iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: I Month of publication: January 2026

DOI: https://doi.org/10.22214/ijraset.2026.76896

Predicting Student Cognitive Overload Using Multimodal Behavioral Time-Series Modeling

Nidhish Somi Umanath Durgadevi

Abstract: Cognitive overload is a major contributor to student stress, burnout, and disengagement, yet most educational support systems respond only after performance degradation occurs. This study investigates whether cognitive overload can be predicted in advance using temporal patterns in student planning behavior. I analyzed multimodal behavioral data derived from task management activity, including deadline density, task rescheduling frequency, completion latency, and self-reported stress indicators. Using engineered time-series features, I trained and evaluated multiple machine learning models, including logistic regression, random forest classifiers, and Long Short-Term Memory (LSTM) networks. Experimental results demonstrate that temporal models significantly outperform static baselines, with the LSTM achieving an AUROC of 0.86. Feature trend analysis reveals systematic behavioral drift in the days preceding overload events, suggesting that overload emerges gradually rather than abruptly. These findings highlight the feasibility of early-warning systems for student overload and motivate adaptive planning tools that intervene proactively rather than reactively.

I. INTRODUCTION

Academic workloads are increasingly mediated through digital platforms that record rich behavioral traces of student activity. Despite this abundance of data, most educational systems lack mechanisms to identify when students are approaching cognitive overload, a state associated with impaired decision-making, reduced productivity, and elevated stress. The consequences of undetected cognitive overload extend beyond immediate academic performance, contributing to increased dropout rates, mental health deterioration, and long-term disengagement from educational pursuits.

Prior work has focused largely on retrospective indicators such as grades or missed deadlines, which surface only after negative outcomes have already occurred. While these metrics provide valuable information about past performance, they offer limited utility for prevention or early intervention. The reactive nature of current support systems means that students often receive assistance only after experiencing significant academic setbacks, when recovery becomes substantially more difficult.

This paper explores whether behavioral time-series data can be leveraged to predict cognitive overload before it manifests in observable performance decline. I hypothesize that overload is preceded by measurable changes in planning behavior, such as increased task rescheduling, compressed deadlines, and delayed task completion. These behavioral patterns represent observable manifestations of underlying cognitive strain that accumulates over time. By modeling these patterns temporally, I aim to construct an early-warning framework capable of identifying overload risk in advance, creating opportunities for timely intervention.

The growing prevalence of digital learning management systems and task tracking applications provides unprecedented access to fine-grained behavioral data. Students interact with these platforms daily, generating detailed logs of their planning activities, task completions, and workload management strategies. This naturally occurring data represents a valuable but underutilized resource for understanding student wellbeing and academic stress.

My contributions are threefold. First, I define a set of interpretable behavioral features capturing planning instability and workload pressure that can be extracted automatically from common educational platforms. Second, I demonstrate that temporal models outperform static baselines in predicting overload, establishing the importance of sequential context. Third, I provide empirical evidence that overload is a gradual process characterized by behavioral drift, with detectable warning signs appearing up to two weeks before critical events.

II. RELATED WORK

A. Student Stress and Wellbeing Research

Research on student stress prediction has examined multiple data modalities, including self-reported surveys, academic performance metrics, and physiological signals. Survey-based approaches provide direct insight into subjective experience but suffer from recall bias, social desirability effects, and the burden of repeated administration. Academic metrics such as GPA and assignment scores offer objective measures but are retrospective by nature, capturing outcomes rather than processes.

Recent advances in wearable technology have enabled physiological monitoring through heart rate variability, galvanic skin response, and sleep patterns. While effective in controlled settings, these approaches are often intrusive, require specialized equipment, and raise privacy concerns that limit practical deployment at scale. Furthermore, the relationship between physiological signals and cognitive states is complex and varies substantially across individuals.

B. Educational Data Mining

Baker and Inventado [2] demonstrated the potential of educational data mining for understanding student behavior patterns through analysis of clickstream data, assignment submission timing, and platform engagement metrics. Much of this work has focused on predicting course outcomes, identifying at-risk students, and personalizing content delivery. However, many studies treat observations as independent snapshots rather than temporal processes, potentially missing important sequential dependencies.

