Curriculum-Aware Contrastive Learning Bi-Directional Gated Recurrent Unit for Student Learning Course Recommendation and Classification

Joshi Vinay Kumar
Department of Computer Science and Engineering, Dr. M. G. R. Educational and Research Institute, Chennai, Tamil Nadu, India
joshivinay0506@gmail.com (corresponding author)

Dahlia Sam
Department of Information Technology, Dr. M. G. R. Educational and Research Institute, Chennai, Tamil Nadu, India
dahliasam@drmgrdu.ac.in

Received: 28 August 2025 | Revised: 14 October 2025 and 5 November 2025 | Accepted: 6 November 2025

DOI: https://doi.org/10.48084/etasr.14369

ABSTRACT

Student learning course recommendation is a system that recommends the appropriate next course for students based on their interests and performance in previous courses. In a course recommendation system, after a student completes a suggested course, the classification task involves predicting whether the student would recommend the course to others. However, traditional course recommendation models rely on collaborative filtering or static rules, and graph-based attention models frequently ignore curriculum structure and sequential dependencies between courses, which leads to irrelevant or misaligned recommendations that affect student engagement and academic progression and fail to reflect a student's learning path. To overcome these challenges, a Curriculum-Aware Contrastive Learning–Bi-directional Gated Recurrent Unit (CACL-Bi-GRU) is proposed for student learning course recommendation and classification to improve model performance by capturing personalized learning patterns and curriculum dependencies. The proposed CACL-Bi-GRU is utilized to capture both future and past dependencies in sequential course history and to predict the next course based on sequential student enrolment. The proposed model integrates contrastive learning to differentiate semantically relevant courses, whereas the Bi-GRU model learns forward and backward dynamics. The proposed model achieves high values for the Hit Ratio of top-K items (HR@K) at HR @10, Normalized Discounted Cumulative Gain of top-K items (NDCG@K) at NDCg@10, and Mean Reciprocal Rank (MRR) values of 91.85, 65.87, and 56.97, respectively, compared to the Contrastive Learning and Graph convolutional-based Attention Decay network (CLGADN). The proposed model output ensures that associating curriculum-aware context and sequential modeling improves both the relevance and accuracy of course recommendations.

Keywords-Bi-directional Gated Recurrent Unit (Bi-GRU); course history; curriculum-aware contrastive learning; student learning; recommendation

I. INTRODUCTION

The rapid advancement of data and communication technologies has significantly contributed to the emergence and increasing popularity of online education [1]. The process of self-regulation enables students to manage their learning activities and monitor their progress to enhance academic performance during a course [2]. Nevertheless, teachers find it difficult to determine which course resources and activities should be transformed into actionable recommendations to help students enhance their learning and performance [3]. As an extension of traditional education, e-learning represents a progressive model of online instruction that complements classroom-based approaches [4]. It is designed to enhance the learning experience of students who participate in various programs [5] providing a dynamic environment that allows automatic adjustment of content or integration of courseware to meet student needs [6]. However, many educational platforms still rely on static content and assume that all learners are homogeneous, making it difficult to meet the needs of each student [7]. To overcome this limitation, modern e-learning environments are increasingly using learner-generated data to personalize instruction. Online learners generate large volumes of data with complex structures that reflect their learning behaviors, helping identify individual learning patterns [8]. These data are then utilized by educational systems for performance assessment and monitoring [9].

In educational environments, several factors positively influence students' learning processes and improve their academic outcomes [10]. Students learn best when they are actively engaged in the learning process, a concept, often described as the process of information acquisition [11]. In course recommendation systems, course vectors are used in an inner product function to predict scores and generate a final recommendation list [12]. Methods such as collaborative filtering are commonly applied to build these recommendation systems [13]. In traditional recommendation systems, learning activities are suggested based on student characteristics and learning behaviors, often focusing only on the knowledge explicitly covered in courses [14]. For example, students in an Artificial Intelligence (AI) course are expected to develop skills that extend beyond the course content; however, most existing recommendation methods provide recommendations based solely on course content knowledge [15]. To address these limitations, modern e-learning platforms leverage Learning Analytics (LA) frameworks to analyze data from student interactions, providing actionable and personalized course recommendations [16]. In Massive Open Online Course (MOOC) platforms, course recommendation is treated as a sequential problem that aims to propose the most suitable next course to learners based on their history of completed courses and profile information [17].

