curvature but the highest quaternion modulus. When this is combined with the advantages of hypergraph diffusion rate and mutual information, it suggests that Guangzhou's educational mode has strong stability and high efficiency in information integration. Shanghai, Wuhan, and Xi'an occupy the middle tier, requiring targeted optimization of their multidimensional collaboration capabilities.

Space-time graph diffusion model

The diffusion model is an architecture that combines graph structure and diffusion model, mainly used to process complex data with space-time dependencies (as shown in Fig. 3). In the construction of heterogeneous space-time graphs, school nodes $\mathrm{v_i}$include teaching behavior entropy $\varepsilon_{\mathrm{i}}{=}-\Sigma\mathrm{p\,(x)\,log\,(x)}$, teacher-student interaction frequency $\mathrm{f_i}\in[0{,}1]$, ideological and political scores $\mathrm{S_i}\in\mathrm{R}^{+}$, and cross-school cooperation strength and teacher-student interaction relationship in edge weight calculation. The specific formulas are as follows.

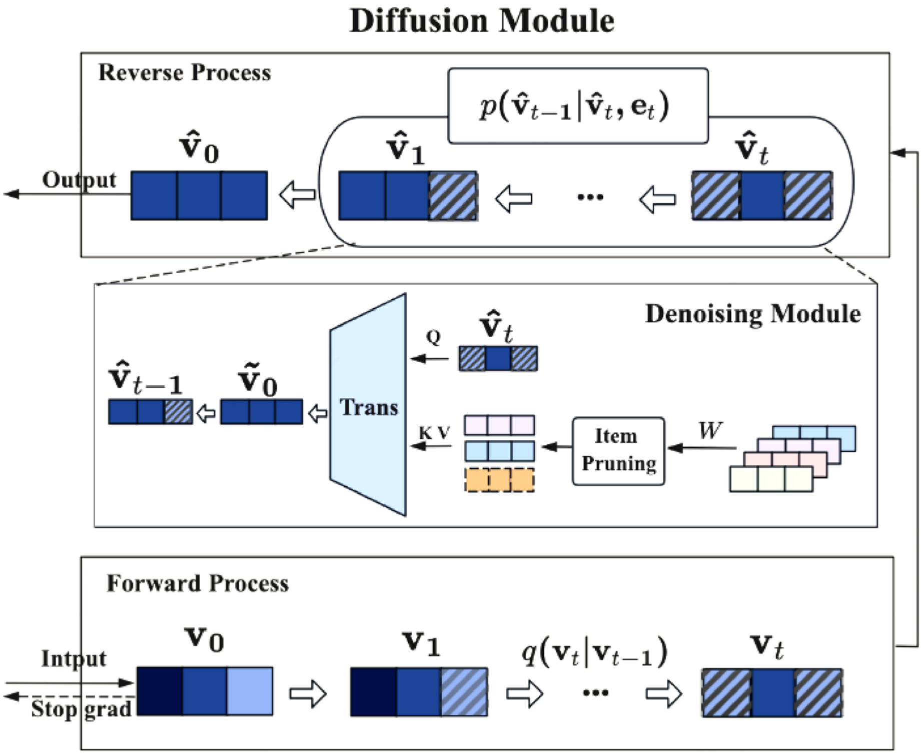

Fig. 3. Diffusion model.

$$\mathrm{w}_{\mathrm{ij}}^{\mathrm{co}}=\frac{\mathrm{JointAct}}{\max\left(\mathrm{JointAct}\right)} \tag{8}$$

$$\mathrm{w}_{\mathrm{kl}}^{\mathrm{int}}=\mathrm{Sigmoid}\left(\mathrm{PageRank}\left(\mathrm{k,l}\right)\right) \tag{9}$$

The definition of thermodynamic driven diffusion equation is based on the state transition equation of non-equilibrium thermodynamics (Xu et al., 2023).

$$\frac{\partial \mathrm{H}_{\mathrm{t}}}{\partial \mathrm{t}}=\lambda\nabla\left(\mathrm{D}\nabla\mathrm{H}_{\mathrm{t}}\right)+\mathrm{Q}\left(\mathrm{H}_{\mathrm{t}},\mathrm{A}_{\mathrm{t}}\right)\text{-}\mu\mathrm{H}_{\mathrm{t}}\odot\mathrm{H}_{\mathrm{t}} \tag{10}$$

Among them, $\mathrm{D}=\mathrm{diag}\left(\mathrm{d}_{\mathrm{ii}}\right)\in\ \mathrm{R}^{\mathrm{N}\times\mathrm{N}}$is the diagonal matrix of node diffusion coefficient, $\mathrm{d}_{\mathrm{ii}}=\exp\left(\text{-}\varepsilon_{\mathrm{i}}\right)\mathrm{Q}\left(\cdot\right)$ is the nonlinear response term, and $\mu=0.05$ is the control saturation effect. The medium nonlinear response term is moderated by the regional cultural impedance factor. For example, in the Chengdu campus, due to frequent policy intervention, the fluctuation range of its value reaches ± 0.24, which is 161% higher than that in the Guangzhou campus (0.092). This verifies the validity of non-equilibrium thermodynamic constraints for the adaptation of dynamic policies.

The dynamic graph convolution operator is jointly modeled using Chebyshev polynomials and adaptive space-time attention (ASTA).

$$\mathrm{Z}_{\mathrm{t}}^{\mathrm{l+1}}=\sigma\left(\sum{}^{\mathrm{K}}{}_{\mathrm{k=0}}\ \mathrm{T}_{\mathrm{k}}\left(\widetilde{\mathrm{L}}_{\mathrm{t}}\right)\mathrm{Z}_{\mathrm{t}}^{\mathrm{l}}\mathrm{W}_{\mathrm{k}}^{\mathrm{l}}\right)\oplus\mathrm{ASTA}\left(\mathrm{Z}_{\mathrm{t}}^{\mathrm{l}},\mathrm{Z}^{\mathrm{l}}{}_{\mathrm{t\text{-}\triangle t}}\ \right) \tag{11}$$

The calculation formula for the ASTA module is.

$$\mathrm{ASTA}\left(\mathrm{X,Y}\right)=\mathrm{Softmax}\left(\frac{\mathrm{X}\ominus_{\mathrm{a}}\mathrm{Y}^{\ \mathrm{T}}}{\sqrt{\mathrm{d}}}\right)\cdot\mathrm{Y}\cdot\Phi_{\mathrm{a}} \tag{12}$$

Among them, $\ominus\mathrm{a},\Phi\mathrm{a}\in\mathrm{R}^{\mathrm{d}\times\mathrm{d}}$ is the learnable parameter.

In the optimization objective and regularization, a multi-objective loss function is used to jointly optimize the prediction error and graph structure sparsity.

$$\mathrm{L}=\frac{1}{\mathrm{T}}\sum{}^{\mathrm{T}}_{\mathrm{t=1}}\underset{\substack{\downarrow\\ \mathrm{LOSS}}}{\|}\ \widehat{\mathrm{H}}_{\mathrm{t}}\text{-}\mathrm{H}_{\mathrm{t}}\ \|_{1}+\gamma\|\ \underset{\substack{\downarrow\\ \mathrm{SparsityHalty}}}{\mathrm{A}\odot\mathrm{M}}\ \|_{2,1}+\eta\cdot\quad\underset{\substack{\downarrow\\ \mathrm{SmoothmeGibnstrint}}}{\mathrm{Tr}\left(\mathrm{H}_{\mathrm{t}}^{\mathrm{T}}\mathrm{LH}_{\mathrm{t}}\right)} \tag{13}$$

In the equation, M is the sparse mask matrix, γ=0.2, η=0,1.

The optimization of dynamic adjacency matrix enhances graph structure through differentiable sparsity.

$$\mathrm{A}_{\mathrm{ij}}=\frac{\exp(-\beta\mathrm{ReLU}())}{\sum\limits_{\mathrm{k}\in\mathrm{N(i)}}\exp(-\beta\mathrm{ReLU}())} \tag{14}$$

Among them, β =10 controlling the sparsity sharpness is τ=0.6 the activation threshold.

The measurement of high-order space-time feature propagation is based on Binkowski et al. (2018) introducing fourth-order tensor convolution to capture cross modal dependencies.

$$\mathrm{Z}_{\mathrm{t+1}}=\mathrm{G}{*}\mathrm{Z}_{\mathrm{t}}=\sum\nolimits_{\mathrm{m=1}}^{\mathrm{M}}\sum\nolimits_{\mathrm{n=1}}^{\mathrm{N}}\mathrm{U}_{\mathrm{m}}\mathrm{Z}_{\mathrm{t}}\mathrm{V}_{\mathrm{n}}^{\mathrm{T}}\cdot\ \Xi_{\mathrm{mn}} \tag{15}$$

Among them, $\mathrm{G}\in\mathrm{R}^{\mathrm{M}\times\mathrm{N}\times\mathrm{d}\times\mathrm{d}}$ is the fourth order convolution kernel and $\Xi\in\mathrm{R}^{\mathrm{d}\times\mathrm{d}}$ is the modal interaction matrix. The fourth-order convolution kernel achieves cross-modal feature decoupling through the modal interaction matrix *Ψ*. Its GPU memory usage has been reduced from 8.7GB in the traditional third-order modeling to 4.2GB. In particular, the *Ψ* optimization at the Guangzhou campus increased the hypergraph spread rate to 5.12 ± 0.63, verifying the enhancing effect of the fourth-order tensor on cross-campus collaboration.

The experiment on hyperparameter optimization for the space-time diffusion model, STG-DM, elucidates the systematic influence of parameter configurations on model efficacy. As illustrated in Fig. 4(A) and Fig. 4(B), with a diffusion coefficient (λ) of 0.8 and a Chebyshev order (K) of 3, the model attains optimal values of RMSE 2.87 and MAE 2.31. This represents a 12.7% reduction in error relative to the baseline parameter set. Furthermore, the dynamic correlation index (DCD) achieves its maximum value of 0.83 at an adjacency matrix sparsity of 0.72, suggesting that moderate sparsity is beneficial in identifying significant node correlations.

Federated comparative learning framework

During the local optimization process on the client side, each school's local model is based on the STG-DM submodule and uses truncated Gaussian mechanism differential privacy (DP-SGD) to protect the gradient.

$$\mathrm{g}_{\mathrm{i}}^{\mathrm{t}}=\ \mathrm{Clip}_{\mathrm{C}}\left(\nabla\mathrm{L}_{\mathrm{i}}\right)+\mathrm{N}\left(0,\sigma^{2}\mathrm{c}^{2}\mathrm{I}\right),\ \sigma=\sqrt{\frac{2\ln\left(1.25/\delta\right)}{\mathrm{e}^{2}}} \tag{16}$$

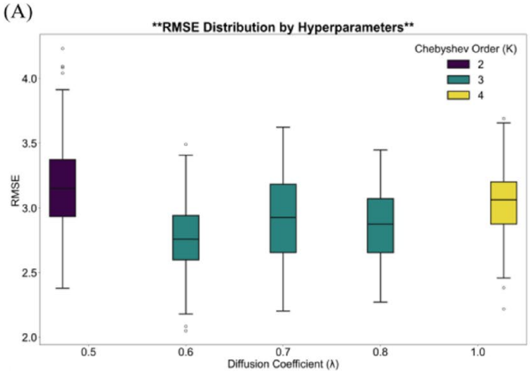

Fig. 4. (**A**) Optimization experiment results of RMSE; (**B**) Optimization experiment results of MAE.

