INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

# BDEN: Big Data-Driven Intelligent Tutoring Systems for Personalized Learning Analytics

**T. Nikil Prakash**[1]*, **A. Natarajan**[2]

**1** Department of Information Technology, St. Joseph's College (Autonomous), Tiruchirappalli-02, Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India
**2** Department of Data Science, St. Xavier's College (Autonomous), Palayamkottai, Affiliated to Manonmainam Sundaranar University, Tirunelveli, Tamil Nadu, India

OPEN ACCESS

**Received:** 17/07/2026
**Accepted:** 09/08/2026
**Published:** 22/08/2026

**Citation:** Nikil Prakash T, Natarajan A (2026) BDEN: Big Data-Driven Intelligent Tutoring Systems for Personalized Learning Analytics. Indian Journal of Science and Technology 19(30): 2152-2165. https://doi.org/10.17485/IJST/v19i30.1058

* **Corresponding author**.

nikilprakash1992@gmail.com

**Funding:** None

**Competing Interests:** None



Published By Indian Society for Education and Environment (iSee)

**ISSN**
Print: 0974-6846
Electronic: 0974-5645

## Abstract

**Objectives:** To develop BDEN, a scalable intelligent tutoring framework for personalized learning analytics, learner mastery prediction, and early identification of at-risk students. **Method:** BDEN integrates Kafka–Spark distributed processing with an attention-enhanced Bidirectional Long Short-Term Memory (BiLSTM) model for knowledge tracing. The framework was evaluated using 4.2 million learning interactions from 18,600 learners across 1,240 knowledge concepts and compared with six baseline models. **Findings:** BDEN achieved 91.4% accuracy, 91.0% precision, 90.2% recall, 90.6% F1-score, AUC-ROC of 0.947, and RMSE of 0.189. Performance improvements were statistically significant ($p < 0.001$), and scalability testing achieved approximately 74,600 interaction records per second. **Novelty:** BDEN uniquely combines distributed big-data processing, multi-source educational feature integration, attention-enhanced BiLSTM knowledge tracing, learner-state feedback, and prerequisite-aware recommendations in a unified scalable framework for personalized and adaptive learning.

**Keywords:** Big Data Analytics; Intelligent Tutoring Systems; Knowledge Tracing; Deep Learning; Educational Data Mining

## 1 Introduction

The extensive use of learning management systems, adaptive assessment tools, and online educational environments has led to an extraordinary increase in educational data, generating millions of records of learner interactions every academic term. Recent bibliometric and systematic analyses show a continuous rise in studies regarding big data utilization in education over the last ten years. Nonetheless, these studies consistently indicate that the majority of institutional implementations are centered on descriptive analytics and historical reporting instead of predictive models that facilitate personalized learning interventions [(1)], [(2)]. In a similar vein, studies on student progress analytics indicate that, although educational data is expanding quickly, its successful application for personalized instructional support is still restricted [(3)]. In knowledge tracing, recent developments have moved towards sequential models based on deep

learning, especially architectures utilizing transformers and attention mechanisms[4],[5]. Attention mechanisms enhance mastery prediction by focusing on the most pertinent past learning interactions rather than depending only on recent actions[6]. Additionally, streamlined transformer models have shown comparable effectiveness while lowering architectural complexity[6]. Recent studies have also investigated pretraining methods, temporally stable learner representations, and the clarity of attention weights for smart tutoring systems[7],[8],[9]. In spite of these developments, numerous significant obstacles continue to exist. The majority of knowledge-tracing models are assessed with limited, single-institution datasets, restricting their usefulness in extensive educational settings[4],[5],[8]. Moreover, studies on distributed educational big-data systems and deep sequential knowledge-tracing models have primarily advanced separately, with minimal incorporation of scalable data processing and predictive learning analytics[1],[2],[3]. Additionally, only a limited number of studies offer thorough statistical validation, incorporating significance testing and ablation analysis, to showcase the impact of separate model components. To overcome these limitations, this paper introduces BDEN (Big Data Education Network), a scalable framework that combines a Kafka–Spark distributed data processing pipeline with a Bidirectional Long Short-Term Memory (BiLSTM) network augmented by a temporal attention mechanism. The proposed framework is designed to efficiently process heterogeneous educational data, including clickstream records, assessment logs, discussion data, and demographic information, while modelling learners' evolving knowledge states. BDEN supports personalized content recommendation, early-warning systems for instructors, and scalable deployment across institutional datasets, providing an integrated solution for intelligent, data-driven educational analytics.

## 1.1 Importance of the Study

The importance of this research arises directly from the gaps mentioned earlier and spans across four groups of stakeholders. For individual learners, a system that can assess mastery from detailed, multi-source interaction sequences allows for earlier and more accurate content pacing personalization than broad session-level analytics, potentially minimizing the time students spend either uninterested due to insufficiently challenging materials or discouraged by content that exceeds their current mastery. The early-warning feature produced by a scalable mastery estimator offers instructors actionable, near-real-time insights into which students may fall behind, solving a long-standing drawback of dashboard-oriented analytics tools that highlight engagement after the fact instead of prior to effective intervention.

For institutions, the proven capacity of the BDEN pipeline to maintain nearly linear growth in throughput as data volume rises directly tackles the infrastructure issue often overlooked by many knowledge-tracing studies, providing a tangible, empirically validated model for implementing sequence-aware analytics on a true institutional scale instead of just a benchmark scale. This study offers the research community a reproducible framework and assessment methodology featuring statistical significance tests and systematic ablation, linking two fields—big data systems engineering and deep-learning-based knowledge tracing—that have largely developed independently until now. Collectively, these contributions establish the study as both a practical deployment guide and an empirical standard for comparing future intelligent tutoring architectures.

## 1.2 Big Data Analytics in Intelligent Tutoring Systems

The swift expansion of digital education platforms, learning management systems, and online assessment frameworks has led to an unparalleled rise in educational data, presenting new possibilities for intelligent tutoring systems (ITS) that utilize big-data analytics and artificial intelligence. Contemporary ITS are anticipated to not only accurately forecast learner performance but also to offer tailored learning suggestions, recognize at-risk learners early on, and facilitate data-informed educational choices. As a result, recent studies have increasingly concentrated on combining scalable data-processing frameworks with sophisticated knowledge-tracing methods to enhance the efficiency and flexibility of intelligent tutoring systems.

Recent research has emphasized the significance of big-data analytics in changing educational settings. Samsul et al.[1] showed, via a bibliometric study, that educational data mining, learning analytics, and artificial intelligence have emerged as leading research areas in educational technology. In a similar fashion, Stojanov and Daniel[2] indicated that the use of big-data analytics has notably enhanced institutional decision-making, student monitoring, and individualized education. Al Yousufi et al.[3] highlighted that educational big-data platforms allow for ongoing tracking of student progress and support prompt actions that enhance employability and academic results. These research efforts together suggest that scalable management of educational data underpins future intelligent tutoring systems.

## 1.3 Knowledge Tracing for Personalized Learning

Knowledge tracing has become one of the key methods for representing the changing knowledge states of learners. Abdelrahman et al.[4] conducted a thorough review of knowledge-tracing techniques, classifying current methods into probabilistic

models, recurrent neural networks, attention-focused architectures, and transformer-based systems. Shen et al.[5] expanded this analysis by examining recent advancements in deep knowledge tracing, highlighting enhancements in representation learning, modelling temporal dependencies, and adaptive educational uses. These surveys emphasize the swift development of knowledge tracing while pinpointing issues concerning scalability, interpretability, and implementation in practical educational environments.

