Research on Enhancing Deep Learning Model Performance on MOOCs Using Data Enrichment with Graph Neural Networks

Kha Nguyen[1,3], Phuong Nguyen[1,3], Thu Nguyen[1,3], Khoa Tan VO[1,3], Phuc Nguyen[2,3], Hong-Tri Nguyen[4]
Tu-Anh Nguyen-Hoang[1,3]
[1]Faculty of Information Science and Engineering, University of Information Technology, Ho Chi Minh City, Vietnam
[2]Faculty of Information Systems, University of Economics and Law, Ho Chi Minh City, Vietnam
[3]Vietnam National University, Ho Chi Minh City, Vietnam
[4]Aalto University, Finland
{21520949,21521308}@gm.uit.edu.vn, {thunta,khoavt}@uit.edu.vn, phucnq@uel.edu.vn, hong-tri.nguyen@aalto.fi
anhnht@uit.edu.vn

***Abstract*—High dropout rates are a persistent challenge faced by Massive Open Online Courses (MOOCs). Research in this area is often hindered by the massive volume and the incomplete and inconsistent nature of learner data. This paper introduces Graph Convolutional Network for Imputation (GCN-I), a graph-based method that leverages relationships between learners and courses to fill in missing values and enhance data quality. By integrating GCN-I into deep learning pipelines, we improve classification performance on large and sparse datasets. Experiments on the MOOCCubeX dataset demonstrate significant improvements, with F1-scores reaching up to 0.92 for label D, accuracy up to 95%, and AUC-ROC scores of 0.99, outperforming traditional imputation methods. These results are of great significance for predicting course completion outcomes, enabling timely intervention strategies to help reduce dropout rates.**

***Index Terms*—MOOCs, graph neural networks, data enrichment, data imputation, deep learning**

I. INTRODUCTION

Alongside the rapid development of MOOCs, the dropout rate has increasingly become a major concern [1]. Key contributing factors include low learning motivation, limited interaction between learners and instructors, difficulties in following lecture content, unclear course objectives, and time constraints. These issues often lead to reduced learner engagement, increased procrastination, and, in many cases, course abandonment [2]. Therefore, early prediction of a learner's likelihood to complete a course is crucial for minimizing risks and effectively addressing this problem[3].

Numerous studies have focused on analyzing students' learning behaviors to gain a better understanding of learning motivation and to enhance the prediction of course completion outcomes. However, this task is far from straightforward, as data from MOOC platforms is massive in scale and highly heterogeneous, with a significant proportion of missing and noisy values caused by various factors. These data quality issues pose major challenges for predictive models to achieve accurate results. Common imputation techniques—such as replacing missing values with zeros, mean values, or interpolation—are widely applied to address these problems [4].

Nevertheless, most existing studies still fall short in terms of predictive accuracy, especially when it comes to scalability and generalization. A key reason is that many studies are based on small-scale or less realistic datasets[5]. Therefore, when these conventional methods are applied to large, complex, and realistic datasets such as MOOCCubeX, they may distort the original data values and negatively affect model performance [6]. In addition, the risk of overfitting remains a major obstacle in learning outcome prediction research, primarily because traditional data processing methods introduce bias into the input data, which then propagates through machine learning models—ultimately resulting in output that lacks meaningful value[7][8]. For this reason, it becomes essential to address a significant research challenge: improving data quality and mitigating overfitting in large and diverse datasets like MOOCCubeX.

To address this challenge, we propose a data enrichment strategy that leverages the relationships between entities to enhance input quality for deep learning models. Specifically, our method imputes missing values through a graph-based approach, enabling the model to effectively identify and exploit complex connections among learners, courses, and learning activities. Utilizing a two-stage Generator-Discriminator mechanism, this technique continuously refines the missing values based on feedback from the discriminator. As a result, the input data quality is significantly improved. When integrated with deep learning models, this approach substantially enhances model performance, especially in handling large-scale and incomplete MOOC datasets—where traditional methods often fall short.

The remainder of this paper is organized as follows: Section II reviews related research on data imputation using graph neural networks and deep learning applications in MOOCs. Section III introduces our proposed GCN-I method, including the graph-based data processing strategy and its integration into deep learning pipelines. Section IV presents the experimental setup, outlines the characteristics of the dataset, describes the evaluation metrics employed, and details the comparisons with several baseline imputation methods. This section also provides an analysis and discussion of the results, emphasizing the advantages of the proposed approach while acknowledging its potential limitations that warrant further investigation. Finally, Section V concludes the study and proposes several promising directions for future research.

