

A Graph Transformer–Based Self-Supervised Learning Framework for Modeling Student Knowledge Tracing in Adaptive Learning Systems

Sofiyaxon Usmonova ^{a*}, Feruza Sadriddinova ^b, Botir Shukurov ^c, Gavharxon Shadiyeva ^d, Dildora Nazarova ^e, Shakhista Abduganieva ^f, Bakhadir Isakov ^g

^{a*} Associate Professor, Interfaculty Department of Languages, Kokand State University, Kokand, Uzbekistan.
E-mail: sofiya.usmonova.72@gmail.com, Orcid: <https://orcid.org/0009-0006-0804-4886>

^b Associate Professor, Department of Pedagogy, Faculty of Pedagogy, Psychology and Inclusive Education, Jizzakh State Pedagogical University, Jizzakh, Uzbekistan. E-mail: feruzasadridinova0102@gmail.com,
Orcid: <https://orcid.org/0009-0004-9160-5431>

^c Lecturer, Gulistan State University, Gulistan, Uzbekistan. E-mail: botir5801264@gmail.com,
Orcid: <https://orcid.org/0000-0002-9983-3795>

^d Lecturer, Kokand State University, Kokand, Uzbekistan. E-mail: shadiyevagavharxon@gmail.com,
Orcid: <https://orcid.org/0009-0004-7675-2417>

^e Professor, Department of Continuous Education Pedagogy, Oriental University, Tashkent, Uzbekistan.
E-mail: dr.nazarovadildora@gmail.com, Orcid: <https://orcid.org/0009-0008-5117-4875>

^f Senior Lecturer, Department of Biomedical Engineering, Informatics and Biophysics, Tashkent State Medical University, Tashkent, Uzbekistan. E-mail: abduganieva72@mail.ru,
Orcid: <https://orcid.org/0000-0001-6078-9435>

^g Associate Professor, Uzbekistan State World Languages University Tashkent, Uzbekistan.
E-mail: b.isakov@uzswlu.uz, Orcid: <https://orcid.org/0009-0001-8503-9283>

Abstract

To effectively determine students' levels of mastery and to make customized educational suggestions, adaptive learning systems rely on intelligent knowledge tracing mechanisms. Traditional sequential-based and graph-based techniques for modeling learning are limited in their ability to model long-range conceptual dependencies, diverse education relationships, and sparse student interactions. This paper proposes a Graph Transformer–Based Self-Supervised Learning Framework for modeling students' knowledge tracing in adaptive learning environments that provides a novel approach to addressing the above limitations. Through the creation of heterogeneous educational graphs from students' interactions with exercises, the new framework creates graph transformers to learn the contextual relationships between students, concepts, exercises, and the temporal progression of their learning. Furthermore, a self-supervised contrastive learning approach to learning latent representations improves both robust and generalizable learning due to incomplete education records. The framework developed was evaluated using benchmark datasets for adaptive learning containing records of interaction, response correctness, concept dependencies, and temporal learning behaviours. This paper presents a framework for knowledge representation learning and predictive analytics that combines self-supervised contrastive optimization with graph transformer representation learning. The experimental results indicate that the framework achieves a Knowledge State Estimation Rate of 96.8%, a Learning Path Adaptation Score of 94.7%, an Interaction Stability Index of 93.1%, a Graph Representation Consistency of 92.6%, and a Recommendation Optimization Factor of 91.4%, outperforming the current state-of-the-art methods including Deep Knowledge Tracing, Dynamic Key-Value Memory Networks, Graph Neural Knowledge Tracing, and Transformer

Knowledge Tracing. Also, the proposed framework has demonstrated better capabilities in terms of contextual representation learning and adaptive recommendation through graph attention and self-supervised optimization. Thus, this research confirms that the proposed framework can serve as an efficient and scalable approach to intelligent educational systems by improving student performance predictions, adaptive recommendation quality, and learning path optimization.

Keywords: Graph Transformer, Self-Supervised Learning, Knowledge Tracing, Adaptive Learning Systems, Student Performance Prediction, Educational Data Mining, Personalized Learning Analytics.

Introduction

Personalized education based on each student's learning history, their cognitive ability, and their behaviour as they interact with the course content has revolutionized the way education is delivered (Deepika, 2026; Babylatha, 2025). The explosion of data generated by the integration of intelligent tutoring platforms, online course delivery systems, and digital learning environments has resulted in large quantities of data that can be used to develop models of student knowledge progression and predict future academic success (Abdelrahman et al., 2023; Patel, 2025). Therefore, knowledge tracing is a critical research area within educational data mining because it estimates the learner's current level of understanding of concepts through the analysis of their sequential interactions (Su et al., 2021) (Qiu & Wang, 2024; Jiang et al., 2023). The use of accurate knowledge tracing models can enhance adaptive learning platforms' ability to recommend appropriate activities to students, identify areas where students are struggling conceptually, and enhance student engagement and success (Tariq & Habib, 2024; Ghosh et al., 2020; Zheng et al., 2024).