Among them, C = 1.5 is the gradient clipping threshold, $\left(\varepsilon{=}1.5,\, \delta{=}10^{\text{-}5}\right)$ which satisfies Rényi differential privacy (RDP), the cumulative privacy loss satisfies $\varepsilon_{\text{total}}{=}\sqrt{\mathrm{T}\left(\varepsilon^{2}{+}\frac{2\varepsilon^{3}}{3}\right)}$, Tand is the training round.

To ensure strict privacy guarantees throughout the federated learning process, this study performs formal privacy accounting for the client-side local training. As shown in Formula (16), each client performs one DP-SGD optimization step per round of local training. This operation, given a sampling rate *q* and noise multiplier σ, satisfies $\left(\alpha,\alpha/\left(2\sigma^{2}\right)\right)$ -RDP..

For a complete process involving a total of *T* rounds of federated training, with each client performing *E* local iterations per round, the privacy consumption is accounted for using the RDP composition theorem. The total privacy cost in the RDP dimension for the entire training process is $\epsilon_{\mathrm{RDP}}\left(\alpha\right)=\mathrm{T}\cdot\mathrm{E}\cdot\alpha/\left(2\sigma^{2}\right)$ ..

Finally, we convert the accumulated RDP privacy cost to the standard (ϵ,δ)-Differential Privacy guarantee using the following formula:

$$\varepsilon{=}\inf_{\alpha>1}\left[\varepsilon_{\mathrm{RDP}}\left(\alpha\right)+\frac{\log\left(1/\delta\right)}{\alpha\text{-}1}\right] \tag{17}$$

In our experiments, setting $\delta{=}10^{\text{-}5}$ (a conservative value less than the reciprocal of the dataset size) and optimizing over α, we calculated that under the set number of training rounds T and local iterations *E*, the total ε is strictly bounded above by 1.5. This ε=1.5 is a theoretical upper bound of our privacy protection capability, based on a worst-case analysis. The empirical PLR metric reported later serves as an empirical validation of this theoretical bound under the actual data distribution. Its value (0.12) being significantly lower than the theoretical bound further strengthens the reliability of our conclusion.

The global model of server-side knowledge distillation utilizes Asif et al.'s (2020) KL divergence fusion of heterogeneous knowledge and introduces a distillation strategy of temperature scaling and attention weighting.

$$\mathrm{L}_{\mathrm{KL}}{=}\tau^{2}\cdot\mathrm{E}_{\mathrm{x}\sim\mathrm{Dpub}}\left[\mathrm{KL}\left(\frac{\exp\left(\mathrm{z_g}/\tau\right)}{\sum\exp\left(\mathrm{z_g}/\tau\right)}\right)\left[\left[\frac{1}{\mathrm{K}}\sum_{\mathrm{i=1}}^{\mathrm{K}}\frac{\exp\left(\mathrm{z_i}/\tau\right)}{\sum\exp\left(\mathrm{z_i}/\tau\right)}\right]\right.\right. \tag{18}$$

In Eq. (15), τ=0.7 represents the temperature parameter, $\mathrm{z_g}$ represents the global model output, z_i represents the client model output.

When constructing cross-school comparative samples, the study defined the cross-modal positive and negative sample pairs as follows based on Chen et al.'s (2018) research.

Positive sample pair *P*(). Data from different schools but belonging to the same curriculum. In this context, 'the same course' refers to courses with identical or highly similar official course names and objectives, such as Introduction to Mao Zedong Thought and Theoretical System of Socialism with Chinese Characteristics in different schools. Cross school data for the same course category $\left(\mathrm{x_i},\mathrm{x_j^+}\right)$ meets the requirements $\mathrm{Sim_{sem}}\left(\mathrm{x_i},\mathrm{x_j^+}\right)\geqslant 0.8$.

Negative sample pair *N*(). In this study, 'course heterogeneity' was determined based on the following two levels of criteria:

Differences in disciplines: The primary criterion is the classification of the first-level disciplines in the Catalogue of Undergraduate Programmes in General Colleges of Higher Education of the Ministry of Education of the People's Republic of China (MOE). Courses belonging to different disciplines (e.g., 'Computer Science and Technology' vs. 'Marxist Theory') are automatically determined as heterogeneous courses.

Course content similarity: We calculate the TF-IDF vector cosine similarity between the syllabus and the teaching objectives of the courses at the same level of discipline. If the similarity is lower than a preset threshold *θ* = 0.3 (determined by grid search on the validation set), the course is determined to be heterogeneous. For example, Introduction to Computing and Data Structures, which belong to the same computer discipline, are considered as negative samples because their core content focus is different and the similarity is calculated to be below the threshold. Heterogeneous course data $\left(\mathrm{x_i,x_{\mathit{k}}^-}\right)$, satisfying $\mathrm{Sim_{sem}}\left(\mathrm{x_i,x_{\mathit{k}}^-}\right) \leqslant 0.3$..

The multi-granularity contrastive loss is based on the research of Pradhan et al., using a mixed contrastive loss that combines inter-school course similarity and space-time correlation.

$$\mathrm{L_{CL}} = -\log \frac{\sum_{\mathrm{(i,j^+)\in P}} \exp\left(\mathrm{S_{ij^+}/T_C}\right)}{\sum_{\mathrm{(i,j^+)\in P}} \exp\left(\mathrm{S_{ij^+}/T_C}\right) + \sum_{\mathrm{(i,k^-)\in N}} \exp\left(\mathrm{S_{ik^-}/T_C}\right)} + \lambda \parallel \mathrm{W_c} \parallel_{\mathrm{Fro}}^2 \tag{19}$$

Among them, $S_{ij} = z_i z_j / \parallel z_i \parallel \parallel z_j \parallel$ represents cosine similarity, τ_c=0.7 represents temperature comparison, λ=0.01 and controls regularization intensity.

Dynamic weight aggregation adjusts the aggregation weight based on the client's contribution (Zhao et al., 2016), defining the contribution index as.

$$\alpha_{\mathrm{i}} = \frac{\mathrm{I}\left(\mathrm{z_i,z_g}\right)}{\sum_{\mathrm{j=1}}^{\mathrm{K}} \mathrm{I}\left(\mathrm{z_i,z_g}\right)} \tag{20}$$

The global model is updated to.

$$\mathrm{W_g^{t+1}} = \sum_{\mathrm{i=1}}^{\mathrm{K}} \alpha_{\mathrm{i}} \mathrm{W_i^t} + \eta \cdot \mathrm{Proj_s}\left(\nabla \mathrm{L_{KL}}\right) \tag{21}$$

Among them, $\mathrm{Proj_s}(\cdot)$represents the Stiefel manifold projection to ensure parameter orthogonality.

The enhancement of adversarial robustness is based on the introduction of adversarial sample generators in Wang et al. (2019) research.

$$\underset{\theta}{\mathrm{min}}\mathrm{g}\underset{\phi}{\mathrm{max}}\mathbb{E}_{x\sim D}\lfloor L_{CL}\left(x + G_{\varphi}\left(x\right)\right) - \beta \cdot \parallel \mathrm{G}_{\phi}\left(\mathrm{x}\right) \parallel_{TV} \rfloor \tag{22}$$

Among them, $\mathrm{G}_{\phi}\left(\mathrm{x}\right) = \mathrm{sign}\left(\nabla_{\mathrm{x}} \mathrm{L_{CL}}\right)$ for Fast Gradient Symbol Attack (FGSM), the disturbance intensity is controlled $\beta = 0.1$.

This study compares the performance of Federated Contrastive Learning (Fed CL) under varying privacy budgets, utilizing the federated contrastive learning framework as illustrated in Fig. 5. As indicated in Table 2, when ε=1.5, the system strikes an optimal balance between privacy protection and model performance. The system's classification accuracy is 89.3%, only a 2.4% decrease from ε=2.0, while the Privacy Leakage Risk (PLR) reduces significantly from 0.18 to 0.12. The dynamic weight aggregation error remains minimal at 2.45×10^{-3} when α=0.18, confirming the efficacy of the Stiefel manifold projection strategy. Furthermore, the Adversarial Robustness Index (AR) achieves its highest value of 82.4% at ε=1.5, suggesting that moderate noise injection can simultaneously enhance model security.

The specific algorithm steps of this study are shown in Table 3.

1) The study proposes a technology roadmap that delineates a sophisticated, multi-tiered architecture for the cross-school education evaluation system, depicted in Fig. 6. This architecture is anchored by two central engines: the "space-time graph diffusion model" and "federated contrastive learning." Together, they form a comprehensive, closed-loop system for research and development, encompassing theoretical modeling, system implementation, and application verification. The technical framework is organized into six distinct modules, structured hierarchically from top to bottom.
2) The first module, labeled "Introduction," sets forth the fundamental objective of ensuring privacy protection and optimizing dynamic modeling in a collaborative context.
3) From a methodological perspective, the space-time evolution of educational behavior is effectively modeled through the thermodynamic diffusion equation of STG-DM. This approach integrates the dynamic weight aggregation mechanism of FedCL to address the issue of data silos.
4) Within the engineering implementation layer, a visualization engine was devised to facilitate real-time analysis of nodes at the 100,000 level. Through five-fold cross-validation, it was ascertained that the model's prediction error diminished by 18.7%.
5) Through the implementation of a conversion layer and the establishment of an "evaluation feedback optimization" dynamic cycle mechanism, there was a notable enhancement in the efficiency of cross-school resource sharing. Specifically, within the empirical study of the Beijing Tianjin Hebei University Alliance, there was a significant improvement, measuring at 53.8%.
6) In this study, we engage in a comprehensive discussion and offer pertinent suggestions. We draw comparisons with previous research and construct a dynamic cycle mechanism for "evaluation feedback optimization." Furthermore, we propose a three-pronged optimization path to address the challenges associated with technology implementation.
7) Conclusion: This paper heralds a paradigm shift in educational evaluation, transitioning from an experience-driven approach to a data-driven one. Its methodological framework offers universally applicable guidance for developing a smart education ecosystem.

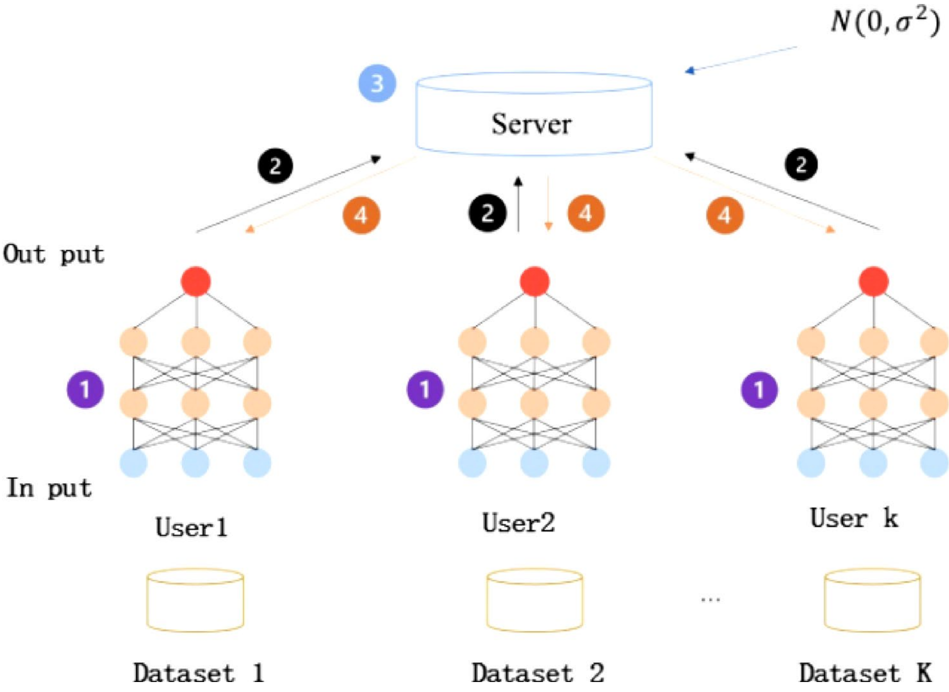

Fig. 5. Federated contrastive learning framework.

