Scientific Journal of Computer Science
futuristech PT. Teknologi Futuristik Indonesia
Date of publication April 18, 2026, date of current version April 18, 2026.
Digital Object Identifier 10.64539/sjcs.v2i2.2026.453

Review

Graph Neural Networks (GNN) and Long Short-Term Memory (LSTM) for Forecasting Learner Attrition: A Systematic Review

Chinedu Cory Otuya[1,*], Afolayan Ayodele Obiniyi[2], Joseph Sunday Igwe[3]

[1] Africa Centre of Excellence on Technology Enhanced Learning, National Open University of Nigeria, Plot 91, Abuja 900108, Nigeria; e-mail: ace2225013@noun.edu.ng, ccotuya@nuc.edu.ng (C.C. Otuya).
[2] Department of Computer Science, Federal University Lokoja, Lokoja-Okene Expressway, Felele, Lokoja, Kogi State, Nigeria; e-mail: aaobiniyi@gmail.com (A. A. Obiniyi).
[3] Department of Computer Science, Ebonyi State University, P.M.B. 053, Abakaliki, Ebonyi State, Nigeria; e-mail: igwejoesun@ebsu.edu.ng (J. S. Igwe).
[*] Correspondence

The authors received no financial support for the research, authorship, and/or publication of this article.

Abstract: The issue of learner attrition is a long-standing problem in Open and Distance Learning (ODL) settings where the lack of physical interaction and flexibility exacerbates the risk of disengagement. The use of Deep Learning (DL) techniques for forecasting complexity of behavioural and relational patterns of educational data has grown in usage. While Artificial Intelligence, DL in particular offers superior accuracy in forecasting attrition, the selection of appropriate techniques that addresses temporal sequence and relational patterns remains a critical gap due to inductive biases of ODL settings. This paper performs a systematic review based on the Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA) tool in order to synthesize the current body of knowledge regarding the use of LSTM and GNN in forecasting attrition. The peer-reviewed articles were located in major digital databases and filtered based on predetermined inclusion and exclusion criteria. The review evaluated model archetypes, data properties, metrics of evaluation, and performance results. Results showed that LSTM models were more useful in learning temporal patterns of engagement, whereas GNN models were efficient at learning relational and social learning patterns. Nevertheless, differences in datasets, validation procedures and evaluation metrices made it difficult to directly compare the results. The study identified methodological gaps of single models and recommended the use of hybrid methods for increased accuracy. The review gave consolidated information that direct researchers and institutions in the selection of suitable hybrid deep learning model in forecasting learner attrition.

Keywords: Learner attrition; Deep learning; Graph neural networks; Long short-term memory; PRISMA.

1. Introduction

The issue of learner attrition is a long-standing problem in Open and Distance Learning (ODL), settings where the flexibility of delivery, the lack of physical interaction, and the diversity of learners' population compound the chances of dropping out or attrition [1]-[4]. In contrast to conventional classroom learning, online courses and Massive Open Online Courses (MOOCs) are prone to high dropout rates which is a major waste of institutional resources and institutional breakdown in supporting students [3], [5]-[8]. It is crucial to examine the key determinants of this attrition in the fact that these are the direct determinants of the institutional coefficient of efficiency [3]. As a result, educational data mining (EDM) has become a fundamental part of modern academic retention practices because it has led to the creation of strong early-warning systems [1], [2], [4], [5], [9]-[14].

The initial predictive systems were mainly based on traditional machine learning classifiers that extensively used the non-dynamic socio-demographic and engagement aspects, with moderate predictive accuracy [5], [13], [15]. Attrition is never the consequence of one factor; instead, it develops along complicated behavioural paths and social contacts [2], [3]. This has seen the emergence of the deep learning architectures that have the capacity to model the high dimensional digital trace data of Learning Management Systems (LMS) [11], [12]. Among them, Long Short-Term Memory (LSTM) networks and Graph Neural Networks (GNNs) are two paradigms of modelling that are fundamentally different.

LSTM networks were specifically designed to learn temporal data causalities that exist in sequential data through the use of an elaborate gating structure that controls the retention and modulation of memory cells [2], [16]. They demonstrate a reasonable level of expertise in tracking the temporal dynamics of a single learner, therefore, revealing the occurrence of such phenomena as gradual burnout by a gradual decrease in the frequency of logins or progressive changes in activity profiles [4], [17]. However, learner success is not a discrete chronological event but is deeply interred in both social and structural networks. Theoretical perspectives of academic integration presuppose that peer influence and collaborative networks can be regarded as the salient predictors of persistence [17]. Traditional sequential models on the contrary are essentially node-blind to these relational dynamics - a limitation that GNNs are intended to overcome.

Graph Neural Network are educational information that is presented in the form of complex networks of interconnected objects, such as students, courses, and other units [7], [12]. Such networks represent the relationships, or contagion effects, where the exit of one player increases the probability of exit of the proximate collaborators of the player [7], [18]. More recent studies have examined the concept of multi-topology or dual-graph design to simultaneously measure endogenous behavioural properties and exogenous social ties [5], [12], [18]. Moreover, the predictability of dropouts has been increasing over time, which is due to the development of spatio-temporal GNN models and interpretable relational graph convolutional processes [10], [19], [20]. Other similar hybrid graph enhanced architectures have been successfully applied in advanced predictive tasks, such as chemical structure based activity forecasting, hence demonstrating the scalability and extensiveness of graph neural network architectures to a range of applications [21]. Beyond dropout forecasting, graph neural network designs have also been used to construct individualized learning groups in massively open online courses, as well as to suggest courses, thus further demonstrating their capabilities with respect to modelling intricate networks of educational interactions [6].

