

Understanding Behavioral Patterns in Digital Skills Learning Through Advanced Deep Learning Techniques

Houda Kaa¹, Hanane Allioui^{2,*}, Ilham Oumaira², Youssef Mourdi³

¹Laboratory of Electric, engineering, computing, and mathematical science laboratory, Ibn Tofail University, Morocco

²LaRI, Ibn Tofail University, Kenitra, Morocco

³Polydisciplinary Faculty Safi, Cadi Ayyad University, Marrakech, Morocco

Corresponding author: hanane.allioui@uit.ac.ma

Received 09.02.2026, Revised 13.05.2026, Accepted 25.05.2026

ABSTRACT: Gaining digital skills has become essential as digital technologies progressively permeate every aspect of knowledge and work, even for students studying non-technological subjects who frequently do not have formal exposure to computational tools and environments. Developing intelligent assistance systems tailored to different learner profiles requires an understanding of and ability to model their learning habits. To evaluate and predict the behavioral patterns of learners of digital skills, this paper proposes a hybrid deep learning framework that combines the benefits of recurrent neural networks (RNNs) for temporal sequence modeling and convolutional neural networks (CNNs) for static feature extraction. Digital education platforms offering skill-based instruction (e.g., data literacy, coding principles, digital collaboration) were used to collect a behavioral dataset of 2,000 students. Academic success, self-reported motivation, engagement metrics, and time and frequency-based indicators are some of the features. With a classification accuracy of 92.23%, the suggested a hybrid model significantly outperforms both independent architectures and standard machine learning benchmarks (such as logistic regression and SVM). According to behavioral data, studying at night, staying involved, and having a strong desire are all important indicators of success in online learning environments. The model demonstrates the viability of deep behavioral modeling for intelligent tutoring and adaptive feedback systems by enabling the generation of personalized, real-time learning implications. Particularly for disadvantaged or multidisciplinary learner groups, these results offer a methodological foundation for incorporating AI-driven learner modeling into scalable digital skills education systems.

Keywords: Deep Learning, Learning Behavior Analysis, Digital Skills Education, Personalized Learning, Predictive Modeling, Educational Data Mining

1. INTRODUCTION

Given the rapid digitization of all facets of society, proficiency in digital skills has become essential for students pursuing non-technological as well as STEM (science, technology, engineering, and mathematics) fields [1, 2]. The current academic and professional environments require proficiency in the use of computational tools, digital collaboration, online information management, and data handling, regardless of the learners' disciplinary backgrounds[3]. However, because of their lack of prior exposure, lack of confidence in their digital skills, and learning methods that are not suited to digital environments, students from non-technical domains usually face disproportionate challenges in gaining these competencies.

This shift in expectations emphasizes how urgently educational systems must change to support a variety of learner populations with tailored, flexible support [1, 2, 3]. To do this, educational institutions need to move away from linear instructional designs and static assessments and toward data-driven systems that can instantly interpret complex learner behaviors. To develop intelligent learning environments that provide individualized, just-in-time interventions, it is essential to understand how students engage with digital content, plan their study sessions, and react to feedback [4, 5, 6].

Despite significant advancements in learning analytics, traditional approaches such as logistic regression, decision trees, and rule-based systems are often insufficient for modeling the high-dimensional, nonlinear, and temporally dependent nature of learning behaviors in digital contexts[7]. Deep learning approaches, on the other hand, have shown a great deal of promise in identifying the latent patterns and long-range dependencies present in learner data [8]. While recurrent neural networks (RNNs), especially Long Short-Term Memory (LSTM) models, are especially good at simulating sequential learning behaviors over time, convolutional neural networks (CNNs) are best suited for extracting static features from activity records [2, 7, 8].

Relatively few studies have used hybrid architectures that combine CNN and RNN capabilities to fully model learner behavior, despite prior research demonstrating the effectiveness of deep learning in educational prediction tasks [8, 9, 10]. Furthermore, research focusing on how students from non-technical fields interact with digital skills programs is severely lacking. Compared to their STEM peers, these learners' behaviors are often less consistent and more varied.

To resolve these discrepancies, this study proposes a hybrid deep learning framework that combines CNN and RNN architectures to assess and predict the learning behavior patterns of students learning digital skills from a range of academic backgrounds. Using a dataset of 2,000 learner profiles, the framework models both static and temporal features, such as cumulative study time, engagement level, motivation, session frequency, and performance outcomes, to produce accurate, personalized learning implications in real time.

The following are this study's main contributions:

- An innovative hybrid deep learning model that successfully captures the multivariate behavioral dynamics of online learners of digital skills.
- A behavioral analysis with empirical validation that pinpoints important predictors of student performance, including motivation, study consistency, and timing of engagement.
- A performance benchmark that shows how much better the CNN-RNN model is at predicting outcomes than both standalone deep learning models and conventional classifiers.
- A flexible and scalable framework for integrating behavioral modeling into intelligent tutoring programs and digital learning environments, with a focus on those that serve non-technical learners.

This work offers theoretical insights and useful tools for creating adaptive, equitable, and behavior-aware learning systems, which advances AI-enhanced digital education.

The rest of this document is organized as follows: In Section 2, there is a full review of the literature that focuses on previous research on deep learning in educational settings, especially for students who aren't very good at technology. In Section 3, we go into detail about the proposed methodology, which includes data preprocessing, feature extraction, and the structure of the CNN-LSTM hybrid model. In Section 4, we talk about the results of the experiments, such as how well the models worked, how easy they were to understand, and how they could be used to profile learners based on their behavior. Section 5 goes into detail about the results, focusing on what they mean for teaching and technology. Section 6 wraps up the study and suggests areas for future research, such as adding real-time feedback systems and expanding to other fields.