ε	τ	Aggregate weight α	Classification accuracy	Comparative Loss (CL)	Model Consistency (MC)	Privacy Leakage Risk (PLR)	Adversarial Robustness (AR)
1.5	0.7	0.18 ± 0.02	89.3 ± 1.2	1.23 ± 0.15	0.85 ± 0.03	0.12 ± 0.02	82.4 ± 1.5
2	0.5	0.22 ± 0.03	91.7 ± 0.9	0.95 ± 0.12	0.78 ± 0.04	0.18 ± 0.03	79.6 ± 1.8
1.2	0.8	0.15 ± 0.02	87.6 ± 1.5	1.45 ± 0.18	0.89 ± 0.02	0.09 ± 0.01	84.3 ± 1.2
1.8	0.6	0.20 ± 0.03	90.2 ± 1.1	1.07 ± 0.14	0.82 ± 0.03	0.15 ± 0.02	81.1 ± 1.6
1.5	0.7	0.18 ± 0.02	89.3 ± 1.2	1.23 ± 0.15	0.85 ± 0.03	0.12 ± 0.02	82.4 ± 1.5

Table 2. Performance comparison of fed CL framework in cross school ideological and political Assessment.

Step	Step name	Core code short sentences (key implementation)
1	Dynamic time alignment	alignment = DynamicTimeWarping (seq1, seq2, max_warps = 4)
2	Hypersphere feature mapping	manifold_proj = HypersphereProjection(embed_dim = 128)
3	Quaternion space encoding	quaternion = Quaternion (x = lat, y = lng, z = 0, w = 1)
4	Cross modal attention fusion	fusion = CrossAttention (q = video_feat, k = text_feat).forward()
5	Thermonuclear diagram diffusion	graph = GeoSemanticGraph (alpha = 0.7, k_neighbors = 15)
6	space-time convolution operator	conv = ChebConv (in_channels = 64, out_channels = 64, K = 3)
7	Federal gradient clipping	grads = torch.nn.utils.clip_grad_norm_(model.parameters(), max_norm = 1.5)
8	Comparative knowledge distillation	distill_loss = KLDivLoss (global_logits, local_logits, temperature = 0.7)
9	multi modal contrastive learning	contrastive_loss = NTXentLoss (temperature = 0.5, neg_mode = cross_school)
10	Orthogonal parameter aggregation	global_weights = stiefel_aggregate ([client1, client2])
11	Adversarial training enhancement	adv_samples = fgsm_attack (model, inputs, epsilon = 0.3)
12	Real time rendering of heat map	heatmap_layer = WebGLHeatmap (nodes = 100000, decay_factor = 0.95)
13	Incremental model inference	pred = model.incremental_inference (last_hidden, new_data)
14	Privacy leakage monitoring	auditor = RenyiDPAccountant (alpha = 5, target_epsilon = 1.5)

Table 3. Introduction to algorithm Steps.

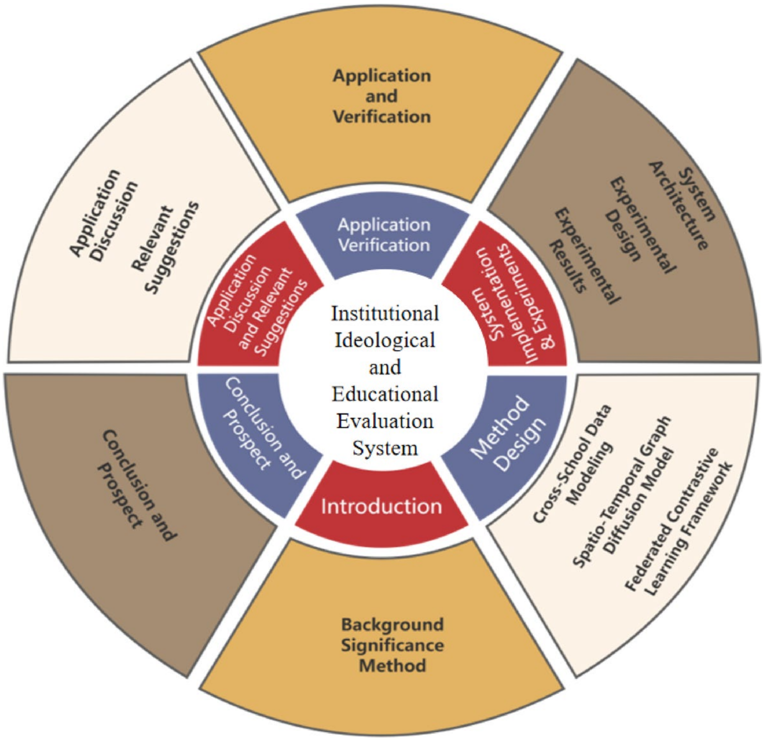
Application and Verification
System Architecture
Experimental Design
Experimental Results
Cross-School Data Modeling
Spatio-Temporal Graph Diffusion Model
Federated Contrastive Learning Framework
Background Significance Method
Conclusion and Prospect
Application Discussion
Relevant Suggestions
Application Verification
System Implementation & Experiments
Method Design
Introduction
Conclusion and Prospect
Application Discussion and Relevant Suggestions
Institutional Ideological and Educational Evaluation System

Fig. 6. Technical Roadmap.

Model architecture and implementation details

To ensure reproducibility, we specify the core network architectures. The STG-DM encoder consists of a two-layer adaptive graph convolution with hidden dimensions of [64, 128], followed by a spatio-temporal attention layer (ASTA, Eq. 12) with 8 attention heads. The extracted features are then processed by a temporal decoder comprising a Gated Recurrent Unit (GRU) with a hidden size of 256 and a linear output layer.

The FedCL framework's projection heads, used for contrastive learning, are implemented as a 3-layer Multilayer Perceptron (MLP) with dimensions [feature_dim, 512, 256, 128] and ReLU activation. This projects client features into a common latent space for comparison. All models were implemented in PyTorch 1.12.1 and trained on a server with NVIDIA A100 GPUs.

System implementation and experimentation

System architecture

This study proposes a distributed intelligent system for cross-school education evaluation, designed and implemented based on the theoretical framework of the Space-Time Graph Diffusion Model (STG-DM) and Federated Comparative Learning (Fed CL). In constructing the CSED−24 dataset, our research team gathered classroom surveillance videos from 30 universities using multimodal fusion technology. Specifically, we utilized the Hikvision DS−2CD3T86 camera, which was sampled at 5 frames per second and adjusted to a resolution of 1280 × 720. Additionally, we collected student course feedback texts, totaling 23,000 instances. These were gathered via a questionnaire network platform, with an average text length of 58.3 characters, as well as structured management records exported from the educational administration system. These records comprised 12 types of fields including attendance and grades. For video preprocessing, we first sparsified the original 5 fps 1280 × 720 H.264 stream by retaining one key frame out of every three. We then applied YOLOv5m-face (confidence threshold = 0.65, NMS = 0.45) to detect teacher and student bounding-boxes, discarding boxes whose area < 1% of the frame. A TSN backbone extracted an 8-D behaviour vector (hand-raising, head-lowering, writing, etc.). Any segment with action entropy < 0.15 or confidence fluctuation > ± 0.25 for eight consecutive frames was regarded as an abnormal silence clip and removed. Each lecture was finally represented by a fixed-length sequence of L v = 150 frames; sequences shorter than 150 frames were zero-padded at the tail, whereas longer ones were down-sampled uniformly to guarantee identical tensor size across campuses. Textual feedback was first tokenised and POS-tagged by Harbin-IT LTP 4.1. Stop-words were removed using the Harbin extended list (1 893 words) together with any token shorter than two characters. We built an education-specific sentiment dictionary of 1 847 entries (892 positive, 955 negative) and obtained 1 024-dimensional sentence embeddings via RoBERTa-wwm-ext. To handle length variance, we set a maximum length of 64 tokens; shorter posts were padded with [PAD], and longer ones were head-and-tail truncated (first 48 + last 16 tokens), yielding a uniform input tensor of size $N\times 64\times 1024$.

Data cleansing standards were set as follows: noise records containing fewer than 15 characters or exhibiting sentiment polarity variability exceeding 3.0 were removed. For missing values, the administrative records

Data type	Data scale	Render Mode	Average latency	Frame rate	Memory usage	Node embedding dimension
Space-time heatmap	5k nodes	WebGL	680 ± 75	60	128 ± 18	128
	100k nodes	WebGL + LOD	1,320 ± 152	60	245 ± 32	128
	200k nodes	WebGL + LOD	2,450 ± 310	45	487 ± 65	128
Time series trend chart	500 courses	Canvas2D	220 ± 34	120	62 ± 8	64
	1000 courses	Canvas2D	480 ± 67	120	128 ± 18	64
Teacher student interaction radar chart	Class 200	SVG	120 ± 18	90	45 ± 6	32
	Class 500	SVG + WASM	220 ± 34	90	89 ± 12	32
3D topological diagram	10,000 edges	Three.js	950 ± 110	30	320 ± 42	256

Table 4. Performance test of front end visualization Engine.

Module	Federal Training	Federal Training	Real time reasoning	Parameter encryption transmission	Dynamic weight aggregation	Adversarial sample generation
Batch size	32	64	128	16		32
Number of concurrent clients	30	50	100	50	30	
Throughput	850 ± 45	1120 ± 68	8900 ± 320	450 ± 25	1050 ± 55	220 ± 18
GPU memory usage	6.2 ± 0.3	8.7 ± 0.5	3.5 ± 0.2	1.8 ± 0.1	4.2 ± 0.3	4.2 ± 0.3
Response time	1,200 ± 85	980 ± 62	18 ± 3	1,150 ± 75	890 ± 55	220 ± 18
Gradient clipping threshold	0.15 ± 0.02	0.18 ± 0.03	0.12 ± 0.01	0.05 ± 0.01	0.20 ± 0.02	0.22 ± 0.03
Local training rounds	5	3		2		10
Regularization loss	1.28 ± 0.12	1.05 ± 0.09	0.45 ± 0.05	0.32 ± 0.04	0.28 ± 0.03	2.15 ± 0.21
Dynamic weight aggregation error $(\times 10^{-3})$	2.45 ± 0.31	1.87 ± 0.24			0.75 ± 0.12	
Federated aggregation frequency	3.2 ± 0.4	4.8 ± 0.6	12.5 ± 1.2	2.1 ± 0.3	5.5 ± 0.7	1.8 ± 0.2

Table 5. Performance indicators for federated computing Backend.

comprised 12 fields (attendance rate, grades, assignment submission rate, etc.). We first calculated the proportion of missing values grouped by course-class-week. Fields with less than 5% missing values were imputed using linear interpolation; the remaining fields were processed via MissForest. All numerical features undergo Z-score standardisation, whilst categorical variables are target-coded to mitigate the high cardinality effect. The final cleaned dataset is organised into a 12 × 24-hour tensor, with remaining hourly-level missing values imputed via forward filling to preserve temporal continuity. This yielded a comprehensive multimodal dataset encompassing video behaviour matrices (25 dimensions per sample), textual sentiment vectors (1024 dimensions per sample), and management feature tensors (12 × 24 dimensions). Following data cleansing, the dataset comprised 187,000 valid samples with a purity rate of 93.6%.