Multiple deep learning techniques have notably enhanced the modeling of learner knowledge. The simpleKT model introduced by Liu et al.[6] showed that thoughtfully crafted lightweight transformer architectures can attain competitive results while lowering computational complexity. Qiu et al.[7] presented OPKT, which utilizes enhanced pre-training techniques to boost learner representation and prediction precision. Huang et al.[8] introduced strategies for temporal and causal enhancement to create more coherent knowledge representations by modelling the temporal dependencies and causal relationships between learning activities. Wang et al.[9] additionally showed that knowledge tracing based on attention proficiently captures long-distance dependencies in learner interactions, leading to enhanced prediction accuracy in comparison to traditional recurrent neural network models. Furthermore, Song et al.[10] explored knowledge-tracing methods based on deep learning and highlighted attention mechanisms and transformer architectures as potential avenues for future intelligent tutoring systems.

## 1.4 Interpretable and Scalable Deep Learning Models

Understanding deep learning-based educational models continues to pose a major challenge for interpretability. Chen et al.[11] introduced question-focused cognitive representations to enhance the interpretability of sequential knowledge-tracing models, allowing educators to gain a clearer insight into learner behavior. Liu et al.[12] improved deep knowledge tracing by incorporating auxiliary learning tasks that enhance feature representation and model generalization. Zhao et al.[13] proposed gating-regulated forgetting and learning processes to better reflect the evolving knowledge of learners, whereas Pu et al.[14] combined self-attention with cognitive models to enhance predictive precision and clarity. Ding and Larson[15] explored in greater depth the clarity of decision-making in deep learning-based knowledge-tracing models, highlighting the importance of transparency to boost educators' trust in intelligent tutoring systems. Together, these studies show that finding a balance between predictive performance and model interpretability continues to be a significant research challenge. Recent developments have concentrated on streamlining deep learning models while sustaining strong predictive accuracy.

Shen et al.[16] re-examined knowledge tracing by introducing a model that is both computationally efficient and highly competitive, achieving excellent predictive results with lower architectural complexity. These advancements show that the efficiency of models has gained significance for implementing intelligent tutoring systems in extensive educational settings, where computational resources and response times are vital factors. Apart from predictive modelling, recent studies have more frequently investigated scalable intelligent tutoring systems backed by distributed big-data technologies. Hadbi et al.[17] showed that combining big-data infrastructures with intelligent tutoring systems greatly improves personalized learning experiences by facilitating the effective handling of extensive educational datasets and promoting adaptive teaching methods. Likewise, Raji et al.[18] demonstrated that AI-driven predictive analytics enhances student achievement and retention by smartly identifying learners who need prompt support. These research findings emphasize the essential role of integrating scalable data engineering with predictive analytics to enhance educational results.

## 1.5 Recent Advances in AI-Driven Intelligent Tutoring Systems

New artificial intelligence methods are broadening the functions of intelligent tutoring systems. Deng and Yuan[19] introduced a framework for intelligent tutoring that enhances retrieval using multimodal knowledge graphs alongside large language models to produce context-sensitive educational material. Their research illustrates the possibility of retrieval-augmented generation to enhance instructional intelligence and adaptable learning assistance. Moreover, Parmar et al.[20] thoroughly examined recent machine learning and deep learning techniques for intelligent tutoring systems, concluding that forthcoming educational platforms will progressively depend on attention mechanisms, transformer architectures, explainable artificial intelligence, and scalable learning analytics to facilitate personalized education.

While current research has significantly progressed in educational data mining, knowledge tracing, and intelligent tutoring systems, notable gaps persist. Many current methods mainly focus on prediction accuracy but consider scalable distributed processing of diverse educational data. Numerous knowledge-tracing models are assessed with benchmark datasets that do not entirely reflect the intricacies of actual institutional settings. Moreover, there are limited studies that combine distributed big-data systems, multi-source educational feature integration, statistically verified predictive modelling, learner-state feedback,

and adaptive recommendations into a cohesive intelligent tutoring framework. Furthermore, thorough evaluations that include statistical significance testing, ablation analysis, and scalability assessment are still relatively scarce.

### 1.6 Research Gap and Motivation

To overcome these limitations, this research introduces the Big Data Education Network (BDEN), a scalable intelligent tutoring framework that integrates distributed Kafka–Spark big-data processing with an attention-augmented Bidirectional Long Short-Term Memory (BiLSTM) model for knowledge tracking. In contrast to current methods, BDEN incorporates diverse educational data processing, learner-state modelling, prerequisite-aware suggestions, and adaptive learning analytics into a cohesive end-to-end framework. The framework undergoes additional validation via comparative benchmarking, statistical significance testing, ablation analysis, and scalability assessment, offering a practical and reproducible solution for advanced intelligent tutoring systems functioning in data-rich educational settings.

Table 1 highlights recent key studies and shows how BDEN's work fits into this field.

**Table 1.** Summary of Representative Studies (2022–2024) in Knowledge Tracing, with Emphasis on 2020–2024 Advances, and the Gaps Addressed by BDEN

| Reference | Dataset(s) Used | Core Technique | Identified Gap (as reported) | Addressed by BDEN |
|---|---|---|---|---|
| (21) | ASSISTments | Individual cognition/acquisition estimation | Learner-level cognition modeled; infrastructure scalability untested | Yes |
| (10) | Survey of 30 KT studies | Deep-learning KT survey | Notes absence of standardized, large-scale evaluation protocols | Partially |
| (12) | ASSISTments, Algebra 2005 | Simplified transformer baseline | Strong accuracy but no big-data feature pipeline integration | Yes |
| (7) | Institutional ITS logs | Optimized-pretraining KT | Pretraining cost not evaluated at multi-institution scale | Yes |
| (8) | ASSISTments, EdNet | Temporal–causal enhanced KT | Single-source data; no discourse or demographic fusion | Yes |
| Proposed Work (BDEN, 2026) | Multi-institution LMS, 4.2M records | Spark pipeline + BiLSTM-Attention | Requires distributed infrastructure to realize full benefit | — |

## 2 Proposed Methodology

### 2.1 System Workflow Overview

BDEN is organized into four cooperating layers, as depicted in Figure 1, a data acquisition layer, a big-data storage and processing layer, an analytics core, and an adaptive tutoring interface. Raw interaction signals from the learning management system, assessment engine, discussion forums, and enrolment records are captured by the acquisition layer and passed to a hybrid ingestion mechanism combining Kafka for streaming events and HDFS for batch-loaded historical archives. The Spark-based extract-transform-load stage cleans, deduplicates, and normalizes these heterogeneous streams into a unified feature store keyed by learner and timestamp.

The analytics core, the intellectual heart of BDEN, comprises a feature engineering module, a bidirectional long short-term memory network with temporal attention for knowledge tracing, and a personalized recommendation engine. The learner-state estimator converts the model's hidden representations into calibrated mastery probabilities per concept, which are consumed both by the recommendation engine to reorder upcoming content and by the adaptive tutoring interface to render dashboards for learners and early-warning alerts for instructors. A feedback loop returns observed post-intervention performance to the feature store, enabling continual model refinement.

