

Hybrid CNN–RNN Model for Automated Student Engagement Analysis

Dr. Arman Petrosyan

Department of Artificial Intelligence and Robotics
Armenian Institute of Computational Science Yerevan, Armenia

Dr. Mariam Hakobyan

Center for Intelligent Data Systems
Caucasus AI Research Institute Yerevan, Armenia

VOLUME 03 ISSUE 08 (2026)

Published Date: 11 August 2026 // Page no.: - 35-40

ABSTRACT

Student engagement is a multidimensional construct encompassing behavioral, cognitive, and emotional participation in learning activities. In e-learning environments, identifying engagement continuously and objectively is challenging because conventional assessment methods depend heavily on self-reports, instructor observations, activity logs, or post-course outcomes. This study proposes a hybrid Convolutional Neural Network–Recurrent Neural Network (CNN–RNN) model for automated student engagement analysis in digital learning environments. The proposed architecture integrates CNN-based representation learning for extracting discriminative patterns from learner-related visual or interaction features with RNN-based temporal modeling for identifying changes in engagement over sequential learning activities. The conceptual framework is grounded in prior research on e-learning models, learning analytics, online engagement, cognitive engagement, and automated behavioral engagement detection. The methodology defines an engagement-processing pipeline consisting of data acquisition, preprocessing, feature representation, CNN-based spatial or local-pattern extraction, RNN-based temporal modeling, engagement classification, and analytical interpretation. The proposed approach is designed to distinguish persistent engagement patterns from short-term behavioral fluctuations. The analysis indicates that combining spatial feature extraction with temporal sequence modeling can provide a more comprehensive representation of student engagement than isolated static classification approaches. The study further identifies challenges associated with multimodal data quality, class imbalance, interpretability, privacy, and generalizability across different e-learning contexts. The resulting framework provides a technically grounded foundation for intelligent learning analytics systems capable of supporting early intervention and adaptive educational decision-making.

Keywords: Student Engagement, E-Learning, Hybrid Deep Learning, CNN–RNN, Learning Analytics, Behavioral Analysis, Educational Data Mining, Deep Neural Networks, Engagement Detection.

INTRODUCTION

1.1 Background

The expansion of digital technologies has transformed the delivery, monitoring, and evaluation of education. Contemporary e-learning platforms generate substantial quantities of learner interaction data, including participation records, discussion activity, navigation behavior, content access, and other digital traces. Digital technologies therefore create opportunities for moving beyond conventional end-of-course evaluation toward continuous analysis of learner behavior and engagement. Haleem et al. emphasize the increasingly important role of digital technologies in transforming educational practices and supporting data-informed learning environments (Haleem et al., 2022).

Student engagement is particularly important because participation in an online learning environment cannot be adequately represented by academic outcomes alone.

Learners may obtain similar grades while demonstrating substantially different levels of behavioral, emotional, or cognitive participation. Pilotti et al. identify cognitive, emotional, and behavioral dimensions as important components of engagement in asynchronous learning environments (Pilotti et al., 2017). Similarly, Redmond et al. conceptualize online engagement through a structured framework for higher education, indicating that engagement should be examined as a multidimensional phenomenon rather than as a single observable action (Redmond et al., 2018).

Learning analytics provides an important technical foundation for analyzing these dimensions. Johar et al. demonstrate the relevance of learning analytics for understanding student engagement and improving learning performance (Johar et al., 2023). However, conventional analytical approaches may have difficulty representing sequential behavioral changes. A learner's engagement at one point in time does not necessarily predict engagement several minutes or activities later.

1.2 Problem Statement

The principal problem addressed in this study is the automated identification of student engagement from sequential e-learning data. Traditional analytical methods frequently treat learner observations as independent records, potentially overlooking temporal dependencies. For example, repeated interaction with learning content may indicate sustained engagement, whereas the same number of interactions distributed irregularly across a course may represent a different behavioral pattern.

Kew and Tasir demonstrate the importance of analyzing cognitive engagement in e-learning discussion forums, where the content and nature of learner participation provide information beyond simple activity counts (Kew and Tasir, 2021). Okur et al. further demonstrate the feasibility of detecting behavioral engagement automatically, supporting the development of computational systems for engagement recognition (Okur et al., 2017). These studies motivate architectures capable of extracting complex features while preserving their temporal relationships.

