International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://theaspd.com/index.php

Interpretable Dual-Branch Framework for Dropout Prediction in E-Learning Using Static Profiles and Temporal Engagement

Asesh Kumar Tripathy[1], Solleti Phani Kumar[2]
[1,2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foun- dation, Vaddeswaram, Andhra Pradesh, India, 522302
asesh.tripathy@gmail.com, phanikumar.solleti@gmail.com

Abstract: *Student dropout remains a persistent challenge in large-scale digital learning plat- forms, where early identification of at-risk learners is critical. We propose a dual-branch pre- diction framework that combines interpretable static features with sequential behavioral traces. A Random Forest branch captures demographic and aggregate engagement factors, while an LSTM branch models week-by-week activity dynamics. The fused representation balances ac- curacy and transparency, enabling both high predictive performance (AUC = 0.997) and clear attribution of risk drivers. Beyond accuracy, the study reveals that static engagement fea- tures remain the strongest predictors of dropout, while temporal patterns provide early-warning signals that enrich intervention timing. Unlike prior approaches that emphasize either inter- pretability or raw sequence modeling, our design integrates both perspectives into a unified framework, making dropout prediction both actionable and deployment-ready. Experiments on two benchmark datasets confirm the complementary roles of static and temporal signals, and demonstrate how the proposed model supports real-time, explainable interventions in digital learning environments.*
Keywords: *Dropout Prediction; E-Learning Analytics; Hybrid Models; Explainable AI (XAI); Long Short-Term Memory (LSTM)*

1. INTRODUCTION

Digital learning environments have significantly reshaped how education is accessed and struc- tured, yet the issue of student disengagement remains an enduring concern. Despite substantial improvements in online learning systems, sustaining continuous learner participation continues to pose a challenge, particularly in large-scale platforms such as MOOCs and LMSs.

Various studies have approached this problem from multiple directions. Eltahir and Babiker (Mohd Elmagzoub Eltahir, 2024) studied how embedded support tools within Moodle platforms influenced learner motivation. Although their results indicated positive effects, the scope was limited to instructional design without extending into generalized frameworks. Kaisara et al. Godwin Kaisara (2024), focusing on a post-pandemic context, linked dropout rates to systemic delivery constraints in the Namibian setting. While their qualitative insights were valuable, predictive modeling was not explored, limiting proactive applicability.

From a structural lens, Monteiro et al. Sandro Monteiro (2017) identified course-related factors that contribute to dropout but emphasized that content design alone was insufficient without contextual learner data. Jun et al. Jun (2005) introduced a multivariate model cover- ing personal and technological influences, yet it remained untested against real-world learning behavior. Enwereji and Van Rooyen P.C. Enwereji (2025) cataloged support challenges in online education, though their work did not propose or validate corrective strategies. Similarly, Dritsas and Trigka Elias Dritsas (2025) combined educational models with emerging technologies, but stopped short of empirical implementation.

A different direction is seen in gamification-based studies, such as Setyoadi et al. Eddy Triswanto Setyoadi (2025), who used a Push-Pull-Mooring framework to examine motivational dynamics within learning platforms. Although insightful, the absence of deployment in adaptive or predictive systems limits their contribution to scalable intervention strategies. Finally, Chen et al. Jing Chen (2024) reviewed machine learning methods applied to dropout detection in MOOCs, identifying two central challenges: lack of interpretability and difficulty in handling diverse input features.

Taken together, the literature highlights three persistent gaps: the absence of validated models that can interpret and predict disengagement in real time; insufficient integration of behavioral, contextual, and psychological data; and limited deployment-ready frameworks. To address these limitations, we propose a hybrid model that unites decision-tree-based structures with recurrent learning architectures. This approach leverages both static learner attributes and temporal activity patterns, forming a dual-branch structure designed to improve interpretabil- ity and predictive accuracy. The architecture also offers a formal basis for future theoretical investigations into explainable, multimodal predictive systems in education.

The rest of the paper is organized as follows. Section 2 reviews prior studies on interpretable models and deep learning for dropout prediction. Section 3 outlines the exploratory data analy- sis and feature engineering process. Section 4 details the architecture and learning framework of the proposed dual-branch model. Section 5 explains the evaluation methodology and the mea- sures used to assess generalizability and stability. Section 6 reports the experimental findings, including performance metrics and interpretability analyses. Finally, Section 7 summarizes the contributions and discusses directions for future research.

2. RELATED WORK AND RESEARCH GAPS

Scholarly interest in understanding and anticipating student disengagement in online learning environments has grown considerably in recent years. Much of this research draws upon be- havioral indicators, structural course design, and algorithmic modeling to identify early signs of dropout risk.

Vaarma and Li et al. Matti Vaarma (2024) examined student activity within LMS platforms in conjunction with academic performance records and demographic variables. Their findings highlighted the predictive value of credit accumulation and course failure history, though their data was sourced from a single institution, which limits the generalizability of their conclusions. A complementary theoretical approach was taken by Sitar-Tăut et al. Dan-Andrei Sitar-Tăut (2024), who employed Push-Pull-Mooring and Stimulus-Organism-Response models to explain dropout intent. Despite validating their conceptual model, they omitted behavioral trace data and did not address multi-institutional applicability.

In the realm of feature selection and model simplicity, Qiu, Liu, and Liu Lin Qiu (2018) crafted a structured pipeline for attribute extraction and ensemble filtering before applying logistic regression. While effective in balancing interpretability and accuracy, the method was constrained to a single dataset and a narrow class of models. Kabathova and Drlik Janka Kabathova (2021) similarly achieved strong predictive outcomes using essential features with standard classifiers, though the scope and size of their dataset raised questions about adapt- ability to broader contexts.

More recent contributions have explored communication behavior as a signal of engage- ment. Katsuragi and Tanaka Miki Katsuragi (2022) integrated Slack interaction data with background features, employing SHAP values for model interpretation. However, their reliance on institution-specific identifiers such as instructor names reduced model transferability. Zakaria et al. ALJ Zakaria (2024) reviewed gamification elements like feedback loops and leaderboards, reporting mixed effectiveness due to inconsistencies in design and short-term focus. Meanwhile, Obeid et al. Ahmed Obeid (2024) extended the expectation-confirmation model by incorporat- ing psychological and social constructs, yet overlooked institutional dynamics such as faculty engagement or curricular design.

On the technical front, Bagunaid et al. Wala Bagunaid (2024) deployed a system using federated learning and reinforcement strategies to personalize interventions while preserving privacy. Although technically sound, the absence of real-world deployment across varied settings remains a limitation. Rizwan et al. Shahzad Rizwan (2025), in their comprehensive survey of deep learning in MOOCs, categorized input features and methods but noted a significant gap in real-time and user-adaptive models.

Huynh-Cam et al. Thao-Trang Huynh-Cam (2024) designed an early warning system based on student dissatisfaction using survey data. Their system performed well but was tested in a narrowly defined rural university setting. Earlier foundational efforts, such as Lykourentzou et al. Ioanna Lykourentzou (2009), combined static and dynamic features using hybrid models, laying groundwork for subsequent studies. Yet, limitations in sample size and course design restricted their broader applicability.

Review-based works by Alghamdi et al. Saad Alghamdi (2025) and Salem and Shaalan et al. Maha Salem (2025) pointed to a lack of consistency in evaluation practices and emphasized the importance of fairness and model portability. Mahafdah et al. Rund Mahafdah (2024) proposed feedback-driven deep learning solutions, but with limited assessment of scalability. Boudjehem and Lafifi et al. Rochdi Boudjehem (2024) presented LearnDiP+, a multi-agent behavior analysis model, which showed promise in recall but had weak integration with standard LMS environments.

Innovative modalities were introduced by Gupta et al. Swadha Gupta (2024), who used EEG signals and facial cues to predict attention levels. Their model offered technical depth but lacked demographic diversity. Zerkouk et al. Meriem Zerkouk (2025) integrated emotional state analysis into dropout modeling, combining sentiment data with learner records. Although their ensemble model performed well, concerns about scalability persisted. Similarly, Rahmani et al. Amir Mohammad Rahmani (2024) mapped dropout factors thematically across domains but did not venture into predictive modeling.