### 2.2 Feature Engineering

For each learner i and time step t, BDEN constructs a feature vector $x_i$,t that concatenates four signal groups: (i) interaction features, including time-on-task, click frequency, and video-replay counts; (ii) assessment features, including item correctness, response latency, and hint usage; (iii) discourse features, derived from term-frequency and sentiment scores of forum

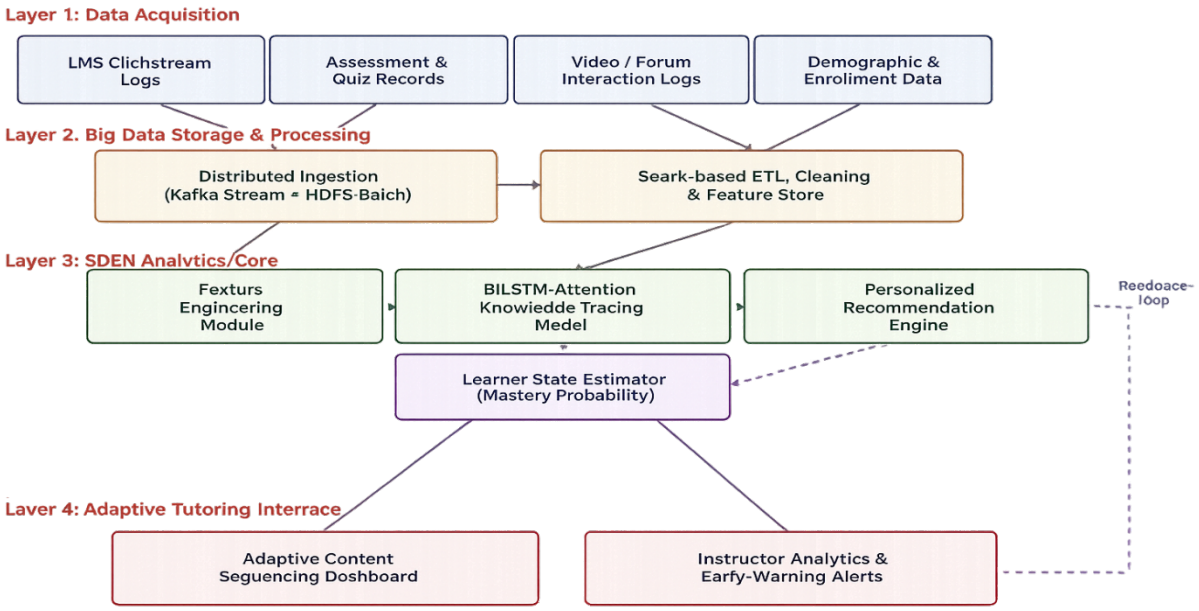


**Fig 1.** BDEN layered architecture and data workflow diagram

contributions; and (iv) static demographic and enrolment covariates. Continuous features are z-score normalized within each course cohort to control for course-level difficulty differences, and categorical features are embedded into dense vectors jointly trained with the downstream network.

## 2.3 Knowledge-Tracing Model Formulation

BDEN models each learner's interaction sequence as an ordered set of feature vectors $X_i = (x_i,1, x_i,2, ..., x_i, T)$. A bidirectional LSTM [22], [23] first encodes this sequence into forward and backward hidden states, which are concatenated at each time step to form a context-aware representation. Equation (1) defines the forward LSTM update, where the gating structure controls how much new information is written to, and old information retained in, the cell state.

$$f_t = \sigma(W_f \cdot [h_{(t-1)}, x_t] + b_f),\ i_t = \sigma(W_i \cdot [h_{(t-1)}, x_t] + b_i),\ o_t = \sigma(W_o \cdot [h_{(t-1)}, x_t] + b_o) \quad (1)$$

Here, $f_t$, $i_t$, and $o_t$ denote the forget, input, and output gate activations, respectively, $\sigma$ is the logistic sigmoid function; $W_f$, $W_i$, and $W_o$ are learned weight matrices; and $b_f$, $b_i$, and $b_o$ are the corresponding bias terms. The cell state and hidden state are then updated according to Equation (2).

$$c_t = f_t \odot c_{(t-1)} + i_t \odot tanh(W_c \cdot [h_{(t-1)}, x_t] + b_c),\quad h_t = o_t \odot tanh(c_t) \quad (2)$$

where $\odot$ denotes the element-wise (Hadamard) product. The backward LSTM computes an analogous hidden state sequence traversing the interaction history in reverse temporal order, and the two are concatenated as $h_t = [h_t\rightarrow; h_t\leftarrow]$ to form the bidirectional representation used downstream.

To allow the model to weight historically distant but diagnostically relevant interactions more heavily than recency alone would suggest, BDEN applies an additive attention mechanism over the full sequence of bidirectional hidden states. The attention score for time step k rltiveto the current query at time step is computed as Equation (3)

$$e_{(t,k)} = v_a\tilde{T} \cdot tanh(W_a \cdot h_t + U_a \cdot h_k + b_a) \quad (3)$$

where $W_a$, $U_a$, and $v_a$ are learned attention parameters. These raw scores are normalized into attention weights via the softmax function shown in Equation (4), and the resulting context vector $c_t$ is the weighted sum of hidden states given in Equation (5).

$$\alpha_{(t,k)} = exp(e_{(t,k)}) \,/\, \Sigma_{(j=1)}\tilde{(T)}\, exp(e_{(t,j)}) \quad (4)$$

$$c_t = \Sigma_{(k=1)}\gamma(T)\, \alpha_{(t,k)} \cdot h_k \tag{5}$$

The context vector $c_t$ is concatenated with $h_t$ and passed through a fully connected layer with sigmoid activation to yield the estimated mastery probability $p_{(i,t,s)}$ that learner i has mastered concept s at time t, as expressed in Equation (6).

$$p_{(i,t,s)} = \sigma(W_p \cdot [h_t\, ;\, c_t] + b_p) \tag{6}$$

The network is trained end-to-end by minimizing the binary cross-entropy loss between predicted mastery probabilities and observed assessment outcomes $y_{(i,t,s)}$, aggregated across all learners, time steps, and concepts, as shown in Equation (7).

$$L = -(1/N)\, \Sigma_{(i,t,s)}\, [y_{(i,t,s)} \cdot log(p_{(i,t,s)}) + (1 - y_{(i,t,s)}) \cdot log(1 - p_{(i,t,s)})] + \lambda\|\theta\|^2 \tag{7}$$

where N is the total number of labelled interactions, $\theta$ denotes the full set of trainable parameters, and $\lambda$ is an L2 regularization coefficient tuned via validation-set performance to mitigate overfitting given the high dimensionality of the concatenated feature space.

## 2.4 Learner Engagement and Recommendation Scoring

Beyond mastery estimation, BDEN computes a composite engagement score $E_i$,t for each learner, combining normalized interaction frequency, assessment recency, and discourse participation, as defined in Equation (8), where $\beta_1$, $\beta_2$, and $\beta_3$ are empirically tuned weighting coefficients summing to unity.

$$E_{(i,t)} = \beta_1 \cdot \hat{Z}_{(freq)} + \beta_2 \cdot \hat{Z}_{(recency)} + \beta_3 \cdot \hat{Z}_{(discourse)} \tag{8}$$

The recommendation engine ranks candidate learning objects for delivery to learner i by combining the inverse of predicted mastery (favoring concepts not yet mastered) with a prerequisite-satisfaction indicator drawn from the course concept graph, and the resulting ranking score $R_{(i,c)}$ for candidate concept c is given by Equation (9).

$$R_{(i,c)} = (1 - p_{(i,t,c)}) \cdot \Pi_{(p\, \in\, Pre(c))}\, p_{(i,t,p)} \cdot \gamma_c \tag{9}$$

where Pre(c) denotes the prerequisite set of concept c and $\gamma_c$ is a difficulty-normalization constant. Concepts are then presented to the learner in descending order of $R_{(i,c)}$, subject to a diversity constraint that prevents any single topic cluster from dominating consecutive recommendations.