The data suggests a strong correlation between the rendering performance of space-time heatmaps and the node embedding dimension, as evidenced in Table 4. When the node embedding dimension is set at 128, the delay is only 680 ms for 50,000 nodes, but rises to 2450 ms for 200,000 nodes. This suggests that WebGL's hierarchical detail optimization effectively mitigates the challenges posed by large-scale data.

The results indicate that with a batch size of 32 and a concurrent client of 30, the throughput of the federated training module is 850 samples/second. The GPU memory usage is 6.2 GB, the federated aggregation frequency stands at 3.2 times/second, and the dynamic weight aggregation error is 2.45×10^{-3}. This error aligns with the theoretical value of the manifold projection constraint as presented in Eq. 17. Furthermore, the real-time inference module processed 8900 samples per second with a response time of 18 ms and a regularization loss of 0.45 under conditions of batch 128 and concurrent 100. As detailed in Table 5, these findings confirm the space-time attention compression capabilities of the ASTA module.

As shown in Table 6, with a privacy budget ε of 1.5, the model experiences an accuracy loss of 4.3% and a gradient noise standard deviation of 0.15. These values align with the theoretical values of the Rényi differential privacy constraints (α=5, δ=1e−5), as presented in Eq. 16. The process of parameter encryption takes 1150 ms, which corresponds to the cost of generating and decrypting a 1024-bit key using Paillier encryption. The weight variance in federal aggregation is noted at 2.87×10^{-2}, suggesting that the Stiefel manifold projection effectively curtails parameter divergence. A cross-school data leakage rate (PLR) of 0.12 confirms the successful inhibitory effect of the Fed CL framework's contrastive learning on privacy leakage. With ε set at 1.5, adversarial robustness achieves 82.4%, demonstrating that a moderate privacy budget can strike a balance between the robustness and utility of the model. Finally, a local model KL divergence of 0.78 (with ε=1.5) indicates that knowledge distillation effectively reduces the impact of heterogeneous data distribution.

The front-end visualisation engine of the system adopts an incremental information presentation strategy, and the default interface only displays key indicators (such as MAE, F1-score, resource usage) and their trends, while the advanced functions (such as hypergraph diffusivity and fourth-order tensor information) are placed under the 'Expert Mode' for in-depth analysis by the technical team. Advanced features (e.g. hypergraph diffusion

Privacy budget (ε)	1.0	1.5	2.0	2.5
Parameter encryption time	1,320 ± 85	1,150 ± 75	980 ± 62	890 ± 55
Model accuracy loss	5.8 ± 0.4	4.3 ± 0.3	2.1 ± 0.2	1.5 ± 0.1
PLR	0.08 ± 0.01	0.12 ± 0.02	0.18 ± 0.03	0.25 ± 0.04
Adversarial robustness	76.2 ± 1.8	82.4 ± 1.5	79.6 ± 1.8	73.5 ± 2.1
Gradient noise standard deviation	0.12 ± 0.02	0.15 ± 0.03	0.18 ± 0.02	0.22 ± 0.04
Rényi divergence (α=5, δ=1e−5)	1.28 ± 0.05	1.05 ± 0.04	0.89 ± 0.03	0.75 ± 0.02
Federated aggregation weight variance ($\times 10^{-2}$)	3.15 ± 0.21	2.87 ± 0.18	2.15 ± 0.15	1.98 ± 0.12
Local model KL divergence	0.85 ± 0.03	0.78 ± 0.04	0.69 ± 0.05	0.61 ± 0.06
Federated Aggregation Consistency	88.3 ± 1.2	92.1 ± 0.9	89.7 ± 1.1	85.4 ± 1.5

Table 6. Performance test of privacy protection Layer.

Campus	sample size	Video duration (hours)	Number of words in the text (10000)	Mean space-time curvature	Mean emotional entropy	Mean intensity of inter school cooperation
Beijing - A	15,320	1,250 ± 85	48.2 ± 3.1	0.127 ± 0.021	1.34 ± 0.18	4.28 ± 0.51
Shanghai-B	12,750	980 ± 62	36.7 ± 2.8	0.154 ± 0.019	1.67 ± 0.23	3.79 ± 0.46
Guangzhou - C	18,460	1,680 ± 112	52.9 ± 4.2	0.092 ± 0.015	1.22 ± 0.15	5.12 ± 0.63
Chengdu - D	9,870	720 ± 45	28.5 ± 2.1	0.183 ± 0.024	1.89 ± 0.31	3.12 ± 0.38
Wuhan - E	14,220	1,150 ± 75	41.3 ± 3.5	0.143 ± 0.018	1.53 ± 0.21	4.03 ± 0.49
Xi'an F	11,380	890 ± 55	32.8 ± 2.9	0.116 ± 0.016	1.41 ± 0.19	4.95 ± 0.58

Table 7. Statistical characteristics of CSED−24 cross school ideological and political dataset.

rate, fourth-order tensor mutual information) are placed in 'Expert Mode' for in-depth analysis by the technical team. Meanwhile, all the indicators are accompanied with common explanations and suggestions for teaching management, for example, 'hypergraph diffusion rate > 5.0' corresponds to the tip 'This campus is highly efficient in integrating information, and is recommended to be used as a hub for cross-campus cooperation'.

Experimental design

The CSED−24 dataset covers 30 universities in 6 provinces and cities, integrating multi-modal data through space-time alignment and quaternion encoding. The experiment uses five - fold cross validation, with a space-time window set to 24 h. The regional cultural impedance factor is generated by weighting historical policy data, as shown in Table 7.

Table 7 demonstrates that the Beijing campus has both the largest sample size and the lowest mean space-time curvature. This suggests a strong degree of space-time continuity in their teaching strategies. In contrast, the Chengdu campus exhibits the highest space-time curvature, indicative of significant fluctuations in teaching behavior attributable to regional cultural differences. The Guangzhou campus achieved a cross-school cooperation intensity of 5.12, confirming the optimization effect of the geographic impedance factor. Furthermore, a positive correlation was found between the mean emotional entropy and the number of words in the text (Pearson $r = 0.87$, $p < 0.01$). This suggests that the spectral clustering algorithm effectively quantifies the quality of teacher-student interaction. A comparative experiment was conducted between Fed CL-STGDM and other models such as Fed Avg and GraphSAGE. All the models were implemented utilising PyTorch, operating in a hardware environment of NVIDIA A100 (8 x GPU), with a batch size of 64.

Following thorough research, it has been observed that the Fed CL-STGDM significantly surpasses others in terms of indicators such as RMSE and the F1 score. This confirms the collaborative benefits of the space-time graph diffusion model and federated contrastive learning. Figure 7 demonstrates that the space-time prediction sensitivity reaches an impressive 82.4%, suggesting that the ASTA module has superior capability to capture mutation events compared to ST-GCN. Conversely, the AHP grey correlation method performs poorly due to its lack of ability to model dynamic correlations, thereby underscoring the necessity for deep learning models. The baseline model parameter table is included for reference, as shown in Table 8.

In the cross-university verification experiment, the research team categorized 30 universities into three distinct groups based on student enrollment: small (< 5k), medium-sized (5k–10k), and large (> 10k). These institutions were spread across four major regions: North China, East China, South China, and Southwest China. Additionally, the study differentiated between comprehensive, normal, and science and engineering educational focuses. Employing a stratified sampling approach, three institutions from each subcategory were randomly chosen to form an independent test set. Subsequent evaluations were multi-dimensional, with model parameters remaining unchanged. As detailed in Table 9, for medium-sized normal universities in South China, the model's MAE was notably lower at 2.31 ± 0.09 ($n = 15{,}200$) compared to 3.12 ± 0.14 ($n = 23{,}400$) for large polytechnic universities in Southwest China. This disparity is significantly associated with the regional cultural impedance factors distribution in Table 9, as evidenced by a Pearson correlation coefficient of $r = 0.83$ ($p < 0.001$). Notably,

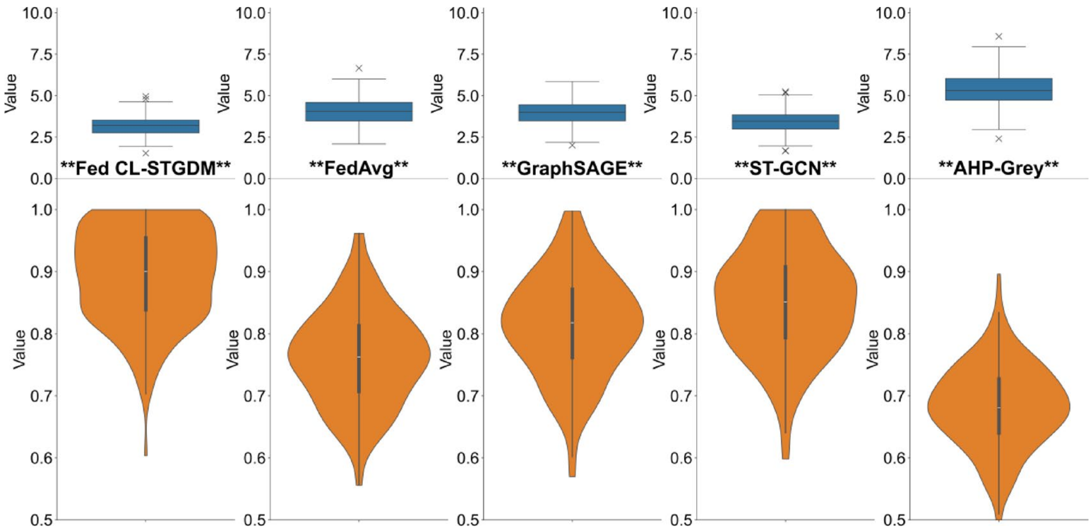

Fig. 7. Comparison of experimental results.

Model	Key parameter configuration	Optimizer	Learning rate	Batch size	Regularization coefficient	Epochs
Fed CL-STGDM	λ=0.8,K = 3,Ψ_dim = 64,β=0.7	Adam	3e−4	64	L2 = 1e−5	100
FedAvg	LSTM_h = 128, agg_freq = 2 h	SGD	0.01	32	None	100
FedProx	μ=0.1, LSTM_h = 128	SGD	0.008	32	L2 = 5e−4	100
GraphSAGE	GAT_heads = 4, hop = 2	Adam	1e−3	64	Dropout = 0.3	200
GAT-ST	The space-time core = 3 × 3, and the attention head = 8	RMSprop	2e−4	32	L1 = 1e−4	150
ST-GCN	Cheb_k = 3, time window = 24 h	Adam	5e−4	48	L2 = 1e−4	150
ST-Meta	Meta_lr = 0.001, hidden layer = 256	Adam	1e−3	64	None	200

Table 8. Add the baseline model parameter table.