II. RELATED WORK

A. Graph neural networks

Graph Networks represent a powerful concept in artificial intelligence and machine learning for modeling complex structured data, where the relationships between entities are as important as the entities themselves. In this framework, entities are depicted as nodes, while the connections between them are represented as edges [9]. A prominent realization of this approach is Graph Neural Networks (GNNs), a specialized class of deep learning models designed to process graph-structured data. GNNs facilitate the exchange and aggregation of information across nodes via edges, enabling the model to learn rich, meaningful representations for both individual nodes and the entire graph. These models have found extensive applications in domains such as social network analysis, behavior prediction, and recommendation systems [10][11][12].

In addition to social network analysis and recommendation systems, another significant application of GNNs is the imputation of missing values in graph data, particularly when relationships between objects are critical. Unlike traditional statistical methods or matrix-based deep learning models, GNNs can leverage the entire graph structure to learn contextual representations, thereby improving accuracy when handling missing data. As a result, GNNs can integrate information from neighboring nodes to enhance prediction accuracy while learning latent representations that provide deeper insights into the data structure. Specifically, the study [13] demonstrates that handling missing data using Graph Convolutional Networks (GCNs) yields highly promising results on datasets related to environmental, health, and physical sciences. This has laid the groundwork for us to apply this approach to enhance datasets in the field of education.

GNNs can effectively analyze complex graph-structured data and have been successfully applied in areas such as social network analysis and incomplete dataset processing. More importantly, GNNs offer significant advantages for enhancing deep learning models, particularly in scenarios with high rates of missing data. This capability is especially relevant for MOOCs, where processing incomplete, multi-dimensional data is crucial.

B. Deep learning

In recent years, educational data mining has become a dynamic and rapidly growing research field, with numerous studies exploring new methodologies, primarily aimed at reducing dropout rates. This section provides an overview of some of the most advanced and commonly used deep learning models to outline the problem and propose effective solutions.

Artificial Neural Networks (ANNs) are widely used in learner behavior research, particularly within the field of Educational Data Mining (EDM) [14]. ANNs typically consist of three layers of artificial neurons that process and transmit information. Their strength lies in handling large datasets and uncovering hidden patterns, which helps predict academic outcomes, identify at-risk learners, and personalize learning paths [15]. Additionally, Deep Neural Networks (DNNs) and the CNN-LSTM model are two of the most commonly used methods in studies predicting academic outcomes during the period from 2019 to 2023. Research applying deep learning techniques such as Convolutional Neural Networks (CNNs), DNNs, and Long Short-Term Memory networks (LSTMs) have achieved high accuracy rates exceeding 90%, while some other studies have reported accuracy ranging from 60% to 90% [16]. Among these, the CNN-LSTM model [17] — a hybrid of CNN and LSTM — allows simultaneous extraction of spatial and temporal features from data.

The 4-layer stacked LSTM model is a deep learning architecture composed of four sequentially connected LSTM layers. This model is designed to capture long-term dependencies in time-series data by preserving information over extended sequences. Stacking multiple LSTM layers enables the model to learn more complex features, improving its effectiveness in analyzing sequential data such as student learning behaviors in online education environments. This approach is commonly used for predicting academic results, studying learner behavior, and early identification of at-risk students due to its ability to detect latent trends and long-term relationships within learning activities [18][19].

Furthermore, several other deep learning models have also demonstrated effectiveness in predicting learner outcomes. The Bidirectional Long Short-Term Memory (BiLSTM) model stands out for its ability to process sequential data in both forward and backward directions. Especially when combined with an attention mechanism [20], BiLSTM has become one of the most favored approaches because it not only captures bidirectional information but also focuses on the most important features. Recent studies indicate that integrating BiLSTM with attention significantly enhances prediction accuracy, allowing the system to learn from historical data while identifying key factors influencing academic performance.

III. PROPOSED METHOD

Figure 2 provides a comprehensive overview of the system and its operational workflow. The system comprises four main stages: Data Collection, Data Processing, Predictive Model Training, and Performance Analysis. Each stage plays a critical role in ensuring high-quality input data, accurate predictive models, and meaningful analytical results.