1.3 Research Relevance and Objectives

The relevance of automated engagement analysis lies in its potential to provide timely evidence for instructional intervention. Alhothali et al. identify machine learning as an important analytical direction for predicting student outcomes in online courses (Alhothali et al., 2022). A system that can detect engagement changes before they become reflected in final academic outcomes could support earlier pedagogical responses.

The primary objective of this study is to develop a conceptual hybrid CNN-RNN architecture for automated student engagement analysis. The specific objectives are to establish a preprocessing and representation pipeline, extract engagement-related patterns through CNN layers, model sequential dependencies using RNN layers, classify engagement states, and evaluate the theoretical and practical implications of the hybrid architecture.

1.4 Scope and Significance

The study focuses on e-learning environments in which learner activity can be represented as sequential digital observations. The proposed architecture is applicable to interaction sequences, discussion-related features, behavioral traces, or other structured representations of learner activity. The significance of the approach lies in combining local feature extraction and temporal dependency modeling rather than relying exclusively on either static feature classification or sequential modeling.

A broader systems perspective is also relevant because intelligent educational infrastructures increasingly require coordinated analytical layers. Although outside the educational domain, the multi-layer coordination principle described by Geo Philip and Paulson illustrates how complex intelligent systems can integrate multiple analytical and operational layers rather than treating individual components independently (Geo Philip and Paulson, n.d.). In the present context, this principle

motivates a layered engagement-analysis architecture in which feature extraction, temporal modeling, classification, and interpretation operate as interconnected stages.

LITERATURE REVIEW

Research on student engagement in e-learning establishes that engagement is multidimensional and context-dependent. Pilotti et al. distinguish cognitive, emotional, and behavioral engagement, demonstrating why a single activity metric may provide an incomplete representation of learner participation (Pilotti et al., 2017). Redmond et al. similarly provide an online engagement framework for higher education, reinforcing the need for structured conceptualization of engagement rather than simplistic measurement (Redmond et al., 2018).

The development of learning analytics has created opportunities to operationalize these concepts through digital learner traces. Johar et al. systematically examine learning analytics approaches for student engagement and their relationship with learning performance (Johar et al., 2023). This direction is particularly relevant to automated engagement detection because computational models can transform large-scale behavioral observations into interpretable engagement indicators.

Kew and Tasir provide a more focused perspective by examining cognitive engagement in e-learning discussion forums through content analysis (Kew and Tasir, 2021). Their work indicates that engagement cannot always be inferred from the frequency of participation. The substance and characteristics of learner contributions can provide additional evidence about cognitive involvement. Consequently, a robust automated system should ideally capture multiple feature types rather than relying exclusively on raw interaction counts.

Okur et al. directly support the computational detection perspective through behavioral engagement detection in naturalistic educational settings (Okur et al., 2017). Their work is particularly relevant to the proposed model because it demonstrates that observable learner behavior can serve as an input for automated engagement recognition.

Alhothali et al. examine machine learning techniques for predicting student outcomes in online courses (Alhothali et al., 2022). Although outcome prediction differs from engagement detection, the study supports the broader proposition that machine learning can extract useful patterns from online learning data.

Blagoev et al. propose an e-learning model based on knowledge of learners, highlighting the importance of learner-specific information in the design of intelligent educational environments (Blagoev et al., 2021). This provides theoretical support for adaptive engagement analysis in which learner observations are interpreted within a broader learner model.

Haleem et al. describe the growing role of digital technologies in education, establishing the technological context in which automated learning analytics can operate (Haleem et al., 2022). Collectively, the literature demonstrates the importance of engagement analytics but also reveals a methodological gap: the integration of complex feature extraction with explicit temporal modeling remains a valuable

direction for automated engagement analysis.

The proposed hybrid CNN-RNN framework addresses this gap by combining two complementary analytical functions. CNN processing is used to transform structured learner representations into higher-level features, while RNN processing models how these features evolve over time. This layered analytical design is conceptually consistent with the broader principle of coordinating multiple intelligent processing layers described by Geo Philip and Paulson (n.d.), although the present study applies that principle specifically to educational analytics.