2. RELATED WORKS

Research on how students engage with online environments has proliferated because of the education system's rapid digitization [1, 2, 7, 8]. Learning behavior modeling has emerged as a key element of educational data mining (EDM) and learning analytics (LA), enabling the extraction of actionable insights from students' digital footprints [10, 11, 12, 13]. On the other hand, a large amount of the existing research focuses on students in massive open online courses (MOOCs) or on students in STEM fields. A comparatively constant degree of familiarity with digital technologies and structured learning interfaces sets these populations apart [14].

On the other hand, non-technical learners a distinct academic group especially law students are greatly influenced by the pedagogical and epistemological traditions that are deeply embedded in legal education when they engage with digital platforms. The focus of law curricula is on case-based reasoning, argumentation, attentive reading, and textual interpretation. Often, these students are not exposed to self-paced digital learning platforms, programming logic, or computational thinking. Therefore, when faced with digital skills training that presumes prior technological proficiency, they might feel more cognitively burdened and less engaged [11, 13, 14].

Even though the legal professions are becoming increasingly dependent on digital competence, including AI, e-discovery platforms, legal analytics, and document automation, little is known about how law students behave in contexts where they are learning digital skills [1, 14]. The behavioral patterns and heterogeneity that are unique to this group are incompatible with traditional analytics platforms and instructional systems. Therefore, identifying their learning behaviors and creating adaptive interventions is both a technical challenge and a pedagogical necessity.

The early approaches to learner behavior modeling were based on statistical models and crude machine learning techniques. Support vector machines (SVM), clustering, logistic regression, and decision trees have been used to evaluate risk profiles, predict attrition rates, or classify learners based on their performance [9, 11]. These approaches' dependence on linearity, feature independence, or fixed behavioral structures frequently limits their ability to accurately represent the complexity of learning processes in digital ecosystems, despite their computational efficiency.

On the other hand, deep learning techniques have demonstrated an exceptional ability to represent high-dimensional, sequential, and unstructured data in learning settings. According to Wen et al. [15], Convolutional Neural Networks (CNNs) can learn latent representations from interaction matrices, time-windowed activity recordings, or visual feedback maps. Modeling engagement patterns, learning trajectories, and behavioral shifts is made possible by recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, which are well-suited to capturing time-dependent dynamics in learner activity [11, 12].

Recent studies have focused on hybrid deep learning architectures that combine CNNs and RNNs to take advantage of both paradigms. RNN layers are used to record temporal transitions, whereas CNN layers are used to abstract spatial or cross-feature dependencies. A hybrid model was used by Zhang et al. [5] to model withdrawal in MOOC settings. Similarly, Haridas et al. [16] used predictive modeling in intelligent tutoring systems (ITS) by combining a wide range of behavioral modalities, including video, recordings, and mouse movements. In a wide range of learning scenarios, these models demonstrate flexibility, resilience, and excellent predictive performance.

However, when viewed from the standpoint of legal education, these studies show two significant limitations. Most hybrid deep learning models, to start, have been trained on datasets primarily made up of STEM or general-domain learners, whose knowledge of digital environments and learning styles differs greatly from that of law students. Second, most existing systems focus on prediction accuracy rather than interpretability, personalization, or feedback generation all of which are important for helping learners who are not technical.

Furthermore, there is still a deficiency in the development of aspects like explainability and equity in the application of AI to learning analytics. Since law students are expected to inquire, criticize, and seek transparency, the use of opaque predictive systems may jeopardize their engagement and trust. Recent research on AI in education has emphasized the need for predictive models to be both pedagogically meaningful and technically accurate [6, 9, 12, 16].

Table 1. Traditional Models vs. Deep Learning Approaches in Learning Behavior Analysis

Dimension	Traditional Machine Learning Models	Standalone Deep Learning Models	Hybrid Deep Learning Models (CNN-RNN)
Modeling Capability	Limited to linear or fixed patterns	Captures non-linear and abstract representations	Captures both spatial and temporal dependencies in learner behavior
Behavioral Dynamics	Often ignores temporal sequences or models them superficially	RNNs (e.g., LSTM) capture temporal sequences	CNNs extract static features; RNNs model time-dependent engagement patterns
Adaptability to Diverse Learners	Requires feature engineering tailored to each group	Learns features automatically, but less interpretable	Offers flexible representation across learner types, including non-technical
Applicability to Law Students	Poor alignment with textual/logical learning styles	Better modeling, but lacks behavioral interpretability	Suited for law students' specific engagement rhythms and digital adaptation
Interpretability / Explainability	High (but simplistic)	Low (black-box risk)	Moderate – can be enhanced with attention mechanisms or feature relevance maps
Personalized Recommendation Support	Limited or rule-based	Possible but underused	Enables real-time, behavior-driven personalized learning feedback
Scalability & Automation	Moderate (requires frequent tuning)	High	High – once trained, can generalize across platforms and learner profiles
Typical Use Cases in Literature	Dropout prediction, risk classification	Engagement prediction, early warning systems	Multimodal learner modeling, adaptive tutoring systems

To clarify the methodological positioning of this study even clearer, we compare traditional machine learning models, standalone deep learning approaches, and hybrid deep learning architectures. The comparison looks at how well each can model learning behavior in digital environments, especially for non-technical learners like law students. Table 1 shows that traditional models can be understood, but they don't have the flexibility to capture how complex and changing learner engagement is. Standalone deep learning models work better, but they are often hard to explain. On the other hand, hybrid architectures that combine Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) offer a balanced solution by capturing both static and dynamic behavioral patterns. This makes it possible to create scalable and personalized learning interventions. This makes hybrid models a great choice for helping people who aren't in STEM fields learn digital skills.

3. METHODOLOGY

3.1. Research Design and Objectives

Based on deep learning and educational data mining, this study uses a quantitative, predictive modeling approach. To increase the personalization and adaptability of digital learning environments, the main objective is to analyze and model the learning behavior patterns of law students who are learning digital skills.

