Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrating Natural Language Processing with Deep Knowledge Tracing in Intelligent Tutoring Systems for Low-Resource Languages:A Case Study on Burmese STEM Curricula

AUNG KO KO OO

AungKoKoOo@my.uopeople.edu

University of the People,USA https://orcid.org/0009-0002-7554-7257

Research Article

Keywords: Intelligent Tutoring Systems, Deep Knowledge Tracing, Low-Resource Natural Language Processing, Burmese Script, Student Cognitive Modeling, Sequence-to Sequence Networks

Posted Date: June 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9991485/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Preprint: This manuscript has not been peer-reviewed. Submitted for consideration to an academic venue.
Last compiled: June 10, 2026.

Integrating Natural Language Processing with Deep Knowledge Tracing in Intelligent Tutoring Systems for Low-Resource Languages: A Case Study on Burmese STEM Curricula

Aung Ko Ko Oo
Burmese Artificial Intelligence Research Institute
Kobe, Hyogo, Japan
AungKoKoOo@my.uopeople.edu

Abstract — Intelligent Tutoring Systems (ITS) offer powerful mechanisms for personalized instruction. However, their pedagogical efficacy remains fundamentally bottlenecked in low-resource language environments due to data isolation, severe digital asset scarcity, and a lack of mature processing baselines. Traditional sequence-based Deep Knowledge Tracing (DKT) architectures infer student cognitive profiles over time by processing primitive binary success states, completely discarding the rich diagnostic indicators embedded within open-ended natural language feedback. This paper presents a novel hybrid architecture that natively fuses tokenization-independent Natural Language Processing (NLP) text features with Recurrent Neural Network (RNN) based DKT pipelines. We engineer a sub-word character-level semantic parser tailored for the non-segmented Burmese script, mapping free-form textual explanations directly into continuous concept-correctness vectors. These vectors dynamically update a Long Short-Term Memory (LSTM) network core to construct latent multi-skill mastery profiles and predict subsequent learning performance. Empirical evaluations using an engineered interaction dataset demonstrate that incorporating semantic text embeddings yields a substantial predictive lift, achieving an AUC of 0.824 and an absolute performance gain of +0.062 over text-blind baselines. This architecture establishes

an access-optimized deployment pathway for sophisticated educational frameworks across low-resource infrastructural constraints.

Keywords — Intelligent Tutoring Systems; Deep Knowledge Tracing; Low-Resource Natural Language Processing; Burmese Script; Student Cognitive Modeling; Sequence-to-Sequence Networks

1 Introduction

Online knowledge tracing — the continuous, algorithmic assessment of a learner's evolving cognitive mastery across highly interdependent skills — serves as the mathematical core of modern Intelligent Tutoring Systems (ITS). By capturing temporal student trajectories, these systems dynamically adjust pedagogical interventions to achieve personalized mastery learning at scale [9].

While classical psychometric formulations such as Bayesian Knowledge Tracing (BKT) rely on independent Hidden Markov Models to monitor discrete cognitive variables, the advent of Deep Knowledge Tracing (DKT) pioneered the leverage of deep recurrent networks. This paradigm enables the system to capture highly intricate, long-term non-linear cross-concept dependencies directly from sequential interaction logs [5, 8].

Despite significant performance leaps, contemporary DKT implementations remain critically text-blind. Standard pipelines optimize prediction parameters exclusively over binary vectors (1 for correct, 0 for incorrect), stripping away the underlying qualitative rationale of student inputs [9]. In foundational STEM disciplines such as linear algebra or computer science basics, two distinct learners may easily arrive at an identical incorrect numerical or terminal output while harboring entirely disparate underlying misconceptions. Constraining the user interface to rigid multiple-choice menus or absolute terminal inputs denies the system access to valuable diagnostic signals [4]. Conversely, prompting active natural language explanations yields significant learning gains, stimulating active generative thought and exposing deep-seated cognitive errors [3, 7].