The first stage involves collecting data from MOOCCubeX, a large-scale dataset that records students' learning activities along with course information provided by multiple universities. These data points are crucial for building models and

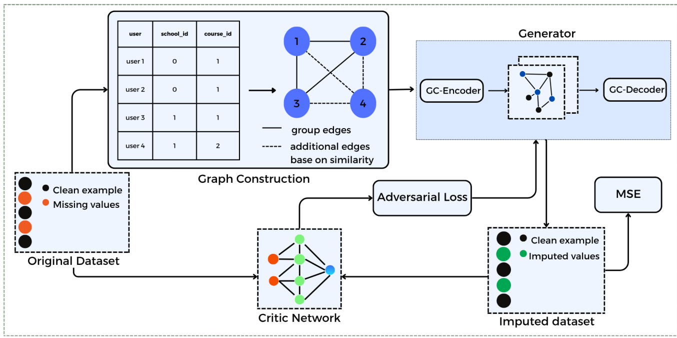

Fig. 1. GCN-Imputer Architecture

predicting academic outcomes. After collection, preprocessing begins with data profiling and imputing missing values using GCNs.

To ensure data quality and readiness for analysis, the evaluation process includes four key steps: Data Overview, Exploratory Data Analysis (EDA), Data Quality Assessment, and Data Preprocessing. This process starts with examining dataset size, data types, and missing value rates to detect incomplete or invalid entries. Statistical summaries and visualizations provide insights into data distributions and attribute correlations. The first stage of quality assessment focuses on completeness and consistency, enabling us to fully understand the dataset's condition and make necessary adjustments. The preprocessing step then improves data integrity by detecting and removing outliers using IQR and Boxplot methods, minimizing the influence of extreme values on model accuracy. Features are normalized using MinMax Scaling to scale values between 0 and 1, reducing disparities caused by differing ranges. Unlike conventional approaches, we conduct a second round of data evaluation after these steps to assess the degree of improvement and fully understand the dataset before proceeding with our proposed imputation method.

To address issues related to sparse and incomplete datasets, we propose an imputation method named GCN-I, where nodes represent learners and edges denote shared attributes such as school, enrolled courses, or data similarity. If edge density is low, additional edges are added based on similarity metrics. The imputation mechanism is built upon a two-stage Generator–Discriminator architecture (Figure 1) utilizing graph convolutional layers. Unlike conventional GCNs, our architecture enhances imputation accuracy through adversarial learning, enabling the generator to iteratively refine missing values based on feedback from the discriminator. The alternating training of these two components ensures optimal results. Model performance is evaluated using Mean Squared Error (MSE) between true and imputed values. After the proposed imputation process, data quality is reassessed both through direct and indirect metrics, allowing for close monitoring of improvements and ensuring clean, optimized input for downstream machine learning pipelines.

Once the data is finalized, it is split into three subsets with a ratio of 8:1:1 (Train, Validation, Test). To mitigate the effects of class imbalance, we examine label distribution across the datasets and apply SMOTE to generate synthetic samples for minority classes. With the input data fully prepared, we train four deep learning models: 4-layer stacked LSTM, GRU, RNN, and BiLSTM.

The GCN-I model addresses missing data through a dual-component architecture: a Generator for data imputation and a Critic to evaluate the imputed values. The Generator uses an encoder-decoder framework with residual connections, where each part contains two GCNConv layers enhanced with BatchNorm1d, ReLU, and Dropout. The Critic, a three-layer GCN network, distinguishes between real and imputed data, outputting a final scalar value while applying normalization and dropout for stability.

The main hyperparameters used in GCN-I are summarized in Table I. These were chosen based on experimental tuning to ensure optimal imputation quality.

TABLE I
HYPERPARAMETERS FOR THE GCN-I MODEL

Parameter	**Value**
Network Architecture	
`hidden_channels`	128
`dropout`	0.3
Training Process	
`optimizer`	Adam
`learning_rate`	0.01
`weight_decay`	5×10^{-5}
`epochs`	100
`early_stop_patience`	10
`gradient_penalty_coeff`	10

IV. EXPERIMENTS

A. Environment

Our experiments utilize the MOOCCubeX dataset from China's XuetangX platform[21]. To ensure consistency and focus on the early learning phase, we filtered and aggregated data from users who were active during the first four weeks after enrollment. A summary of the aggregated data is presented in Table II, highlighting the key features. The dataset includes 16,267 learners, 174 courses, and 915 institutions. Course completion outcomes are categorized into five groups: A (Excellent), B (Good), C (Pass), D (Fail), and E (Incomplete). Among them, learners with average results — group B (3.37%), C (6.34%), and D (6.27%) — account for a significantly smaller proportion compared to the two dominant groups: A (22.88%) and E (61.14%). This imbalance presents a major challenge for machine learning models, as they tend to favor the majority classes, which reduces their ability to accurately predict the minority ones. To address this issue, techniques such as oversampling will be explored and applied in the subsequent stages of the research.