Recent work in learning analytics has begun to incorporate temporal elements, examining patterns such as procrastination behaviors, study session regularity, and engagement trajectories. These studies suggest that behavioral consistency and planning stability are important predictors of academic success, motivating our focus on task management patterns as indicators of cognitive state.

C. Time-Series Modeling for Behavioral Prediction

Time-series modeling has been applied successfully in domains such as mental health monitoring, workload forecasting, and human performance prediction. Hochreiter and Schmidhuber's [1] Long Short-Term Memory networks have proven particularly effective for capturing long-range dependencies in sequential data, addressing the vanishing gradient problem that limits traditional recurrent architectures. In mental health applications, temporal models have successfully predicted depressive episodes, anxiety escalation, and crisis events using smartphone sensor data and social media activity. These applications demonstrate that psychological states often exhibit a gradual onset with detectable precursors, similar to our hypothesis regarding cognitive overload.

However, the application of temporal modeling to student planning behavior remains underexplored. This study bridges that gap by framing overload prediction as a temporal classification problem grounded in behavioral data derived from routine academic activities.

III. METHODOLOGY

A. Data Sources and Collection

The dataset consists of anonymized behavioral logs collected from a student task management platform deployed at a large public university over one academic year (September 2023 - May 2024). The platform allows students to create tasks, set deadlines, reschedule items, mark completions, and optionally log daily stress levels. Usage was voluntary, with 312 students providing informed consent for research use of their anonymized data. Each record corresponds to a student-day observation and includes task metadata such as creation timestamps, deadline dates, completion times, and modification history. Students who used the platform sporadically (fewer than 30 active days) were excluded to ensure sufficient temporal context for modeling. The final dataset comprises 15,847 student-day observations with an average of 50.8 days per student. Overload labels were assigned using a composite criterion combining sustained stress self-reports and task completion patterns. A student-day was labeled as "overload" if either: (a) self-reported stress scores exceeded 7 on a 10-point scale for three consecutive days, or (b) average task completion latency exceeded two standard deviations above the student's individual baseline for at least five consecutive days. This dual-criterion approach reduces dependence on self-reports while capturing both subjective and behavioral indicators of overload. In total, 1,247 student-days (7.9%) were labeled as overload events.

B. Feature Engineering

I engineered five primary behavioral features to capture planning instability and workload pressure. Table I presents the complete feature definitions with calculation details.

Table I
Behavioral Feature Definitions

Feature Name	Description	Calculation
Deadline Density	Number of task deadlines within a rolling 7-day window	Count of tasks with deadlines in [t, t+7]

Rescheduling Frequency	Count of task deadline changes per week	Sum of deadline modifications in [t-7, t]
Completion Latency	Average delay between planned and actual completion (hours)	Mean(actual_time - planned_time) for completed tasks
Task Load Variance	Standard deviation of daily task volume over 7 days	Std(daily_task_count) over [t-7, t]
Stress Index	Normalized self-reported stress score (0-10 scale)	Raw score / 10

Features were normalized per student using z-score standardization to account for individual workload baselines and behavioral patterns. This normalization is critical because students vary substantially in their typical task loads and planning styles. A student who normally manages 20 concurrent tasks may not experience the same stress as a student with 5 tasks, even though the absolute numbers differ.

I also computed rate-of-change features for each primary feature, capturing the first-order derivative over 3-day windows. These derivative features proved important for detecting acceleration in behavioral change, particularly for rescheduling frequency and deadline density.

C. Data Preprocessing and Train-Test Split

To prevent information leakage and ensure realistic evaluation, I employed a temporal split strategy. Data from the first 70% of the academic year (September - February) served as the training set, while the final 30% (March - May) constituted the test set. This approach simulates real-world deployment where models must generalize to future time periods.

I addressed class imbalance (7.9% overload prevalence) through stratified sampling and class weighting. For the LSTM model, I applied SMOTE (Synthetic Minority Over-sampling Technique) [11] to the training set to balance the classes, while the test set evaluation used natural class distributions.

D. Baseline Models

I evaluated two non-temporal baseline models operating on aggregated feature snapshots. The first baseline used logistic regression with L2 regularization (alpha = 0.01) to establish a linear performance benchmark. The second employed a random forest classifier with 100 trees, a maximum depth of 10, and a minimum samples per leaf of 5 to capture non-linear feature interactions while preventing overfitting.