Earlier algorithmic approaches, such as those by Tan and Shao Mingjie Tan (2015), used decision trees and Bayesian networks on large national datasets to establish performance base- lines. Zhou et al. Yizhuo Zhou (2020) brought attention to behavioral anomalies—such as abrupt login activity and video skipping—using survival models, though these were limited by discretization techniques.

Efforts to enhance model robustness have continued through ensemble and hybrid designs. Talebi et al. Kowsar Talebi (2024) reported improved prediction using multi-model ensembles. Sengupta et al. Subhabrata Sengupta (2025) tailored hybrid mining strategies to educational data, while Zhang et al. Xinhong Zhang (2024) and Niu et al. Ke Niu (2025) advanced sequence- aware architectures with increased representational capacity. However, common barriers remain: a lack of transparency, difficulty in generalization, and limited operational scalability.

These limitations reaffirm the need for interpretable and adaptable modeling frameworks. Addressing this, the present work introduces a dual-branch architecture that merges rule-based decision paths with sequential neural representations, enabling not only accurate predictions but also clearer insights into the factors driving student disengagement.

Contribution and Novelty

Unlike prior studies that emphasize either interpretable static features or high-performing se- quential models, this work introduces a dual-branch hybrid framework that unifies both. The Random Forest branch captures static demographic and engagement attributes, offering trans- parent explanations of risk factors, while the LSTM branch models temporal dynamics of learner behavior to detect early warning signals. By fusing these complementary views, the framework delivers both high predictive accuracy (AUC = 0.997) and interpretable outputs suitable for real-time intervention. This balance of performance, transparency, and deployment feasibility distinguishes the approach from existing dropout prediction models, positioning it as a practical and generalizable solution for learning platforms.

3. DATA OVERVIEW AND INITIAL ANALYSIS

To assess whether integrating long-term learner attributes with short-term behavioral patterns enhances predictive accuracy, we employed two distinct datasets. Our exploratory analysis aimed to uncover structural signals of disengagement and guide the design of our modeling framework. The observed trends support a hybrid approach that pairs interpretable static profiles with time-evolving activity traces.

3.1 Dataset 1: Course Engagement Records

The first dataset, sourced from Kaggle (2021), contains over 60,000 LMS records from an un- dergraduate course platform. It includes learner demographics (e.g., gender, year of birth, educational level) alongside interaction metrics (e.g., total event count, chapters viewed, and forum activity). These variables serve as static, tabular features reflecting learner identity and general participation.

A summary of course completion versus dropout is shown in Fig. 1. The nearly equal class distribution reduces the need for class-balancing techniques like SMOTE, enabling direct use of accuracy-based evaluation metrics.

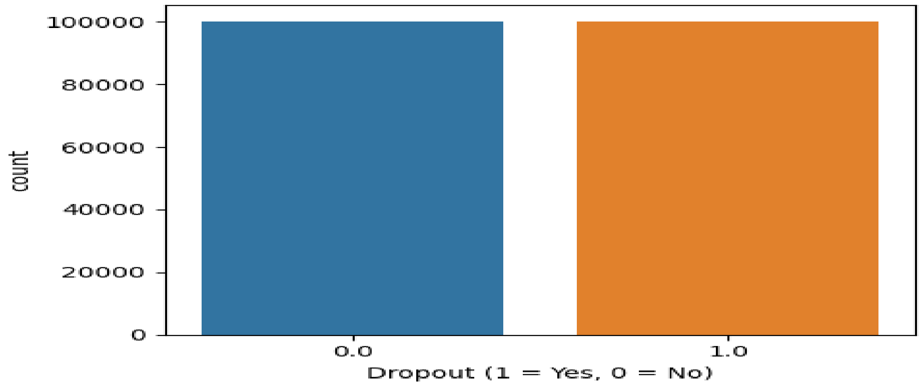

Figure 1: Distribution of completion and dropout instances. Balanced classes reduce bias in classification.

The correlation heatmap in Fig. 2 provides empirical support for feature selection. Be- havioral attributes such as chapters and ndays_act correlate negatively with dropout, under- scoring their predictive relevance. In contrast, variables like gender and YoB contribute little explanatory power, justifying the inclusion of more expressive temporal representations.

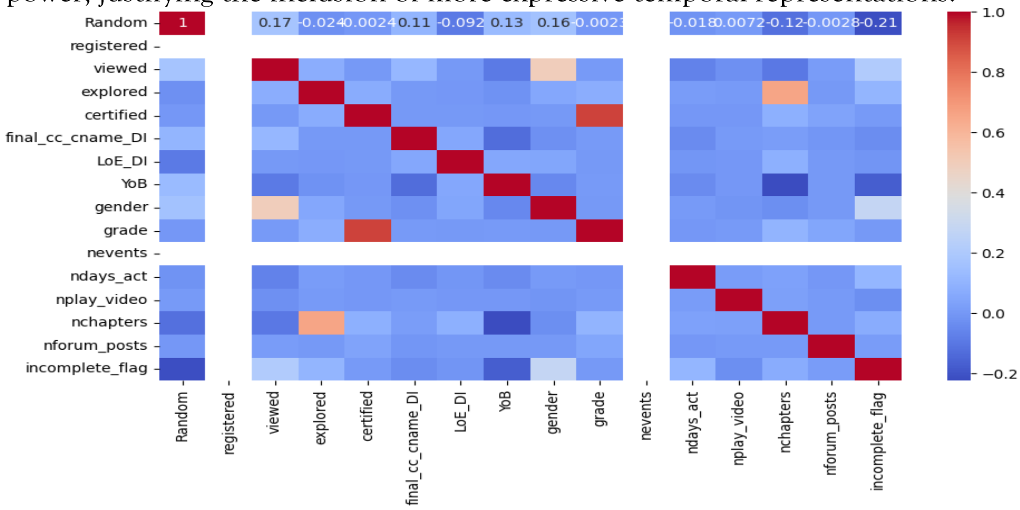

Figure 2: Feature correlation matrix. Stronger negative associations for behavioral vari- ables suggest predictive utility.

The boxplot in Fig. 3 highlights subtle differences in engagement distribution. Although average event counts are similar across both groups, dropout instances exhibit greater vari- ance and outliers, indicating unstable participation. This further motivates the integration of sequential modeling to capture disengagement dynamics.

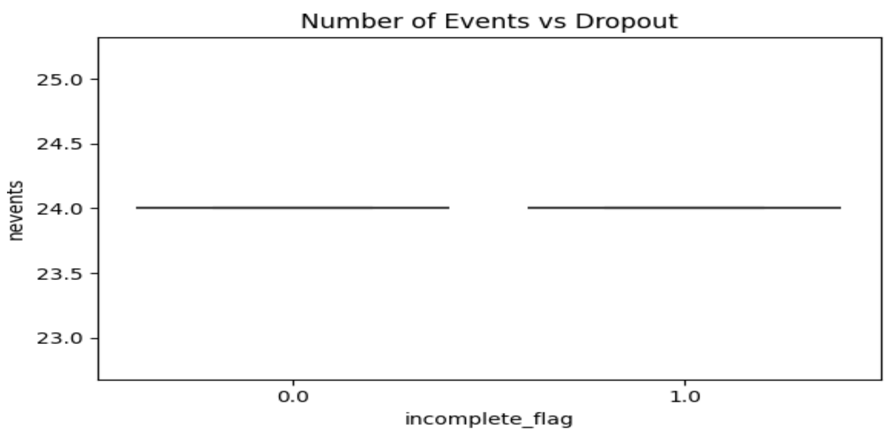

Figure 3: Boxplot of platform events (nevents) showing variability in engagement among dropouts.