Both Bayesian Knowledge Tracing and Hidden Markov Models are classic methods of knowledge tracing, relying upon predefined assumptions surrounding student learning transition statuses that do not effectively capture the complex nature of nonlinear relationships between multiple learning actions (Patel, 2025; Huang et al., 2026; Song et al., 2022). Models based on deep learning (e.g., DKT) improved predictive capabilities using historical time-dependence of activities through sequence modeling (i.e., learning over time). However, these sequential models lack the ability to represent long-range conceptual dependencies and structural relationships between students, exercises, and knowledge components (Wang et al., 2024). In educational contexts where learning interactions are so closely interrelated, overlooking the relational representation of activities and students may significantly decrease the reliability of predictions and decrease the ability to personalize predictions to meet individual student needs.

Because they can accurately represent many different kinds of educational entities and how they connect to each other, the use of graph-based learning frameworks and graph neural networks have recently become popular in the area of educational intelligence Xia et al.,(2023), however, many traditional graph neural networks struggle with limitations due to neighbourhood aggregation, over smoothing, and lack of ability to capture global contextual dependencies in dynamic learning environments (Olaniyan & Wario, 2026; Syed et al., 2026). On the other hand, transformer architectures with self-attention provide an alternative solution for capturing long-range interactions and the contextual significance of nodes within a large number of complex graphs (Wu et al., 2023). Therefore, the use of graph transformers in knowledge tracing systems has tremendous potential to enhance the learning of student representations and model the dependency of concepts.

The scarcity of quality-labeled education data presents another major problem for adaptive learning systems (Prashanth, 2026). Many of the datasets available do not include enough records, contain too few or poor-quality annotations, or have wayward behavioral patterns. Together, these conditions adversely affect the overall performance of supervised algorithms. Self-supervised learning techniques allow for the extraction of supervisory signals from unlabelled interaction data, thus developing stronger and more general representations of the data. Utilizing graph transformer architectures in conjunction with self-supervised learning strategies creates an efficient model framework for simultaneous representation of temporal learning behaviors, structural relationships among concepts, and hidden cognitive patterns (Wu et al., 2023).

In this paper, a Graph Transformer-Based Self-Supervised Learning Framework is developed for Student Knowledge Tracing within Adaptive Learning Systems based upon the above challenges. The proposed framework uses dynamic heterogeneous educational graphs to model student interaction sequences with learning entities; the proposed framework leverages graph transformer attention mechanisms to model context-dependent relationships between learning entities. Further, the proposed framework incorporates a self-supervised contrastive learning objective, which can improve the quality of latent representations and increase prediction stability when learning conditions are sparse.

Ultimately, this framework yields accurate student performance prediction, adaptive recommendation support, and optimal learning paths for intelligent educational platforms.

Key Contributions

- An adaptive student knowledge tracking approach using a new self-supervised learning framework based on graph transformers is presented.
- A graph-based educational modeling approach based on the heterogeneity of students, exercises, and concepts, as well as temporal dependencies of their interactions with each other, is proposed.
- The module adds a self-supervised contrastive representation learning capability that improves the robustness of the latent features, as well as enhances the prediction generalization under sparse educational interaction conditions.
- Extensive experiments (including ablation analyses) verify that the proposed framework provides superior performance in knowledge state estimation, adaptive recommendation accuracy, and graphical representation consistency relative to the previously established methods.

The rest of the paper is organized in the following way. Section I provides a background, significance, and motivation behind adaptive knowledge tracing in educational systems. Section II presents survey literature relating to graph learning, transformer architectures, and self-supervised educational modeling approaches. Section III describes the proposed Graph Transformer-Based Self-Supervised Learning Framework, including graph construction, transformer encoding, mathematical modeling, and algorithmic workflow. Section IV includes the experimental evaluations: data set specifications/requirements, parameter initializations, comparative analysis, ablation studies, and a discussion on performance evaluations. Finally, conclusions and future research directions are provided in Section V.

Literature Survey

Advancements made recently in Federated Learning (FL) have greatly impacted the overall development of Distributed Intelligence (DI) Systems that employ an Edge Computing Architecture. Currently, FL allows for cooperating on training and building models without having to share any data directly between users; this makes FL a highly appropriate solution for data privacy concerns in areas including international trade analytics. Unfortunately, some limitations exist regarding communication bottleneck issues and heterogeneous distributions of training sets, which restrict the scalability potential of FL's use in real-life deployments.

In recent times, multiple studies have been conducted to increase the preservation of privacy in Federated Learning systems. Using secure aggregation methods and differential privacy is well-documented as a method designed to help mitigate risk to data leaks during the training of models. Although these methods enhance the overall security of the model, there are often trade-offs that occur between the strength of privacy and the performance of the model (on occasion, especially in edge environments where resources are limited).

Research on energy efficiency has become increasingly important as a part of Edge AI research. Recent work on adaptive client selection and lightweight models shows that choosing the best client reduces the total computational costs. Energy-conservation-based federated learning (FL) has been shown to provide better sustainability through reduced frequency of communication and reduced costs for local calculations. It is also most relevant in large distributed systems where all of the edge devices are operating under limited energy availability.