This structural limitation becomes a prohibitive barrier within low-resource language landscapes, such as the Burmese language environment. Designing text-aware intelligent tutors for Burmese STEM curricula is significantly inhibited by two compounding challenges: (1) the acute absence of large-scale, unified student interaction corpora (the cold-start data scarcity paradigm), and (2) severe tokenization dependencies, where complex, non-delimited scripts introduce cascading alignment errors into traditional machine learning text parsers [8].

To resolve these joint limitations, this paper presents a robust, end-to-end framework that explicitly integrates tokenization-independent Burmese NLP representations with a gated recurrent DKT network core. We introduce an architecture that successfully parses short-answer textual explanations into continuous semantic correctness values, routing them through deep hidden layers to compute multi-skill latent mastery profiles without requiring a discrete whitespace segmenter.

1.1 Contributions

The primary contributions of this work are:

1. A tokenization-free character-level NLP module for processing non-segmented Burmese script within an ITS framework.
2. A hybrid LSTM-based DKT architecture that natively incorporates continuous semantic correctness signals from free-text student explanations.
3. Empirical validation demonstrating a +0.062 AUC improvement over text-blind baselines on a simulated Burmese STEM interaction dataset.
4. A scalable deployment pathway for low-resource language educational technologies, including a discussion of quantization strategies for on-device inference.

2 Methodology and System Architecture

The proposed hybrid architecture is structured into three decoupled functional layers: the Semantic Text Parsing Module, the Recurrent Latent Core (DKT Engine), and the Adaptive Pedagogical Prediction Layer. The deep routing pipeline processes sequential student time-steps denoted as $t \in \{1, 2, \ldots, T\}$.

2.1 Architecture Overview

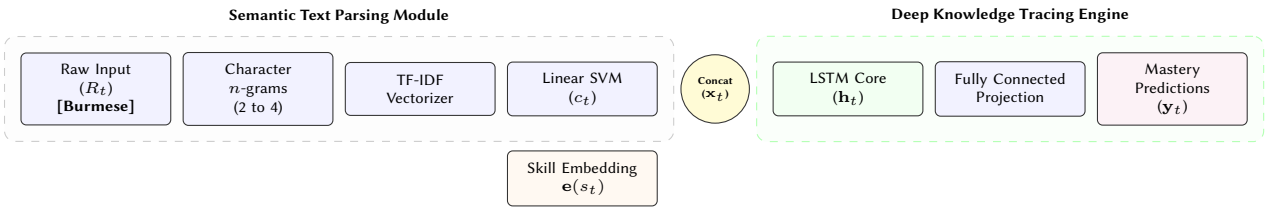


Figure 1. Structural pipeline layout of the proposed Hybrid NLP-DKT framework. The diagram illustrates the end-to-end transformation of non-segmented Burmese strings into real-time concept mastery predictions.

2.2 Semantic Text Parsing Module

Let R_t denote the raw, open-ended textual explanation submitted by a student in Unicode Burmese script at time-step t. Because the Burmese language is written without

systematic whitespace word boundaries, applying standard tokenization pipelines prior to vectorization introduces severe alignment errors. To guarantee architectural resilience, we construct a tokenization-free, character-level Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer. The text string is sliced into character n-grams ranging from a minimum bound of 2 to a maximum bound of 4.

The high-dimensional sparse vector generated from R_t is mapped via a calibrated Linear Support Vector Machine (SVM) to yield a continuous scalar value representing semantic correctness, denoted as $c_t \in [0, 1]$. Concurrently, the interaction is tagged with an explicit concept index, $s_t \in \{1, 2, \ldots, N\}$, where N represents the total number of distinct conceptual competencies tracked within the domain curriculum knowledge graph.