The related works relevant to this research are reviewed, highlighting their key contributions as well as their limitations. Authors in [18] developed a Contrastive Learning and Graph Convolution-based Attention Decay Network (CLGADN) to improve fairness and account for diverse learner knowledge backgrounds. CLGADN employed contrastive learning to identify differences in learners' prior knowledge, and a monotonic attention decay mechanism was integrated to model the knowledge-forgetting curve. This combination allowed the model to capture the temporal and relational dynamics and to better understand how learners acquire new courses relative to previous knowledge. However, CLGADN relied on the quality of the learner-item graph, which could be noisy or incomplete, affecting the model's ability to learn accurate representations. Authors in [19] implemented EduGraph for MOOC course recommendation systems using learning path hypergraphs. Learners were denoted as hyperedges and courses as vertices. Framelet-based hypergraph convolution with low-pass filters highlighted similarities and enhanced distinct learning paths. EduGraph employed a two-hypergraph learning model for vertex and hyperedge encoding, with collaborative data used to refine learner preference embeddings. Nevertheless, EduGraph required detailed learner course histories to construct meaningful hypergraphs, which led to the cold start problem. Authors in [20] introduced Decision Tree (DT), Gradient Boosting (GB), and Random Forest (RF) models to predict and recommend MOOCs. In the data preprocessing stage, cleaning and scaling adjusted class distributions to prevent poor generalization and model bias. Sequential forward selection was then utilized to select relevant features for target classes, which were subsequently used to train the classifiers. However, these models considered only static features, failing to capture temporal or sequential dependencies in learning behavior.

Authors in [21] developed a heterogeneous graph-enhanced interactive recommendation scheme for online learning using a deep Q-learning neural network. A primary heterogeneous graph was created to represent associations between courses, teachers, and concepts. The course representations were generated using a Graph Attention Network (GAN), and student latent interests were derived from their interactive courses, which were then passed into a double-dueling deep Q-network for interaction-based course recommendations. Nevertheless, recommendations for new users with no interaction history were suboptimal. Authors in [22] implemented a Course Recommendation model based on Learner-Course Relation Prediction (CR-LCRP) with multigranularity data augmentation. Multi-source interactive data on courses, learner–learner comparisons, and course–course similarities were combined to construct a heterogeneous information network. Both explicit and implicit features were extracted, enabling the model to learn high-quality representations of learners and course performance. However, balancing fine- and coarse-grained features was challenging, which sometimes led to overemphasis on one type. Authors in [23] developed log activity-based transformer encoders to predict students' final performance using data generated through a Learning Management System (LMS). Authors in [24] implemented a time-aware transformer with knowledge graphs to focus on temporal user features. A time-aware positional encoding module captured the order and timing of courses in learners' sequences.

The problem statement, objectives, and contributions are discussed next. Traditional course recommendation systems rely on collaborative filtering or static rules, which fail to consider the context and sequence of previously selected courses. These methods suggest courses based on historical enrolment patterns or similarities between students, often resulting in recommendations that are misaligned with students' academic progression. Such systems also ignore contextual factors, including learning paths, areas of interest, and past performance, which are essential for providing personalized course suggestions. The objective of this research is to develop a Curriculum-Aware Contrastive Learning integrated with a Bi-directional Gated Recurrent Unit (CACL-Bi-GRU) for efficient student learning course recommendation and classification. The Bi-GRU captures sequential and bi-directional dependencies in students' learning histories whereas CACL highlights the curricular relations and semantic similarities among courses, ensuring that the recommendations are aligned with learning outcomes. By integrating sequence modeling with curriculum-aware context, the proposed method enhances both student academic progression and recommendation accuracy.

The main contributions of this research are as follows:

- This study proposes CACL-Bi-GRU for student learning course recommendation and classification, improving system performance and supporting student skill development.

- The proposed model ensures curriculum consistency, reinforces academic performance, and enhances learners' skill acquisition.
- Preprocessing techniques such as stemming, tokenization, and stop-word removal improve text consistency, allowing the model to focus on meaningful terms for more accurate classification.

II. PROPOSED METHODOLOGY

This research proposes CACL-Bi-GRU to classify and recommend learning courses for students, primarily using MOOC and Coursera datasets to validate model performance. By capturing sequential learning patterns and curriculum dependencies, the proposed CACL-Bi-GRU offers personalized and context-aware course recommendations, minimizing misalignment or inappropriate suggestions. Preprocessing techniques such as tokenization, stop-word removal, stemming, and data cleaning are used to eliminate redundant text and improve contextual understanding. The preprocessed text is then used in the feature extraction stage, after which the proposed model performs classification and course recommendation. Figure 1 illustrates the proposed methodology.