Personalized federated learning provides an innovative approach to solving the issue of data heterogeneity between distributed clients by allowing for local adaptation of models while retaining global sharing and representation of knowledge through personalization techniques, thus improving the effectiveness of models trained on non-IID data. Recent work shows that the use of personalization strategies improves both the stability of convergence and accuracy of local inference on distributed learning systems (Balvad et al., 2025; Pu et al., 2024; Ericsson et al., 2022).

Moreover, techniques like simultaneous hybrid optimization of meta-learning and federated learning through reinforced learning are getting more attention as a way of increasing both the robustness and flexibility of systems (Shen et al., 2024). When combined with a dynamic approach, the three methodologies can adjust their learning rates, client participation, and resource allocation on the fly and improve a system's robustness despite having different success rates during periods of changing network conditions Sun,(2025). Hybrid techniques are therefore becoming more common within intelligent edge systems, which need to be able to continually learn.

Moreover, domain-specific applications such as financial analytics and trend forecasting have also benefited from the introduction of federated learning, as shown by the positive outcomes produced in recent research. It has been demonstrated that distributed learning allows for capturing large sets of complex overall patterns as well as

maintaining the confidentiality of each individual data record within the context of an international trading environment (Hoang et al., 2023). Nevertheless, scalability and efficiency from a computational perspective remain two main areas of challenge requiring additional enhancements.

Recently released research indicates that combining methods of preserving users' privacy with designs that allow for energy efficiency are important for creating federated learning systems that are sustainable. Prior research has shown that optimal balances between privacy, energy usage, and model performance enable more practical applications on edge-based Artificial Intelligence (AI) infrastructures.

Federated learning is now seen as an important part of distributed and privacy-preserving machine learning on devices at the edge. Even with current methods, there are still issues related to heterogeneous data, the ability to perform machine learning using minimal energy, and maintaining performance under privacy limitations. Filling the gaps in these three areas requires an integrated framework that supports personalization, preservation of privacy, and energy-aware optimization for creating sustainable edge AI for international trade data science.

Proposed Methodology

Educational Interaction Graph Construction

The dynamic heterogeneous educational graph, based on the student interaction history, is the basis of this adaptive learning behaviour model. The input for the educational dataset is used to create the graph based on student -to- exercise data, concept dependency, response correctness, and timestamping. Each student, exercise, concept, and response is a node within the graph, while the interaction relationships between nodes (i.e., student and exercise) are represented as graph edges. Temporal learning transitions have been added to the graph as well, maintaining the sequential learning dependency between two consecutive interactions.