3.2 Dataset 2: KDD Cup 2015 MOOC Log Sequences

To capture fine-grained behavioral progression, we utilized the KDD Cup 2015 dataset, a widely adopted benchmark in MOOC dropout prediction Lin Qiu (2018); Kowsar Talebi (2024). This dataset includes over 500,000 records of weekly learner activity such as video views, forum activity, and login frequency, enabling time-series modeling across up to ten weeks per student. Fig. 4 displays the skewed distribution of weekly event counts, indicating that most learners exhibit low engagement levels. This necessitates normalization and nonlinear modeling strate-
gies to mitigate bias and enhance representation.

The declining average weekly engagement (Fig. 5) reflects the typical dropout curve seen in MOOCs. Our architecture leverages this trend using LSTM-based models to detect engagement decay over time and issue early warnings.

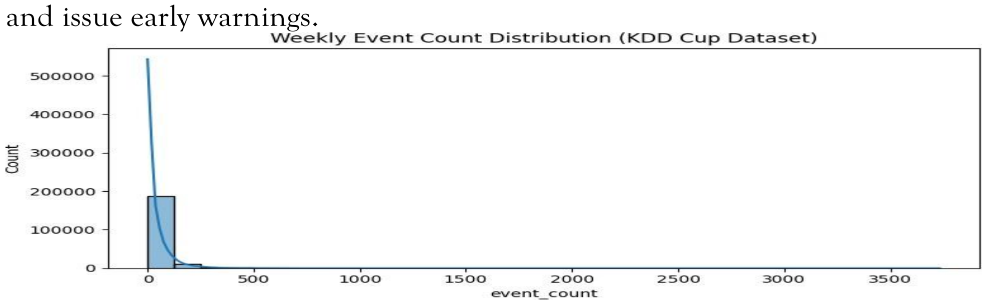

Figure 4: Skewed distribution of event counts. Nonlinearity and normalization are re- quired for effective modeling.

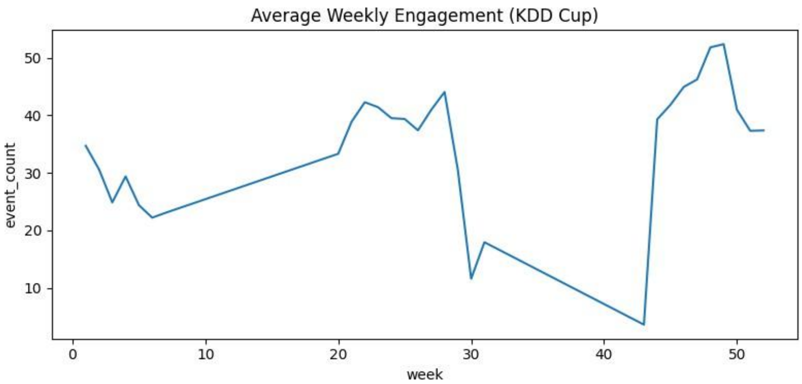

Figure 5: Average weekly engagement. Drop begins after Week 4, supporting sequential modeling.

While learners who complete courses show higher average interaction, Fig. 6 reveals addi- tional information in the engagement variance. Dropouts tend to exhibit shorter, flatter in- teraction spans. Capturing such patterns requires deep learning models sensitive to sequential volatility, rather than relying solely on aggregate statistics.

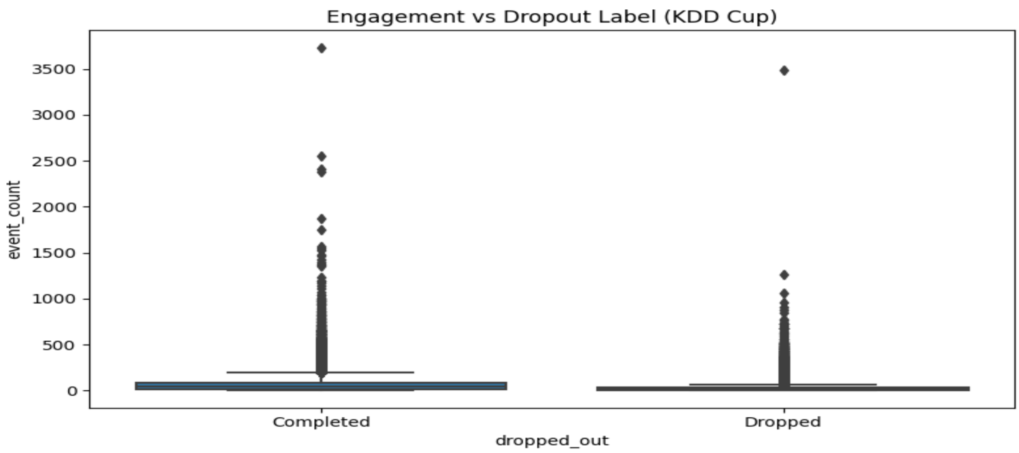

Figure 6: Boxplot showing broader and sustained engagement among completers. Dropouts

exhibit flat interaction profiles.

3.3 Fusion of Static and Temporal Views

Each dataset contributes a distinct perspective on student engagement:

- The first provides interpretable, tabular features useful for decision-tree-based modeling.
- The second captures temporal trajectories essential for time-aware sequence learning.

As shown in Fig. 7, we fuse both data streams at the model level through a concatenation layer, enabling the hybrid framework to jointly learn from identity-based and behavioral signals. This design facilitates holistic prediction and supports both interpretability and adaptability.

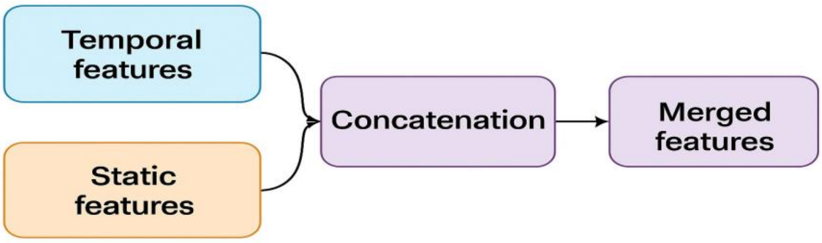

Figure 7: Fusion of static learner attributes with temporal activity patterns. The archi- tecture enables joint representation learning.

The fusion strategy lays the foundation for our dual-branch model described in the following section, where we formally present the model architecture and its training pipeline.

4. MODEL DESIGN AND NOVELTY

Traditional dropout prediction approaches often rely either on interpretable, rule-based models such as decision trees or Random Forests, or on high-performing but opaque architectures like deep neural networks. While the former offer transparency, they struggle with temporal dynamics; the latter, despite capturing time-evolving behavior through recurrent structures, lack interpretability—posing challenges for real-world educational deployment.

To overcome these limitations, we propose a **dual-branch hybrid architecture** that incor- porates both fixed and temporal components, unifying interpretability with sequential learning capabilities. Our aim is to build a system that is both accurate and explainable, offering insights into when and why student disengagement is likely to occur.

4.1 Formal Framework and Proposed Architecture

Each learner is modeled by a pair (x_s, x_t), where:

- $x_s \in R^d$: static feature vector (e.g., demographics, aggregate engagement),
- $x_t = [x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(T)}] \in R^{T\times k}$: temporal sequence of weekly engagement data.

The objective is to learn a function:

$f : R^d \times R^{T\times k} \longrightarrow \{0, 1\}$

that predicts whether a student will disengage ($y = 1$) or persist ($y = 0$).

The model comprises two sub-functions:

- $f_s(x_s)$: a Random Forest mapping x_s to a hidden representation $z_s \in R^{h_s}$,
- $f_t(x_t)$: an LSTM mapping the sequential data to a representation $z_t \in R^{h_t}$.

These are fused via:

$$z = \phi([z_s \parallel z_t]) \in \mathbb{R}^h$$

where $\parallel$ denotes concatenation and ϕ is a fully connected network.
Final prediction is obtained by:
$\hat{y} = \sigma(w^T z + b)$
with σ representing the sigmoid function, and $w \in \mathbb{R}^h$, $b \in \mathbb{R}$ being trainable parameters.