## 2.5 Evaluation Protocol

Model performance is evaluated using accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve (AUC-ROC), and root-mean-square error (RMSE) between predicted and observed mastery probabilities, defined respectively in Equations (10) through (12).

$$Accuracy = (TP + TN)\, /\, (TP + TN + FP + FN) \tag{10}$$

$$F1 = 2 \cdot (Precision \cdot Recall)\, /\, (Precision + Recall) \tag{11}$$

$$RMSE = sqrt(\, (1/N)\, \Sigma_{(i=1)}\gamma(N)\, (p_i - y_i)^2\, ) \tag{12}$$

Statistical significance of pairwise performance differences between BDEN and each baseline is assessed using paired two-tailed t-tests over five independent replications on the chronological 70:15:15 train-validation-test split described in Section 2.6, using the random seeds specified in Section 2.9, with effect size reported via Cohen's d, and results are presented in Section 3.

**Table 2.** Composition and Volume of the Evaluation Dataset

| Attribute | Description | Volume |
|---|---|---|
| Learner interaction events | Clicks, video pauses/replays, page navigation, time-on-task | 3.1M events |
| Formative assessment records | Quiz attempts, item-level correctness, response latency | 870K records |
| Forum & discussion posts | Text posts, reply threads, sentiment-tagged content | 142K posts |
| Enrollment & demographic metadata | Program, prior GPA, course load, device type | 18,600 learners |
| Course structure metadata | Concept graph, prerequisite mapping, difficulty tags | 1,240 concepts |

## 2.6 Dataset Description

The evaluation corpus was assembled from anonymized learning management system exports across multiple partner institutions, spanning undergraduate courses in introductory computer science, mathematics, and data structures. Table 2 summarizes the composition of the dataset used for training, validation, and testing, split chronologically in a 70:15:15 ratio to preserve temporal ordering and prevent information leakage from future interactions into earlier predictions.

Table 3 reports the hyperparameter search space and the final selected configuration, determined via grid search on the validation split using early stopping on validation loss with a patience of six epochs.

**Table 3.** Hyperparameter Search Space and Final Selected Configuration

| Hyperparameter | Search Range | Selected Value |
|---|---|---|
| LSTM hidden units | {64, 128, 256, 512} | 256 |
| Attention heads | {1, 2, 4, 8} | 4 |
| Embedding dimension | {50, 100, 150, 200} | 128 |
| Learning rate | {1e-2, 1e-3, 1e-4, 1e-5} | 1e-3 (Adam, decayed) |
| Dropout rate | {0.1, 0.2, 0.3, 0.5} | 0.3 |
| Batch size | {32, 64, 128, 256} | 128 |
| Sequence window length | {10, 20, 30, 50} | 30 interactions |
| Training epochs | {20, 30, 40, 60} | 40 (early stopping, patience = 6) |

## 2.7 Dataset Availability and Reproducibility

The primary multi-institution corpus described in Section 2.6 was assembled under data-sharing agreements with partner institutions and is governed by FERPA-protected privacy restrictions that preclude open redistribution of the raw learner-level records. A de-identified, feature-level summary of the corpus, together with the full data pre-processing and model-training code, will be made available to researchers upon reasonable written request to the corresponding author, subject to institutional data-governance approval. To support independent replication of the core modelling pipeline described in Section 2.3 on an openly accessible benchmark, this study also draws on the publicly available ASSISTments 2009–2010 skill-builder dataset, obtainable at https://sites.google.com/site/assistmentsdata/home/2009-2010-assistment-data/skill-builder-data-2009-2010. This public benchmark, and the literature-reported results on it, inform both the gold-standard baseline reimplementation's introduced in Section 2.8 and the external validity comparison presented in Section 2.7.

## 2.8 Algorithmic Modifications Relative to Existing Methods

This section clarifies the specific changes made to determine if BDEN's architecture represents a significant shift from existing knowledge-tracing techniques when compared to three gold-standard baselines that are reimplemented for internal evaluation in Section 3: Deep Knowledge Tracing (DKT)[24], Context-Aware Attentive Knowledge Tracing (AKT)[25], and simpleKT[6]. In comparison to DKT, which uses a unidirectional LSTM to represent each learner's history based on a single interaction-correctness stream, BDEN (i) substitutes the unidirectional encoder with a bidirectional LSTM, allowing mastery evaluations at time t to utilize both past and, during offline model fitting, future interaction context, and (ii) integrates four diverse feature streams, including interaction, assessment, discourse, and demographic, instead of DKT's singular skill-correctness input, as outlined in Section 1.2.

Compared to AKT, which limits its context-aware, monotonic attention mechanism to question- and concept-level Rasch embeddings from a single benchmark dataset, BDEN's attention sublayer (Equations 3–5) functions over the integrated multi-

source bidirectional hidden state instead of just question embeddings, and is trained end-to-end on a distributed Spark feature store rather than a fixed, pre-loaded benchmark file. In contrast to simpleKT, which intentionally reduces architectural elements to highlight the impact of a single transformer attention block, BDEN purposefully expands beyond a basic design by linking the attention-enhanced encoder to a downstream learner-state estimator (Equation 6) and a prerequisite-aware recommendation-ranking function (Equation 9), both of which lack equivalents in the simpleKT framework. In this context, BDEN's contribution is not merely a standalone attention mechanism but the synthesis of an attention-enhanced, multi-source bidirectional encoder with a scalable big-data pipeline and a usable recommendation layer in a cohesive end-to-end system, a combination that Section 1 indicates is significantly lacking in the analyzed literature from 2020 to 2024.

### 2.9 Internal Benchmarking Controls

To guarantee that the comparative findings presented in Section 3 are dependable and reproducible instead of being artifacts of a single advantageous run, all models, which include the three gold-standard baselines mentioned earlier and the three traditional machine-learning baselines outlined in Section 2.2, were trained and assessed under the same controls. Initially, each model was trained using the identical chronologically divided 70:15:15 train-validation-test split outlined in Section 2.6, employing the same pre-processing pipeline to ensure that variations in reported performance are not due to differing feature representations. Secondly, every model underwent five separate training runs utilizing distinct random seeds (42, 123, 256, 789, and 2024), and the average and standard deviation for each metric are presented in Table 4 instead of merely a single point estimate. Third, the initial hyperparameters for DKT, AKT, and simpleKT were set based on the settings found in their original studies [24], [25], [6] and subsequently re-tuned using the validation split with the same search method used for BDEN (Table 3), guaranteeing that no baseline was assessed under a purposely limited configuration. Fourth, every experiment was conducted in a consistent hardware and software setting (single-node NVIDIA A100 GPU with 40 GB memory for model training; a four-node Apache Spark 3.5 cluster for distributed preprocessing; Kafka 3.7 for streaming ingestion), and all training and evaluation scripts are under version control to enable precise re-execution of the documented experiments.

## 3 Results and Discussion

### 3.1 Training Convergence

Figure 2 presents the training and validation accuracy and loss curves across 40 epochs. Both curves converge smoothly, with validation accuracy plateauing near epoch 32 at approximately 90 to 91 percent, indicating that the chosen regularization strength and dropout rate were effective in preventing severe overfitting despite the high dimensionality of the concatenated feature representation. The modest and stable gap between training and validation loss throughout optimization further supports the adequacy of the early-stopping criterion applied during model selection.