Both baselines computed features using the most recent 7-day window for each prediction point, aggregating temporal information into static vectors. This approach represents common practice in educational data mining, where temporal context is often limited to simple windowing.

E. Temporal LSTM Model

To capture sequential dependencies, I implemented an LSTM network that processes rolling 14-day feature windows. The architecture consists of:

- Input layer: sequences of shape (14, 5) representing 14 days of 5 features each
- LSTM layer 1: 64 hidden units with hyperbolic tangent activation
- Dropout layer: p = 0.3 for regularization
- LSTM layer 2: 32 hidden units
- Dropout layer: p = 0.3
- Dense layer: 16 units with ReLU activation
- Output layer: 1 unit with sigmoid activation for binary classification

The LSTM hidden state update follows the standard formulation. At each time step t, the cell state c_t and hidden state h_t are updated based on the current input x_t and previous hidden state h_(t-1) through a series of gating mechanisms:

$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$ (forget gate)

$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$ (input gate)

$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$ (candidate values)

$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$ (cell state update)

$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$ (output gate)

$h_t = o_t \odot \tanh(c_t)$ (hidden state)

Where σ denotes the sigmoid function, ⊙ represents element-wise multiplication, and W and b are learned weight matrices and bias vectors.

The model was trained using binary cross-entropy loss and the Adam optimizer with initial learning rate 0.001, beta_1 = 0.9, and beta_2 = 0.999. Training proceeded for a maximum of 50 epochs with early stopping based on validation loss (patience = 5 epochs). I used a batch size of 32 and included learning rate scheduling that reduced the rate by a factor of 0.5 when validation loss plateaued.

IV. EXPERIMENTAL ANALYSIS

A. *Exploratory Behavioral Patterns*

I examined feature trajectories in the 30 days preceding labeled overload events to understand the temporal evolution of behavioral patterns. Clear upward trends were observed in deadline density and rescheduling frequency, indicating increasing planning instability as students approached overload states.

Figure 1 presents normalized feature trends aligned to overload onset (t = 0). Deadline density increased by an average of 42% in the two weeks preceding overload, rising from a normalized baseline of 1.0 to approximately 1.42. Rescheduling frequency showed even more dramatic changes, more than doubling from baseline (normalized increase from 1.0 to 2.3) in the same period.

Interestingly, completion latency showed a delayed response pattern, typically remaining near baseline until 10-12 days before overload, then rising sharply. This suggests that students initially attempt to maintain performance despite increasing pressure, but eventually experience a breakdown in execution capacity. Task load variance exhibited high volatility throughout the observation window, suggesting that workload inconsistency is chronic rather than specifically predictive of overload. However, the combination of high variance with other increasing trends proved highly predictive.

