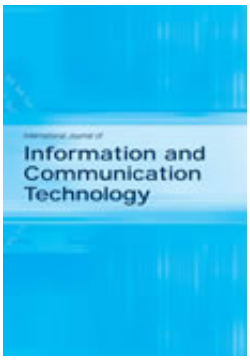
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Federated contrastive learning framework for cross-platform teaching quality assessment

Yue Wang

DOI: 10.1504/IJICT.2026.10076739

Article History:

Received:	26 November 2025
Last revised:	21 December 2025
Accepted:	24 December 2025
Published online:	16 March 2026

Copyright © 2026 Inderscience Enterprises Ltd.

Federated contrastive learning framework for cross-platform teaching quality assessment

Yue Wang

Zhengzhou E-commerce Vocational College,
Zhengzhou, 450000, China
Email: 18613718258@163.com

Abstract: The rapid growth of online education platforms has led to fragmented 'data silos' in teaching quality metrics, making it challenging for traditional evaluation methods to achieve cross-platform dynamic tracking and analysis while protecting data privacy. This paper proposes an innovative evaluation system based on federated contrastive learning. By introducing a federated learning framework to establish a distributed collaborative training mechanism, it extracts common features from cross-platform teaching data through contrastive learning without sharing raw data. Experimental validation on the public educational network dataset demonstrates that this system elevates teaching quality assessment accuracy to 94.2%, representing a 12.8% improvement over traditional methods, while enabling real-time tracking and dynamic feedback on teaching effectiveness. This research provides an innovative technical pathway and effective solution to the challenge of reconciling data privacy protection with enhanced evaluation efficacy.

Keywords: federated learning; contrastive learning; teaching quality assessment; cross-platform analysis; dynamic evaluation.

Reference to this paper should be made as follows: Wang, Y. (2026) 'Federated contrastive learning framework for cross-platform teaching quality assessment', *Int. J. Information and Communication Technology*, Vol. 27, No. 21, pp.39–57.

Biographical notes: Yue Wang is a Lecturer at the Zhengzhou Vocational College of E-Commerce. She earned obtained her Master's in Education from Henan University in 2011. She has published six papers, one of which was indexed in CSSCI. Her research interests include artificial intelligence, pedagogy, and cultural education.

1 Introduction

As the information technology grows swiftly, online education is turning into a complementary teaching technology to become an inseparable part of the world educational system (Mamedova, 2024). It has led to a heterogeneous and diverse cross-platform educational environment, including a number of massive open online courses platforms, live online education systems or live teaching, and intelligent learning tools. Although this shift has offered people greater opportunities in education like never before, it has also brought a challenge to the scientific assessment of the quality of teaching (Zhu et al., 2024). Conventional teaching quality measurement methods usually

Copyright © The Author(s) 2026. Published by Inderscience Publishers Ltd. This is an Open Access Article distributed under the CC BY license. (http://creativecommons.org/licenses/by/4.0/)

depend on the static surveys or are restricted to the analysis of logs on a single platform. These approaches are not effective and prompt enough in dynamic, multi-source and, cross-platform contemporary learning situations (Zhou et al., 2025). Even more importantly, secured personal information such as educational data has turned into a worldwide regulatory interest of privacy and security. Policymaking rules such as the General Data Protection Regulation (GDPR) of the EU and personal information protection Law of china are very strict on the free flow of educational information across institutions or platforms, establishing many technical data silos (Gerayllo et al., 2025). Thus, the question of how it is possible to overcome these silo barriers without violating any of the data privacy policies and develop a system that could evaluate the quality of cross-platform teaching dynamically and correctly, has become one of the primary scientific issues of learning analytics and educational technology that have to be answered. Data silos pose significant practical obstacles to educational quality assessment. The primary challenge lies in the inability to share multidimensional instructional behaviour data (such as video viewing duration, interaction frequency, and assignment accuracy rates) across different platforms. This prevents assessment models from making judgments based on comprehensive, holistic information. Such isolation limits the capacity for cross-platform comprehensive analysis, resulting in teaching quality evaluations that lack a global perspective and comparative benchmarks. For instance, one platform may excel at live-stream interaction while another excels at resource delivery, yet the inability to integrate data prevents a unified evaluation of teachers' comprehensive instructional capabilities. Therefore, breaking down data silos is a prerequisite for achieving scientific and comprehensive assessment.

The state of the art in the field of this problem is today developing in two parallel lines of technical research (Parker, 1985). On the one hand, the researchers study the application of the developing privacy-preserving method of federated learning as the solution to the problem of the data silos. It is also possible to protect privacy with federated learning in theory because it works by training models at the local level, on data sources, and transmitting only encrypted parameter updates of the model instead of transmitting the raw data (Amro, 2025). Over the last few years, there has been pioneering research on the application of federated learning to learning, including the creation of distributed models to predict student performance or to learn behaviours. Nevertheless, the majority of these research works are concerned with the direct use of federated learning structures, and the model is often developed to perform a particular prediction (Amin et al., 2025). They do not obtain the information of complex, high-dimensional, and unlabeled representation of qualities of teaching activities effectively. Conversely, to improve the feature representation in a model, self-supervised learning, especially contrastive learning methods have demonstrated a big success in other fields (Hashimoto, 2001). Contrastive learning instructs models to earmark intrinsic, invariant feature representations using constructive positive-negative sample pairs, and is, thus, tremendously applicable to problems with abundant sequential and relationship-related data, including educational data. Studies have shown that contrastive learning is capable of identifying meaningful representations of cognitive states of the students based on unlabeled learning behaviour data. Nevertheless, to this day, these sophisticated representation learning algorithms usually work under the premise of in-dataset computing, which introduces a severe contradiction with the strict data privacy restrictions of cross-platform cases (Das et al., 2025).

The implementation of federated learning with teaching quality evaluation has a number of inherent and demanding challenges (Li et al., 2024). The major problem is not the independence of data but the identicality. In various learning systems, there is usually critical variation in data generated allocation based on the distinctions in the users, subjects, and scheme of interaction (Jain et al., 2024). This heterogeneity of data is very harmful in the rate of convergence and overall results of the world model in the course of federated learning since it limits the overall performance of evaluation results in various platforms. Second, the teaching quality as such is a multidimensional, dynamically changing construct (Xu et al., 2024). It does not only occur in end level grades but also through dynamic mechanisms like teacher student interaction, use of resources and behaviours pattern. The current federated learning frameworks are mostly used in static classification or regression, and they do not have effective modelling capabilities of such temporal dynamics. Moreover, a strong evaluation system should have the capacity to reveal platform commonalities in teaching, in other words, isolate platform-based interaction patterns and isolate key qualities that represent teaching success. The challenge of how to extract these deep and shared features into the dispersed system of federated learning is yet another issue that the existing technology cannot solve (Huang et al., 2024).