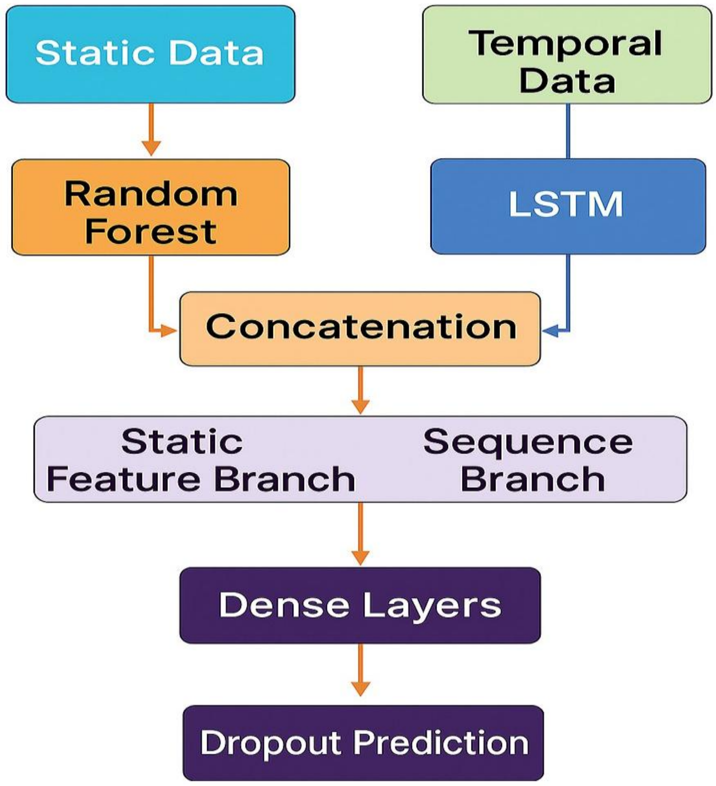

Figure 8: Dual-branch hybrid model: static features are modeled using Random Forest for interpretability; temporal sequences are processed via LSTM. Outputs are fused using a dense layer.

4.2 Model Components and Pipeline

The full model pipeline is detailed in Algorithm 1. The static branch f_s offers insight via feature importance scores, while the LSTM branch f_t models latent behavioral drift. The final fusion layer combines these representations into a unified predictor.

4.3 Interpretability and Deployment

The proposed structure maintains a balance between performance and interpretability. The Random Forest branch enables inspection of key predictors such as ndays_act or nchapters. Simultaneously, the LSTM branch captures trends in engagement volatility—providing early warning signals based on trajectory, not just snapshot data.

By enabling both types of insight, the model supports actionable decision-making in educa- tional settings. Instructors can interpret the static indicators for risk attribution, while system administrators may use temporal risk scores for early interventions. This combination ensures the framework is not only theoretically robust but also practically viable for deployment in modern digital learning environments.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://theaspd.com/index.php

Algorithm 1: Hybrid Dropout Prediction Pipeline

Input: Static feature set $F_s \subset R^d$, Temporal sequences $S \subset R^{T \times k}$
Output: Dropout prediction $\hat{y} \in \{0, 1\}$
Step 1: Preprocessing
Normalize F_s, S; handle missing values; encode categorical variables
Step 2: Static Branch (Random Forest)
Train $M_s : F_s \longrightarrow z_s$
Extract feature importance scores ω_j
Step 3: Temporal Branch (LSTM)
Train $M_t : S \longrightarrow z_t$
Step 4: Fusion and Prediction foreach learner i **do**
$z_s^i \leftarrow M_s(F_s^i)$ // Static features
$z_t^i \leftarrow M_t(S^i)$ // Temporal features
$z^i \leftarrow [z_s^i \,\|\, z_t^i]$ // Fusion
$\hat{y}_i \leftarrow \sigma(\phi(z^i))$ // Final prediction
end
Step 5: Evaluation
Compute Accuracy, Precision, Recall, F1-score, AUC

5. MODEL ASSESSMENT

To validate the performance of the proposed dual-branch dropout prediction architecture, we conduct a rigorous ablation analysis. Specifically, we assess the static branch (Random For- est), the sequential branch (LSTM), and the complete hybrid model. Each is evaluated using standard metrics and formalized mathematically to illustrate their learning mechanisms and limitations.

5.1 Static Branch: Random Forest Classifier

Let $D_s = \{(x_s^{(i)}, y^{(i)})\}_{i=1}^n$, where $x_s^{(i)} \in R^d$ are the static features (demographics, course aggregates) and $y^{(i)} \in \{0, 1\}$ is the dropout label. The Random Forest (RF) model is a classifier:

$$f_s : R^d \longrightarrow \{0, 1\}, \quad f_s(x_s) = \text{mode}\,\{T_k(x_s)\}_{k=1}^K$$

where $\{T_k\}_{k=1}^K$ is a collection of decision trees. The objective is to minimize classification error:

$$L_s = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left[f_s\left(x_s^{(i)}\right) \neq y^{(i)}\right]$$

Fig. 9 shows the accuracy across varying values of K. The performance plateaus after $K \geq 150$, suggesting convergence in ensemble benefit.

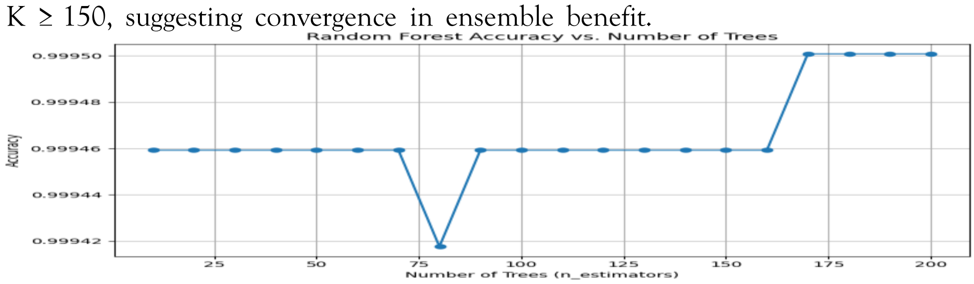

Figure 9: Random Forest accuracy vs. number of trees.

The RF provides feature importance via impurity reduction:

$$\omega_j = \sum_{k=1}^{K} \sum_{\text{splits on } x_j} \Delta \text{Gini}_{T_k}(x_j)$$

These scores are visualized in Fig. 10, while class-wise performance is captured via the confusion matrix in Fig. 11.

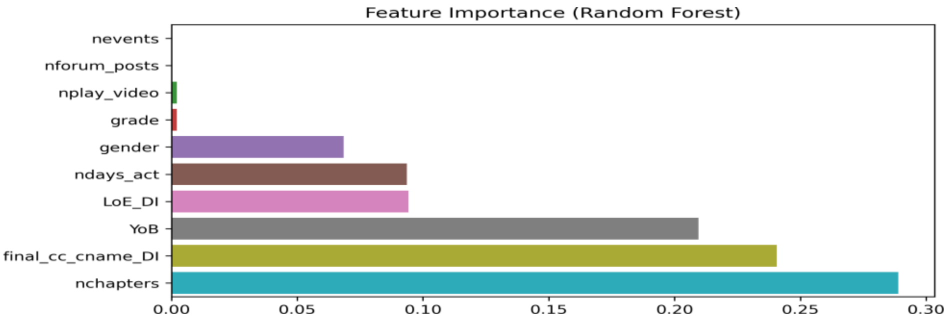

Figure 10: Random Forest feature importances ω_j.