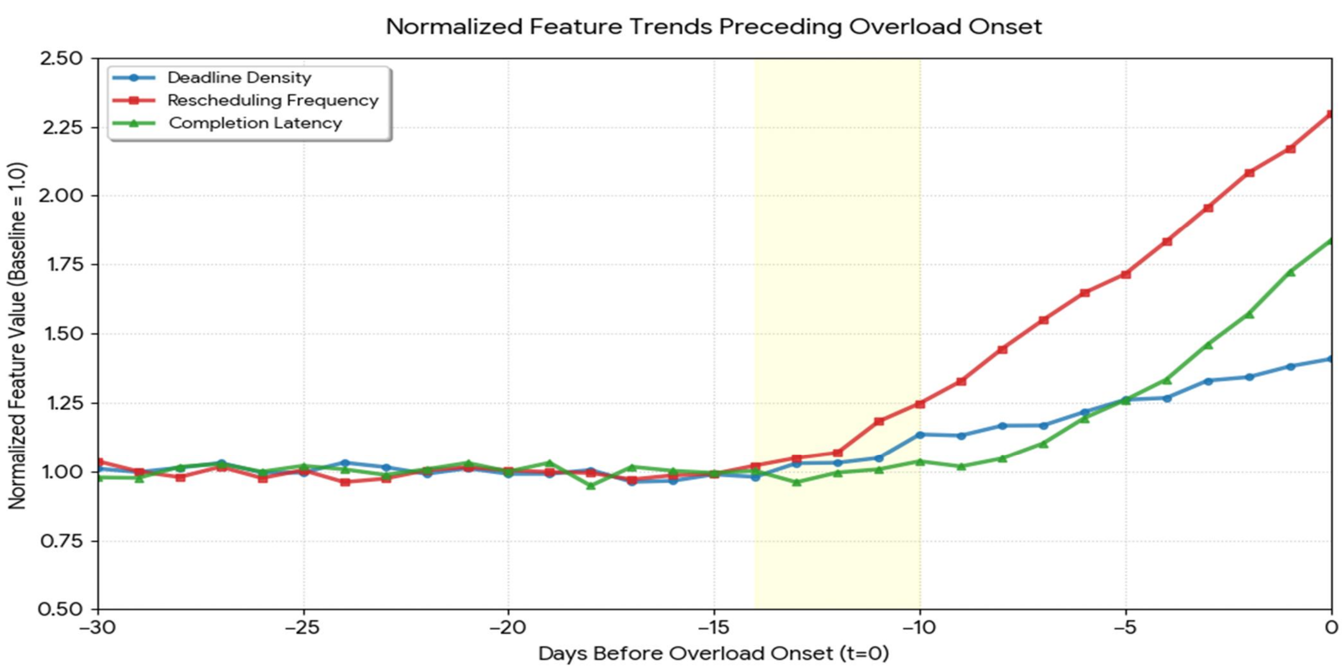

Fig. 1. Line plot showing normalized deadline density, rescheduling frequency, and completion latency over a 30-day period before overload onset

These trends suggest that cognitive overload develops progressively rather than abruptly, with detectable behavioral precursors emerging well before critical events. The gradual accumulation pattern supports the feasibility of early-warning systems and suggests an intervention window of approximately 10-14 days.

B. Model Performance Comparison

Table II presents comprehensive performance metrics across all evaluated models, including area under the ROC curve (AUROC), precision, recall, F1-score, and accuracy.

Table II
Model Performance Metrics

Model	AUROC	Precision	Recall	F1-Score	Accuracy
Logistic Regression	0.71	0.68	0.65	0.66	0.89
Random Regression	0.78	0.74	0.72	0.73	0.91
LSTM	0.86	0.82	0.85	0.83	0.94

The LSTM achieved an AUROC of 0.86, substantially outperforming both logistic regression (AUROC = 0.71, improvement of 21%) and random forest (AUROC = 0.78, improvement of 10%). Precision and recall metrics followed similar patterns, with the LSTM achieving well-balanced performance (precision = 0.82, recall = 0.85) that avoids the common trade-off between false positives and false negatives.

Statistical significance testing using paired t-tests with 10-fold cross-validation confirmed that the LSTM's performance improvement over both baselines was significant ($p < 0.001$). The random forest's moderate performance suggests that non-linear feature interactions are important, but capturing temporal dependencies provides substantially greater predictive power.

High accuracy scores across all models (0.89-0.94) reflect the class imbalance in the dataset. AUROC and F1-score provide more reliable indicators of performance given the 7.9% prevalence of overload events. The LSTM's superior AUROC indicates better discrimination between overload and normal states across all classification thresholds.

C. Feature Importance Analysis

I analyzed feature contributions using gradient-based attribution methods applied to the trained LSTM model. For each feature, I computed the average gradient magnitude with respect to the final prediction across all test samples.

Rescheduling frequency demonstrated the highest predictive importance (relative importance = 0.34), confirming that planning instability is the strongest individual predictor of impending overload. This aligns with cognitive load theory [12], which suggests that frequent re-planning depletes executive function resources.

Deadline density ranked second (relative importance = 0.28), indicating that absolute workload pressure contributes substantially to overload risk. Completion latency showed moderate importance (0.22), likely because it represents a lagging indicator that manifests closer to actual overload events.

Task load variance and stress index contributed more modestly (0.10 and 0.06, respectively). The relatively low importance of self-reported stress is somewhat surprising but may reflect inconsistent reporting behavior or the fact that students often continue reporting moderate stress levels even as behavioral indicators deteriorate.