Although the number of ensemble approaches that combine Convolutional Neural Networks (CNNs), Gated Recurrent Units (GRUs), or Bidirectional GRUs (BiGRUs) with Graph Convolutional Networks (GCNs) continues to increase [8], [22], there is limited empirical research that explicitly compares the predictive power of pure LSTM models to that of GNNs in the same experimental conditions. Specifically, hybrid architectures incorporating both LSTM and GNN elements have also been proposed in different areas with superior performance in terms of simultaneous temporal and relational features learning [23]. Using the PRISMA guidelines, the review also outlines the current trends, strengths, limitations, and research gaps in the existing literature. Isolating the architectural paradigms and keeping the preprocessing and validation protocols consistent, the study aims at improving methodological transparency on whether the temporal depth or the relational breadth is the more effective framework in the context of the attrition forecasting [1], [24]. Despite the various DL models suggested to tackle the problem of attrition, prevalent literature often evaluate individual models [2] and hybrid models [1]. Rather than treating LSTM and GNN as competing techniques, this study aims to conduct a comparative synthesis of their respective inductive biases. Roh et al. [7] and Li et al. [18] argues that misalignment of models to dataset causes underperformance, highlighting specifically in the case where pure LSTM was applied to an environment where the primary driver of attrition was peer-influence or social network. The author described this situation as "node-blind" specifically in recall because its chronological architecture is misaligned with the relational structure of the data. Lower performance of DL models may also be as a result of limited scope of the dataset. In [2], the authors deployed an LSTM in an ODL context but only achieved an accuracy of 0.57. While in [25] class imbalance using a single metric (like basic Accuracy) causes the model to appear misaligned or fail to capture the actual at-risk students, necessitating a multi-metric approach. Thus by isolating models and evaluating how they interact with unique educational datasets, this review seeks to clarify the specific conditions under which temporal depth or relational breadth is most effective, ultimately guiding the development of robust, complementary hybrid architectures in identifying at-risk learners at an early age to institute suitable intervention.

The rest of this paper presents Section 2 outlining the method, which includes research questions and search strings. Section 3 presents an overview of the LSTM and GNN paradigms. Section 4 provides the comparative results, statistical synthesis, and addresses the research questions. Finally, Section 5 concludes the study with implications and future research directions.

2. Method

This systematic review was prepared in accordance with the Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA) guidelines, which made it a transparent, objective, and reproducible synthesis of the deep-learning applications to forecasting attrition. The stages of the systematic review process, including identification, screening, eligibility, and inclusion, are illustrated in Figure 1. The research was guided by the following research questions:

Table 1. Inclusion and Exclusion Eligibility Criteria.

Criteria Type	**Inclusion Criteria**	**Exclusion Criteria**
Domain	Educational settings (ODL, MOOCs, Higher Ed), or parallel attrition domains	General NLP or computer vision tasks unrelated to human attrition/dropout.
Architecture	Employs LSTM, GNN, or hybrid combinations (e.g., CNN-LSTM, GNN-LSTM).	Traditional Artificial Intelligence.
Outcomes	Must report quantitative classification metrics (Accuracy, Precision, Recall, F1, AUC).	Qualitative studies, conceptual frameworks without empirical validation.
Publication	Peer-reviewed journal articles and conference proceedings (2021–2026).	Preprints, non-peer-review articles.

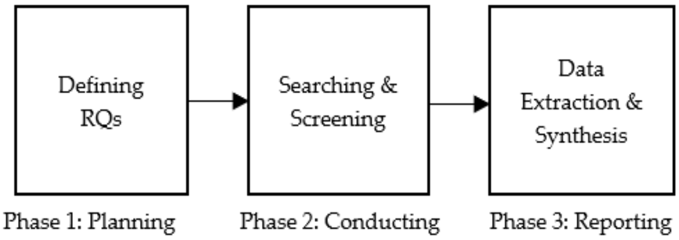

Figure 1. Systematic Review stages diagram.

a) What are the current architectural trends in applying LSTM and GNN for forecasting learner attrition?
b) How does the performance of LSTM and GNN models compare across DL evaluation metrics?

2.1. Search Strategy and Information Sources

A structured literature search was carried out in the key academic databases, which included, IEEE Xplore, ProQuest, ScienceDirect, ACM Digital Library, SpringerLink, and Google Scholar. Such databases have been chosen to ensure that engineering, computer science, and educational technology scholarship is fully covered. The search strategy included Boolean operators and domain-specific keywords including learner attrition, student dropout, dropout forecasting, deep learning, Long Short-Term Memory, LSTM, and Graph Neural Network, GNN, MOOC, and Open and Distance Learning. The search timeframe covered the years between January 2021 and March 2026, which allowed capturing recent developments in deep-learning designs and changing inductive biases applicable to sequential and graph-based modelling [2], [26], [27]. Besides searching databases, backward and forward citation of the highly cited and relevant studies was conducted to determine additional publications that met the eligibility criteria. Peer-reviewed journal articles and conference proceedings were taken into consideration only hence ensuring academic and technical strength.

2.2. Inclusion and Exclusion Criteria

The studies were incorporated in case they reported empirical studies that applied pure or hybrid LSTM or GNN architectures to conduct predictive modelling tasks, especially in the area of learner attrition. It was also taken into account to use hybrid models based on complementary architecture, which included CNN, GRU, BiGRU, or Graph Convolutional Networks when it was a main part of the predictive framework [1], [8], [28]. The main area of concern was the attrition of learners in Open and Distance Learning (ODL) and Massive Open Online Courses (MOOCs) [7], [10]. However, research that examines the corresponding attrition predictive situations, including employee attrition [27], was included when they presented transferrable DL methodologies that could be applied to educational forecasting modelling. Human-centric disengagement often follows similar patterns of longitudinal and relational decay patterns within corporate and academic networks. Thus, the research [27] provided critical architectural insights applicable to learner attrition. Finally, the studies had to be eligible, i.e. supply quantifiable performance measures, such as but not limited to Accuracy, Precision, Recall, F1-score, or ROC-AUC, to support systematic comparison. The studies were not included in case they did not contain enough architectural information, had no empirical verification, were non-peer-reviewed editorials or opinion papers, and were not in English. These criteria were used to guarantee technical comparability and methodological consistency between the ultimate study pool (Table 1).