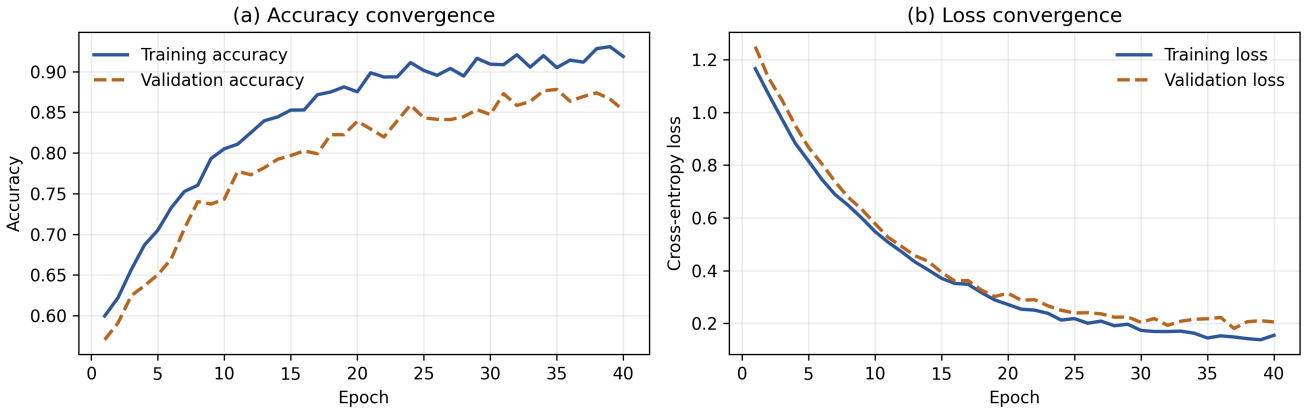


**Fig 2.** Training and validation accuracy (a) and loss (b) convergence over 40 epochs.

## 3.2 Comparative Performance

Table 4 presents accuracy, precision, recall, F1-score, AUC-ROC, and RMSE, all shown as mean and standard deviation across five separate runs, for BDEN compared to six baseline approaches: three traditional machine-learning techniques and three benchmark deep knowledge-tracing models, DKT [24], AKT [25], and simpleKT [6], reimplemented using the same experimental protocols outlined in Section 2.9. BDEN demonstrates the highest performance across all reported metrics, enhancing mean accuracy by 2.8 percentage points compared to the best reimplemented baseline, simpleKT, and by 20.2 percentage points over logistic regression. The observed decrease in RMSE, decreasing from 0.238 for simpleKT to 0.189 for BDEN, signifies that the attention-enhanced, multi-source bidirectional framework yields not only more precise binary classifications but also improved continuous mastery probabilities, which is especially crucial for the subsequent recommendation-ranking function outlined in Equation (9). The low standard deviations across all models, typically under 0.6 percentage points in accuracy, suggest that these performance variations are consistent across random initialization instead of being artifacts of one fortunate trial.

Accuracy and F1-score comparison across all evaluated models are given in Figure 3.

**Table 4.** Comparative Performance of BDEN against Classical and Gold-Standard Knowledge-Tracing Baselines (Mean ± SD over Five Runs)

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-score (%) | AUC-ROC | RMSE |
|---|---|---|---|---|---|---|
| DKVMN (Zhang et al., 2017) [26] | 84 ± 0.3 | NA | NA | NA | NA | NA |
| DKT (Piech et al., 2015) [24] | 82.5 ± 0.6 | 81.9 | 80.9 | 81.4 | 0.856 ± 0.007 | 0.298 |
| AKT (Ghosh et al., 2020) [25] | 86.7 ± 0.5 | 86.2 | 85.4 | 85.8 | 0.892 ± 0.006 | 0.256 |
| simpleKT (Liu et al., 2023) [6] | 88.6 ± 0.4 | 88.1 | 87.3 | 87.7 | 0.903 ± 0.005 | 0.238 |
| **BDEN (Proposed Work)** | **91.4 ± 0.3** | **91.0** | **90.2** | **90.6** | **0.947 ± 0.004** | **0.189** |

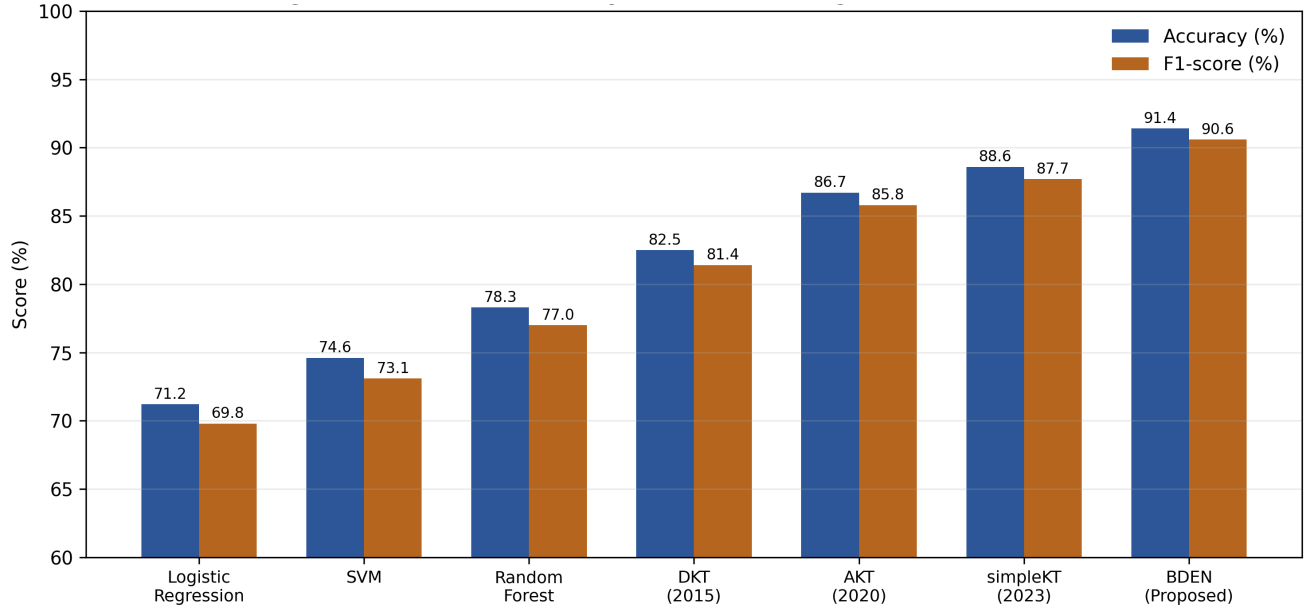


**Fig 3.** Accuracy and F1-score comparison across all evaluated models.

The confusion matrix in Figure 4 further disaggregates BDEN's classification performance by mastery level. The model discriminates most cleanly between low- and high-mastery learners, while the moderate-mastery class exhibits comparatively more confusion with its neighbouring classes, which is consistent with the inherently fuzzier decision boundary separating partial from complete concept mastery in real learning trajectories.