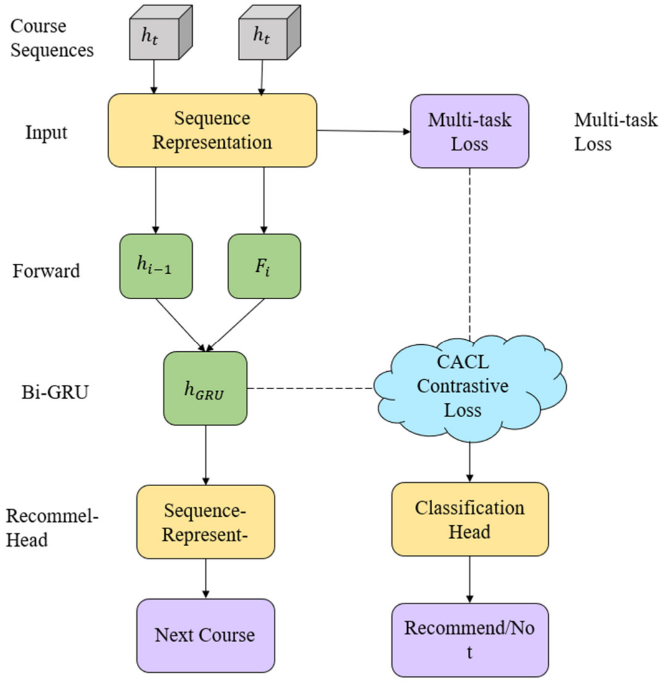

Fig. 1. Architecture of the proposed CACL-Bi-GRU model.

A. *Dataset Acquisition*

In this research, MOOC and Coursera datasets are used for student learning course recommendation and classification, described below:

- MOOC dataset [25]: This dataset, collected from XuetangX, one of the largest MOOC platforms in China, includes 82,535 users and 1,302 courses, totaling 458,454 user-course interactions. An individual user is registered for a minimum of three courses, with an average course enrolment of 5.55 per user.
- Coursera dataset [26]: This dataset consists of 890 courses and six columns, summarized in Table I.

TABLE I. COURSERA DATASET DESCRIPTION

Column	Description
Course name	The title of the course
Organization	The institution offering the course
Certificate	Types of certificates available for the course
Rating	Ratings assigned to each course
Course enrolled	Number of students registered or enrolled in the course
Difficulty level	The course level

B. *Preprocessing*

The preprocessing stage applies tokenization, stop-word removal, and stemming, described below:

- Tokenization: Each sentence in the text is split into small words or chunks, such as symbols, keywords, or phrases.
- Stop-word removal: Stop words are removed to reduce text size and eliminate terms that do not convey meaningful information.
- Stemming: Stemming converts words to their root forms to group similar words together, improving processing speed and consistency. For example, the words fast, faster, and fastest are all put into the same category.

C. *Feature Extraction*

Word2Vec generates dense vector representations of words based on their surrounding context, allowing the model to classify sentiment even when distant vocabulary is employed to express similar meaning. Moreover, it captures semantic relationships among words, producing a feature set with emotional depth and contextual information from student reviews. Word2Vec has two types of architecture to generate similar words: skip-gram and Continuous Bag of Words (CBOW). The CBOW model predicts a target word given its context, whereas the skip-gram predicts surrounding words from a target word. In this study, the skip-gram model receives an input word and identifies other words with similar meanings. The window size defines the maximum distance between the target word and context words. For example, a window size of five represents a model that looks at five words before and after the target word. Moreover, each word is indicated as a 300-dimensional dense vector, balancing model performance and computational cost. This representation is well-suited for generating vector representations of a large number of words in unstructured text.

D. *Classification and Recommendation*

The GRU branch is utilized to enrich the temporal representation and improve the performance of course classification and recommendation. A Bi-GRU processes data in both forward and backward directions, capturing sequential dependencies more effectively than a unidirectional GRU. By exploiting information from the entire course sequence, the Bi-GRU accurately classified the courses. The GRU includes the input x_t, the prior hidden state h_{t-1}, the update gate z_t^i, the reset gate r_t and the new hidden state h_t. Detailed explanations of these states are provided below.

1) Update Gate

An update gate z_t^i is utilized to determine the balance between retaining past data and integrating new data. The update gate z_t^i is formulated as:

$$z_t^i = \sigma(W_z^i x_t + U_z^i h_{t-1} + b_z^i) \tag{1}$$

where W_z^i denotes the weight matrix, b_z^i denotes the bias, $\sigma(x) = 1/(1 + exp(-x))$ represents the sigmoid function, which maps values to the range [0, 1].

2) Candidate Gate

The candidate hidden state $\hat{h}_t^i$ integrates the current input and a transformed version of the previous hidden state through the reset gate r_t^i. It is computed as:

$$\hat{h}_t^i = \tanh\left(W_h^i x_t + U_h^i\left(r_t^i \odot h_{t-1}\right) + b_h^i\right) \tag{2}$$

where $\odot$ denotes the element-wise multiplication and b_h^i denotes the bias.