2.3. Data Extraction and Synthesis

Data extracted from the final pool of articles included model archetypes (e.g., SIG-Net, CNN-BiGRU-GCN), data properties (clickstream data, social interaction graphs), and performance outcomes [12], [13], [28]. We categorized the findings into two primary modeling paradigms: Temporal Sequence Modeling (focusing on LSTM gating) and Relational Graph Modeling (focusing on GNN message passing) [16], [26].

2.4. Model Archetypes, Data Properties, Metrics of Evaluation, and Performance Results

The review presented a spectrum of paradigms of models, including standard Long Short-Term Memory networks (LSTMs) with univariate time-series data [4] and more advanced relational networks like Study-GNN [13]. Data properties are largely divided into chronological

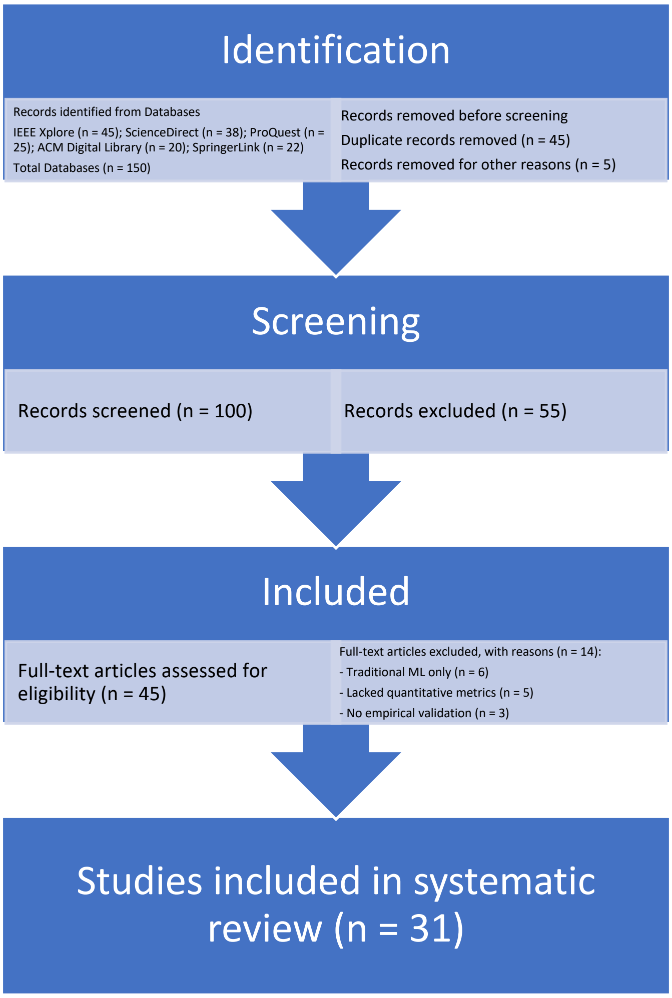

Figure 2. PRISMA Flow Diagram.

clickstream logs and interaction matrices [29]. Evaluation metrics in the reviewed literature prioritized the F1-score and Recall over Accuracy due to class imbalance inherent in attrition datasets [18], [24]. The literature brought out three salient architectural trends. To begin with, LSTM-only models are primarily used to sequential LMS logs, and clickstream data, showing a high level of success in the modelling of longitudinal engagement patterns [12], [13]. Second, graph-based models which were purely GNN-based models, used graph convolution and structural relational aggregation to learn peer influence and structural dependencies based on graph-structured representations of student-to-student or student-to-course interactions [7], [10]. Third, hybrid architectures that include convolutional, recurrent, and graph-based architectures, including CNN-BiGRU-GCN variants that were designed to trade-off temporal depth and relational representation ability [8], [28]. The performance measures reported in the literature have shown variations with F1-score and ROC-AUC being a consistent performance metrics documented in the literature, however, variation in evaluation protocols, data partitioning methods, and class imbalance managing factors restricts the straightforward quantitative comparability.

2.5. Data Analysis

The analytical approach used by the study focused on comparative thematic evaluation across architectural paradigms. Models were examined with respect to their inductive biases, representational capacity, computational complexity, and reported predictive performance using evaluation metrics. Particular emphasis was placed on understanding how sequential dependency modelling in LSTMs contrasted with relational dependency modelling in GNNs. This involved an investigation of the "forget-gate" mechanics of LSTMs [16], [30] against the "neighbourhood aggregation" or "message passing" techniques of GNNs [1], [12], [26]. Performance was synthesized by grouping studies based on dataset scale and the degree of social interaction features present. Trend analysis further indicated a growing shift toward graph-based approaches in recent years, reflecting increased recognition of relational dynamics in learner attrition behaviour. Nevertheless, LSTM-based architectures remain prominent due to their effectiveness in modelling temporal engagement trajectories.

2.6. Validity and Reliability

Internal validity was strengthened through adherence to PRISMA guidelines, predefined eligibility criteria, and structured screening procedures (Figure 2). Also to ensure validity, the selection process involved cross-verification of studies against established EDM benchmarks [9], [14], [24]. Cross-database searches and citation tracking reduced the likelihood of omission bias. Reliability was supported through the consistent application of a standardized data extraction template and careful verification of architectural descriptions, reported metrics, such as k-fold cross-validation or hold-out testing, as exemplified in recent baseline studies [4], [5]. Despite these measures, variability in reporting standards and dataset characteristics across studies constitutes an inherent limitation affecting cross-study comparability. The absence of standardized benchmark datasets further constrains direct performance aggregation.