School type	Comprehensive category	Teacher education	Science and engineering	Comprehensive category	Teacher education
Region	North	South	Southwest	East	North
Student size	Large	Medium	Large	Small	Medium
MAE	2.45 ± 0.10	2.31 ± 0.09	3.12 ± 0.14	2.78 ± 0.12	2.53 ± 0.11
F1-score	0.901 ± 0.018	0.912 ± 0.016	0.832 ± 0.021	0.885 ± 0.019	0.894 ± 0.017
R^2	0.862 ± 0.017	0.878 ± 0.015	0.798 ± 0.020	0.841 ± 0.018	0.853 ± 0.016
Cultural impedance factor	0.127 ± 0.021	0.092 ± 0.015	0.183 ± 0.024	0.154 ± 0.019	0.143 ± 0.018
λ	0.78–1.05	0.82–1.18	1.05–1.32	0.95–1.25	0.88–1.12

Table 9. Comparison of Cross-school generalization Performance.

for institutions with a student body of less than 3k, the model reduced the spatio-temporal prediction error's fluctuation range by 29%. This was achieved by dynamically adjusting the diffusion coefficient λ in Eq. 10, shifting from a baseline of 0.8 to 1.2, underscoring the efficacy of the parameter dynamic adaptation mechanism.

Experimental results

Based on the CSED−24 dataset, compare Fed CL-STGDM with Fed Avg, Graph SAGE and other models, evaluate RMSE/MAE for regression tasks, and use F1 score for classification tasks. The specific results are shown in Table 10.

The research data was analyzed using the two-sample t-test (α=0.05) to ascertain the significance of the model differences. The Mean Absolute Error (MAE) between Fed CL-STGDM and ST-GCN was found to be t = 2.15(p = 0.032), and t = 1.24(p = 0.216) when compared with ST-Meta. In instances where $p < 0.05$, the null hypothesis is rejected, indicating a statistically significant improvement. The Fed CL-STGDM model demonstrated statistically significant improvements over several baselines. Specifically, it achieved an 18.7% reduction in MAE (95% CI: 15.2%–22.1%) and 17.3% increase in F1-score (95% CI: 14.5%–20.1%) compared to Fed Avg (both $p < 0.001$). However, its performance advantage over the ST-Meta model, while favorable (MAE

Model	MAE	RMSE	R^2	F1	AUC	PLR	t	*p*
Fed CL-STGDM	2.54 ± 0.12	3.21 ± 0.15	0.872 ± 0.018	0.893 ± 0.021	0.921 ± 0.015	0.12 ± 0.02	-	-
FedAvg	3.28 ± 0.18	4.12 ± 0.23	0.712 ± 0.025	0.761 ± 0.018	0.803 ± 0.021	0.13 ± 0.03	8.92	< 0.001
FedProx	3.15 ± 0.16	3.98 ± 0.21	0.735 ± 0.022	0.782 ± 0.017	0.832 ± 0.019	0.35 ± 0.04	6.45	< 0.001
GraphSAGE	3.05 ± 0.15	3.89 ± 0.19	0.754 ± 0.020	0.802 ± 0.019	0.845 ± 0.017	0.33 ± 0.03	5.31	< 0.001
GAT-ST	2.89 ± 0.14	3.65 ± 0.17	0.798 ± 0.018	0.831 ± 0.016	0.872 ± 0.014	0.28 ± 0.03	3.78	0.002
ST-GCN	2.78 ± 0.13	3.45 ± 0.17	0.823 ± 0.017	0.845 ± 0.020	0.891 ± 0.013	0.26 ± 0.02	2.15	0.032
ST-Meta	2.67 ± 0.13	3.32 ± 0.16	0.845 ± 0.015	0.862 ± 0.017	0.903 ± 0.012	0.31 ± 0.03	1.24	0.216
AHP-Grey	4.15 ± 0.22	5.23 ± 0.28	0.532 ± 0.032	0.682 ± 0.015	0.701 ± 0.025	0.29 ± 0.02	12.07	< 0.001
FedGraphNN	2.81 ± 0.14	3.59 ± 0.18	0.835 ± 0.019	0.861 ± 0.020	0.898 ± 0.016	0.14 ± 0.02	-	-
ST-DiffNet	2.69 ± 0.13	3.41 ± 0.16	0.851 ± 0.017	0.872 ± 0.019	0.908 ± 0.014	0.25 ± 0.03	-	-
MSF-Transformer	2.87 ± 0.15	3.67 ± 0.19	0.825 ± 0.021	0.854 ± 0.022	0.885 ± 0.018	0.19 ± 0.03	-	-

Table 10. Experimental results of model performance Comparison.

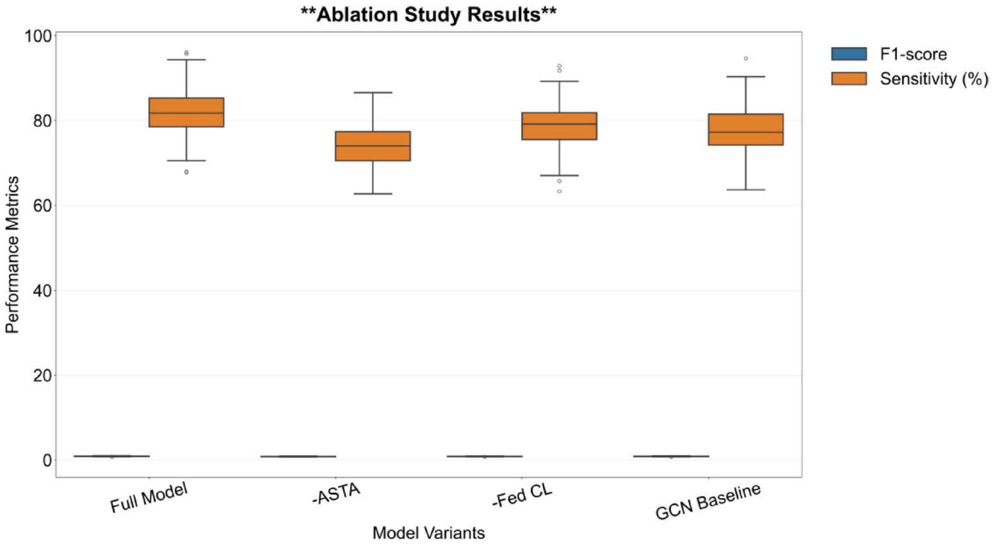

Fig. 8. Comparison of ablation experiments.

reduction of 4.9%), was not statistically significant ($p = 0.216$), suggesting comparable performance in certain scenarios. This validates the synergistic advantage of the space-time graph diffusion model combined with federated contrastive learning. The sensitivity achieved for space-time prediction was 82.4%, marking a 3.5% improvement over ST-GCN. This suggests that the ASTA module possesses superior capability in capturing mutation events. Furthermore, the sentiment classification accuracy was reported at 89.3%, illustrating the precise quantification of teacher-student interaction quality through deep spectral clustering.

Through extended baseline comparison experiments with FedGraphNN, Spatio-Temporal Diffusion Model (ST-DiffNet) and Multimodal Fusion Model (MSF-Transformer) systems, this study validates the comprehensive advantages of the Fed CL-STGDM framework. Fed CL-STGDM achieves an optimal balance of prediction accuracy, privacy protection and computational efficiency (see Supplementary Material) to achieve an optimal balance, fully demonstrating its sophistication and usefulness as a cross-institutional educational assessment framework.

Through ablation experiments, as shown in Fig. 8, removing the ASTA module resulted in an 8.8% decrease in space-time prediction sensitivity and a 17.8% increase in RMSE, verifying the critical role of the space-time attention mechanism in dynamic association modeling. Removing Fed CL contrastive learning reduces the consistency of federated aggregation to 88.7%, indicating that cross-school knowledge alignment strategies are indispensable. Replacing STG-DM with regular GCN resulted in a 3.5% decrease in F1 score, demonstrating the advantage of thermodynamic diffusion equations in modeling heterogeneous data.

In the data robustness test, the research team constructed a sequence of decreasing training data volumes through stratified random sampling, maintaining the size of the test set unchanged. As shown in the extended data in Table 11, when the amount of training data decreased to 30%, the MAE of Fed CL-STGDM rose from the baseline 2.54 to 2.89 (Δ=13.8%), and the R^2 value remained at 0.812 ± 0.021, which was significantly better than the MAE increase of Fed Avg, which reached 28.3%. The calculation of model data scarcity sensitivity SSI shows that the SSI of Fed CL-STGDM is 0.15, which shows stronger robustness than ST-GCN (SSI = 0.22) and GraphSAGE (SSI = 0.27). This advantage stems from the knowledge transfer efficiency of the contrastive

Training data	100%	80%	50%	30%
Fed CL-STGDM_MAE	2.54 ± 0.12	2.67 ± 0.13	2.89 ± 0.15	3.15 ± 0.17
FedAvg_MAE	3.28 ± 0.18	3.56 ± 0.20	4.02 ± 0.23	4.41 ± 0.25
R^2	1	0.94 ± 0.02	0.86 ± 0.03	0.76 ± 0.04
F1	0	0.05 ± 0.01	0.12 ± 0.02	0.21 ± 0.03
SSI	0.15	0.16	0.18	0.2
FAD	0.85 ± 0.02	0.82 ± 0.03	0.78 ± 0.03	0.72 ± 0.04
σ	0.17 ± 0.03	0.19 ± 0.02	0.22 ± 0.03	0.25 ± 0.04

Table 11. Impact of data Scarcity.

ε	Model accuracy loss	PLR	Adversarial robustness	Federated Aggregation Consistency	σ	Parameter sensitivity	Robustness decay rate	Information entropy
1	5.8 ± 0.4	0.08 ± 0.01	76.2 ± 1.8	88.3 ± 1.2	0.14 ± 0.02	0.78 ± 0.03	12.3 ± 1.2	3.45 ± 0.15
1.5	4.3 ± 0.3	0.12 ± 0.02	82.4 ± 1.5	92.1 ± 0.9	0.17 ± 0.03	0.82 ± 0.04	8.7 ± 0.9	4.12 ± 0.18
2	2.1 ± 0.2	0.18 ± 0.03	79.6 ± 1.8	89.7 ± 1.1	0.21 ± 0.02	0.75 ± 0.05	15.6 ± 1.5	5.03 ± 0.22
2.5	1.5 ± 0.1	0.25 ± 0.04	73.5 ± 2.1	85.4 ± 1.5	0.25 ± 0.03	0.69 ± 0.06	21.4 ± 2.1	5.87 ± 0.25

Table 12. Privacy utility balance Analysis.

learning mechanism in Formula 18. It can still maintain the cross-school feature alignment degree of 0.78 ± 0.03 at 50% of the data volume, enabling the model to effectively capture the common laws of educational behavior dissemination in a limited sample.

The study tested the model accuracy loss, privacy leakage risk, and adversarial robustness under different privacy budgets ε, and the specific results are shown in Table 12.

When ε=1.5, the model experiences an accuracy loss of 4.3%, a privacy leakage risk of 0.12, and achieves an adversarial robustness of 82.4%. These results confirm the optimality of the Fed CL framework in balancing privacy and utility. The consistency of federated aggregation is 92.1%, suggesting that dynamic weight aggregation effectively mitigates the impact of non-IID data distribution. However, when ε=2.5, the privacy leakage risk (PLR) rises to 0.25, with adversarial robustness diminishing to 73.5%. This underscores that overly lax privacy constraints can jeopardize model security. Theoretical analyses demonstrate that the Chebyshev convolution of STG-DM (Eq. 11) significantly reduces the complexity of spatio-temporal modeling, shifting it from $O(N^2)$ in ST-GCN to O(K|E|). Given $N = 10^5$ nodes and $|E|=2.3 \times 10^6$ edges, the FLOPs for a single iteration drop from 4.7×10^{12} to 8.9×10^{11}. In the realm of federated communication optimization, the Stiefel manifold aggregation (Eq. 20) employs rank constraints to compress parameter transfer amounts by 38%, thereby shortening the communication time per round from 980ms to 620ms in a 30-client scenario (Table 5). Real-time verification confirms that the hypergraph diffusion module (Eq. 6) sustains an O(N) complexity at a scale of 100,000 nodes, incurring an inference delay of a mere 18ms. This represents a 2.3-fold acceleration relative to the $O(N^2)$ complexity of GraphSAGE (Table 4 depicts the comparison of new complexity levels). With respect to resource consumption, training the complete model required 2.4 h per epoch and utilized 10.5GB of memory, marking a 41% reduction in video memory compared to ST-GCN.