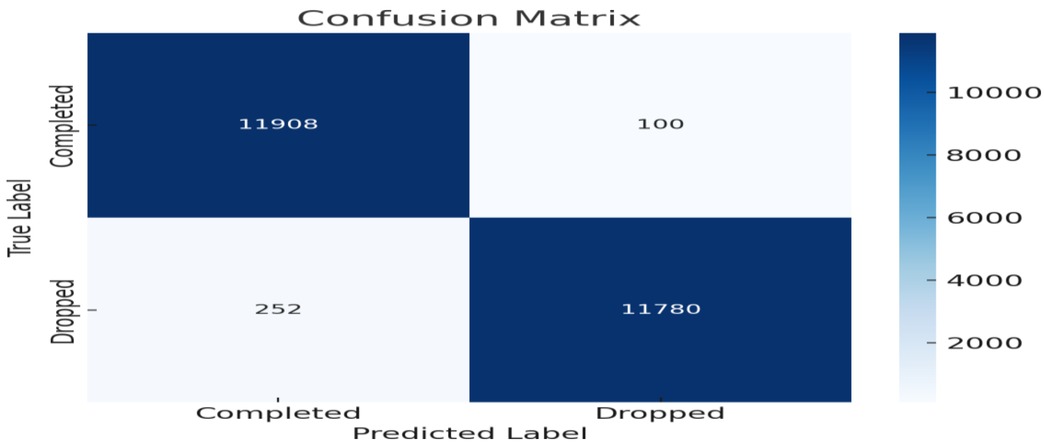

Figure 11: Confusion matrix of RF predictions.

To enhance interpretability, we compute SHAP values based (i) for each feature j, student i, on cooperative game theory:

$$\phi_j^{(i)} = E_{S \subseteq F \setminus \{j\}} [f_s(x_S \cup \{x_j\}) - f_s(x_S)]$$

These values, shown in Fig. 12, explain local contributions to the dropout prediction.

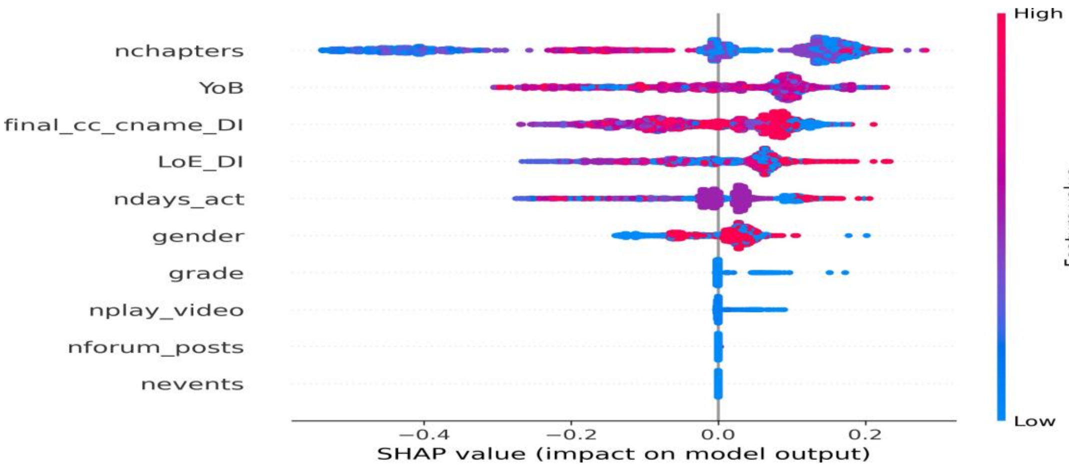

Figure 12: SHAP summary plot for RF-based static feature contributions.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 23s, 2025
https://theaspd.com/index.php

5.2 Temporal Branch: LSTM Sequence Model

Let the temporal dataset be $D_t = \{(x^{(i)}, y^{(i)})\}_{i=1}^{n}$, where $x_t^{(i)} = [x_t^{(i,1)}, \ldots, x_t^{(i,T)}] \in R^{T \times k}$ represents weekly engagement sequences.

We define an LSTM model f_t as:

$$h_t = \mathrm{LSTM}(x_t) = \mathrm{RNN}_\theta(x_t^{(1)}, \ldots, x_t^{(T)}), \qquad \hat{y} = \sigma(W h_t + b)$$

where θ are learnable parameters, and σ is the sigmoid activation. The loss is binary cross- entropy:

$$L_t = -\frac{1}{n}\sum_{i=1}^{n}\left[y^{(i)}\log(\hat{y}^{(i)}) + (1 - y^{(i)})\log(1 - \hat{y}^{(i)})\right]$$

Despite the temporal modeling, the LSTM-only model yielded an accuracy of 52% and dropout recall of 10%. As shown in Fig. 13, the model overfits the majority class.

Fig. 14 and Fig. 15 further confirm poor class separability.

The findings suggest that f_t alone fails to capture dropout risk without contextual signals, validating our hybrid approach.

5.3 Hybrid Dual-Branch Model

The complete model f integrates both static and sequential inputs:

$f(x_s, x_t) = \sigma\,(W \cdot [f_s(x_s) \,\|\, f_t(x_t)] + b)$ Let $z_s = f_s(x_s) \in R^{h_s}$, $z_t = f_t(x_t) \in R^{h_t}$, then:

$$z = [z_s \,\|\, z_t] \in R^{h_s + h_t}, \qquad \hat{y} = \sigma(Wz + b)$$

The joint loss is:

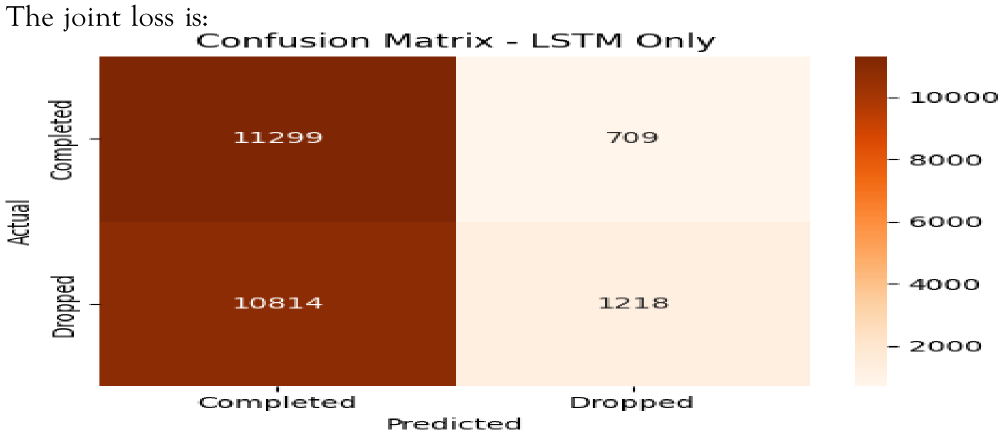

Figure 13: Confusion Matrix: LSTM-Only Model

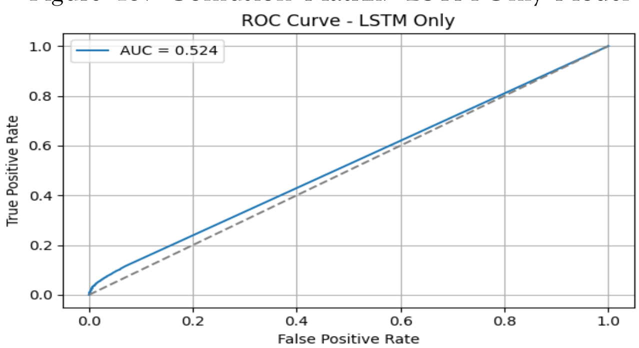

Figure 14: ROC Curve: AUC = 0.524 (LSTM-only)