2.7. Ethical Consideration

This review adhered to ethical guidelines by ensuring proper attribution of all reviewed works and focusing on anonymized or peer-reviewed datasets. As the study moves toward practical application in ODL, it emphasizes the importance of explainability in GNNs and LSTMs to prevent algorithmic bias during student intervention [7], [10], [27].

3. LSTM and GNN Overview

The systematic review identified two distinct pathways through which deep learning addresses the attrition problem.

Table 2. Comparative Summary of LSTM and GNN in Attrition Research.

Key Feature	LSTM Paradigm	GNN Paradigm	Key References
Data Structure	Sequential/Time-series	Graph/Relational	[16], [26]
Inductive Bias	Chronological dependencies	Structural connectivity	[7], [17]
Strengths	Individual burnout patterns	Social/Network contagion	[4], [10]
Input Type	LMS Log sequences	Interaction Adjacency Matrices	[13], [29]
Explainability	Attention over time	Relational edge importance	[7], [27]

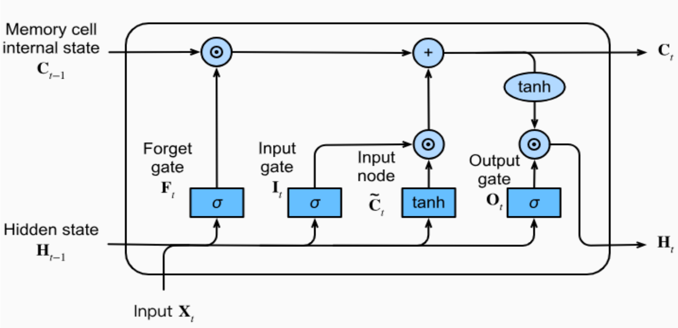

Figure 3. Schematic Diagram of a typical LSTM [16].

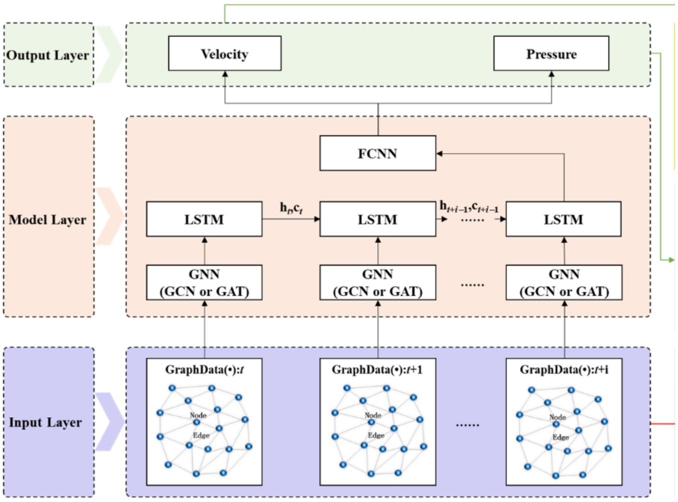

Figure 4. Schematic Diagram of Hybrid LSTM-GNN [19].

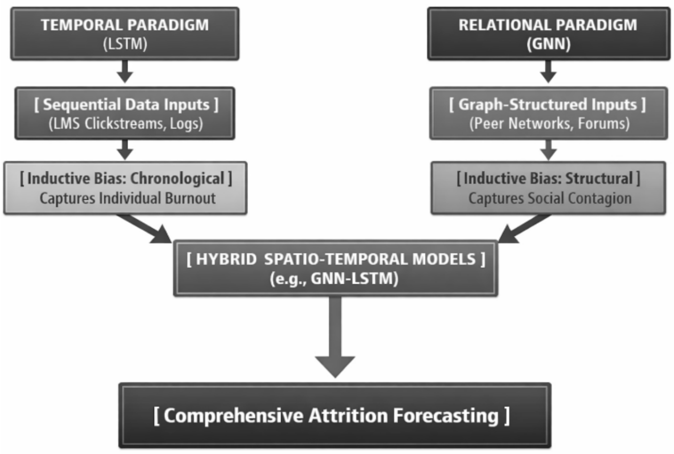

Figure 5. Conceptual Framework.

3.1. Temporal Modelling with LSTM

LSTM networks have become the baseline for sequential student data [4], [31]. By utilizing input, forget, and output gates, these models effectively process the "temporal trajectory" of students [16], as illustrated in Figure 3. As noted by Umar et al., the ability to remember long-term engagement trends allows LSTMs to outperform traditional machine learning in identifying "slow burnout" patterns [4].

3.2. Relational Modelling with GNN

GNNs represent the "structural shift" in attrition research. Rather than viewing a student in isolation, GNNs model the student as a node within an interaction graph [7], where information is propagated through neighboring nodes via message passing and aggregation mechanisms, as illustrated in Figure 6. Frameworks like SIG-Net use student interaction graphs to capture the social contagion of dropout [7]. Studies show that relational features often provide higher sensitivity (recall) because disengagement frequently spreads through peer networks [1], [10]

3.3. Systematic Comparison of Architectures

Based on the PRISMA review, Table 2 summarizes the key differences between LSTM and GNN paradigms in attrition research. LSTM focuses on sequential data and chronological dependencies to model individual behavior patterns, while GNN captures relational structures and network interactions. These differences are reflected in their input types and explainability approaches, highlighting their complementary strengths in modeling attrition phenomena.

3.4. Emerging Hybrid and Spatio-Temporal Models

Recent research from 2025–2026 suggests a convergence of these paradigms. Hybrid models such as CNN-GNN-LSTM [15] and CNN-BiGRU-GCN [28] are being deployed to capture both the local feature extraction and the spatio-temporal dynamics of learner behaviour [6], [19], [30]. It is necessary to differentiate the inductive biases of temporal sequencing (LSTM) and relational connectivity (GNN) and combine them into hybrid forecasting models, as illustrated in Figure 4 and Figure 5.