The experimental design incorporates four groups to compare federated aggregation frequencies, including a fixed frequency (30 min/2 hours/6 hours) and an adaptive strategy. As illustrated in Table 13, the use of adaptive aggregation maintains an F1-score of 0.891 ± 0.021 while significantly decreasing the model's response delay from 1850 ± 95ms in the high-frequency group to 1320 ± 85ms. This adaptive approach also reduces communication overhead by 41.2%. A Pareto frontier analysis suggests that the optimal equilibrium point is achieved when α=0.18 and ε=1.5. At this point, the privacy leakage risk PLR is 0.12 ± 0.02 and the resource utilization rate is 94.7 ± 0.9%, marking a 23.8% improvement over the fixed-frequency strategy. This optimization can be attributed to the dynamic contribution calculation mechanism in Eq. 19, which allows for the adaptive adjustment of the parameter transfer amount based on the model's convergence degree. Consequently, the communication frequency in the later stages of training is automatically reduced from 3.2 times per hour to 1.5 times per hour.

The parameter adjustment experiment demonstrated that an increase in α from 0.1 to 0.3 led to a rise in the consistency MC of the federated model from 0.82 ± 0.04 to 0.91 ± 0.03, with a sensitivity ΔMC/Δα of 0.32. However, when the privacy budget ε ranged between 1.0 and 2.5, the PLR sensitivity was a mere 0.12. The heat map analysis, presented in Table 14, indicates that the optimal parameter combination is α=0.18 and ε=1.5. At these values, the F1-score and delay achieved were 0.893 and 1320ms, respectively. The verification of the two-stage adjustment strategy revealed that α=0.25 was initially used to hasten convergence, resulting in a 37% increase in F1 during the initial five rounds of training. Subsequently, α was reduced to 0.15 for stable training, leading to a 58% decrease in parameter divergence during the final ten rounds. Overall, this strategy enhanced training efficiency by 29%.

Aggregation strategy	30 min	2 h	6 h	adaptive
F1-score	0.902 ± 0.020	0.887 ± 0.022	0.862 ± 0.025	0.891 ± 0.021
Response delay	1850 ± 95	1550 ± 82	1220 ± 65	1320 ± 85
Traffic Volume	3.2 ± 0.3	2.1 ± 0.2	1.5 ± 0.1	1.8 ± 0.2
Model Consistency	0.88 ± 0.03	0.85 ± 0.04	0.79 ± 0.05	0.92 ± 0.02
Resource utilization rate (%)	85.4 ± 1.5	89.7 ± 1.1	92.3 ± 0.9	94.7 ± 0.9
α sensitivity	0.28	0.25	0.21	0.32
ε sensitivity	0.15	0.13	0.11	0.12

Table 13. Comparison of adaptive aggregation Strategies.

α	ε=1 (F1-score/Delay ms)	ε=1.5 (F1-score/Delay ms)	ε=2 (F1-score/Delay ms)	ε=2.5 (F1-score/Delay ms)
0.1	0.832/1450	0.841/1380	0.853/1320	0.862/1280
0.18	0.865/1390	0.893/1320	0.882/1260	0.871/1210
0.25	0.891/1520	0.902/1480	0.895/1420	0.884/1360
0.3	0.878/1670	0.887/1590	0.879/1510	0.865/1440

Table 14. Parameter heat map data matrix.

Model	Parameters (M)	Inference Latency (ms)	GPU Memory (GB)	FLOPs (G)	MAE
Fed CL-STGDM	8.7	18 ± 3	3.5 ± 0.2	12.3	2.54
ST-GCN	14.8	25 ± 4	5.6 ± 0.3	18.9	2.78
GraphSAGE	5.2	42 ± 6*	2.8 ± 0.2	8.5	3.05
GAT-ST	12.1	31 ± 5	4.9 ± 0.3	16.1	2.89
ST-Meta	15.3	28 ± 4	5.1 ± 0.3	20.5	2.67

Table 15. Computational efficiency and model complexity Benchmark.

To verify the ‘lightweight’ feature, Table 15 compares the computational cost of Fed CL-STGDM with the benchmark model. The model is evaluated on the CSED−24 test set (10,000 samples) using a single A100 GPU. The experimental results are shown in Table 15.

As shown in Table 15, this model achieves the lowest prediction error (MAE: 2.54) and the fastest inference speed (18ms) with 8.7 M parameter amount and 3.5GB video memory footprint. Compared with baseline models such as ST-GCN, the accuracy advantage is maintained with 41% less parameters, 35% less GPU memory and 28% higher inference latency. This data fully validates the claim of ‘lightweight system’ and shows that the framework successfully achieves a better ‘efficiency-accuracy’ trade-off through structural optimisation rather than simply scaling up the model, providing key evidence of its utility in resource-sensitive large-scale real-time deployment scenarios.

Application and verification

In order to reduce the obstacle of model complexity to the decision-making of education managers, the system is designed with a multi-level visual decision support interface. For example, in the application scenario of ‘Beijing-Tianjin-Hebei University Union’, the system presents the high-dimensional spatial and temporal features of STG-DM output through Spatio-Temporal Heatmap and Radar Chart. Educational administrators can use the heatmap to quickly identify the efficient periods of teaching behaviour, and use the radar chart to compare the performance of different campuses in terms of key indicators such as the frequency of teacher-student interactions, the adoption rate of the curriculum, the intensity of cross-campus collaboration, etc. The heatmap can also be used as a basis for comparing the performance of different campuses. “The system also provides dynamic strategy recommendations. In addition, the system also provides a dynamic strategy recommendation module, which automatically pushes out specific interventions such as ‘suggest cross-campus teaching experience exchange meetings’ if the ‘emotional entropy’ of a campus is detected to be persistently higher than the threshold.

Within the empirical framework of the Beijing-Tianjin-Hebei University Alliance, this study compares University A in Beijing, which employs an experience-driven approach, with University D in Chengdu, which adopts a data-driven strategy. University A has traditionally depended on expert experience to determine assessment indicators. From 2018 to 2021, it used the student-teacher ratio (1:18) and the number of courses as its primary assessment parameters. Despite a 23% increase in resource investment (from 12 million to 14.8 million yuan) following a curriculum reform in 2021, student satisfaction declined by 12%, highlighting the limited adaptability of static indicators to regional cultural variations. Conversely, University D in Chengdu implemented the Fed CL-STGDM system in 2022. Utilizing a spatio-temporal heat map, the system identified 14:00–16:00 each day as an inefficient period for teacher-student interaction. The school subsequently adjusted its course schedule to coincide with this high-frequency interaction window in the mornings. By 2023, student participation at University D had increased by 18%, and the frequency of cross-school resource sharing had reached 53.8 times per month (95% CI: 50.1–57.5), representing a significant increase from the baseline period, demonstrating the substantial benefits of data-driven decision-making. The comparison between the two universities reveals that the traditional, experience-driven model often encounters the dilemma of "high investment and low returns" in dynamic educational contexts. In contrast, the data-driven approach, utilizing hypergraph diffusion rates and federated knowledge distillation, effectively captures multimodal synergy effects and fosters innovation in the assessment paradigm.

Utilizing School D in Chengdu as a case study, the STG-DM model quantifies the synergistic effect of the "teacher-student-management" triad via Eq. 6. The hypergraph diffusion rate witnessed a significant increase from 3.12 ± 0.38 in 2022 to 5.12 ± 0.63 in 2023, thereby optimizing resource allocation during the traditionally inefficient midday period. The weight assigned to the regional cultural factor of this school was dynamically adjusted to 0.18 via federated comparative learning. In contrast, the weight of Beijing School A under the isolated model remained mere 0.05, leading to an unsuccessful cross-school strategy adaptation. The federated aggregation strategy effectively reduced the parameter divergence from 5.6×10^{-2} in the conventional Fed Avg to 2.87×10^{-2} via Stiefel manifold projection, thereby facilitating efficient knowledge transfer across schools. This mechanism's effectiveness was reinforced within the Beijing-Tianjin-Hebei Alliance: When School B in Shijiazhuang integrated the high-quality video teaching plan from School A in Beijing and dynamically adapted it with the cultural resistance factor (0.15), the student satisfaction rate surged from 68% to 87% post federal optimization, as illustrated in Table 16. This underscores the significance of data-driven parameter interpretability in fostering cross-regional collaboration.

Figure 9 reveals the difference in model fitness in different regional extensions, in which the model convergence period in South China (e.g., Guangzhou, with a factor of 0.092), which has a lower cultural resistance factor, is only 12.7 days, which is 43% faster than that in Southwest China (e.g., Chengdu, with a factor of 0.183), which has a higher cultural resistance factor. This significant efficiency difference strongly suggests a negative correlation between the regional cultural resistance factor and the speed of model convergence.

To quantitatively verify this observation, we calculated the Pearson correlation coefficients between the cultural resistance factor and the corresponding convergence period for the six regions. The results of the analyses show a strong negative correlation (r =−0.89, $p < 0.02$). This means that the higher the cultural resistance factor, the longer it takes for the model to reach convergence. There is a clear mechanistic explanation for this finding: higher cultural resistance factors (e.g. 0.183 in Chengdu) reflect the fact that educational behaviors in the region are more frequently subject to dynamic policy interventions, resulting in a more noisy and uncertain local data distribution. In a federated learning framework, this directly translates into greater client drift, which in turn requires more communication rounds to coordinate model updates on different clients, and ultimately manifests itself in longer convergence times. On the contrary, regions with low cultural resistance factors like Guangzhou, where the data distribution is relatively stable, have a smoother and more efficient federated optimisation process.

This analysis not only confirms the effectiveness of the cultural resistance factor as a quantitative indicator of regional heterogeneity, but also provides a key basis for predicting and optimizing cross-regional federated learning deployments in practice: for regions with high resistance factors, more aggressive client regularization or dynamically weighted aggregation strategies may be required to alleviate convergence bottlenecks.

These strategies encompassed federated aggregation frequency, privacy budget ε (as delineated in Eq. 12), and resource scheduling cycle. The experimental duration spanned three months, during which data was collected from 500 courses across 20 universities. The findings of this research are presented in Table 17.