METHODOLOGY

3.1 Research Design

This study adopts a model-development research design focused on constructing a conceptual deep-learning framework for automated student engagement analysis. The methodology translates the multidimensional engagement concept into a computational pipeline:

Learner Data → Preprocessing → Feature Representation → CNN Feature Extraction → RNN Temporal Modeling → Engagement Classification → Analytical Interpretation

The architecture is designed for sequential e-learning observations. Each learner is represented by a sequence of observations collected across defined time intervals or learning events.

3.2 Data Representation

Let a learner sequence be represented as:

[
 $X = \{x_1, x_2, \dots, x_T\}$
]

where (x_t) represents the learner's feature vector at time (t), and (T) represents the number of observations.

Potential feature categories include interaction frequency, content-access behavior, discussion activity, task participation, response patterns, and other available learning-platform indicators. These variables should not automatically be interpreted as engagement labels. Instead, they constitute observable evidence from which an engagement model estimates latent engagement states.

A hypothetical learner sequence might contain low activity during the first three intervals, increased content interaction during subsequent intervals, and sustained participation during a discussion activity. A temporal model should distinguish this progression from a learner exhibiting the same total number of interactions but in a highly irregular sequence.

3.3 Preprocessing

Preprocessing consists of data cleaning, missing-value treatment, normalization, temporal alignment, and

sequence construction. Numerical features should be normalized to reduce differences in scale. Categorical features can be encoded into machine-readable representations.

Temporal segmentation is particularly important. If observations are excessively aggregated, short-term changes in engagement may disappear. Conversely, extremely fine-grained sequences may introduce noise and increase computational requirements.

Class imbalance should also be examined because highly engaged students may be more prevalent than disengaged learners, or vice versa. An imbalanced training distribution can cause a classifier to favor the majority engagement category.

3.4 CNN-Based Feature Extraction

The CNN component extracts higher-level representations from the structured learner input. Rather than treating each behavioral variable as an isolated predictor, convolutional operations can identify local combinations of features.

A generic convolution operation can be expressed as:

[

$$h_i = f(W * x_i + b)$$

]

where (W) represents learnable filters, (x_i) is the local input region, (b) is a bias term, and (f) represents a nonlinear activation function.

In the proposed framework, CNN layers identify local patterns such as combinations of content interaction, discussion activity, and task behavior. Pooling or dimensionality-reduction operations can then produce compact representations.

3.5 RNN-Based Temporal Modeling

The CNN-generated feature representations are passed to an RNN component. The RNN maintains a hidden representation of previous observations:

[

$$h_t = f(W_{xx}x_t + W_{hh}h_{t-1} + b)$$

]

where (h_t) represents the current hidden state and (h_{t-1}) represents historical information.

This mechanism enables the system to model changes in engagement. For example, a temporary decline in activity following a period of high participation should not necessarily be classified as sustained disengagement. The temporal context provides a basis for distinguishing transient deviations from persistent behavioral patterns.

3.6 Engagement Classification

The final representation can be passed through a dense

classification layer. For a binary classification task:

$$\begin{aligned} &[\\ &\hat{y} = \sigma(W_{ch_T} + b_c) \\ &] \end{aligned}$$

where ($\hat{y}$) represents the estimated engagement probability and (h_T) represents the final temporal representation.

For multiple engagement states, such as low, moderate, and high engagement, a softmax function can be used.

3.7 Evaluation Framework

Model performance should be assessed using accuracy, precision, recall, F1-score, and confusion matrices. Accuracy alone is insufficient when engagement classes are imbalanced. Precision measures the reliability of positive engagement predictions, while recall measures the ability to identify relevant engagement states. The F1-score balances these two properties.

A practical evaluation should additionally examine temporal robustness, cross-course generalization, and sensitivity to missing or noisy learner data. These considerations are important because an engagement model that performs well on one course may not necessarily generalize to another course with different teaching practices or student populations.

3.8 Interpretability and System-Level Integration

Automated engagement detection should not be treated as a purely predictive task. Instructors need evidence explaining why a learner was classified as disengaged. Therefore, feature-level analysis, temporal visualization, and confidence scores can accompany model predictions.

The layered organization of the proposed architecture also supports modular implementation. Similar to the coordination principle discussed by Geo Philip and Paulson (n.d.), separate processing layers can be connected into an integrated intelligent system. In educational applications, this means that data ingestion, representation learning, temporal analysis, and instructional decision support can remain functionally distinct while contributing to one analytical workflow.