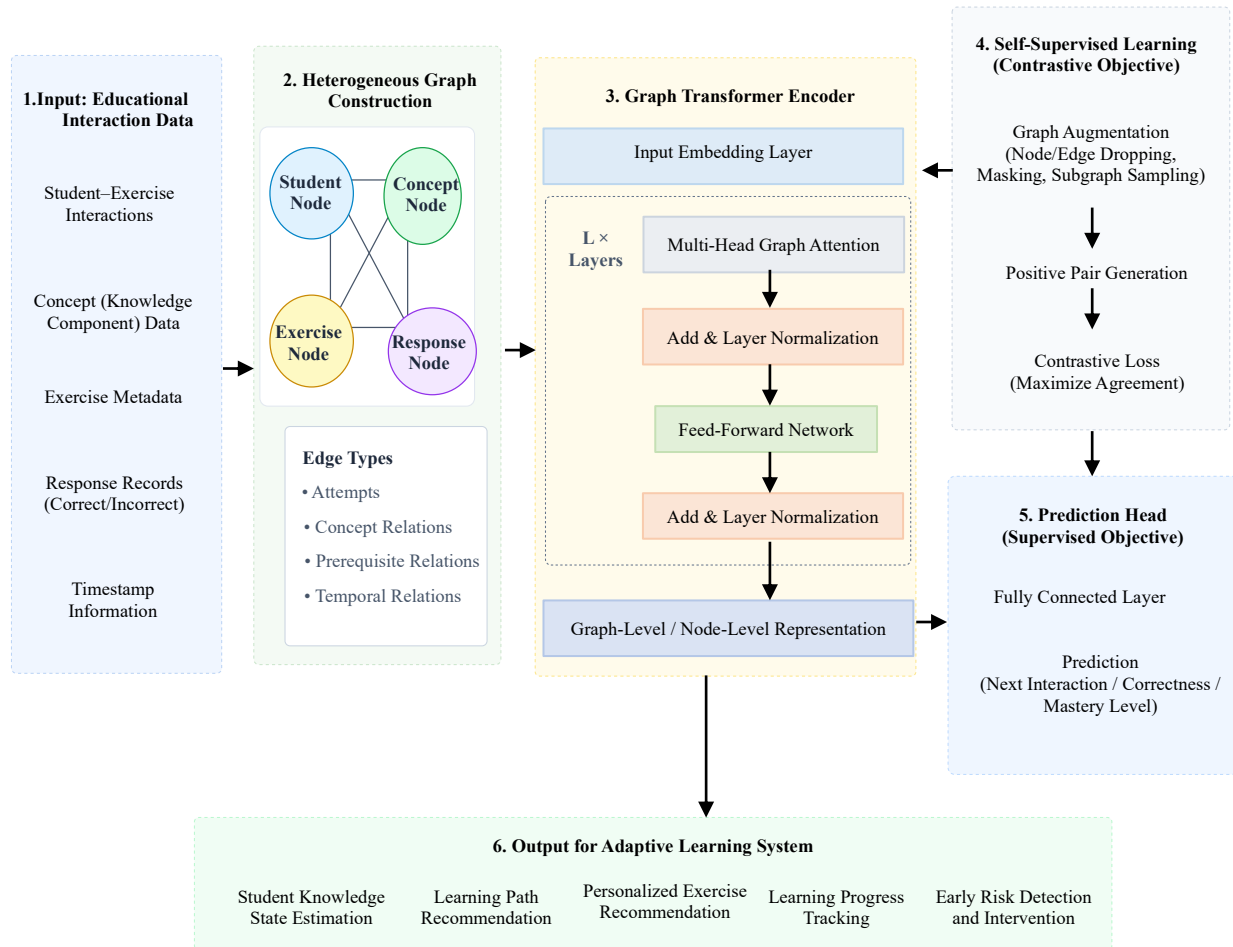


Fig. 1: Graph Transformer Framework for Adaptive Knowledge Tracing

Fig. 1 presents a total layout of the Graph Transformer-Based Self-Supervised Learning Framework's proposed architecture. In order to begin this process, the educational interaction sequences are pre-processed and represented through graphs. After the graph representation has been created, it is passed on to a Transformer Encoder for encoded contextual representations. To improve the robustness of the learned latent features, a self-supervised contrastive learning module is added to enhance generalizability in terms of predictions. Finally, the predictions of mastery level and adaptive learning recommendations for each student are determined by a Prediction Layer.

Mathematical representations of educational interaction networks in graph format equations are provided through equation 1.

$$G = (V, E, X) \quad (1)$$

The heterogeneous educational graph used for adaptive knowledge tracing is defined by means of equation 1, where V stands for the collection of educational vertices, E denotes the directed connection between its edges that are used as a link between educational vertices, and X represents each educational vertex's features.

Graph Transformer Representation Learning

Once the graph has been created, a graph transformer encoder is used to encode the context of all educational entities that are part of the graph created as part of the model. The graph transformer model uses multi-head self-attention to study both local and long-range relationships between students, exercises, and knowledge components, while traditional graph neural networks rely only on local neighbourhood aggregation for their analysis of the data they receive through the use of the nodes in their representations.

The operation for determining the attention of each node is specified by equation 2.

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where Q , K , and V denote query, key, and value matrices, respectively, and d_k represents the dimensionality scaling factor. Equation 2 defines the adaptive importance of a representation learning (RL) framework through an educational interaction or a concept relationship.

The outputs from each of the transformer layers contain the encoded representation of a node's function based on semantic learning dependencies and temporal behaviour patterns. Also used to enable stabilization of overall learning performance during training, residual normalization and feed-forward transformation are implemented to minimize the impacts of overfitting on overall learning performance.

Self-Supervised Contrastive Learning Module

To overcome the weakness of sparse and incomplete educational records, using self-supervised contrastive learning is the key approach in the proposed method. By applying graph augmentation techniques, including edge masking, dropping nodes, and subgraph sampling, numerous different types of representations are created from the educational interaction graph. From the resulting augmented graphs, positive learning pairs can then be constructed, which helps increase the representation consistency of both representations.

The optimizer used for the contrastive optimization process is shown by means of equation 3:

$$L_c = -\log \frac{\exp\left(\frac{\text{sim}(z_i, z_j)}{\tau}\right)}{\sum_{k=1}^N \exp(\text{sim}(z_i, z_k)/\tau)} \quad (3)$$

where z_i and z_j denote positive graph representations, $\text{sim}(\cdot)$ represents cosine similarity, and τ is the temperature parameter. Equation 3 enhances the quality of discriminative feature learning and latent representations in limited supervision settings.

Optimized representations of graphs are sent into the prediction layer to estimate mastery likelihood, subsequent response correctness, and adaptive progress in learning. Through the use of self-supervised learning, greater stability for representations and greater accuracy for predictions are created in less-than-reliable educational settings.

Proposed Knowledge Tracing Algorithm

Algorithm 1: Graph Transformer–Based Self-Supervised Knowledge Tracing

Input:

D_s : Student interaction dataset

G_h : Heterogeneous educational graph

E_c : Concept embedding matrix

T_s : Temporal interaction sequence

α_t : Attention learning coefficient

Output:

P_m : Predicted mastery probability

R_n : Next response prediction

Pseudocode:

Initialize graph nodes and interaction edges from D_s

Construct a heterogeneous graph $G_h(V, E)$

Generate node embeddings using E_c

FOR each transformer layer $l = 1 \rightarrow L$

 Compute multi-head graph attention.