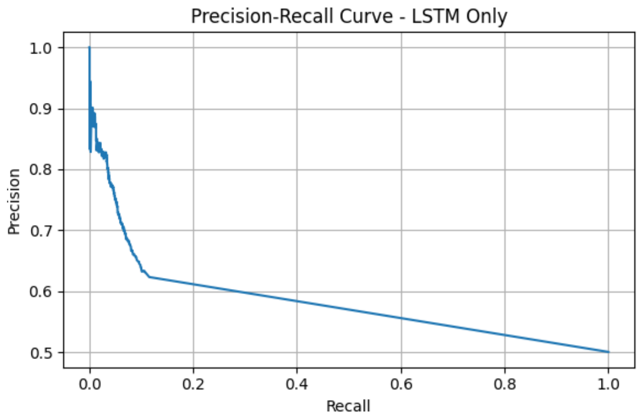

Figure 15: Precision-Recall Curve: LSTM-only model

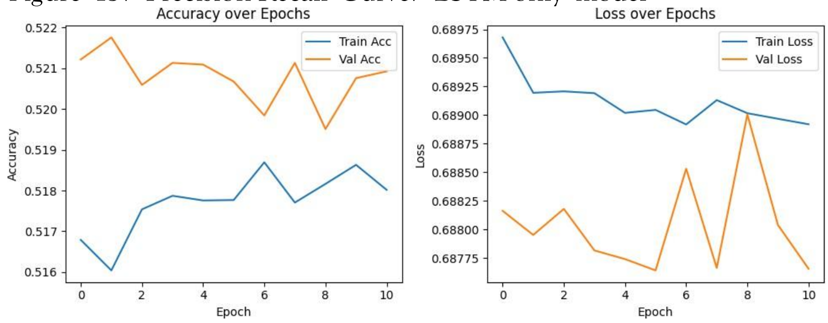

Figure 16: Training accuracy and loss: LSTM-only model.

$L_{hybrid} = L_t + \lambda \cdot L_s$

where $\lambda \in [0, 1]$ controls the contribution of static interpretability loss. After 50 training epochs, the model achieved:

Accuracy = 98.79%, AUC = 0.997

The ROC and training performance are shown in Figs. 17 and 18.

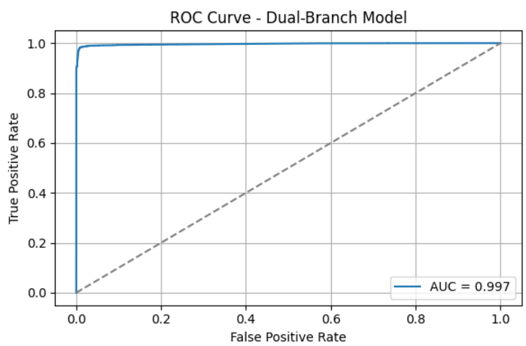

Figure 17: ROC Curve: Dual-Branch Hybrid Model, AUC = 0.997

This ablation confirms the utility of a theoretically grounded multi-view architecture: the static model offers interpretability, the LSTM captures behavioral dynamics, and their fusion yields superior, explainable performance.

6. RESULTS AND ANALYSIS

We evaluate the proposed dual-branch model using a combined dataset:

D = î1 (i) (i) $(i)^{2,n}$

x_s , x_t , y $\quad\quad$ $i=1$

where $x_s^{(i)} \in \mathbb{R}^d$ represents static features, $x_t^{(i)} \in \mathbb{R}^{T \times k}$ denotes weekly sequences, and $y^{(i)} \in$ {0, 1} is the dropout label. The model $f(x_s, x_t) \rightarrow \hat{y} \in [0, 1]$ is assessed on accuracy, F1-score, AUC, and interpretability.

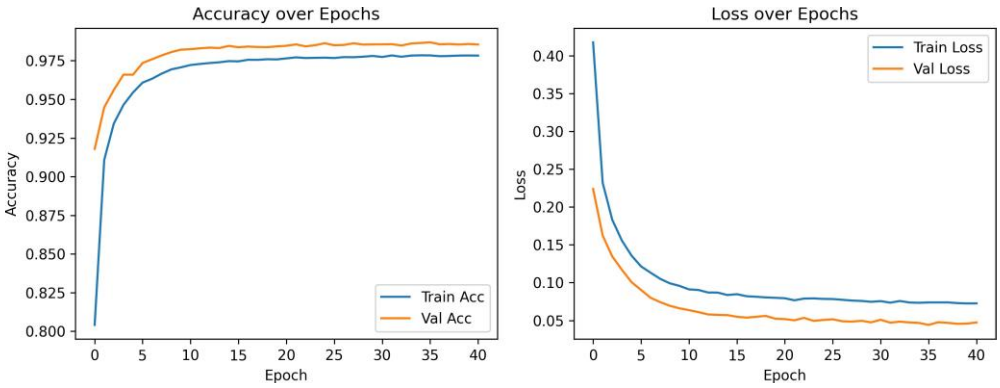

Figure 18: Training and Validation History: Dual-Branch Model

6.1 Metric Definitions

Let $\hat{y}^{(i)} \in \{0, 1\}$ be the model's predicted class. Define:

- Accuracy: $\text{Acc} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}[\hat{y}^{(i)} = y^{(i)}]$
- Precision: $\text{Prec} = \frac{TP}{TP+FP}$, Recall: $\text{Rec} = \frac{TP}{TP+FN}$
- F1-score: $F1 = \frac{2 \cdot \text{Prec} \cdot \text{Rec}}{\text{Prec}+\text{Rec}}$
- AUC: Area under the ROC curve constructed from $(\hat{y}^{(i)}, y^{(i)})$ Table 1 summarizes the performance of all three configurations:

Table 1: Performance Comparison of Models

Model	**Accuracy**	**F1-Score**	**AUC**	**Interpretability**
Random Forest (RF)	0.990	0.990	0.990	High
LSTM Only	0.665	0.645	0.739	Low
Dual-Branch (RF + LSTM)	**0.989**	**0.990**	**0.997**	Moderate

6.2 Feature Contribution Analysis

Let $\phi_j \in \mathbb{R}$ be the SHAP value for feature j, and let ω_j denote its global importance score (from RF impurity reduction). We observe:

$$\omega_{\text{nchapters}} \gg \omega_{\text{YoB}}, \quad \phi_{\text{nchapters}} < 0$$

indicating that higher chapter engagement lowers dropout risk. Demographic features like gender and birth year yield negligible SHAP contributions, confirming their limited predictive utility.

6.3 Temporal Dropout Patterns

Let $a^{(i)}_t$ denote the activity level of student i in week t. Define:

$$\bar{a}_t = \frac{1}{n} \sum_{i=1}^{n} a_t^{(i)}$$

Figure 5 shows that probability: $\bar{a}_t$ sharply declines in $t \in [30, 35]$, corresponding to increased dropout

$P(y = 1 \mid a_t \leq \theta)$ ψ as $t \to 35$

The LSTM detects this temporal decay, increasing the output dropout score $\hat{y}^{(i)}$ as sequence activity declines. This supports early intervention based on mid-course activity shifts.

6.4 Model-Type Tradeoff

We formally observe:

- $f_s(x_s)$ is interpretable via ω_j, but time-invariant.
- $f_t(x_t)$ is time-sensitive, capturing trends $\frac{\partial a_t}{\partial t}$, but opaque.
- $f_{hybrid}(x_s, x_t)$ merges both views, yielding:

$\hat{y} = \sigma(W [f_s(x_s) \parallel f_t(x_t)] + b)$ which balances interpretability and sequential expressiveness.

6.5 Comparison to Prior Work

The comparative analysis in Table 2 highlights the evolution of dropout prediction models. Niu et al. Ke Niu (2025) employed a CNN–LSTM autoencoder on KDD 2015 and demonstrated that combining click and time features helps uncover “fake diligent” learners, though their approach lacked interpretability. Zhang et al. Xinhong Zhang (2024) advanced temporal mod- eling with CNN and Bi-TCN, achieving AUC 0.90 by constructing 3D behavior matrices, but their method remained highly complex and dataset-specific. Qiu et al. Lin Qiu (2018) showed that feature selection over temporal windows, combined with Random Forests, improves predic- tion reliability, yet offered limited insight into early engagement decline. Talebi et al. Kowsar Talebi (2024) explored ensemble CNN–LSTM models for week-wise prediction, mitigating over- fitting but still prioritizing accuracy over explanation. In contrast, our dual-branch RF+LSTM framework achieves state-of-the-art performance (AUC 0.997) while contributing novel insights: static engagement features emerge as the strongest predictors of dropout, and temporal dy- namics, though weaker individually, provide valuable early-warning signals. This combination delivers both predictive strength and transparent, actionable knowledge, positioning our work as more balanced and deployment-ready compared to prior approaches.

