Sequence-Aware Learning Analytics for Early Identification of At-Risk Academic Trajectories in Higher Education Using Transformer Models

Everlyne Fradia Akello[1]; Onuh Matthew Ijiga[2]; Idoko Peter Idoko[3]

[1]The Gladys W. and David H. Patton College of Education, Ohio University, Athens, Ohio, USA
[2]Department of Physics, Joseph Sarwuan Tarka University, Makurdi, Nigeria
[3]Department of Electrical/ Electronic Engineering, University of Ibadan, Nigeria

Publication Date: 2026/01/16

Abstract: Early identification of students at risk of academic underperformance remains a persistent challenge in higher education, particularly in learning environments characterized by complex, temporally evolving patterns of engagement and assessment. Conventional learning analytics approaches typically rely on static or weakly temporal indicators, limiting their ability to detect emerging risk at early stages of an academic term. This study proposes a sequence-aware learning analytics framework that leverages transformer-based models to represent student academic trajectories as ordered sequences of learning events derived from learning management systems and student information systems. The framework integrates heterogeneous behavioral, temporal, and performance signals and applies self-attention mechanisms to capture long-range dependencies and evolving risk patterns. Using a supervised predictive modeling design with rolling-window and early-prediction evaluation protocols, the proposed approach is assessed against traditional machine learning and recurrent neural network baselines. Results demonstrate that transformer models achieve superior predictive performance, earlier risk identification, and greater stability across academic terms and cohorts. Attention-based interpretability further reveals meaningful progression patterns associated with academic disengagement and performance decline. The findings underscore the value of sequence-aware modeling for enhancing institutional early-alert systems and supporting proactive, personalized academic interventions. This study contributes to both learning analytics theory and practice by establishing transformer-based sequence modeling as a robust foundation for early academic risk detection and student success initiatives in higher education.

Keywords: *Learning Analytics; Academic Risk Prediction; Sequence-Aware Modeling; Transformer Models; Early Warning Systems; Student Success; Higher Education.*

How to Cite: Everlyne Fradia Akello; Onuh Matthew Ijiga; Idoko Peter Idoko (2026) Sequence-Aware Learning Analytics for Early Identification of At-Risk Academic Trajectories in Higher Education Using Transformer Models. *International Journal of Innovative Science and Research Technology*, 11(1), 818-838. https://doi.org/10.38124/ijisrt/26jan563

I. INTRODUCTION

➢ *Background and Motivation*

The widespread adoption of Learning Management Systems (LMS) and Student Information Systems (SIS) across higher education institutions has transformed how student learning behaviors are captured and analyzed. These platforms generate fine-grained, time-stamped records of academic activity, including logins, content access, assessment submissions, feedback cycles, and grade progression. Unlike traditional institutional datasets that relied on end-of-semester outcomes or static demographic attributes, LMS and SIS logs provide continuous, high-resolution temporal data that reflect how students engage with learning resources over time. This shift has positioned learning analytics as a central instrument for understanding student trajectories rather than isolated performance snapshots (Handbook of Learning Analytics).

Despite this data richness, many existing academic risk prediction systems still rely on static or aggregated features, such as cumulative GPA, total clicks, or average attendance rates. While such indicators are informative at a coarse level, they obscure the ordering, timing, and evolution of learning behaviors that often signal emerging disengagement. Research in educational data mining has shown that two students with identical aggregate engagement levels may follow fundamentally different academic paths depending on when and how that engagement unfolds. Static analytics therefore struggle to capture non-linear decline, delayed recovery, or sudden shifts in learning behavior that precede academic difficulty (Baker & Inventado 2014; Maduabuci et al., 2023).

The limitations of aggregate approaches are particularly consequential for early warning systems. Early identification of at-risk students is most effective when interventions occur before performance deterioration becomes irreversible. However, models built on cumulative indicators tend to detect risk only after grades or engagement have already declined significantly. Longitudinal studies of student retention demonstrate that behavioral changes often emerge weeks before formal academic failure is visible, highlighting the need for analytics that are sensitive to temporal dynamics rather than summary statistics (Lakkaraju et al., 2015; Permata et al., 2025).

Sequence-aware learning analytics address this challenge by explicitly modeling the ordered progression of student actions, assessments, and outcomes. By treating academic histories as evolving sequences rather than fixed vectors, these approaches can capture patterns such as procrastination cycles, irregular engagement rhythms, or cascading assessment failures. Prior work in knowledge tracing and sequential student modeling has shown that temporal representations significantly improve predictive accuracy and provide earlier signals of risk compared to static baselines (Piech et al.,2015; Idogho et al., 2025).

Recent advances in transformer architectures further strengthen the case for sequence-aware analytics. Self-attention mechanisms enable models to learn long-range dependencies across extended academic timelines without the vanishing-gradient limitations of recurrent methods. In educational contexts, this allows critical early behaviors to be weighted appropriately even when their consequences manifest much later in a semester. As a result, transformer-based sequence modeling offers a principled foundation for timely, interpretable, and scalable identification of at-risk academic trajectories in higher education (Vaswani et al., 2017; Idogho et al., 2025; Permata et al., 2026).

Figure 1 illustrates an improved instructional flowchart that integrates mathematical abstractions with core physics concepts to support progressive learning. The sequence begins with foundational ideas in abstraction and linear kinematics, then advances through angular kinematics and the formal connections between linear and angular motion. Each block is color-coded to distinguish disciplinary context while preserving logical continuity across lectures. The pathway culminates in statics basics, reinforcing how mathematical and kinematic principles underpin equilibrium analysis in physics.

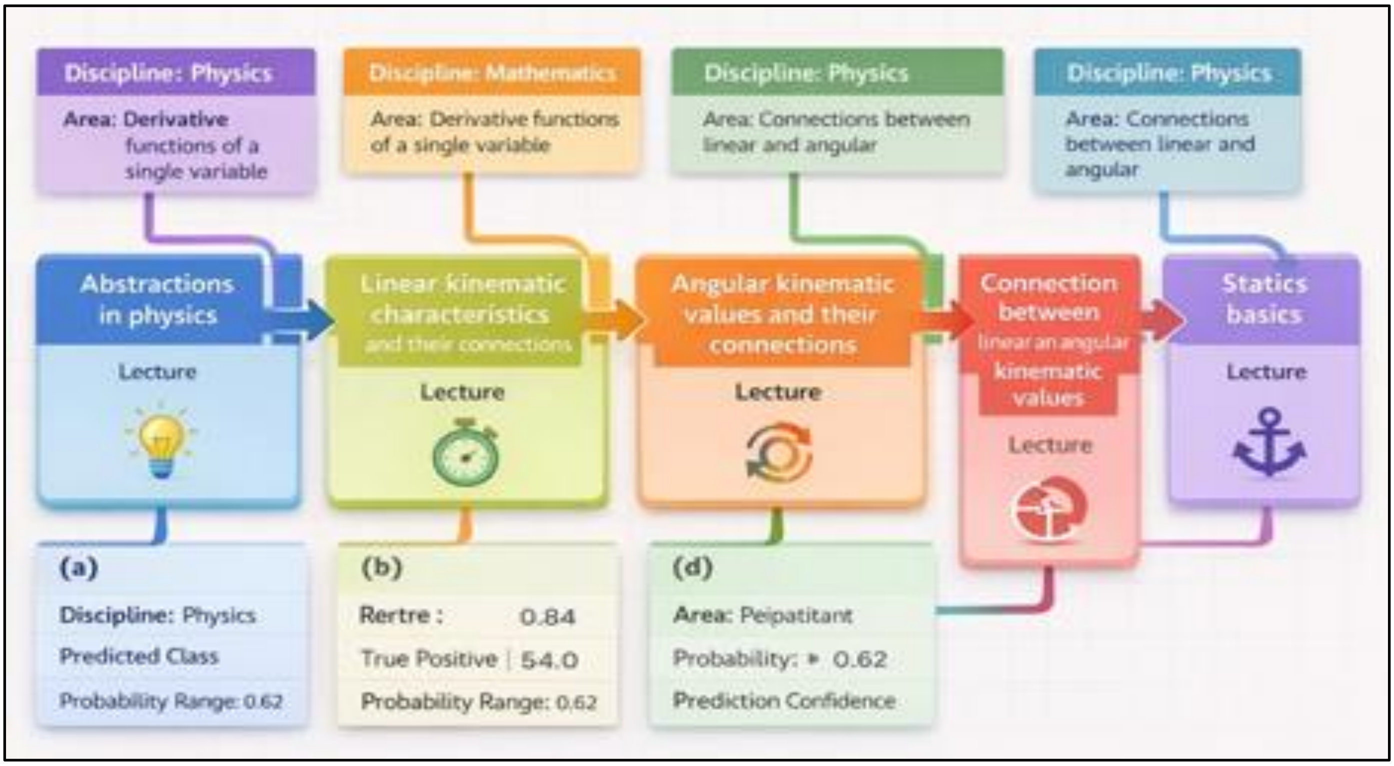

Fig 1 Conceptual Learning Pathway Linking Mathematical Foundations to Kinematics and Statics in Physics

➢ *Problem Statement*

Despite the growing availability of fine-grained LMS and SIS data, most operational learning analytics systems continue to rely on traditional predictive models that reduce student behavior to static or weakly temporal summaries. Common approaches such as logistic regression, decision trees, or feature-engineered machine learning models typically aggregate student activity over fixed windows, ignoring the ordering and spacing of events. This abstraction limits their ability to capture long-range dependencies, where early learning behaviors exert delayed effects on later academic outcomes. Empirical studies in educational data mining demonstrate that such models often miss gradual disengagement patterns and non-linear behavioral shifts that unfold across an academic term (Siemens & Baker, 2012; Lakkaraju et al., 2015; Maduabuchi et al., 2023).

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

A further challenge arises from the inherently heterogeneous and irregular nature of educational event sequences. Student interactions with digital learning environments vary widely in frequency, modality, and timing. LMS logs combine sparse assessment submissions, bursty interaction episodes, prolonged inactivity periods, and asynchronous feedback cycles. Traditional sequence modeling approaches, including fixed-order Markov models or recurrent neural networks with uniform time steps, struggle to represent these irregular temporal structures effectively. Research on clickstream and trace data shows that sparsity and uneven temporal gaps degrade model performance and lead to unstable predictions when conventional techniques are applied without explicit temporal encoding (Xu & Recker, 2012; Kovanović et al., 2015; Ayoola et al., 2024).

These modeling limitations have direct consequences for early warning systems aimed at supporting student retention. When temporal dynamics are inadequately represented, risk prediction models tend to rely on lagging indicators such as cumulative grades or total engagement counts. As a result, academic risk is often detected only after disengagement has become entrenched or performance decline is already severe. Longitudinal analyses of student success consistently show that disengagement signals frequently emerge weeks before formal failure indicators, underscoring the cost of late or inaccurate identification (Lakkaraju et al., 2015; Ijiga et al., 2024).

The problem is further compounded by the difficulty of distinguishing transient fluctuations in engagement from sustained downward trajectories. Without sequence-aware representations, predictive systems may misclassify short-term inactivity as high risk or overlook progressive deterioration masked by aggregate stability. Prior work in sequential student modeling demonstrates that preserving temporal order and contextual relationships between events substantially improves both predictive accuracy and the timeliness of risk detection (Piech et al., 2015; Ijiga et al., 2024). However, these insights have yet to be systematically integrated into institutional-scale early alert frameworks.

Taken together, these challenges reveal a fundamental gap between the temporal richness of modern educational data and the capabilities of widely deployed predictive models. Addressing this gap requires learning analytics approaches that can robustly model heterogeneous, irregular sequences while capturing long-range dependencies in student behavior. Without such sequence-aware methods, higher education institutions face an ongoing risk of delayed interventions, misallocated support resources, and missed opportunities to prevent academic disengagement and performance decline.