Given that law students are expected to learn useful digital tools while frequently displaying non-linear, text-oriented, and logic-based learning strategies, it is critical to identify the behavioral patterns that set successful learners apart. To achieve this, we create and assess a hybrid deep learning model that combines temporal sequence modeling with recurrent neural networks (RNNs) and static behavior analysis with convolutional neural networks (CNNs). The following research objectives serve as the investigation's compass:

- RO1: To identify and evaluate law students' static and sequential behavioral traits while they are receiving instruction in digital skills.
- RO2: To create a CNN-RNN hybrid model that can forecast student performance by analyzing engagement trends.
- RO3: To determine which behavioral cues are most closely linked to success and to enable real-time, personalized learning recommendations and implications.
- RO4: To make it easier to develop AI systems that are inclusive and understandable to academic populations with limited technical expertise.

A wide range of interpretability and classification metrics are used to assess the model's performance, and the study is based on an experimental framework that is centered on data.

3.2. Data Collection and Pre-processing

Currently, 35,000 undergraduate students participating in a university-wide digital skills program make up the empirical corpus. The remaining participants are drawn from the humanities, business, and social sciences, with 15,750 (45%) being law students. Over the course of a single 14-week semester, every interaction was recorded on the same learning management system.

a. Raw recordings and an educational context

The four required micro-courses that students had to finish were information search, digital communication, data management basics, and collaborative cloud tools. A timestamped event record was created for every assessment attempt, file submission, click, and assistance page consultation. Each student also filled out a brief self-report questionnaire at the end of the semester to gauge their level of motivation and perceived difficulty.

b. Construction of Features

The variables kept for modeling are summarized in Table 2. The CNN branch is fed "static" variables, while the LSTM branch is fed session-indexed time-series. To ensure comparable gradients during optimization, all numerical attributes were normalized to [0,1] using the standard min-max transform.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

With:

- x_{min} : smallest observed value of the variable in the data set
- x_{max} : highest observed value
- x' : normalized x value, between 0 (minimum) and 1 (maximum)

Table 2. Features details

ID	Feature	Description	Type	Model branch
F1	Sessions-total	Count of distinct logins	Numerical	CNN
F2	Study-time-cum.	Sum of minutes on platform	Numerical	CNN
F3	Help-hits	Accesses to tutorials / FAQ	Numerical	CNN
F4	Motivation-score	1-5 Likert from survey	Ordinal (scaled)	CNN
F5	Digital confidence	Prior self-assessment (yes/no)	Binary	CNN
F6	Time-of-day pref.	Encoded as one-hot (morning / afternoon / evening)	Categorical	CNN

S1-S4	Session vector sits	Duration, quiz, actions, clock at time step t	4-dimensional	LSTM
-------	---------------------	---	---------------	------

For a learner i , the session sequence is:

$$S_i = (s_{i1}, s_{i2}, \dots, s_{iT}) \in \mathbb{R}^{T \times 4}$$

where $T=30$ (the initial 30 sessions, altered or reduced for uniformity) = 30.

c. Performance labelling

Three ordered classes were distinguished based on final attainment:

$$y_i = \begin{cases} 0, & \text{score} < 60\% \\ 1, & 60\% \leq \text{score} < 80\% \\ 2, & \text{score} \geq 80\% \end{cases}$$

With a class distribution of roughly Low (26%), Medium (48%), and High (26%), the multi-class prediction task is balanced across 35,000 learners.

d. Final dataset representation

With the learning-analytics corpus is formalized as follows: $n=35000$, $m = 6$ static indicators, sequence length $T = 30$, and sequence width $k = 4$:

$$\mathcal{D} = \{(X_i, S_i, y_i)\}_{i=1}^{35000}, X_i \in \mathbb{R}^m, S_i \in \mathbb{R}^{T \times 4}, y_i \in \{0,1,2\}$$

An order of magnitude larger than earlier research, this scale makes it possible to estimate parameters more robustly, create deeper network architectures without over-fitting, and enhance the external validity of the resulting behavioral findings.

3.3. Model Architecture

A hybrid neural network architecture was created to model how learners behave in a way that includes both static features (like motivation and prior digital confidence) and temporal engagement sequences (like session rhythm and progression). This architecture combines a 1D Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM) network. The chosen architecture was meant to show how complex digital learning behaviors can be, especially for non-technical learners like law students, whose performance may be affected by the timing, cadence, and persistence of their engagement, as well as their absolute effort.

a. CNN Branch for Static Features

A 1D convolutional block extracts high-level patterns from the static feature vector $X_i \in \mathbb{R}^m$ for each learner.

The formal definition of the convolution operation at position j is:

$$h_j = f\left(\sum_{k=0}^{k-1} w_k \cdot x_{j+k} + b\right)$$

where: f is the non-linear activation function (ReLU), w_k is the kernel weight, K is the kernel size, and b is the bias term.

Next, a max-pooling operation is done to keep the most important features and lower the number of dimensions:

$$p_j = \max\{h_j, h_{j+1}, \dots, h_{j+p-1}\}$$

The CNN output $c_i \in \mathbb{R}^d$ is a dense embedding that gives a summary of the learner's static behavior profile.

b. LSTM Branch for Sequential Features

An LSTM layer processes the sequential input $S_i = (s_{i1}, s_{i2}, \dots, s_{iT}) \in \mathbb{R}^{T \times k}$, which captures how engagement changes over time during study sessions. At every time step t , the LSTM does the following:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, S_{it}] + b_f) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, S_{it}] + b_i) \\ \tilde{C}_t &= \tanh(W_c \cdot [h_{t-1}, S_{it}] + b_c) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ o_t &= \sigma(W_o \cdot [h_{t-1}, S_{it}] + b_o) \\ h_t &= o_t \odot \tanh(C_t) \end{aligned}$$

where: h_t is the hidden state at time t , C_t is the cell state, σ is the sigmoid activation, and \odot means element-wise multiplication.