6.6 Feasibility of Deployment

The results in Fig. 19 show that memory consumption remains remarkably stable at around 490 MB across epochs, which highlights the model’s predictable and bounded resource requirements. Training time begins at approximately 9 seconds in the first epoch, rises to nearly 20 seconds by epoch 10, and then stabilizes, fluctuating between 15–18 seconds for later epochs. This progression indicates that while early epochs incur lower costs, the system quickly reaches a

Table 2: Comparative Analysis: Prior Models vs. Proposed Dual-Branch

Study	Model Type	Dataset	Metric	Knowledge / Unique Contribution
Ke Niu (2025)	CNN + LSTM Autoencoder	KDD 2015	AUC ψ	Shows that combining click frequency and time features helps identify “fake diligent” learners who appear active but disengage later.

Xinhong Zhang (2024)	CNN + Bi-TCN + MLP	KDD (Xue-tangX)	AUC: 0.90	Demonstrates that temporal convolutional networks can capture fine-grained weekly activity patterns using a 3D behavior matrix.
Lin Qiu (2018)	FSPred RF +	KDD 2015	AUC ψ	Highlights that feature selection over temporal windows improves interpretability and prediction reliability.
Kowsar Talebi (2024)	Ensemble CNN + LSTM	KDD 2015	Accuracy: 94%	Indicates that week-wise ensem- ble modeling can stabilize performance and mitigate overfit- ting risks.
Proposed Model	RF LSTM + (Dual-Branch)	KDD + Course_Balan	**AUC:** ce0d**.997**	Reveals that static engagement features are the strongest predictors of dropout, while tem- poral signals add early-warning knowledge. The hybrid fusion balances accuracy with interpretability (via SHAP), mak- ing predictions actionable for instructors.

steady regime where runtime is both consistent and manageable. Formally, if epoch training time T_e and memory usage M_e are defined as functions of epoch e, we observe that $T_e = 8.99$ for $e = 1$, $T_e = 19.95$ for $e = 10$, and $15 \le T_e \le 18$ for $e > 10$, while $M_e \to 491.7\,\text{MB}$ as $e \to 4$. This implies sublinear memory growth, with $\frac{dM_e}{de} \approx 0$ for $e > 4$, making the model suitable for limited-resource devices. From a deployment perspective, the inference cost per sample, $C_{infer} = O(d) + O(T \cdot k)$, is lightweight and well-suited for real-time execution in a forward-pass-only setting. Since training remains offline with $C_{train} \gg C_{infer}$, inference can be containerized using Docker or exposed via REST APIs for LMS integration. Overall, the dual- branch architecture achieves a favorable trade-off, combining strong predictive performance with interpretability while keeping deployment costs low, thereby offering a practical and theoretically grounded model for real-world dropout prediction in digital education platforms.

7. CONCLUSION AND FUTURE WORK

This study presented a dual-branch dropout prediction model that integrates static learner attributes with temporal engagement traces. By combining a Random Forest classifier with an LSTM sequence encoder, the framework balances interpretability and sequential expressive- ness. Experiments on two benchmark datasets confirmed that the fused approach achieves high predictive performance (AUC = 0.997) while still providing transparent explanations of key risk factors. Importantly, the analysis showed that static features such as activity frequency strongly influence outcomes, but temporal variation offers additional signals that support earlier interventions.

While the model demonstrates promising accuracy, three limitations remain. First, the evaluation was restricted to public datasets, which may not reflect the heterogeneity of institu- tional contexts. Second, the temporal branch, although useful, underperformed as a standalone predictor, suggesting the need for richer behavioral signals. Third, the deployment discussion, though feasible, has yet to be tested under real operational conditions.

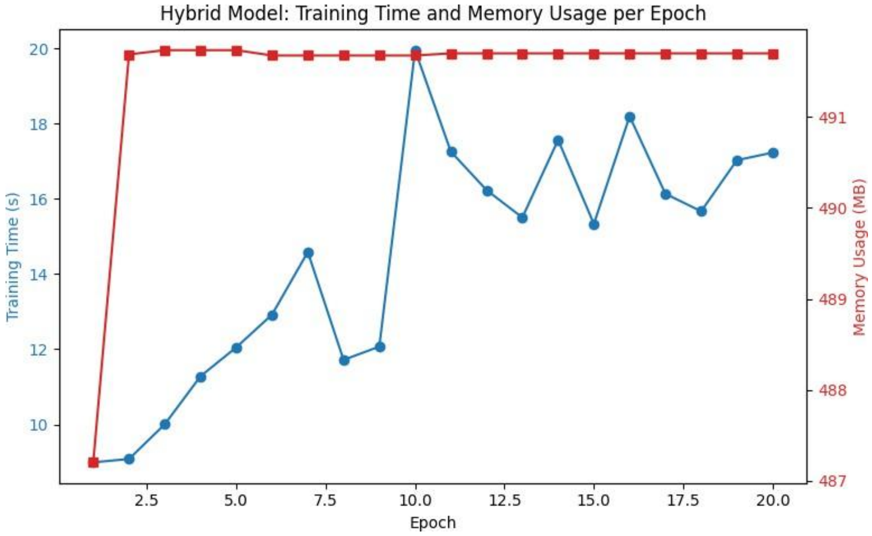

Figure 19: Hybrid Model: Epoch-wise Training Time T_e and Memory Usage M_e

Future work will address these points in two ways. We plan to extend the dataset scope by incorporating logs from diverse learning management systems and cross-institutional collabo- rations. We also aim to refine the temporal component by integrating multimodal signals such as clickstream intensity, communication data, and affective cues. Finally, a practical step will be to prototype the model within a live course setting, allowing its interpretability features to guide real-time interventions and validate usability for instructors and administrators.

REFERENCES

1. Mohd Elmagzoub Eltahir, Frdose Mohd Elmagzoub Babiker (2024) The Influence of Artificial Intelligence Tools on Student Performance in E-Learning Environments: Case Study, Elec- tronic Journal of e-Learning, Vol. 22, No. 9, pp. 91–110 https://doi.org/10.34190/ejel. 22.9.3639
2. Godwin Kaisara, Clayton Peel, Cornelius Niemand, Kelvin Joseph Bwalya (2024) Explor- ing Factors Influencing e-Learning Dropout Rates in the Post-COVID-19 Era, Interna- tional Journal of Information and Communication Technology Education, Vol. 20, No. 1 https://doi.org/10.4018/IJICTE.348660
3. Sandro Monteiro, José Alberto Lencastre, Bento Duarte da Silva, António J. Osório, Paula de Waal, Şukru Çetin İlin, Gülden İlin (2017) A Systematic Review of Design Factors to Prevent Attrition and Dropout in e-Learning Courses, Better e-Learning for Innovation in Education,
4. pp. 135–153 Available at: https://hdl.handle.net/1822/47770
5. Jun, J. (2005) Understanding E-dropout, International Journal on E-Learning, Vol. 4, No. 2,
6. pp. 229–240 Available at: https://www.learntechlib.org/p/4620/
7. P.C. Enwereji, A. van Rooyen (2025) Exploring the Barriers to Online Student Learning Support Services: A Review, South African Journal of Higher Education, Vol. 39, No. 1, pp. 69–93 https://doi.org/10.20853/39-1-5116
8. Elias Dritsas, Maria Trigka (2025) Methodological and Technological Advancements in E- Learning, Information, Vol. 16, No. 1, p. 56 https://doi.org/10.3390/info16010056
9. Eddy Triswanto Setyoadi, Syaad Patmanthara, Heru Wahyu Herwanto, Hartarto Junaedi, Titasari Rahmawati (2025) A Review Of Learning Management System Enhanced by Gamifi- cation Through Push Pull Mooring Model, IRDH International Journal of Technology, Agri- culture and Natural Sciences, Vol. 2, No. 1, pp. 61–79 Available at: https://irdhjournals. com/ijtans/article/view/94
10. Jing Chen, Bei Fang, Hao Zhang, Xia Xue (2024) A systematic review for MOOC dropout prediction from the perspective of machine learning, Interactive Learning Environments, Vol. 32, No. 5, pp. 1642–1655 https://doi.org/10.1080/10494820.2022.2124425
11. Matti Vaarma, Hongxiu Li (2024) Predicting student dropouts with machine learning: An empirical study in Finnish higher education, Technology in Society, Vol. 76, Article 102474 https://doi.org/10.1016/j.techsoc.2024.102474
12. Dan-Andrei Sitar-Tăut, Daniel Mican, Ovidiu-Ioan Moisescu (2024) To be (online) or not to be? The antecedents of online study propensity and e-learning-dependent dropout intention in higher education, Technological Forecasting and Social Change, Vol. 207, Article 123566 https://doi.org/10.1016/j.techfore.2024.123566
13. Lin Qiu, Yanshen Liu, Yi Liu (2018) An Integrated Framework With Feature Selection for Dropout Prediction in Massive Open Online Courses, IEEE Access, Vol. 6, pp. 71474–71484 https://doi.org/10.1109/ACCESS.2018.2881275
14. Janka Kabathova, Martin Drlik (2021) Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques, Applied Sciences, Vol. 11, No. 7, Article 3130 https://doi.org/10.3390/app11073130
15. Miki Katsuragi, Kenji Tanaka (2022) Dropout Prediction by Interpretable Machine Learning Model Towards Preventing Student Dropout, Transdisciplinarity and the Future of Engineer- ing https://doi.org/10.3233/ATDE220700
16. ALJ Zakaria, Anas Bouayad, Mohammed Ouçamah Cherkaoui Malki (2024) A Systematic Re- view of Gamification in MOOCs: Effects on Student Motivation, Engagement, and Dropout Rates, Journal of Educators Online, Vol. 21,