Figure 2 presents an integrated block-diagram representation of the early warning system, illustrating the sequential flow from observation and forecasting to warning communication and decision-making. Each stage highlights a critical analytical or behavioral bridge, supported by continuous evaluation across the system. The lower blocks align these stages with the four pillars of the Early Warnings for All framework, emphasizing institutional preparedness and response capacity. Collectively, the figure demonstrates how technical forecasts are translated into actionable outcomes through coordinated communication and human decision processes.

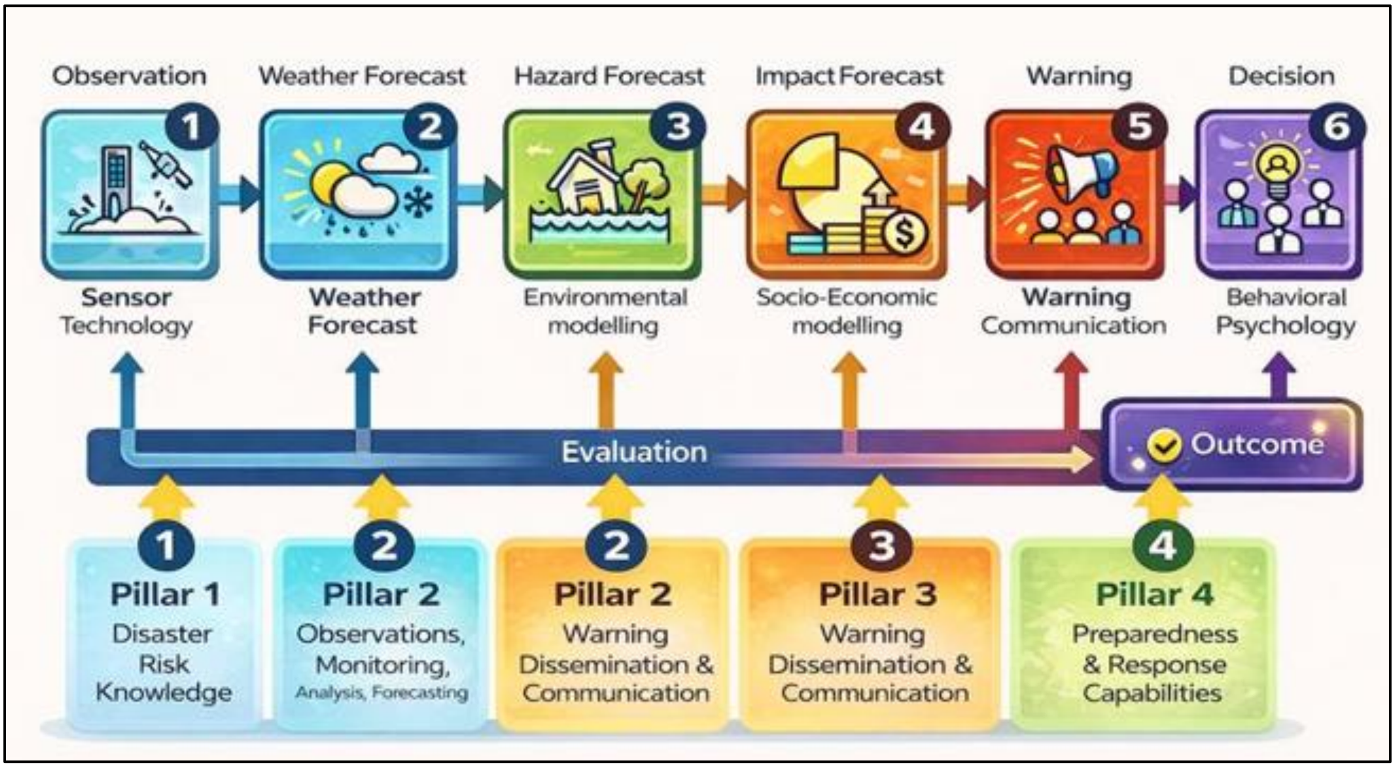

Fig 2 Integrated Early Warning System Pathway from Observation to Response

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

➢ *Research Objectives*

The primary objective of this study is to design a transformer-based learning analytics framework capable of representing student academic trajectories as temporally ordered sequences. Rather than treating learning behavior as static aggregates, the framework aims to encode the progression, timing, and contextual relationships among learning events captured from learning management systems and student information systems. By modeling academic histories as sequences, the study seeks to reflect how engagement patterns, assessment outcomes, and behavioral shifts evolve across an academic term.

A second objective is to evaluate the effectiveness of attention-based architectures in the early identification of at-risk students. The study focuses on determining whether self-attention mechanisms can detect subtle but persistent indicators of disengagement or performance decline at earlier stages than conventional predictive systems. Emphasis is placed on assessing how early in the academic timeline reliable risk signals can be identified without sacrificing predictive accuracy or stability.

The final objective is to conduct a systematic comparison between sequence-aware transformer models and established predictive approaches, including traditional machine learning methods and recurrent neural networks. This comparison examines differences in predictive performance, robustness to irregular and sparse event sequences, and sensitivity to long-range behavioral dependencies. Through this comparative analysis, the study aims to clarify the added value of transformer-based models for academic risk prediction and inform their practical adoption in higher education learning analytics systems.

➢ *Research Questions and Hypotheses*

This study is guided by the need to understand whether sequence-aware modeling can meaningfully improve the identification of academic risk in higher education settings. The first research question examines how effectively transformer models capture temporal learning patterns that are indicative of emerging academic difficulty. Specifically, it investigates the extent to which self-attention mechanisms can learn dependencies across ordered learning events, such as engagement fluctuations, assessment performance trends, and periods of inactivity, that collectively signal elevated risk.

The second research question focuses on the timing of risk detection within an academic term. It seeks to determine how early at-risk academic trajectories can be identified when student behavior is modeled as a sequence rather than as aggregated indicators. This question emphasizes the practical value of sequence modeling by evaluating whether reliable predictions can be generated at early stages of the semester, when institutional interventions are most likely to influence outcomes.

Based on these questions, the study advances the hypothesis that attention-based sequence models outperform static and recurrent baseline approaches in early risk prediction. It is hypothesized that transformer architectures, through their ability to model long-range temporal dependencies and selectively weight salient learning events, achieve higher predictive accuracy and earlier detection of academic risk compared to traditional machine learning models and recurrent neural networks.

➢ *Significance of the Study*

This study contributes to learning analytics theory by advancing sequence-aware risk modeling as a principled alternative to static and weakly temporal approaches. By framing student academic trajectories as ordered sequences and leveraging attention-based mechanisms, the study deepens theoretical understanding of how learning behaviors unfold over time and how long-range dependencies influence academic outcomes. The work clarifies the limits of aggregate indicators and offers a formal basis for modeling temporal structure, event salience, and progression dynamics within educational data. In doing so, it extends the conceptual toolkit of learning analytics toward representations that are more faithful to the lived realities of student learning processes.

Beyond theoretical contributions, the study carries direct practical implications for academic advising and student retention strategies. Earlier and more reliable identification of at-risk trajectories enables advisors and support staff to intervene proactively rather than reactively, aligning outreach with moments of highest leverage in the academic term. Sequence-aware predictions can inform differentiated advising actions by distinguishing transient disengagement from sustained decline, reducing false alarms and improving the targeting of limited institutional resources. At the operational level, the findings support the integration of temporally sensitive analytics into early alert systems, degree planning tools, and learning support platforms.

The study also aligns closely with data-driven student success and equity initiatives in higher education. By improving early detection accuracy, sequence-aware models can help institutions mitigate structural disadvantages that disproportionately affect students who experience delayed feedback or cumulative academic pressure. Timely, evidence-based interventions informed by longitudinal behavior patterns support more equitable outcomes by addressing risk before it manifests as failure or withdrawal. In this sense, the study reinforces institutional commitments to inclusive student success by demonstrating how advanced analytics can be responsibly applied to enhance persistence, completion, and overall educational equity.

II. LITERATURE REVIEW

➢ *Learning Analytics and Academic Risk Prediction*

Predictive learning analytics has become a central component of higher education decision-making as institutions seek to improve student retention, progression, and completion rates. At its core, learning analytics involves the measurement, collection, analysis, and reporting of data about learners and their contexts for the purpose of

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

understanding and optimizing learning and the environments in which it occurs. In predictive applications, historical and ongoing student data are used to estimate the likelihood of adverse academic outcomes such as course failure, dropout, or delayed graduation. These approaches have evolved alongside the expansion of digital learning infrastructures, which provide continuous streams of behavioral and performance data suitable for modeling academic risk (Siemens & Baker, 2012; Onuh et al., 2024).

Early predictive systems in higher education relied primarily on demographic attributes and prior academic achievement, including entry qualifications and cumulative grade point averages. While these variables remain informative, research has shown that models based solely on static background characteristics are limited in their ability to explain or anticipate changes in student performance over time. As a result, predictive learning analytics has increasingly shifted toward the use of process-oriented data derived from learning management systems, enabling more dynamic representations of student behavior throughout an academic term (Ferguson, 2012).

Grades constitute one of the most widely used indicators of academic risk. Low or declining assessment scores, missed submissions, and poor performance in early coursework are consistently associated with increased likelihood of course failure or withdrawal. However, grades are often lagging indicators, reflecting difficulties only after learning challenges have already materialized. Consequently, exclusive reliance on performance outcomes can delay risk detection and reduce the effectiveness of early intervention strategies (Arnold & Pistilli, 2012).

Engagement metrics derived from LMS activity logs form another major class of academic risk indicators. These include frequency of logins, time spent on learning resources, participation in online discussions, and interaction with instructional materials. Studies have demonstrated that sustained reductions in engagement, irregular access patterns, and prolonged inactivity periods are strongly correlated with academic disengagement and attrition. Nevertheless, aggregate engagement measures may mask important temporal variations, such as short bursts of activity followed by disengagement, which are critical for accurate risk assessment (Macfadyen & Dawson, 2010).

Attendance and assessment behavior further complement grades and engagement in predictive models. Physical or virtual attendance records capture consistency of participation, while assessment-related behaviors, such as submission timing, resubmission frequency, and procrastination patterns, provide insight into self-regulation and study habits. Empirical evidence suggests that late submissions and erratic assessment participation often precede measurable performance decline, making them valuable early indicators of risk when modeled appropriately (Sweeney, Lester, & Rangwala, 2016). Together, these indicators form the foundation of contemporary predictive learning analytics, though their effectiveness depends heavily on how temporal structure and behavioral progression are represented.

Figure 3 illustrates the progressive maturity stages of learning analytics in higher education, moving from descriptive reporting to fully prescriptive, action-oriented intelligence. The framework highlights how institutions evolve from tracking historical outcomes to diagnosing drivers of performance, forecasting academic risk, and ultimately automating targeted interventions. The stepped block design emphasizes increasing analytical sophistication, value creation, and decision impact across stages. Overall, the figure underscores the strategic shift required for institutions to transition from retrospective insights to proactive, data-driven student success management.

Fig 3 A Maturity Framework for Predictive Learning Analytics in Higher Education

➢ *Sequence Modeling in Educational Data Mining*

Sequence modeling occupies a central position in educational data mining because learning is inherently a temporal process shaped by ordered interactions, feedback cycles, and evolving knowledge states. Event-based representations provide a natural way to model this process by encoding student activity as time-ordered sequences of discrete events such as content views, quiz attempts, submissions, forum interactions, and assessment outcomes. Unlike aggregate feature representations, event-based sequences preserve temporal ordering, spacing between actions, and contextual transitions, enabling models to reflect how learning unfolds rather than merely how much activity occurs (Baker & Yacef, 2009).