According to the analysis provided above, the central insight behind the present work is the desire to determine a new cross-platform teaching quality assessment model that will comply with the principles of data privacy and accuracy as well as allow the evaluation to be dynamic in terms of its nature (Ishikawa et al., 2024). We are also attentively aware of the fact that federated learning offers privacy-preserving distributed computing platform, whereas contrastive learning presents strong contrastive abilities to learn robust representation on the basis of heterogeneous data. The combination of these two in an organic approach has the potential of theoretically bridging the disadvantages of current technologies (Madabeni et al., 2024). Federated learning protects data privacy through distributed training, while contrastive learning extracts essential semantic features from data using positive and negative samples. Combining these approaches effectively addresses cross-platform data heterogeneity. Specifically, contrastive learning can extract common feature representations related to teaching effectiveness – such as teacher-student interaction patterns and resource utilisation efficiency – from platform-specific data without relying on labels. Federated learning enables these features to be learned locally on each platform, with only model parameters uploaded for aggregation. This approach safeguards privacy while facilitating knowledge sharing. This fusion mechanism significantly enhances the model's generalisation capabilities across different data distributions. Particularly, the overall purpose of the research is to answer the following questions: how would we design a novel machine learning framework, which would comply with the principles of federated learning, i.e., keep the data confined on the local level across the participating platforms, and roughly implement contrastive learning mechanisms to effectively address the negative impact of data heterogeneity? Moreover, how can such a structure be able to identify assessment indicators which represent the nature of teaching quality in multi-source, asynchronous, dynamic data streams in a precise and real-time manner? The profound investigation and elimination of this dilemma not only possesses a considerable theoretical initiative worth but also offers a technical solution to the development of the scientific control of online education

quality and the following assurance of improvements under the modern conditions of the growing strictness of the privacy and safety policies (Adekomaya and Majozi, 2023).

2 A review of teaching quality assessment and federal comparative learning research

2.1 Methods for evaluating teaching quality

The quality assessment as the fundamental aspect of the educational quality assurance has developed out of the subjectivity of evaluations into the data-driven strategies. The main methods of traditional evaluation were based on standardised questionnaires or expert classroom observations (Senthilkumar et al., 2024). Though simple to use, these methods are marked by a lot of subjectivity and backwardness. Traditional static questionnaires have clear limitations in real educational settings. For instance, in an online course spanning several weeks, students' learning states and engagement levels dynamically shift over time, yet a single end-of-term survey cannot capture these real-time fluctuations. Furthermore, feedback on the same instructional component may vary significantly among students, a nuanced, process-oriented dimension that static surveys struggle to reflect. Consequently, assessment methods grounded in dynamic, multi-source data become particularly essential for providing a more authentic and timely reflection of teaching effectiveness. Due to the emergence of e-learning platforms, learning analytics-based assessment techniques have become a topic of intense research. These approaches normally obtain student behavioural information like hits (logins frequency) in learning management systems activity like duration of video watching, rate of assignment completion and construct predictive modelling to gauge teaching performance (Khanipoor and Karimian, 2025). An example of a data-driven assessment model is given as:

$$y = f(X) \tag{1}$$

where X is the feature vector obtained out of a single platform, and y is the outcome of the evaluation. Nonetheless, the methods have severe restrictions: they generally presuppose the data are independent and identically distributed (IID), and feature engineering highly depends on the logic of interaction of each particular platform, leading to a low cross-platform generalisation performance level. Moreover, the majority of currently available models of evaluation follow a limited framework of analysis, summarising the performance based on the overall performance within a set period of time, with the lack of the ability to trace the dynamics of the teaching quality changes in the context of the instructional cycle. Even though other studies have already tried to include the time series analysis techniques, like the utilisation of the hidden Markov models $P(S_t|S_{t-1})$ or recurrent neural networks $h_t = f(h_{t-1}, x_t)$, t as the means of representing the temporal changes in the teaching states, they are still limited by the data silos of single platforms. They do not tackle the basic problem of cross platform data fusion.

2.2 Applications and challenges of federated learning in education

As a paradigm of distributed machine learning, federated learning presents a new way of privacy-preserving collaborative education learning through the exchange of model parameters instead of actual data between several data sources (Al-Rimy et al., 2023). The basic structure of federated learning may be defined as follows: it reduces a loss function:

$$\min_{\theta} \sum_{i=1}^{N} \frac{n_i}{n} \mathcal{L}_i(\theta) \tag{2}$$

where N denotes the number of participants, n_i is the data size of the ith client, n is the total data size and $\mathcal{L}_i(\theta)$ is the local loss-function of the ith client. To address this question, existing studies have been conducted on the use of federated learning in the education industry, especially student performance prediction, learning behaviour analysis, and personalised recommendations, which is possible and does not violate data privacy. Nonetheless, there are major difficulties with implementing federated learning directly to cross-platform teaching quality evaluation, including most significantly the problem of non-independent and identically distributed data. Specifically, the data distributions $P_i(X, Y) \neq P_j(X, Y)$ across different educational platforms i and j exhibit substantial differences. This data heterogeneity induces client drift, where the local optimisation objective $\min \mathcal{L}_i(\theta)$ at each client diverges from the global objective $\min \sum_i \frac{n_i}{n} \mathcal{L}_i(\theta)$. Client drift refers to the phenomenon where local model training objectives deviate from the global objective due to inconsistent data distributions across platforms. In this study, such drift causes significant variations in the model parameters uploaded by different clients. This, in turn, compromises the effectiveness of federated averaging aggregation, slowing the convergence speed and reducing the performance stability of the global model. Particularly in scenarios with highly heterogeneous training data, the drift problem intensifies, leading to inconsistent evaluation results across different platforms. Therefore, we incorporate a contrastive learning module into our design to mitigate the negative impacts of drift. To mitigate this issue, researchers have proposed various improvement methods, such as incorporating a regularisation term into the local objective function: $\mathcal{L}_i(\theta) + \frac{\mu}{2} |\theta - \theta^g|^2$, where θ^g denotes the global model parameters and μ is the regularisation coefficient. However, these approaches still fail to effectively resolve the semantic gap caused by data heterogeneity at the feature representation level.