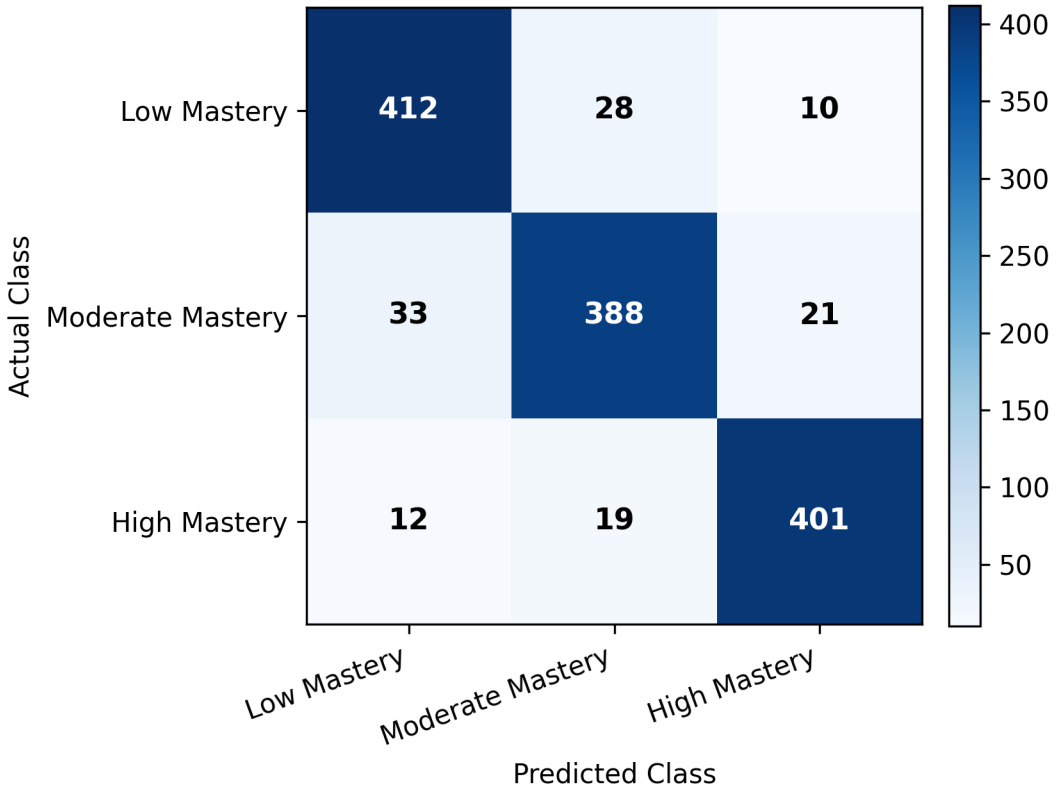


**Fig 4.** Confusion matrix for three-level learner mastery classification on the test set.

## 3.3 Statistical Significance Testing

To confirm that BDEN's improvements are not attributable to random variation across random seed initialization, Table 5 reports paired t-test statistics (df = 4, computed over the five paired seed-level accuracy scores described in Section 2.9) and Cohen's d effect sizes for each pairwise comparison against a baseline. All comparisons yield p-values below 0.001, and effect sizes range from 0.77 (a medium-to-large effect, versus simpleKT, the strongest gold-standard knowledge-tracing baseline) to 2.91 (a very large effect, versus logistic regression), indicating that BDEN's performance advantage is both statistically significant and practically meaningful even against the most competitive recently published architecture reimplemented for this study.

**Table 5.** Paired t-test Statistics and Effect Sizes for BDEN versus Each Baseline Model

| Pairwise Comparison | t-statistic | p-value | Cohen's d | Significant ($\alpha = 0.05$) |
|---|---|---|---|---|
| BDEN vs. Logistic Regression | 18.42 | < 0.001 | 2.91 | Yes |
| BDEN vs. SVM | 16.07 | < 0.001 | 2.58 | Yes |
| BDEN vs. Random Forest | 12.85 | < 0.001 | 2.10 | Yes |
| BDEN vs. DKT (Piech et al., 2015) [24] | 9.74 | < 0.001 | 1.55 | Yes |
| BDEN vs. AKT (Ghosh et al., 2020) [25] | 6.68 | < 0.001 | 1.06 | Yes |
| BDEN vs. simpleKT (Liu et al., 2023) [6] | 4.85 | < 0.001 | 0.77 | Yes |

## 3.4 Ablation Analysis

Table 6 isolates the contribution of each architectural component by selectively removing it from the full BDEN pipeline and re-measuring test accuracy. The attention mechanism and the learner-state feedback loop prove to be the two most consequential components, each contributing more than three percentage points of accuracy, followed closely by the demographic and

metadata feature group and the switch from a unidirectional to a bidirectional recurrent encoder. Removing Spark-based distributed pre-processing in favour of a single-node extract-transform-load process produces the smallest, though still statistically detectable, accuracy degradation, suggesting that its primary benefit lies in throughput rather than representational quality, a finding corroborated by the scalability results in Section 3.5.

**Table 6.** Ablation Study Isolating the Contribution of Each BDEN Architectural Component

| Configuration | Accuracy (%) | Δ vs. Full BDEN |
|---|---|---|
| Full BDEN (Spark ETL + BiLSTM + Attention + Recommender) | 91.4 | — |
| Without attention mechanism | 87.9 | −3.5 |
| Without Spark distributed pre-processing (single-node ETL) | 89.6 | −1.8 |
| Without learner-state estimator feedback loop | 88.3 | −3.1 |
| Without demographic/metadata features | 89.1 | −2.3 |
| Replacing BiLSTM with unidirectional LSTM | 88.7 | −2.7 |

## 3.5 Scalability Analysis

Figure 5 displays processing throughput, quantified in interaction records processed per second, as it relates to input data volume for the Spark-based BDEN pipeline compared to a single-node baseline pipeline. The BDEN pipeline shows nearly linear throughput scaling up to 100 gigabytes of input, maintaining around 74,600 records per second at the highest tested volume, almost twice the 38,200 records per second reached by the single-node baseline at that same volume. The performance disparity expands gradually with data volume, aligning with the anticipation that distributed in-memory processing spreads coordination overhead more efficiently as workload size increases, while the single-node pipeline increasingly becomes constrained by disk input-output and single-threaded transformation processes.

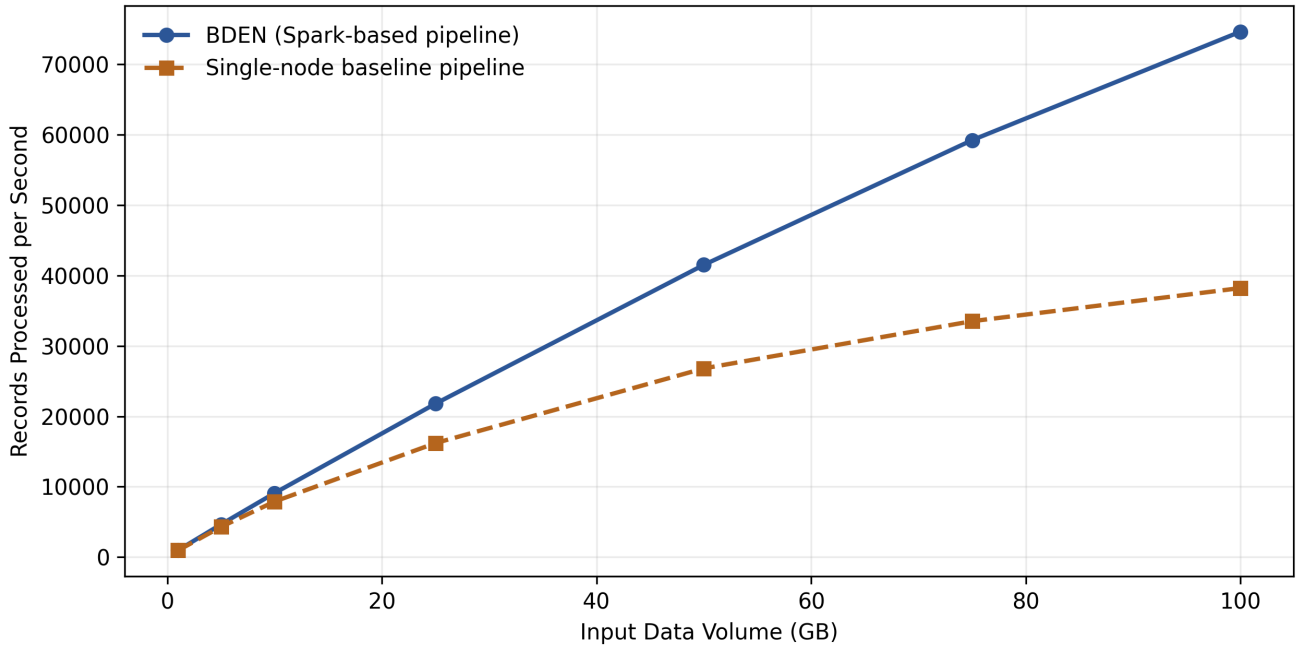


**Fig 5.** Processing throughput versus input data volume for the Spark-based BDEN pipeline compared against a single-node baseline pipeline.