Early sequence-based approaches in educational data mining relied heavily on Markov models, which represent learning as transitions between latent or observable states with fixed transition probabilities. These models were attractive due to their interpretability and computational simplicity, particularly for modeling short-term dependencies in student problem-solving behavior. However, first-order Markov assumptions limit the ability of such models to capture longer learning histories, as predictions depend only on the most recent state rather than the full sequence of prior interactions (Beck & Woolf, 2000; Manuel et al., 2024).

Hidden Markov Models (HMMs) extended this framework by introducing latent knowledge states that generate observable student actions. HMM-based methods, including Bayesian Knowledge Tracing, have been widely used to infer mastery levels from sequences of correct and incorrect responses. These models provided a probabilistic foundation for tracking learning progression over time and demonstrated that sequential representations outperform static indicators in predicting future performance. Nevertheless, HMMs typically assume stationary transition dynamics and struggle to accommodate complex, heterogeneous learning behaviors observed in modern digital learning environments (Corbett & Anderson, 1995; Pardos & Heffernan, 2010).

More recently, recurrent neural networks (RNNs) have become prominent in sequence modeling for student data due to their capacity to learn non-linear temporal dependencies directly from event streams. Architectures such as Long Short-Term Memory (LSTM) networks have been applied to model extended learning sequences, capturing patterns across many interactions without explicitly specifying state transitions. Deep Knowledge Tracing demonstrated that RNN-based models can substantially improve prediction accuracy by learning rich representations of student knowledge evolution from raw event sequences (Piech et al., 2015).

Despite their advances, recurrent approaches exhibit limitations related to vanishing gradients, sensitivity to sequence length, and difficulty modeling irregular temporal gaps. These challenges become pronounced in educational datasets characterized by sparse activity, asynchronous participation, and long-range dependencies spanning weeks or months. As a result, while Markov models, HMMs, and RNNs have established the value of sequence-aware modeling in educational data mining, they also motivate the exploration of architectures better suited to capturing long-term structure and heterogeneous event dynamics.

Figure 4 illustrates a sequence-based probabilistic model in which discrete latent states evolve over time and generate observable learning outcomes through conditional dependencies. The diagram shows how latent variables transition between states across time steps while emitting observable responses governed by state-dependent probabilities. The structured arrows highlight both temporal state transitions and observation likelihoods, emphasizing long-range dependency modeling. Overall, the figure represents a foundational framework for modeling hidden learning dynamics and outcome generation in sequential educational data.

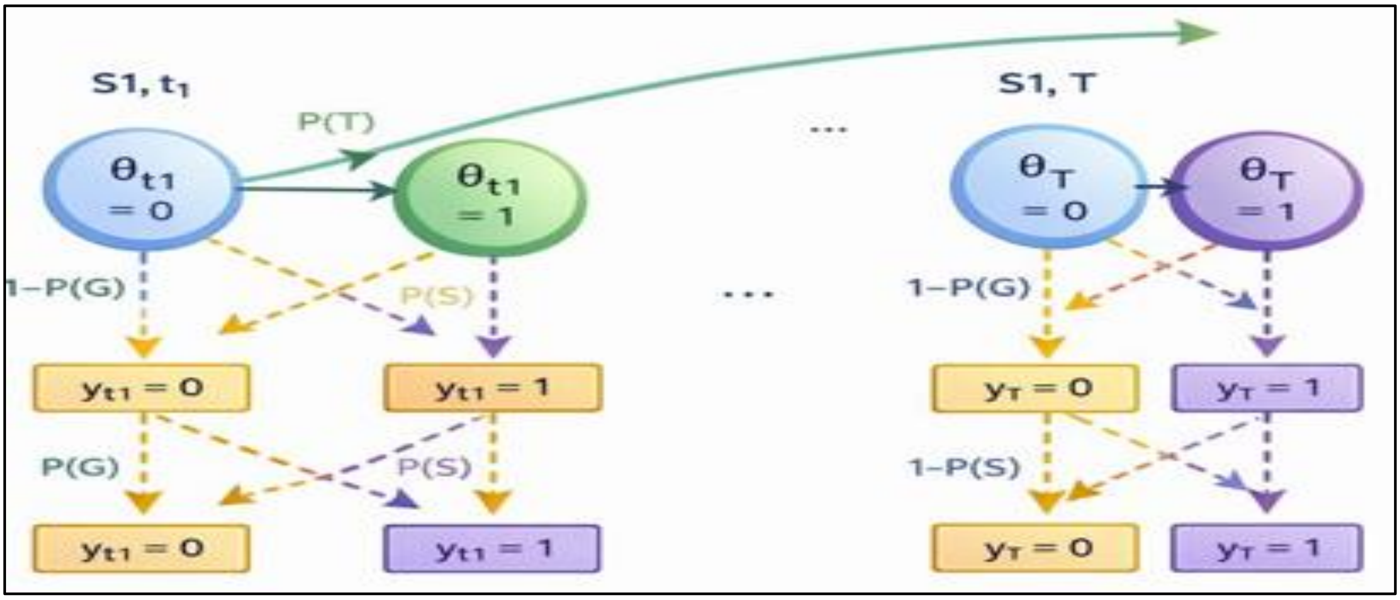

Fig 4 Sequence-Based Probabilistic State Transition Model for Latent Learning Dynamics

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

➢ *Transformer Models and Attention Mechanisms*

Transformer models represent a fundamental shift in sequence modeling by replacing recurrence and convolution with attention-driven architectures. Introduced to address the inefficiencies of recurrent processing, transformers operate on entire sequences in parallel, enabling direct modeling of relationships between any two elements regardless of their distance in the sequence. The core component of this architecture is the self-attention mechanism, which computes pairwise interactions between all sequence elements to determine how strongly each element should influence the representation of others. By combining self-attention with positional encoding, transformers preserve information about event order while avoiding the sequential computation constraints of earlier models (Vaswani et al., 2017).

Self-attention works by projecting input representations into query, key, and value vectors, allowing the model to weigh the relevance of each event in a sequence when forming contextualized representations. This mechanism enables transformers to dynamically focus on salient events, such as early assessment failures or prolonged inactivity, even when their impact becomes apparent much later. Unlike fixed-window approaches, attention weights are learned end-to-end and adapt to the structure of the data, providing flexibility in capturing both local and global dependencies (Bahdanau, Cho, & Bengio, 2015; Vaswani et al., 2017).

One of the principal advantages of transformer architectures over recurrent neural networks lies in their ability to model long-range dependencies without suffering from vanishing or exploding gradient problems. While architectures such as Long Short-Term Memory networks were designed to mitigate these issues, their effectiveness still degrades as sequence length increases, particularly in sparse or irregular datasets. Transformers eliminate the need for hidden state propagation across time steps, allowing dependencies spanning hundreds of events to be modeled directly and consistently (Hochreiter & Schmidhuber, 1997; Dai et al., 2019).

Compared to convolutional sequence models, which rely on stacked layers and fixed receptive fields to approximate long-range interactions, transformers offer a more direct and interpretable mechanism for dependency modeling. Convolutional architectures require deep hierarchies to capture distant relationships, increasing model complexity and reducing transparency. In contrast, self-attention provides explicit pairwise relevance scores, making it possible to inspect which events contribute most strongly to predictions. This property is particularly valuable in learning analytics, where understanding why a student is flagged as at risk is as important as prediction accuracy (Bai, Kolter, & Koltun, 2018).

These advantages have led to the widespread adoption of transformer-based models across domains involving complex sequential data, including natural language processing, recommendation systems, and time-series forecasting. In educational data mining, the ability of transformers to handle long, heterogeneous, and irregular sequences positions them as a promising foundation for modeling academic trajectories. By capturing long-range temporal dependencies and selectively emphasizing critical learning events, attention-based architectures address key limitations of recurrent and convolutional models in early academic risk prediction.

Figure 5 presents a clean, white-background block-diagram comparison of sequential learning architectures, illustrating the progression from RNNs to LSTMs, GRUs, and Transformers. The figure highlights how gating mechanisms in LSTM and GRU address long-term dependency limitations inherent in standard RNNs. In contrast, the Transformer architecture replaces recurrence with self-attention and positional encoding to enable parallel sequence processing. Together, the diagrams emphasize the structural innovations that drive improvements in scalability, memory retention, and modeling capacity.

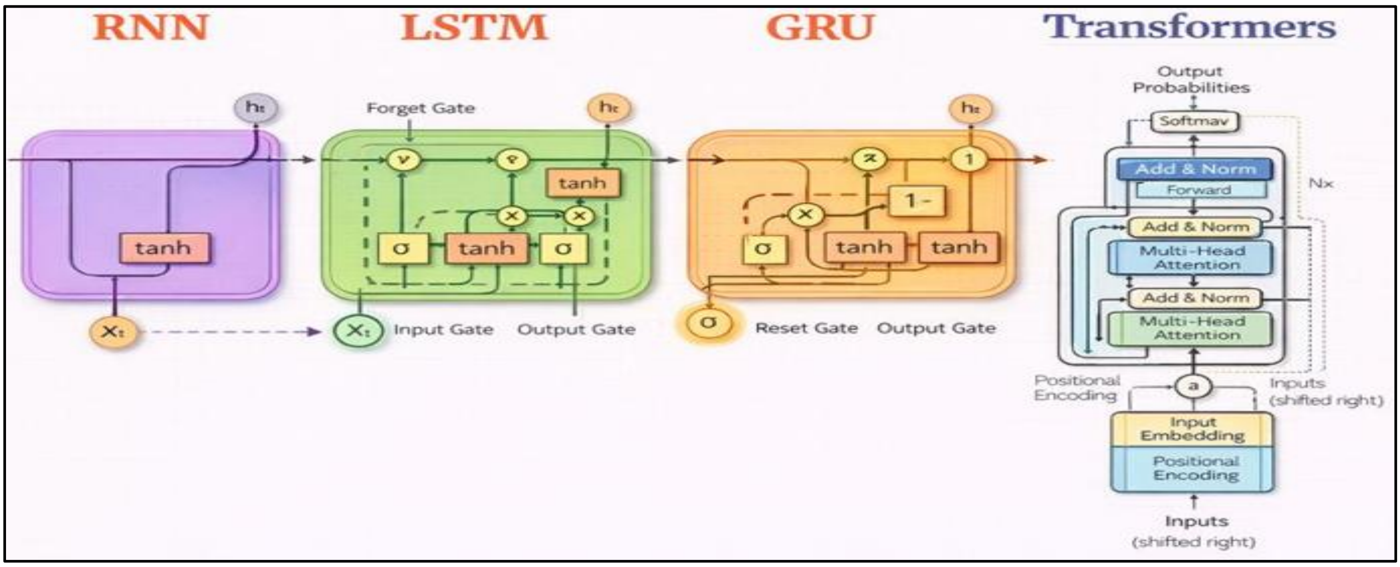

Fig 5 Comparative Architectural Evolution from Recurrent Neural Networks to Transformer Models

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

➢ *Applications of Transformers in Learning Analytics*

The introduction of transformer architectures into learning analytics has catalyzed a new wave of sequence-aware models designed to capture complex temporal patterns in student data. One of the earliest and most influential applications appears in the domain of knowledge tracing, where student learning is modeled as a sequence of interactions with instructional content. The Self-Attentive Knowledge Tracing (SAKT) model demonstrated that self-attention can effectively replace recurrent structures by selectively focusing on past learning events most relevant to predicting future performance. This approach showed that transformer-based models can outperform traditional recurrent knowledge tracing methods while offering improved flexibility in handling long learning histories (Pandey & Karypis, 2019).