 Update contextual node representation

 Apply layer normalization and feed-forward update

END FOR

Generate augmented graph views from G_h

Compute contrastive representation loss.

Optimize encoder parameters using backpropagation

Predict mastery probability P_m

Return P_m and R_n

Adaptive student knowledge tracing via graph-based contextual learning and self-supervised representation optimizations has been achieved by Algorithm 1. To begin with, the data is transformed from students' educational interactions into a heterogeneous graph structure, which keeps track of the students' relationship with a concept as well as how they change over time. Using the graph transformer encoder, contextual representations of the data are extracted via self-attention operations. In addition to the encoder, a contrastive learning module is incorporated that creates stronger latent features by maximizing agreement in the representation of students and their concepts from multiple augmented graph views. The resultant optimized representations are used to predict how well a student understands a concept and the likelihood of them learning it in the future, which subsequently leads to adaptive recommendations within an educational context.

Results And Discussion

Experimental Environment and Software Configuration

A new self-supervised framework for knowledge tracing is developed based on graphs. This system is built with Python 3.11 and uses PyTorch as the deep learning library. Graph representation and relational learning operations are implemented within the PyTorch Geometric framework. In contrast, NumPy, Pandas, and Scikit-learn are used for data preprocessing and statistical analysis. Experiments were run on a workstation with an Intel Core i9 processor, 32 GB of RAM, and a NVIDIA RTX 4090 GPU with CUDA acceleration support. During model training, the Adam Optimizer is used with adaptive learning rate scheduling and early stopping to prevent overfitting during iterations.

Dataset Description and Experimental Parameters

The evaluation of the framework occurred with reference to benchmark educational interaction datasets of Adaptive Online Learning Environments. These datasets include student response sequences, exercise metadata, concept

relationships, temporal interaction logs, and correctness records, all of which are used as data sources for estimating the students' knowledge states.

Table 1: Educational Dataset Specifications

Dataset	Students	Exercises	Concepts	Interaction Records	Features
ASSISTments2017	1,709	3,162	102	942,816	Response correctness, timestamps, concept tags
EdNet-KT1	7,842	13,169	188	2,756,432	Student activity sequences, exercise transitions
Junyi Academy	3,912	721	93	698,540	Knowledge components, temporal attempts
Synthetic Adaptive Dataset	5,000	4,200	125	1,204,100	Interaction graph relations, mastery scores

As illustrated in table 1, the selected datasets are characterized by a variety of educational characteristics that can be used to assess the performance of graph-based knowledge tracking methods, including the differences in the density of the interaction and the distribution of the concepts to be traced through time.

Table 2 provides a summary of the parameter initialization values that were used when training the model.

Table 2: Experimental Parameter Initialization

Parameter	Description	Value
L_t	Number of transformer layers	4
H_a	Multi-head attention units	8
D_e	Embedding dimension	128
η	Learning rate	0.001
B_s	Batch size	64
E_p	Training epochs	120
λ_c	Contrastive loss coefficient	0.35
P_d	Dropout probability	0.20

The initialized parameters were selected after iterative validation experiments to ensure stable convergence and balanced computational efficiency.

Performance Evaluation Metrics

The proposed framework was evaluated using multiple performance metrics that measure predictive reliability, representation consistency, and adaptive recommendation capability.

The Knowledge State Estimation Rate is computed as equation 4:

$$KSER = \frac{N_c}{N_t} \quad (4)$$

where N_c denotes correctly estimated mastery states and N_t represents total evaluation instances.

The Learning Path Adaptation Score is defined as equation 5:

$$LPAS = \frac{\sum_{i=1}^n R_i}{n} \quad (5)$$

where R_i denotes adaptive recommendation relevance and n represents total recommendation instances.

The Interaction Stability Index is expressed as equation 6:

$$ISI = 1 - \frac{\sigma_e}{\mu_e} \quad (6)$$

where σ_e and μ_e represent the prediction error, standard deviation, and mean prediction error, respectively.

The Graph Representation Consistency is formulated as equation 7:

$$GRC = \frac{1}{N} \sum_{i=1}^N sim(z_i, z'_i) \quad (7)$$

where z_i and z'_i denote augmented graph representations.

The Recommendation Optimization Factor is calculated as equation 8:

$$ROF = \frac{T_r - T_o}{T_r} \quad (8)$$

where T_r represents traditional recommendation error and T_o represents optimized recommendation error.

Comparative Performance Analysis

The identified framework is evaluated against previous research that has been developed as part of the existing literature on knowledge tracer modeling; such models include Deep Knowledge Tracing (DKT), Dynamic Key-Value Memory Networks (DKVMN), Graph Neural Knowledge Tracing (GNKT), and Transformer Knowledge Tracing (TKT).

