

Hybrid Human–Machine Intelligence for Smart Learning: A Multidisciplinary Approach to Context-Aware, Adaptive, and Explainable Educational Systems [†]

Nisreen A. Alzahrani ^{1,*} , Yerragolla Hareesh Kumar ²  and Gopalappa Bindusree ³ 

¹ Department of Information Systems, College of Computer and Information Sciences, Majmaah University, Majmaah 15341, Saudi Arabia

² School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu, Fukushima 965-8580, Japan; yerragollahareeshkumar@gmail.com

³ School of Computer Science and Engineering, Vellore Institute of Technology, Vellore 632014, India; bindusree.g2025@vitstudent.ac.in

* Correspondence: n.alzhrani@mu.edu.sa

[†] Presented at the 8th International Global Conference Series on ICT Integration in Technical Education & Smart Society, Aizuwakamatsu City, Japan, 20–26 January 2026.

Abstract

The growing demand for personalized learning in digital education has encouraged researchers to explore artificial-intelligence-driven recommendation systems. However, most existing systems rely solely on either content-based or collaborative filtering methods, limiting their adaptability and performance in real-world educational environments. Moreover, such approaches often face challenges related to data sparsity, cold-start issues, and limited explainability, reducing their effectiveness in diverse learner contexts. This paper introduces a novel Hybrid Human–Machine Intelligence Framework designed to build context-aware, adaptive, and transparent learning recommendation systems. Using behavioral and demographic data from the Open University Learning Analytics Dataset (OULAD), the framework employs a multi-task deep learning model to jointly predict each learner’s engagement level and preferred learning modality. These predictions serve as the foundation for a weighted hybrid recommendation engine that integrates content-based, collaborative, and knowledge-based filtering to generate highly personalized suggestions. An explainable reasoning module further translates the hybrid scores into clear, pedagogically meaningful justifications to enhance trust and instructional relevance. Experimental results demonstrate that the proposed model effectively mitigates cold-start issues and outperforms traditional machine learning baselines, achieving an RMSE of approximately 0.09, an MAE of 0.05, and an R2 value close to 0.99. The findings highlight the potential of hybrid human–machine intelligence to bridge cognitive understanding and computational learning, thereby advancing the design of adaptive and interpretable smart education systems.

Keywords: hybrid recommendation systems; personalized learning; explainable artificial intelligence; adaptive learning models; learning analytics



Academic Editors: Debopriyo Roy,
George Fragulis and Peter Ilic

Published: 20 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

In today’s world, education has moved beyond classrooms into digital spaces where learners interact with large volumes of online material. With the growth of e-learning platforms, the abundance of resources makes it difficult for learners to identify content suited to their goals, motivating the development of intelligent recommendation systems [1,2].

Recommender systems, widely used in entertainment and retail, can personalize learning and improve engagement by analyzing student behavior and performance [2–4].

However, existing educational recommendation models still face major limitations. Content-based filtering often produces repetitive suggestions, while collaborative filtering suffers from data sparsity and cold-start problems [2,5]. Knowledge-based methods require expert rules that are difficult to scale for large datasets. Most current systems rely on a single strategy and lack deeper behavioral prediction or clear explanation mechanisms, creating a gap in developing adaptive and interpretable recommendation frameworks.

Recent works emphasize the need for hybrid systems that integrate multiple recommendation approaches with human reasoning. Studies show that combining algorithmic prediction with human feedback increases reliability and transparency in adaptive learning environments [6,7]. Building on these ideas, this work introduces a Hybrid Human–Machine Intelligence Framework that merges data-driven behavioral modeling with rule-based reasoning to generate personalized, context-aware, and explainable recommendations.

Using demographic and interaction data from the Open University Learning Analytics Dataset (OULAD), the proposed system first predicts learner engagement and preferred learning modality using a multi-task deep learning model. These predictions form the basis for a hybrid recommendation engine that blends content-based, collaborative, and knowledge-based filtering through adaptive weighting. An explainable reasoning module then converts internal scores into clear justifications to help educators and learners understand why a resource was recommended.

The goal of this framework is to deliver an educational recommendation system that is adaptive, interpretable, and aligned with learner behavior. Experimental results demonstrate improved accuracy, reduced cold-start issues, and more meaningful recommendations compared with traditional approaches, supporting the development of intelligent, learner-centered digital learning environments.

