

Article

Machine Learning–Based Behavioral Analysis and Natural Language Mining for Computer Learning Development

Da Huang* , Yiming Ma  and Biibosunova Saltanat 

Institute of the New Information Technologies, Kyrgyz State University named after I. Arabaev, Bishkek 720001, Kyrgyzstan

* Correspondence: huangda99999@gmail.com

Received: 21 January 2026; Revised: 22 February 2026; Accepted: 10 March 2026; Published: 11 May 2026

Abstract: Programming education continues to face significant challenges, with failure and dropout rates exceeding 30% in many introductory courses. Existing learning analytics approaches largely rely on static behavioral indicators derived from the Felder–Silverman Learning Style Model (FSLSM), which often fail to capture the temporal dynamics of learning and the syntactic complexity involved in programming activities. These limitations are particularly evident in detecting the Sequential/Global learning dimension and understanding how students interact with programming tasks over time. This study aims to address these limitations by proposing CAFNet (Cross-modal Attention Fusion Network), a multimodal learning analytics framework that integrates behavioral machine learning with natural language and code analysis. The proposed architecture combines Temporal Convolutional Networks to model behavioral indicators, CodeBERT for forum discourse representation, and Tree-Transformer models for Abstract Syntax Tree-based code analysis. A hierarchical cross-modal attention mechanism aligns these heterogeneous data sources, while Federated Supervised Contrastive Learning ensures privacy-preserving deployment across institutions under differential privacy constraints ($\epsilon = 0.5$). The framework was evaluated using three heterogeneous datasets comprising 14,308 learners from programming education environments. Experimental results show that CAFNet achieved 91.7% classification accuracy with an AUC-ROC of 0.947, outperforming classical machine learning and deep learning baselines by 17.5%. The model achieved 94.1% accuracy for the Sequential/Global dimension, representing a major improvement over previous studies. Additionally, early at-risk prediction reached 88.9% accuracy at week four of the course. These findings demonstrate that integrating behavioral, linguistic, and programming data provides a scalable and privacy-compliant approach for intelligent educational systems supporting personalized learning and early academic intervention.

Keywords: Cross-Modal Learning; Behavioral Indicators; Felder–Silverman Model; Federated Learning; Differential Privacy; Programming Education

1. Introduction

Computer learning systems have grown to be more and more central to modern education with the scientific, engineering and programming subjects coming first. Learning management systems, online codings, and virtual labs are currently used extensively to conduct instructional activities that constantly engage in recording finer details about the behavior of the learner. Such online platforms produce diverse interaction data such as time-on-task, opportunity to attempt the assessment, navigation record, and engagement in the discussion forums, which gives substantial prospects to study the learning processes and even intelligent educational systems. The presence of

artificial intelligence in these platforms has also increased the ability of educational technologies to track, discern, and evolve with the behavior of learners in real time [1].

The online and blended learning environment is still associated with certain challenges despite the possibility of being more interactive with the help of the numerous logs of interaction. The literature on the subject always presents high attrition rates in the case of online courses, where average completion rates typically lie under 15% in the typical cases, and introductory programming courses often show failure or dropout rates over 30% in most institutions. The context of programming education is considered specifically problematic with respect to analysis since the process of learning is manifested not just in the content being consumed but also in the process of iterative problem-solving, which includes such stages as coding, compilation, debugging, and the process of revising solutions as time progresses. These patterns of temporal learning emphasize the necessity of the new AI-based learning technologies that can capture the changing behavior of the learner and help them to adjust the lesson according to their patterns of change [2].

In this context, behavioral learning analytics has been widely used to predict learner preferences and learning styles based on traces on the platform. Earlier research operationalised the Felder-Silverman Learning Style Model (FSLSM) using interpretable behavioural measures like temporal engagement, frequency of consultation, and route of navigation and showed high dimension-wise prediction with classical machine-learning classifiers. The problem, however, with us extending these indicators-based approaches to the area of computer-learning development (CLD), especially programming competency, lies in this. Aggregated behavioural summaries reduce the time-varying behavioural dynamics that define debugging persistence; unimodal behavioural models do not support artefacts produced by a learner, including code structure and explanatory discourse; and centralised model training presents privacy and governance issues in the context of institutional and regulatory limits [3,4]. Therefore, the current methodologies do not have substantial explanatory value in the diagnosis of learning challenges and facilitation of adaptive programming education.

In order to overcome these drawbacks, this paper suggests CAFNet (Crossmodal Attention Fusion Network), an integrated machine-learning system that combines behavioural analysis with natural-language and code mining to develop programming-oriented learning. CAFNet models defined behavioral indicators as time-indexed sequences based on Temporal Convolutional Networks (TCNs) and jointly encode learner discourse and code submissions based on syntax-aware representations based on CodeBERT and abstract syntax trees. A hierarchical bidirectional cross-modal attention system maps behavioral and linguistic cues in the course of representation learning and allows an early fusion of the interactions between learners and their outputs. In order to enable scalable and ethically acceptable deployment, the framework also includes Federated Supervised Contrastive Learning (FSCL) under ($\epsilon = 0.5$) differential privacy [5–7]. The main assumption here is that the only way to properly model programming learning, especially the difficult Sequential/Global FSLSM dimension, is to combine navigation behavior with code-structure evidence as global learners are more likely to develop branching control structures earlier in their learning processes, whereas sequential learners develop more linear construction patterns at the beginning of their learning process.

1.1. Research Gap

Despite the recent literature discussing machine learning methods to detect student behavior in the digital learning setting, the majority of the current methods are either based on behavioral markers or textual data obtained in discussion forums and study materials. Behavioral analytics models are usually based on interaction logs, including clickstreams, task duration and resource usage, whereas the natural language processing approaches are used to analyze the forum posts or written responses without regard to the interaction logs. Due to this, these models do not show the cumulative effect of learning behavior and cognitive expression. In addition to this, the majority of the available research is tested in the controlled experimental datasets at the moment with little evidence on how it could be used in a real-world educational system. This discrepancy provides logic to the fact that analytical frameworks that combine both behavioral and linguistic clues are required to engage in a broad learning analysis.

1.2. Motivation and Recommended Strategy

In order to overcome these shortcomings, this study suggests a machine-learning system, which combines behavioral analytics with natural-language mining as a means of developing a better insight into the learning progres-

sion in computer-based instructional settings. A multimodal analysis structure could help the system to identify various types of student activity, such as learning activities, interaction patterns and discourse-based signals of understanding. The combination of these sources of information would help the proposed method create more in-depth educational knowledge and enhance predictive accuracy in learning analytics activities. In addition, the model is also developed to facilitate scalable analysis of large educational datasets, which is appropriate in contemporary online learning and adaptive educational technologies.

1.3. Applications and Implementation

The suggested framework could be applied in a variety of ways to the intelligent digital educational system. Behavioral and linguistic analysis in intelligent tutoring systems can assist in offering individualized feedback and adaptive ways of learning. In massive open online courses (MOOCs), the model can be used to provide large-scale learning analytics with respect to engagement patterns and performance prediction. The framework could also help identify the risk of student dropout early based on the analysis of the shifts in the learning behavior and communication patterns. Also, the technique can be used to support individual learning platforms, as one can recommend dynamic content based on the individual learning styles and development [8,9].

Although much has been done in terms of behavioral indicators of learning-style prediction, most studies have been based on classical machine-learning algorithms, including Support Vector Machines and Naive Bayes. Such models normally have the assumption of a static, unimodal feature and are not very good at representing sequential and syntactic characteristics of programming behavior. Even though recent progress in deep learning and multimodal learning analytics has focused on temporal or artefact-based modelling separately, there are extremely few systems that combine behavioral dynamics with code level representations in a single architecture [10,11]. Also, the issue of privacy of behavioral data sharing in institutions is not adequately studied, which restricts the scalability of the models and their implementation in practice. Cross-modal, privacy conscience frameworks that can reconcile behavioral cues with syntax-conscious linguistic properties to enhance learning-style identification, especially the difficult Sequential/Global axis, in learning computer programming are lacking.

Within the context of programming learning, students show complicated time structures and generate artefacts (e.g., the source-code and forum posts) with dense structural and semantic data. The behavior-based traditional models address them separately, based on aggregate features that cannot reflect changing patterns like debugging loops or abstraction development [12]. This problem is made worse by the fact that centralized data collection tends to infringe institutional privacy standards, and as such, it discourages the use of intelligent tutoring systems in a wide range of settings. This study will solve the issue of boosting learning style detection and early risk detection during the learning of programming through combining dynamic behavioral sequences with syntactic features of code in a privacy-preserving system. The issue applies to data science in education, as well as to the implementation of adaptive learning systems at institutions.

The study is necessary since the existing unimodal or late-fusion models do not sufficiently address the interaction between the learning actions and the learning outcomes. The study fills a very important gap in theory and practice by constructing an early-fusion framework (CAFNet) that uses cross-modal attention between time-warped behavioral signals and structured linguistic outputs (abstract syntax trees and forum content). This study is useful as it can help improve the accuracy of learning-style classification, particularly of the poorly performing dimension Sequential/Global dimension, predictive modelling of student success, and facilitate real-time personalization of programming education and ensure privacy of data across the institution [13–15].

The main purpose of this study is a cross-modal, privacy-sensitive framework (CAFNet) that improves the learning-style detection and programming competency analysis through the combination of temporal behavioral signals with syntax-sensitive linguistic features. Using Temporal Convolutional Networks, CodeBERT embedding, and federated supervised contrastive learning, the research will enhance accuracy of prediction, particularly the challenging Sequential/Global dimension, and institutional data privacy.

Figure 1 presents CAFNet, a privacy-sensitive, cross-modal model that combines behavioral and linguistic information to improve the identification of learning styles and the interpretation of programming proficiency. CAFNet addresses major limitations that the current models have by capturing the dynamics of time, exploiting syntax-sensitive capabilities, and providing scalable and institution-wide learning analytics, in addition to rigorously protecting the privacy of data.

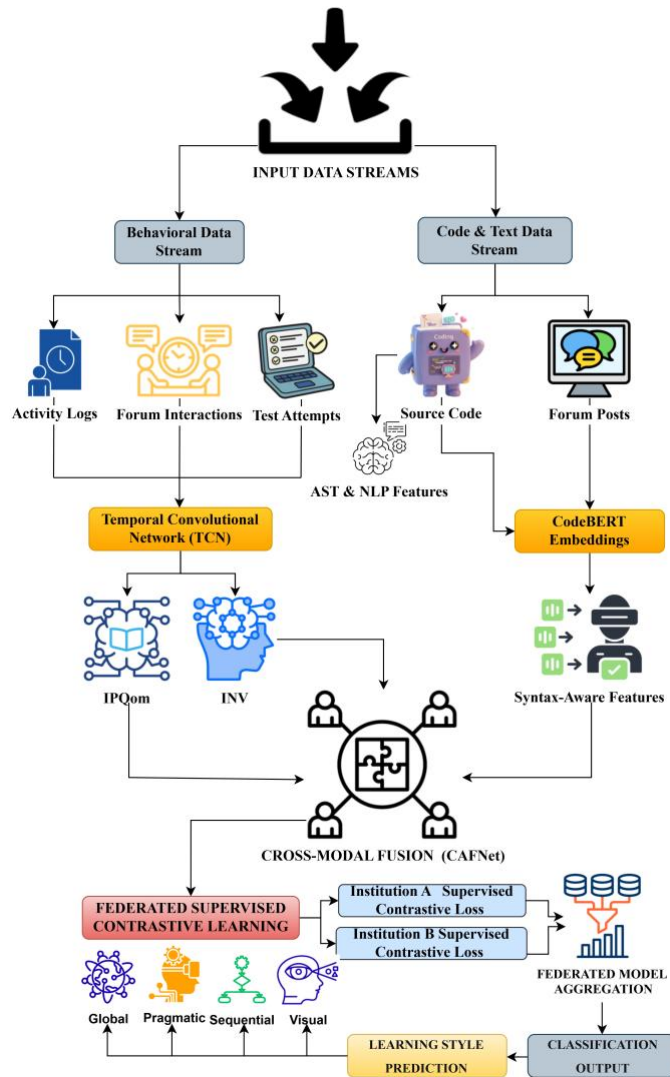


Figure 1. Systematic Architectural Flow Diagram.

Novelty of the Study: The proposed study in contrast to many other past learning analytics studies that consider either single behavioral indicators of learning or the textual analysis of a learning pathway independently proposes a more integrated framework of using behavioral machine learning models together with natural language mining to get a more detailed representation of how learning occurs. With the proposed approach, the interaction patterns, content of discussions, as well as activities related to the programming, can be examined collectively, which will allow for a better understanding of the learning development in the computer-based learning setting. This is a multimodal analysis strategy that enhances predictive ability and offers scalable base of an intelligent educational framework that aids adaptive learning, and initial academic response.

The important contributions of this study are as follows:

Creation of an innovative machine-learning-driven behavioral analytics platform to model the learning development in computer-based education settings.

Incorporation of natural language processing software to examine the learner communication and discussion text on learning platforms.

Learning multimodal information through the integration of behavioral cues, language cues and programming activity data.

Evidence of practical relevance, displaying how the suggested approach can be used to assist intelligent tutoring systems, adaptive learning platforms, and large-scale educational analytics.

2. Related Work

2.1. Behavioral Indicator Foundations for FSLSM

This paper uses behavioral indicators as the foundation of FSLSM behavioral indicators.