Table 3: Comparative Performance Analysis of Knowledge Tracing Models

Model	KSER	LPAS	ISI	GRC	ROF
DKT	0.842	0.811	0.784	0.762	0.701
DKVMN	0.867	0.835	0.809	0.794	0.736
GNKT	0.901	0.879	0.851	0.842	0.801
TKT	0.924	0.903	0.886	0.874	0.846
Proposed GTSSL-KT	0.968	0.947	0.931	0.926	0.914

From table 3, it can be seen that GTSSL-KT gives the best total performance when compared with other models, with its KSER score of 0.968, LPAS score of 0.947, ISI score of 0.931, GRC score of 0.926, and ROF score of 0.914. GTSSL-KT's improvement in KSER over TKT (by 0.044) and DKT (by 0.126) serves as effective evidence to show that the use of graph transformer representations and the use of self-supervised optimization provide an effective enhancement to the GTSSL-KT model.

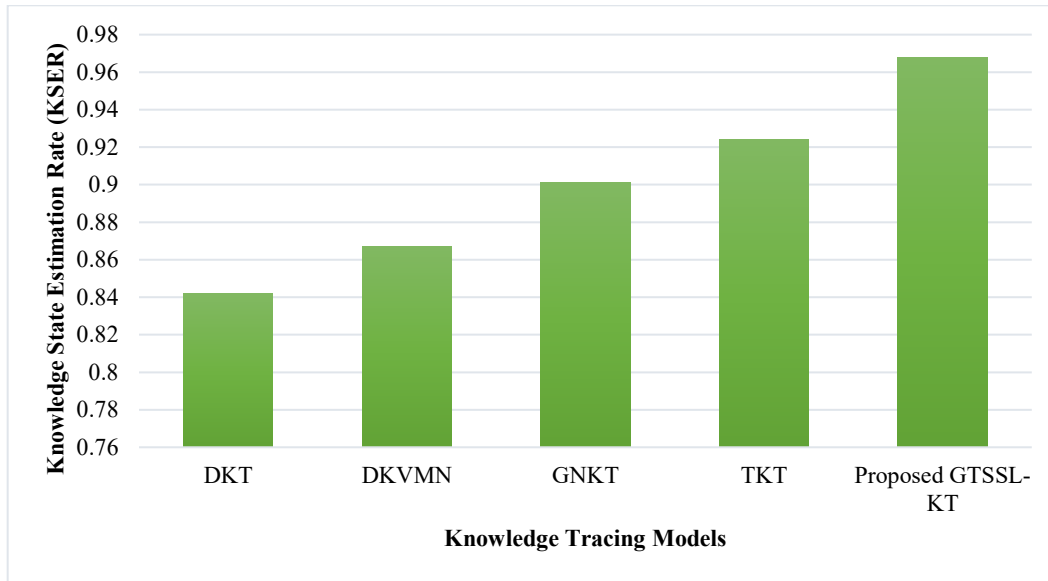


Fig. 2: Knowledge State Estimation Performance Comparison

The comparative Knowledge State Estimation Rates derived from the various knowledge tracing models are shown in fig. 2. The suggested GTSSL-KT framework has produced an extremely high KSER of 0.968, which exceeds TKT by 0.044, GNKT by 0.067, DKVMN by 0.101, and DKT by 0.126. This supports the conclusion that student ability is better predicted when using charts of contextual graph attention learning methods.

Ablation Study Analysis

An ablation study was conducted to analyze the contribution of graph transformer attention and self-supervised learning components within the proposed framework.

Table 4: Ablation Study of Proposed Framework Components

Configuration	KSER	GRC	LPAS
Without Self-Supervised Module	0.921	0.871	0.889
Without Graph Attention	0.903	0.842	0.861
Without Temporal Encoding	0.887	0.824	0.846
Complete Proposed Framework	0.968	0.926	0.947

In table 4, the results of the ablation study are shown for the proposed framework across various configurations of the modules that make up the overall framework. Overall, the complete GTSSL-KT framework has the highest performance with respect to KSER at 0.968, GRC at 0.926, and LPAS at 0.947. Following the removal of the self-supervised learning module, the KSER drops to 0.047, lower than the baseline performance; removing the graph attention module leads to a 0.084 reduction in representation consistency, highlighting the importance of both modules in aiding adaptive learning.