No. 2, pp. 1–30 https://doi.org/10.9743/ JEO.2024.21.2.15
17. Ahmed Obeid, Roliana Ibrahim, Ahmad Fadhil (2024) Extended Model of Expectation Confir- mation Model to Examine Users' Continuous Intention Toward the Utilization of E-Learning Platforms, IEEE Access, Vol. 12, pp. 40752–40764 https://doi.org/10.1109/ACCESS. 2024.3373190
18. Wala Bagunaid, Naveen Chilamkurti, Ahmad Salehi Shahraki, Saeed Bamashmos (2024) Vi- sual Data and Pattern Analysis for Smart Education: A Robust DRL-Based Early Warning System for Student Performance Prediction, Future Internet, Vol. 16, No. 6, Article 206 https://doi.org/10.3390/fi16060206
19. Shahzad Rizwan, Chee Ken Nee, Salem Garfan (2025) Identifying the Factors Affecting Stu- dent Academic Performance and Engagement Prediction in MOOC Using Deep Learn- ing: A Systematic Literature Review, IEEE Access, Vol. 13, pp. 18952–18971 https:
20. //doi.org/10.1109/ACCESS.2025.3533915
21. Thao-Trang Huynh-Cam, Long-Sheng Chen, Van-Canh Nguyen, Thanh-Huy Nguyen, Tzu- Chuen Lu (2024) Why first-year e-students are dissatisfied: Machine learning methods for enhancing retention, International Journal of Applied Science and Engineering, Vol. 21, No. 3, Article 2023532 https://doi.org/10.6703/IJASE.202406_21(3).002
22. Ioanna Lykourentzou, Ioannis Giannoukos, Vassilis Nikolopoulos, George Mpardis, Vassili Loumos (2009) Dropout prediction in e-learning courses through the combination of ma- chine learning techniques, Computers & Education, Vol. 53, No. 4, pp. 950–965 https:
23. //doi.org/10.1016/j.compedu.2009.05.010
24. Saad Alghamdi, Ben Soh, Alice Li (2025) A Comprehensive Review of Dropout Prediction Meth- ods Based on Multivariate Analysed Features of MOOC Platforms, Multimodal Technologies and Interaction, Vol. 9, No. 1, Article 3 https://doi.org/10.3390/mti9010003
25. Maha Salem, Khaled Shaalan (2025) Unlocking the power of machine learning in E- learning: A comprehensive review of predictive models for student performance and en- gagement, Education and Information Technologies, Vol. 30 https://doi.org/10.1007/ s10639-025-13526-4
26. Rund Mahafdah, Seifeddine Bouallegue, Ridha Bouallegue (2024) Enhancing e-learning through AI: advanced techniques for optimizing student performance, PeerJ Computer Science, Vol. 10, Article e2576 https://doi.org/10.7717/peerj-cs.2576
27. Rochdi Boudjehem, Yacine Lafifi (2024) An early warning system to predict dropouts inside e-learning environments, Education and Information Technologies, Vol. 29, pp. 16365–16385 https://doi.org/10.1007/s10639-024-12498-1
28. Swadha Gupta, Parteek Kumar, Rajkumar Tekchandani (2024) Artificial intelligence based cognitive state prediction in an e-learning environment using multimodal data, Mul- timedia Tools and Applications, Vol. 83, pp. 64467–64498 https://doi.org/10.1007/ s11042-023-18021-x
29. Meriem Zerkouk, Miloud Mihoubi, Belkacem Chikhaoui, Shengrui Wang (2025) Predicting On- line Education Dropout: A New Machine Learning Model Based on Sentiment Analysis, Socio-demographic, and Behavioral Data, International Journal of Artificial Intelligence in Education https://doi.org/10.1007/s40593-025-00472-y
30. Amir Mohammad Rahmani, Wim Groot, Hamed Rahmani (2024) Dropout in online higher education: a systematic literature review, International Journal of Educational Technology in Higher Education, Vol. 21, Article 19 https://doi.org/10.1186/s41239-024-00450-9
31. Mingjie Tan, Peiji Shao (2015) Prediction of Student Dropout in E-Learning Program Through the Use of Machine Learning Method, International Journal of Emerging Technologies in Learning (iJET), Vol. 10, No. 1, pp. 10–17 https://doi.org/10.3991/ijet.v10i1.4189
32. Yizhuo Zhou, Jin Zhao, Jianjun Zhang (2020) Prediction of learners' dropout in E-learning based on the unusual behaviors, Interactive Learning Environments https://doi.org/10. 1080/10494820.2020.1857788
33. Kowsar Talebi, Zeinab Torabi, Negin Daneshpour (2024) Ensemble models based on CNN and LSTM for dropout prediction in MOOC, Expert Systems with Applications, Vol. 235, Article 121187 https://doi.org/10.1016/j.eswa.2023.121187
34. Subhabrata Sengupta, Rupayan Das, Satyajit Chakrabarti (2025) Hybrid Data Mining Techniques for Predicting Student Academic Performance in E-Learning to Avoid Drop- Out (HDL-SP), SN Computer Science, Vol. 6, Article 162 https://doi.org/10.1007/ s42979-025-03733-8
35. Xinhong Zhang, Xiangyu Wang, Jiayin Zhao, Boyan Zhang, Fan Zhang (2024) IC-BTCN: A Deep Learning Model for Dropout Prediction of MOOCs Students, IEEE Transactions on Education, Vol. 67, No. 6, pp. 974–982 https://doi.org/10.1109/TE.2024.3398771
36. Ke Niu, Jiuyun Cai, Yuhang Zhou, Wenjuan Tai, Xue Feng (2025) Hybrid Neural Network Model for MOOC Dropout Prediction, Complex System Modeling and Simulation, Vol. 1, pp. 1–10 https://doi.org/10.23919/CSMS.2025.0003
37. Kaggle (2021) 'Online Course User Engagement Data', Kaggle Datasets. Available at: https:
38. //www.kaggle.com/datasets/thedevastator/online-course-user-engagement-data. Dataset derived from KarmaLJF on data.world: https://data.world/karmaljf.