B. Evaluation Metrics

The process of data quality assessment and preprocessing is conducted through an iterative cycle, incorporating multi-

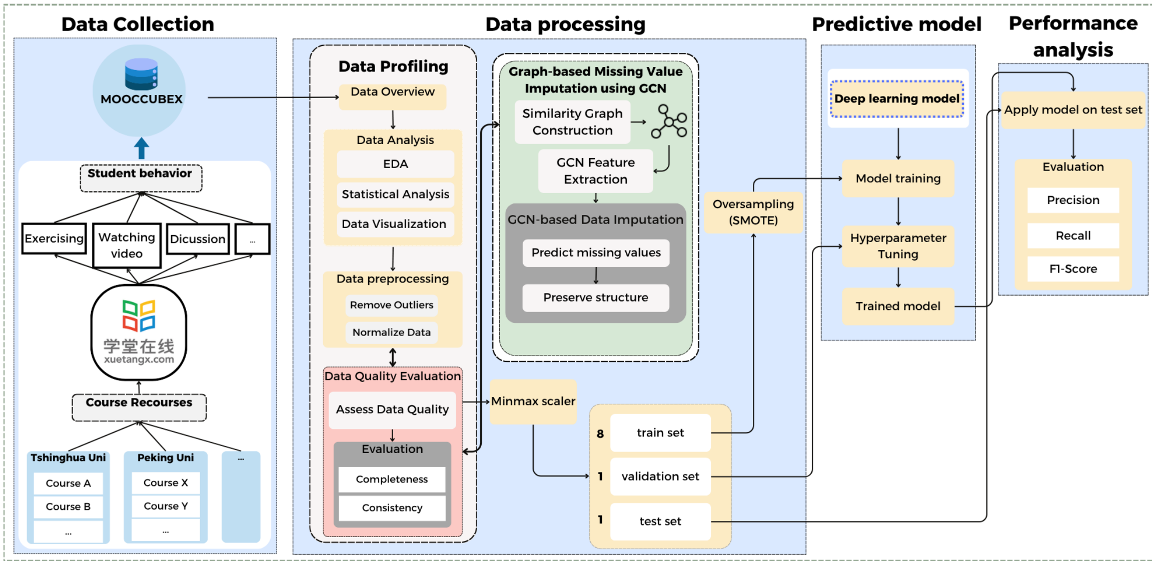

Fig. 2. System Architecture

TABLE II
DATASET DESCRIPTION

Feature	Type	Description
User	Object	Users enrolled
Course	Object	Courses have been enrolled
School	Object	School's learner
Gender	Integer	Gender's learner
Enroll time	Datetime	Timestamp enrolls in a course
Comment count week i	FLoat	Number of comments in week i.
Reply count week i	FLoat	Number of replies in week i
Questions done week i	FLoat	Number of questions completed in week i.
Attempts count week i	FLoat	Number of attempts in week i
Correct answers week i	FLoat	Number of correct answers in week i
Total score week i	FLoat	Total score in week i
User watching time week i	FLoat	Total time (in hours) watching content in week i
Classification	Object	Completion status of the course

* i represents the week number (1, 2, 3, and 4).

dimensional evaluations, including *completeness* and *consistency*, to identify the most optimal method. The detailed formulas for each metric are presented as follows:

$$\text{Completeness}_{\text{df}} = \frac{\text{Total non-missing values}}{\text{Total rows} \times \text{Total columns}} \times 100\% \quad (1)$$

$$\text{Consistency}_{\text{df}} = \frac{\text{Rule-compliant records}}{\text{Total records}} \times 100\% \quad (2)$$

To assess the performance of the classification models, we utilize the F1-Score, a widely adopted metric that harmonizes Precision and Recall into a single measure. This metric is particularly effective in scenarios involving class imbalance and has been employed in various studies for evaluating model performance, such as in the work of Pagano [22].