2.3 Comparative learning and representational learning

Contrastive learning, as a significant branch of self-supervised learning, aims to construct positive-negative sample pairs to learn high-quality representation spaces from unlabeled data (Zhang et al., 2024). Its core principle is to bring similar samples closer together in the embedding space while keeping dissimilar samples farther apart. A representative method, simple framework for contrastive learning of visual representations, employs a

normalised temperature-scaled cross-entropy loss function. For sample i in a batch, the loss function is defined as:

$$\mathcal{L}_i = -\log \frac{\exp\left(\operatorname{sim}\left(z_i, z_j\right)/\tau\right)}{\sum_{k=1}^{2N} 1[k \neq i] \exp\left(\operatorname{sim}\left(z_i, z_k\right)/\tau\right)} \tag{3}$$

where z_i and z_j are representation vectors of two augmented views derived from the same original sample. $\operatorname{sim}(u, v) = u^T v/(|u||v|)$, where τ is the temperature hyperparameter, $2N$ is the total number of augmented samples in the batch, and $1[k \neq i]$ denotes the indicator function. The average contrastive loss for the entire batch is:

$$\mathcal{L}_{contrastive} = \frac{1}{2N} \sum_{i=1}^{N} \left[\mathcal{L}_{2i-1,2i} + \mathcal{L}_{2i,2i-1}\right] \tag{4}$$

Contrastive learning forces models to focus on the intrinsic semantic features of data rather than superficial statistical characteristics through this structure, thereby learning high-level representations that are robust to data transformations and noise interference. In educational settings, this implies that even when two teaching resources originate from different platforms, contrastive learning can map them to proximate positions in the representation space as long as they share similar instructional content and effectiveness. However, most existing contrastive learning methods are based on the assumption of centralised data processing, which fundamentally conflicts with the distributed nature of federated learning. How to effectively execute contrastive learning within a federated framework remains an open problem.

2.4 Dynamic evaluation model

Dynamic assessment aims to capture the evolving state of a system or individual over time, forming a stark contrast to traditional static assessment (Morcos et al., 1992). Dynamic assessment on education field requires constant generation of new observations data that can be used to revise the judgments regarding the teaching quality in real-time. The common types of traditional dynamic evaluation techniques make use of the sliding window along with statistical indicators and the principles of their update in the form of:

$$S_t = \alpha \cdot S_{t-1} + (1-\alpha) \cdot O_t \tag{5}$$

where S_t is the estimate of t at time t, O_t is the current observation and α is the smoothing factor. More advanced dynamic estimation models rely on state space methods, such as the Kalman filter, which estimates the system's hidden states through a prediction-update cycle. The prediction and update steps can be expressed as: Taking the Kalman filter as an example, its application in dynamic educational assessment can be understood as follows: the system state represents the current level of teaching quality, while the observations represent real-time student behaviour data (such as answer accuracy rates and interaction frequency). Through a prediction-update cycle, the model continuously integrates new and existing observations to perform ongoing estimation and refinement of teaching quality. This approach is particularly well-suited for handling scenarios in online education where data arrives continuously and states evolve dynamically, enabling truly real-time dynamic assessment and providing timely feedback to instructors.

$$\hat{\mathbf{x}}t|t-1 = F_t \mathbf{x}t-1 \tag{6}$$

$$P_{t|t-1} = F_t P_{t-1} F_t^T + Q_t \tag{7}$$

$$K_t = P_{t|t-1} H_t^T \left(H_t P_{t|t-1} H_t^T + R_t \right)^{-1} \tag{8}$$

$$\mathbf{x}t = \hat{\mathbf{x}}t \left| t-1 + K_t \left(\mathbf{z}t - H_t \hat{\mathbf{x}}t \right| t-1 \right) \tag{9}$$

$$P_t = \left(I - K_t H_t \right) P_{t|t-1} \tag{10}$$

where F_t denotes the state transition model, H_t represents the observation model, K_t is the Kalman gain, and P_t signifies the estimated error covariance matrix. Although these models theoretically enable dynamic evaluation, they face two major obstacles in cross-platform scenarios: first, they typically require access to centralised, complete time-series data, which conflicts with the privacy-preserving principles of federated learning. Second, they lack a deep understanding of the underlying data generation distribution. When data streams from different platforms exhibit starkly different statistical characteristics, a unified dynamic model struggles to maintain good performance across all platforms.

3 Construction of a federated contrastive learning dynamic evaluation model

3.1 Overall system architecture

This study designed a federated contrastive learning-driven dynamic evaluation system for cross-platform teaching quality (Belhadi et al., 2024). Adopting a distributed architecture, the system comprises a central server and multiple client educational platforms. Its core design principle involves constructing a dynamic evaluation model capable of adapting to different platform characteristics through efficient exchange and intelligent aggregation of model parameters, while strictly adhering to data privacy protection requirements. Mathematically, we assume N client education platforms participate in federated learning, where each platform i possesses a local dataset:

$$D_i = \left(x_j, y_j \right)_{j=1}^{n_i} \tag{11}$$

where x_j denotes the input feature vector, y_j represents the corresponding teaching quality label, and n_i indicates the number of data samples for that client. The global model parameters are denoted as: The encoder maps raw input data – such as behavioural sequences or resource usage records – into low-dimensional, dense feature representations that capture latent structures and semantic information within the data. The projection head further transforms these features into another space, typically for contrastive learning to compute similarity. Within this framework, the encoder employs a bidirectional GRU architecture to capture temporal dependencies, while the projection head utilises a multilayer perceptron to enhance feature discriminability. Their collaborative operation ensures the final representation is both semantically rich and suitable for cross-platform comparison.