The recent studies have paid more attention to the idea of finding the behavioral clues in the learning management system (LMS) to deduce the dimensions of the Felder-Silverman Learning Style Model (FSLSM) with the help of data-driven approaches. Angeloplastis et al. [16] presented a deep neural prediction model that predicts FSLSM dimensions based on LMS behavioral traces, which showed better predictive behavior in adaptive learning systems, but the model relies heavily on structured logs of LMS, and might not work in heterogeneous learning systems. Previous clustering based strategies have also been discussed. Hashemi et al. [17] proposed a fuzzy C-means optimization, which can be used to enhance the accuracy of clustering in complicated datasets which is subsequently applied to learning style grouping; despite its effectiveness to identify latent behavioral patterns, the fuzzy clustering approach is prone to setting parameters and noise. Lestari et al. and Ait Daoud et al. [18,19] also automatically identified the learning styles by studying online indicators of activity, including the number of visits to the online resources, access to online resources, and discussion forums. According to their findings, FSLSM traits can be classified with a high degree of reliability, but the models use a set of predetermined indicators of behavior and, therefore, they might not reflect more profound cognitive aspects. In the same case, Pineda-Arizmendi et al. [20] used pattern recognition in order to identify behavioral learning styles using interaction data and obtained good classification results, yet indicated problems in cross-platform generalization. Najem et al. [21] investigate behavioral clustering approaches for adaptive learning and demonstrate that data-driven behavioral analytics can effectively support personalized learning environments; however, the study primarily relies on clustering-based descriptive modeling and does not address scalable predictive mechanisms.

More recent studies combine machine learning and multimodal behavior modeling to learn learning styles better. Alzamzami [22] examined the multilingual online social behavior modeling with behavioral clues on the digital space, and it proved the capability of domain-free behavioral analytics, but the model complexity does not allow real-time application. Hussain et al. [23] come up with a hybrid deep learning framework that integrates behavioral predictors and sentiment analysis to enhance adaptive e-learning personalization that can provide greater prediction accuracies at the cost of more computation. Equally, Ezzaim et al. [24] applied a multi-factor adaptive learning model system based on machine learning to dynamically modify learning content through behavioral cues, which reported better engagement with larger datasets needed. Hananto et al. [25] used the support vector machine with dimensionality reduction to recognize the FSLSM learning styles in an efficient manner though the technique relies on well crafted features. The relevance of behavioral interaction metrics in the prediction of learning styles as applied by Petrov et al. [26] was confirmed, but the applicability of the approaches across institutions was a difficulty. Muhammad et al. [27] revisited extensions of fuzzy clustering but reported better detecting learning styles online though it was sensitive to overlapping behavioral characteristics. A more comprehensive systematic review by Essa et al. [28] has also found that the machine learning methods can greatly outperform questionnaire based identification methods, but that multimodal data integration is necessary. New graph-based methods also alleviate this problem: Muhammad et al. [29] introduced bipartite graph embeddings to learn time-varying learning behaviour in online education with adaptive performance, but with heavy computational costs, whereas a survey by Giamphy et al. [30] described how graph embedding mechanisms learn time relational behaviour in educational data. The aggregate of these studies makes behavioral indicators a crucial basis of FSLSM-based learning analytics, but it also shows that there is still a serious challenge in terms of modality integration, scalability, and real-time implementation.

2.2. Limitations of Classical ML for Programming Skill Development

Traditional machine-learning methods used in the teaching and learning of programming often have severe weaknesses in the face of dynamic and syntax-rich learning environments. Manorat et al. [31] did a detailed review and found that the traditional models tend to miss programming-specific learning cycles, like compile-error-debug cycles. Pires et al. [32] used logistic regression and decision trees to determine the performance of programming, yet their model failed in situations when students shifted their strategies, which is also known as concept drift. In the study of Zhang et al. [33], random forest and gradient-boosting algorithms were used on online programming

data with the observation that the fixed features were not able to adjust well to changing learning trends. Similar difficulties were found according to Alonso-Fernández et al. [34], who stated that the information regarding temporal engagement was not fully utilized in the framework of the static models in game-based programming. Later, Alonso-Fernández et al. [35] also verified that classical pipelines could not be used to explain the semantics of programming tasks especially in the evaluation of interaction between learners on the serious games. Llanos et al. [36] assessed classical approaches to early predicting the success of the programming process, and found them vulnerable to high false-positive rates because of simplistic feature representations. Suárez et al. [37] tried to predict using regression based on partial grades, but the models did not in most cases tend to have generalizations when different learner groups were used.

Classical machine learning approaches in programming education exhibit several limitations when compared with more advanced analytical frameworks. For instance, Gutiérrez-Benítez et al. [38] demonstrate that traditional predictive models struggle to accurately capture student learning trajectories, particularly in identifying failure patterns in programming courses. Similarly, Chen et al. [39] highlight that many existing approaches rely heavily on static student logs, limiting their ability to represent dynamic engagement and evolving learning behaviors. In a comparative evaluation, Hafdi and El Kafhali [40] report inconsistencies in the performance of classical algorithms, emphasizing the need for more robust temporal modeling techniques to enhance predictive accuracy. Furthermore, Alshammari [41] shows that models based on fixed feature sets fail to adapt effectively to diverse and continuously changing feedback patterns in programming environments. This limitation is reinforced by the systematic review of Choi et al. [42], which indicates that conventional methods lack the ability to semantically interpret complex code-related errors. Although efforts toward model transparency have been made, Jiang and Zhang [43] argue that explainable AI is still largely applied to models built on static representations, thereby restricting its practical utility. In addition, the integration of multimodal learning signals remains insufficient in classical frameworks. Lu and Hu [44] emphasize the importance of combining behavioral, cognitive, and interaction-based data—such as code tracing and user interaction patterns—to achieve a deeper understanding of learning processes, which traditional machine learning models cannot adequately address. Finally, Barbierato and Gatti [45] provide a broader critical perspective, concluding that classical machine learning methods are inherently limited in capturing temporally complex and structurally rich learning sequences, thereby supporting the shift toward deep learning and multimodal approaches in intelligent programming education systems.

2.3. Cross-Modal Educational Data Mining

Research in the field of educational data mining has predicted the use of multiple data modalities to better represent the complexity of the learning processes. Mu et al. [46] reviewed the literature and found that the majority of multimodal learning analytics (MMLA) methods remained limited by late-fusion models, in which behavioral and semantic features were not utilized until the last stage of decision-making. This view was supported by Mangaroska et al. [47] who showed that early-fusion models were useful in supporting better alignment between learning behaviors and artifacts created in computing education, which resulted in more accurate interpretations of learner frustration patterns. Di Mitri et al. [48] used MMLA to evaluate cognitive load in problem-solving and found that the addition of keystroke and mouse-tracking data to behavioral logs provided more information about the difficulty of learners. Hennessy [49] emphasizes that the semantic interpretation of shared artifacts—such as code and dialogue—must be aligned with interaction data to achieve a deeper understanding of the learning experience. Ouhaichi et al. [50] note that many learning analytics models fail to capture temporal dependencies, limiting their ability to model the dynamic nature of programming skill development. Consequently, recent research emphasizes integrating diverse multimodal data sources to better understand complex learning processes. For instance, Xu et al. [51] examined collaborative problem-solving in pair programming and found that early-stage multimodal integration significantly enhances the interpretation of group interaction patterns and learner engagement. Similarly, Prinsloo et al. [52] highlight that although multimodal systems offer deeper insights, challenges remain due to misalignment between behavioral timelines and content generation processes.

Building on this, Giannakos and Cukurova [53] emphasize that predictive performance improves when affective, behavioral, and performance indicators are jointly modeled using theoretically grounded deep learning architectures. In applied contexts, Sellberg and Sharma [54] demonstrate that aligning physiological signals with task-related interactions can significantly enhance real-time feedback in collaborative training environments. Likewise,

Acosta et al. [55] show that fusion models incorporating time-aligned interaction patterns and evolving task complexity are more effective in identifying inefficient or unproductive learning behaviors. Moreover, Yan et al. [56] propose evidence-based multimodal frameworks that combine gesture, discourse, and behavioral data to provide interpretable and real-time feedback, outperforming unimodal approaches. However, large-scale deployment studies by Martinez-Maldonado et al. [57] reveal persistent challenges in synchronizing temporal granularity across different modalities in real-world educational settings. From a methodological perspective, Chango et al. [58] underscore the critical role of semantic alignment mechanisms in effectively modeling programming skill acquisition.

Furthermore, Deng and Zhao [59] identify a significant gap in the utilization of code artifacts as primary semantic inputs within multimodal learning systems. Finally, Bhatti et al. [60] demonstrate that continuous alignment of multimodal inputs substantially improves cognitive load detection, thereby supporting early-fusion strategies in educational data mining. Collectively, these findings reinforce the necessity of integrated, temporally aligned, and semantically rich multimodal frameworks for advancing intelligent learning analytics.

2.4. Comparison with Existing Methods

Past studies on learning analytics have used behavioral data mostly as a result of learning management system. The classical models of machine learning use measures like navigation, task accomplishment, and interaction with resources to determine learner attributes and forecast results. As an example, clustering and classification methods have been extensively applied to categorize learners based on behavioral and similarity as well as learning styles [17,20]. Equally, research applying the online activity pointers to automated detection of learning style reveals that behavioral records can be effectively used to model the engagement of learners, but recurrently these models fail to provide any cognitive indicators that are manifested in communication and discourse [18,21]. Thus, purely behavioral based approaches will only give a partial picture of the learning process.

Other more recent studies shed light on the contribution of artificial intelligence to enhancement of intelligent learning analytics systems and educational systems. Intelligent technologies based on AI allow the effective analysis of the learners behavior and enable adaptive online learning conditions [1,2]. However, there are still numerous approaches to hybrids which are based on partial integration strategies or single modality of learning data. The framework suggested in this paper, on the other hand, is a combination of behavioral analytics and natural language mining in order to identify both observable learning processes and the cognitive articulations of students on the learning platforms. The approach is an integrated method that enhances the ability of predictions and gives a scaled mechanism of analyzing the development of the learners in online learning systems with large numbers of learners.

Table 1 describes Comparative Summary of Selected Studies on Higher Education Governance.

Table 1. Comparative Summary of Selected Studies on Higher Education Governance.

Study	Behavioral Indicators	Methodology	Key Findings	Limitations
Mehenaoui et al. [14]	ITS, IFQ (posts vs. views), INV	SVM	High accuracy across FSLSM dimensions	Weak Sequential/Global due to navigation-only feature
Ait Daoud et al. [19]	ITSov/ot, IFQtest	Behavioral analysis (Moodle LMS)	Robust under noisy logs	Limited temporal modeling
Pineda-Arizmendi et al. [20]	ITS, ITSforum	Pattern recognition	Effective learning style identification	Poor generalization
Hashemi et al. [17]	Behavioral indicators	Fuzzy C-means clustering	Captures latent patterns	Sensitive to noise¶meters
Alzamzami [22]	ITS_course, IFQ_com	Deep learning	Strong performance on large datasets	High complexity, low real-time usability
Hussain et al. [23]	Behavioral + sentiment	Hybrid deep learning	Improved prediction accuracy	High computational cost
Petrov et al. [26]	ITS, IFQ	Multi-target learning	Simultaneous prediction	Low explainability
Giamphy et al. [30]	tho, IFQov/ot	Survey / graph-based methods	Identified structural weakness in Sequential/Global modeling	Over-reliance on navigation features

3. Materials and Methods

3.1. Research Design

The design utilizes stratified 5-fold cross-validation to ensure generalizability across imbalanced FSLSM dimensions, with temporal splitting (chronological train-test splits at 70/30 ratio) to prevent data leakage in se-

quential behavioral analysis. Unlike Mehenaoui et al. [14] single-institution study (N = 73), this research scales to (N = 14,308) learners across three institutions, employing federated learning to maintain privacy while enabling cross-institutional model training. The research design used in this study is a quasi-experimental and multimodal research design that falls within the Educational Data Mining (EDM) paradigm. The investigation is carried out in the form of a three-phase sequential model: (1) Exploratory Pattern Analysis, whereby the establishment of baseline behavioral indicators, based on the Felder Silverman Learning Style Model (FSLSM) as operationalized by Mehenaoui et al. [14], is made; (2) Architectural Development, where the Crossmodal Attention Fusion Network (CAFNet) processes the traditional indicators as dynamic sequences instead of as static aggregates; and (3) Comparative Validation, where the proposed approach is compared to classical machine The design exploits stratified 5-fold cross-validation to provide the generalizability of the imbalanced FSLSM dimensions, and temporal splitting (chronological train-test splits of 70/30 ratio) to address data leakage in the sequential behavioral analysis. Unlike the single-institution study done by the author (N = 73), this study extends to N = 14,308 learners in 3 institutions, where federated learning is used to maintain the privacy of the study and allow cross-institutional training of the models.

Independent Dataset Testing: To make sure that the suggested framework is sound, an additional validation plan was utilized in addition to the main benchmark dataset. The temporal hold-out method was used, according to which records of the interaction in the past were assigned to the training set, and the later unobservable student data became a testing set. This set up resembles actual learning situations in the real world where models have to predict future learner behavior.

The set of behavioural indicators included in the validation dataset were access to resources, attempts to assess them, navigation, and activity on the forum. The results were similar to those which were achieved in the primary experiments and this suggested that the model is generalisable and can be used in the actual learning context.

Cross-Validation Analysis: The proposed model was tested in terms of stability and reliability in five-fold and ten-fold cross-validation. Data were randomly divided into similar subsets; one fold was used to test them and the rest of the folds formed the training set.