3) Reset Gate

The reset gate r_t^i controls the extent to which the previous hidden state is forgotten:

$$r_t^i = \sigma(W_r^i x_t + U_r^i h_{t-1} + b_r^i) \tag{3}$$

where W_r^i, U_r^i, and b_r^i are the weight matrices and bias for the reset gate.

4) Forward and Backward Gates

A conventional GRU predicts the next state based on prior sequential information. In a Bi-GRU, input x_t passes through both forward and backward hidden layers, capturing dependencies from the entire sequence. The forward layer processes the current and previous time steps, whereas the backward layer processes the current and next time steps.

At any time t, the input to the hidden layer has two sources: (1) $\overrightarrow{h}_{t-1}^{(i)}$, the transformed hidden state from the previous time step, and (2) $h_t^{(t-1)} = \left[\overrightarrow{h}_t^{(i-1)}; \overleftarrow{h}_t^{(i-1)}\right]$, the concatenation of forward and backward hidden states from the previous layer. The forward and backward hidden states are computed as:

$$\overrightarrow{h}_t^{(i)} = f\left(\overrightarrow{W}^{(i)} h_{t-1} + \overrightarrow{V}^{(i)} \overrightarrow{h}_{t-1}^{(i)} + \overrightarrow{b}^{(i)}\right) \tag{4}$$

$$\overleftarrow{h}_t^{(i)} = f\left(\overleftarrow{W}^{(i)} h_t^{(i-1)} + \overleftarrow{V}^{(i)} \overleftarrow{h}_{t+1}^{(i)} + \overleftarrow{b}^{(i)}\right) \tag{5}$$

where f is the activation function, $\overrightarrow{W}^{(i)}$, $\overrightarrow{V}^{(i)}$, $\overrightarrow{b}^{(i)}$ are forward layer weights and bias, and $\overleftarrow{W}^{(i)}$, $\overleftarrow{V}^{(i)}$, $\overleftarrow{b}^{(i)}$ are backward layer weights and bias. Moreover, $\overrightarrow{h}_t^{(i)}$ and $\overleftarrow{h}_t^{(i)}$ represent the forward and backward hidden states, respectively. The bi-directional structure captures both forward and backward dependencies in student learning sequences, significantly improving the network's processing ability to process nonlinear data. Hence, the Bi-GRU structure is essential for accurate course recommendation.

E. Curriculum-Aware Contrast Learning

Although the Bi-GRU model captures sequential context, it does not inherently account for whether a learning path follows a valid curriculum, such as the completion of prerequisites. To address this limitation, CACL is introduced after Bi-GRU processing. GRU is a type of Recurrent Neural Network (RNN) primarily used for the classification of sequential data, including human emotions and student learning behaviors. Compared to traditional deep learning models such as RNN and Long Short-Term Memory (LSTM), GRU offers computational efficiency and simplicity while maintaining strong temporal modeling capabilities. The Bi-GRU improves this ability by processing sequences in both forward and backwards directions, which is particularly effective for course recommendation, where both past and future contexts matter. Understanding the learning context, such as course prerequisites and the order of completed courses, is essential in this domain. Bi-GRU enables this by learning bi-directional dependencies from the course sequences. It captures how earlier courses influence later courses and vice versa, making it suitable for modeling curriculum-driven student behaviors, which are expressed in (6) to (8):

$$\begin{cases} h_{s_1} = \mathrm{BiGRU}(S_1), \\ h_{s_2} = \mathrm{BiGRU}(S_2), \\ h_{s_3} = \mathrm{BiGRU}(S_3) \end{cases} \tag{6}$$

$$L_{CACL} = \\ -\log\left(\frac{\exp\left(\mathrm{sim}(h_{s_1}, h_{s_2})/T\right)}{\exp\left(\mathrm{sim}(h_{s_1}, h_{s_2})/T\right) + \exp\left(\mathrm{sim}(h_{s_1}, h_{s_3})/T\right)}\right) \tag{7}$$

$$L_{total} = L_{rec} + L_{cls} + \lambda L_{CACL} \tag{8}$$

F. Integration of Contrastive Learning and Bi-GRU

In the proposed CACL-Bi-GRU model, the Bi-GRU first processes the sequential course enrollment data to produce hidden representations that capture both forward and backward dependencies in a student's learning history. These representations are then used in contrastive learning to distinguish between similar and dissimilar course sequences.