2. Related Works

Over the last decade, many studies have explored the development of intelligent recommendation systems to improve learning experiences and deliver personalized educational content. Researchers have aimed to design adaptive systems that can analyze student behavior, preferences, and performance to provide the most relevant materials. Early works such as those by Tang and McCalla [8] introduced intelligent tutoring and recommendation systems that could adapt to learners' evolving needs by tracking their interaction patterns and suggesting suitable learning resources. Similarly, Drachsler and Verbert [2] highlighted the role of recommender systems in supporting learning by emphasizing the balance between system automation and learner control. These studies established the foundation for data-driven personalization in education, which continues to influence recent hybrid learning architectures.

Traditional recommendation systems, including content-based and collaborative filtering approaches, have been successfully applied to e-learning platforms but still face notable challenges. Herlocker et al. [9] discussed how collaborative filtering models often struggle with the cold-start problem, while Bobadilla et al. [3] emphasized that content-based models suffer from over-specialization, repeatedly recommending similar materials. Manouselis and Costopoulou [10] introduced multi-criteria recommendation frameworks to improve accuracy by incorporating contextual and learner-specific attributes. However, these systems still rely heavily on static data and fail to capture the complex, evolving context of real learning environments. These limitations motivated the emergence of hybrid systems that integrate multiple recommendation strategies.

Hybrid and context-aware systems combine the strengths of different algorithms and adjust to the learner's context dynamically. Ricci et al. [11] proposed an architecture that uses both user similarity and resource features to enhance recommendation precision. Lu et al. [6] later extended this concept through a multi-model adaptive framework that selects algorithms according to data availability and quality. Li and Zhao [1] demonstrated that including human feedback within algorithmic reasoning can significantly improve the trust and reliability of adaptive learning systems. These works collectively indicate that hybrid frameworks offer a more flexible and scalable solution for personalized education, aligning with the direction of the system proposed in this study.

In parallel, explainable artificial intelligence has gained attention as an essential factor in educational recommender systems. Lu and Zhang [8] introduced a hybrid explainable model combining knowledge graphs and deep learning to generate transparent and interpretable recommendations for students and educators. Verbert et al. [7] focused on the visualization of learning data through dashboards, helping users understand how recommendations are produced. Drachsler and Greller [12] emphasized the importance of privacy, transparency, and ethical responsibility in learning analytics. Chen et al. [13] developed a feedback-driven explainable approach where learner satisfaction actively refines system recommendations. These studies underline the growing need for interpretability and human involvement in AI-driven education systems.

More recently, researchers have begun to incorporate human-in-the-loop approaches to balance machine learning automation with human expertise. Khamparia and Pandey [14] presented a hybrid human–AI tutoring framework in which teacher feedback refines machine recommendations in real time. Kumar et al. [15] proposed an adaptive recommendation model that integrates machine predictions with teacher evaluations to reduce algorithmic bias and maintain pedagogical relevance. These works collectively highlight that human feedback enhances explainability and ensures that machine intelligence aligns with educational objectives. Building on this foundation, the present study introduces a Hybrid Human–Machine Intelligence Framework that unifies computational and human reasoning to produce adaptive, transparent, and context-aware learning recommendations.

3. Proposed Framework

The proposed framework is developed and evaluated using the Open University Learning Analytics Dataset (OULAD), a large-scale educational dataset widely used in learning analytics research. OULAD contains demographic, academic, and detailed click-stream records for 32,593 students across 22 module–presentation offerings. It provides rich behavioral data reflecting how learners interact with activities such as discussion forums, quizzes, reading materials, and supplementary resources in the Virtual Learning Environment (VLE).

This dataset is suitable for multi-task modeling because it captures both the background characteristics of each learner and their engagement patterns across multiple resource types. These attributes collectively form the foundation for constructing structured learner representations and training the multi-task behavioral prediction model described in the following section.