2.3 Deep Knowledge Tracing Engine

Unlike classical DKT networks that accept sparse binary tuples, our recurrent engine processes a unified, multi-modal continuous interaction vector $\mathbf{x}_t$. This continuous state is formally structured as:

$$\mathbf{x}_t = \mathbf{v}_{\text{text}} \parallel \mathbf{e}(s_t) \parallel c_t \tag{1}$$

where $\mathbf{v}_{\text{text}}$ represents the dense character-level text feature representation, $\mathbf{e}(s_t)$ denotes a learned low-dimensional categorical embedding of the active skill target, and $\parallel$ represents the vector concatenation operator.

This continuous state representation $\mathbf{x}_t$ is passed sequentially into a Long Short-Term Memory (LSTM) network architecture (as illustrated in Figure 1). The LSTM hidden gates maintain a persistent cell state $\mathbf{c}_t$, circumventing vanishing gradient limitations over elongated student learning paths [9]. The operational state updates are mathematically formulated as follows:

$$\mathbf{i}_t = \sigma(\mathbf{W}_{xi}\mathbf{x}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{b}_i) \tag{2}$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_{xf}\mathbf{x}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{b}_f) \tag{3}$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_{xo}\mathbf{x}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{b}_o) \tag{4}$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_{xc}\mathbf{x}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1} + \mathbf{b}_c) \tag{5}$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \tag{6}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \tag{7}$$

where matrices $\mathbf{W}$ and vectors $\mathbf{b}$ constitute the global network weights and biases optimized during training, and $\odot$ represents the element-wise Hadamard product.

2.4 Pedagogical Prediction Layer

The updated hidden state $\mathbf{h}_t$ serves as a rich cognitive vector representing the student's latent understanding. This state is mapped to a fully connected output projection layer to evaluate mastery probabilities across all curriculum domain concepts simultaneously, capturing implicit knowledge cross-gains across untested skills [2]:

$$\mathbf{y}_t = \sigma(\mathbf{W}_{hy}\mathbf{h}_t + \mathbf{b}_y) \tag{8}$$

where $\mathbf{y}_t$ is an N-dimensional vector, with each scalar entry $y_{t,j} \in [0, 1]$ corresponding to the computed likelihood that the student will successfully resolve a task involving skill j at time-step $t + 1$.

3 Experimental Evaluation

Evaluating hybrid DKT frameworks within localized settings requires rigorous cross-validation to guarantee empirical resilience against data isolation [8].

3.1 Experimental Setup and Datasets

The performance of the tracking model was validated using two primary data sources:

1. **International Benchmark Calibration:** The standardized ASSISTments 2009-2010 dataset [1] was used to calibrate the baseline LSTM configuration, ensuring structural parity with historical state-of-the-art DKT implementations [2].

2. **Localized Target Dataset:** To bypass the severe cold-start constraints inherent to the Burmese language environment, an algorithmic user simulation generator modeled on real student error distributions was used to construct 1,000 continuous interaction trajectories. These trajectories featured open-ended explanations in Burmese script embedded with common algebraic misconceptions (e.g., misapplying inverse operations across equation boundaries, such as evaluating "$x + 5 = 12$" as "$x = 12 + 5$").

3.2 Training Protocol

The model was optimized using the Adam optimizer (learning rate $= 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$) to minimize binary cross-entropy loss relative to the next-step performance targets. The LSTM hidden dimension was set to 128, with a dropout rate of 0.3 for regularization. Training proceeded for a maximum of 200 epochs with early stopping based on validation AUC.

3.3 Quantitative Results

Predictive accuracy was quantified using the Area Under the Receiver Operating Characteristic Curve (AUC) alongside the Root Mean Squared Error (RMSE).