C. Baseline models comparison

TABLE III
PREPROCESSING STRATEGIES AND BASELINE MODELS

	Preprocessing						Model			
	Raw	LD	Mean	Median	KNN	GCN	LSTM	GRU	RNN	BiLSTM
B1	x						x	x	x	x
B2		x					x	x	x	x
B3			x				x	x	x	x
B4				x			x	x	x	x
B5					x		x	x	x	x
Ours						x	x	x	x	x

Note: LD stands for Listwise-deletion.

Table III summarizes the preprocessing strategies and deep learning models evaluated in this study, including five baseline methods (B1 to B5) and our proposed method (Ours). B1 uses input data where missing values are imputed with zero to assess model performance on minimally processed data. B2 applies conditional listwise deletion, removing records with 50% or more missing values. Baselines B3, B4, and B5 use Mean, Median, and KNN imputation techniques, respectively, before training the models. Each method is tested with four deep learning architectures: LSTM, GRU, RNN, and BiLSTM, allowing us to compare how preprocessing impacts prediction accuracy.

Our proposed method employs GCN-based imputation applied consistently across all four deep learning models, enabling a direct comparison with baseline approaches. This comprehensive evaluation helps identify the best preprocessing strategy and the most effective deep learning model for predicting student academic outcomes.

D. Result

This study evaluated six popular missing value imputation methods: Raw Data, Listwise Deletion, Mean Imputation, Median Imputation, K-Nearest Neighbors (KNN), and Graph Convolutional Network (GCN), using three main criteria: data completeness and consistency (Direct Evaluation), accuracy and classification performance (Reliability), and alignment with actual outcomes (Relevance).

Results indicate that raw data, with only 39.38% completeness, yielded poor model performance (accuracy of 0.65) and low F1-Scores for challenging labels D and E. Listwise Deletion, while achieving 100% completeness by discarding incomplete records, significantly reduced dataset size, impairing model learning and performance, particularly for minority label D. Mean and Median imputation preserved dataset size and efficiently handled missing values but introduced bias by reducing data variability and distorting natural distributions. Although these methods improved accuracy and F1-scores over raw data and Listwise Deletion, their effectiveness remained limited, especially for simpler models like RNN or imbalanced labels. In terms of AUC-ROC, Mean, Median, and Listwise Deletion scored between 0.83 and 0.94, notably lower than KNN and GCN, which both achieved 0.99, demonstrating superior capability in preserving data relevance and enhancing predictive performance.

The KNN method delivered promising results. In terms of accuracy and F1-score, most models significantly outperformed the Mean, Median, and Listwise Deletion methods. Notably, the F1-scores for Label E exceeded 0.90 across models; however, performance on more challenging labels like Label D remained limited, with an F1-score of only 0.69 on the RNN model.

The GCN-I method demonstrates superior performance in predicting learning outcomes, particularly for the minority "Fail" class (D), which constitutes only 6.27% of the dataset and is challenging to predict. Unlike traditional approaches such as Mean, Median, or KNN that treat students independently, GCN-I models the relational graph among students sharing similar attributes (e.g., same course, institution, or learning behavior). This enables effective data enrichment for students with missing information by leveraging connected nodes within the graph.

Furthermore, GCN-I effectively identifies abnormal learning behaviors—typical indicators of at-risk students, such as reduced interaction or disengagement. Its Generator–Discriminator architecture ensures missing data imputation is both accurate and consistent with the global graph structure, reducing noise in inputs for downstream deep learning models.

As a result, deep learning models like GRU and BiLSTM achieve impressive metrics: GRU reaches 95% accuracy, BiLSTM 91%, and the F1-score for the "Fail" class improves to 0.92, significantly outperforming traditional methods (ranging