3.1. Student Data Modeling and Pre-Processing

The first stage of the proposed framework focuses on building structured learner representations using the Open University Learning Analytics Dataset (OULAD). This large-scale dataset comprises records from 32,593 registered students enrolled across 22 distinct module–presentation combinations, providing a rich basis for modeling learner behavior and performance within the Virtual Learning Environment (VLE). It captures comprehen-

sive academic, demographic, and behavioral information for each student mathematically represented as a feature vector

$$X_i = [d_{i,1}, \dots, d_{i,p}, c_{i,1}, \dots, c_{i,k}], \quad (1)$$

where $d_{i,j}$ denotes demographic attributes (e.g., gender, region, qualification, age range), and $c_{i,j}$ represents behavioral interaction features extracted from clickstream logs.

Each learner is represented as a structured feature vector defined in Equation (1) and combines static demographic features such as course module, presentation term, gender, region, highest educational qualification, socio-economic band, age range, number of previous attempts, and total studied credits with dynamic behavioral features derived from VLE clickstream activity. These behavioral variables record how frequently students interact with digital learning materials, including discussion forums, quizzes, web links, learning resources, and supplementary pages, thereby reflecting both their engagement intensity and learning behavior. To determine each learner's preferred learning modality, the raw OULAD activity types (e.g., oucontent, forumng, quiz, subpage, oucollaborate) were first mapped to human-readable modality categories such as Course Reading Material, Discussion Forum, Interactive Quiz, Supplementary Page, and Collaborative Workspace. For each student, the total number of clicks associated with each modality category was computed from the VLE interaction logs. The modality with the highest cumulative click count was then assigned as the learner's preferred learning modality. This maximum interaction modality strategy is widely adopted in learning analytics to infer dominant behavioral preferences from clickstream data.

To compute the learner's engagement level, the overall interaction intensity of each student was quantified using the cumulative VLE activity across all resource categories. For a given student s_i , the total engagement score was calculated as

$$E_i = \sum_{j=1}^k C_{i,j}, \quad (2)$$

The overall learner engagement is computed using Equation (2). Where $C_{i,j}$ denotes the raw click count of student i for the j -th learning resource type. Based on the raw engagement score E_i , learners were classified into five discrete engagement categories using fixed rule-based thresholds: No Engagement ($E_i = 0$), Low ($0 < E_i \leq 50$), Moderate ($50 < E_i \leq 200$), High ($200 < E_i \leq 600$), and Very High ($E_i > 600$). This rule-based discretization exactly follows the engagement labeling implemented in the preprocessing pipeline.

A representative portion of the raw dataset showing demographic attributes and sample interaction counts is presented in Table 1. The aggregated modality-wise click features used for this computation are illustrated in Table 2.

Table 1. Sample student demographic attributes with representative VLE interaction counts from the OULAD.

Code Module	Presentation	Student ID	Gender	Region	Education	Forum Clicks
AAA	2013J	11391	M	East Anglian	HE Qualification	193
AAA	2013J	28400	F	Scotland	HE Qualification	417
AAA	2013J	30268	F	North West	A Level	126
AAA	2013J	31604	F	South East	A Level	634

Table 2. Aggregated clickstream interaction features across multiple learning resource types from the OULAD.

Student ID	CRM	DTR	Forum	WebLink	Glossary	Quiz	LR File	Webinar	Shared	Supp Page
11391	553	0	193	5	0	0	13	0	0	32
28400	537	10	417	48	0	0	12	0	0	87
30268	66	0	126	4	0	0	4	0	0	22
31604	836	2	634	90	1	0	19	0	0	144
32885	494	0	194	14	4	0	45	0	0	79

To ensure analytical consistency and prepare the data for modeling, several preprocessing operations are applied. First, missing numerical values are replaced with the mean of the corresponding feature, while missing categorical fields are filled using the mode (most frequent category) to preserve the statistical distribution of the dataset. Continuous attributes, including studied credits and VLE click counts, are normalized to the range [0, 1] using the min–max transformation:

$$\tilde{X}_{i,j} = \frac{X_{i,j} - \min(X_j)}{\max(X_j) - \min(X_j)}, \quad (3)$$

Continuous features are normalized using the min–max transformation defined in Equation (3). Where X_j denotes the set of all values for feature j across all learners. This normalization prevents attributes with larger numeric scales from dominating the learning process and ensures uniform contribution of all input features.