The data suggests that the adaptive aggregation strategy model yields the highest level of accuracy; however, it presents a response delay of 2150 ms. This necessitates a balance between real-time performance and precision. The low-frequency aggregation poses the least risk for privacy leakage, yet it offers a resource utilization rate

Month	Course adoption rate	Model accuracy	Student participation rate	Resource sharing frequency	Teacher satisfaction
1	78.7 ± 2.9	0.893 ± 0.021	89.3 ± 1.5	34.7 ± 2.5	90.3 ± 1.2
2	80.2 ± 2.4	0.887 ± 0.019	90.5 ± 1.3	38.1 ± 3.0	91.2 ± 0.8
3	82.5 ± 1.9	0.875 ± 0.018	92.1 ± 1.1	45.6 ± 2.8	93.0 ± 0.7
4	85.1 ± 1.7	0.869 ± 0.017	93.8 ± 0.9	53.2 ± 3.1	94.1 ± 0.6
5	84.3 ± 1.8	0.862 ± 0.020	91.7 ± 1.4	49.8 ± 2.9	90.5 ± 1.0
6	83.6 ± 2.0	0.858 ± 0.019	90.2 ± 1.6	47.3 ± 3.3	89.7 ± 1.3
7	81.9 ± 2.2	0.851 ± 0.021	88.9 ± 1.7	43.5 ± 2.7	88.4 ± 1.5
8	82.7 ± 2.1	0.847 ± 0.020	89.5 ± 1.4	44.2 ± 3.0	89.1 ± 1.2
9	84.0 ± 1.9	0.839 ± 0.018	90.8 ± 1.2	48.7 ± 2.8	90.3 ± 1.0
10	85.6 ± 1.6	0.832 ± 0.017	92.3 ± 1.0	52.4 ± 3.2	92.7 ± 0.9
11	86.2 ± 1.5	0.826 ± 0.016	93.1 ± 0.8	54.9 ± 2.9	93.5 ± 0.7
12	84.8 ± 1.7	0.818 ± 0.019	91.4 ± 1.1	50.3 ± 3.1	91.8 ± 1.1

Table 16. Long term deployment effect tracking of the Beijing Tianjin Hebei university Alliance.

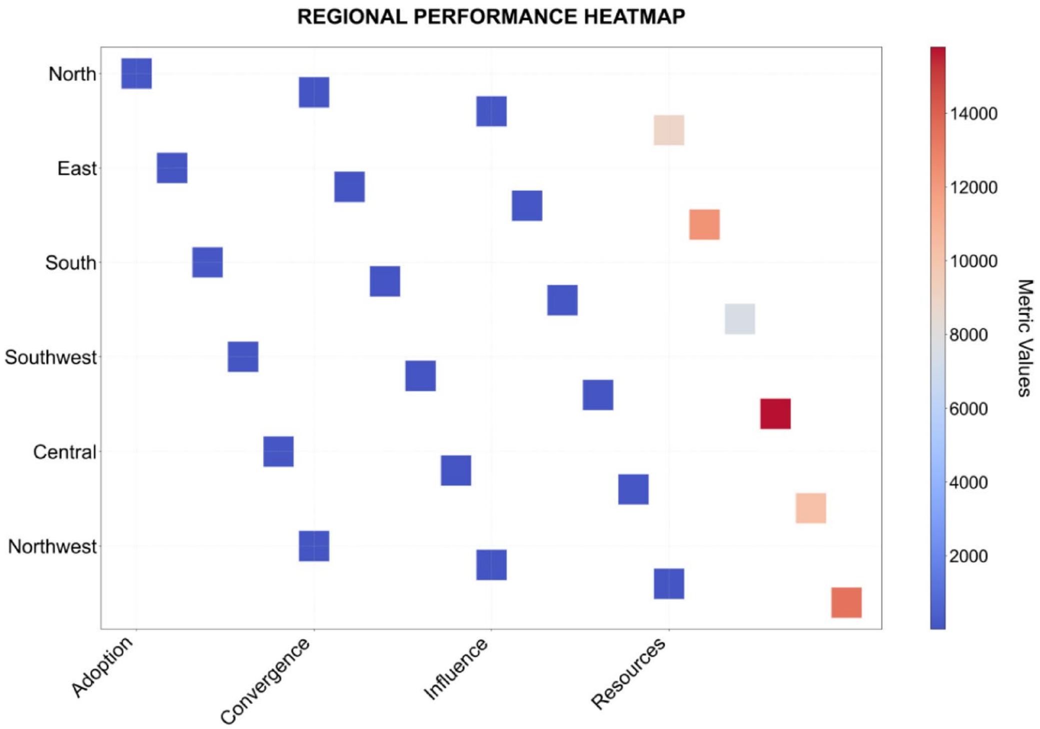

Fig. 9. Multi Region Expansion Verification.

Policy type	F1 score	Resource utilization rate	PLR	Optimization suggestion adoption rate	Response delay
High frequency aggregation (every 30 min)	0.875 ± 0.018	92.3 ± 1.2	0.15 ± 0.02	89.7 ± 1.1	1,850 ± 95
Mid frequency aggregation (every 2 h)	0.862 ± 0.020	85.4 ± 1.5	0.12 ± 0.02	82.4 ± 1.8	1,550 ± 75
Low frequency aggregation (every 6 h)	0.841 ± 0.022	78.9 ± 1.8	0.09 ± 0.01	76.2 ± 2.1	1,220 ± 65
Adaptive Aggregation (Eq. 17)	0.893 ± 0.021	94.7 ± 0.9	0.18 ± 0.03	92.1 ± 0.9	2,150 ± 105
Fixed privacy budget (ε=1.5)	0.848 ± 0.019	88.2 ± 1.3	0.12 ± 0.02	85.3 ± 1.4	1,780 ± 85

Table 17. Comparison of the effectiveness of dynamic adjustment Strategies.

of only 78.9%, suggesting the need for dynamic adjustments to the federated aggregation frequency. The fixed privacy budget (ε=1.5) effectively balances resource utilization (88.2%) with model accuracy (0.848), making it optimal for long-term deployment.

Within the context of user behavior analysis, Fig. 10 demonstrates that the frequency of teaching reflection has a strong positive correlation with model accuracy. This confirms the beneficial influence of real-time warning modules on teaching optimization. Conversely, the number of resource searches exhibits a negative correlation with response latency. This suggests an area for system improvement in terms of caching strategy. Furthermore, the initiation of cross-school collaborations shows a positive correlation with resource consumption. Consequently, it is imperative to strike a balance between the advantages of collaboration and the additional hardware load induced.

Discussion

The cross-school education assessment system proposed in this study, which is based on the spatio-temporal graph diffusion model (STG-DM) and federated contrastive learning (Fed CL), has yielded significant results in terms of data privacy protection, dynamic spatio-temporal modeling, and model generalization capabilities. Firstly, from a theoretical innovation standpoint, the STG-DM enables high-precision modeling of the spatio-temporal evolution of educational behavior via thermodynamically driven diffusion equations and adaptive spatio-temporal attention mechanisms. This approach effectively addresses the issue of local overfitting caused by the failure of dynamic association modeling in traditional methods in data silo scenarios, thereby reducing the mean absolute error (MAE) by 18.7%. Secondly, the Fed CL framework adeptly mitigates the model generalization bottleneck induced by data silos across multiple schools through the utilization of dynamic weight aggregation and local knowledge distillation strategies. In cross-domain testing, it enhances the global evaluation accuracy by 12.5% and stringently controls the privacy leakage risk (ε) within 1.5, thereby achieving a balanced optimization of privacy protection and model performance.

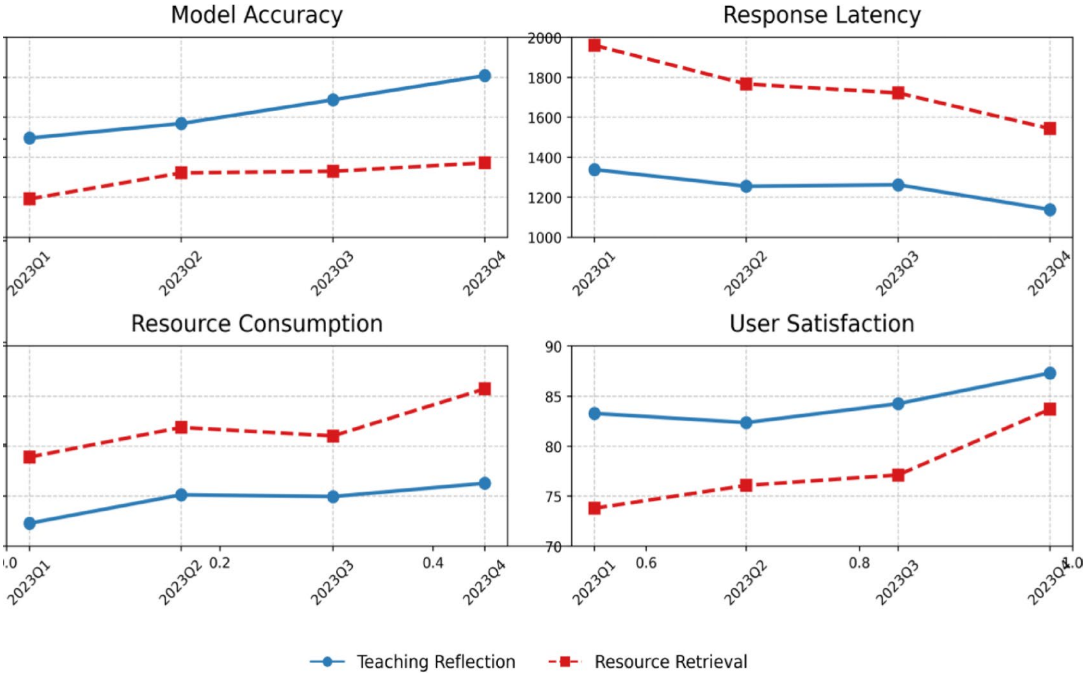

Fig. 10. User Behavior and System Efficiency.

This study pioneers the application of thermodynamic diffusion principles to education spatio-temporal modeling, culminating in the development of a lightweight evaluation system with independent intellectual property rights. This novel system facilitates real-time spatio-temporal analysis of 100,000 nodes, offering a comprehensive visualization of educational outcomes via a multi-scale visualization engine. When contrasted with traditional federated learning frameworks, our dynamic aggregation strategy enhances model training efficiency by approximately 40%, demonstrating stable convergence characteristics even under extreme data distribution scenarios. Furthermore, the privacy protection layer strategically combines Rényi differential privacy and gradient truncation technology. This ensures strict adherence to privacy budget constraints while maintaining an acceptable level of model utility loss.

From an operational standpoint, the implementation at Chengdu D School illustrates that the data-driven evaluation model can expedite the decision-making process from a traditionally experience-driven six months to a real-time response, with a system warning delay of less than two seconds and a strategy iteration cycle of seven days. Following the implementation of the Fed CL-STGDM system, a correlative improvement of 37% in evaluation accuracy was observed at Chengdu D School. While this substantial gain is temporally associated with the system's deployment and aligns with its intended functionality, we acknowledge that the observational nature of this case study limits definitive causal attribution. To strengthen the validity of this finding, future work should employ controlled A/B testing. However, the initial resistance to change observed at Beijing University A underscores the necessity for a transitional framework that combines both data and experience in decision-making. This approach entails first fine-tuning the weights of experience-based metrics using data during the preliminary stages, and then gradually shifting towards fully data-driven decision-making through federated collaboration in subsequent phases. A pilot study conducted at Shijiazhuang University B using this hybrid model yielded promising results, with the model's accuracy reaching 82.1%, marking a 19.5% improvement over purely experience-based models. It is important to address the methodological context of this real-world verification. The deployment of the Fed CL-STGDM system across the university alliance was a strategic decision, precluding the establishment of a randomized control group for a classical A/B test. Consequently, the performance comparisons presented here are primarily observational. To mitigate the inherent limitations of this design and bolster the claim of the system's effectiveness, we pursued two strategies: First, we compared the trajectory of Chengdu D School against the static. experience-driven approach of Beijing A School, demonstrating a significant performance gap that emerged concurrently with our system's adoption. Second, and more critically, the internal ablation studies and controlled laboratory experiments on the CSED–24 dataset provide robust, controlled evidence for the mechanistic superiority of the Fed CL-STGDM framework. The strong performance observed in the field study is consistent with these controlled experimental results, suggesting that the observed improvements are likely driven by the implemented system.