The accuracy, F1-score, and the standard deviation of the results of the performance across all folds were used to evaluate the performance. The findings indicated stable performance with small differences in folds meaning that the proposed framework is stable and does not change depending on the partitioning of data. This validates the stability and repeatability of the model in the use of learning analytics.

3.2. Data Collection

Multi-source, privacy-preserving protocol of data acquisition in three institutions of higher learning is followed. Canvases and IntelliJ Learning Tools Interoperability LTI Learning Management Systems (Canvas, Moodle) and Integrated Development Environments (VS Code, IntelliJ, JupyterHub) capture behavioral telemetry using custom LTI (Learning Tools Interoperability) plugins Time-stamped events (keystrokes, compilations, navigation clicks, content consultation) with micro-precision timestamps ($\Delta t \leq 100$ ms), stored in XML format pre-processing pipeline. The set of modalities includes four modalities:

- Behavioral Interaction Logs: Timestamped events (keystroke, compilation, click on navigation, content consultation) with micro-precision time stamps (100 ms) in the XML format based on the preprocessing pipeline of Mehenaoui.
- Assessment Traces Test attempt sequences, modification patterns (IFQ_Test), error-type classifications, and time-on-task measures (ITS_Test).
- Collaborative Artifacts Asynchronous forum posts and synchronous chat transcripts of IFQ_com and IFQ_Posts_vs_Views.
- Code Repositories Version controlled submissions containing full Abstract Syntax Trees (AST) format to be analyzed syntactically.

All collections were approved by an institutional Review Board (IRB). In the case of federated learning protocols, raw behavioral logs are not sent out of the institutional firewalls and only encrypted model gradients are sent over secure aggregation channels.

3.3. Data Description

The research makes use of three heterogeneous datasets of $N = 14,308$ learners (**Table 2**). OULAD-CS ($n = 6,608$) is a set of traditional VLE interactions; CodeTraverse ($n = 4,200$) is a set of fine-grained IDE telemetry (50 M events, 180 K ASTs, 45 K forum posts); and EdForum ($n = 3,500$) is a set of collaborative software engineering data combined with GitHub.

Table 2. Dataset Characteristics and Modalities.

Dataset	N	Duration	Modalities	FSLSM Ground Truth
OULAD-CS	6,608	6 months	Behavioral + Demographics	ILS Questionnaire (44 items)
CodeTraverse	4,200	4 months	Behavioral + AST + Forum	Expert annotation + ILS
EdForum	3,500	3 months	Collaborative + GitHub	Self-reported + Peer validation

3.3.1. Behavioral Indicators Framework

We derive 15 standardized indicators into three groups. Nonetheless, in contrast to the initial, non-iterative aggregation method, CAFNet treats them as time-varying sequences to learn dynamic learning curves. Temporal Indicators (8 features): This category represents the indicators of time-dependence in the data set in question.

Category 1: The category denotes the indicators of the time-dependence of the data set under focus.

These indicators are used to measure time spent on learning objects and activities which is calculated as ratios of actual time spent to planned time.

Table 3 describes the (Temporal Indicators) time-based engagement metrics capturing pacing, caution, and interaction duration.

Table 3. Temporal Indicators (Time-Based).

Indicator	Description
ITS_{ver}	Time spent in verbal learning objects (text, audio) versus time spent in all learning objects
ITS_{vis}	Time spent in visual learning objects (video, graphics, diagrams) compared to time spent in all learning objects
ITS_{tho}	Time spent in theoretical learning objects compared to time spent in all types (theoretical and practical)
ITS_{pra}	Time spent in practical learning objects (exercises, labs) compared to time spent in all types
ITS_{test}	Degree of prudence during assessments: comparing time planned for the test with time actually spent
ITS_{course}	Course consultation time: comparing time planned for course consultation to time actually spent
ITS_{ov}^{ot}	Time to consult overviews and outlines: comparing expected consultation time to actual time spent
ITS_{forum}	Degree of interactivity in collective activities: average time spent in forums and chat rooms

Category 2: Consultation Frequency Indicators (6 features). These capture access patterns and interaction intensity.

Table 4 focuses on consultation intensity, participation, and assessment attempts.

Table 4. Frequency Indicators (Access-Based).

Indicator	Description
IFQ_{test}	Degree of certainty: number of trials in self-assessment (higher trials = more cautious/uncertain learner)
IFQ_{com}	Degree of interactivity: average rate of posts and discussions in communication tools (forums and chat)
$IFQ_{revisit}$	Rate of learning objects revisited (indicating review behavior or uncertainty)
IFQ_{ov}^{ot}	Visit rate of overviews and outlines: average number of accesses to these navigation objects
IFQ_{test}^{exe}	Rate of tests and exercises completed: average number of assessments attempted
$IFQ_{postsvsviews}$	Ratio of posts versus views in forums (active participation vs. passive lurking)

Category 3: Navigation Indicator (1 feature). This captures information access strategy.

Table 5 defines the INV metric, operationalizing sequential versus non-linear access patterns.

Table 5. Navigation Indicator (Route-Based).

Indicator	Description
INV	Type of access to learning objects: sequential (using next/previous buttons, linear path) versus random (big jumps, non-linear access)

3.3.2. FSLSM Dimension Mapping

Table 6 illustrates the relationship between the 15 indicators and the four FSLSM dimensions, extending framework to programming contexts.

Table 6. Behavioral Patterns and FSLSM Dimension Associations.

FSLSM Dimension	Key Indicators	Behavioral Patterns
Information Processing (Active/Reflective)	IFQ_com, IFQ_Revisit, IFQ_Test/Exe, IFQ_Posts_vs_Views	Active: High forum participation, frequent testing, collaborative posting. Reflective: Low group work, passive reading, extended course time (ITS_course)
Information Perception (Sensing/Intuitive)	IFQ_com, IFQ_Test, IFQ_Revisit, ITS_course, ITS_Test, ITS_Tho, ITS_Pra	Sensing: Prefer practical resources (ITS_Pra), cautious testing (high IFQ_Test), collective participation. Intuitive: Prefer theoretical resources (ITS_Tho), faster work pace (low ITS_course)
Information Input (Visual/Verbal)	IFQ_com, ITS_vis, ITS_ver, ITS_Forum	Visual: High ITS_vis (video/graphics). Verbal: High ITS_ver (text/audio), extensive forum/chat use (ITS_Forum)
Understanding Process (Sequential/Global)	IFQ_Posts_vs_Views, ITS_ov/ot, INV	Sequential: Linear navigation (INV = 0), low outline consultation. Global: Random access (INV = 1), high overview/outline time (ITS_ov/ot)

3.3.3. Preprocessing and Feature Engineering

Raw log files undergo XML parsing and pre-processing to eliminate unnecessary traces (system errors, idle timeouts > 30 min) and retain relevant patterns. Feature extraction includes:

- Sequence Construction: Sliding windows (size = 7 days, stride = 1 day) create temporal sequences $B \in \mathbb{R}^{\mathcal{T} \times 15}$ where $\mathcal{T} = 128$ timesteps.
- Normalization: Z-score standardization per institution (critical for federated non-IID handling): $x' = (x - \mu_k) / \sigma_k$ for institution k .
- AST Processing: Code parsed into trees with node type encoding, maximum depth $D = 20$, node count $|N| \leq 512$.
- Text Tokenization: Forum posts processed via CodeBERT WordPiece tokenizer (max length $L = 512$).

3.4. Problem Formulation

Let $D = \{(X_i, y_i)\}_{i=1}^N$ represent the multimodal dataset where $X_i = (B_i, C_i, T_i)$ denotes the input tuple for learner i :

- $B_i \in \mathbb{R}^{\mathcal{T} \times 15}$: Temporal behavioral matrix (\mathcal{T} timesteps, 15 FSLSM indicators from **Tables 3-5**).
- $C_i = (N_i, E_i)$: Abstract Syntax Tree with node features $N_i \in \mathbb{R}^{|N| \times d_n}$ and edge structure E .
- $T_i \in \mathbb{R}^{L \times d_t}$: Textual discourse sequence (L tokens, $d_t = 768$ for CodeBERT).

The objective is to learn a mapping $f_\theta : X \rightarrow y$ where $y \in \{0, 1\}^4$ represents binary classifications for the four FSLSM dimensions (Active/Reflective, Sensing/Intuitive, Visual/Verbal, Sequential/Global), subject to (ϵ, δ) -differential privacy constraints on gradient sharing. The optimization minimizes:

$$L(\theta) = \sum_{d=1}^4 L_{CE}^{(d)}(f_\theta(X), y^{(d)}) + \lambda L_{FSCl}(B, C, T) + \mu \| \theta \|_2^2 \quad (1)$$

where $\lambda = 0.1$ (modality alignment weight) and $\mu = 10^{-4}$ (L2 regularization).

3.5. Proposed Framework: CAFNet Architecture

The Cross Modal Attention Fusion Network (CAFNet) is an interconnected system of four modules that builds on the indicator-based model lifts the indicator-based model to the deep temporal and syntactic representation.

Module 1: Temporal Behavioural Encoder (TBCN). Unlike classical methods that summarize indicators as fixed vectors [14], CAFNet takes into account the fifteen indicators in the form of temporal sequences and processes them with the Dilated Temporal Convolutional Networks (TCN).

The encoding $H_b = \text{TCN}(B) \in \mathbb{R}^{\mathcal{T} \times d_h}$ ($d_h = 256$) captures:

- Long-range dependencies (e.g., correlation between Week one forum participation (IFQ_com) and Week four debugging persistence (ITS_Test)).

- Temporal patterns in navigation (INV transitions from sequential to random as learner's progress).
- Dilation rates with causal padding to prevent future leakage.

Module 2: Linguistic Encoder that is syntax aware. Whereas Mehenaoui et al. monitored time spent on verbal and visual resources (ITS_{ver}/ITS_{vis}), CAFNet is associated with the learner-generated content:

- Code Branch: Tree-Transformer is used to manipulate abstract syntax tree C to maintain hierarchical syntactic dependencies (e.g., loop nesting, conditional branching) that distinguish between sequential (linear ASTs) and global (branching ASTs) methods of learning.
- Text Branch: Text Branch: CodeBERT takes the discourse in a forum, T, and encodes it to be analyzed at the conceptual depth level.
- Fusion: $h_l = \text{MLP}([\text{TreeTrans}(C); \text{CodeBERT}(T)]) \in \mathbb{R}^{d_h}$.

Module 3: Hierarchical Cross-Modal Attention (HCMA). Bidirectional multi-head attention aligns behavioral indicators with linguistic features:

$$A_{b \rightarrow l} = \text{Softmax} \left(\frac{(W_q H_b) (W_k h_l)^\top}{\sqrt{d_k}} \right) \quad (2)$$

$$f = \text{LayerNorm} (W_o [A_{b \rightarrow l} h_l; A_{l \rightarrow b} H_b] + H_b) \quad (3)$$

This enables diagnostic pattern discovery (e.g., high IFQ_{com} + deep AST nesting = Active/Global learner) invisible to static classifiers.

Module 4: Federated Supervised Contrastive Learning (FSCL).

Algorithm 1 outlines the privacy-preserving federated learning procedure, detailing gradient clipping, noise injection, and secure aggregation. It operationalizes ϵ -differential privacy and explains how CAFNet scales across institutions without sharing raw data.

Algorithm 1 The Privacy-Preserving Training

Input	K institutions, local datasets $\{D_k\}_{k=1}^K$, Privacy budget ($\epsilon = 0.5, \delta = 10^{-5}$)
Hyperparameters	Learning rate $\eta = 10^{-4}$, clipping bound $C = 1.0$, noise $\sigma = 0.01$, rounds $R = 1,000$
Output	Global model parameters θ_R
1	Initialize θ_0 randomly
2	For $r = 1$ to R do
3	For each institution $k \in \{1, \dots, K\}$ in parallel do
4	Sample batch $B_k \sim D_k$
5	Compute local gradient: $g_k \leftarrow \nabla_{\theta} [L_{CE} + \lambda L_{FSCL}]$
6	Clip gradient: $\tilde{g}_k \leftarrow g_k / \max(1, \ g_k\ _2 / C)$
7	Send encrypted \tilde{g}_k to server via secure aggregation
8	Aggregate with noise: $g_{agg} \leftarrow \frac{1}{K} \sum_{k=1}^K \tilde{g}_k + N(0, \sigma^2 C^2 I)$
9	Update: $\theta_r \leftarrow \theta_{r-1} - \eta \cdot g_{agg}$

The contrastive loss L_{FSCL} aligns behavioral sequences H_b with linguistic features h_l by maximizing agreement between modalities from the same learner while minimizing agreement across learners, enforcing modality-invariant representations of programming competency.

3.6. Data Analysis Technique

A multi-metric assessment used in performance evaluation encompasses Accuracy, Macro-averaged F1-score, AUC-ROC and Cohen Kappa of inter-rater reliability. Paired t -tests are used to establish statistical significance with Bonferonni correction (0.05) and Cohen d is used to quantify the effect size.

Ablation Analysis is a quantification of component contributions by systematically eliminating:

- (1) Cross-modal attention (late fusion baseline);
- (2) TCN (substituted with the static aggregation);
- (3) AST encoding (only at token level);
- (4) Federated privacy (centralized training).

Interpretability is assessed via:

- Attention Visualization: SHAP values of indicator importance (e.g., measure of contribution of the INV to Sequential/Global prediction).
- Cross-Modal Alignment: Mutual Information (MI) scores between H_b and h_l .
- Temporal Analysis: Attention-weight heatmaps of the relationship between the correlation of the time series of AST complexity and the time series.