Contrastive learning measures the similarity between Bi-GRU embeddings of positive and negative course pairs. The contrastive loss L_{CACL} enables the model to minimize the distance between positive pairs and maximize the distance from negative pairs:

$$L_{CACL} = -\log\frac{\exp\left(\mathrm{sim}(h_i, h_j^+)/T\right)}{\sum_{k=1}^{N}\exp(\mathrm{sim}(h_i, h_k)/T)} \tag{9}$$

where h_i and h_j^+ are Bi-GRU embeddings of positive pairs, sim denotes cosine similarity, and N is number of courses in the batch.

Finally, the total loss integrates contrastive loss with the recommendation and classification loss L_{rec} using a weighted sum:

$$L_{total} = \alpha L_{rec} + (1 - \alpha) L_{CACL} \tag{10}$$

This integration ensures that the Bi-GRU learns representations that are both temporally coherent and semantically discriminative, enhancing the relevance and accuracy of course recommendations.

III. RESULTS AND DISCUSSION

The proposed method is simulated using a Python environment with 64 GB RAM, an Intel i5 processor, and Windows 10 operating system. The model is applied to classify and recommend courses using the MOOC and Coursera datasets. Performance is evaluated using metrics such as Mean Reciprocal Rank (MRR), Hit Ratio of top-K items (HR@K), and Normalized Discounted Cumulative Gain of top-K items (NDCG@K):

$$\mathrm{MRR} = \frac{1}{|U_{test}|}\sum_{u\in U_{test}}\frac{1}{\mathrm{rank}_u} \quad (11)$$

$$\mathrm{HR@K} = \frac{1}{|U_{test}|}\sum_{u\in U_{test}}\mathrm{hits@K} \quad (12)$$

$$\mathrm{NDCG@K} = \frac{r(1)+\sum_{i=2}^{\mathrm{K}}\frac{r(i)}{log_2 i}}{\sum_{i=1}^{|\mathrm{REL}|}\frac{r(i)}{log_2(i+1)}} \quad (13)$$

where $r(i)$ denotes the relevance score of the course at position i in the ranking. In this study, REL represents the binary relevance between predicted and actual course recommendations: $\mathrm{REL} = 1$ if a student actually enrolled in the recommended course, and $\mathrm{REL} = 0$ otherwise. HR@K measures the proportion of recommended courses that match a student's actual enrollments within the top-K recommendations. NDCG@K evaluates how well the model ranks truly relevant courses among the top-K, normalizing the ranking to account for position.

A. Performance Analysis

Tables II-IV present the quantitative evaluation of the proposed model using the MOOC and Coursera datasets. Performance metrics include HR@k, NDCG@k, and MRR, as well as training time, inference time, memory consumption, and t-test results.

TABLE II. PERFORMANCE ANALYSIS OF VARIOUS FEATURE EXTRACTION MODELS

Dataset	Method	HR@k	NDCG@k	MRR
MOOC dataset	TF-IDF	81.43	57.21	49.63
	GloVe	85.67	60.45	52.74
	BoW	87.38	62.19	54.18
	Word2Vec	91.85	65.87	56.97
Coursera dataset	TF-IDF	79.14	55.38	47.21
	GloVe	83.26	58.64	50.13
	BoW	85.72	60.18	52.07
	Word2Vec	80.23	71.63	61.68

Table II shows the performance of different feature extraction methods, including Term Frequency–Inverse Document Frequency (TF-IDF), GloVe, Bag of Words (BoW), and Word2Vec. Word2Vec achieves the best results for both the MOOC and Coursera datasets (e.g. HR@K = 91.85 for MOOC). This result is attributed to Word2Vec's ability to learn semantic word embeddings that capture contextual meaning, improving recommendation accuracy compared to frequency-based methods such as TF-IDF and BoW. Overall, Word2Vec outperforms other feature extraction methods by focusing on semantic relationships among words, which enhances contextual understanding.

TABLE III. PERFORMANCE ANALYSIS OF VARIOUS CLASSIFICATION MODELS

Dataset	Method	HR@k	NDCG@k	MRR
MOOC dataset	GRU	83.72	59.04	51.26
	LSTM	85.91	60.82	53.14
	Bi-GRU	87.89	62.71	54.19
	Proposed	91.85	65.87	56.97
Coursera dataset	GRU	74.15	66.32	56.48
	LSTM	76.12	67.81	58.14
	Bi-GRU	78.25	68.37	59.78
	Proposed	80.23	71.63	61.68

Table III compares classification models, including GRU, LSTM, Bi-GRU, and the proposed CACL-Bi-GRU. The proposed method achieves the highest HR@K, NDCG@K, and MRR values for both datasets. By combining Word2Vec embeddings with CACL-Bi-GRU, the model efficiently captures both temporal and contextual relationships among courses, improving ranking accuracy and effectively learning long-term dependencies.