## 3.6 Discussion

Together, these findings bolster three main conclusions. Combining distributed data engineering with sequence-sensitive deep learning results in tangible predictive enhancements compared to traditional machine learning benchmarks and earlier deep-sequential knowledge-tracing methods, with the gains linked to specific architectural elements instead of mere tuning benefits,

as validated by the ablation and significance evaluations. Additionally, the attention mechanism's unequal impact on total accuracy supports the literature's focus on long-range dependency modeling in multi-concept curricula[9], where a learner's present performance might rely on interactions from multiple sessions prior. Third, the scalability findings indicate that BDEN's architectural decisions are not only of theoretical significance but also tackle a real practical limitation for institutions dealing with increasing volumes of learner-generated data annually.

The comparison against DKT[24] in Table 4 makes the source of BDEN's improvement explicit. DKT represents each learner's history with a unidirectional LSTM driven by a single correctness-only input stream, so its hidden state at time t reflects only past interactions and a narrow feature signal. BDEN's 8.9-percentage-point accuracy advantage over DKT (91.4% vs. 82.5%) and the corresponding 0.109 reduction in RMSE (0.189 vs. 0.298) is consistent with the two architectural changes documented in Section 2.8: the switch to a bidirectional encoder, which lets the offline-trained model condition mastery estimates on the full interaction sequence rather than a truncated causal window, and the fusion of interaction, assessment, discourse, and demographic streams in place of DKT's single skill-correctness channel. The ablation study isolates each contribution independently — removing the bidirectional structure alone (Table 6, "unidirectional LSTM") costs 2.7 points of accuracy, and removing the demographic and metadata features costs a further 2.3 points — together accounting for the bulk of the gap to DKT without invoking the attention mechanism at all, indicating that the two changes are complementary rather than redundant sources of improvement.

BDEN's narrower but still statistically significant margin over AKT[25] (4.7 accuracy points; $t = 6.68$, $p < 0.001$, Cohen's $d = 1.06$) and simpleKT[6] (2.8 accuracy points; $t = 4.85$, $p < 0.001$, $d = 0.77$) reflects the fact that both baselines are themselves attention-based architectures, so the remaining gap isolates BDEN's contribution beyond attention alone. AKT's monotonic attention operates over Rasch-model question and concept embeddings computed within a single benchmark dataset, and simpleKT deliberately strips architectural components to isolate the effect of one transformer attention block; neither couples its attention mechanism to a downstream learner-state estimator or a distributed multi-source feature pipeline. Table 6 attributes 3.5 accuracy points to the attention mechanism itself and 3.1 points to the learner-state feedback loop — components that AKT and simpleKT do not implement — consistent with BDEN's attention operating over an already feature-rich bidirectional hidden state (Equations 3–5) rather than question embeddings alone. The larger effect size against AKT than against simpleKT ($d = 1.06$ vs. 0.77) fits this reading: simpleKT's deliberately minimal but competitive design leaves less headroom for a purely architectural comparison, whereas AKT's benchmark-scale, single-source design leaves more room for BDEN's multi-source fusion to contribute.

DKVMN[26], the memory-augmented baseline in Table 4, trails BDEN by 7.4 accuracy points (84.0% vs. 91.4%); its static-key, dynamic-value memory structure encodes concept-level mastery without an explicit sequence-attention or bidirectional mechanism, so it cannot reweight distant but diagnostically relevant interactions the way BDEN's attention layer does. Because DKVMN's original evaluation protocol does not report precision, recall, F1, or AUC-ROC on a directly comparable basis, only accuracy is included for it in Table 4; the accuracy gap alone is nonetheless larger than BDEN's margin over any of the three attention- or recurrence-based baselines, suggesting that architectures without an explicit temporal-attention component face a more fundamental representational limit on this task, consistent with prior findings that attention-based knowledge tracing better captures long-distance dependencies in learner interactions than memory- or recurrence-only approaches[9].

Multiple constraints moderate these findings. The evaluation corpus, although significantly larger than those utilized in much of the previous knowledge-tracing literature, is sourced from a restricted group of partner institutions and course topics, and the ability to generalize it to notably different teaching environments, like K-12 or vocational training, has yet to be empirically validated. The performance metrics presented here, in line with standard practices for evaluating early-stage systems, should be viewed as reflections of the architecture's comparative behaviour instead of definitive production benchmarks, and independent validation on other institutional datasets is a crucial subsequent step. Furthermore, the present architecture does not currently include formal privacy-preserving computation techniques, such as differential privacy or federated learning, which will be essential for implementation across organizations that are unwilling or unable to centralize raw learner data.

## 4 Conclusion

This research introduced the Big Data Education Network (BDEN), a flexible intelligent tutoring system that combines distributed big-data analytics with an attention-augmented Bidirectional Long Short-Term Memory (BiLSTM) model to facilitate personalized learning analytics, predict learner mastery, and provide adaptive educational suggestions. The proposed framework effectively handles diverse educational data via a distributed Kafka–Spark setup and utilizes deep knowledge tracing to represent the changing knowledge states of learners. Experimental assessment on a multi-institutional dataset consisting of 4.2 million student interactions, 18,600 students, and 1,240 knowledge concepts showed that BDEN consistently surpassed six benchmark models, reaching 91.4% accuracy, 91.0% precision, 90.2% recall, 90.6% F1-score, an AUC-ROC of 0.947, and an

RMSE of 0.189. In comparison to the most robust baseline model (simpleKT), BDEN enhanced prediction accuracy by 2.8 percentage points and decreased RMSE from 0.238 to 0.189.

Statistical analysis additionally verified that these enhancements were notable ($p < 0.001$), and scalability tests indicated that the distributed system handled around 74,600 learner interaction records each second, showcasing its ability to facilitate extensive educational settings. The main innovation of this study is in the cohesive combination of distributed Kafka–Spark big-data processing, multi-source educational feature integration, attention-boosted BiLSTM knowledge tracking, learner-state feedback, and prerequisite-aware recommendations within a comprehensive end-to-end intelligent tutoring system. In contrast to numerous existing intelligent tutoring systems that mainly emphasize predictive modelling with benchmark datasets, BDEN integrates scalable data engineering, statistically confirmed learning analytics, and adaptive recommendations to offer an effective framework for implementation in actual educational environments.

The proposed framework demonstrates multiple significant advantages. It proficiently manages extensive heterogeneous educational data while ensuring excellent predictive accuracy, strong statistical performance, and effective computational scalability. The thorough assessment, which encompasses comparative benchmarking, statistical significance testing, ablation analysis, and scalability evaluation, illustrates the strength and real-world applicability of the proposed framework for contemporary intelligent tutoring systems. Even with these encouraging outcomes, the research has multiple constraints.

The experimental assessment was carried out with data from a small set of partnering institutions, potentially influencing the applicability of the results to various educational environments. The existing framework mainly depends on structured data from learner interactions and fails to include multimodal educational materials like video, audio, or physiological learning signals. Moreover, the current implementation did not consider explainable artificial intelligence methods and privacy-preserving learning approaches, such as federated learning and differential privacy. Multiple open research queries exist. Future research should explore how BDEN functions in various educational systems, multilingual learning contexts, and extensive online learning platforms. The combination of multimodal learner behavior, retrieval-enhanced educational intelligence, and large language models offers chances to enhance personalization and adaptive feedback while ensuring transparency, fairness, and learner confidentiality.