Categorical variables such as gender, region, highest education, age band, disability status, and final result are encoded numerically using label encoding. This approach was selected instead of one-hot encoding to avoid excessive dimensionality resulting from multi-class expansion while maintaining computational efficiency for deep learning.

All demographic and behavioral attributes are then merged using the unique identifier `id_student` to form a unified processed dataset:

$$D = \{(X_i, Y_i)\}_{i=1}^N, \quad (4)$$

The processed dataset is formally defined in Equation (4). where the target vector $Y_i = (y_i^{mod}, y_i^{eng})$ contains the learning modality label and engagement category for student i . The modality label is derived using the maximum interaction modality rule described above, while the engagement label represents the learner's cumulative activity level, ranging from No Engagement, Low, Moderate, High, to Very High.

After preprocessing, the raw OULAD data are transformed into a structured and numerically consistent representation suitable for multi-task deep learning. Each learner is represented by a single feature vector containing encoded demographic fields and normalized behavioral values. Behavioral features such as forum, quiz, web link, and supplementary page interactions are included as normalized components of X_i , with their scaled values computed using

$$\tilde{C}_{i,j} = \frac{C_{i,j} - \min(C_j)}{\max(C_j) - \min(C_j)}, \quad (5)$$

Behavioral click features are normalized using the transformation defined in Equation (5). Where $C_{i,j}$ is the raw click count for student i in resource category j , and C_j is the set of click counts for that feature across all students.

This structured and pre-processed representation captures each student's behavioral and academic profile in a balanced numerical format. It forms the foundational input for the multi-task deep learning-based behavioral prediction model described in the next stage of the framework.

3.2. Multi-Task Deep Learning Model for Behavioral Prediction

The second stage of the proposed framework employs a multi-task deep learning model designed to jointly predict two key behavioral outcomes for each learner: the engagement level and the preferred learning modality. This approach allows the system to learn the intrinsic relationship between how frequently students interact within the Virtual Learning Environment (VLE) and the type of learning activity they most engage with. By sharing representations across these related tasks, the model improves predictive performance and generalization by enabling both tasks to benefit from common behavioral patterns.

Let the processed dataset be represented as

$$D = \left\{ \left(X_i, y_i^{mod}, y_i^{eng} \right) \right\}_{i=1}^N, \quad (6)$$

The multi-task learning dataset is formally represented in Equation (6). Where $X_i \in \mathbb{R}^d$ is the d -dimensional feature vector for student s_i . Both y_i^{mod} and y_i^{eng} are expressed as one-hot encoded vectors, where the modality label represents categories such as reading material, forum, quiz, or collaborative content, and the engagement label denotes levels ranging from very low to very high. The overall predictive mapping function is defined as

$$F_{\theta}(X_i) = \left(\hat{y}_i^{mod}, \hat{y}_i^{eng} \right), \quad (7)$$

The predictive mapping function of the multi-task model is defined in Equation (7). Where F_{θ} is the neural network parameterized by weights θ , and each $\hat{y}_i$ is a probability distribution obtained from the softmax output.

The model architecture consists of two shared hidden layers that capture generalized behavioral patterns, followed by two task-specific output branches responsible for predicting the learning modality and engagement level, respectively. Formally,

$$\begin{aligned} h_1 &= \text{ReLU}(W_1 X_i + b_1), \\ h_2 &= \text{Dropout}(0.3, \text{ReLU}(W_2 h_1 + b_2)), \\ \hat{y}_i^{mod} &= \text{Softmax}(W_m h_2 + b_m), \\ \hat{y}_i^{eng} &= \text{Softmax}(W_e h_2 + b_e), \end{aligned} \quad (8)$$

The shared and task-specific layers of the multi-task network are formally defined in Equation (8). Where W and b denote the trainable weight and bias matrices. The *ReLU* activation introduces non-linearity for richer feature learning, while dropout with a probability of 0.3 randomly deactivates neurons during training to reduce overfitting. Each softmax layer outputs a probability vector, and the predicted class corresponds to the index with the highest probability.