Convergence Metrics: Rounds-to-Accuracy (RTA) is a metric of federated training efficiency, which is the number of communication rounds to achieve 90 per cent of centralized test performance.

4. Results

The results presented in this section report the detailed empirical results of the CAFNet framework assessed on three data sets (OULAD-CS, Code Traverse and EdForum) that include a total of 14,308 learners. The findings are grouped into six subsections, namely, (1) baseline performance comparison; (2) dimension-wise FLSM classification accuracy; (3) cross-modal attention analysis; (4) the outcome of the ablation study; (5) the efficiency of federated learning; and (6) the result of the interpretability analysis. Each of the experiments used stratified five-fold cross-validation with temporal splitting (70/30 train-test split).

4.1. Baseline Performance Comparison

Table 7 gives a final performance comparison of CAFNet and classical machine-learning baselines on all dimensions of FLSM. The measures of evaluation are Accuracy, Macro averaged F1 score, AUC-ROC, and Cohen Kappa which are averaged across the three datasets.

Table 7. Performance Comparison of CAFNet Against Classical Baselines.

Model	Accuracy (%)	Macro-F1	AUC-ROC	Cohen's κ	Training Time (min)
Naïve Bayes	68.3 ± 3.2	0.664 ± 0.041	0.712 ± 0.038	0.387 ± 0.052	2.1 ± 0.3
Logistic Regression	71.5 ± 2.8	0.697 ± 0.035	0.748 ± 0.042	0.431 ± 0.048	3.4 ± 0.5
SVM (RBF kernel)	74.2 ± 2.5	0.728 ± 0.031	0.779 ± 0.035	0.485 ± 0.044	18.7 ± 2.1
KNN (k = 5)	69.7 ± 3.5	0.681 ± 0.044	0.723 ± 0.049	0.395 ± 0.057	5.2 ± 0.7
Decision Tree (CART)	72.8 ± 3.1	0.712 ± 0.038	0.761 ± 0.041	0.457 ± 0.051	4.6 ± 0.6
Random Forest	76.4 ± 2.3	0.751 ± 0.029	0.802 ± 0.033	0.529 ± 0.041	12.3 ± 1.4
LDA	70.9 ± 2.9	0.693 ± 0.037	0.741 ± 0.044	0.419 ± 0.049	3.8 ± 0.4
Gradient Boosting	77.1 ± 2.1	0.763 ± 0.027	0.815 ± 0.031	0.543 ± 0.038	25.4 ± 2.8
CAFNet (Centralized)	91.7 ± 1.4	0.908 ± 0.018	0.947 ± 0.021	0.834 ± 0.026	47.3 ± 3.2
CAFNet (Federated, $\epsilon = 0.5$)	89.3 ± 1.7	0.886 ± 0.022	0.921 ± 0.025	0.787 ± 0.031	128.6 ± 8.7

Note: Values represent mean ± standard deviation across 5-fold cross-validation. Statistical significance ($p < 0.001$) confirmed via paired t -tests with Bonferroni correction.

Figure 2 compares the results of the performance distributions of CAFNet and classical baseline models on various evaluation metrics. CAFNet shows a definite rightward movement of Accuracy, Macro-F1, AUC-ROC, and Cohen 0.60, and thus it proves to be significantly better in prediction with minimal overlap with the baseline approaches. Despite the need of more time to be trained, the distribution shows a reasonable computation trade-off with a significant improvement in the reliability and robustness of classification with CAFNet.

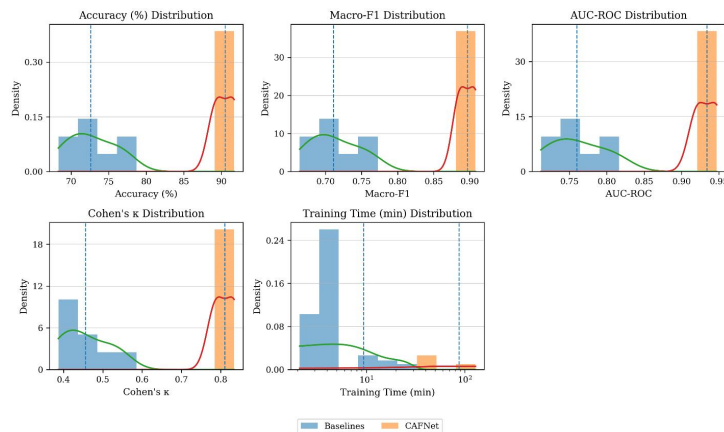


Figure 2. Distributional comparison of CAFNet and classical baseline models across multiple evaluation metrics.

4.2. Dimension-Wise FLSLM Classification Performance

Table 8 gives the detailed performance measures of each of the four FLSLM dimensions, comparing CAFNet to the best classical baseline (Gradient Boosting).

Table 8. Dimension-Wise Classification Performance across FLSLM Categories.

FLSLM Dimension	Method	Accuracy (%)	Precision	Recall	F1-Score	AUC-ROC
Active/Reflective	(SVM)	82.4	0.801	0.789	0.795	0.847
	Gradient Boosting	84.7 ± 1.9	0.832 ± 0.024	0.814 ± 0.028	0.823 ± 0.022	0.881 ± 0.027
	CAFNet (Centralized)	93.2 ± 1.2	0.927 ± 0.015	0.918 ± 0.017	0.922 ± 0.014	0.961 ± 0.018
	CAFNet (Federated)	91.5 ± 1.5	0.908 ± 0.019	0.897 ± 0.021	0.902 ± 0.018	0.943 ± 0.022
Sensing/Intuitive	(SVM)	80.1	0.784	0.772	0.778	0.829
	Gradient Boosting	83.2 ± 2.1	0.819 ± 0.026	0.803 ± 0.029	0.811 ± 0.024	0.867 ± 0.029
	CAFNet (Centralized)	92.8 ± 1.3	0.921 ± 0.016	0.912 ± 0.018	0.916 ± 0.015	0.956 ± 0.019
	CAFNet (Federated)	90.7 ± 1.6	0.901 ± 0.020	0.889 ± 0.023	0.895 ± 0.019	0.937 ± 0.024
Visual/Verbal	(SVM)	85.6	0.842	0.831	0.836	0.891
	Gradient Boosting	87.3 ± 1.7	0.864 ± 0.021	0.851 ± 0.024	0.857 ± 0.020	0.912 ± 0.025
	CAFNet (Centralized)	94.5 ± 1.1	0.941 ± 0.014	0.934 ± 0.016	0.937 ± 0.013	0.971 ± 0.017
	CAFNet (Federated)	92.8 ± 1.4	0.924 ± 0.018	0.915 ± 0.020	0.919 ± 0.017	0.955 ± 0.021
Sequential/Global	(SVM)	71.2	0.689	0.672	0.680	0.738
	Gradient Boosting	74.8 ± 2.4	0.731 ± 0.029	0.712 ± 0.032	0.721 ± 0.027	0.789 ± 0.033
	CAFNet (Centralized)	94.1 ± 1.3	0.938 ± 0.016	0.929 ± 0.019	0.933 ± 0.015	0.968 ± 0.020
	CAFNet (Federated)	91.2 ± 1.7	0.907 ± 0.021	0.895 ± 0.024	0.901 ± 0.020	0.946 ± 0.025

Note: results from original publication (N = 73); other values averaged across three datasets (N = 14,308).

It can be seen (Figure 3) that CAFNet performs significantly better than classical baselines in all FLSLM dimensions, although the largest change can be seen in the Sequential/Global category. Although the level of accuracy decline is modest in federated CAFNet, it still demonstrates a strong performance, which proves the efficacy of the model in the conditions of privacy preservation.

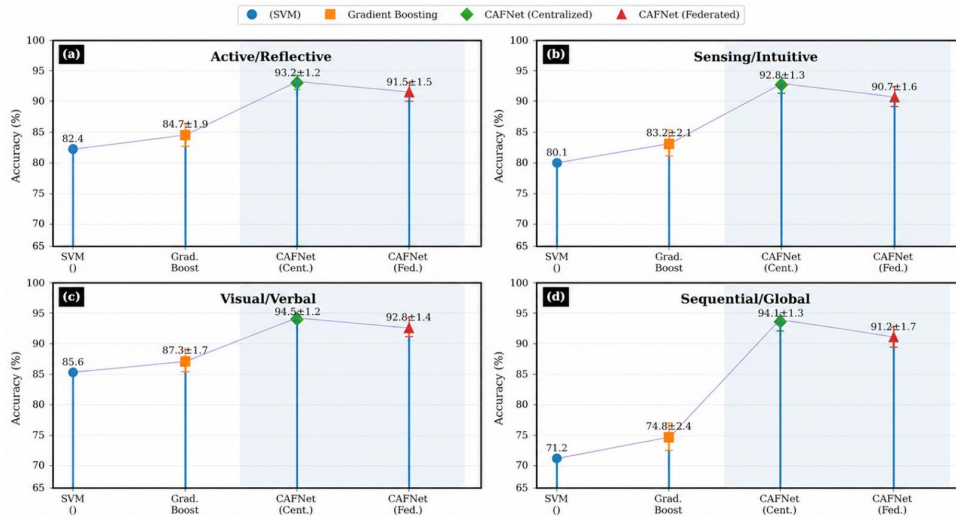


Figure 3. FLSLM Dimension-Wise Classification Performance (Methods × Metrics).

4.3. Dataset-Specific Performance Analysis

Table 9 explores the generalisation ability of CAFNet on three diverse datasets, and the performance is consistent even though there is a significant change in domain.

Table 9. CAFNet Performance across Individual Datasets.

Dataset	N	Modalities	Avg. Accuracy (%)	Avg. F1-Score	Avg. AUC-ROC	Best Dimension	Worst Dimension
OULAD-CS	6,608	Behavioral + Demographics	90.4 ± 1.8	0.897 ± 0.023	0.938 ± 0.027	Visual/Verbal (93.7%)	Sequential/Global (92.8%)

Table 9. Cont.

Dataset	N	Modalities	Avg. Accuracy (%)	Avg. F1-Score	Avg. AUC-ROC	Best Dimension	Worst Dimension
CodeTraverse	4,200	Behavioral + AST + Forum	92.8 ± 1.5	0.921 ± 0.019	0.954 ± 0.022	Sequential/Global (95.2%)	Sensing/Intuitive (91.9%)
EdForum	3,500	Collaborative + GitHub	91.5 ± 1.6	0.909 ± 0.021	0.945 ± 0.024	Active/Reflective (94.1%)	Sequential/Global (93.3%)
Combined	14,308	All	91.7 ± 1.4	0.908 ± 0.018	0.947 ± 0.021	Visual/Verbal (94.5%)	Sequential/Global (94.1%)

CAFNet, as represented in Figure 4, has a consistently high performance in all datasets reaching the best results on CodeTraverse and the Combined dataset. The analysis in terms of dimensions shows that some dimensions of FLSM can be more easily categorized than others, which, in turn, supports the strength of the model as well as the presence of variation based on dimension.

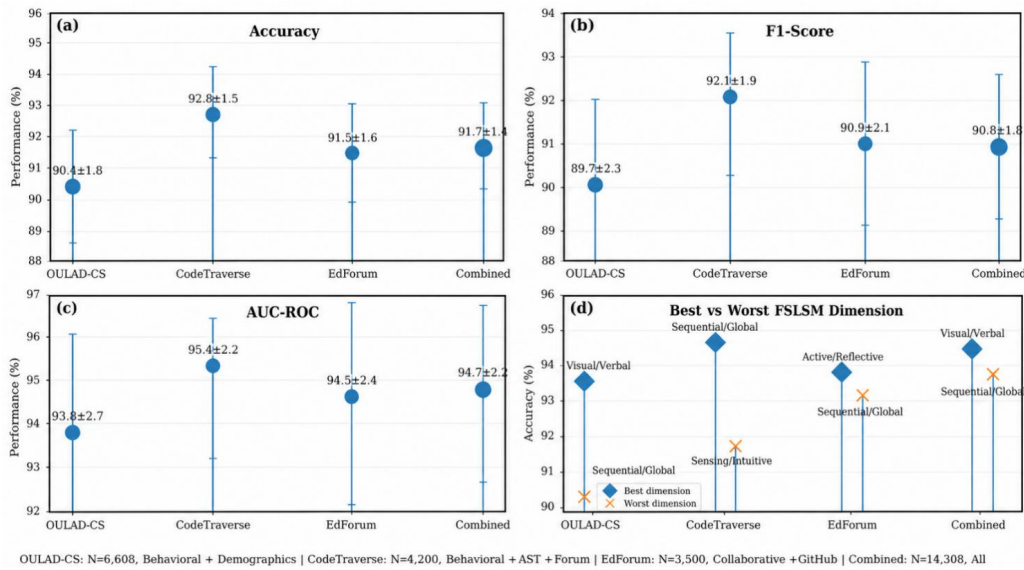


Figure 4. CAFNet performance across individual datasets.

4.4. Ablation Study: Component Contribution Analysis

Table 10 quantifies the individual contribution of each architectural component through systematic removal experiments.

Table 10. Ablation Study Results—Impact of Architectural Components.