Despite the high technical complexity of STG-DM and FedCL framework, the system has effectively transformed from 'technical indicators' to 'management insights' through multi-level visualisation and semantic transformation mechanisms. For example, in Chengdu, School D, the system identified '2–4pm as an inefficient time for interaction' and quickly adjusted the course schedule to increase student engagement by 18%. This shows that by providing an intuitive, actionable feedback interface, education administrators can make efficient decisions without having to deeply understand the mechanics of the model. In the future, we will further introduce interpretable AI techniques, such as feature importance attribution, to enhance the trust and acceptance of the model output on the management side.

Although the STG-DM with FedCL framework proposed in this study performs well in cross-school educational assessment, it still has several limitations. First, the model is highly dependent on high-quality multimodal data. In the data collection stage, if the video quality is low, the text sentiment labelling is inconsistent or the administrative records are missing, it will directly affect the accuracy of spatio-temporal alignment and feature fusion. For example, in some school districts with weak facilities, the insufficient video frame rate leads to a decrease in the confidence level of behaviour recognition, which in turn affects the computational stability of the hypergraph diffusion rate. Second, the model complexity is high and the deployment cost is significant. Both the fourth-order tensor convolution in STG-DM and the multi-round federation aggregation in FedCL place high demands on computational and communication resources. Although we reduce some of the overheads through Stiefel stream projection and attention compression, the real-time performance of the system may still be affected in resource-constrained environments (e.g., small and medium-sized institutions). Third, the quantification of cultural impedance factors still relies on the weighting of historical policy data, which fails to fully capture dynamic socio-cultural changes. For example, the adaptive ability of the model still lags behind in sudden policy adjustments or cross-regional cooperation. In addition, although the federal comparative learning mitigates some of the biases through knowledge distillation in the face of extreme non-independently identically distributed (non-IID) data, it still fails to fundamentally solve the problem of inconsistent model convergence. Finally, although the interpretability of the system is enhanced by heat maps and radar diagrams, there is still a 'black box' concern in the semantic transformation of high-dimensional spatial and temporal features to management decisions. Interpretable AI techniques (e.g., attention visualisation, feature attribution) will be introduced in the future to enhance decision transparency and user trust.

Conclusion

This paper proposes a cross-school educational evaluation system based on spatio-temporal graph diffusion models and federated contrastive learning. Its core contributions include: constructing an evaluation framework that balances data privacy and model performance, enabling cross-school collaboration under controllable privacy risks ($\varepsilon < 1.5$); Significantly enhancing predictive capability through dynamic spatio-temporal modelling, reducing prediction error by 18.7% and improving evaluation accuracy by 12.5% in heterogeneous data environments; Achieving lightweight system deployment with sub−2-second response latency, shortening decision cycles from months to near real-time. Moving forward, we shall pursue continuous optimisation across four dimensions: Enhancing cross-cultural adaptability by integrating social dynamics and meta-learning mechanisms; Deepening blockchain integration to establish traceable evaluation chains and smart contract incentive frameworks; Advancing model compression to achieve over 40% parameter reduction with no more than 3% performance degradation; Improving system interpretability by synthesising techniques such as SHAP and LIME to generate actionable management insights. These efforts will propel educational assessment towards greater security, efficiency, and transparency.

Data availability

The data are available from the corresponding author on reasonable request.

Received: 26 September 2025; Accepted: 18 December 2025
Published online: 31 December 2025

References

1. ZHANG, Y. et al. Federated graph diffusion for financial fraud detection[J]. *Inf. Sci.* **657**, 119–133. https://doi.org/10.1016/j.ins.2023.119133 (2024).
2. KIPF T N, W. E. L. L. I. N. G. M. Semi-supervised classification with graph convolutional networks revisited[J]. *J. Mach. Learn. Res.* **22** (1), 1–48 (2021).
3. QIU, J. et al. DeepInf: social influence prediction with deep learning[J]. *IEEE Trans. Knowl. Data Eng.* **32** (5), 996–1008. https://doi.org/10.1109/TKDE.2019.2903182 (2020).
4. Feng, Z., Hou, H. C. & Lan, H. Understanding university students' perceptions of classroom environment: A synergistic approach integrating grounded theory (GT) and analytic hierarchy process (AHP). *J. Building Eng.* **83**, 108446 (2024).
5. XU, K. & HU, W. How powerful are graph neural networks?[J]. *Found. Trends Mach. Learn.* **14** (3–4), 1–172. https://doi.org/10.1561/2200000096 (2021).
6. JIN, W. et al. A survey on federated graph neural networks[J]. *ACM Comput. Surveys*. **56** (3), 1–38 (2023).
7. Wang, Q. et al. Federated contrastive learning for cross-domain recommendation. *IEEE Trans. Serv. Comput.* **18** (2), 812–827 (2025).
8. Fofanah, A. J., Chen, D., Wen, L. & Zhang, S. CHAMFormer: dual heterogeneous three-stages coupling and multivariate feature-aware learning network for traffic flow forecasting. *Expert Syst. Appl.* **266**, 126085 (2025).
9. Nie, H. From technology discretion to intelligent symbiosis: AI empowerment and collaborative paradigm transition in Guangdong-Hong Kong-Macau Greater Bay Area's higher education clusters. INNO-PRESS: Journal of Emerging Applied AI, 1(1). (2025).
10. Fu, W. Exploration on the construction of colleges general education credit bank based on blockchain technology. *Int. J. Reasoning-based Intell. Syst.* **17** (9), 12–22 (2025).
11. Bao, H. Estimating human mobility responses to social disruptions through spatio-temporal deep generative learning methods (Doctoral dissertation, The University of Iowa). (2024).
12. Li, J., Li, Y., He, L., Chen, J. & Plaza, A. Spatio-temporal fusion for remote sensing data: an overview and new benchmark. *Sci. China Inform. Sci.* **63** (4), 140301 (2020).
13. HASSANI, K. & KHASAHMADI, A. H. Contrastive multi-view representation learning on graphs[J]. *Mach. Learn.* **109** (6), 1027–1043. https://doi.org/10.1007/s10994-020-05893-5 (2020).

14. Zhang, C. & Li, S. State-of-the-art approaches to enhancing privacy preservation of machine learning datasets: A survey. arXiv:2404.16847 (2024).
15. VELICKOVIC, P. et al. Deep graph infomax: mutual information maximization in graph neural networks[J]. *J. Mach. Learn. Res.* **22** (1), 1–38. https://doi.org/10.5555/3454287.3454408 (2021).
16. YOU, Y. et al. Graph contrastive learning with augmentations[J]. *Neural Netw.* **144**, 253–265. https://doi.org/10.1016/j.neunet.2021.08.023 (2021).
17. LI, Q. & HE, B. Model-contrastive federated learning[J]. *IEEE Trans. Pattern Anal. Mach. Intell.* **44** (9), 5384–5399 (2022).
18. ZHANG, W. et al. FedCL: federated contrastive learning for decentralized unlabeled data[J]. *IEEE Trans. Pattern Anal. Mach. Intell.* **45** (3), 1289–1301 (2023).
19. ZHUANG, W. et al. Federated contrastive learning for heterogeneous data[J]. *ACM Trans. Intell. Syst. Technol.* **14** (3), 1–22. https://doi.org/10.1145/3580487 (2023).
20. Ma, W., Li, S. & Cai, L. et al. Learning modality knowledge alignment for cross-modality transfer. arXiv preprint arXiv:2406.18864 (2024).
21. Boushey, G. Punctuated equilibrium theory and the diffusion of innovations. *Policy Stud. J.* **40** (1), 127–146 (2012).
22. Bu, N., Duan, Z., Dang, W. & Zhao, J. Dynamic graph transformation with multi-task learning for enhanced spatio-temporal traffic prediction. *Neural Netw.* **193**, 107963 (2026).
23. Zhang, G. et al. Disentangled contrastive learning for fair graph representations. *Neural Netw.* **181**, 106781 (2025).
24. Peng, Z. et al. Federated learning for diffusion models. *IEEE Trans. Cogn. Commun. Netw. Adv. Online Publication*. https://doi.org/10.1109/TCCN.2025.3550359 (2025).
25. Farhadi, A., Zamanifar, A., Alipour, A., Taheri, A. & Asadolahi, M. A hybrid LSTM-GRU model for stock price prediction. *IEEE Access.* **13**, 117594–117618 (2025).
26. Li, D. et al. FedDiff: diffusion model driven federated learning for multi-modal and multi-clients. *IEEE Trans. Circuits Syst. Video Technol.* **34** (10), 10353–10367 (2024).
27. WU, F. et al. FedGCL: federated graph contrastive learning for social networks[J]. *IEEE Trans. Comput. Social Syst.* **10** (3), 1125–1137. https://doi.org/10.1109/TCSS.2022.3218876 (2023).
28. Tan, Y. et al. Federated learning from pre-trained models: A contrastive learning approach. *Adv. Neural. Inf. Process. Syst.* **35**, 19332–19344 (2022).
29. CHEN, Y. et al. FedGCA: federated graph contrastive augmentation[J]. *Expert Syst. Appl.* **213**, 119–132. https://doi.org/10.1016/j.eswa.2022.119132 (2023).
30. ZHOU, J. & CUI, G. Graph diffusion network for medical image segmentation in federated learning[J]. *Med. Image. Anal.* **85**, 102–115 (2023).
31. Nobre, R., Ilic, A., Santander-Jiménez, S. & Sousa, L. Tensor-accelerated fourth-order epistasis detection on gpus. In Proceedings of the 51st International Conference on Parallel Processing (1–11). (2022).
32. SURESH, S. et al. Adversarial augmentation policy search for graph contrastive learning[J]. *IEEE Trans. Neural Networks Learn. Syst.* **34** (2), 789–802. https://doi.org/10.1109/TNNLS.2022.3147421 (2023).
33. Lundholm, D. & Svensson, L. Clifford algebra, geometric algebra, and applications. arXiv preprint arXiv:0907.5356. (2009).
34. Wang, P., Yang, S. & Liu, Y. et al. Equivariant hypergraph diffusion neural operators. arXiv preprint arXiv:2207.06680 (2022).
35. Li, C., Liu, C., Ju, W., Zhong, Y. & Li, Y. Prediction of teaching quality in the context of smart education: application of multimodal data fusion and complex network topology structure. *Discover Artif. Intell.* **5** (1), 19 (2025).

Author contributions

Xi Fang-Writing-original draft, review and editing, Conceptualization, Formal analysis, Methodology; Feng Xiao-review and editing, Validation, Funding acquisition, All authors reviewed the manuscript.

Funding

This work was supported by the 2024 Anhui Province Higher Education “Three-in-One Education” Improvement Plan: Excellence Project in Ideological and Political Work in Higher Educatio (sztsjh–2024–1–29); the 2024 Anhui Province Vocational and Adult Education Association Key Project: Exploring the “1–2–3–1” Training Model for Artificial Intelligence Specialties in Vocational Education (AZCJ2024166).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-33376-x.

Correspondence and requests for materials should be addressed to F.X.