The network is trained using a composite loss function that combines the categorical cross-entropy losses of both prediction tasks:

$$\mathcal{L} = \alpha \mathcal{L}_{mod} + \beta \mathcal{L}_{eng}, \quad (9)$$

where

$$\mathcal{L}_{mod} = -\sum_i y_i^{mod} \log(\hat{y}_i^{mod}), \quad \mathcal{L}_{eng} = -\sum_i y_i^{eng} \log(\hat{y}_i^{eng}). \quad (10)$$

The overall training objective is defined in Equation (9), where the task-specific categorical cross-entropy losses are given in Equation (10). Equal task weights ($\alpha = \beta = 1$) are used because both tasks are pedagogically important and exhibit similar class distributions. Model parameters are optimized using the Adam algorithm with adaptive learning rates.

For training and evaluation, the processed OULAD is divided into training and testing subsets with an 80–20 ratio, maintaining class balance through stratified sampling. Batch normalization is applied to stabilize training, and early stopping is employed to prevent overfitting. Performance is assessed using accuracy, precision, recall, and Root Mean Square Error (RMSE). RMSE is computed using the probability outputs of the softmax layer against the one-hot encoded labels:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\hat{y}_i - y_i\|^2} \quad (11)$$

Model performance is evaluated using the Root Mean Square Error defined in Equation (11). The model achieved 99.46% accuracy for modality prediction and 99.15% for engagement classification, with an RMSE of 0.0919, indicating highly consistent and reliable predictions across unseen data.

Table 3 presents sample prediction outcomes for selected students, where the predicted labels correspond to the decoded class names obtained from the softmax output.

Table 3. Predicted learning modality and engagement level derived from the multi-task deep learning model.

Student ID	Predicted Learning Modality	Predicted Engagement Level
11391	Course Reading Material	High
28400	Collaborative Workspace	Moderate
30268	Discussion Forum	Low
31604	Course Reading Material	Very High

The results clearly indicate that the multi-task deep learning model successfully learns the underlying behavioral relationships among students. By jointly predicting engagement and learning modality, the model provides a coherent understanding of how learners interact within the virtual classroom. These predicted behavioral attributes form the primary input for the Hybrid Recommendation Engine, which transforms them into personalized and pedagogically relevant learning recommendations.

3.3. Hybrid Learning Recommendation Engine

This stage of the framework converts the predicted behavioral profiles of learners into concrete and personalized learning recommendations. Using the outputs of the multitask deep learning model, the predicted preferred activity type and engagement level, a hybrid learning recommendation engine is developed by integrating three complementary strategies: Content-Based Filtering (CBF), Collaborative Filtering (CF), and Knowledge-Based Filtering (KBF). These components jointly exploit individual learner preferences, historical interaction patterns, and rules to generate adaptive and meaningful recommendations.

A content metadata table, denoted as C , is constructed from the VLE dataset, where each entry corresponds to a learning resource and is represented as

$$C = \{(\text{content_id}_j, \text{type}_j, \text{title}_j, \text{difficulty}_j, \text{duration}_j)\}_{j=1}^J. \quad (12)$$

The content metadata structure is formally defined in Equation (12). The raw activity_type values are first mapped to human-readable labels (e.g., Course Reading Material,

Interactive Quiz, Discussion Forum) using a predefined mapping dictionary. A reverse mapping dictionary is subsequently used to associate each human-readable label with its original activity code, ensuring consistent comparison between predicted learner preferences and content modalities. The predicted learner information is stored in the table `preds_df`, which includes `id_student`, `Predicted_Modality_Human`, and `Predicted_Engagement`.

For each learner s_i and content item r_j , three independent relevance scores are computed.

(a) **Content-Based Filtering (CBF):**

CBF estimates the similarity between the learner's predicted modality and the modality of the content item. The score is defined as

$$S_{CBF}(s_i, r_j) = \begin{cases} 1.0, & \text{if } raw_mod_i = raw_modality_j, \\ 0.7, & \text{if both modalities are semantically related,} \\ 0.3, & \text{otherwise.} \end{cases} \quad (13)$$

The content-based similarity between a learner and a resource is computed using Equation (13). This step prioritizes resources that directly match or closely relate to the learner's predicted preference and is implemented through the `cbf_score` and `compute_cbf` functions.