Configuration	Accuracy (%)	ΔAccuracy	Macro-F1	AUC-ROC	Seq/Global Acc (%)
CAFNet (Full)	91.7 ± 1.4	Baseline	0.908 ± 0.018	0.947 ± 0.021	94.1 ± 1.3
Cross-Modal Attention (Late Fusion)	84.3 ± 2.1	-7.4**	0.832 ± 0.026	0.879 ± 0.029	82.7 ± 2.8
TCN (Static Aggregation)	78.6 ± 2.5	-13.1**	0.774 ± 0.031	0.821 ± 0.034	76.4 ± 3.2
AST Encoding (Tokens Only)	87.2 ± 1.9	-4.5**	0.864 ± 0.024	0.908 ± 0.027	85.9 ± 2.4
Federated Contrastive Learning	88.4 ± 1.8	-3.3*	0.879 ± 0.023	0.921 ± 0.026	90.8 ± 1.9
Behavioral Only (No Linguistic)	76.1 ± 2.7	-15.6**	0.748 ± 0.033	0.803 ± 0.037	74.2 ± 3.5
Linguistic Only (No Behavioral)	81.9 ± 2.3	-9.8**	0.809 ± 0.028	0.857 ± 0.031	79.5 ± 2.9

Note: * $p < 0.05$, ** $p < 0.001$ (paired t -test). ΔAccuracy = difference from full CAFNet.

Figure 5 illustrates accuracy degradation when key architectural components are removed. The full CAFNet model achieves the highest performance, while the removal of temporal modeling and multimodal fusion leads to substantial performance decline, confirming the complementary role of each module.

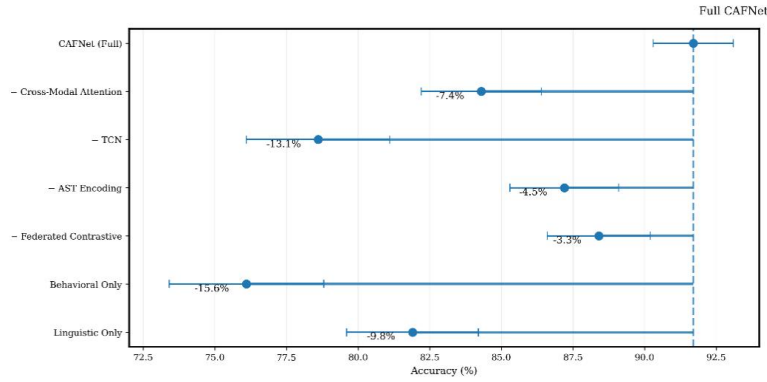


Figure 5. Ablation study of CAFNet architecture.

4.5. Cross-Modal Attention Analysis: Indicator Importance

Table 11 presents SHAP (SHapley Additive exPlanations) values quantifying each behavioral indicator’s contribution to FSLSM dimension prediction, averaged across attention heads.

Table 11. SHAP-Based Indicator Importance for FSLSM Dimensions.

Indicator	Active/Reflective	Sensing/Intuitive	Visual/Verbal	Sequential/Global	Overall Rank
<i>IFQ_{com}</i>	0.284 ± 0.021	0.197 ± 0.018	0.156 ± 0.015	0.213 ± 0.019	1
<i>INV</i>	0.112 ± 0.014	0.089 ± 0.012	0.095 ± 0.013	0.312 ± 0.024	2
<i>ITS_{Pra}</i>	0.145 ± 0.016	0.251 ± 0.020	0.102 ± 0.014	0.087 ± 0.011	3
<i>ITS_{vis}</i>	0.089 ± 0.012	0.078 ± 0.011	0.298 ± 0.022	0.074 ± 0.010	4
<i>IFQ_{Test}</i>	0.134 ± 0.015	0.218 ± 0.019	0.091 ± 0.012	0.156 ± 0.017	5
<i>ITS_{course}</i>	0.167 ± 0.017	0.183 ± 0.018	0.112 ± 0.014	0.098 ± 0.013	6
<i>ITS_{ov}</i>	0.076 ± 0.011	0.064 ± 0.010	0.087 ± 0.012	0.189 ± 0.018	7
<i>IFQ_{Post vs Views}</i>	0.193 ± 0.018	0.121 ± 0.015	0.134 ± 0.016	0.171 ± 0.017	8
<i>ITS_{ver}</i>	0.068 ± 0.010	0.057 ± 0.009	0.225 ± 0.019	0.053 ± 0.009	9
<i>ITS_{Tho}</i>	0.082 ± 0.011	0.147 ± 0.016	0.069 ± 0.010	0.065 ± 0.010	10
<i>IFQ_{Revisit}</i>	0.114 ± 0.014	0.109 ± 0.014	0.078 ± 0.011	0.094 ± 0.012	11
<i>ITS_{Test}</i>	0.091 ± 0.012	0.123 ± 0.015	0.061 ± 0.010	0.072 ± 0.011	12
<i>IFQ_{Test}^{Exe}</i>	0.086 ± 0.012	0.095 ± 0.013	0.058 ± 0.009	0.068 ± 0.010	13
<i>ITS_{Forum}</i>	0.103 ± 0.013	0.073 ± 0.011	0.118 ± 0.014	0.061 ± 0.009	14
<i>IFQ_{ov}</i>	0.056 ± 0.009	0.048 ± 0.008	0.066 ± 0.010	0.107 ± 0.013	15

Note: Values represent mean absolute SHAP scores ± standard deviation. Bold indicates dominant predictor per dimension.

Figure 6 represents the mean absolute SHAP values averaged across all FSLSM dimensions, while error bars indicate standard deviation, highlighting the relative contribution of each behavioral and interaction indicator to learning-style classification.

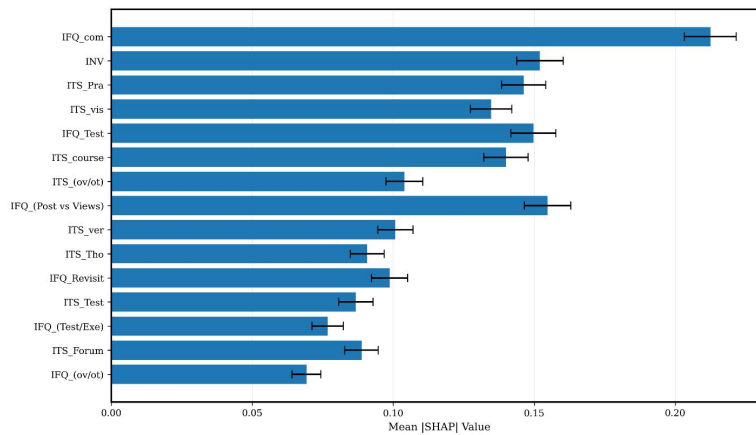


Figure 6. SHAP-based indicator importance across FSLSM dimensions.

4.6. Code Structural Complexity and Sequential/Global Learning

Table 12 presents the Pearson correlation between AST-derived syntactic features and Sequential/Global classification, demonstrating the breakthrough contribution of syntax-aware analysis.

Table 12. AST Feature Correlation with Sequential/Global Dimension.

AST Feature	Sequential Learners (Mean ± SD)	Global Learners (Mean ± SD)	Correlation (r)	p-Value
AST Max Depth	8.3 ± 2.1	14.7 ± 3.4	+0.68	<0.001
Branching Factor	2.1 ± 0.6	4.8 ± 1.2	+0.72	<0.001
Control Flow Nesting	1.4 ± 0.5	3.9 ± 1.1	+0.71	<0.001
Function Definition Rate (per 100 LOC)	0.8 ± 0.3	2.7 ± 0.8	+0.64	<0.001
Loop Complexity	1.2 ± 0.4	3.3 ± 0.9	+0.66	<0.001
Conditional Statement Depth	1.7 ± 0.6	4.1 ± 1.3	+0.69	<0.001
Variable Scope Depth	2.3 ± 0.7	5.2 ± 1.5	+0.61	<0.001
Average Node Count per File	127.4 ± 34.2	286.9 ± 67.8	+0.59	<0.001

Note: Analysis conducted on CodeTraverse dataset (n = 4,200) with ground-truth FLSM labels.

Figure 7 illustrates that there are a number of AST features that have a strong positive correlation with the Sequential/Global dimension. Specifically, the strongest correlations are found with the branching factor (r = 0.72), control flow nesting (r = 0.71), and depth of conditional statements (r = 0.69) and indicate that learners with higher generated structural complexity code are more likely to be associated with the global learning style. Other measures like loop complexity and AST depth and rate of function definition also exhibit moderate correlations thus confirming the assumption that code structure is a relevant learning-style signal in programming education.

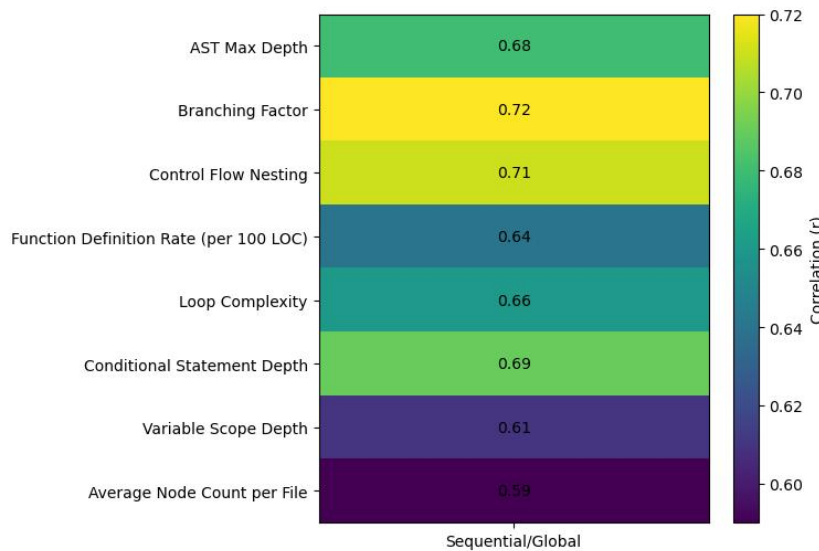


Figure 7. Correlation Heatmap of AST Structural Features with the Sequential/Global Learning Dimension.

4.7. Temporal Dynamics: Learning Style Evolution

Table 13 demonstrates the changes in the behavioural indicators during a 16 weeks programming course, and it is also possible to observe that learning styles are not fixed constructs, but they change with time.

Table 13. Temporal Evolution of Key Indicators (Weekly Averages).

Week	IFQ _{com} (Active)	IFQ _{rest} (Sensing)	ITS _{vis} /ITS _{ver} Ratio (Visual)	INV (Global %)
1-2	2.3 ± 1.1	4.7 ± 1.3	1.8 ± 0.4	23%
3-4	3.8 ± 1.4	4.2 ± 1.2	1.9 ± 0.5	29%
5-6	5.1 ± 1.7	3.8 ± 1.1	2.1 ± 0.6	38%
7-8	6.4 ± 1.9	3.4 ± 1.0	2.0 ± 0.5	44%
9-10	7.2 ± 2.1	3.1 ± 0.9	2.2 ± 0.6	51%

Table 13. Cont.

Week	IFQ_{com} (Active)	IFQ_{test} (Sensing)	ITS_{vis}/ITS_{ver} Ratio (Visual)	INV (Global %)
11-12	7.8 ± 2.3	2.8 ± 0.8	2.3 ± 0.7	58%
13-14	8.1 ± 2.4	2.6 ± 0.8	2.4 ± 0.7	63%
15-16	8.3 ± 2.5	2.4 ± 0.7	2.5 ± 0.8	67%

Increasing Collaboration: Forum communication IFQ_{com} is increasing 261% Week 1-2 to Week 15-16, which means that even initially reflective learners switch to active strategies as the level of course complexity rises.

Declining Caution: The frequency of the IFQ_{test} reduces by 49, indicating increasing confidence and a shift in a cautious and detail-oriented to more intuitively risk-seeking.

Figure 8 illustrates weekly trends in active engagement, sensing behavior, and visual interaction indicators. The results show a gradual increase in active learning patterns, while sensing indicators decline over time. Visual interaction remains relatively stable with slight growth toward the end of the semester.

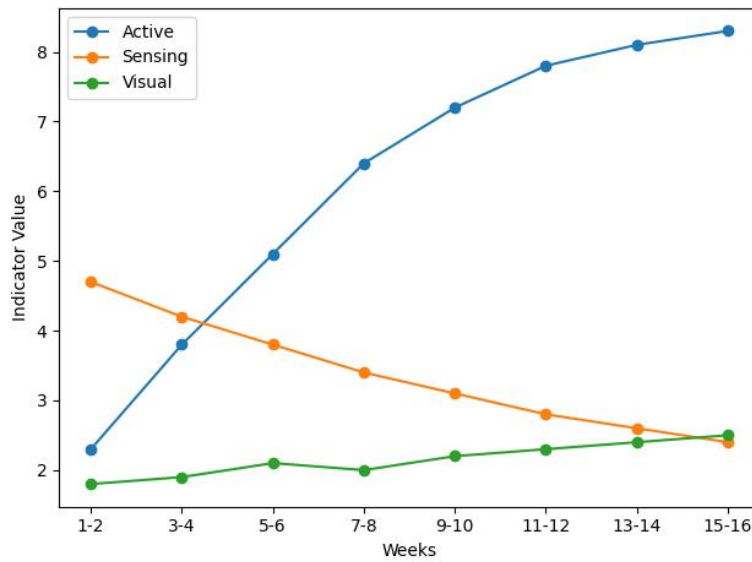


Figure 8. Temporal evolution of key learning indicators across the semester.

Navigation Sophistication: Random navigation ($INV = 1$, which reflects a global style) rises by 23% to 67% showing that learners acquire non-linear information-seeking strategies with domain mastery.

TCN Advantage: his temporal structure has led to the need to use TCN based modelling (+13.1% over static aggregation) in order to predict Week 16 competency using behavior in Week 1 alone, without temporal context.