4. Result and Discussion

The PRISMA screening procedure resulted in the inclusion of 31 peer-reviewed studies published between 2021 and 2026 that satisfied the predefined eligibility criteria (Figure 2). The selected literature represents empirical

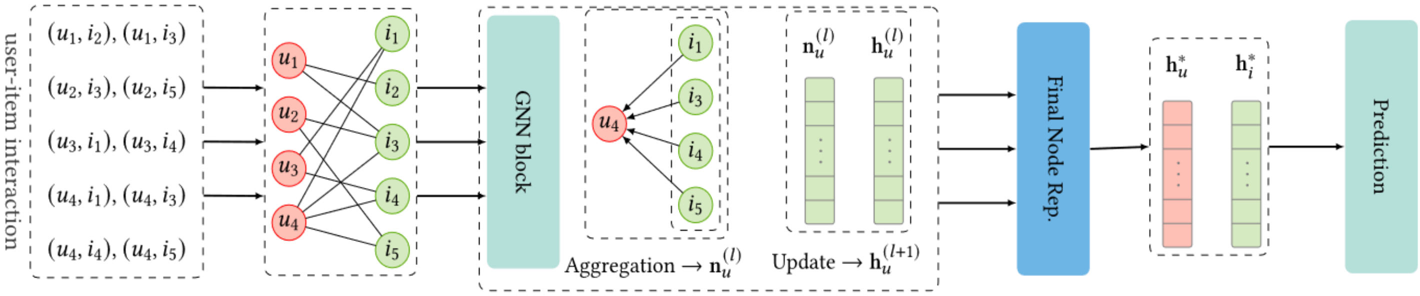

Figure 6. Schematic Diagram of Typical flow of GNN [32].

Table 3. Summary of Metrics across Key Studies.

Study/ Model	Architecture Type	Dataset Domain	Accuracy (%)	Precision	Recall	F1-Score	ROC-AUC	Validation Method
[2]	LSTM	ODL	0.57	—	—	—	—	Train/Test Split
[7]	SIG-Net (GNN)	MOOC (KDD Cup)	0.89	0.865	0.894	0.879	—	Cross-Validation
[12]	Dual GNN	Online Learning (OULAD)	0.93	0.915	0.902	0.908	—	Cross-Validation
[5]	Hybrid LSTM + 2D-CNN	Campus Multi-source Data	0.90	0.885	0.892	0.889	—	Cross-Validation
[13]	Random Forest (RF)	LMS Clickstream	0.91	0.89	0.91	0.9	0.94	10-Fold CV
[9]	Random Forest	Higher Education	0.709	—	—	—	—	10-Fold CV
[14]	WResNeXt-MJ	Educational Environment	0.98	0.984	—	—	—	10-Fold CV
[15]	Regularized Regression	Higher Education (US)	—	—	—	—	0.88	10-Fold CV
[24]	XGBoost	Employee Attrition	0.955	—	0.984	—	—	Cross-Validation
[18]	Study-GNN	Educational Network	> GCN	—	—	—	—	Cross-Validation
[21]	Dumpling GNN (Hybrid)	Chemical Structure	—	0.412	—	—	0.784	Cross-Validation
[33]	GNN-LSTM	Fire Weather Index	—	0.85	—	—	0.96	Cross-Validation
[25]	7 Classifier Framework	Distance Learning	—	—	—	—	—	Cross-Validation

Table 4. Statistical Summary of Model Performance Metrics.

Metric	Minimum	Maximum	Mean	Std. Dev.
Accuracy	0.57	0.98	0.86	0.13
Precision	0.87	0.98	0.91	0.04
Recall	0.89	0.98	0.92	0.04
F1-Score	0.88	0.91	0.89	0.01
ROC-AUC	0.78	0.96	0.89	0.08

applications of Long Short-Term Memory (LSTM), Graph Neural Networks (GNN), and hybrid deep learning architectures applied to forecasting attrition and related problems. The results synthesize quantitative performance outcomes, architectural trends, and methodological implications aligned with the objectives of this systematic review.

It is important to mention that the statistical aggregations used in Table 4 are purely descriptive because of the high heterogeneity of the datasets involved in the reviewed studies, the type of features used, and validation procedures. As a result of high heterogeneity of the dataset sizes, types of features, and validation procedures, these aggregated measures cannot be used as a direct benchmark comparison. In addition, because of the severe imbalance between classes attrition.

4.1. LSTM vs GNN Comparative Performance

Table 3 summarizes evaluation metrices obtained from selected studies on application of LSTM and GNN. In the sampled literature, predictive performance of the models has high level of validity due to the characteristics of the dataset, validation procedures, and architectural inductive biases. Accuracy values have been reported as high as 0.57 in early LSTM deep learning model deployed in an Open and Distance Learning dataset [2] and as high as 0.98 in more sophisticated deep -learning models on complex educational data [14]. The statistical aggregation summarized in Table 4 indicates a mean accuracy of 0.86, suggesting that deep learning approaches generally outperform traditional machine learning baselines previously used for attrition forecasting [9], [15]. The value of precision and recall also have comparatively high ranges, with the mean values of precision and recall as 0.91 and 0.92 respectively. Studies employing relational graph modelling, particularly SIG-Net and dual-GNN frameworks, consistently reported higher recall scores, reflecting improved sensitivity in identifying at-risk learners within socially connected learning environments [7], [12]. The result confirms the theoretical assumption that dropout behaviour spreads via interaction networks as opposed to individual temporal patterns emerging only [1], [10]. The average F1-score of 0.89 also indicate a balanced classification performance although the imbalance between the classes in attrition datasets as very strong [18], [24]. The average values of ROC-AUC is 0.89, which shows that the models have strong discriminatory abilities and that the deep-learning approaches are already mature enough to predict educational analytics.