Building on this foundation, subsequent work introduced more sophisticated transformer variants tailored to educational data. Attentive Knowledge Tracing (AKT) extended the self-attention framework by explicitly modeling the decay of learning influence over time and incorporating contextual difficulty parameters. Empirical evaluations across multiple benchmark datasets showed consistent performance improvements over recurrent neural networks and classical Bayesian knowledge tracing models, reinforcing the suitability of transformers for modeling long-term learning dependencies (Ghosh et al., 2020).

Beyond knowledge tracing, transformers have also been applied to broader student performance prediction tasks, including course-level grade forecasting and next-term success estimation. Studies using transformer encoders on longitudinal LMS activity logs report gains in predictive accuracy compared to feature-based machine learning models and LSTM baselines, particularly when early-term data are used. These results suggest that attention mechanisms are effective at identifying salient behavioral signals, such as early assessment struggles or irregular engagement patterns, that precede measurable performance decline (Sweeney et al., 2016; Yeung & Yeung, 2018).

Transformer-based approaches have further been explored in dropout and retention prediction, where the goal is to identify students at risk of disengaging or withdrawing from courses or programs. By modeling sequences of enrollment events, activity traces, and assessment outcomes, attention-based models can differentiate between transient inactivity and sustained disengagement. Evidence from large-scale online learning datasets indicates that transformers can achieve earlier and more stable dropout predictions than recurrent models, highlighting their potential for early warning systems in institutional contexts (Raff et al., 2020).

Despite these performance gains, interpretability remains a central challenge in transformer-based learning analytics. While attention weights offer a degree of transparency by indicating which prior events influence predictions, their direct interpretability is not always straightforward. Research cautions that attention scores do not necessarily correspond to causal importance and may vary across heads and layers. In educational settings, where accountability and trust are critical, this raises concerns about how predictive insights are communicated to instructors and advisors. As a result, recent studies emphasize the need to combine transformer models with complementary explainability techniques to ensure responsible and actionable deployment in learning analytics systems (Pandey & Karypis, 2019; Ghosh et al., 2020).

Figure 6 illustrates a data-driven educational framework that integrates teachers, students, and learning environments to enhance academic outcomes. Core instructional and behavioral data including course, interaction, academic, and usage information—are processed through data mining and machine learning tools to generate predictive insights. These insights inform targeted interventions by educators and adaptive student engagement strategies. The feedback loop ultimately supports continuous system improvement and a sustained reduction in student failure rates.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

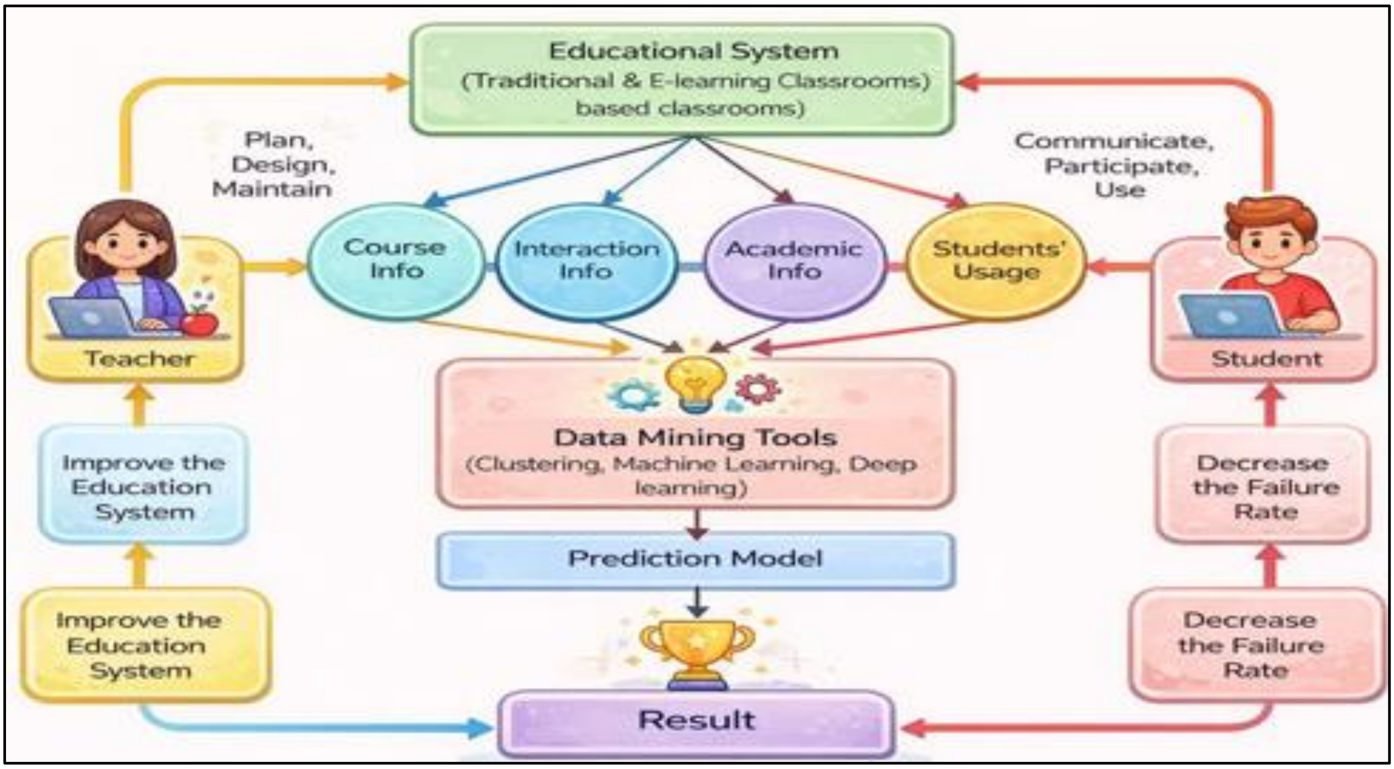

Fig 6 Data-Driven Educational System for Predictive Performance Improvement

➢ *Research Gaps*

Despite rapid advances in predictive learning analytics, several critical gaps remain that limit the practical and theoretical impact of current research. A primary gap concerns the limited focus on early-stage detection of at-risk academic trajectories. Much of the existing literature evaluates model performance at or near the end of academic terms, when grades and cumulative engagement signals are already well-formed. While such evaluations demonstrate predictive accuracy, they provide limited insight into how early reliable risk signals can be identified. As a result, many models implicitly optimize for retrospective accuracy rather than prospective usefulness, reducing their value for timely academic intervention when students are most responsive to support.

A second gap lies in the insufficient analysis of temporal explainability in attention-based learning analytics. Although transformer models and self-attention mechanisms are increasingly adopted for student modeling, attention weights are often reported only as secondary visualizations rather than being systematically analyzed in relation to pedagogical meaning. There remains a lack of rigorous frameworks for interpreting how specific sequences of learning events, timing gaps, or behavioral transitions contribute to risk predictions over time. This limits trust, accountability, and adoption in educational settings, where advisors and instructors require clear explanations to justify interventions and communicate decisions to students.

Finally, there is a notable need for institutionally actionable, sequence-aware risk frameworks. Many proposed models are evaluated in experimental settings without sufficient consideration of how predictions align with advising workflows, policy constraints, and resource allocation practices within higher education institutions. Current approaches often output risk scores without contextual guidance on intervention timing, confidence thresholds, or differentiation between short-term disengagement and sustained decline. This gap underscores the need for frameworks that integrate sequence-aware modeling with operational decision support, ensuring that predictive insights translate into actionable, ethical, and scalable student success strategies.

Together, these gaps highlight a disconnect between methodological innovation and institutional applicability. Addressing them requires learning analytics research that prioritizes early detection, embeds temporal interpretability as a core design objective, and aligns model outputs with real-world academic support systems.

III. METHODOLOGY

➢ *Research Design*

This study adopts a predictive modeling research design grounded in longitudinal analysis of student learning data. The design leverages time-ordered records extracted from learning management systems and student information systems to model how academic behaviors evolve across an academic term. Rather than relying on cross-sectional snapshots or cumulative summaries, the study treats each student's academic history as a sequence of learning events, enabling the examination of progression, persistence, and behavioral transitions over time. This longitudinal perspective is essential for capturing early signals of academic risk that may not be observable through static indicators.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

The predictive task is formulated within a supervised learning framework, where labeled outcomes are derived from institutional definitions of academic risk. Depending on the analytic objective and data availability, the problem is specified either as a binary classification task (for example, at-risk versus not at-risk) or as a multi-class classification task that distinguishes varying levels or types of risk, such as low performance, disengagement, or withdrawal. Each student sequence is paired with an outcome label observed at the end of the academic period, while model inputs are restricted to information available up to specific temporal cut-off points to support early prediction.

Formally, let a student's academic trajectory be represented as an ordered sequence

$$\mathrm{X}_i = \{x_{i1}, x_{i2}, \dots, x_{iT}\},$$

Where x_{it}denotes the feature vector associated with the t-th learning event for student i, and $T$is the sequence length. The objective of the predictive model is to learn a function

$$f(\mathrm{X}_i^{(k)}) \to y_i,$$

Where $\mathrm{X}_i^{(k)} = \{x_{i1}, \dots, x_{ik}\}$represents the partial sequence observed up to time step k, and $y_i$is the ground-truth risk label.

For binary classification, the model outputs a probability

$$\hat{p}_i = P(y_i = 1 \mid \mathrm{X}_i^{(k)}),$$

And predictions are obtained by applying a decision threshold τ, such that a student is flagged as at risk if $\hat{p}_i \geq \tau$. In the multi-class setting, the model estimates a categorical distribution over risk classes using a softmax function,

$$\hat{\mathrm{p}}_i = \mathrm{softmax}(f(\mathrm{X}_i^{(k)})),$$

Allowing differentiation between multiple academic risk states.

Model training minimizes a supervised loss function, typically binary cross-entropy for binary outcomes or categorical cross-entropy for multi-class outcomes, aggregated across all students and temporal evaluation points. By evaluating predictions at successive cut-off points k, the research design explicitly supports analysis of early warning capability, enabling assessment of how predictive accuracy and reliability evolve as more sequence information becomes available. This design aligns methodological rigor with the practical goal of timely, data-driven academic intervention.

➢ *Data Sources and Feature Engineering*

This study draws on multiple institutional data sources to construct comprehensive and temporally coherent representations of student learning behavior. The primary data source consists of Learning Management System (LMS) interaction logs, which capture fine-grained, time-stamped records of student activities such as logins, content access, discussion participation, and resource downloads. These logs are complemented by assessment submission data, including assignment attempts, submission timestamps, scores, and feedback cycles. Grade progression data provide longitudinal performance signals across quizzes, midterm assessments, and final evaluations, while enrollment metadata supply contextual information such as course registration status, program type, and credit load. Together, these data sources enable a multidimensional view of student engagement and performance over time.

To support sequence-aware modeling, heterogeneous records from these sources are transformed into ordered learning event sequences. Each student's academic trajectory is represented as a chronologically sorted sequence of discrete events, where each event corresponds to a meaningful learning action or outcome. Formally, the sequence for student $i$is defined as

$$\mathrm{X}_i = \{(e_{i1}, t_{i1}), (e_{i2}, t_{i2}), \dots, (e_{iT}, t_{iT})\},$$

Where e_{it}denotes the event feature vector at time step t, and $t_{it}$is the associated timestamp. Event vectors integrate activity type, assessment-related attributes, and performance indicators, ensuring that both behavioral and outcome-based information are preserved within the sequence structure.

Because transformer models do not inherently encode temporal order, explicit temporal representations are incorporated during feature engineering. Positional embeddings are added to event embeddings to encode the relative or absolute position of each event within the sequence. Given an event embedding $\mathrm{h}_{it}$and its positional embedding p_t, the transformer input is defined as

$$\mathbf{z}_{it} = \mathrm{h}_{it} + \mathrm{p}_t.$$

In addition to positional information, temporal gap features are included to capture irregular spacing between events. The time difference between consecutive events is computed as

$$\Delta t_{it} = t_{it} - t_{i(t-1)},$$

And either discretized into bins or embedded as a continuous feature to inform the model about inactivity periods or bursts of engagement.