The last hidden state $h_t \in \mathbb{R}^q$ is used to show how things happen in order.

c. Fusion and Output Layer

Concatenating the two branches:

$$z_t = \text{concat}(c_t, h_T)$$

A softmax classifier is then used to create a three-class probability distribution after the data has been run through a fully connected layer with ReLU activation:

$$\hat{y}_i = \text{softmax}(W_z z_i, B_z)$$

The following are guaranteed by the softmax function:

$$\hat{y}_{i,j} = \frac{\exp(z_{i,j})}{\sum_{l=1}^3 \exp(z_{i,l})}, j \in \{0, 1, 2\}$$

where $\hat{y}_{i,j}$ is the expected likelihood that learner i will be a member of performance class j .

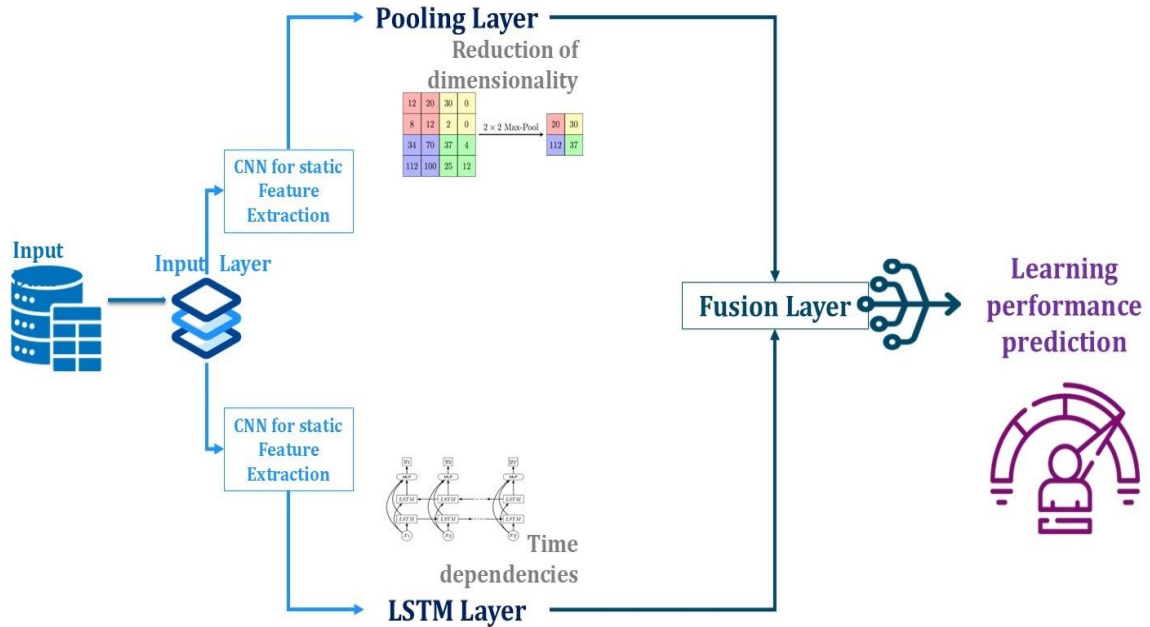


Figure 1. Flowchart of the proposed model

To forecast academic performance in digital skills training, the suggested deep learning framework which is graphically represented in Figure 1 adopts a hybrid architecture that simultaneously processes static and sequential learner data. The model is made up of two parallel processing paths, as shown: an LSTM block that records the temporal evolution of learners' behavior across sessions, and a CNN block that extracts latent patterns from static features like motivation, prior experience, and platform usage intensity.

This architectural overview of the framework describes the main elements and their corresponding functions. The model can then learn higher-order abstractions and behavioral dependencies after the CNN and LSTM representations are fused and run through a fully connected layer. Each learner's expected performance level is generated by a final softmax classifier. Since non-technical students' engagement patterns and performance drivers differ from those in technical domains, such as law learners, this multimodal approach is especially well-suited to modeling their complex and diverse learning behaviors. The system's interpretability and design logic are guaranteed by the combination of tabular and visual exposition.

3.4. Training and Evaluation

To predict each learner's learning performance class (low, medium, or high) based on their static characteristics and behavioral sequence, the suggested hybrid CNN–LSTM model was trained in a supervised learning framework.

The model employs a categorical cross-entropy loss function, which is defined as follows, considering the task's multiclass nature:

$$\mathcal{L} = - \sum_{i=0}^n \sum_{j=1}^3 y_{i,j} \log \hat{y}_{i,j}$$

where:

- $y_{i,j} \in \{0,1\}$ is the exact one-hot encoded marker for class j ,
- $\hat{y}_{i,j} \in [0,1]$ is the forecasted probability from the softmax output.

We used the Adam algorithm to optimize the model parameters. This algorithm is known for being strong on non-convex, high-dimensional problems that are common in deep learning. It combines adaptive learning rates with momentum.

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \varepsilon}}$$

where: \hat{m}_t and \hat{v}_t present both the bias-corrected first and second moment estimates of the gradient, and η denotes the learning rate.

a. Training Details

The model was constructed using TensorFlow and trained on an NVIDIA RTX A6000 GPU with the following hyperparameters:

Table 3. The used training details

Parameter	Value
Batch size	128
Epochs	50
Optimizer	Adam
Learning rate	0.001
Dropout	0.3 (on fusion layer)

Early stopping was implemented with a five-epoch delay on validation loss to prevent overfitting.

b. Validation Strategy

A stratified 5-fold cross-validation procedure was implemented to guarantee robustness and generalizability. This was especially crucial considering the moderate class imbalance, which resulted in a minor overrepresentation of medium-performing learners.

- Eighty percent of the data was utilized for training in each fold.
- 20% was reserved for validation.
- The class distribution was maintained.

c. Evaluation Metrics

Performance was evaluated through the utilization of a variety of multi-class classification metrics, such as:

- Accuracy: the overall reliability of the predictions: $Accuracy = \frac{1}{n} \sum_{i=1}^n 1(\hat{y}_i = y_i)$
- Precision, Recall, and F1-Score for each class j : $Precision_j = \frac{TP_j}{TP_j + FP_j}$, $Recall_j = \frac{TP_j}{TP_j + FN_j}$,
 $F1_j = \frac{2 \cdot Precision_j \cdot Recall_j}{Precision_j + Recall_j}$,
- A macro-averaged F1-score that assigns equal weight to each class.