The result helps to investigate which of the two predictive variables, relational breadth or temporal depth is more important. The data shows that while LSTMs measures the timing of "when" of attrition happens, GNNs better captures the "why" by modelling student-to-student relationship. The review also establishes that the inductive bias of the LSTMs on the chronological dependencies is most efficient in click-stream logs as indicated by Liu et al. [13], who had an ROC-AUC of 0.94. On the other hand, on platforms like MOOCs where peer communication is widespread, relational bias of GNNs is better. Roh et al. [7] utilized the SIG -Net (a GNN) to report an F1-score of 0.88, which proves that structural connectivity is an effective predictor of dropout contagion. Also, the study established which method among the two, temporal sequencing and relational modelling, has a better predictive value in predicting learner attrition. The result shows that there is no consistent predominance between the two paradigms, the comparative advantages are based on the structural features of the dataset and the contextual learning situations [25]. LSTM models proved to have consistent predictive performance when trained on sequential LMS click -stream and longitudinal patterns of engagement. Their gating systems were successful in capturing time-related dependencies that were relevant to the gradual disengagement profiles, thus supporting previous results of sequential behavioural modelling in educational contexts [4], [16].

The comparative analysis of activity-decline trends has supported the fact that LSTM models are especially effective in identifying the progressive changes of behavioural patterns before dropout events [13], [18]. On the other hand, GNN models showed better recall and sensitivity measures in a situation where the relational interaction data were present. Models like the SIG-Net model treat learners as nodes in interaction graphs and uses the message-passing and neighbourhood-aggregation mechanisms to model the peer-influence effects [7], [26]. Experimental data indicate that the contagious nature of disengagement often spreads through network effects, and, therefore, graph-based research is more likely to produce improved minority-class classification [10], [20]. These are direct answers to the research question which was set out in the first part of the Section 2, that is the relative effects of temporal depth and relational breadth. Temporal models are good in the individual development of behaviour and the graph models are better at the collective structural behaviour; this shows that they complement each other and not competing.

4.2. Discussion of Methodological Gaps

The statistical summary presented in Table 4 shows a high Mean Accuracy of 0.86 but also a high Standard Deviation 0.13, which indicates that there is a lot of inconsistency in how attrition is defined and measured in different institutions. This finding supports the critique by Takaki et al. [25], who proposed that the use of a single metric is inadequate when there is an imbalance in the data in the distance learning settings. This necessitates an urgent requirement of Explainable Artificial Intelligence (XAI) in such predictive models to ensure that the interventions are accurate, as well as transparent and ethical [10], [24].

Results further indicate strong mediation of the architectural success through the characteristics of datasets. Studies using a largely chronological click-stream log favoured LSTM-based solutions, due to their inherent inductive bias of sequential dependency modelling [16], [29]. On the other hand, the data sets that included collaboration networks, course relationships, or peer-interaction graphs had better performance when operated in GNN frameworks [5], [12]. The growing popularity of temporal graph networks highlights a new point of intersection between these paradigms of method. The studies of spatio-temporal GNN have reported superior predictive robustness in heterogeneous datasets suggesting that the combination of both temporal memory mechanisms and relational aggregation reduce limitations related to isolated architectures [11], [19]. This direction is in line with the larger developments in graph-learning studies, which have demonstrated scalability to other fields outside education, such as recommender systems and scientific predictive tasks [21], [31].

Despite encouraging predictive outcomes, the systematic synthesis showed that there are still some methodological issues. To begin with, no standardized benchmark datasets exist, which does not allow making a direct quantitative comparison of studies, a problem also observed in larger machine learning evaluations [27], [28]. Second, evaluation metrics are inconsistently reported which limits reproducibility and prevents aggregation of meta-analyses. Although F1 -score and ROC-AUC were commonly provided, some of the studies only used accuracy measures, which may be inaccurate when it comes to class-imbalance conditions [18], [24]. Also, explainability is an issue that is not yet resolved, especially with graph-based architectures, whose decision-making is not as transparent as sequential attention mechanisms. Explainable GNN methods are thus becoming a critical research agenda towards ethical intervention systems [10], [24].

4.3. Explainability XAI in Temporal and Relational Models

One of the major impediments to institutional adoption of Deep Learning is the black box nature of these models [10], [24]. The differences between the two paradigms of explainable AI (XAI) are inherently different. Explainability in LSTM architectures is usually provided by temporal attention mechanisms that indicate when the engagement of a student started to worsen (e.g., indicating a particular week of missed assignments). In contrast, GNNs are based on structural explainers (including GNNExplainer) in order to detect critical edges and node effects, exposing who or what relational factors (e.g. a disconnected peer group) were causing the attrition risk [7], [27]. To intervene ethically, both the chronological trigger as detected by LSTMs and the contextual social trigger as detected by GNNs are needed by the institutions.

5. Conclusion

This systematic review validates transition from traditional Machine Learning to Deep Learning paradigms, namely LSTM and GNN, which has radically changed the dynamics of forecasting learner attrition in ODL settings. Through the synthesis of empirical evidence from 2021–2026, a key finding identified in the research is that architectural effectiveness of data topology. LSTMs are the gold standard when it comes to single-temporal burnout trajectories, whereas GNNs have better sensitivity to social network and peer-influence dropouts. The line of research made a case for a hybrid spatio-temporal architecture as a more resilient system for heterogeneous LMS environments. Nonetheless, to transition these models beyond the confines of individual studies and into the general institutional practices, four key obstacles need to be overcome, including standardizing benchmark datasets, cross-institutional testing, introducing XAI to establish ethical transparency, and training models to operate in noisy data settings. In conclusion, the convergence of temporal memory and relational aggregation represents the next phase of intelligent early-warning systems as illustrated in the conceptual framework in Figure 4. properly match their model choice to their respective model of instruction, focusing on GNNs in collaborative models and LSTMs to forecast behaviour longitudinally. In so doing, the institutions are able to harness the predictive power of Artificial Intelligence / Deep Learning in forecasting attrition and mitigating dropout rate of at-risk learners.