(b) **Collaborative Filtering (CF):**

Since explicit ratings are unavailable, collaborative filtering is modeled using implicit feedback derived from interaction frequency. The total popularity of each resource is computed as

$$popularity_j = \sum_i \text{sum_click}(s_i, r_j), \quad (14)$$

which is then normalized using min-max scaling:

$$S_{CF}(r_j) = \frac{popularity_j - \min(popularity)}{\max(popularity) - \min(popularity) + \epsilon}, \quad (15)$$

The resource popularity is computed using Equation (14) and subsequently normalized to obtain the collaborative filtering score defined in Equation (15). Where ϵ is a small constant to prevent division by zero. This component promotes content that is widely accessed by the learner community and is implemented using the `compute_cf` function.

(c) **Knowledge-Based Filtering (KBF):**

KBF incorporates rule-based pedagogical constraints to preserve alignment with the learner's predicted modality. The knowledge-based score is assigned as

$$S_{KBF}(s_i, r_j) = \begin{cases} 1.0, & \text{if } raw_modality_j = raw_mod_i, \\ 0.2, & \text{otherwise.} \end{cases} \quad (16)$$

The knowledge-based relevance score is computed according to Equation (16). This mechanism ensures that alternative learning resources remain visible while giving priority to content that matches the learner's dominant learning behavior. It is implemented in the `compute_kbf` function.

(d) **Hybrid Score Fusion and Ranking:**

The final relevance score for each learner-resource pair is obtained by weighted fusion of the three components:

$$S_{Hybrid}(s_i, r_j) = 0.4 \cdot S_{CBF}(s_i, r_j) + 0.4 \cdot S_{CF}(r_j) + 0.2 \cdot S_{KBF}(s_i, r_j). \quad (17)$$

The final recommendation score is computed using the weighted fusion defined in Equation (17). These weights were selected empirically to balance personalization, community trends, and pedagogical consistency. The combined scores are sorted in descending order, and the top- N resources ($N = 3$ in this work) are selected as the final recommendations using the `combine_scores` and `build_recommender` functions.

To efficiently handle large-scale data, recommendations are generated in mini-batches using the `recommend_in_batches` procedure, which produces a comprehensive recommendation table `recs_all` containing, for each learner, recommended content identifiers, titles, estimated duration, and hybrid scores.

3.4. Explainable Recommendation Output

The final stage of the framework is dedicated to generating recommendations that are transparent, interpretable, and pedagogically aligned. While the hybrid recommendation engine determines which learning resources best match a student's predicted behavior, this stage focuses on clearly communicating the reasoning behind those suggestions. The goal is to ensure that both learners and educators understand the evidence supporting each recommendation, reinforcing trust and instructional validity.

To explain how the system derives these recommendations, the reasoning process is grounded in the hybrid score computed for each resource. For a student s_i , each candidate resource r_j receives a score based on a weighted combination of content similarity, collective usage, and instructional logic:

$$S_{\text{Hybrid}}(s_i, r_j) = 0.4 S_{\text{CBF}}(s_i, r_j) + 0.4 S_{\text{CF}}(r_j) + 0.2 S_{\text{KBF}}(s_i, r_j). \quad (18)$$

The final ranking score is computed using Equation (18). This formulation ensures that multiple dimensions of learning behavior contribute to the final ranking. The content-based similarity score $SCBF$ captures how closely the resource aligns with the student's predicted modality. The collaborative score SCF reflects how frequently the resource is accessed by learners with comparable engagement patterns. The knowledge-based score $SKBF$ maintains pedagogical consistency by prioritizing resources appropriate for the learner's study trajectory.

Consider the behavioral profile of student ID 28400 from Table 3. The multi-task model predicts a Collaborative Workspace learning preference and a Moderate engagement level. When the hybrid score is computed for various resources, discussion forums and shared lesson pages obtain the highest values because they closely match the predicted modality and are widely used by students with similar engagement intensity.

The explainable module transforms these internal computations into a clear justification, such as:

“Discussion forums and shared lesson pages are suggested because they align with your collaborative learning preference and are frequently accessed by students who show similar engagement levels.”

This form of explanation links the recommendation directly to the behavioral evidence identified by the model and the scoring functions that drive the ranking. It also provides educators with actionable insight, helping them monitor engagement trends, identify support needs, and refine course materials.