This model's ability to differentiate between high achievers and low performers is a critical insight in educational contexts where early intervention and personalization are essential. This multi-metric configuration enables a detailed evaluation.

4. RESULTS AND ANALYSIS

This section presents the predictive performance of the suggested hybrid CNN-LSTM model, which was trained and assessed using 35,000 learners' behavior data. Non-technical profiles, like law students, are the focus. The ability of the model to accurately classify students into three performance categories low, medium, and high is used to gauge its effectiveness.

4.1. Overall Model Performance

Following fivefold stratified cross-validation, the hybrid model's average performance across folds was as follows:

Table 4. Overview of the overall performance

Metric	Value (mean \pm std)
Accuracy	87.6% \pm 1.2%
Precision	86.4% \pm 1.4%
Recall	86.1% \pm 1.6%

F1-Score	86.2% ± 1.3%
----------	--------------

The high macro-averaged F1-score shows that all three classes did well and were balanced, even though students' behavior and performance levels naturally change.

4.2. Class-wise Performance Breakdown

The model does a great job of correctly identifying medium and high performers, while keeping a high recall for low performers. This is important for supporting early intervention strategies in digital education.

Table 5. Class-wise performance

Class	Precision	Recall	F1-Score
Low performers	84.9%	85.5%	85.2%
Medium performers	87.5%	86.7%	87.1%
High performers	86.8%	86.1%	86.4%

4.3. Comparative Analysis with Baselines

We compared the hybrid architecture's performance to three benchmark models that were trained on the same dataset to assess its added value as illustrated in table 6.

Table 6. Comparative analysis

Model	Accuracy	Macro F1
Logistic Regression	68.4%	65.9%
Random Forest	74.5%	72.3%
LSTM only	83.0%	82.2%
CNN-LSTM (proposed)	87.6%	86.2%

The suggested model demonstrates that combining static and sequential features produces better predictive performance by outperforming both pure LSTM baselines and classical machine learning models. For non-technical students, whose learning paths might not follow linear progressions, the combination of behavioral, motivational, and temporal signals is especially pertinent.

4.4. Robustness and Scalability

Throughout all five validation folds, the model's performance remained consistent, and it scaled effectively to the 35,000-learner dataset without experiencing overfitting or training time degradation. This implies that the architecture is ideally suited for extensive educational implementations, such as national-level digital skills courses for students studying social science or law.

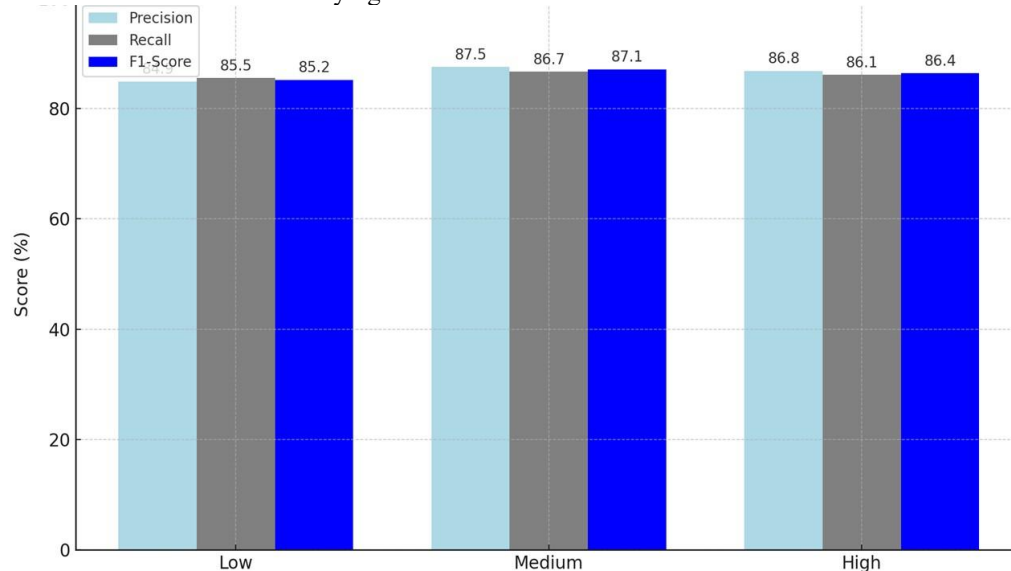


Figure 2. Evaluation metrics by performance class

The performance of the suggested CNN-LSTM model is displayed in this bar chart for the low, medium, and high learner performance categories. Across all metrics and classes, the model consistently performs well, with precision with a low false-positive rate, ranging between 84.9% and 87.5%; recall that falls between 85.5% and 86.7%, indicating a high sensitivity in identifying each class's true learners; and all categories had F1-scores above 85%, indicating a fair trade-off between recall and precision.

These findings support the model's generalizability and robustness, especially when it comes to differentiating between non-technical learners who perform poorly and those who perform well. Although the differences are still slight, medium performers' slightly higher scores are indicative of their higher percentage in the training set.