TABLE IV. PERFORMANCE ANALYSIS OF TRAINING TIME, INFERENCE TIME, MEMORY USAGE AND T-TEST

Dataset	Method	Training time (s)	Inference time (s)	Memory usage (KB)	t-test
MOOC dataset	GRU	158.24	9.52	3214	0.042
	LSTM	176.85	10.67	3540	0.038
	Bi-GRU	169.12	9.93	3381	0.031
	Proposed	142.67	8.34	2978	0.04
Coursera dataset	GRU	194.30	10.85	3657	0.046
	LSTM	208.91	12.16	3922	0.041
	Bi-GRU	199.42	11.02	3738	0.034
	Proposed	173.56	9.61	3326	0.05

Table IV presents comparisons of training time, inference time, memory consumption, and t-test results. The proposed CACL-Bi-GRU achieves reduced training time, lower memory usage, and faster inference compared to GRU, LSTM, and Bi-GRU models.

The CACL-Bi-GRU achieves lower training time because the CACL model focuses only on informative contextual patterns, reducing redundant computations. Selective learning accelerates convergence, improves efficiency, and minimizes resource usage without increasing model complexity. Consequently, the proposed model is both faster and more resource-efficient than conventional LSTM and GRU methods.

B. Comparative Analysis

The comparative analysis includes existing methods such as CLGADN [18] and CR-LCRP [22], evaluated against the proposed model using the MOOC dataset. Table V presents the comparative results. The proposed model outperforms the existing approaches across all evaluation metrics. It achieves HR@10 = 91.85, NDCG@10 = 65.87, and MRR = 56.97, showing significant improvement over CLGADN and CR-LCRP.

TABLE V. COMPARATIVE ANALYSIS OF PROPOSED AND EXISTING METHODS ON THE MOOC DATASET

Method	Dataset	HR@10	NDCG@10	MRR
CLGADN [18]	MOOC	74.93	48.35	41.45
CR-LCRP [22]		72.25	49.86	NA
Proposed		91.85	65.87	56.97

C. *Discussion*

The advantages of the proposed CACL-Bi-GRU model and the limitations of existing approaches are discussed in this section. Existing approaches show some limitations. For example, CLGADN [18] depends on the quality of the learner–item graph; if this graph is noisy or incomplete, the resulting representations become inaccurate. EduGraph [19] requires detailed learner–course histories to build meaningful hypergraphs, making it difficult to solve the cold start problem for new users or courses. Similarly, heterogeneous graph models [22] rely on historical data, leading to suboptimal recommendations for new users with no interaction history. To address these limitations, this study proposes the CACL-Bi-GRU model for student course recommendation and classification. The model enhances system performance by integrating CACL with a Bi-GRU, enabling it to capture both temporal dependencies and curriculum consistency in course sequences. The proposed method achieves HR@10 = 91.85, NDCG@10 = 65.87, and MRR = 56.97 on the MOOC dataset, outperforming existing approaches.

IV. CONCLUSION

This research introduced the Curriculum-Aware Contrastive Learning Bi-Directional Gated Recurrent Unit (CACL-Bi-GRU) model to address the limitations of traditional course recommendation systems that ignore course order and context. In the proposed course recommendation framework, CACL-Bi-GRU provides personalized course suggestions by analyzing students' past learning courses and predicting the next relevant courses. Unlike collaborative or static approaches, the proposed model focuses on sequential dependencies in students' course histories and learning behaviors, supporting both classification (recommend or not recommend) and next-course prediction tasks. The CACL component enhances discrimination by aligning contextually relevant course sequences while separating inappropriate ones, which improves the recommendation accuracy. During preprocessing, tokenization, stop-word removal, and stemming are applied to ensure meaningful text inputs that enhance model performance. Experimental evaluation demonstrates that the proposed method achieves high Hit Ratio at top-10 (HR@10), Normalized Discounted Cumulative Gain at top-10 (NDCG@10), and Mean Reciprocal Rank (MRR) values of 91.85, 65.87, and 56.97, respectively, on the Massive Open Online Course (MOOC) dataset. In the future, incorporating additional side information such as student grades, course transparency indicators, and curriculum structures could further improve adaptability and effectiveness in continuous learning environments.

DATA AVAILABILITY STATEMENT

The datasets analyzed in the current study are publicly available. The MOOC dataset can be accessed at https://data.mendeley.com/datasets/v398vj34h6/1, and the Coursera dataset is available at https://www.kaggle.com/datasets/siddharthm1698/coursera-course-dataset.