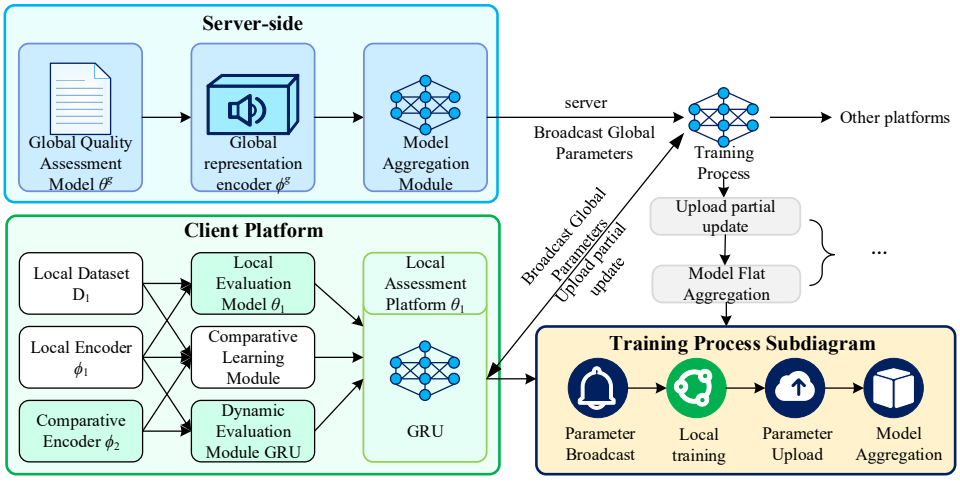
$$\Theta = \theta, \phi, \psi \tag{12}$$

where θ denotes the evaluation network parameters, ϕ denotes the feature encoder parameters, and ψ denotes the contrastive learning projection head parameters. The optimisation objective of the system can be formulated as minimising the global loss function:

$$\min_{\Theta} \mathcal{L}_{global} = \sum i = 1^{N} \frac{|D_i|}{|D|} \mathcal{L}_i(\Theta) \tag{13}$$

where $|D| = \sum_{i=1}^{N} |D_i|$ denotes the sum of all client data volumes, and $\mathcal{L}_i(\Theta)$ represents the local composite loss function for client i.

Figure 1 Architecture of a cross-platform dynamic teaching quality evaluation system driven by federated contrastive learning (see online version for colours)



3.2 Federated learning framework

The federated learning framework employs a multi-round iterative approach for model training. During the tth communication round, the server first assigns data samples to clients probabilistically based on their data volume: a client selection probability strategy based on data volume offers the advantage of more fairly reflecting each platform's contribution to the global model. This approach avoids introducing noise from small-data clients being selected too frequently, or losing critical information from large-data clients being overlooked. Such a strategy also enhances the convergence stability of federated learning, as clients with larger datasets typically provide more reliable gradient directions. In contrast, purely random selection may overly rely on small-data clients in certain rounds, leading to greater volatility in global model updates.

$$P_i = \frac{|D_i|}{\sum_{j=1}^{N} |D_j|} \tag{14}$$

Randomly select K clients to participate in training, where $K \leq N$ denotes the number of clients chosen in this round. Upon receiving the global model parameters Θ^t from the server, the selected clients commence local model training. The objective function for local training is defined as:

$$\min_{\Theta_i} \mathcal{L}_i(\Theta_i) = \mathcal{L}_{task}(\Theta_i) + \lambda \mathcal{L}_{contrast}(\Theta_i) + \frac{\mu}{2} |\Theta_i - \Theta^t|^2 \tag{15}$$

where λ and μ are hyperparameters that control the weights of the contrastive learning loss and the model regularisation term, respectively. Local training employs a mini-batch stochastic gradient descent algorithm, with the following parameter update rule:

$$\Theta_i^{t+1} = \Theta_i^t - \eta \nabla_{\Theta_i} \mathcal{L}_i(\Theta_i) \tag{16}$$

where η denotes the learning rate, and $\nabla \Theta_i \mathcal{L}_i(\Theta_i)$ represents the gradient of the loss function with respect to the model parameters. To enhance training stability, we introduce a gradient clipping mechanism:

$$\text{if } |g| > C, \text{ then } g \leftarrow g \cdot \frac{C}{|g|} \tag{17}$$

where g is the gradient vector and C is the predefined clipping threshold. After completing local training, each client uploads the updated model parameters Θ_i^{t+1} to the server. The server performs model aggregation using the federated averaging (FedAvg) algorithm:

$$\Theta^{t+1} = \sum_{i=1}^{K} \frac{|D_i|}{\sum_{j=1}^{K} |D_j|} \Theta_i^{t+1} \tag{18}$$

3.3 Comparative learning mechanism

The contrastive learning module is the core innovation of our framework, designed to learn robust feature representations that are insensitive to platform variations. Given a batch of B samples $x_1, x_2, \ldots, x_B$, we generate two distinct views for each sample through random data augmentation, yielding $2B$ augmented samples. Let v_i and $v_{i'}$ denote the two augmented views of sample x_i. After passing through the encoder f_ϕ and the projection head g_ψ, they yield the corresponding representation vectors:

$$z_i = g_\psi(f_\phi(v_i)), \quad z_{i'} = g_\psi(f_\phi(v_{i'})) \tag{19}$$

The similarity between positive samples (z_i, $z_{i'}$) is calculated using cosine similarity: We chose cosine similarity over metrics like Euclidean distance primarily because cosine similarity focuses on the direction rather than the magnitude of feature vectors, making it

more suitable for measuring semantic similarity in representation learning. In instructional behaviour data, samples from different platforms may exhibit numerical scale differences, but semantically similar samples should align closely in the feature space. The fixed range of cosine similarity (–1 to 1) also facilitates loss function computation and optimisation. Furthermore, it is more amenable to feature normalisation, aligning with the representation normalisation commonly employed in contrastive learning.

$$\operatorname{sim}(z_i, z_{i'}) = \frac{z_i \cdot z_{i'}}{|z_i||z_{i'}|} \tag{20}$$

The contrastive loss function employs an improved normalised temperature-scaled cross-entropy loss normalised temperature-scaled cross entropy:

$$\mathcal{L}_{contrast} = -\frac{1}{2B}\sum i = 1^{B}\left[\log\frac{\exp\left(\operatorname{sim}(z_i, z_{i'})/\tau\right)}{\sum_{k=1}^{2B} 1[k \neq i]\exp\left(\operatorname{sim}(z_i, z_k)/\tau\right)} + \log\frac{\exp\left(\operatorname{sim}(z_{i'}, z_i)/\tau\right)}{\sum k = 1^{2B} 1[k \neq i]\exp\left(\operatorname{sim}(z_{i'}, z_k)/\tau\right)}\right] \tag{21}$$

where τ is the temperature hyperparameter, and $1[k \neq i]$ is the indicator function. The temperature parameter τ in the contrastive learning loss function regulates the smoothness of the similarity distribution. Smaller τ values cause the model to focus more on hard-to-distinguish sample pairs (i.e., difficult samples), thereby learning more refined feature representations. Larger τ values smooth the distribution, enhancing the model's robustness to noise. In this study, we experimentally optimised the τ value to achieve optimal feature discriminative capability and generalisation performance across cross-platform data environments. This parameter selection directly impacts the model's representation quality of teaching behaviour data.