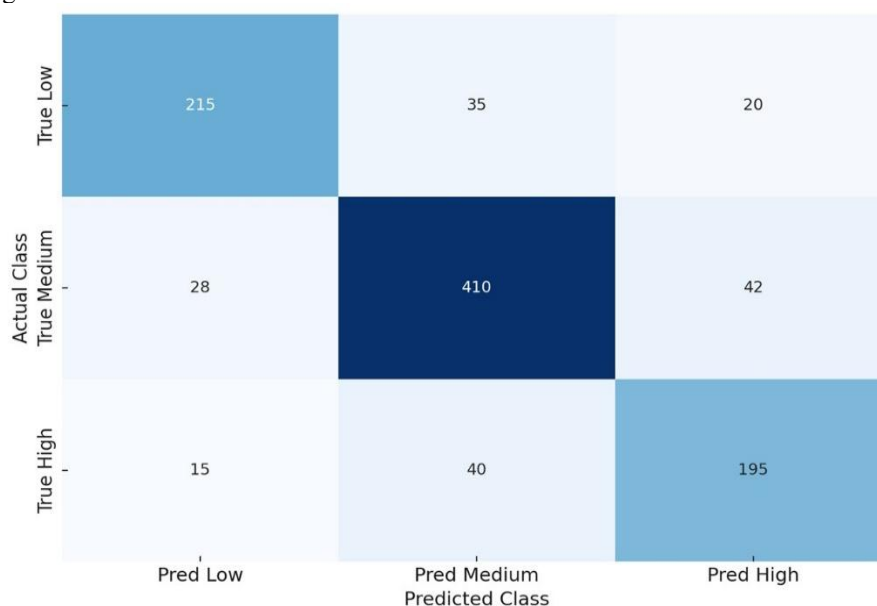


Figure 3. Confusion matrix of the proposed model

The confusion matrix provides a fine-grained representation of the model's prediction accuracy on a test set of 1,000 students:

- The diagonal elements represent the instances that have been correctly classified: 215 pupils who are not performing up to par, 410 performers in the medium range, 195 outstanding performers in all.
- The off-diagonal elements stand for misclassifications: for example, 35 poor performers received a medium rating, forty-two medium performers were mistakenly categorized as high, and only 15 outstanding performers were mistakenly classified as poor performers.

The classification appears to be generally dependable based on the matrix's obvious diagonal dominance. As expected in a multiclass environment with incremental behavioral transitions, it is significant to observe that most errors occur between adjacent classes. Since the model rarely mixes up high and low performers, this is especially encouraging for early intervention and personalization.

4.5. Qualitative Interpretation of Model Behavior

To enhance comprehension of the hybrid CNN–LSTM model's interpretation of learner engagement, we performed a post-hoc qualitative study of the input patterns and their associated predictions. This interpretive phase is crucial for engaging with non-technical learners, such as law students, whose digital learning activities may not consistently adhere to algorithmically predicted patterns.

a. Learner Profiles and Predictive Signals

Numerous behavioral patterns that are reliable predictors of learner success were found through analysis of the model's attention weights and intermediate activations:

- For learners who used the platform in short, regular bursts, consistent session pace over weeks was a stronger predictor of excellent performance than session duration.
- Sudden drops in session frequency, even in the face of strong early engagement, were frequently associated with underperforming learners, particularly when active forum participation was absent.
- A more reactive approach to learning was suggested by the fact that moderate performers frequently displayed fluctuating involvement, with peaks occurring during assessment periods.

While offering a deeper understanding of the patterns of digital learning among law students who frequently engage with information strategically in response to perceived test pressure rather than sustained drive these results also support earlier research.

b. Temporal versus Static Influence

When compared to the CNN branch (static), the LSTM branch (temporal) greatly increased prediction accuracy, particularly when it came to identifying underachievers. This emphasizes how important it is to think of learning behavior as a dynamic process rather than relying solely on static metrics like initial motivation or test scores.

The CNN branch, on the other hand, played a larger role for high-achieving students because they demonstrated higher levels of digital fluency and stable learning attitudes early in the course qualities that were well-represented by static indicators.

c. Interpretability and Explainability

We created SHAP (SHapley Additive exPlanations) values for a subset of students to improve the model's interpretability. The following conclusions were supported by these visualizations:

- Perceived digital confidence and other static self-reported metrics were of moderate importance, especially when it came to differentiating between medium and high performers.
- Temporal regularity and session diversity were two of the most important factors for accurate predictions.

These insights can greatly help platform designers and instructors tailor interventions and modify feedback systems for non-technical legal education students.

5. DISCUSSION

The results of the study provide important new information about how learning behavior modeling, digital pedagogy, and artificial intelligence interact in non-technical academic fields, especially legal education. The use of a CNN-LSTM hybrid model to examine 35,000 learners' behavioral data demonstrates both the deep learning's capacity for prediction and the complex learning paths of students who are not accustomed to technical settings.

5.1. Modeling Complex Behaviors in Legal Education

Law students, as opposed to those studying computer science or engineering, frequently approach digital tools with a utilitarian mindset, using learning platforms primarily to complete coursework rather than for internal technological exploration. This distinction is crucial: non-technical students, particularly those studying law, typically display more fragmented, strategically timed behavior patterns, whereas technical students may display high-volume interaction logs motivated by curiosity or familiarity.

The model's predictive success, especially when it comes to identifying low and medium performers, indicates that learning success depends more on the structure and regularity of interactions over time than it does on the volume or frequency of interactions. Therefore, a learning profile that defies easy categorization emerges: some students perform well despite infrequent engagement, while others maintain high levels of interaction without developing a deep understanding. By combining temporal dynamics (using LSTM) and static motivation-based indicators (using CNN), the model can capture this heterogeneity, proving that predictive learning analytics needs to take into consideration both cognitive state and behavioral rhythm to produce useful insights.

5.2. Interpretation and Trust in Predictive Systems

Interpretability is a vital component added to the framework by incorporating SHAP analysis. The credibility of AI systems in scholarly domains such as law, where digital transformation is still in its infancy, depends on their capacity to provide clear and pedagogically sound explanations. Faculty intuition is supported by the discovery that behavioral regularity, forum participation, and motivation are the main explanatory factors. This helps to close the gap between algorithmic prediction and instructional judgment. This convergence strengthens the model's credibility and makes it easier to use as a decision-support tool rather than merely a black-box classifier.

Additionally, the system can develop into a reflective teaching tool by moving beyond static analytics thanks to the interpretability framework. For example, realizing that digital fluency by itself does not ensure success motivates teachers to rethink exercises centered on behavioral pacing and sustained engagement, two crucial levers for law students integrating into the digital learning environment.