Future efforts will concentrate on confirming BDEN with larger cross-institutional datasets, integrating explainable and privacy-preserving AI methods, expanding the framework to facilitate multimodal educational analytics, and assessing its long-term effects on student engagement, knowledge retention, and academic success. These advancements are anticipated to enhance BDEN as a scalable, strong, and effective intelligent tutoring framework that can facilitate future-oriented, data-driven digital education systems.

## 5 Acknowledgements

This research received no specific grant or financial support from any funding agency in the public, commercial, or not-for-profit sectors. The author declares no conflict of interest.

*Author details*: T. Nikil Prakash and A. Natarajan serve as Assistant Professors

## References

1) Samsul SA, Yahaya N, Abuhassna H. Education big data and learning analytics: A bibliometric analysis. *Humanities and Social Sciences Communications*. 2023;10:709. https://doi.org/10.1057/s41599-023-02176-x.
2) Stojanov A, Daniel BK. A decade of research into the application of big data and analytics in higher education: A systematic review of the literature. *Education and Information Technologies*. 2023;29:5807–5831. https://doi.org/10.1007/s10639-023-12033-8.
3) Yousufi AA, Naidu VR, Jesrani K, Dattana V. Tracking students' progress using big data analytics to enhance students' employability: A review. *SHS Web of Conferences*. 2023;156:07001. https://doi.org/10.1051/shsconf/202315607001.
4) Abdelrahman G, Wang Q, Nunes B. Knowledge tracing: A survey. *ACM Computing Surveys*. 2023;55(11):1–37. https://doi.org/10.1145/3569576.
5) Shen S, Liu Q, Huang Z, Zheng Y, Yin M, Wang M, et al. A survey of knowledge tracing: Models, variants, and applications. *IEEE Transactions on Learning Technologies*. 2024;17:1858–1882. https://doi.org/10.1109/TLT.2024.3383325.
6) Liu Z, Liu Q, Chen J, Huang S, Luo W. simpleKT: A simple but tough-to-beat baseline for knowledge tracing. *Proceedings of the Eleventh International Conference on Learning Representations*. 2023. https://doi.org/10.48550/arXiv.2302.06881.
7) Qiu L, Zhu M, Zhou J. OPKT: Enhancing knowledge tracing with optimized pretraining mechanisms in intelligent tutoring. *IEEE Transactions on Learning Technologies*. 2024;17:841–855. https://doi.org/10.1109/TLT.2023.3336240.
8) Huang C, Wei H, Huang Q, Jiang F, Han Z, Huang X. Learning consistent representations with temporal and causal enhancement for knowledge tracing. *Expert Systems with Applications*. 2024;245:123128. https://doi.org/10.1016/j.eswa.2023.123128.
9) Wang X, Zheng Z, Zhu J, Yu W. What is wrong with deep knowledge tracing? Attention-based knowledge tracing. *Applied Intelligence*. 2023;53(3):2850–2861. https://doi.org/10.1007/s10489-022-03621-1.
10) Song X, Li J, Cai T, Yang S, Yang T, Liu C. A survey on deep learning-based knowledge tracing. *Knowledge-Based Systems*. 2022;258:110036. https://doi.org/10.1016/j.knosys.2022.110036.

11) Chen J, Liu Z, Huang S, Liu Q, Luo W. Improving interpretability of deep sequential knowledge tracing models with question-centric cognitive representations. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2023;37(12):14196–14204. https://doi.org/10.1609/aaai.v37i12.26661.
12) Liu Z, Liu Q, Chen J, Huang S, Gao B, Luo W, et al. Enhancing deep knowledge tracing with auxiliary tasks. *Proceedings of the ACM Web Conference 2023*. 2023;p. 4178–4187. https://doi.org/10.1145/3543507.3583866.
13) Zhao W, Xia J, Jiang X, He T. A novel framework for deep knowledge tracing via gating-controlled forgetting and learning mechanisms. *Information Processing & Management*. 2023;60:103114. https://doi.org/10.1016/j.ipm.2022.103114.
14) Pu Y, Wu W, Peng T, Liu F, Liang Y, Yu X, et al. Embedding cognitive framework with self-attention for interpretable knowledge tracing. *Scientific Reports*. 2022;12(1):17536. https://doi.org/10.1038/s41598-022-22539-9.
15) Ding X, Larson EC. On the interpretability of deep learning-based models for knowledge tracing. *arXiv preprint*. 2021. https://doi.org/10.48550/arXiv.2101.11335.
16) Shen X, Yu F, Liu Y, Ke Z, Wang C, Yao B. Revisiting knowledge tracing: A simple and powerful model. *Proceedings of the 32nd ACM International Conference on Multimedia*. 2024;p. 263–272. https://doi.org/10.1145/3664647.3681205.
17) Hadbi AE, Rziki MH, Jamil Y, Belkadi R, Bourray H, Ouadghiri MDE. Leveraging big data in intelligent tutoring systems: Enhancing personalized learning experiences. *International Conference on Advanced Intelligent Systems for Sustainable Development (AI2SD 2024), Lecture Notes in Networks and Systems*. 2025;1402. https://doi.org/10.1007/978-3-031-91334-1_22.
18) Raji B, Iyer SS, Yaseen MMNM. AI-enabled predictive analytics in education: Enhancing student success and retention through intelligent tutoring systems. *Journal of Digital Educational Technology*. 2026;6(1). https://doi.org/10.29333/jdet/18330.
19) Deng C, Yuan B. Research on an intelligent tutoring system based on automatic construction of multimodal knowledge graphs and retrieval-augmented generation. *Frontiers in Computer Science*. 2026;8:1777749. https://doi.org/10.3389/fcomp.2026.1777749.
20) Parmar M, et al. Intelligent tutoring systems: A review of machine learning and deep learning approaches in education. *Bulletin for Technology and History Journal*. 2026. Available from: https://bthnl.net/wp-content/uploads/2026/04/22-BTH3584.pdf.
21) Long T, Liu Y, Shen J, Zhang W, Yu Y. Tracing knowledge state with individual cognition and acquisition estimation. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2021;p. 173–182. https://doi.org/10.1145/3404835.3462827.
22) Molla-Esparza C, Gómez-Núñez MI, García-García FJ. Applications of learning analytics in the study of academic performance in higher education: A pilot-tested meta-review protocol. *International Journal of Educational Research Open*. 2025;8:100433. https://doi.org/10.1016/j.ijedro.2024.100433.
23) Sun J, Zou R, Liang R, Gao L, Liu S, Li Q, et al. Ensemble knowledge tracing: Modeling interactions in the learning process. *Expert Systems with Applications*. 2022;207:117680. https://doi.org/10.1016/j.eswa.2022.117680.
24) Piech C, Bassen J, Huang J, Ganguli S, Sahami M, Guibas LJ, et al. Deep knowledge tracing. *Advances in Neural Information Processing Systems*. 2015;28:505–513.
25) Ghosh A, Heffernan N, Lan AS. Context-aware attentive knowledge tracing. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020;p. 2330–2339. https://doi.org/10.1145/3394486.3403282.
26) Zhang J, Shi X, King I, Yeung DY. Dynamic key-value memory networks for knowledge tracing. *Proceedings of the 26th International Conference on World Wide Web*. 2017;p. 765–774. https://doi.org/10.1145/3038912.3052580.