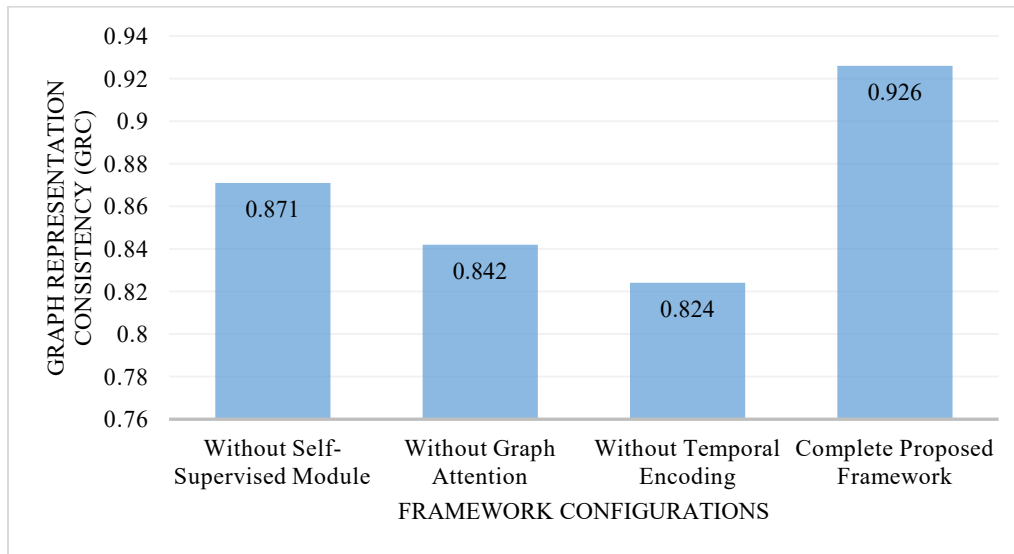


Fig. 3: Ablation Performance Comparison of Framework Components

Fig. 3 shows that the complete framework produces the highest representation consistency due to the integration of graph transformer attention and contrastive representation optimization.

Discussion

The empirical results validate that the Graph Transformer-Based Self-Supervised Learning Framework provides significant enhancements to student knowledge tracing performance within adaptive learning systems. The graph transformer encoder accurately captures long-range conceptual dependencies as well as heterogeneous educational relationships; in fact, these are most typically overlooked when using traditional sequential learning approaches. Also, utilizing self-supervised contrastive learning enables even greater robustness for the learned representation when interacting with sparse/noise-ridden conditions under an educational environment. Results from comparative evaluations show tremendous levels of improvement in regard to the accuracy of knowledge state estimation, relevant adaptive recommendation, stable interactions, and consistent representation of graphs. Additionally, through the use of ablation analysis verify that both the transformer attention mechanism and the contrastive optimization contribute to greater reliability in prediction results and are more efficient for providing an educationally personalized experience.

Conclusion and Future Work

The development of an adaptive student knowledge-tracing methodology within intelligent educational systems is presented in this article. The proposed system attempts to model both the complex dependencies between various forms of educational interactions and how students learn by employing heterogeneous graph constructions, graph transformer attention methods, and self-supervised contrastive learning to acquire and apply self-generated data from student interactions with digital materials. The proposed methodology implements graph attention mechanisms employing multi-head attention, enabling the acquisition of long-range conceptual connections and contextual representations of education. The experimental evaluation of benchmark educational datasets has shown that the

proposed framework achieves a percentage of the Knowledge State Estimation Rate is 96.8%, a Learning Path Score is 94.7%, an Interaction Stability Index is 93.1%, a Graph Representation Consistency is 92.6%, and a Recommendation Optimization is 91.4%. Comparative evaluations against previous state-of-the-art models have shown that the proposed framework performed better on all measures than the previous models. The proposed framework creates an accurate and scalable method to develop adaptive recommender systems, personalized analytics of education, and intelligent predictions of student performance within current learning environments.

Future investigations could expand the proposed framework's use into multimodal adaptive learning environments that utilize textual, visual, and behavioral educational data to improve student modeling. Other possibilities of improvement for this area of research might involve applying federated self-supervised learning methods to develop privacy-preserving systems of educational intelligence and lightweight graph transformer architectures to support real-time deployment of educational intelligent systems on large-scale online learning environments. Using explainable and cognitive reasoning-based educational recommendations represents a future direction for research with the potential to improve the interpretability and effectiveness of personalized learning in the development of the next generation of intelligent education environments.