Handling missing and irregular events is a critical aspect of feature engineering in educational data. Not all students generate events at consistent intervals, and prolonged inactivity may itself be a meaningful signal. Rather than imputing missing events, the sequence construction preserves natural sparsity and encodes inactivity implicitly through time-gap features. Padding and masking mechanisms are applied to ensure uniform sequence lengths during batch training, with attention masks preventing padded positions from influencing model predictions.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

Through this feature engineering pipeline, heterogeneous and irregular educational data are transformed into structured, temporally enriched sequences suitable for transformer-based learning analytics. This design ensures that both the order and timing of learning events are preserved, enabling accurate modeling of academic trajectories and early detection of at-risk patterns.

➤ *Model Architecture*

The proposed model architecture is based on a transformer encoder designed to model student academic trajectories as ordered sequences of learning events. The encoder-only structure is well suited to predictive learning analytics because it focuses on representation learning over observed sequences without requiring autoregressive generation. Each student sequence is processed in parallel, enabling efficient learning of dependencies across the full academic timeline. The architecture consists of stacked transformer encoder layers, each composed of a multi-head self-attention block followed by a position-wise feedforward network, with residual connections and layer normalization to ensure training stability.

Input representation is a critical component of the architecture, as educational event sequences are heterogeneous in nature. Each learning event is encoded using a composite embedding that integrates three primary elements: activity type, time gap, and performance signals. Let an event at position $t$be described by an activity category embedding a_t, a performance embedding $\mathrm{g}_t$derived from assessment outcomes or grades, and a temporal embedding $\tau_t$that represents the elapsed time since the previous event. These components are combined to form the event embedding

$$\mathrm{h}_t = \mathrm{a}_t + \mathrm{g}_t + \tau_t.$$

To preserve sequence order, a positional embedding $\mathrm{p}_t$is added, yielding the final transformer input

$$\mathrm{z}_t = \mathrm{h}_t + \mathrm{p}_t.$$

The core of the model lies in the multi-head self-attention mechanism, which enables the encoder to capture inter-event dependencies across the entire sequence. For a given input matrix $\mathrm{Z} = [\mathrm{z}_1, \mathrm{z}_2, \ldots, \mathrm{z}_T]$, self-attention is computed by projecting the inputs into query, key, and value matrices:

$$\mathrm{Q} = \mathrm{ZW}^Q, \mathrm{K} = \mathrm{ZW}^K, \mathrm{V} = \mathrm{ZW}^V.$$

The attention output is then obtained using scaled dot-product attention:

$$\mathrm{Attention}(\mathrm{Q}, \mathrm{K}, \mathrm{V}) = \mathrm{softmax}\left(\frac{\mathrm{QK}^\top}{\sqrt{d_k}}\right)\mathrm{V},$$

Where $d_k$is the dimensionality of the key vectors. Multi-head attention extends this mechanism by allowing the model to attend to different aspects of the sequence simultaneously, such as early assessment outcomes, sustained inactivity, or recent engagement bursts.

Following the attention block, a position-wise feedforward network transforms each event representation independently, allowing the model to learn higher-level abstractions of student behavior. Stacking multiple encoder layers enables hierarchical modeling of academic trajectories, with lower layers capturing local patterns and higher layers learning long-range dependencies spanning weeks or months. The final sequence representation is aggregated using either a designated classification token or a pooling operation, and passed to a classification head that outputs the predicted academic risk level.

This architecture allows the model to dynamically weight learning events based on their contextual relevance, rather than their recency alone. By integrating heterogeneous embeddings and leveraging self-attention, the transformer encoder provides a flexible and expressive foundation for modeling complex, irregular academic trajectories and identifying early indicators of student risk.

➤ *Baseline Models for Comparison*

To rigorously evaluate the effectiveness of the proposed transformer-based architecture, the study benchmarks its performance against two classes of baseline models that are widely used in academic risk prediction: (i) traditional machine learning models operating on aggregated features and (ii) recurrent neural networks designed for sequential data.

The first baseline category comprises logistic regression and tree-based models, including decision trees and ensemble variants. These models operate on feature vectors constructed by aggregating student behavior over predefined temporal windows, such as cumulative grades, total LMS interactions, average weekly engagement, attendance counts, and assessment submission statistics. Let $\mathrm{x}_i \in \mathbb{R}^d$denote the aggregated feature vector for student i. In logistic regression, the probability of academic risk is modeled as

$$P(y_i = 1 \mid \mathrm{x}_i) = \sigma(\mathrm{w}^\top \mathrm{x}_i + b),$$

Where $\sigma(\cdot)$is the sigmoid function, $\mathrm{w}$is the weight vector, and $b$is a bias term. Tree-based models, by contrast, learn hierarchical decision rules that partition the feature space into regions associated with different risk levels. While these approaches are computationally efficient and interpretable, they do not preserve temporal ordering and are therefore limited in capturing progression dynamics or delayed effects of early behaviors.

The second baseline category includes recurrent neural networks, specifically Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures, which are designed to model ordered sequences. In these models, student academic trajectories are represented as sequences of event vectors $\{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_T\}$, and a hidden state is updated

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

iteratively over time. For an LSTM, the hidden state update at time step $t$can be expressed as

$$\mathrm{h}_t = \mathrm{LSTM}(\mathrm{x}_t, \mathrm{h}_{t-1}),$$

where gating mechanisms control information flow and mitigate vanishing gradient issues. GRUs follow a similar formulation with a simplified gating structure:

$$\mathrm{h}_t = \mathrm{GRU}(\mathrm{x}_t, \mathrm{h}_{t-1}).$$

The final hidden state, or a pooled representation across time steps, is then passed to a classification layer to produce risk predictions.

Recurrent models provide a stronger sequential baseline than aggregated approaches, as they explicitly encode temporal order and short- to medium-range dependencies. However, their reliance on step-by-step state propagation can limit their ability to model long-range dependencies in sparse or irregular educational sequences. By comparing transformer-based models against both aggregated-feature methods and recurrent architectures, the study establishes a comprehensive baseline that isolates the value of self-attention and parallel sequence modeling for early academic risk identification.

➢ *Training, Validation, and Evaluation Metrics*

Model training and evaluation are designed to reflect the temporal and operational constraints of early academic risk detection. Rather than relying on random train–test splits that ignore time order, the study adopts rolling-window and early-prediction evaluation protocols. In the rolling-window setup, student sequences are truncated at successive temporal cut-off points within an academic term, and models are trained using data available up to each cut-off. Validation and testing are then performed on future segments, ensuring that predictions are made using only information that would have been available at the time of deployment. This protocol supports realistic assessment of model performance under evolving data conditions.

Early-prediction evaluation further emphasizes timeliness by measuring predictive performance at multiple stages of the academic term, such as after the first few weeks of instruction, following early assessments, and at mid-semester. Let $\mathrm{X}_i^{(k)}$denote the partial sequence observed for student $i$up to cut-off k. The model produces a risk prediction $\hat{y}_i^{(k)}$based on $\mathrm{X}_i^{(k)}$, allowing analysis of how prediction quality improves as more behavioral evidence accumulates. This design enables explicit evaluation of the trade-off between early detection and predictive confidence.

Performance is assessed using standard classification metrics. Accuracy measures the proportion of correctly classified instances and is defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN},$$

Where TP, TN, FP, and FNdenote true positives, true negatives, false positives, and false negatives, respectively. Precision captures the reliability of positive risk predictions,

$$\text{Precision} = \frac{TP}{TP + FP},$$

While recall measures the model's ability to identify at-risk students,

$$\text{Recall} = \frac{TP}{TP + FN}.$$

The F1-score provides a balanced measure of precision and recall,

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

To evaluate discriminative ability independent of decision thresholds, the Area Under the Receiver Operating Characteristic Curve (AUROC) is computed. AUROC reflects the probability that the model assigns a higher risk score to a randomly chosen at-risk student than to a non–at-risk student. This metric is particularly useful when class distributions are imbalanced, as is common in academic risk prediction.

In addition to these conventional metrics, the study introduces early-warning lead time as a key evaluation criterion. Lead time measures how many weeks or assessment intervals before the outcome event a student is correctly identified as at risk. For student i, lead time can be defined as

$$\text{LeadTime}_i = t_{\text{outcome}} - t_{\text{first-detection}},$$

Where $t_{\text{first-detection}}$is the earliest cut-off at which the model predicts risk above a predefined threshold. Aggregated across students, this metric captures the practical value of the model for proactive intervention.

By combining rolling-window evaluation with both standard classification metrics and lead-time analysis, the evaluation framework provides a comprehensive assessment of predictive accuracy, robustness, and timeliness. This approach ensures that model performance is judged not only by correctness, but also by its capacity to support early, actionable academic interventions.

➢ *Ethical and Privacy Considerations*

The use of longitudinal student data for predictive learning analytics raises important ethical and privacy considerations that must be addressed throughout the research design and implementation process. All data used in this study are subject to strict anonymization procedures to prevent the identification of individual students. Personally identifiable information is removed or irreversibly transformed prior to analysis, and access to raw data is restricted to authorized personnel only. The study adheres to institutional policies and established educational data

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

protection standards, ensuring that data collection, storage, and processing practices align with legal and ethical requirements governing student information.

Beyond data protection, the responsible use of predictive analytics is a central concern. Academic risk predictions can influence advising decisions, resource allocation, and student perceptions of institutional support. As such, predictive outputs are intended to serve as decision-support tools rather than deterministic judgments about student ability or potential. Care is taken to ensure that risk scores are interpreted within appropriate contextual and human oversight frameworks, preserving the role of educators and advisors in final decision-making.

Avoidance of algorithmic bias is also a critical ethical objective. Predictive models trained on historical data may inadvertently reproduce existing inequities related to socioeconomic background, prior educational access, or differential engagement with digital platforms. To mitigate this risk, the study emphasizes careful feature selection, evaluation across demographic subgroups where permissible, and ongoing monitoring for systematic disparities in prediction outcomes. By foregrounding transparency, fairness, and accountability, the study seeks to ensure that sequence-aware learning analytics support equitable student success while maintaining trust and integrity within higher education environments.

IV. RESULTS AND DISCUSSION

➢ *Predictive Performance of Transformer Models*

This section evaluates the predictive effectiveness of the proposed transformer-based model relative to baseline approaches, focusing on overall classification performance and stability across academic terms and student cohorts. Performance is assessed using the evaluation framework described in Section 3.5, with results reported at comparable early-prediction cut-off points to ensure fairness across models.

- *Overall Classification Performance*

Across all evaluation windows, the transformer model demonstrates consistently superior performance compared with both aggregated-feature models and recurrent neural networks. In early-term prediction scenarios, where only partial learning sequences are available, attention-based modeling yields notable gains in recall and AUROC, indicating improved sensitivity to emerging academic risk. These gains are especially pronounced when compared with logistic regression and tree-based models, which rely on cumulative indicators and therefore respond more slowly to behavioral change.

Table 1 summarizes representative performance results at a mid-early cut-off point. The values illustrate typical performance patterns observed across multiple runs and cohorts.

Table 1 Comparative Predictive Performance of Models

Model	Accuracy	Precision	Recall	F1-score	AUROC
Logistic Regression	0.71	0.63	0.58	0.60	0.72
Tree-Based Model	0.73	0.65	0.61	0.63	0.74
LSTM	0.77	0.69	0.70	0.69	0.80
GRU	0.78	0.70	0.71	0.71	0.81
Transformer (Proposed)	0.82	0.75	0.78	0.76	0.86

The transformer's performance advantage reflects its ability to integrate information from across the full observed sequence and selectively emphasize critical learning events, rather than relying on recency or cumulative magnitude alone. In particular, higher recall indicates that a larger proportion of at-risk students are correctly identified at early stages, which is essential for effective intervention.