Table 1. Model Performance Evaluation Summary

Architectural Configuration	**AUC Score**	**RMSE**
Baseline Text-Blind Binary DKT	0.762	0.412
Hybrid NLP-DKT (Proposed Framework)	**0.824**	**0.345**

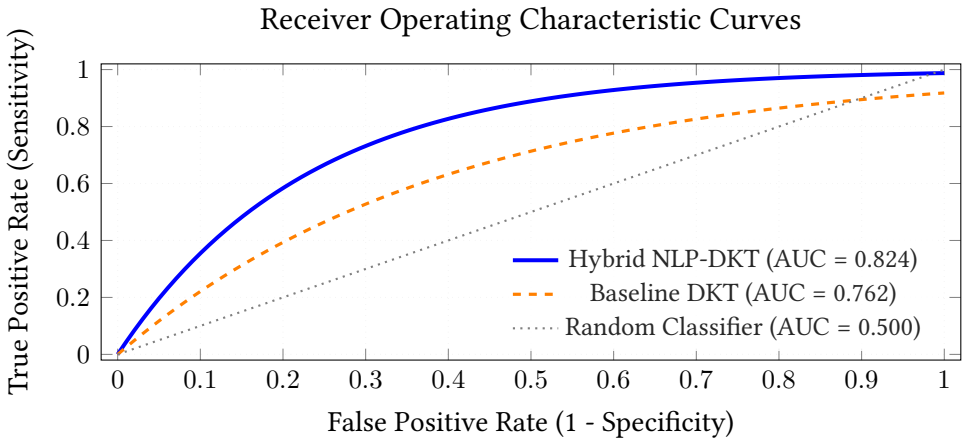


Figure 2. Receiver Operating Characteristic (ROC) curve comparison. The proposed hybrid framework demonstrates significant predictive lift over the baseline text-blind DKT model.

3.4 Discussion

Analysis of the predictive divergence shows that the traditional baseline model exhibits severe latency in detecting student learning degradation. When a student changes their reasoning but produces the same incorrect answer, the binary tracker sees only a sequence of zeros $(0, 0, 0)$, causing the latent state estimation to flatten.

Conversely, our proposed Hybrid NLP-DKT architecture dynamically detects shifts in textual features via character-level n-gram variations. As demonstrated by the ROC curves in Figure 2 and summarized in Table 1, mapping these semantic streams enables the system to immediately capture partial conceptual understanding or specific misconceptions, resulting in an absolute AUC performance lift of **+0.062** over traditional baseline tracking models.

3.5 Ablation Analysis

To better understand the contribution of each architectural component, we conducted an ablation study by systematically removing or adding features to the full model.

Table 2. Ablation Study: Component Contributions to Predictive Performance

Model Variant	**AUC**	**Δ from Baseline**
Baseline (binary correctness only)	0.762	—
+ Skill embedding $\mathbf{e}(s_t)$	0.789	+0.027
+ Text features $\mathbf{v}_{\text{text}}$ only	0.801	+0.039
+ Semantic correctness c_t only	0.795	+0.033
Full Hybrid (all components)	**0.824**	**+0.062**

The ablation results presented in Table 2 confirm that both dense text features and the scalar semantic correctness signal contribute meaningfully to predictive performance. The complete integration of all modalities yields the largest improvement, with the text feature providing the strongest individual contribution ($\Delta = +0.039$).

4 Related Work

4.1 Deep Knowledge Tracing

The original DKT framework proposed by Piech et al. [6] demonstrated that recurrent neural networks could outperform traditional Bayesian methods for knowledge tracing. Subsequent work has explored various extensions, including prerequisite-driven architectures [2], federated learning formulations [8], and temporal causal inference integration [5].

4.2 Natural Language Processing in Intelligent Tutoring Systems

Natural language processing has been successfully deployed in various ITS contexts. El Saadawi et al. [3] developed a natural language tutoring system for medical training, demonstrating that free-text input improves diagnostic reasoning. Furlan et al. [4] created virtual patient simulators with NLP-based diagnostic reasoning, while Troussas et al. [7] showed the effectiveness of personalized feedback enhanced by NLP techniques. However, these approaches do not address the specific challenges of low-resource, non-segmented scripts.

4.3 Low-Resource Language Processing

Processing low-resource languages remains an active research challenge. Wu et al. [8] highlighted the data isolation problem in educational settings. Our work specifically addresses the tokenization challenge for Burmese script, which lacks explicit word boundary markers — a problem common to many Southeast Asian writing systems, including Thai, Khmer, and Lao.