REFERENCES

[1] M. Afzaal, A. Zia, J. Nouri, and U. Fors, "Informative Feedback and Explainable AI-Based Recommendations to Support Students' Self-regulation," *Technology, Knowledge and Learning*, vol. 29, no. 1, pp. 331–354, Mar. 2024, https://doi.org/10.1007/s10758-023-09650-0.

[2] S. Bhaskaran and R. Marappan, "Design and analysis of an efficient machine learning based hybrid recommendation system with enhanced density-based spatial clustering for digital e-learning applications," *Complex & Intelligent Systems*, vol. 9, no. 4, pp. 3517–3533, Aug. 2023, https://doi.org/10.1007/s40747-021-00509-4.

[3] J.-W. Tzeng, N.-F. Huang, A.-C. Chuang, T.-W. Huang, and H.-Y. Chang, "Massive open online course recommendation system based on a reinforcement learning algorithm," *Neural Computing and Applications*, vol. 37, no. 18, pp. 11607–11618, June 2025, https://doi.org/10.1007/s00521-023-08686-8.

[4] W. Deng, P. Zhu, H. Chen, T. Yuan, and J. Wu, "Knowledge-aware sequence modelling with deep learning for online course recommendation," *Information Processing & Management*, vol. 60, no. 4, July 2023, Art. no. 103377, https://doi.org/10.1016/j.ipm.2023.103377.

[5] H. Zhang, X. Shen, B. Yi, W. Wang, and Y. Feng, "KGAN: Knowledge Grouping Aggregation Network for course recommendation in MOOCs," *Expert Systems with Applications*, vol. 211, Jan. 2023, Art. no. 118344, https://doi.org/10.1016/j.eswa.2022.118344.

[6] J. Chinnadurai *et al.*, "Enhancing online education recommendations through clustering-driven deep learning," *Biomedical Signal Processing and Control*, vol. 97, Nov. 2024, Art. no. 106669, https://doi.org/10.1016/j.bspc.2024.106669.

[7] M. Praseptiawan, N. M. Putri, M. F. Damar Muchtarom, M. H. Zakaria, and A. N. Che Pee, "Application of Collaborative Filtering and Explainable AI Methods in Recommendation System Modeling to Predict MOOC Course Preferences," in *2024 2nd International Symposium on Information Technology and Digital Innovation*, Bukittinggi, Indonesia, 2024, pp. 228–233, https://doi.org/10.1109/ISITDI62380.2024.10797073.

[8] C. Troussas, F. Giannakas, C. Sgouropoulou, and I. Voyiatzis, "Collaborative activities recommendation based on students' collaborative learning styles using ANN and WSM," *Interactive Learning Environments*, vol. 31, no. 1, pp. 54–67, Jan. 2023, https://doi.org/10.1080/10494820.2020.1761835.

[9] A. B. Rashid, R. R. R. Ikram, Y. Thamilarasan, L. Salahuddin, N. F. A. Yusof, and Z. B. Rashid, "A Student Learning Style Auto-Detection Model in a Learning Management System," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 11000–11005, June 2023, https://doi.org/10.48084/etasr.5751.

[10] T. Nguyen *et al.*, "H-BERT4Rec: Enhancing Sequential Recommendation System on MOOCs Based on Heterogeneous Information Networks," *IEEE Access*, vol. 12, pp. 155789–155803, 2024, https://doi.org/10.1109/ACCESS.2024.3462830.

[11] X. Xia, "Learning behavior mining and decision recommendation based on association rules in interactive learning environment," *Interactive Learning Environments*, vol. 31, no. 2, pp. 593–608, Feb. 2023, https://doi.org/10.1080/10494820.2020.1799028.

[12] N. S. M. Yusop, M. M. Rosli, N. F. Farid, N. A. S. M. Nazri, N. Jamil, and M. I. Ramli, "Development of a Collaborative Intelligent Individual Education Program System using a Prototyping Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14666–14676, June 2024, https://doi.org/10.48084/etasr.7352.

[13] P. Hao, Y. Li, and C. Bai, "Meta-relationship for course recommendation in MOOCs," *Multimedia Systems*, vol. 29, no. 1, pp. 235–246, Feb. 2023, https://doi.org/10.1007/s00530-022-00989-5.

[14] X. Wang, L. Jia, L. Guo, and F. Liu, "Multi-aspect heterogeneous information network for MOOC knowledge concept recommendation," *Applied Intelligence*, vol. 53, no. 10, pp. 11951–11965, May 2023, https://doi.org/10.1007/s10489-022-04025-x.