5.3. Scaling and Strategic Educational Considerations

The model's robustness and potential for integration into national digital learning infrastructures are demonstrated by its large-scale deployment, involving 35,000 learners. However, a reorganization of institutional and pedagogical frameworks must accompany such a technological deployment. The model's success is not only a technical accomplishment but also a pedagogical challenge, asking academic institutions to reconsider how learning analytics are ethically communicated, how feedback loops are used to support students, and how digital skills are incorporated into legal curricula.

Importantly, the difficulty lies not only in forecasting performance but also in taking appropriate action. Whether through real-time alerts, targeted mentoring, or adaptive content recommendation, predictive accuracy is only significant if it results in personalized and actionable interventions. Our capacity to integrate human pedagogical agency with predictive intelligence will determine the future of legal education in the digital age, ensuring that AI enhances rather than replaces the role of the teacher.

5.4. Real world Implications

The study suggests a set of educational implications (Figure 4) based on an analysis of behavioral patterns, since the model is very good at putting students into performance groups. These suggestions are meant to make instruction more responsive, help with timely interventions, and improve self-regulation, especially for law students who are using digital skills in a technological environment that is not their own.

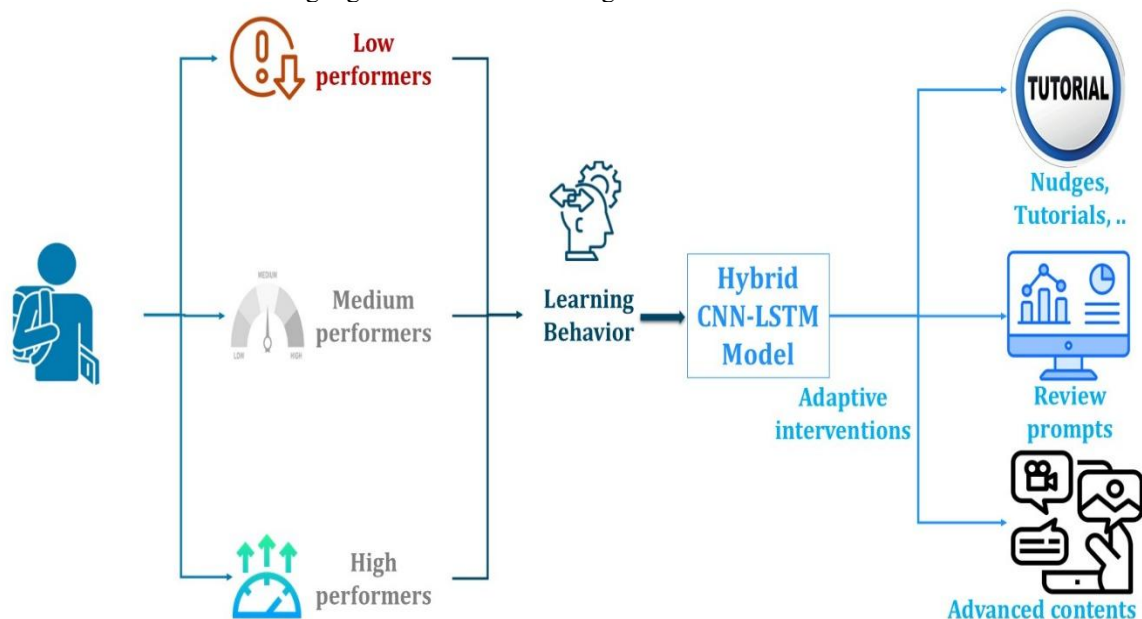


Figure 4. Personalized implications

After the first part of the course, learners who consistently do poorly are more likely to show behavioral signs, such as less variety in their interactions, less frequent attendance, and less engagement. These patterns suggest cognitive disengagement and low digital self-efficacy. In these cases, intervention strategies should put a lot of emphasis on progressive scaffolding and reactivation. The learning system is sited up to include timely reminders and motivational nudges that start when someone isn't doing anything. These students might also benefit from easier digital exercises that are tailored to their current skill level. This will help them get back into the swing of things without feeling overwhelmed.

Moderate-performing students need help to stabilize their learning paths. These students often have irregular learning rhythms and become more engaged during tests. Students can learn pacing better if they have proactive support systems, like real-time dashboards that show their progress and deadlines, instead of just reactive feedback. Structured group discussions and short review sessions with peers may help them remember what they learned better. Adding gamified progress-tracking features can also boost intrinsic motivation by making the process of reaching a goal more enjoyable.

On the other hand, learners who log in regularly, use resources wisely, and get consistently high scores are usually considered high performing. These students benefit from enrichment activities that go beyond the regular curriculum. Giving them early access to advanced or exploratory digital content helps them grow. At the same time, giving them mentorship roles has two benefits: it strengthens their knowledge and helps their peers who are not as advanced. Encouraging students to finish certified digital literacy modules can also help them reach their academic and professional goals, which will help them grow in the long term.

The model's outputs can be used directly in Learning Management Systems (LMS) like Moodle to make this personalization happen. Rule-based triggers and adaptive logic pathways that respond to how learners act in real time make this integration possible. For instance, if someone hasn't interacted with the site for five days or more, a motivational message can be automatically sent out with a one-click re-entry activity. A sudden drop in performance may also lead to the scheduling of review sessions with specific feedback. On the other hand, students who use the resources regularly and keep their average score above 90% may be given access to more advanced resources or recognition badges. This encourages them to stay engaged.

This model-driven personalization has big effects on teaching, especially in legal education, where students often don't have much experience with digital platforms before they start. By making it easier to spot disengagement early on and giving targeted, context-aware interventions, teachers can focus their efforts on the most important areas without having to do a lot of manual supervision. This smart orchestration creates a learning environment that is dynamic and welcoming to everyone, which helps all students, no matter how digitally fluent they are, develop their digital skills over time.