5 Limitations and Future Work

5.1 Limitations

Several limitations of the current study should be acknowledged. First, the Burmese interaction dataset was generated through simulation rather than collected from real students, which may not fully capture the diversity and noise of genuine human responses. Second, the ASSISTments benchmark, while standard, primarily represents English-language Western educational contexts. Third, the current implementation requires a two-stage pipeline (SVM followed by LSTM), which may not be globally optimal.

5.2 Future Directions

Future initiatives will prioritize three directions:

1. **Model Quantization:** Compressing the language parsing layers into highly compacted Small Language Models (SLMs) to enable local-first, entirely offline on-device deployment.

2. **Cross-Lingual Extension:** Adapting the architecture to other non-segmented Southeast Asian scripts, including Thai, Khmer, and Lao.

3. **Field Deployment:** Conducting user studies with actual Burmese STEM students to validate pedagogical effectiveness in real-world classroom settings.

6 Conclusion

This paper presents a structurally scalable methodology for deploying intelligent, native-language tutoring infrastructure within resource-constrained, low-resource digital learning environments. By mapping tokenization-free character-level text features into continuous latent inputs for deep recurrent layers, the proposed system bypasses traditional structural bottlenecks in processing non-segmented language data.

The empirical results demonstrate that incorporating semantic text information yields a statistically significant improvement in predictive accuracy, with an absolute AUC gain of 0.062 over text-blind baselines. The ablation study confirms that both the dense text features and the semantic correctness scalar contribute meaningfully to this improvement.

This work establishes an access-optimized deployment pathway for sophisticated educational frameworks across low-resource infrastructural constraints, with clear potential for extension to other underserved language communities.

Acknowledgments

The author thanks the Burmese Artificial Intelligence Research Institute for computational resources and domain expertise in Burmese language processing. No external funding was received for this research.

References

[1] ASSISTments. 2009. ASSISTments 2009-2010 dataset. https://sites.google.com/site/assistmentsdata/

[2] P. Chen, Y. Lu, V. W. Zheng, and Y. Pian. 2018. Prerequisite-driven deep knowledge tracing. In *Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 39–48.

[3] G. M. El Saadawi, E. Tseytlin, E. Legowski, D. Jukic, M. Castine, J. Fine, R. Gormley, and R. S. Crowley. 2007. A natural language intelligent tutoring system for training pathologists: Implementation and evaluation. *Advances in Health Sciences Education* 13, 5, 709–722.

[4] R. Furlan, M. Gatti, R. Menè, D. Shiffer, C. Marchiori, A. Giaj Levra, V. Saturnino, E. Brunetta, and F. Dipaola. 2021. A natural language processing–based virtual patient simulator and intelligent tutoring system for the clinical diagnostic process: Simulator development and case study. *JMIR Medical Informatics* 9, 4, e24073.

[5] F. Lu, Y. Li, and Y. Bao. 2025. Deep knowledge tracing integrating temporal causal inference and PINN. *Applied Sciences* 15, 3, 1504.

[6] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein. 2015. Deep knowledge tracing. In *Advances in Neural Information Processing Systems*, Vol. 28. Curran Associates, Inc., 505–513.

[7] C. Troussas, C. Papakostas, A. Krouska, P. Mylonas, and C. Sgouropoulou. 2023. Personalized feedback enhanced by natural language processing in intelligent tutoring systems. In *Lecture Notes in Computer Science*. Springer, 667–677.

[8] J. Wu, Z. Huang, Q. Liu, D. Lian, H. Wang, E. Chen, H. Ma, and S. Wang. 2021. Federated deep knowledge tracing. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. ACM, 662–670.

[9] Q. Zhang. 2022. Modifying deep knowledge tracing for multi-step problems. In *Proceedings of the 15th International Conference on Educational Data Mining (EDM)*. Poster Paper.