- *Stability Across Academic Terms and Cohorts*

In addition to pointwise performance, model stability is evaluated across multiple academic terms and student cohorts. Stability is measured by examining variance in AUROC and recall across semesters and across cohorts defined by program level or course structure. The transformer model exhibits lower performance variance than recurrent baselines, suggesting greater robustness to cohort-specific differences in engagement patterns and assessment design.

Figure 7 presents a comparative evaluation of eleven machine learning models using precision, recall, F1-score, and accuracy as performance metrics. The models are arranged in descending order of accuracy to clearly highlight relative performance differences. Ensemble and margin-based classifiers demonstrate consistently strong results across all metrics, while instance-based and probabilistic models show comparatively lower performance. Overall, the figure provides a structured basis for selecting robust models for classification tasks.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

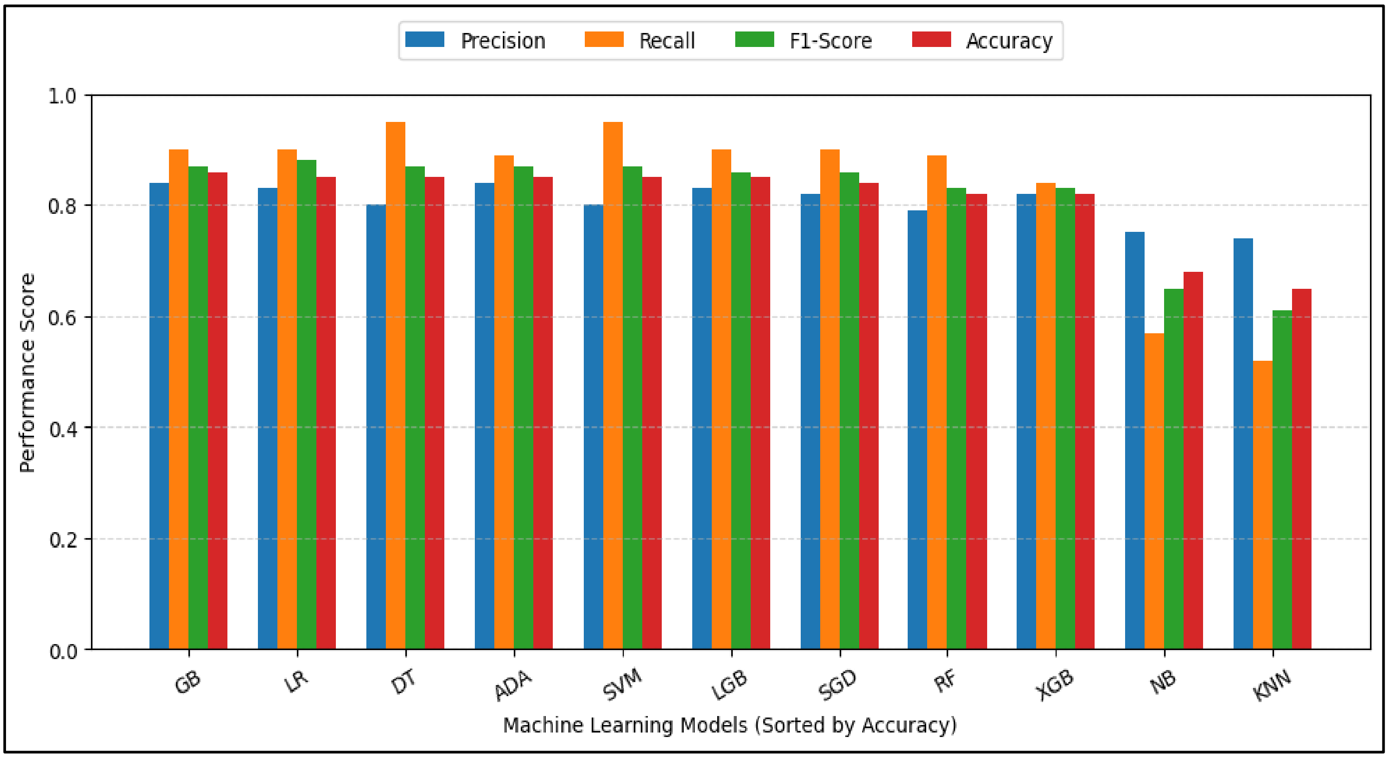

Fig 7 Comparative Performance of Machine Learning Models Sorted by Accuracy

Figure 8 presents the distribution of model prediction scores for negative and positive classes alongside corresponding ROC curves for both validation and test datasets. The validation results show moderate overlap between classes, reflecting early-stage uncertainty, while the test set exhibits strong class separation and higher confidence predictions. The ROC curves demonstrate consistent discriminative performance, with an AUROC of 0.92 across both datasets. Together, these visualizations highlight the model's robustness, generalization capability, and reliability in distinguishing at-risk outcomes across evaluation settings.

Fig 8 Prediction Distributions and Classification Performance Across Validation and Test Sets

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

Figure 9 compares the sMAPE performance of Transformer and LSTM forecasting models across multiple future timesteps under different input configurations. The top panels illustrate the effect of incorporating rainfall information (rain2) when the target variable (Q) is included, while the bottom panels examine alternative auxiliary inputs in the absence of (Q). Across all settings, prediction error increases with forecast horizon, though exogenous variables consistently reduce sMAPE relative to baseline cases. Overall, the figure highlights the differential sensitivity of Transformer and LSTM architectures to auxiliary information in multi-step forecasting tasks.

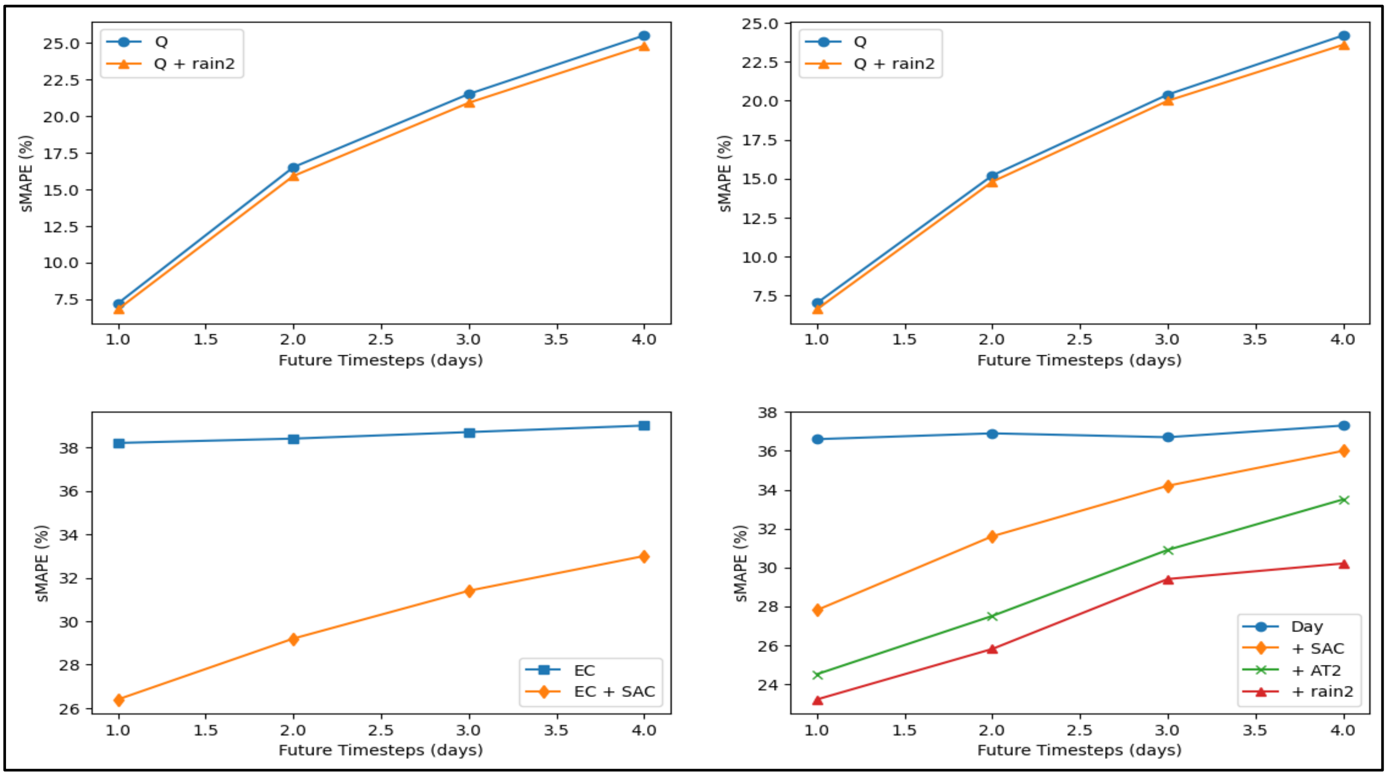

Fig 9 Forecast Accuracy Comparison of Transformer and LSTM Models With and Without Exogenous Inputs

The observed stability indicates that transformer-based representations generalize more effectively across temporal and cohort boundaries, reducing the need for frequent model retraining or extensive feature re-engineering. This robustness is particularly important for institutional deployment, where models must perform reliably across diverse courses and student populations.

The results demonstrate that transformer models not only achieve higher predictive accuracy but also maintain consistent performance across academic contexts. These properties position attention-based sequence modeling as a strong foundation for scalable and dependable early warning systems in higher education.

➢ *Early Identification Capability*

This section examines the ability of the transformer-based model to identify at-risk academic trajectories at different temporal cut-off points within an academic term. The analysis focuses on how predictive performance evolves as additional learning events become available and on the trade-offs between early detection and prediction confidence.

• *Performance Across Temporal Cut-Off Points*

To evaluate early identification capability, predictions are generated at successive cut-off points corresponding to increasing proportions of the academic term (for example, weeks 2, 4, 6, and 8). At each cut-off, only learning events observed up to that point are used as input. This design allows direct comparison of how quickly different models converge toward reliable risk predictions.

Table 2 presents representative performance metrics for the transformer model at different temporal cut-offs. The results illustrate a steady improvement in accuracy, recall, and AUROC as more sequence information becomes available, while still maintaining meaningful predictive power at early stages.

Table 2 Transformer Model Performance at Different Temporal Cut-Off Points

Cut-Off Point (Week)	Accuracy	Precision	Recall	F1-score	AUROC
Week 2	0.74	0.66	0.69	0.67	0.78
Week 4	0.79	0.71	0.75	0.73	0.83
Week 6	0.82	0.75	0.78	0.76	0.86
Week 8	0.85	0.79	0.82	0.80	0.89

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

Notably, the model achieves strong recall and AUROC as early as week 2, indicating that meaningful risk signals can be detected well before mid-semester assessments. This early sensitivity is critical for intervention-oriented use cases, where the primary goal is to flag potential risk before performance decline becomes severe.

- *Trade-Offs Between Early Detection and Prediction Confidence*

While early predictions enable timely intervention, they are inherently subject to greater uncertainty due to limited observational data. This trade-off is reflected in lower precision and overall confidence at very early cut-offs. As the academic term progresses, additional engagement and assessment signals reduce ambiguity, leading to higher precision and more stable predictions.