TABLE IV
EVALUATION OF DATA IMPUTATION METHODS AND DEEP LEARNING MODELS ACROSS QUALITY DIMENSIONS

Dataset	**Direct Evaluation**		**Indirect Evaluation**																			
			Reliability												**Relevance**							
	Completeness	**Consistency**	Accuracy				F1-Score								AUC-ROC							
							Label D				Label E				Label D				Label E			
			RNN	LSTM	BiLSTM	GRU	RNN	LSTM	BiLSTM	GRU	RNN	LSTM	BiLSTM	GRU	RNN	LSTM	BiLSTM	GRU	RNN	LSTM	BiLSTM	GRU
Raw data	39.38%	70.94%	0.65	0.71	0.72	0.81	0.52	0.60	0.56	0.72	0.78	0.85	0.84	0.91	0.87	0.88	0.90	0.94	0.88	0.86	0.91	0.97
Listwise Deletion	100%	100%	0.57	0.64	0.67	0.57	0.55	0.54	0.57	0.54	0.66	0.75	0.77	0.63	0.85	0.83	0.83	0.85	0.81	0.87	0.84	0.84
Mean	100%	100%	0.91	0.85	0.90	0.75	0.55	0.36	0.49	0.47	0.94	0.90	0.93	0.92	0.94	0.90	0.90	0.85	0.98	0.94	0.97	0.96
Median	100%	100%	0.70	0.57	0.66	0.60	0.51	0.41	0.34	0.54	0.81	0.73	0.78	0.75	0.90	0.83	0.87	0.91	0.87	0.87	0.79	0.88
KNN	100%	100%	0.78	0.86	0.80	0.84	0.69	0.71	0.73	0.76	0.90	0.93	0.91	0.94	0.93	0.92	0.92	0.94	0.96	0.97	0.96	0.97
GCN	100%	100%	**0.90**	**0.91**	**0.91**	**0.95**	**0.80**	**0.77**	**0.93**	**0.92**	**0.95**	**0.95**	**0.95**	**0.98**	**0.99**	**0.98**	**0.99**	**0.99**	**0.98**	**0.98**	**0.99**	**0.99**

from 0.36 to 0.76). Additionally, GCN-I enables the highest AUC-ROC of 0.99, reflecting excellent class discrimination. These findings confirm that GCN-I not only enhances input data quality but also substantially improves early warning system performance on large-scale MOOC platforms.

From the above analysis, it can be concluded that GCN-I is the most suitable method for handling datasets with high proportions of missing values. It effectively improves model accuracy, classification performance, and the consistency between input data and prediction results, especially in complex data scenarios.

V. CONCLUSION

This study proposes a graph-based data enrichment method integrated with deep learning models to enhance the prediction of student course completion in MOOC platforms. The proposed approach not only improves early dropout detection accuracy but also enables the timely deployment of personalized intervention strategies, helping to reduce attrition rates and improve learning outcomes. By leveraging relational information among entities such as learners, courses, and learning activities, the model achieves more accurate data imputation than traditional techniques while preserving the structural dependencies inherent in online learning environments.

Unlike conventional methods that treat data points independently, this graph-based approach learns student feature representations within the context of a global learning network, leading to significant improvements in modeling performance. These findings contribute to the growing body of research aimed at improving data quality and predictive accuracy in online education systems, while also demonstrating practical potential for implementation in adaptive learning systems, course recommendation engines, and real-time early warning dashboards for large-scale MOOC platforms.

Future work will focus on optimizing model architectures through hyperparameter tuning and exploring hybrid designs such as Transformer-augmented Graph Neural Networks to better capture complex feature interactions. Furthermore, integrating this approach into real-world applications—such as student analytics dashboards or intelligent tutoring systems on platforms like XuetangX or edX—will be explored to assess its feasibility and scalability. These directions not only validate the practical value of the proposed method but also bridge the gap between academic research and real-world deployment in digital education systems.

ACKNOWLEDGMENT

This research is funded by Vietnam National University HoChiMinh City (VNU-HCM) under grant number DS.C2025-26-12.

REFERENCES

[1] Ahmed A Mubarak, Han Cao, and Weizhen Zhang. Prediction of students' early dropout based on their interaction logs in online learning environment. *Interactive Learning Environments*, 30(8):1414–1433, 2022.

[2] Long Ma and Chei Sian Lee. Leveraging moocs for learners in economically disadvantaged regions. *Education and Information Technologies*, 28(9):12243–12268, 2023.

[3] David Bañeres, M Elena Rodríguez-González, Ana-Elena Guerrero-Roldán, and Pau Cortadas. An early warning system to identify and intervene online dropout learners. *International Journal of Educational Technology in Higher Education*, 20(1):3, 2023.

[4] Zhuojia Xu, Hua Yuan, and Qishan Liu. Student performance prediction based on blended learning. *IEEE Transactions on Education*, 64(1):66–73, 2020.