RESULTS

The proposed hybrid architecture indicates that automated student engagement analysis can benefit from integrating feature extraction and temporal dependency modeling. The principal finding is conceptual: engagement should not be represented solely through isolated learner observations because meaningful engagement patterns often emerge from sequences of behavior. A CNN component addresses the feature-representation problem by identifying combinations of learner characteristics, while an RNN component addresses the temporal problem by preserving information about preceding observations.

The first expected outcome is improved representation of heterogeneous engagement signals. A learner may simultaneously demonstrate frequent content access, limited discussion participation, and irregular task completion. A conventional single-feature approach could interpret these signals inconsistently. CNN-based representation learning can identify combinations of features and transform them into higher-level representations.

The second finding concerns temporal sensitivity. Engagement is dynamic rather than fixed. The RNN component allows the model to distinguish a temporary reduction in activity from a sustained decline. This distinction is important for educational intervention because premature intervention based on one unusual observation may be inappropriate.

The third finding concerns alignment with learning analytics research. Johar et al. emphasize the role of learning analytics in examining engagement and supporting improved learning performance (Johar et al., 2023). The proposed architecture extends this analytical orientation by emphasizing sequential deep-learning representations. Similarly, the behavioral detection perspective of Okur et al. supports the feasibility of automated engagement recognition from observable behavior (Okur et al., 2017).

The fourth finding is that engagement should be treated as multidimensional. The framework can incorporate behavioral indicators while leaving room for cognitive and interaction-oriented features. This is consistent with Pilotti et al.'s treatment of cognitive, emotional, and behavioral engagement (Pilotti et al., 2017) and Kew and Tasir's emphasis on cognitive engagement in discussion environments (Kew and Tasir, 2021).

Finally, the architecture has potential value for early intervention. If sequential predictions identify a persistent downward engagement trajectory, an instructor or learning-support system could prioritize that learner for review. However, predictions should be interpreted as decision-support signals rather than definitive judgments about learner motivation. Blagoev et al.'s learner-oriented e-learning perspective reinforces the importance of contextualizing learner information within an intelligent educational environment (Blagoev et al., 2021).

DISCUSSION

The proposed CNN-RNN architecture contributes to the literature by connecting three important ideas: multidimensional engagement, automated learning analytics, and temporal deep learning. Existing research establishes that engagement contains behavioral and cognitive dimensions and that digital learning environments generate useful evidence for its analysis (Pilotti et al., 2017; Kew and Tasir, 2021). The proposed architecture provides a computational mechanism for combining such evidence into sequential representations.

A major theoretical implication is that engagement analysis should move from static classification toward dynamic state estimation. Redmond et al.'s engagement framework demonstrates the multidimensional nature of online engagement (Redmond et al., 2018), while the proposed model translates this conceptual complexity into an architecture capable of processing sequential observations. The model

therefore treats engagement as a changing condition rather than a permanent learner characteristic.

The practical implication is the possibility of earlier and more targeted instructional intervention. Alhothali et al. demonstrate the value of machine-learning approaches in analyzing online-course data and predicting student-related outcomes (Alhothali et al., 2022). An engagement-focused system can complement such approaches by providing an intermediate behavioral signal before final academic outcomes become available.

Nevertheless, several trade-offs must be recognized. CNN-RNN models can be computationally more demanding than conventional statistical or shallow machine-learning methods. They also require adequate training data and careful parameter selection. If training data are biased toward a particular course, institution, or learner population, the resulting model may perform poorly in other contexts.

Interpretability represents another limitation. A highly accurate prediction is not necessarily pedagogically useful if instructors cannot understand its basis. Because engagement is multidimensional, the model may also produce misleading conclusions if observable activity is treated as a direct substitute for motivation or cognitive investment. A learner with low platform activity may be deeply engaged in offline reflection, while a highly active learner may interact mechanically without meaningful cognitive involvement.

The literature further suggests that learner knowledge and context are important considerations. Blagoev et al. emphasize learner-oriented modeling in e-learning (Blagoev et al., 2021). Consequently, future implementations should incorporate contextual learner information rather than relying solely on universal engagement thresholds.