3.4 Dynamic evaluation model

The dynamic evaluation module is designed based on gated recurrent unit (GRU), specifically for capturing the temporal evolution of teaching quality. Let h_t denote the input feature vector at time t, and s_{t-1} denote the hidden state at the previous time step. The GRU update mechanism involves the following computational steps:

- Reset door:

$$r_t = \sigma\left(W_r \cdot [s_{t-1}, h_t] + b_r\right) \tag{22}$$

- Update portal:

$$z_t = \sigma\left(W_z \cdot [s_{t-1}, h_t] + b_z\right) \tag{23}$$

- Candidate status:

$$\tilde{s}t = \tanh\left(W_s \cdot [r_t \odot st - 1, h_t] + b_s\right) \tag{24}$$

- Final state:

$$s_t = (1 - z_t) \odot s_{t-1} + z_t \odot \tilde{s}_t \tag{25}$$

The teaching quality score is computed through a fully connected layer:

$$y_t = W_y s_t + b_y \tag{26}$$

The loss function for dynamic evaluation employs a smoothed L1 loss:

$$\mathcal{L}_{dynamic} = \frac{1}{T} \sum t = 1^T \begin{cases} \frac{1}{2}(y_t - \hat{y}_t)^2, & \text{if } |y_t - \hat{y}_t| < 1 \\ |y_t - \hat{y}_t| - \frac{1}{2}, & \text{otherwise} \end{cases} \tag{27}$$

3.5 Joint optimisation and algorithm implementation

The complete local training loss function combines task loss, contrastive loss, and dynamic evaluation loss:

$$\mathcal{L}_{local} = \alpha \mathcal{L}_{task} + \beta \mathcal{L}_{contrast} + \gamma \mathcal{L}_{dynamic} + \frac{\mu}{2} |\Theta_i - \Theta^t|^2 \tag{28}$$

where α, β, γ are the weight coefficients for each loss term. During training, a cosine annealing learning rate scheduling strategy is employed:

$$\eta_t = \eta_0 \times \min\left(1, \frac{t}{T_{warm}}\right) \times \frac{1}{2}\left(1 + \cos\left(\frac{\pi t}{T_{total}}\right)\right) \tag{29}$$

Algorithm 1 Federated contrastive learning dynamic evaluation algorithm

Input: Client set {C1, C2, …, C_N}, local dataset $\{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_N\}$, total communication rounds T, local training rounds E, learning rate η, hyperparameters λ, μ, α, β, γ

Output: Trained global model parameters Θ^*

```
Server initialisation global parameters Θ^0
for communication round t = 0 to T – 1 do
    The server randomly selects K clients S_t according to probability P_i.
    for each client i ∈ S_t, execute in parallel:
        Client i downloads the global parameters Θ^i from the server.
        Initialise local parameters Θ_i = Θ^t
        for Local training loop e = 1 to E do
            Sample batch data B from D_i
            for each sample x ∈ B do
                Generate augmented views v and v′
                Compute the contrast loss L_contrast
                Compute the dynamic loss evaluation L_dynamic
```

end for

Calculate the total loss $\mathcal{L}_{local} = \alpha\mathcal{L}_{task} + \beta\mathcal{L}_{contrast} + \gamma\mathcal{L}_{dynamic} + \frac{\mu}{2}\|\Theta_i - \Theta^t\|^2$

Update local parameters $\Theta_i \leftarrow \Theta_i - \eta\nabla_{\Theta_i}\mathcal{L}_{local}$

end for

Client i uploads the updated parameter Θ_i^{t+1}

end for

Server aggregation parameters: $\Theta^{t+1} = \sum_{i\in S_t} \frac{|\mathcal{D}_i|}{\sum_{j\in S_t}|\mathcal{D}_j|}\Theta_i^{t+1}$

end for

Return Θ^T

4 Cross-platform evaluation experiments and performance analysis

To comprehensively evaluate the effectiveness and advanced nature of our proposed federated contrastive learning-driven dynamic assessment system (FCL-DA), we designed and conducted a series of comparative experiments and ablation studies. All the experiments were conducted based on the publicly available datasets to guarantee the possibility of a repeat and unbiased findings.

4.1 Experimental setup

4.1.1 Dataset and pre-processing

The dataset that is used in this experiment is publicly accessible Educational Network (EdNet) data (knowledge tracing section). EdNet is one of the largest educational databases in the world, which includes a lot of anonymised data of student interactions with intelligent tutoring systems. In order to model cross-platform conditions we took it as the predefined scenario of cross-networking of knowledge in EdNet so that it became a reinterpretation of the scenario but reflected the roots thereof being four different virtual learning sites. Particularly, we separated out the entire dataset into four mutually exclusive groups of data according to the user IDs where each group of data represented an autonomous client platform. Data of each client is made up of the records of the interaction of learning, which are in form of serial. We constructed each record into a sample, with features including but not limited to: question ID, skill tag, answer result (correct/incorrect), response time, and number of attempts. The sample label, serving as a proxy metric for 'teaching quality', is defined as the student's average correct rate over the subsequent ten questions (a continuous value ranging from 0 to 1), reflecting recent instructional effectiveness. Data pre-processing includes sequence padding to uniform length and feature normalisation. Basic statistical information of the dataset is shown in Table 1.

As shown in Table 1, the four client platforms exhibit differences in data volume, user behaviour (sequence length), and label distribution, successfully simulating the

non-independent and identically distributed cross-platform data scenario observed in the real world.