6. Declarations

6.1. Author Contributions

Chinedu Cory Otuya: Conceptualization, methodology, formal analysis, investigation, data curation, writing of original draft; **Afolayan Ayodele Obiniyi:** methodology, review and editing, supervision; **Joseph Sunday Igwe:** formal analysis, review and editing. All authors have read and agreed to the published version of the manuscript.

6.2. Institutional Review Board Statement

Not applicable.

6.3. Informed Consent Statement

Not applicable.

6.4. Data Availability Statement

The data presented in this study are available on reasonable request from the corresponding author.

6.5. Acknowledgment

The authors acknowledge the use of Grammarly AI for linguistic refinement and technical spell-checking during the preparation of the original draft. The Authors also acknowledge the support of human and material resources of the Africa Centre of Excellence on Technology Enhanced Learning (ACETEL), National Open University of Nigeria (NOUN).

6.6. Conflicts of Interest

The author declares no conflicts of interest.

7. References

[1] K. R. Kannan, K. T. M. Abarna, and S. Vairachilai, "Graph Neural Networks for Predicting Student Performance: A Deep Learning Approach for Academic Success Forecasting," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12,, no. 1s, pp. 228-232, 2024. https://ijisae.org/index.php/IJISAE/article/view/3410.

[2] J. N. Ndunagu, D. O. Oyewola, F. S. Garki, J. C. Onyeakazi, C. U. Ezeanya, and E. Ukwandu, "Deep Learning for Predicting Attrition Rate in Open and Distance Learning (ODL) Institutions," *Computers*, vol. 13, no. 9, p. 229, Sep. 2024. https://doi.org/10.3390/computers13090229.

[3] C. Otuya, A. A. Obiniyi, J. S. Igwe, D. M. Adayilo, and E. Ladan, "Analysing Critical Determinants of Learner Attrition in Open and Distance Learning: Effect on Coefficient of Efficiency," in *Proceedings of the 1st International Hybrid Conference of Faculty of Science, Federal University Gusau*, 2025. https://doi.org/10.57233/ihcfs.013.

[4] I. Y. Umar, K. I. Musa, and M. Tella, "Long Short-Term Memory (LSTM) Based Model for Forecasting Students' Dropout," *International Journal of Scientific Research and Management Studies (IJSRMS)*, vol. 11, no. 7, pp.124-135, 2025. [Online]. Available: https://www.isroset.org/pdf_paper_view.php?paper_id=3915&14-ISROSET-IJSRMS-10579.pdf.

[5] M. Li, X. Wang, Y. Wang, Y. Chen, and Y. Chen, "Study-GNN: A Novel Pipeline for Student Performance Prediction Based on Multi-Topology Graph Neural Networks," *Sustainability*, vol. 14, no. 13, p. 7965, Jun. 2022. https://doi.org/10.3390/su14137965.

[6] Z. Luo, X. Wang, Y. Wang, H. Zhang, and Z. Li, "A Personalized MOOC Learning Group and Course Recommendation Method Based on Graph Neural Network and Social Network Analysis," *arXiv preprint arXiv:2410.10658*, Oct. 2024. [Online]. Available: http://arxiv.org/abs/2410.10658.

[7] D. Roh, D. Han, D. Kim, K. Han, and M. Y. Yi, "SIG-Net: GNN-Based Dropout Prediction in MOOCs Using Student Interaction Graph," in *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, Avila, Spain: ACM, Apr. 2024, pp. 29–37. https://doi.org/10.1145/3605098.3636002.

[8] X. Wu, Z. Yu, C. Zhang, and Z. Zhiheng, "Research on MOOC Dropout Prediction by Combining CNN-BiGRU and GCN," in *Proceedings of the Fourth International Conference on Computer Vision, Application, and Algorithm (CVAA 2024)*, H. Yuan and L. Leng, Eds., Chengdu, China: SPIE, Jan. 2025, p. 109. https://doi.org/10.1117/12.3055872.

[9] D. K. Dake and C. Buabeng-Andoh, "Using Machine Learning Techniques to Predict Learner Dropout Rate in Higher Educational Institutions," *Mobile Information Systems*, vol. 2022, pp. 1–9, Nov. 2022. https://doi.org/10.1155/2022/2670562.

[10] Y. Guo and Y. He, "MOOC Dropout Prediction Using Explainable Relational Graph Convolution," *IEEE Access*, vol. 13, pp. 204759–204772, 2025. https://doi.org/10.1109/ACCESS.2025.3637528.

[11] Q. Huang and J. Chen, "Enhancing Academic Performance Prediction with Temporal Graph Networks for Massive Open Online Courses," *Journal of Big Data*, vol. 11, no. 1, p. 52, Apr. 2024. https://doi.org/10.1186/s40537-024-00918-5.

[12] Q. Huang and Y. Zeng, "Improving Academic Performance Predictions with Dual Graph Neural Networks," *Complex & Intelligent Systems*, vol. 10, no. 3, pp. 3557–3575, Jun. 2024. https://doi.org/10.1007/s40747-024-01344-z.

[13] Y. Liu, S. Fan, S. Xu, A. Sajjanhar, S. Yeom, and Y. Wei, "Predicting Student Performance Using Clickstream Data and Machine Learning," *Education Sciences*, vol. 13, no. 1, p. 17, Dec. 2022. https://doi.org/10.3390/educsci13010017.

[14] L. Liu and L. Wan, "Innovative Models for Enhanced Student Adaptability and Performance in Educational Environments," *PLoS ONE*, vol. 19, no. 9, p. e0307221, Sep. 2024. https://doi.org/10.1371/journal.pone.0307221.