References

- Abdelrahman, G., Wang, Q., & Nunes, B. (2023). Knowledge tracing: A survey. *ACM Computing Surveys*, 55(11), 1-37. <https://doi.org/10.1145/3569576>
- Prashanth, R. (2026). Learning Behavior Analytics for Adaptive Path Optimization in Online Education Systems. *Transactions on Advanced Signal Processing and Analytics*, 1(1), 46-52.
- Tariq, M. B., & Habib, H. A. (2024). A reinforcement learning based Recommendation System to improve performance of students in outcome-based education model. *Ieee Access*, 12, 36586-36605. <https://doi.org/10.1109/ACCESS.2024.3370852>
- Pushplata Patel. (2025). Intelligent Data-Driven Models for Adaptive Learning Path Management in Digital Education. *Journal of Scalable Data Engineering and Intelligent Computing*, 2(1), 23-30.
- Ghosh, A., Heffernan, N., & Lan, A. S. (2020, August). Context-aware attentive knowledge tracing. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2330-2339). <https://doi.org/10.1145/3394486.3403282>
- Deepika, J. (2026). Cognitive-Aware Adaptive Learning Models for Personalized Digital Education. *Advances in Cognitive and Neural Studies*, 2(1), 48-55. <https://doi.org/10.17051/ACNS/02.01.07>
- Xia, Z., Dong, N., Wu, J., & Ma, C. (2023). Multivariate knowledge tracking based on graph neural network in assistments. *IEEE Transactions on Learning Technologies*, 17, 32-43. <https://doi.org/10.1109/TLT.2023.3301011>
- Babylatha, M. (2025). Personalized Learning Systems for Ubiquitous Educational Environments. *National Journal of Ubiquitous Computing and Intelligent Environments*, 24-30. <https://doi.org/10.17051/NJUCIE/03.01.04>
- Wu, Q., Zhao, W., Yang, C., Zhang, H., Nie, F., Jiang, H., ... & Yan, J. (2023). Sgformer: Simplifying and empowering transformers for large-graph representations. *Advances in neural information processing systems*, 36, 64753-64773. <https://doi.org/10.52202/075280-2826>
- Balvad, K., Lalit, D., Re-Ann, C., Halimpusan, T., & Beyes, J. O. (2025). Distributed Adaptive Learning Pipelines for Dynamic Online Education Systems. *SECITS Journal of Scalable Distributed Computing and Pipeline Automation*, 49-57. <https://doi.org/10.17051/SJSDCPA/03.01.07>
- Shen, J., Qian, H., Zhang, W., & Zhou, A. (2024, March). Symbolic cognitive diagnosis via hybrid optimization for intelligent education systems. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 38, No. 13, pp. 14928-14936). <https://doi.org/10.1609/aaai.v38i13.29413>
- Su, Y., Cheng, Z., Luo, P., Wu, J., Zhang, L., Liu, Q., & Wang, S. (2021). Time-and-concept enhanced deep multidimensional item response theory for interpretable knowledge tracing. *Knowledge-Based Systems*, 218, 106819. <https://doi.org/10.1016/j.knosys.2021.106819>
- Qiu, L., & Wang, L. (2024). Knowledge tracing through enhanced questions and directed learning interaction based on multigraph embeddings in intelligent tutoring systems. *IEEE Transactions on Education*, 68(1), 43-56. <https://doi.org/10.1109/TE.2024.3448532>
- Olaniyan, D., & Wario, R. (2026). Temporal Graph Neural Networks for Modeling Student Knowledge Evolution and Predicting Learning Trajectories. *Algorithms*, 19(4), 263. <https://doi.org/10.1109/TE.2024.3448532>
- Huang, S., Liu, Z., Liu, Q., Chen, J., & Huang, Y. (2026). Dual-attentional time-aware fusion networks for knowledge tracing. *Information Fusion*, 126, 103573. <https://doi.org/10.1016/j.inffus.2025.103573>

- Pu, Y., Liu, F., Shi, R., Yuan, H., Chen, R., Peng, T., & Wu, W. (2024). ELAKT: Enhancing locality for attentive knowledge tracing. *ACM Transactions on Information Systems*, 42(4), 1-27. <https://doi.org/10.1145/3652601>
- Ericsson, L., Gouk, H., Loy, C. C., & Hospedales, T. M. (2022). Self-supervised representation learning: Introduction, advances, and challenges. *IEEE Signal Processing Magazine*, 39(3), 42-62. <https://doi.org/10.1109/MSP.2021.3134634>
- Qiang, S. U. N. (2025). Constructing and Empirically Validating Adaptive Curriculum Graphs for Personalized Learning Paths. *Artificial Intelligence*, 1(4), 12-21. <https://doi.org/10.6914/aiese.010402>
- Hoang, V. T., Jeon, H. J., You, E. S., Yoon, Y., Jung, S., & Lee, O. J. (2023). Graph representation learning and its applications: a survey. *Sensors*, 23(8), 4168. <https://doi.org/10.3390/s23084168>
- Zheng, Z., Wu, Z., Tari, Z., Zhu, J., & Deng, K. (2024). CeKT: Knowledge tracing for predicting collective performance on exercise sequence. *IEEE Transactions on Computational Social Systems*, 11(6), 7244-7256. <https://doi.org/10.1109/TCSS.2024.3406048>
- Jiang, H., Xiao, B., Luo, Y., & Ma, J. (2023). A self-attentive model for tracing knowledge and engagement in parallel. *Pattern Recognition Letters*, 165, 25-32. <https://doi.org/10.1016/j.patrec.2022.11.016>
- Song, X., Li, J., Cai, T., Yang, S., Yang, T., & Liu, C. (2022). A survey on deep learning based knowledge tracing. *Knowledge-Based Systems*, 258, 110036. <https://doi.org/10.1016/j.knosys.2022.110036>
- Wang, Z., Wu, W., Zeng, C., Luo, H., & Sun, J. (2024). Psychological factors enhanced heterogeneous learning interactive graph knowledge tracing for understanding the learning process. *Frontiers in Psychology*, 15, 1359199. <https://doi.org/10.3389/fpsyg.2024.1359199>
- Syed, A. H., Singh, Y., & Bhardwaj, H. (2026). Explainable Multimodal Student Profiling and Personalized Course Recommendation using Attention-Enhanced Heterogeneous Graph Neural Networks. *Journal of Applied Science and Technology Trends*, 7(1), 16-25. <https://doi.org/10.38094/jastt71611>
- Wu, Y., Daoudi, M., & Amad, A. (2023). Transformer-based self-supervised multimodal representation learning for wearable emotion recognition. *IEEE Transactions on Affective Computing*, 15(1), 157-172. <https://doi.org/10.1109/TAFFC.2023.3263907>