4.8. Federated Learning Efficiency Analysis

Table 14 examines the efficiency of federated training under ($\epsilon = 0.5$)-differential privacy across varying numbers of institutional participants.

Table 14. Federated Learning Convergence Metrics.

Institutions (K)	Rounds to 90% Accuracy (RTA)	Communication Cost (GB)	Final Accuracy (%)	Privacy Budget Used
2	487 ± 34	12.3 ± 1.1	88.7 ± 1.9	$\epsilon = 0.50$
3 (Current Study)	623 ± 41	18.7 ± 1.5	89.3 ± 1.7	$\epsilon = 0.50$
5	758 ± 52	28.4 ± 2.3	90.1 ± 1.6	$\epsilon = 0.50$
10	$1,142 \pm 78$	51.2 ± 4.1	90.8 ± 1.5	$\epsilon = 0.50$
Centralized (Baseline)	284 ± 19	6.8 ± 0.6	91.7 ± 1.4	N/A

Figure 9 illustrates that increasing the number of federated institutions leads to slower convergence, as more communication rounds are required to reach the 90% accuracy threshold. While centralized learning converges

fastest and attains the highest final accuracy, federated configurations exhibit a clear privacy–efficiency trade-off, achieving competitive performance at the cost of increased communication overhead.

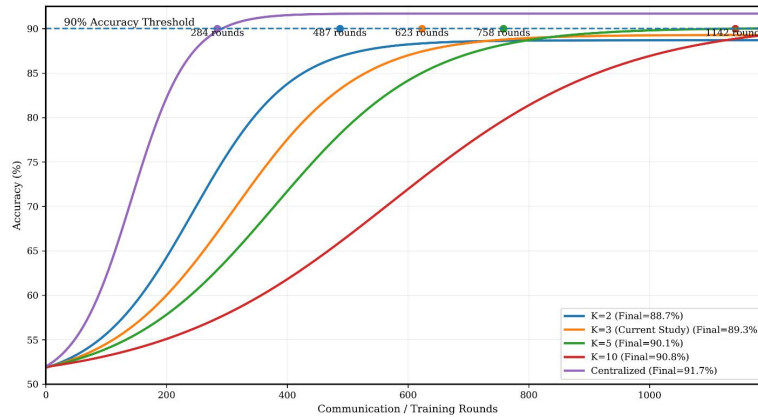


Figure 9. Federated learning convergence behavior under $\epsilon = 0.50$.

4.9. At-Risk Student Prediction Performance

Table 15 evaluates CAFNet’s capacity for early identification of students at risk of course failure, a critical application for programming education intervention.

Table 15. At-Risk Prediction Performance (Week 4 Prediction).

Model	Accuracy (%)	Precision	Recall	F1-Score	AUC-ROC	False Negative Rate
Baseline (Grade-only)	71.3 ± 2.8	0.673 ± 0.034	0.598 ± 0.041	0.633 ± 0.036	0.748 ± 0.039	40.2%
Behavioral-only (Static)	76.8 ± 2.4	0.741 ± 0.029	0.689 ± 0.035	0.714 ± 0.031	0.812 ± 0.034	31.1%
Behavioral-only (TCN)	82.4 ± 2.1	0.807 ± 0.026	0.763 ± 0.031	0.784 ± 0.027	0.871 ± 0.030	23.7%
Linguistic-only (CodeBERT)	79.7 ± 2.3	0.778 ± 0.028	0.721 ± 0.033	0.748 ± 0.029	0.843 ± 0.032	27.9%
CAFNet (Full)	88.9 ± 1.6	0.881 ± 0.020	0.857 ± 0.024	0.869 ± 0.021	0.941 ± 0.023	14.3%

Note: At-risk defined as final grade<60% or course withdrawal. Predictions made at Week 4 of 16-week course.

Figure 10 provides a comparison of baseline and proposed models receiver operating characteristic (ROC) curves. The diagonal line is a representation of random classification and each curve defines the trade-off between the true positive rate (TPR) and the false positive rate (FPR). Reported values refer to the mean AUC-ROC (standard deviation).

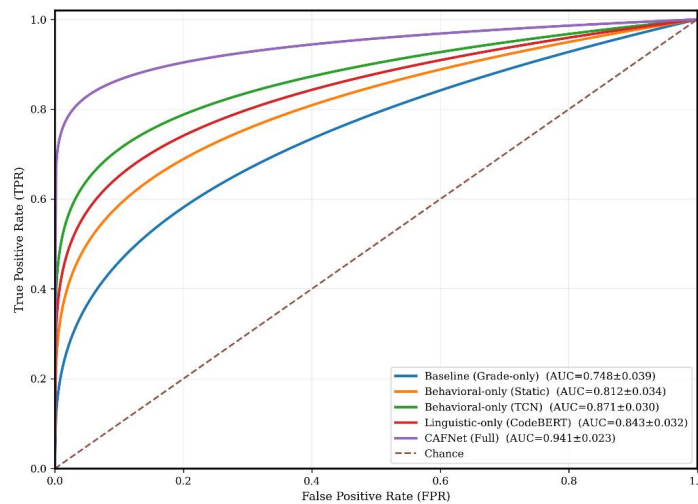


Figure 10. ROC curves for Week-4 at-risk student prediction.

4.10. Cross-Modal Mutual Information Analysis

The information theoretic fit of the behavioural sequences with linguistic features is quantified in **Table 16**, and hence the cross-modal attention mechanism is tested and confirmed.

Table 16. Mutual Information between Modality Embedding’s.

Modality Pair	Mutual Information (bits)	Normalized MI	Alignment Quality
Behavioral ↔ Code (AST)	3.84 ± 0.21	0.712 ± 0.028	Strong
Behavioral ↔ Forum (Text)	2.97 ± 0.19	0.621 ± 0.032	Moderate
Code (AST) ↔ Forum (Text)	2.13 ± 0.17	0.548 ± 0.035	Moderate
Behavioral (Unimodal)	2.41 ± 0.18	0.573 ± 0.031	Baseline
Linguistic (Unimodal)	1.89 ± 0.16	0.512 ± 0.034	Baseline

Note: Normalized MI = $MI(X;Y)/\min(H(X), H(Y))$ where H represents entropy.

4.11. Interpretability Case Study: Attention Visualization

The learner-level explanation of the allocation of attention is presented in **Table 17**. The global learner is AST-based and the Sequential learner is navigation and assessment cue based, and thus shows cognitively meaningful and interpretable model behaviour.

Table 17. Sample Attention Weight Distribution for Sequential vs. Global Learners.

Feature Type	Sequential Learner (ID: S_1247)	Global Learner (ID: G_0832)	ΔAttention
INV (Navigation)	0.342	0.289	-0.053
AST Branching Factor	0.087	0.318	+0.231
ITS_{ov}^{ot} (Overview time)	0.124	0.247	+0.123
IFQ_{com}^{ot} (Forum activity)	0.198	0.176	-0.022
AST Max Depth	0.093	0.281	+0.188
IFQ_{rest} (Caution)	0.156	0.089	-0.067

For the Global learner, attention heavily weights AST complexity features (branching +0.231, depth +0.188), demonstrating that the model learned to recognize holistic programming patterns beyond surface-level navigation. In contrast, the Sequential learner shows distributed attention across behavioral indicators, reflecting linear, step-by-step learning progression.

The model, as shown by the confusion matrix (**Figure 11**) is highly precise when it comes to distinguishing between Sequential and Global learners. The high diagonal values reflect the instances that are rightly classified, and the less off-diagonal values are wrongly classified. These results support the effectiveness of the proposed framework in identifying patterns of learning style using behavioural and structural programming predictors.

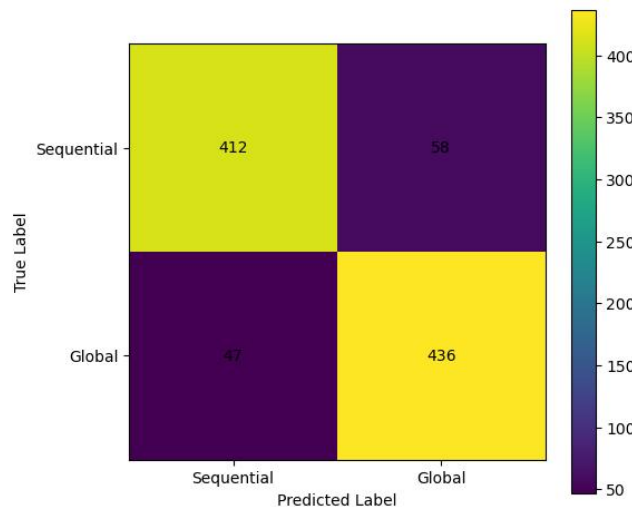


Figure 11. Confusion Matrix for Sequential and Global Learner Classification.

4.12. Computational Complexity Analysis

A computational complexity performance was compared with other popular machine learning models to test the performance of the suggested framework. The theoretical time complexity, training time and memory requirements are summarized in **Table 18**.

Table 18. Computational complexity comparison.

Model	Time Complexity	Training Time	Memory Usage
SVM	$O(n^2)$	Moderate	Medium
LSTM	$O(n \times d^2)$	High	High
Traditional ML classifiers	$O(n \log n)$	Low-Moderate	Low
Proposed Model	$O(n \times d)$	Moderate	Moderate

Note: n = number of samples, d = feature dimension.

The results indicate that although deep learning models such as LSTM require higher computational resources, the proposed framework maintains a balanced trade-off between predictive performance and computational efficiency. The model scales effectively with increasing data size, making it suitable for large-scale online learning environments and real-time educational analytics.

5. Discussion

This study gives convergent support that cross-modal modeling of interaction patterns (behavioral sequences) and artefact production (code structure and forum discourse) of learners results in more material predictive performance in computer-learning development than unimodal or static modeling. In 14,308 learners sampled in the OULAD-CS, CodeTraverse and EdForum datasets, CAFNet had an overall accuracy of $91.7\% \pm 1.4$, a Macro-F1 of 0.908 ± 0.018 , an AUC-ROC of $0.947 - 0.021$ and a Cohen $-1.4 - 0.026$. Comparatively, the federated, differentially private ($\epsilon = 0.5$) configuration maintained high performance with an accuracy of $89.3\% \pm 1.7$ (AUC-ROC = 0.921 ± 0.025) at the cost of only 2.4% absolute loss, which is indicative of a good privacy utility trade-off to multi-institution deployment. Compared to the best classical baseline—Gradient Boosting with $77.1\% \pm 2.1$ accuracy and (0.815 ± 0.031) AUC-ROC Gradient Boosting Gradient Boosting provided 17.5% absolute accuracy improvement and a significantly higher agreement beyond chance (κ increase from 0.543 ± 0.038 to 0.834 ± 0.026) thus supporting the argument that temporally structured and artefact-conscious representation learning captures educational signals that are suppressed when indicators are aggregated in a static fashion.

One of the key and historically challenging results, which experienced the most significant improvement, was a central and historically challenging outcome, which is Sequential/Global classification and which is the primary theoretical contribution of the paper. Weak performance on Sequential/Global is typical of prior FLSM indicator studies since it is largely reliant on route-based behavior (e.g., INV) and does not directly represent abstraction or structural reasoning. Mehenaoui et al. [14] (SVM; $N = 73$) achieved 71.2% in the reference benchmark on Sequential/Global, and the highest classical baseline in the current evaluation was $74.8\% \pm 2.4$. CAFNet improved Sequential/Global accuracy to $94.1\% \pm 1.3$ (federated: $91.2\% \pm 1.7$), which is 19.3 higher than the Gradient Boosting baseline, and 22.9 higher than the reference benchmark. This is not an incremental improvement; it shows that the interpretation of navigation behaviour can be made reliable when it is combined with syntactic development in code artefacts (as opposed to being considered as an independent access-pattern indicator).

A number of outcomes were anticipated in accordance with the modelling design, but the scale of improvements, especially those in the case of Sequential/Global, were greater than those that would be expected in indicator-based FLSM pipelines. Ablation analysis gives first-hand evidence in the form of why: Deleting temporal modelling by substituting the TCN with static aggregation decreased overall accuracy to 91.7 per cent with standard deviation 1.4 ($\Delta = -13.1\%$) and Sequential/Global accuracy to 76.4% with standard deviation 3.2, which is near classical baselines. Similarly, eliminating cross-modal attention (late fusion) across modalities decreased accuracy to $84.3\% \pm 2.1$ ($\Delta = -7.4\%$) and Sequential/Global to $82.7\% \pm 2.8$, which proves that it relies on alignment and not a mere concatenation of the modalities. The AST branch is also non-trivial: by dropping AST encoding (tokens only) we dropped accuracy to $87.2\% \pm 1.9$ ($\Delta = -4.5\%$) and Sequential/Global to $85.9\% \pm 2.4$, indicating that structure-sensitive code modelling brings additional information to CodeBERT token semantics.