[15] C. C. Y. Yang and H. Ogata, "Personalized learning analytics intervention approach for enhancing student learning achievement and behavioral engagement in blended learning," *Education and Information Technologies*, vol. 28, no. 3, pp. 2509–2528, Mar. 2023, https://doi.org/10.1007/s10639-022-11291-2.

[16] N. S. Raj and V. G. Renumol, "An improved adaptive learning path recommendation model driven by real-time learning analytics," *Journal of Computers in Education*, vol. 11, no. 1, pp. 121–148, Mar. 2024, https://doi.org/10.1007/s40692-022-00250-y.

[17] S. Hussain and M. Q. Khan, "Student-Performulator: Predicting Students' Academic Performance at Secondary and Intermediate Level Using Machine Learning," *Annals of Data Science*, vol. 10, no. 3, pp. 637–655, June 2023, https://doi.org/10.1007/s40745-021-00341-0.

[18] W. Ma, W. Chen, L. Lu, and X. Fan, "Integrating learners' knowledge background to improve course recommendation fairness: A multi-graph recommendation method based on contrastive learning," *Information Processing & Management*, vol. 61, no. 4, July 2024, Art. no. 103750, https://doi.org/10.1016/j.ipm.2024.103750.

[19] M. Li, Z. Li, C. Huang, Y. Jiang, and X. Wu, "EduGraph: Learning Path-Based Hypergraph Neural Networks for MOOC Course Recommendation," *IEEE Transactions on Big Data*, vol. 10, no. 6, pp. 706–719, Dec. 2024, https://doi.org/10.1109/TBDATA.2024.3453757.

[20] X. Tang, H. Zhang, N. Zhang, H. Yan, F. Tang, and W. Zhang, "Dropout Rate Prediction of Massive Open Online Courses Based on Convolutional Neural Networks and Long Short-Term Memory Network," *Mobile Information Systems*, vol. 2022, no. 1, May 2022, Art. no. 8255965, https://doi.org/10.1155/2022/8255965.

[21] C. Jin, "MOOC student dropout prediction model based on learning behavior features and parameter optimization," *Interactive Learning Environments*, vol. 31, no. 2, pp. 714–732, Feb. 2023, https://doi.org/10.1080/10494820.2020.1802300.

[22] X. Yu, Q. Mao, X. Wang, Q. Yin, X. Che, and X. Zheng, "CR-LCRP: Course recommendation based on Learner–Course Relation Prediction with data augmentation in a heterogeneous view," *Expert Systems with Applications*, vol. 249, no. C, Sept. 2024, Art. no. 123777, https://doi.org/10.1016/j.eswa.2024.123777.

[23] S. S. Kusumawardani and S. A. I. Alfarozi, "Transformer Encoder Model for Sequential Prediction of Student Performance Based on Their Log Activities," *IEEE Access*, vol. 11, pp. 18960–18971, 2023, https://doi.org/10.1109/ACCESS.2023.3246122.

[24] J. Zhou, G. Jiang, W. Du, and C. Han, "Profiling temporal learning interests with time-aware transformers and knowledge graph for online course recommendation," *Electronic Commerce Research*, vol. 23, no. 4, pp. 2357–2377, Dec. 2023, https://doi.org/10.1007/s10660-022-09541-z.

[25] L. Chaw, "Dataset related to MOOCs." Mendeley Data, 2022, https://doi.org/10.17632/v398vj34h6.1.

[26] S. Amin, M. I. Uddin, W. K. Mashwani, A. A. Alarood, A. Alzahrani, and A. O. Alzahrani, "Developing a Personalized E-Learning and MOOC Recommender System in IoT-Enabled Smart Education," *IEEE Access*, vol. 11, pp. 136437–136455, 2023, https://doi.org/10.1109/ACCESS.2023.3336676.

AUTHORS PROFILE

Joshi Vinay Kumar is a Research Scholar currently pursuing his Ph.D. at Dr. M.G.R Educational and Research Institute, Chennai. His research interests include machine learning, knowledge graphs, information retrieval systems, and e-learning platforms. With 8 years of teaching experience across various institutions, he is committed to academic and research excellence.

Dr. Dahlia Sam, Professor at Dr. M.G.R Educational and Research Institute, Chennai, is an accomplished academician with a Ph.D. in Computer Science and Engineering (2017). With over two decades of experience, she has excelled as an educator, researcher, and freelance web developer. Her contributions include serving as a reviewer for esteemed journals such as Measurement: Sensors (Elsevier) and Transactions on Intelligent Transportation Systems, and editing prominent publications like the Journal of Network and Computer Applications. In addition to her academic achievements, she is part of the university's website management team and has a proven track record in fostering innovation and excellence in both academia and industry.