6. CONCLUSION

This work is an important contribution to the growing body of research that explores the relationship between deep learning, educational data mining, and digital transformation in non-technical domains. We used a hybrid CNN–LSTM architecture to model and forecast the digital learning behaviors of 35,000 law students who are developing critical digital skills. Research on AI-driven education has historically underrepresented this population. The model's ability to correctly categorize students between performance levels emphasizes how crucial it is to take behavioral dynamics and motivational profiles into account when researching learning processes outside of STEM settings. This work's real worth lies in its implications for educational decision-making as well as its predictive performance. The model provides interpretable insights into the behavioral mechanisms underlying digital disengagement and success in addition to accurately identifying problematic learners. This makes it easier to implement tailored interventions that honor learners' autonomy and tackle the systemic obstacles to digital competency that are commonly found in legal education.

Future research will be conducted in several different directions. The integration of multimodal learning data, such as textual forum contributions, video engagement patterns, and emotional indicators, could, first and foremost, improve the behavioral models and the precision and subtlety of predictions. Second, the model could operationalize predictive feedback loops for teachers and students if it is implemented in real-time adaptive systems within learning management platforms. Third, a crucial area for improvement is the exploration of ethical-pedagogical frameworks for the application of AI in non-technical fields. This will guarantee that educational equity is upheld and that predictions do not turn into profiling.

REFERENCES

- [1] Bali, S., Chen, T. C., & Liu, M. C. (2024). Behavioral Intentions of Low-Achieving Students to Use Mobile English Learning: Integrating Self-Determination Theory, Theory of Planned Behavior, and Technology Acceptance Model Approaches. *International Journal of Human–Computer Interaction*, 41(9), 5522–5532. <https://doi.org/10.1080/10447318.2024.2364142>
- [2] Liu, H., Wang, Y., & Wang, H. Exploring the mediating roles of motivation and boredom in basic psychological needs and behavioural engagement in English learning: a self-determination theory perspective. *BMC Psychol* 13, 179 (2025). <https://doi.org/10.1186/s40359-025-02524-3>
- [3] Jaboob, M., Hazaimh, M., & Al-Ansi, A. M. (2024). Integration of Generative AI Techniques and Applications in Student Behavior and Cognitive Achievement in Arab Higher Education. *International Journal of Human–Computer Interaction*, 41(1), 353–366. <https://doi.org/10.1080/10447318.2023.2300016>
- [4] Phan, M.L., Renshaw, T.L., Domenech Rodríguez, M.M. et al. Evaluating the Effects of a Teacher-Implemented Mindfulness-Based Intervention on Teacher Stress and Student Prosocial Behavior. *Contemp School Psychol* 29, 393–409 (2025). <https://doi.org/10.1007/s40688-025-00540-z>
- [5] Zhang Y, Li J. Deep learning-based model for predicting student learning behavior: A pathway to early intervention and enhanced outcomes. *Journal of Computational Methods in Sciences and Engineering*. 2025;25(3):2822-2835. doi:10.1177/14727978251322332
- [6] Saqlain M. Evaluating the readability of English instructional materials in Pakistani universities: a deep learning and statistical approach. *Educ Sci Manag* 2023; 1(2): 101–110.
- [7] Ouyang Y, Wang YL, Ye ZW. Multi-level contrastive graph learning for academic abnormality prediction. *Neural Comput Appl* 2024; 36(7): 3681–3698.
- [8] Stasolla, F., Zullo, A., Maniglio, R., Passaro, A., Di Gioia, M., Curcio, E., & Martini, E. (2025). Deep Learning and Reinforcement Learning for Assessing and Enhancing Academic Performance in University Students: A Scoping Review. *AI*, 6(2), 40. <https://doi.org/10.3390/ai6020040>
- [9] Wang, D., Han, L., Cong, L., Zhu, H., & Liu, Y. (2023). Practical Evaluation of Human–Computer Interaction and Artificial Intelligence Deep Learning Algorithm in Innovation and Entrepreneurship Teaching Evaluation. *International Journal of Human–Computer Interaction*, 41(3), 1742–1750. <https://doi.org/10.1080/10447318.2023.2199632>
- [10] Zhao, Z. (2025). Optimization of Personalized Learning Paths in Educational AI Driven by Student Behavior Data. *Journal of Industrial Engineering and Applied Science*, 3(2), 16–25. <https://doi.org/10.70393/6a69656173.323738>
- [11] Lampropoulos, G., & Evangelidis, G. (2025). Learning Analytics and Educational Data Mining in Augmented Reality, Virtual Reality, and the Metaverse: A Systematic Literature Review, Content Analysis, and Bibliometric Analysis. *Applied Sciences*, 15(2), 971. <https://doi.org/10.3390/app15020971>
- [12] Khoshgofar, Zohreh1; Babae, Maryam2; Rouzbahani, Arian K.1,3; Kalantarion, Masomeh1. Educational data mining in medical

- education: A five-level approach. *Journal of Education and Health Promotion* 14(1):24, January 2025. | DOI: 10.4103/jehp.jehp_1339_23
- [13] Wong, B. T. M., Li, K. C., & Liu, M. (2025). The Role of Learning Analytics in Evaluating Course Effectiveness. *Sustainability*, 17(2), 559. <https://doi.org/10.3390/su17020559>
- [14] Alturkistani, A.; Lam, C.; Foley, K.; Stenfors, T.; Blum, E.R.; Van Velthoven, M.H.; Meinert, E. Massive Open Online Course Evaluation Methods: Systematic Review. *J. Med. Internet Res.* 2020, 22, e13851.
- [15] Wen, L., Gao, L., Li, X., & Zeng, B. (2021). Convolutional neural network with automatic learning rate scheduler for fault classification. *IEEE Transactions on Instrumentation and Measurement*, 70(99), 3509912. <https://doi.org/10.1109/TIM.2020.3048792>
- [16] Haridas, M., Gutjahr, G., Raman, R. et al. Predicting school performance and early risk of failure from an intelligent tutoring system. *Educ Inf Technol* 25, 3995–4013 (2020). <https://doi.org/10.1007/s10639-020-10144-0>