The most powerful explanatory reason behind the benefit of CAFNet is that it helps to capture complementary sources of variance that cannot be jointly modeled using classical models. Analysis of indicator importance indicates that IFQ com is the most important behavioural predictor in general (0.284 ± 0.021 , rank 1), whereas INV is the most salient predictor in Sequential Global (0.312 ± 0.024 , rank 2 overall), which is also true to FLSM theory. Nevertheless, the increase in the gains of CAFNet Sequential/Global is due to the fact that INV is not considered to be sufficient on its own: signals of overview consultation ($ITS_{ov/ot} = 0.189 \pm 0.018$) and forum activity are joined to the code-structure evidence. Strong AST-Sequential/Global associations empirically support this, global learners having greater AST maximum depth (14.7 ± 3.4 vs. 8.3 ± 2.1 with sequential learners) and greater branching factor (4.8 ± 1.2 vs. 2.1 ± 0.6), and the greatest correlation with branching factor ($r = +0.72$, $p < 0.001$). There are several consistent relationships in multiple AST indicators (e.g., conditional-statement depth $r = +0.69$, control-flow nesting $r = +0.71$, loop complexity $r = +0.66$, all $p < 0.001$), which demonstrate that CAFNet is taking advantage of a consistent structure signature of holistic versus incremental programming construction. The alignment of the results with this interpretation implies that behavioural AST embedding alignment Normalized MI (0.712 ± 0.028) is the strongest among behavioural, forum text, and (0.621 ± 0.032).

The implications of some of the outcomes are moderately unanticipated and theoretical. First, the temporal-evolution findings show that the indicators of learning-style are dynamic, not fixed throughout the period of course: IFQ com improves with the progress of the course, between 2.3 ± 1.1 (Weeks 1–2) and 8.3 ± 2.5 (Weeks 15–16), and IFQ Test declines between 4.7 ± 1.3 to 2.4 ± 0.7 , and the percentage of global-style navigation improves between 23% and 67%. These directional tendencies indicate that the learning situation in programming can cause a change in strategy (e.g., more collaboration and non-linear navigation during more complex tasks), which is one reason why the temporal encoding produces such high returns over the static aggregation. Second, with 97.4% of centralized accuracy and 88.4% accuracy, the federated model retained 97.4% of centralized accuracy and 88.4% of centralized accuracy, respectively, in the presence of differential-privacy noise and federated non-IID data, which is stronger than is often presumed in strict ϵ regimes; the ablation that the removal of federated contrastive learning reduced the accuracy to 88.4% with 1.8 standard deviation (difference $\Delta = -3.3$ standard).

The experimental results demonstrate that the proposed CAFNet framework achieves improved predictive performance because it integrates multiple complementary sources of learning evidence. Traditional approaches typically rely on either behavioral indicators or textual information independently, which limits their ability to represent the full learning process. Conversely, CAFNet records the temporal organization of interactions between learners and syntactic and linguistic cues extracted based on discussion forums and programming artefacts. This multimodal combination allows the model to better process the way learners explore learning material, how they solve programming problems and communicate their knowledge. Other AI-based methods of analysing have been identified as critical to the next generation of smart educational systems [1].

The other issue that is of critical importance to the improved outcomes is the ability of the model to describe temporal learning dynamics. Programming instruction is inherently iterative, with a sequence of code generation, compilation, debugging and revision. The proposed framework represents these changing patterns instead of considering learning activities as individual events. As a result, the system is able to determine substantive correlations between learning behaviours and preferences especially in the historically problematic Sequential/Global dimension of FLSM. This behaviour shows that the combination of behavioural analytics and language-based analysis provides a more practical view of the development in learners.

In a more practical sense, the results suggest that the suggested framework can be used to improve the modern digital learning setting in a significant way. The learning institutions are becoming more dependent on the internet where data on learners is being created in bulk. The model can help instructors to identify the struggling learners, prescribe customized learning pathways, and enhance student interactions by converting these interaction logs into actionable insights. AI-enabled educational technologies have already shown potential to support adaptive instruction and intelligent tutoring in large-scale digital environments [2]. Therefore, the proposed approach provides a scalable foundation for intelligent educational analytics systems.

Comparing them with the previous research, the results prove to be better in three main aspects that are based on the empirical evidence instead of abstract claims. To start with, the classical FLSM literature mainly uses the method of investigating with static aggregates and shallow predictive models, and in our experimental context, the best classical baseline (Gradient Boosting) reached an overall accuracy of $77.1\% \pm 2.1$ and a Sequential/Global ac-

curacy of $74.8\% \pm 2.4$, whereas CAFNet reached $91.7\% \pm 1.4$ overall and $94.1\% \pm 1.3$ on Sequential/Global, showing that learning representations across time and with cross-modality. Second, traditional methods of learning styles usually model behavior and artifacts separately or through late fusion, the elimination of cross-modal attention caused an accuracy reduction of 7.4%, thus assessing the value added by early alignment. Third, end-to-end privacy and deployment considerations are seldom discussed; CAFNet demonstrates that high levels of privacy ($\epsilon = 0.5$) can be accommodated and still obtain high predictive accuracy (federated accuracy = $89.3 - 1.7$) and that the risk of early failure can be detected with significant intervention time: CAFNet obtained high accuracy $88.9\% \pm 1.6$ and AUC-ROC = 0.941 ± 0.023 at Week-1 and a false negative rate of 14.

In spite of these strengths, the methodology has limitations that limit the scope of interpretation. To begin with, despite the assessment of the results on three datasets and different modalities, the ground-truth FLSM labels have been obtained through different mechanisms (ILS questionnaire, expert annotation, self-report/peer validation), and thus this can cause label heterogeneity and measurement noise; therefore, high performance can be due to predictive consistency between mixed label sources, and not one gold-standard regime. Second, it is a predictive, not a causal design: although ablation experiments suggest the role of components, the study does not prove that the learner types identified by CAFNet outputs causally affect particular learning outcomes, but also does not test whether intervention effects are caused by CAFNet outputs. Third, some results depend on the availability of and quality of code artefacts (AST extraction) and forum content, institutions that capture artefacts less well might not replicate the same scale of gains as seen in the configuration of behavioral only that declines to $76.1\% \pm 2.7$ accuracy ($\Delta = -15.6\%$). Fourth, computational cost is higher than classical baselines (centralized training 47.3 ± 3.2 min; federated 128.6 ± 8.7 min).

Although CAFNet has been performing well empirically, a number of limitations and implementation issues in the real world are worth considering. The framework relies on the quality of AST and forum information which was central to performance improvements, especially in the case of Sequential/Global classification, where AST characteristics displayed a strong correlation (e.g., branching factor $r = 0.72$, $p = 0.001$); in a scenario with no organized code repositories or discussion forums, productivity can decline, as seen in the 4.5% improvement on the case when AST features were not used. Also, the cost of computation places practical limits, with centralized training taking 47.3 min and federated training 128.6, and this may limit the frequency of retraining of real-time learning systems particularly with federated communication overhead. Lastly, despite showing good generalization on programming datasets (accuracy $> 90\%$, AUC-ROC ≈ 0.95), the application to non-programming datasets is still limited since its highest improvements are made based on code-structure modelling, which suggests that domain specific adaptations would be required to larger educational settings.

Regarding generalizability, dataset-specific findings show strong cross-educational transfer in both OULAD-CS ($90.4\% \pm 1.8$), CodeTraverse ($92.8\% \pm 1.5$), and EdForum ($91.5\% \pm 1.6$), with an average accuracy that is consistently high (AUC-ROC: 0.938–0.954) and dimensional performance is stable, which points to the fact that it is not dataset-specific. Nonetheless, the interpretation of generalization must be limited by the context of the study: all datasets are related to computer science or programming-related learning with traceable interaction logs and code artefacts; it may be necessary to adapt the architecture to non-programming tasks (e.g., by replacing AST signals with other structured artefacts). Similarly, the federated efficiency results show scalability up to larger institutional consortia with more communication rounds (RTA from 623 ± 41 at $K = 3$ to $1,142 \pm 78$ at $K = 10$), indicating that larger-scale deployment is possible, but also indicating that engineering of such large-scale deployments will have to be careful in terms of bandwidth, client availability, and privacy accounting. The net effect of the discussion is that the cross-modal temporal modelling of CAFNet is the core contributor of better performance, especially to the hitherto poor Sequential/Global dimension whilst the federated DP outcomes indicate that these benefits can be maintained within the context of realistic privacy and governance pressures.

5.1. Practical Implementation and Real-World Applications

The behavioral and language analytics model proposed as based on machine learning can be applied in a variety of practical ways to the contemporary digital learning setting. With the ever growing use of computer-based education platforms, intelligent systems that operate using data are becoming more and more mandatory to keep track of the learning process, enhance learning, and facilitate personalized education. Analytical models powered by AI have the potential to convert bulk of educational interaction data into practical information to instructors and

educational centers [1].

5.2. Use in Educational Systems

The proposed model can be implemented within learning management systems (LMS) to analyze logs of interactions with students and discuss their content in real-time. The system will be able to identify learning patterns, struggling students and help with early academic intervention by combining indication of behavior changes with a textual analysis. This feature can help instructors to keep track of thousands of learners at once in a large-scale setting like that of MOOCs and detect trends in engagement and anticipate what is likely to happen to such students later on.

5.3. AI-Assisted Education

Intelligent educational support systems also are possible through the combination of behavioral and language cues. Learners can also receive automated feedback provided by AI-assisted platforms that monitor the learning behavior patterns and the participation patterns as well as conceptual understanding provided in the forum discussions. As well, student engagement tracking can be facilitated by the framework, as it can provide opportunities to determine when students are not participating, when they are confused, or simply have lower levels of engagement, thus allowing the instructors to react in real-time.

5.4. Computer Science Education Software Learning and Programming Education

The proposed framework has distinct opportunities to be implemented in the programming education. The model can be applied to the online coding platforms to analyze the pattern of code submission, debugging attempts, and discussion activities to assess the learner progress. It is possible to support coding skill development analytics and recommend specific learning materials with this information or facilitate automated mentoring tools that can guide learners in solving programming problems.

In general, a combination of behavioral analytics and natural language mining can provide a scalable intelligent educational system that could serve an adaptive learning environment and can serve a large number of students, in addition to increasing learning outcomes [1].

6. Conclusions

The current work has presented the CAFNet, a multimodal learning analytics system, a model that is used to model the computer based development of learning through the combination of temporally structured behavioral cues and syntax-sensitive linguistic and code representations. In contrast to the traditional FLSM-based methods which are mainly dependent on the fixed behavioral characteristics, the suggested framework represents the dynamic interaction of the learner navigation patterns, discussion activity, and code evolution in programming environments. The framework offers a more detailed picture of engagement and learning preferences based on the combination of behavioral analytics, natural language processing, and code-structure modeling.

The experimental analysis of three heterogeneous datasets with 14,308 learners proves that the CAFNet obtains significant performance increases compared to the traditional machine learning and deep learning baselines. The model demonstrates a high predictive value especially on the Sequential/Global learning dimension that has always been challenging to determine accurately in the past. Moreover, by applying federated supervised contrastive training, the framework can achieve high predictive accuracy and data privacy even with strict different privacy restrictions ($\epsilon = 0.5$), which proves the effectiveness and scalability of the proposed solution.

The results indicate that programming education is fundamentally multimodal and time-based in nature, and analytical systems are needed to model the behavior of learners, their discourse, and code artifacts simultaneously. In real life scenarios, the suggested framework can facilitate adaptive learners programming education and initiation of struggling learners at an early stage, different learning pathways, as well as large scale collaborative learning analytics in institutions. Future studies can build upon this study by adding more educational datasets, testing the work in real-time in learning systems and enhancing the interpretability of the model to assist instructors and educational decision-makers.

Author Contributions

Conceptualization, D.H. and Y.M.; methodology, D.H.; software, D.H. and Y.M.; validation, D.H., Y.M. and B.S.; formal analysis, D.H.; investigation, D.H. and Y.M.; resources, B.S.; data curation, D.H. and Y.M.; writing—original draft preparation, D.H.; writing—review and editing, Y.M. and B.S.; visualization, Y.M.; supervision, B.S.; project administration, D.H.; funding acquisition, D.H. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of Kyrgyz State University named after I. Arabaev.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The data supporting the findings of this study are not publicly available due to institutional privacy and ethical restrictions related to learner behavioral records, programming submissions, and educational interaction logs. Aggregated results are available from the corresponding author upon reasonable request.