By combining deep learning predictions with transparent rule-based reasoning, the Explainable Recommendation Output stage ensures that the generated recommendations are understandable, trustworthy, and aligned with the principles of learner-centric educational support.

4. Experimental Results and Discussion

This section presents the experimental evaluation of the proposed Multi-Task Hybrid Deep Learning Model (MT-HDLM) in comparison with five widely used baseline algorithms—Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting Machine (GBM), and Extreme Gradient Boosting (XGB). All models were trained and tested using the same preprocessed OULAD to ensure a fair and consistent comparison.

Three evaluation metrics were selected to comprehensively assess predictive performance: Root Mean Square Error (RMSE) for penalizing large deviations, Mean Absolute Error (MAE) for measuring average prediction accuracy, and the Coefficient of Determination (R^2) for quantifying the proportion of variance explained by each model. The comparative metrics are visualized in Figure 1.

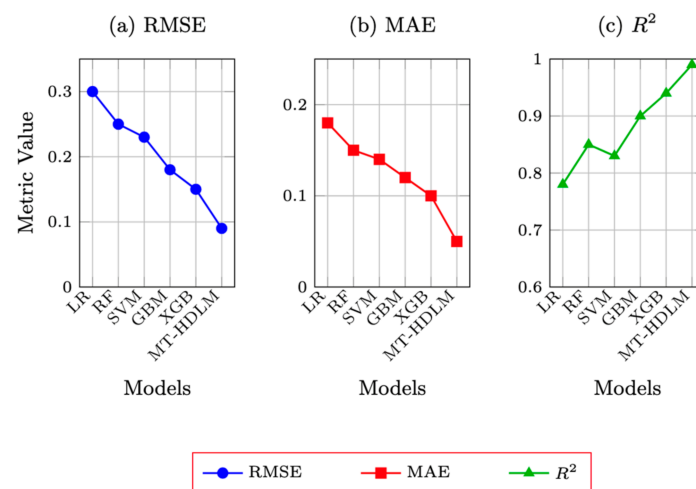


Figure 1. Comparison of RMSE, MAE, and R^2 comparison across models LR, RF, SVM, GBM, XGB, and MT-HDLM, respectively. The blue, red, and green lines represent RMSE, MAE, and R^2 , respectively.

The experimental results demonstrate that the proposed MT-HDLM consistently outperforms all baseline models across all evaluation metrics. It achieved the lowest RMSE (≈ 0.09) and MAE (≈ 0.05), indicating minimal deviation between predicted and actual behavioral outcomes for both learning modality and engagement classification. Additionally, the model obtained the highest R^2 value (≈ 0.99), highlighting its strong ability to capture complex, non-linear, and multi-dimensional patterns in learner behavior. Ensemble-based models such as RF and GBM performed reasonably well but were unable to match the predictive precision of the proposed approach, while linear models like LR showed the weakest generalization capability due to their limited capacity to model non-linearity.

The superior performance of MT-HDLM can be attributed to its multi-task architecture, which jointly optimizes engagement and modality prediction through shared feature representations. This shared-learning structure enables the model to exploit cross-task correlations that single-task baseline models fail to capture. The integration of dropout regularization, adaptive optimization, and careful hyperparameter tuning further stabilized convergence and reduced the risk of overfitting.

These results confirm that combining deep hierarchical representations with multi-task learning significantly enhances behavioral prediction accuracy. Since the recommendation engine relies directly on these predictions, the improved model performance translates

into more reliable, adaptive, and context-aware recommendations within the virtual learning environment.

5. Conclusions and Future Work

This study developed a Hybrid Deep Learning–Based Educational Recommendation Framework using the OULAD. The proposed MT-HDLM model jointly predicts student engagement and preferred learning modalities and integrates these predictions into an explainable hybrid recommendation system. Experimental results show that MT-HDLM outperforms traditional models, achieving lower RMSE and MAE values and a near-perfect R2, demonstrating strong capability in modeling complex learner behavior.

A key strength of the framework is its focus on explainability, ensuring that generated recommendations are transparent and pedagogically meaningful for both learners and instructors. This enhances trust and supports informed decision-making in digital learning environments.