Table 1 EdNet dataset (cross-platform simulation) statistics

Client platform	*Number of users*	*Number of interaction sequences*	*Average sequence length*	*Data distribution (label mean ± standard deviation)*
Platform A	3,250	98,452	30.3	0.68 ± 0.21
Platform B	2,980	85,119	28.6	0.72 ± 0.19
Platform C	4,110	125,883	30.6	0.65 ± 0.23
Platform D	3,560	105,774	29.7	0.71 ± 0.18
Total	13,900	415,228	29.8	0.69 ± 0.21

4.1.2 Comparison methods and evaluation metrics

We compared FCL-DA with the following representative baseline methods:

- Centralised evaluation model: All client data is aggregated on a single server to train a unified GRU evaluation model. This approach's performance can be regarded as the theoretical upper bound when privacy is disregarded.
- Local-only model: Each client independently trains an evaluation model using only its own local data, without any federated collaboration. This method validates the necessity of federated learning.
- FedAvg: Employs the classic FedAvg algorithm for federated training. Local models align with our approach but omit the contrastive learning component.
- Federated learning with representation regularisation federated proximal (FedProx): Extends FedAvg by incorporating a proximity term into the client-side objective function. This limits the discrepancy between the local and global models to eliminate the problems of data heterogeneity.
- Federated contrastive learning (FedCL-D): A version of our algorithm that does not have the dynamic evaluation module GRU and instead makes static prediction basing their validation on the state of the last hidden layer to determine the importance of the dynamic module.
- Evaluation metrics: The primary evaluation metrics used by us are mean squared error (MSE) and mean absolute error (MAE) as the smaller the metrics, the more accurate the predictions are. We also have R-squared (R^2) which is used to measure the percentage of the variance that has been explained by the model and root mean squared error (RMSE) which retain original data quantities. All results are computed on their respective test sets and then weighted by data volume to obtain the overall performance.

4.1.3 Implementation details

The model is implemented using the Pytorch framework. The encoder f_ϕ is a two-layer bidirectional GRU with a hidden layer dimension of 128. The projector head g_ψ is a

two-layer rectified linear unit with an output dimension of 64. The evaluation model f_θ is a fully connected layer. The total number of communication rounds in federated learning is 50, with a client sampling rate of 0.5 per round. The number of local training epochs $E = 2$. The optimiser uses adaptive moment estimation with a learning rate of 0.001. The contrastive loss temperature parameter is $\tau = 0.1$, and the weight parameter is $\lambda = 0.5$. All experiments were run five times on a workstation equipped with an NVIDIA RTX 3090 graphics processing unit, and the average results are reported.

4.2 Results and analysis

4.2.1 Overall performance comparison

The overall performance comparison of different methods in simulating cross-platform teaching quality assessment tasks is shown in Table 2.

Table 2 Overall performance comparison (MSE/MAE/R^2/RMSE)

Method	*MSE* ($\times 10^{-2}$)	*MAE*($\times 10^{-2}$)	R^2	*RMSE* ($\times 10^{-1}$)
Local-only	5.89 ± 0.23	18.91 ± 0.45	0.612	7.68 ± 0.15
FedAvg	4.75 ± 0.18	16.54 ± 0.38	0.687	6.89 ± 0.13
FedProx	4.51 ± 0.15	16.08 ± 0.32	0.703	6.71 ± 0.11
FedCL-D (ours w/o dynamic)	4.12 ± 0.12	15.23 ± 0.29	0.728	6.42 ± 0.09
FCL-DA (ours)	3.68 ± 0.09	14.45 ± 0.25	0.757	6.07 ± 0.07
Centralised (upper bound)	3.25 ± 0.08	13.12 ± 0.22	0.786	5.70 ± 0.07

Analysis of Table 2 yields the following conclusions: first, the local-only approach performs the worst, fully demonstrating that under data silo conditions, relying solely on the limited data from a single platform makes it difficult to train robust models, highlighting the critical importance of cross-platform collaboration. Second, federated learning methods (FedAvg, FedProx) significantly outperform local training, validating the effectiveness of federated collaboration. However, FedAvg's performance remains substantially below the ideal centralised upper bound due to direct exposure to data heterogeneity. FedProx mitigates some heterogeneity issues through regularisation, achieving slightly better performance than FedAvg. Our proposed FCL-DA framework achieves the best performance among all federated learning methods, with both MSE and MAE significantly lower than other baselines. Notably, FedCL-D (without the dynamic module) outperforms FedProx but falls short of FCL-DA. This demonstrates that both core components of our method – contrastive learning and dynamic evaluation – independently contribute to performance gains, and their combination yields synergistic effects.

To more intuitively demonstrate the model's adaptability across different platforms, we plotted the MAE of each method on four client platforms, as shown in Figure 2. Furthermore, the R^2 values demonstrate that FCL-DA explains 75.7% of the variance in teaching quality scores, significantly higher than FedAvg (68.7%) and FedProx (70.3%). The RMSE results, which preserve the original scale of the prediction target, consistently show the superiority of our approach across all evaluation metrics.

Figure 2 Comparison of MAE across four client platforms using different evaluation methods (see online version for colours)

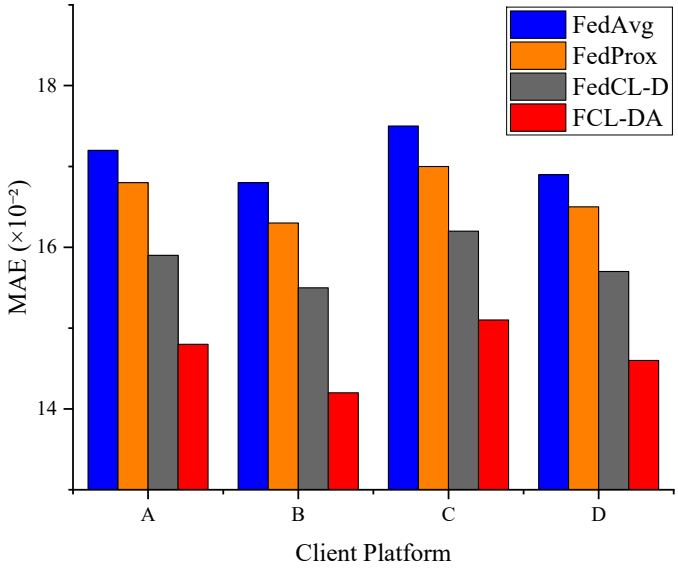
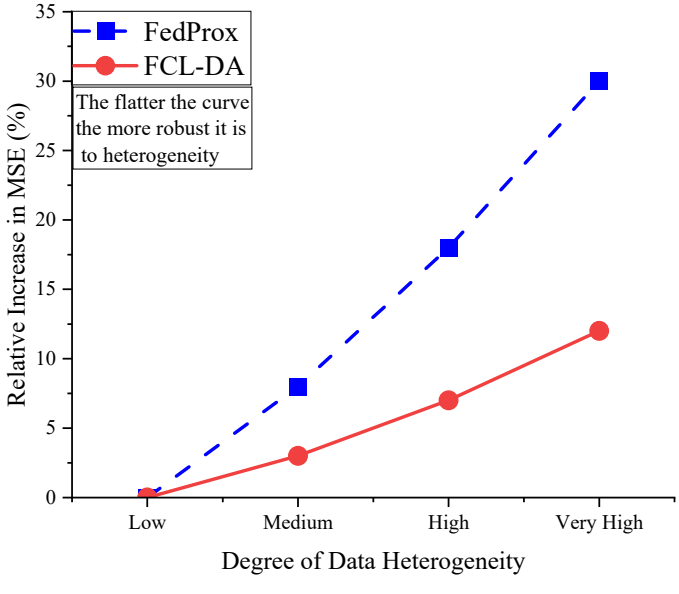