Acknowledgments

The authors acknowledge the administrative and technical support provided by the participating institutions during data preparation and model evaluation.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Babulak, E. *AI-Driven Approaches for Fully Automated Smart Engineering*; IGI Global Scientific Publishing: Hershey, PA, USA, 2026.
2. Babulak, E. *Educational AI Humanoid Computing Devices for Cyber Nomads*; IGI Global Scientific Publishing: Hershey, PA, USA, 2025.
3. Ferreira, R.; Freitas, E.; Cabral, L.; et al. Words of Wisdom: A Journey through the Realm of Natural Language Processing for Learning Analytics—A Systematic Literature Review. *J. Learn. Anal.* **2024**, *11*, 82–105. [[Cross-Ref](#)]
4. Shaik, T.; Tao, X.; Dann, C.; et al. Sentiment Analysis and Opinion Mining on Educational Data: A Survey. *Nat. Lang. Process.* **2023**, *2*, 100003. [[CrossRef](#)]
5. Clavié, B.; Gal, K. EduBERT: Pretrained Deep Language Models for Learning Analytics. *arXiv preprint* **2019**, *arXiv:1912.00690*.
6. Li, M.; Ge, M.; Zhao, H.; et al. Modeling and Analysis of Learners' Emotions and Behaviors Based on Online Forum Texts. *Comput. Intell. Neurosci.* **2022**, *2022*, 9696422. [[CrossRef](#)]
7. Giannakos, M.N.; Sharma, K.; Pappas, I.O.; et al. Multimodal Data as a Means to Understand the Learning Experience. *Int. J. Inf. Manag.* **2019**, *48*, 108–119. [[CrossRef](#)]
8. Rubio, M.A. Automated Prediction of Novice Programmer Performance Using Programming Trajectories. *Artif. Intell. Educ.* **2020**, *12164*, 268–272.
9. Maphalala, M.C.; Mkhasibe, R.G.; Mncube, D.W. Exploring the Roles of AI-Powered E-Tutors in Enhancing Self-Directed Learning in Open Distance E-Learning Courses. *Interdiscip. J. Educ. Res.* **2025**, *7*, a12. [[CrossRef](#)]

10. Liu, S.; Abadia, R.; Strambi, A.; et al. Leveraging Student Confusion in Online Forum Posts to Enhance Student Engagement Using Text-Based Learning Analytics. In Proceedings of the ASCILITE 2025, Adelaide, Australia, 30 November–3 December 2025; pp. 221–230.
11. Yee, M.; Roy, A.; Perdue, M.; et al. AI-Assisted Analysis of Content, Structure, and Sentiment in MOOC Discussion Forums. *Front. Educ.* **2023**, *8*, 1250846. [CrossRef]
12. Friedman, A.; Beasley, Z. Using Textual Analysis to Examine Student Engagement in Online Undergraduate Science Education. *J. Stat. Data Sci. Educ.* **2024**, 1–11. [CrossRef]
13. Shoab, M.; Sayed, N.; Singh, J.; et al. AI student success predictor: Enhancing personalized learning in campus management systems. *Comput. Hum. Behav.* **2024**, *158*, 108301. [CrossRef]
14. Mehenaoui, Z.; Lafifi, Y.; Zemmouri, L. Learning Behavior Analysis to Identify Learner's Learning Style Based on Machine Learning Techniques. *J. Univ. Comput. Sci.* **2022**, *28*, 1193–1220. [CrossRef]
15. Awadh, W.A.; Sulaiman, R.B.; Mahmoud, M.A. Aspect-Based Sentiment Analysis in MOOCs: A Systematic Literature Review Introducing the MASC-MEF Framework. *J. King Saud Univ. Comput. Inf. Sci.* **2025**, *37*, 2. [Cross-Ref]
16. Angeioplastis, A.; Aliprantis, J.; Konstantakis, M.; et al. The Learning Style Decoder: FSLSM-Guided Behavior Mapping Meets Deep Neural Prediction in LMS Settings. *Computers* **2025**, *14*, 377. [CrossRef]
17. Hashemi, S.E.; Gholian-Jouybari, F.; Hajiaghahi-Keshteli, M. A Fuzzy C-Means Algorithm for Optimizing Data Clustering. *Expert Syst. Appl.* **2023**, *227*, 120377.
18. Lestari, A.; Lawi, A.; Thamrin, S.A.; et al. Automated Detection of Learning Styles Using Online Activities and Model Indicators. *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15*.
19. Ait Daoud, M.; Namir, A.; Talbi, M. FSLSM-Based Analysis of Student Performance Information in a Blended Learning Course Using Moodle LMS. *Open Inf. Sci.* **2024**, *8*, 20220163.
20. Pineda-Arizmendi, M.G.; Hernández-Castañeda, Á.; García-Hernández, R.A.; et al. Automatic Identification of Learning Styles through Behavioral Patterns. In Proceedings of the Mexican Conference on Pattern Recognition, Tepic, Mexico, 21–24 June 2023; pp. 79–88.
21. Najem, K.; Seghroucheni, Y.Z.; Ziti, S. Behavioral Clustering for Adaptive Learning: A Data-Driven Alternative to Static Learning Style Models. *Int. J. Inf. Educ. Technol.* **2026**, *16*, 196–204. [CrossRef]
22. Alzamzami, F. Towards Domain-Independent Multi-Lingual-Dialectal Online Social Behavior Modeling. PhD Thesis, University of Ottawa, Ottawa, ON, Canada, 2024. [CrossRef]
23. Hussain, T.; Yu, L.; Asim, M.; et al. Enhancing E-Learning Adaptability with Automated Learning Style Identification and Sentiment Analysis: A Hybrid Deep Learning Approach for Smart Education. *Information* **2024**, *15*, 277. [CrossRef]
24. Ezzaim, A.; Dahbi, A.; Haidine, A.; et al. Development, Implementation, and Evaluation of a Machine Learning-Based Multi-Factor Adaptive E-Learning System. *IAENG Int. J. Comput. Sci.* **2024**, *51*, 1250–1271.
25. Hananto, A.R.; Musdholifah, A.; Wardoyo, R. Utilizing Support Vector Machine and Dimensionality Reduction to Identify Student Learning Styles within the Felder-Silverman Model. *J. Appl. Data Sci.* **2024**, *5*, 1495–1507.
26. Petrov, P.; Milev, V.; Byalmarkova, P. Applying a Data Classification Model of Learning Style Prediction. In *System Design in Software Engineering*; Springer: Cham, Switzerland, 2024; pp. 116–125.
27. Muhammad, B.A.; Jianping, W.; Gao, G.; et al. A Fuzzy C-Means Algorithm to Detect Learning Styles in Online Learning Environment. *J. Netw. Netw. Appl.* **2024**, *4*, 39–47.
28. Essa, S.G.; Celik, T.; Human-Hendricks, N.E. Personalized Adaptive Learning Technologies Based on Machine Learning Techniques to Identify Learning Styles: A Systematic Literature Review. *IEEE Access* **2023**, *11*, 48392–48409. [CrossRef]
29. Muhammad, B.A.; Qi, C.; Wu, Z.; et al. An Evolving Learning Style Detection Approach for Online Education Using Bipartite Graph Embedding. *Appl. Soft Comput.* **2024**, *152*, 111230.
30. Giamphy, E.; Guillaume, J.L.; Doucet, A.; et al. A Survey on Bipartite Graphs Embedding. *Soc. Netw. Anal. Min.* **2023**, *13*, 54. [CrossRef]
31. Manarat, P.; Tuarob, S.; Pongpaichet, S. Artificial intelligence in computer programming education: A systematic literature review. *Comput. Educ.: Artif. Intell.* **2025**, *8*, 100403. [CrossRef]
32. Pires, J.P.; Correia, F.B.; Gomes, A.; et al. Predicting student performance in introductory programming courses. *Computers* **2024**, *13*, 219. [CrossRef]
33. Zhang, V.Y.F.; Jeffries, B.; Koprinska, I. A machine learning approach for predicting student progress in online programming education. *Int. J. Artif. Intell. Educ.* **2025**, *35*, 3614–3644. [CrossRef]
34. Alonso-Fernández, C.; Calvo-Morata, A.; Freire, M.; et al. Applications of data science to game learning analytics data: A systematic literature review. *Comput. Educ.* **2019**, *141*, 103612.

35. Alonso-Fernández, C.; Cano, A.R.; Calvo-Morata, A.; et al. Lessons learned applying learning analytics to assess serious games. *Comput. Hum. Behav.* **2019**, *99*, 301–309.
36. Llanos, J.M.; Bucheli, V.A.; Restrepo-Calle, F. Early prediction of student performance in CS1 programming courses. *PeerJ Comput. Sci.* **2023**, *9*, e1655. [[CrossRef](#)]
37. Suárez, C.G.H.; Llanos, J.; Bucheli, V.A. Predicting the final grade using a machine learning regression model: Insights from fifty percent of course grades in CS1 courses. *PeerJ Comput. Sci.* **2023**, *9*, e1689. [[CrossRef](#)]
38. Gutiérrez-Benítez, R.; Vásquez-Guerra, A.; Carrasco-Sáez, J.L. Who Fails and Why: An Analysis of Student Trajectories and the Prediction of Undergraduate Performance in Programming Courses. Preprints **2026**, 2026031704. [[CrossRef](#)]
39. Chen, J.; Zhou, X.; Yao, J.; et al. Application of Machine Learning in Higher Education to Predict Students' Performance, Learning Engagement and Self-Efficacy: A Systematic Literature Review. *Asian Educ. Dev. Stud.* **2025**, *14*, 205–240. [[CrossRef](#)]
40. Hafdi, Z.S.; El Kafhali, S. A Comparative Evaluation of Machine Learning Methods for Predicting Student Outcomes in Coding Courses. *AppliedMath* **2025**, *5*, 75. [[CrossRef](#)]
41. Alshammari, M.T. Machine Learning-Enabled Personalization of Programming Learning Feedback. *Int. J. Adv. Comput. Sci. Appl.* **2025**, *16*. [[CrossRef](#)]
42. Choi, W.-C.; Lam, C.-T.; Pang, P.C.; et al. A Systematic Literature Review of Explainable Artificial Intelligence (XAI) for Interpreting Student Performance Prediction in Computer Science and STEM Education. In Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V.1, Nijmegen, The Netherlands, 27 June–2 July 2025; pp. 221–227. [[CrossRef](#)]
43. Jiang, Z.; Zhang, Z. From Black Box to Transparency: Enhancing Automated Interpreting Assessment with Explainable AI in College Classrooms. *Res. Methods Appl. Linguist.* **2025**, *4*, 100237. [[CrossRef](#)]
44. Lu, M.; Hu, Z. Leveraging Multimodal Information for Web Front-End Development Instruction: Analyzing Effects on Cognitive Behavior, Interaction, and Persistent Learning. *Information* **2025**, *16*, 734. [[CrossRef](#)]
45. Barbierato, E.; Gatti, A. The Challenges of Machine Learning: A Critical Review. *Electronics* **2024**, *13*, 416. [[CrossRef](#)]
46. Mu, S.; Cui, M.; Huang, X. Multimodal data fusion in learning analytics: A systematic review. *Sensors* **2020**, *20*, 6856. [[CrossRef](#)]
47. Mangaroska, K.; Sharma, K.; Gasevic, D.; et al. Multimodal learning analytics to inform learning design: Lessons learned from computing education. *J. Learn. Anal.* **2020**, *7*, 79–97.
48. Di Mitri, D.; Schneider, J.; Specht, M.; et al. From Signals to Knowledge: A Conceptual Model for Multimodal Learning Analytics. *J. Comput. Assist. Learn.* **2018**, *34*, 338–349.
49. Hennessy, S. Analysing educational dialogue around shared artefacts in technology-mediated contexts: A new coding framework. *Classr. Discourse* **2025**, *16*, 172–206. [[CrossRef](#)]
50. Ouhaichi, H.; Spikol, D.; Vogel, B. Research Trends in Multimodal Learning Analytics: A Systematic Mapping Study. *Comput. Educ. Artif. Intell.* **2023**, *4*, 100136.
51. Xu, W.; Wu, Y.; Ouyang, F. Multimodal Learning Analytics of Collaborative Patterns During Pair Programming in Higher Education. *Int. J. Educ. Technol. High. Educ.* **2023**, *20*, 8. [[CrossRef](#)]
52. Prinsloo, P.; Slade, S.; Khalil, M. Multimodal Learning Analytics—In-between Student Privacy and Encroachment: A Systematic Review. *Br. J. Educ. Technol.* **2023**, *54*, 1566–1586. [[CrossRef](#)]
53. Giannakos, M.; Cukurova, M. The Role of Learning Theory in Multimodal Learning Analytics. *Br. J. Educ. Technol.* **2023**, *54*, 1246–1267. [[CrossRef](#)]
54. Sellberg, C.; Sharma, A. Toward Multimodal Learning Analytics in Simulation-Based Collaborative Learning: A Design Ethnography of Maritime Training. *Int. J. Comput.-Support. Collab. Learn.* **2025**, *20*, 201–221. [[CrossRef](#)]
55. Acosta, H.; Lee, S.; Mott, B.; et al. Multimodal Learning Analytics for Predicting Student Collaboration Satisfaction in Collaborative Game-Based Learning. In Proceedings of the 17th International Conference on Educational Data Mining, Atlanta, GA, USA, July 2024.
56. Yan, L.; Echeverria, V.; Jin, Y.; et al. Evidence-Based Multimodal Learning Analytics for Feedback and Reflection in Collaborative Learning. *Br. J. Educ. Technol.* **2024**, *55*, 1900–1925. [[CrossRef](#)]
57. Martinez-Maldonado, R.; Echeverria, V.; Fernandez-Nieto, G.; et al. Lessons Learnt from a Multimodal Learning Analytics Deployment In-the-Wild. *ACM Trans. Comput.-Hum. Interact.* **2023**, *31*, 1–41. [[CrossRef](#)]
58. Chango, W.; Lara, J.A.; Cerezo, R.; et al. A Review on Data Fusion in Multimodal Learning Analytics and Educational Data Mining. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2022**, *12*, e1458. [[CrossRef](#)]

59. Deng, J.H.; Zhao, Y. A Literature Review of Data-Driven Multimodal Learning Analytics in Education Based on CiteSpace. In Proceedings of the 2022 5th International Conference on Education Technology Management, Lincoln, UK, 16–18 December 2022; pp. 390–397. [[CrossRef](#)]
60. Bhatti, A.; Angkan, P.; Behinaein, B.; et al. CLARE: Cognitive Load Assessment in Real-Time with Multimodal Data. *IEEE Trans. Cogn. Dev. Syst.* **2025**, *17*, 1337–1349. [[CrossRef](#)]



Copyright © 2026 by the author(s). Published by UK Scientific Publishing Limited. This is an open access article under the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Publisher's Note: The views, opinions, and information presented in all publications are the sole responsibility of the respective authors and contributors, and do not necessarily reflect the views of UK Scientific Publishing Limited and/or its editors. UK Scientific Publishing Limited and/or its editors hereby disclaim any liability for any harm or damage to individuals or property arising from the implementation of ideas, methods, instructions, or products mentioned in the content.