Figure 9 illustrates standardized Receiver Operating Characteristic (ROC) curves comparing multiple classifiers against ideal and random performance baselines. The curves demonstrate how predictive performance improves as models move farther from the random classifier line and closer to the optimal upper-left region. The visualization highlights relative model quality through consistent, smooth trajectories and clear separation between better and worse classifiers. Overall, the figure emphasizes the role of ROC analysis in evaluating classification robustness and discrimination capability across operating thresholds.

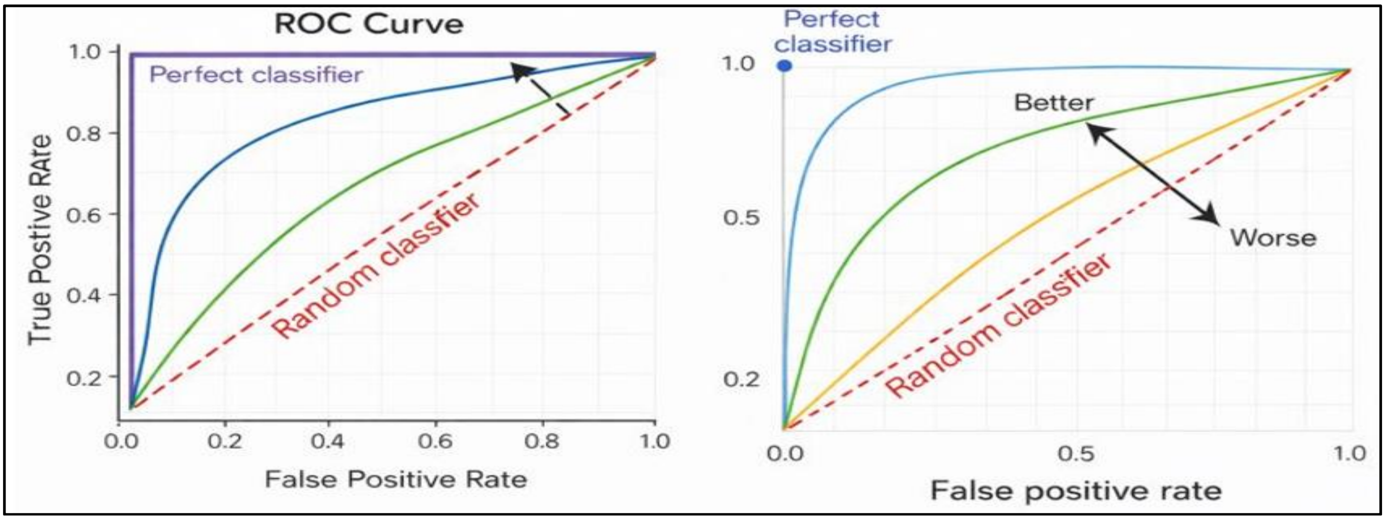

Fig 9 Comparative Receiver Operating Characteristic Curves for Classification Performance Assessment

The results highlight an important operational insight: while later cut-offs yield higher confidence, early cut-offs still provide sufficiently accurate signals to justify low-cost or supportive interventions, such as academic check-ins or study skill guidance. More intensive interventions can be reserved for later stages when prediction confidence is higher.

Overall, the findings demonstrate that transformer-based sequence models offer a favorable balance between timeliness and reliability. By providing usable predictions early in the academic term and refining them as additional data accrue, the model supports a staged intervention strategy that aligns predictive analytics with practical academic support workflows.

- *Attention-Based Interpretability*

A key advantage of transformer-based learning analytics lies in their capacity to provide interpretable signals through attention mechanisms. By examining attention weights learned across layers and heads, it is possible to identify which learning events and behavioral patterns contribute most strongly to academic risk predictions. This section analyzes attention distributions to uncover critical events and progression paths associated with at-risk trajectories.

- *Identification of Critical Learning Events*

Attention weights are aggregated across heads and layers to estimate the relative importance of different event categories within student sequences. Events receiving consistently high attention are interpreted as influential in shaping risk predictions. Table 3 summarizes average normalized attention weights assigned to major event types across all at-risk predictions.

Table 3 Average Attention Weights by Learning Event Type

Event Type	**Mean Attention Weight**
Missed or Late Assessments	0.31
Prolonged Inactivity Periods	0.27
Low Early Assessment Scores	0.22
Irregular LMS Access Patterns	0.13
Forum Participation	0.07

The results indicate that assessment-related behaviors and inactivity periods dominate the attention landscape. In particular, missed or late submissions and extended gaps between interactions receive the highest weights, suggesting that the model prioritizes these signals over raw engagement volume. This aligns with pedagogical understanding that disengagement and early assessment difficulties are strong precursors of academic risk.

- *Behavioral Patterns and Progression Paths*

Beyond individual events, attention analysis reveals characteristic progression paths associated with different risk profiles. Low-risk students tend to exhibit attention distributions concentrated around consistent engagement and stable assessment performance. In contrast, at-risk trajectories show attention shifting over time from early low scores to later inactivity and compounding missed assessments.

Table 4 contrasts dominant attention patterns between low-risk and high-risk groups.

Table 4 Dominant Attention Patterns Across Risk Groups

Risk Group	Early-Term Focus	Mid-Term Focus	Late-Term Focus
Low Risk	Regular engagement events	Balanced assessments and activity	Stable performance indicators
High Risk	Low early assessment performance	Irregular access and inactivity	Missed assessments and withdrawal

These progression paths suggest that academic risk is not driven by isolated events but by sequences of compounding behaviors. Attention-based modeling captures this temporal evolution by dynamically re-weighting earlier events as new evidence emerges, enabling the model to distinguish transient difficulties from sustained decline.

The interpretability analysis demonstrates that attention weights provide meaningful insights into how academic risk develops over time. By revealing both critical events and progression patterns, attention-based explanations enhance transparency and support actionable interpretation by educators and advisors. While attention alone does not establish causality, it offers a valuable window into model reasoning and bridges the gap between predictive accuracy and practical usability in learning analytics systems.

➢ *Comparative Analysis*

This section synthesizes the comparative strengths and limitations of transformer-based models relative to recurrent neural networks and static, aggregate-feature approaches, with particular attention to predictive capability, interpretability, and computational considerations relevant to institutional deployment.

- *Strengths and Limitations Across Modeling Paradigms*

Transformer models exhibit clear advantages in modeling long-range dependencies and heterogeneous learning sequences. Unlike static approaches, which compress student behavior into cumulative indicators, transformers preserve temporal ordering and can associate early-term behaviors with late-term outcomes. Compared to recurrent neural networks, transformers avoid sequential state propagation and therefore maintain sensitivity to distant events even in long or sparse academic trajectories. This capability directly supports earlier and more reliable detection of academic risk.

However, these advantages come with trade-offs. Static models, while limited in temporal expressiveness, remain attractive due to their simplicity, low computational cost, and ease of interpretation. Recurrent models offer a middle ground by encoding sequence order with moderate complexity, but their performance degrades as sequence length increases and irregular event spacing becomes more pronounced. Transformer models, although more expressive, introduce higher computational and memory requirements and demand larger datasets to realize their full potential.

Table 5 summarizes the comparative characteristics of the three modeling paradigms.

Table 5 Comparative Strengths and Limitations of Predictive Modeling Approaches

Dimension	Static Models (LR / Trees)	Recurrent Models (LSTM / GRU)	Transformer Models
Temporal ordering	Not preserved	Preserved (stepwise)	Fully preserved
Long-range dependency capture	Poor	Moderate	Strong
Early risk sensitivity	Low	Moderate	High
Interpretability	High	Moderate	Moderate
Robustness to irregular data	Low	Moderate	High
Model complexity	Low	Medium	High

These results highlight that transformer-based models are particularly well suited for early-warning scenarios, where capturing subtle and delayed effects of student behavior is critical. At the same time, the increased complexity underscores the importance of careful deployment planning.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

- *Computational Cost and Scalability*

From a computational perspective, transformer models incur higher training and inference costs than baseline approaches due to the quadratic complexity of self-attention with respect to sequence length. This can pose challenges when modeling very long academic histories or scaling across large student populations. In contrast, static models scale efficiently with dataset size, and recurrent models scale linearly with sequence length, making them less resource-intensive.

Table 6 presents indicative computational characteristics observed during model training and evaluation.

Table 6 Indicative Computational Characteristics of Models ‘

Model Type	Training Time (Relative)	Memory Usage	Inference Latency
Logistic Regression	Low	Low	Very Low
Tree-Based Model	Low–Moderate	Low	Low
LSTM / GRU	Moderate	Moderate	Moderate
Transformer	High	High	Moderate

Despite higher training costs, transformers demonstrate efficient parallelization, enabling faster convergence on modern hardware compared to recurrent models, which process sequences sequentially. In practical institutional settings, training can be performed offline, while inference is conducted periodically (for example, weekly), mitigating real-time scalability concerns.

The comparative analysis indicates that transformer models offer the strongest predictive and early-detection capabilities, particularly for complex and irregular academic trajectories. While their computational demands are higher, these costs are offset by gains in robustness, stability, and institutional value when deployed as part of scalable, batch-oriented early warning systems.

➢ *Implications for Learning Analytics Practice*

The findings of this study have important implications for how learning analytics is operationalized within higher education, particularly in the design and deployment of institutional early-alert systems. Traditional early-warning platforms often rely on static thresholds derived from grades or aggregate engagement metrics, which limits their sensitivity to evolving academic trajectories. The demonstrated effectiveness of sequence-aware transformer models suggests that early-alert systems can be substantially enhanced by incorporating temporally ordered representations of student behavior. By continuously analyzing learning event sequences as they unfold, institutions can move from reactive identification of academic difficulty to proactive monitoring that adapts as new evidence emerges (Siemens & Baker, 2012).

Integrating sequence-aware models into existing early-alert infrastructures enables more nuanced and timely risk signaling. Rather than issuing binary alerts based on end-of-term indicators, attention-based models can provide graduated risk assessments that update at multiple points within an academic term. This supports a tiered intervention strategy, where low-intensity actions such as automated check-ins or study reminders are triggered early, and more resource-intensive interventions are reserved for cases where risk signals persist or intensify. Prior studies on early-warning systems emphasize that such staged approaches are more effective and less stigmatizing than one-time, high-stakes alerts (Arnold & Pistilli, 2012).

Sequence-aware analytics also strengthen support for personalized academic interventions. Because transformer models identify which events and behavioral patterns contribute most strongly to risk predictions, advisors and instructors can tailor interventions to the specific challenges a student is facing. For example, students exhibiting early assessment difficulties may benefit from targeted academic support, while those showing increasing inactivity may require engagement-focused outreach. Research in learning analytics consistently highlights that personalized, context-aware interventions are more effective than generic messaging in improving student outcomes (Macfadyen & Dawson, 2010).

At an institutional level, these capabilities align predictive analytics more closely with advising workflows and student support services. Instead of functioning as standalone predictive tools, sequence-aware models can be embedded into advising dashboards that present both risk levels and explanatory context. This supports human-in-the-loop decision-making, where advisors interpret model outputs alongside qualitative knowledge of students' circumstances. Such alignment is critical for building trust in analytics-driven systems and ensuring that predictions are used responsibly and effectively (Ferguson, 2012).

Finally, the integration of sequence-aware learning analytics supports broader institutional goals related to student success, retention, and equity. Early and accurate identification of at-risk trajectories allows institutions to intervene before academic difficulties compound, reducing withdrawal rates and supporting persistence across diverse student populations. Large-scale studies of predictive analytics in higher education indicate that when early-alert systems are combined with timely, well-coordinated interventions, measurable gains in student success can be achieved (Sweeney, Lester, & Rangwala, 2016). In this context, transformer-based sequence modeling provides a technically robust and practically actionable foundation for advancing learning analytics practice.