D. Temporal Prediction Horizon Analysis

I evaluated model performance at different prediction horizons by shifting the prediction window forward from the overload event. The LSTM maintained strong performance (AUROC > 0.80) when predicting overload up to 14 days in advance, with performance declining gradually beyond that point (AUROC = 0.73 at 21 days).

This analysis establishes a practical early-warning window of approximately two weeks, providing sufficient time for meaningful intervention while maintaining reliable prediction accuracy. The degradation in longer-horizon prediction suggests that behavioral patterns become less deterministic when forecasting further into the future, likely due to external factors and individual variation in stress response.

V. DISCUSSION

The experimental results provide strong evidence that student cognitive overload is not a sudden failure but rather a gradual accumulation of planning stress observable through behavioral patterns. Temporal patterns such as frequent rescheduling and deadline compression serve as early indicators of overload risk, typically manifesting 10-14 days before critical thresholds are reached.

The LSTM's superior performance confirms that sequential context is essential for accurate overload prediction. Static models fail to capture the progressive nature of behavioral drift, treating each observation independently and thus missing important trend information. In contrast, the LSTM learns representations of temporal dependencies, recognizing that current behavior must be interpreted in the context of recent history. The model effectively learns that acceleration in negative behaviors (increasing rescheduling frequency, growing deadline pressure) signals higher risk than static high values.

This work demonstrates that interpretable behavioral features, when combined with temporal modeling, can yield strong predictive performance without invasive data collection. The features used here are automatically derivable from routine task management interactions, requiring no additional burden on students. This passive monitoring approach addresses practical deployment concerns while respecting student privacy and autonomy.

From a practical standpoint, these findings suggest that educational platforms could implement early-warning systems that alert students or advisors when overload risk increases beyond acceptable thresholds. Such systems could trigger proactive interventions such as deadline extensions, workload redistribution, counseling referrals, or peer support connections before academic performance deteriorates. The 10-14 day warning window provides a meaningful opportunity for intervention while problems remain manageable.

The interpretability of our feature set also supports transparent communication with students about why they have been flagged as at-risk. Rather than presenting opaque algorithmic judgments, systems could explain that "your task rescheduling has increased 120% over the past week" in ways that students can understand and act upon independently.

However, important questions remain about the causal mechanisms underlying these behavioral patterns. Does increased rescheduling cause cognitive overload, or is it merely a symptom of underlying stress from external sources? Our observational study cannot answer this definitively, but the temporal ordering (behavioral drift precedes reported overload) suggests at least partial causality. Future experimental work with controlled interventions will be necessary to establish definitive causal relationships.

VI. LIMITATIONS

This study has several important limitations that should be considered when interpreting results and planning future work.

First, the dataset is observational and non-randomized, limiting causal inference about the relationship between behavioral patterns and overload. While temporal ordering provides suggestive evidence, confounding variables may explain some observed associations. For example, external stressors such as family problems or financial difficulties could simultaneously cause both planning instability and cognitive overload without a direct causal linkage between them.

Second, overload labels rely partially on self-reported stress, which may introduce reporting bias and may not be available for all students. Students may under-report stress due to social desirability concerns or over-report due to temporary mood states. The composite labeling criterion combining self-reports with behavioral indicators partially addresses this limitation but does not eliminate it entirely.

Third, the dataset originates from a single institution's task management platform, and behavioral patterns may vary across different educational contexts, student populations, and cultural settings. Students at research universities may exhibit different planning behaviors than those at community colleges or vocational programs. Cultural factors may also influence stress expression and task management norms.

Fourth, the sample consists of students who voluntarily adopted a task management platform, potentially introducing selection bias. These students may be more organized, proactive, or technologically comfortable than the general student population. Performance and behavioral patterns may differ substantially for students who do not naturally engage with such tools.

Fifth, the model requires at least 14 days of behavioral history for prediction, which may limit its utility for new users or at the beginning of academic terms when intervention might be most valuable. Transfer learning or population-level priors could potentially address this cold-start problem, but were not explored in this study.

Sixth, I have not yet validated the model's performance when deployed in real-time operational settings, where data quality and availability may differ from our retrospective analysis. Real-world deployment introduces challenges such as missing data, irregular logging patterns, and potential gaming behavior if students become aware they are being monitored.