[15] S. C. Matz, C. S. Bukow, H. Peters, C. Deacons, A. Dinu, and C. Stachl, "Using Machine Learning to Predict Student Retention from Socio-Demographic Characteristics and App-Based Engagement Metrics," *Scientific Reports*, vol. 13, no. 1, p. 5705, Apr. 2023. https://doi.org/10.1038/s41598-023-32484-w.

[16] A. Zhang, Z. C. Lipton, Mu Li, and Alexander J. Smola, "Long Short-Term Memory (LSTM)," in *Dive into Deep Learning*. [Online]. Available: https://d2l.ai/chapter_recurrent-modern/lstm.html.

[17] N. A. B. Hasman, N. B. B. A. Mustafa, and M. N. Chik, "Deep Learning Models Performance Comparison for Solar Energy Generation Forecasting in a Large-Scale Solar Farm," *IOP Conference Series: Earth and Environmental Science*, vol. 1560, no. 1, p. 012041, Nov. 2025. https://doi.org/10.1088/1755-1315/1560/1/012041.

[18] X. Li, Y. Zhang, H. Cheng, M. Li, and B. Yin, "Student Achievement Prediction Using Deep Neural Network from Multi-Source Campus Data," *Complex & Intelligent Systems*, vol. 8, no. 6, pp. 5143–5156, Dec. 2022. https://doi.org/10.1007/s40747-022-00731-8.

[19] W. Guo, C. Cheng, C. Huang, Z. Lu, K. Chen, and J. Ding, "A Spatio-Temporal Graph Neural Network for Predicting Flow Fields on Unstructured Grids with the SUBOFF Benchmark," *Journal of Marine Science and Engineering*, vol. 13, no. 9, pp. 1647–1670, 2025. https://doi.org/10.3390/jmse13091647.

[20] L. Yan, *et al.*, "Streamflow Prediction of Spatio-Temporal Graph Neural Network with Feature Enhancement Fusion," *Symmetry*, vol. 18, no. 2, p. 240, Jan. 2026. https://doi.org/10.3390/sym18020240.

[21] S. Xu, L. Xie, R. Dai, and Z. Lyu, "Dumpling GNN: Hybrid GNN Enables Better ADC Payload Activity Prediction Based on the Chemical Structure," *International Journal of Molecular Sciences*, vol. 26, no. 10, p. 4859, May 2025. https://doi.org/10.3390/ijms26104859.

[22] C. Yang, P. Jin, and Y. Chen, "Leveraging Graph Neural Networks and Gated Recurrent Units for Accurate and Transparent Prediction of Baseball Pitching Speed," *Scientific Reports*, vol. 15, no. 1, p. 7745, Mar. 2025. https://doi.org/10.1038/s41598-025-88284-x.

[23] M. S. Sonani, A. Badii, and A. Moin, "Stock Price Prediction Using a Hybrid LSTM-GNN Model: Integrating Time-Series and Graph-Based Analysis," *arXiv preprint arXiv:2502.15813*, Feb. 2025. https://doi.org/10.48550/arXiv.2502.15813.

[24] C. Makanga, *et al.*, "Explainable Machine Learning and Graph Neural Network Approaches for Predicting Employee Attrition," in *Proceedings of the 2024 Sixteenth International Conference on Contemporary Computing*, Noida, India: ACM, Aug. 2024, pp. 243–255. https://doi.org/10.1145/3675888.3676058.

[25] P. Takaki, M. L. Dutra, G. De Araújo, and E. M. D. S. Júnior, "A Proposed Framework for Evaluating the Academic-Failure Prediction in Distance Learning," *Mobile Networks and Applications*, vol. 27, no. 5, pp. 1958–1966, Oct. 2022. https://doi.org/10.1007/s11036-022-01965-z.

[26] S. Abadal, A. Jain, R. Guirado, J. López-Alonso, and E. Alarcón, "Computing Graph Neural Networks: A Survey from Algorithms to Accelerators," *ACM Computing Surveys*, vol. 54, no. 9, pp. 1–38, Dec. 2022. https://doi.org/10.1145/3477141.

[27] J. Cherian and R. Kumar, "Fundamentals of Machine Learning," in *Machine Learning Fundamentals*, 2023, pp. 147–174. https://doi.org/10.1007/978-3-031-22206-1_6.

[28] D. Möller, "Machine Learning and Deep Learning," in *Machine Learning and Deep Learning*, 2023, pp. 347–384. https://doi.org/10.1007/978-3-031-26845-8_8.

[29] J. Bhanbhro, A. A. Memon, B. Lal, S. Talpur, and M. Memon, "Speech Emotion Recognition: Comparative Analysis of CNN-LSTM and Attention-Enhanced CNN-LSTM Models," *Signals*, vol. 6, no. 2, p. 22, May 2025. https://doi.org/10.3390/signals6020022.

[30] G. Wei, *et al.*, "Drilling and Completion Condition Recognition Algorithm Based on CNN-GNN-LSTM Neural Networks and Applications," *Processes*, vol. 13, no. 4, p. 1090, Apr. 2025. https://doi.org/10.3390/pr13041090.

[31] L. C. Lamb, A. Garcez, M. Gori, M. Prates, P. Avelar, and M. Vardi, "Graph Neural Networks Meet Neural-Symbolic Computing: A Survey and Perspective," *arXiv preprint arXiv:2003.00330*, Jun. 2021. https://doi.org/10.48550/arXiv.2003.00330.

[32] S. Wu, F. Sun, W. Zhang, X. Xie, and B. Cui, "Graph Neural Networks in Recommender Systems: A Survey," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1–37, May 2023. https://doi.org/10.1145/3535101.

[33] S. A. Shahriar, Y. Choi, and R. Islam, "Advanced Deep Learning Approaches for Forecasting High-Resolution Fire Weather Index (FWI) over CONUS: Integration of GNN-LSTM, GNN-TCNN, and GNN-DeepAR," *Remote Sensing*, vol. 17, no. 3, p. 515, Feb. 2025. https://doi.org/10.3390/rs17030515.