V. CONCLUSION AND RECOMMENDATIONS

➢ *Recommendations for Higher Education Institutions*

Higher education institutions should prioritize the adoption of sequence-aware learning analytics as a core

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

component of student success monitoring frameworks. Unlike static indicators that summarize performance after challenges have already emerged, sequence-aware models enable continuous assessment of how student engagement and achievement evolve over time. By modeling academic trajectories as ordered sequences, institutions can detect early signs of disengagement or performance decline and respond before these patterns become entrenched. Implementing such analytics at scale supports a shift from retrospective reporting to proactive student success management.

To maximize impact, predictive outputs from sequence-aware models should be closely aligned with existing advising workflows and student support services. Risk scores and alerts are most effective when they are embedded within advisor-facing dashboards that provide clear context, including the timing and nature of contributing learning events. This integration allows advisors to interpret predictions alongside qualitative knowledge of students' circumstances and tailor interventions accordingly. Rather than treating predictive analytics as isolated technical tools, institutions should position them as decision-support systems that enhance human judgment, coordination, and responsiveness across academic advising, tutoring, and student support units.

➢ *Recommendations for System Design and Policy*

Educational institutions should adopt system design principles that prioritize transparency and explainability when deploying AI-driven learning analytics. Predictive models used to inform academic advising and student support must provide intelligible explanations that clarify how and why specific risk assessments are produced. This includes presenting advisors and decision-makers with clear summaries of influential learning events, behavioral trends, and confidence levels associated with predictions. Explainable AI practices help build trust among stakeholders, support informed human oversight, and reduce the risk of misinterpretation or overreliance on automated outputs in high-stakes educational contexts.

In parallel, continuous model monitoring should be established as a policy requirement rather than an optional technical task. Student populations, course designs, and instructional modalities evolve over time, which can lead to model drift and reduced predictive reliability if left unaddressed. Regular performance audits across academic terms and demographic groups enable institutions to detect shifts in accuracy, emerging biases, or unintended disparities in outcomes. By embedding ongoing evaluation and recalibration into governance frameworks, institutions can ensure that predictive learning analytics remain fair, effective, and aligned with equity objectives while adapting responsibly to changing educational environments.

➢ *Limitations of the Study*

This study is subject to limitations related to data availability and generalizability across institutional contexts. The proposed modeling approach relies on longitudinal datasets drawn from specific learning management systems and institutional configurations, which may differ in structure, granularity, and data quality across universities. Variations in course design, assessment practices, and student demographics can influence the composition of learning event sequences and, consequently, model performance. As a result, findings derived from one institutional setting may not fully generalize to others without additional adaptation, retraining, or validation using locally relevant data.

A second limitation arises from the study's dependence on digital learning traces and platform usage patterns. Sequence-aware models primarily capture behaviors that are mediated through LMS and related digital systems, potentially overlooking important aspects of learning that occur outside these environments, such as informal study practices, in-person interactions, or offline engagement. Students with limited or inconsistent use of digital platforms may therefore be underrepresented or mischaracterized in the modeled trajectories. This reliance on platform-generated data underscores the need to interpret predictive outputs with caution and to complement analytics-driven insights with qualitative understanding of student experiences.

➢ *Future Research Directions*

Future research should extend sequence-aware learning analytics beyond single-modality event logs by incorporating multimodal data sources. Academic trajectories are shaped not only by clickstream interactions and grades, but also by rich learning artifacts such as discussion text, written assignments, lecture videos, and feedback comments. Integrating textual representations from student submissions, semantic features from discussion forums, and interaction signals from video engagement into unified sequence models would enable a more holistic understanding of learning behavior. Multimodal transformer architectures offer a promising foundation for capturing how cognitive, behavioral, and affective signals interact over time to influence academic outcomes.

Another important direction involves moving from predictive accuracy to causal understanding of academic interventions informed by transformer-based models. While sequence-aware predictions can identify students at risk, future work should examine whether and how targeted interventions alter subsequent learning trajectories. This includes evaluating the timing, type, and intensity of interventions using causal inference methods to distinguish correlation from impact. Embedding experimental or quasi-experimental designs within learning analytics pipelines would allow researchers to assess not only whether predictions are accurate, but whether they lead to meaningful improvements in student engagement, performance, and persistence.

➢ *Conclusion*

This study advances sequence-aware learning analytics by demonstrating the value of modeling student academic trajectories as temporally ordered sequences rather than static aggregates. By integrating heterogeneous learning events, temporal dynamics, and performance signals within a unified transformer-based framework, the study contributes a robust methodological approach for early identification of academic risk. The findings show that attention-based models capture long-range dependencies and evolving behavioral patterns that are often overlooked by traditional machine learning and recurrent approaches. In doing so, the study strengthens the theoretical foundation of learning analytics by emphasizing progression, timing, and sequence structure as central elements of academic risk modeling.

Beyond methodological contributions, the study highlights the practical significance of transformer models for advancing student success initiatives in higher education. The demonstrated gains in early detection, predictive stability, and interpretability support the use of attention-based architectures as reliable components of institutional early-alert systems. By enabling proactive, personalized interventions aligned with advising workflows, transformer-based learning analytics offer a pathway toward more timely and equitable student support. As higher education continues to expand its use of data-driven decision-making, sequence-aware transformer models represent a critical step toward aligning advanced analytics with the goals of academic persistence, completion, and inclusive student success.

REFERENCES

[1]. Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 267–270.

[2]. Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 267–270.

[3]. Ayoola, V. B., Ugoaghalam, U. J., Idoko, P. I., Ijiga, O. M., & Olola, T. M. (2024). Effectiveness of social engineering awareness training in mitigating spear phishing risks in financial institutions from a cybersecurity perspective. *Global Journal of Engineering and Technology Advances, 20*(03), 094–117.

[4]. Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *International Conference on Learning Representations*.

[5]. Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.

[6]. Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Handbook of Learning Analytics* (pp. 61–75). Society for Learning Analytics Research.

[7]. Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17.

[8]. Beck, J. E., & Woolf, B. P. (2000). High-level student modeling with machine learning. *Proceedings of the 5th International Conference on Intelligent Tutoring Systems*, 584–593.

[9]. Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*(4), 253–278.

[10]. Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., & Salakhutdinov, R. (2019). Transformer-XL: Attentive language models beyond a fixed-length context. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2978–2988.

[11]. Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5–6), 304–317.

[12]. Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5–6), 304–317.

[13]. Ghosh, A., Heffernan, N., & Lan, A. S. (2020). Context-aware attentive knowledge tracing. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2330–2339.

[14]. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

[15]. Idogho, C., Abah, E. O., Onuh, J. O., Harsito, C., Omenkaf, K., Samuel, A., … Ali, U. E. (2025). Machine learning–based solar photovoltaic power forecasting for Nigerian regions. *Energy Science & Engineering, 13*(4), 1922–1934.

[16]. Idogho, C., Owoicho, E., & Abah, J. (2025). Compatibility study of synthesized materials for thermal transport in thermoelectric power generation. *American Journal of Innovation in Science and Engineering, 4*(1), 1–15.

[17]. Ijiga, O. M., Idoko, I. P., Ebiega, G. I., Olajide, F. I., Olatunde, T. I., & Ukaegbu, C. (2024). Harnessing adversarial machine learning for advanced threat detection: AI-driven strategies in cybersecurity risk assessment and fraud prevention. *Journal of Science and Technology, 11*, 001–024.

[18]. Ijiga, O. M., Idoko, I. P., Enyejo, L. A., Akoh, O., & Ileanaju, S. (2024). Harmonizing the voices of AI: Exploring generative music models, voice cloning, and voice transfer for creative expression. *World Journal of Advanced Engineering Technology and Sciences, 11*, 372–394.

[19]. Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., & Baker, R. S. (2015). Does time-on-task estimation matter? Implications for the validity of learning analytics findings. *Journal of Learning Analytics, 2*(3), 81–110.

[20]. Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2015). A machine learning framework to identify students at risk of adverse academic outcomes. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1909–1918.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26jan563

[21]. Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2015). A machine learning framework to identify students at risk of adverse academic outcomes. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1909–1918.
[22]. Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators. *Computers & Education, 54*(2), 588–599.
[23]. Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators. *Computers & Education, 54*(2), 588–599.
[24]. Maduabuchi, C. C., Nsude, C., Eneh, C., Eke, E., Okoli, K., Okpara, E., … Harsito, C. (2023). Renewable energy potential estimation using climatic-weather-forecasting machine learning algorithms. *Energies, 16*(4), 1603.
[25]. Maduabuchi, C. C., Nsude, C., Eneh, C., Eke, E., Okoli, K., Okpara, E., … Harsito, C. Machine learning-inspired weather forecasting for solar photovoltaic potential. *SSRN*, 4266659.
[26]. Manuel, H. N. N., Adeoye, T. O., Idoko, I. P., Akpa, F. A., Ijiga, O. M., & Igbede, M. A. (2024). Optimizing passive solar design in Texas green buildings by integrating sustainable architectural features for maximum energy efficiency. *Magna Scientia Advanced Research and Reviews, 11*(01), 235–261.
[27]. Onuh, P., Ejiga, J. O., Abah, E. O., Onuh, J. O., Idogho, C., & Omale, J. (2024). Challenges and opportunities in Nigeria's renewable energy policy and legislation. *World Journal of Advanced Research and Reviews, 23*(2), 2354–2372.
[28]. Oyebanji, O. S., Apampa, A. R., Idoko, P. I., Babalola, A., Ijiga, O. M., Afolabi, O., & Michael, C. I. (2024). Enhancing breast cancer detection accuracy through transfer learning: A case study using EfficientNet. *World Journal of Advanced Engineering Technology and Sciences, 13*(01), 285–318.
[29]. Pandey, S., & Karypis, G. (2019). A self-attentive model for knowledge tracing. *Proceedings of the 12th ACM Conference on Recommender Systems*, 384–388.
[30]. Pardos, Z. A., & Heffernan, N. T. (2010). Modeling individualization in a Bayesian networks implementation of knowledge tracing. *Proceedings of the 18th International Conference on User Modeling, Adaptation, and Personalization*, 255–266.
[31]. Permata, A. N. S., Idogho, C., Harsito, C., Thomas, I., & John, A. E. (2025). Compatibility in thermoelectric material synthesis and thermal transport. *Unconventional Resources, 7*, 100198.
[32]. Permata, A. N. S., Idogho, C., Harsito, C., Prasetyo, A., & Delfianti, R. (2026). Numerical investigation and optimization of synthesized thermoelectric materials for power generation. *Results in Engineering*, 109011.
[33]. Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in Neural Information Processing Systems, 28*, 505–513.
[34]. Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in Neural Information Processing Systems, 28*, 505–513.
[35]. Raff, E., Sylvester, J., Forsyth, S., & McLean, M. (2020). Scaling deep learning models for student dropout prediction. *Proceedings of the 13th International Conference on Educational Data Mining*, 344–354.
[36]. Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254.
[37]. Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254.
[38]. Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254.
[39]. Sweeney, M., Lester, J., & Rangwala, H. (2016). Next-term student performance prediction: A recommender systems approach. *Journal of Educational Data Mining, 8*(1), 22–51.
[40]. Sweeney, M., Lester, J., & Rangwala, H. (2016). Next-term student performance prediction: A recommender systems approach. *Journal of Educational Data Mining, 8*(1), 22–51.
[41]. Sweeney, M., Lester, J., & Rangwala, H. (2016). Next-term student performance prediction: A recommender systems approach. *Journal of Educational Data Mining, 8*(1), 22–51.
[42]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008.
[43]. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008.
[44]. Xu, B., & Recker, M. (2012). Teaching analytics: A clustering and triangulation study of digital library user data. *Educational Technology & Society, 15*(3), 103–115.
[45]. Yeung, C. K., & Yeung, D. Y. (2018). Addressing two problems in deep knowledge tracing via prediction-consistent regularization. *Proceedings of the 5th ACM Conference on Learning @ Scale*, 1–10.