Future work will explore temporal modeling using recurrent or transformer-based networks, integration of additional contextual features such as assessment feedback or interaction patterns, and deployment in real-time adaptive learning systems. Testing the framework across diverse MOOC platforms or institutional datasets will further validate its generalizability and practical impact.

Author Contributions: Conceptualization, N.A.A. and Y.H.K.; methodology, Y.H.K. and G.B.; software, Y.H.K. and G.B.; validation, N.A.A., Y.H.K. and G.B.; formal analysis, Y.H.K. and G.B.; investigation, N.A.A., Y.H.K. and G.B.; resources, N.A.A.; data curation, Y.H.K. and G.B.; writing—original draft preparation, Y.H.K. and G.B.; writing—review and editing, N.A.A., Y.H.K. and G.B.; visualization, Y.H.K. and G.B.; supervision, N.A.A.; project administration, N.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study used a publicly available anonymized secondary dataset and did not involve direct interaction with human participants or animals.

Informed Consent Statement: Not applicable. This study used a publicly available anonymized secondary dataset. No participants were directly recruited, and no identifiable personal information was used.

Data Availability Statement: Data Availability Statement: The data analyzed in this study are publicly available in the Open University Learning Analytics Dataset (OULAD) at <https://research.stem.open.ac.uk/ouanalyse/dataset/> (Accessed on 1 June 2026). The dataset description is available at <https://doi.org/10.1038/sdata.2017.171> (Accessed on 1 June 2026).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, C.; Zhao, G. Machine-Learning-Based Imputation Method for Filling Missing Values in Ground Meteorological Observation Data. *Algorithms* **2023**, *16*, 422. [\[CrossRef\]](#)
2. Drachsler, H.; Verbert, K. Panorama of recommender systems to support learning. *Comput. Educ.* **2015**, *84*, 1–13.
3. Bobadilla, J.; Ortega, F.; Hernando, A.; Gutiérrez, A. Recommender systems survey. *Knowl.-Based Syst.* **2013**, *46*, 109–132. [\[CrossRef\]](#)
4. Shokeen, S.; Rana, C. A study on features of social recommender systems. *Artif. Intell. Rev.* **2020**, *53*, 965–988.
5. Ricci, F.; Rokach, L.; Shapira, B. *Recommender Systems Handbook*; Springer: New York, NY, USA, 2015.
6. Lu, J.; Wu, D.; Mao, M.; Wang, W.; Zhang, G. Recommender system application developments: A survey. *Decis. Support Syst.* **2022**, *193*, 113719.

7. Verbert, K.; Duval, E.; Santos, O.C. Learning dashboards: Visualizing and awareness tools for learners. In *Proceedings of LAK'12*; ACM: New York, NY, USA, 2012.
8. Lu, J.; Zhang, G. Explainable Artificial Intelligence for Intelligent Education Systems: A Hybrid Approach. *IEEE Trans. Learn. Technol.* **2021**, *14*, 559–572.
9. Tang, T.; McCalla, G. Smart recommendation for an evolving e-learning system: Architecture and experiment. *Int. J. E-Learn.* **2005**, *4*, 105–129.
10. Manouselis, N.; Costopoulou, C. Analysis and classification of multi-criteria recommender systems. *World Wide Web* **2007**, *10*, 415–441. [[CrossRef](#)]
11. Herlocker, J.L.; Konstan, J.A.; Terveen, L.G.; Riedl, J.T. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.* **2004**, *22*, 5–53. [[CrossRef](#)]
12. Drachsler, H.; Greller, W. Privacy and analytics—It's a DELICATE issue: A European framework for learning analytics. *Educ. Technol. Soc.* **2016**, *19*, 42–52.
13. Chen, L.; Wang, F.; Yu, X. Feedback-based explainable recommendation for online learning. *Comput. Hum. Behav.* **2021**, *124*, 106898.
14. Khamparia, A.; Pandey, B. A hybrid human–AI intelligent tutoring system for adaptive learning. *Educ. Inf. Technol.* **2022**, *27*, 7453–7471.
15. Kumar, S.; Singh, R.; Pande, B. Hybrid adaptive models for personalized learning recommendation. *Appl. Soft Comput.* **2023**, *138*, 110209.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.