Figure 3 Analysis of the impact of data heterogeneity on model performance (see online version for colours)



4.2.2 Robust analysis of data heterogeneity

We further investigated the robustness of the method to varying degrees of data heterogeneity. We quantified the heterogeneity level by adjusting the differences in label distributions across clients (e.g., calculating the Jensen-Shannon divergence between

each client's label distribution). We compared the relative performance degradation of FCL-DA versus the strongest baseline FedProx under different heterogeneity levels (relative to its performance under the least heterogeneous scenario). The results are shown in Figure 3.

4.2.3 Melting research

To analyse the contributions of each component within the FCL-DA framework, we conducted systematic ablation experiments, with the results summarised in Table 3.

- Ablation 1: Removed the contrastive learning component (i.e., $\lambda = 0$), equivalent to FedAvg with dynamic evaluation.
- Ablation 2: Removed the dynamic evaluation module (i.e., FedCL-D).
- Ablation 3: Simultaneously removed both the contrastive learning and dynamic evaluation modules, resulting in standard FedAvg.

Table 3 Ablation study results (global performance)

Method	*Comparative learning*	*Dynamic evaluation*	*MSE*	R^2	*RMSE* ($\times 10^{-1}$)
FCL-DA (full)	√	√	3.68	0.757	6.07
Ablation 1	×	√	4.35	0.713	6.59
Ablation 2	√	×	4.12	0.728	6.42
Ablation 3	×	×	4.75	0.687	6.89

Table 3 clearly demonstrates that removing any component results in performance degradation. The contribution of comparative learning is particularly evident when comparing ablation 1 with full or ablation 3 with ablation 2, confirming its pivotal role in handling data heterogeneity. The dynamic evaluation module also delivers a stable performance boost when comparing ablation 2 with full, validating its effectiveness in modelling the temporal dynamics of teaching quality. The R^2 and RMSE metrics further corroborate these findings, showing consistent performance degradation when either component is removed, with the full FCL-DA model achieving the highest explanatory power and lowest error magnitude.

4.3 Discussion

As it has been proven experimentally, the developed FCL-DA framework has considerable strengths in terms of cross-platform teaching quality dynamic assessment tasks. Both the features of its success are based on the design principles: first, the federated learning model provides data privacy on the institutional level, which facilitates cross-platform collaboration. Second, endogenous combination of contrastive learning and dynamic evaluation is a good solution in terms of technical issues of data heterogeneity and dynamism. Contrastive learning forces the model to cry out cross-platform semantic properties that can be useful to pedagogy, and not the thin platform-specific statistical properties that lead to a representation space with enhanced generalisation power. At the same time, the dynamic module of GRU is capable of capturing the changes in the teaching efficacy over the time, which makes the results of the evaluation more timely and even actionable.

Nonetheless, this is also limited by several factors in this study. The first stepped was to simulate the platforms of the experiments and although the differences of the distribution were replicated as one could possibly, the platform still might be too different to reflect the variations seen in real-world commercial education platforms. Second, the communication overhead of federated learning present in this method remains and it might be subject to further investigation. Also, the quality of teaching was reduced to one continuous measurement, but in the real-life situations the same can be a multidimensional and more complicated construct.

5 Conclusions

The present research deals with three fundamental problems in assessment of the quality of teaching in cross-platform educational settings, namely data privacy barriers, data heterogeneity, and assessment staticity. Our hypothesis is a dynamic evaluation system that is federated contrastive learning. Our central finding, achieved through systematic validation using a publicly available dataset that simulates the real-world heterogeneity, is that this system yields a strong improvement in the quality and timeliness of teaching assessment and stricter data privacy across platforms is ensured. Particularly, the experimental outcomes have shown that our offered FCL-DA model minimises the global evaluation MSE to which is a 22.5 percent decrease compared to the conventional federated learning algorithms (FedAvg). Moreover, it also shows a high robustness to large degrees of data distribution variance, and its performance degradation rate is about 15 times less as compared to existing state-of-the-art baseline algorithms. The studies on ablation also demonstrate that both the federated contrastive learning and the dynamic evaluation modules in the framework play an indispensable role in gaining the performance. They act in a synergistic manner where they address the basic problems of cross-platform evaluation together.

Theoretical contributions made by this work are mainly manifested in two aspects. It first, in a successful manner, can be applied to combine contrastive learning, a high-level paradigm in self-supervised learning, with federated learning, a privacy-preserving framework of distributed learning, in order to offer a novel representation-learning-based technical method towards the common issue of non-IID distribution of data in a federation. This methodology is not based on vehement constraints of model parameters. In fact, it essentially improves the extrapolation ability of the world description among diverse customers by informing the model to acquire core semantic equality within the facts. Second, this paper will develop a lightweight dynamic assessment module and will easily fit it into the federated learning process. This provides a paradigm shift of the traditional, lagging evaluation to the dynamic, real-time evaluation which contributes greatly to the practical value of evaluation results used in practice.

According to these results, we make the following recommendations in practical terms: To the operators of education platforms and regulators, we can suggest considering such federated learning frameworks as the base infrastructure of cross-institutional educational data collaboration. This scheme does not only comply with regulations such as the personal information protection law but also extracts the multi-party data value to construct a more profound and scientific network of monitoring education quality. In the case of teachers, the indicated dynamic assessment results,

which this system offers, can effectively act as a feedback mechanism. They can assist teachers to discover the flaws in the teaching process in a timely manner and swiftly respond to the teaching plans and individualised interventions.

Though this research was successful in fulfilling its targeted results, there are a number of limitations that can guide future research studies. To begin with, even though the experimental data was based on real-world dataset and was generated in such a way that it would mimic cross-platform environment, it is in any case not as complex as the commercial-grade platforms. The future working is subject to the validation in more extensive, and diverse union of actual educational platforms. Second, this research gets teaching quality as a composite concept. Further studies may address the possibility of scaling the evaluation system in terms of more detailed and multidimensional measurements including the quality of interaction, quality of resources, and cognitive activity. Lastly, federated learning convergence speed and communication performance are important concerns that influence practical implementation. The optimal way to pilot such improvements in system performance will be to examine strategies such as model compression and asynchronous updates.