[5] Mohammad Alshehri, Ahmed Alamri, Alexandra I Cristea, and Craig D Stewart. Towards designing profitable courses: predicting student purchasing behaviour in moocs. *International Journal of Artificial Intelligence in Education*, 31:215–233, 2021.

[6] Fisnik Dalipi, Ali Shariq Imran, and Zenun Kastrati. Mooc dropout prediction using machine learning techniques: Review and research challenges. In *2018 IEEE global engineering education conference (EDUCON)*, pages 1007–1014. IEEE, 2018.

[7] Yaohao Peng and Mateus Hiro Nagata. An empirical overview of nonlinearity and overfitting in machine learning using covid-19 data. *Chaos, Solitons Fractals*, 139:110055, 2020.

[8] Samuel Yeom, Irene Giacomelli, Matt Fredrikson, and Somesh Jha. Privacy risk in machine learning: Analyzing the connection to overfitting. In *2018 IEEE 31st Computer Security Foundations Symposium (CSF)*, pages 268–282, 2018.

[9] Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

[10] Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. Graph neural networks in recommender systems: a survey. *ACM Computing Surveys*, 55(5):1–37, 2022.

[11] Liqun Gao, Haiyang Wang, Zhouran Zhang, Hongwu Zhuang, and Bin Zhou. Hetinf: social influence prediction with heterogeneous graph neural network. *Frontiers in Physics*, 9:787185, 2022.

[12] Shengjie Min, Zhan Gao, Jing Peng, Liang Wang, Ke Qin, and Bo Fang. Stgsn—a spatial–temporal graph neural network framework for time-evolving social networks. *Knowledge-Based Systems*, 214:106746, 2021.

[13] Indro Spinelli, Simone Scardapane, and Aurelio Uncini. Missing data imputation with adversarially-trained graph convolutional networks. *Neural Networks*, 129:249–260, 2020.

[14] Hajra Waheed, Saeed-Ul Hassan, Naif Radi Aljohani, Julie Hardman, Salem Alelyani, and Raheel Nawaz. Predicting academic performance of students from vle big data using deep learning models. *Computers in Human behavior*, 104:106189, 2020.

[15] Orlando Bisacchi Coelho and Ismar Silveira. Deep learning applied to learning analytics and educational data mining: A systematic literature review. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 28, page 143, 2017.

[16] Balqis Albreiki, Nazar Zaki, and Hany Alashwal. A systematic literature review of student'performance prediction using machine learning techniques. *Education Sciences*, 11(9):552, 2021.

[17] Abdulaziz Salamah Aljaloud, Diaa Mohammed Uliyan, Adel Alkhalil, Magdy Abd Elrhman, Azizah Fhad Mohammed Alogali, Yaser Mohammed Altameemi, Mohammed Altamimi, and Paul Kwan. A deep learning model to predict student learning outcomes in lms using cnn and lstm. *IEEE Access*, 10:85255–85265, 2022.

[18] Ashima Kukkar, Rajni Mohana, Aman Sharma, and Anand Nayyar. Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms. *Education and Information Technologies*, 28(8):9655–9684, 2023.

[19] Junru Ren and Shaomin Wu. Prediction of user temporal interactions with online course platforms using deep learning algorithms. *Computers and Education: Artificial Intelligence*, 4:100133, 2023.

[20] Bashir Khan Yousafzai, Sher Afzal Khan, Taj Rahman, Inayat Khan, Inam Ullah, Ateeq Ur Rehman, Mohammed Baz, Habib Hamam, and Omar Cheikhrouhou. Student-performulator: student academic performance using hybrid deep neural network. *Sustainability*, 13(17):9775, 2021.

[21] Jifan Yu, Yuquan Wang, Qingyang Zhong, Gan Luo, Yiming Mao, Kai Sun, Wenzheng Feng, Wei Xu, Shulin Cao, Kaisheng Zeng, et al. MOOCCubeX: A large knowledge-centered repository for adaptive learning in MOOCs. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 4643–4652, 2021.

[22] Tiago P Pagano, Rafael B Loureiro, Fernanda VN Lisboa, Rodrigo M Peixoto, Guilherme AS Guimarães, Gustavo OR Cruz, Maira M Araujo, Lucas L Santos, Marco AS Cruz, Ewerton LS Oliveira, et al. Bias and unfairness in machine learning models: a systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. *Big data and cognitive computing*, 7(1):15, 2023.