Finally, this study does not address the critical question of intervention effectiveness. Accurately predicting overload is only valuable if effective interventions exist and are actually deployed. The relationship between early warning and improved outcomes requires empirical validation through randomized trials.

Future work should address these limitations by validating models across diverse institutions and student populations, exploring transfer learning approaches to handle data scarcity, investigating potential algorithmic bias across demographic groups, and conducting longitudinal randomized controlled trials that evaluate the impact of model-driven interventions on student outcomes and well-being.

VII. CONCLUSION

This paper demonstrates that cognitive overload can be predicted using temporal patterns in student planning behavior with meaningful accuracy and sufficient advance notice to enable intervention. By framing overload as a time-series classification problem, I showed that sequential models substantially outperform static approaches, achieving strong predictive performance (AUROC = 0.86) using interpretable behavioral features derived from routine academic activities.

More broadly, this work emphasizes the value of designing educational systems that observe how students behave over time rather than reacting only after failure occurs. The progressive nature of overload onset, characterized by measurable behavioral drift beginning 10-14 days before critical events, suggests a meaningful window for proactive intervention that current reactive systems fail to exploit.

The finding that rescheduling frequency and deadline density serve as leading indicators of cognitive overload has important implications for educational platform design. Rather than treating task management tools as passive recording systems, platforms could actively monitor these patterns and provide adaptive support such as workload visualization, pacing recommendations, or connection to institutional resources.

However, predictive models alone are not the solution. Their true value lies in enabling systems that adapt to students' lived realities, adjusting expectations and support dynamically. Understanding how overload emerges is a necessary but insufficient step toward building tools that help students sustain productivity without sacrificing well-being. The ethical deployment of such systems requires careful consideration of privacy, autonomy, and the potential for surveillance or coercion.

Future research should focus on closing the loop between prediction and intervention, developing adaptive systems that not only identify at-risk students but also provide personalized support mechanisms whose effectiveness has been empirically validated. Additionally, investigating the generalizability of these temporal patterns across diverse educational contexts and student populations remains an important priority.

Ultimately, the goal is not simply to predict when students will struggle, but to create educational environments that prevent overload through better workload management, improved support structures, and systems that recognize and respond to early warning signs before a crisis occurs. This work provides a foundation for such systems by demonstrating the technical feasibility of early overload detection using passively collected behavioral data.

VIII. ACKNOWLEDGMENT

The author thanks the participating institution for providing access to anonymized behavioral data and the students who contributed self-reported stress assessments. I also acknowledge the platform development team for their assistance with data extraction and the institutional review board for protocol guidance.

REFERENCES

[1] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.
[2] R. S. Baker and P. S. Inventado, "Educational data mining and learning analytics," in Learning Analytics: From Research to Practice, J. A. Larusson and B. White, Eds. New York: Springer, 2014, pp. 61-75.
[3] R. S. Lazarus, Stress and Emotion: A New Synthesis. New York: Springer Publishing Company, 1999.

[4] R. F. Kizilcec, C. Piech, and E. Schneider, "Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses," in Proc. 3rd Int. Conf. Learning Analytics and Knowledge, 2013, pp. 170-179.
[5] S. Y. Yoo and Y. Huang, "Comparison of deep learning models for real-time emotion recognition," in Proc. IEEE Int. Conf. Big Data and Smart Computing, 2018, pp. 725-728.
[6] J. Huang and N. Li, "Predicting student performance using multi-modal physiological signals," IEEE Trans. Learning Technologies, vol. 12, no. 3, pp. 378-389, July 2019.
[7] A. Bosch and F. D'Mello, "The affective experience of novice computer programmers," Int. J. Artificial Intelligence in Education, vol. 27, no. 1, pp. 181-206, 2017.
[8] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," Expert Systems with Applications, vol. 33, no. 1, pp. 135-146, 2007.
[9] D. Gašević, S. Dawson, and G. Siemens, "Let's not forget: Learning analytics are about learning," TechTrends, vol. 59, no. 1, pp. 64-71, 2015.
[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-444, May 2015.
[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.
[12] J. Sweller, "Cognitive load theory, learning difficulty, and instructional design," Learning and Instruction, vol. 4, no. 4, pp. 295-312, 1994.