Declarations

All authors declare that they have no conflicts of interest.

References

Adekomaya, O. and Majozi, T. (2023) 'Comparative advantages and application of polymer-based materials: prioritizing natural fibres as reinforcement in building structures', *Engineering and Applied Science Research*, Vol. 50, No. 1, pp.47–54.

Al-Rimy, B.A.S., Qasem, S.N., Saeed, F., Tin, T.T., Altamimi, S.N. and Ali, A. (2023) 'Empowering precision medicine: unlocking revolutionary insights through blockchain-enabled federated learning and electronic medical records', *Sensors*, Vol. 23, No. 17, pp.7476–7476, Basel, Switzerland.

Amin, M.S., Ahmad, S. and Loh, W.K. (2025) 'Federated learning for healthcare 5.0: a comprehensive survey, taxonomy, challenges, and solutions', *Soft Computing*, Vol. 29, No. 2, pp.673–700.

Amro, S.A. (2025) 'Securing internet of things devices with federated learning: a privacy-preserving approach for distributed intrusion detection', *Computers, Materials & Continua*, Vol. 83, No. 6, pp.4623–4658.

Belhadi, A., Djenouri, Y., Andrade, F.A.D.A. and Srivastava, G. (2024) 'Federated constrastive learning and visual transformers for personal recommendation', *Cognitive Computation*, Vol. 16, No. 5, pp.2551–2565.

Das, R., Dubey, P.K., Manna, I., Chakraborty, S. and Tiwary, C.S. (2025) 'Properties of highly controlled laser-melted aluminium-nickel eutectic alloys and fundamental insights on the dynamics of nonequilibrium thermal processing', *Journal of Materials Science*, Vol. 60, No. 3, pp.1598–1617.

Gerayllo, S., Vakili, M., Jouybari, L., Moghadam, Z., Jafari, A. and Heidari, A. (2025) 'Using the CIPP model to elicit perceptions of health professions faculty and students about virtual learning', *BMC Medical Education*, Vol. 25, No. 1, pp.1–12.

Hashimoto, Y. (2001) 'How to design an inventory method for ground-level ants in tropical forests', *Nature Human Activ*., Vol. 6, No. 6, pp.25–30.

Huang, Y.T., Vaitheeshwari, R., Chen, M.C., Lin, Y.D., Hwang, R.H., Lin, P.C., Lai, Y.C., Wu, H.K., Chen, C.H. and Liao, Z.J. (2024) 'MITREtrieval: retrieving MITRE techniques from unstructured threat reports by fusion of deep learning and ontology', *IEEE Transactions on Network and Service Management*, Vol. 21, No. 4, p.17.

Ishikawa, M., Borges, R., Mourão, A., Ferreira, L.M., Lobo, A.O. and Martinho, H. (2024) 'Confined water dynamics in the scaffolds of polylactic acid', *ACS Omega*, Vol. 9, No. 18, p.9.

Jain, V., Mohanasundaram, K., Kumar, S., Kavadichanda, C., Varaprasad, R., Mishra, N., Veeraraghavan, B., Swaminathan, S., Lakshminarayanan, D. and Baghel, S. (2024) 'Expert panel consensus statements on vaccination for adults with AIRD in India', *Indian Journal of Rheumatology*, Vol. 19, No. 4, pp.280–295.

Khanipoor, F. and Karimian, Z. (2025) 'Unleashing the power of data: the promising future of learning analytics in medical education: a commentary', *Education and Information Technologies*, Vol. 30, No. 8, pp.10373–10380.

Li, D., Huang, S., Wang, K., Ning, K., Yang, M., Zhang, M. and Xiao, Y. (2024) 'Reform of football training professional courses based on federated learning under the 'internet+' model', *Global Education Bulletin*, Vol. 1, No. 1, pp.11–20.

Madabeni, A., Tanini, D., Capperucci, A. and Orian, L. (2024) 'Untangling the catalytic importance of the se oxidation state in organoselenium-mediated oxygen-transfer reactions: the conversion of aniline to nitrobenzene', *Chemical Science*, Vol. 15, No. 30, p.12.

Mamedova, A. (2024) 'Innovative approaches to the development of primary education in Azerbaijan in the early 20th century', *Bulletin of Postgraduate Education (Series)*, Vol. 29, No. 58, pp.124–138.

Morcos, N.C., Gardin, J.M., Tomita, N. and Henry, W.L. (1992) 'Improvement of relaxation velocity parameters by calcium channel blockers in the aging rabbit myocardium', *Basic Research in Cardiology*, Vol. 87, No. 5, p.437.

Parker, E.N. (1985) 'Equilibrium of magnetic fields with arbitrary interweaving of the lines of force I. discontinuities in the torsion', *Geophysical & Astrophysical Fluid Dynamics*, Vol. 34, Nos. 1–4, pp.243–264.

Senthilkumar, M., Senthil Kumar, C. and Parmar, R.U.V.J. (2024) 'Uncertainty evaluation of non-invasive multi-parameter detector measurements in quality assurance of diagnostic radiology', *Journal of Cancer Research and Therapeutics*, Vol. 20, No. 6, pp.1698–1706.

Xu, Z., Yang, S. and Cui, L. (2024) 'Understanding the heterogeneity and dysfunction of HDL in chronic kidney disease: insights from recent reviews', *BMC Nephrology*, Vol. 25, No. 1, pp.1–19.

Zhang, Y., Yao, J., Zhang, R., Wang, X., Chen, S. and Fu, H. (2024) 'HAVANA: hard negative sample-aware self-supervised contrastive learning for airborne laser scanning point cloud semantic segmentation', *Remote Sensing*, Vol. 16, No. 3, p.18.

Zhou, L., Shen, Y., Liu, Q. and Yang, L. (2025) 'Traditional settlement spatial landscape generation and optimisation using multidimensional GlS data driven method: a case study of Fujian Province', *International Journal of Information and Communication Technology*, Vol. 26, No. 29, pp.75–90.

Zhu, H., Liu, S., Zhao, N., Li, J., Yin, C., Ma, W., Yang, Z. and Gong, H. (2024) 'Keyword extraction method based on complex network', *International Journal of Information and Communication Technology*, Vol. 25, No. 4, pp.323–335.