Privacy and responsible data governance are also essential. Automated engagement analysis involves behavioral information that may influence instructional decisions. Predictions should therefore be used to support human judgment rather than replace it. The multi-layer intelligent-system perspective offered by Geo Philip and Paulson (n.d.) is useful here as an architectural analogy: effective intelligent systems require coordination among analytical components rather than isolated prediction mechanisms.

Overall, the proposed architecture provides a strong methodological direction but should be empirically validated across multiple courses, learner populations, and engagement definitions before being deployed for high-impact educational decisions.

CONCLUSION

This study proposed a hybrid CNN-RNN model for automated student engagement analysis in e-learning environments. The framework combines CNN-based feature extraction with RNN-based temporal modeling to represent both local behavioral relationships and changes in engagement over time. Its theoretical foundation is

supported by research emphasizing multidimensional engagement, learning analytics, learner modeling, cognitive participation, and automated behavioral detection.

The principal contribution is a structured computational perspective in which engagement is modeled as a dynamic and multidimensional phenomenon rather than as a static activity score. The architecture can support continuous monitoring, engagement-state classification, and potentially earlier instructional intervention.

However, successful implementation requires careful treatment of data imbalance, learner diversity, privacy, interpretability, and cross-context generalization. Future research should empirically compare the proposed CNN-RNN architecture with alternative machine-learning and deep-learning models using standardized engagement datasets. Further work should also investigate multimodal engagement representation, explainable predictions, adaptive learner-specific modeling, and real-time integration with learning management systems. Such developments could strengthen the role of intelligent learning analytics as a decision-support mechanism for personalized and responsive e-learning.

REFERENCES

1. A. Alhothali, M. Albsisi, H. Assalahi, and T. Aldosemani, "Predicting student outcomes in online courses using machine learning techniques: A review," *Sustain.*, vol. 14, no. 10, pp. 6199, 2022.
2. I. Blagoev, G. Vassileva, and V. Monov, "A model for e-learning based on the knowledge of learners," *Cybernet. Inf. Technol.*, vol. 21, no. 2, pp. 121-135, 2021.
3. A. Haleem, M. Javaid, M. A. Qadri, and R. Suman, "Understanding the role of digital technologies in education: A review," *Sustain. Oper. Comput.*, vol. 3, no. 4, pp. 275-285, 2022.
4. N. A. Johar, S. N. Kew, Z. Tasir, and E. Koh, "Learning analytics on student engagement to enhance students' learning performance: A systematic review," *Sustain.*, vol. 15, no. 10, pp. 7849, 2023.
5. S. N. Kew and Z. Tasir, "Analysing students' cognitive engagement in e-learning discussion forums through content analysis," *Knowl. Manage. E-Learn.*, vol. 13, no. 1, pp. 39-57, 2021.
6. E. Okur, N. Alyuz, S. Aslan, U. Genc, C. Tanriover and A. A. Esme, "Behavioral engagement detection of students in the wild," in *Artif. Intell. Edu.: 18th Int. Conf.*, Wuhan, China, Jun. 28-Jul. 01, 2017, pp. 250-261.
7. M. Pilotti, S. Anderson, P. Hardy, P. Murphy, and P. Vincent, "Factors related to cognitive, emotional, and behavioral engagement in the online asynchronous classroom," *Int. J. Teach. Learn. High. Edu.*, vol. 29, no. 1, pp. 145-153, 2017.
8. P. Redmond, L. Abawi, A. Brown, R. Henderson, and A. Heffernan, "An online engagement framework for higher education," *Online Learn. J.*, vol. 22, no. 1, pp.

9. K. Ramamurthy, R. K. Konduru and N. Amanmadov, "EvoGraphCoder: An Evolutionary Graph-Reasoning Framework for Self-Adaptive Software Engineering," in IEEE Access, vol. 14, pp. 63063-63076, 2026, doi: 10.1109/ACCESS.2026.3686019.
10. Geo Philip, Paulson, Integrated Intelligent Building Energy Management: A Multi-LayerFramework for Renewable Energy, HVAC Optimization, and SmartElectrical Network Coordination. Available at SSRN: <https://ssrn.com/abstract=6993209> or <http://dx.doi.org/10.2139/ssrn.6993209>