WILEY | Hindawi

Research Article

A Unified Interpretable Intelligent Learning Diagnosis Framework for Learning Performance Prediction in Intelligent Tutoring Systems

Zhifeng Wang [iD],[1] **Wenxing Yan,**[1] **Chunyan Zeng** [iD],[2] **Yuan Tian,**[1] **and Shi Dong**[1]

[1]*School of Educational Information Technology, Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan 430079, China*
[2]*Hubei Key Laboratory for High-efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei University of Technology, Wuhan 430068, China*

Correspondence should be addressed to Zhifeng Wang; zfwang@ccnu.edu.cn and Chunyan Zeng; cyzeng@hbut.edu.cn

Intelligent learning diagnosis is a critical engine of intelligent tutoring systems, which aims to estimate learners' current knowledge mastery status and predict their future learning performance. The significant challenge with traditional learning diagnosis methods is the inability to balance diagnostic accuracy and interpretability. Although the existing psychometric-based learning diagnosis methods provide some domain interpretation through cognitive parameters, they have insufficient modeling capability with a shallow structure for large-scale learning data. While the deep learning-based learning diagnosis methods have improved the accuracy of learning performance prediction, their inherent black-box properties lead to a lack of interpretability, making their results untrustworthy for educational applications. To settle the abovementioned problem, the proposed unified interpretable intelligent learning diagnosis framework, which benefits from the powerful representation learning ability of deep learning and the interpretability of psychometrics, achieves a better performance of learning prediction and provides interpretability from three aspects: cognitive parameters, learner-resource response network, and weights of self-attention mechanism. Within the proposed framework, this paper presents a two-channel learning diagnosis mechanism LDM-ID as well as a three-channel learning diagnosis mechanism LDM-HMI. Experiments on two real-world datasets and a simulation dataset show that our method has higher accuracy in predicting learners' performances compared with the state-of-the-art models and can provide valuable educational interpretability for applications such as precise learning resource recommendation and personalized learning tutoring in intelligent tutoring systems.

1. Introduction

The integration of modern information technology with traditional teaching and education methods has promoted the emergence of smart education, which refers to a new form of education that relies on modern information technology in the education field [1]. It applies new technologies such as artificial intelligence, virtual reality, cloud computing, and big data to improve the quality and effectiveness of education [2]. As a vital part of smart education, the intelligent tutoring system can support monitoring and measuring the learning process and digging out the learning patterns of learners [3], as well as effectively presenting the results to teachers and learners [4]. Correspondingly, intelligent learning diagnosis [5] is one of the key components of the intelligent tutoring system [6], and it supports the intelligent tutoring system through data mining and educational theories technically. To provide learners with personalized learning tutoring services, the intelligent learning diagnosis employs educational psychology and data mining technologies to diagnose the learner's learning state and predict possible future learning performances by collecting and analyzing the exercise records and behavioral data, which are generated by the learner during the learning

process. In both online and offline education scenarios, the number of teachers is far lower than the number of learners. There is an urgent need for intelligent means to assist teachers in improving efficiency and provide targeted guidance for learners to bridge this resource mismatch. As a result, intelligent learning diagnosis has been applied in many e-learning systems such as Coursera, EDX, Khan Academy, and offline classes, and it has brought great progress to the education industry [7]. While facilitating the renewal of teachers' teaching philosophy and driving smart learning systems to provide large-scale personalized learning services, the intelligent learning diagnosis can help teachers provide appropriate guidance based on learners' characteristics, as well as help learners gain a more objective and clear understanding of their learning progress and status, enabling learners to regulate and standardize their learning independently.

The key to intelligent learning diagnosis is to mine the learner's potential learning state, including their mastery state of knowledge points and learning ability, then predict the learner's performance on particular learning tasks. For this pivotal problem, two types of solutions are widely developed at present: psychometrics-based methods [8–15] and deep learning-based methods [16–23]. First, psychometrics-based methods appeared earlier and have been developed for more than 40 years. They usually describe the features of the learner and the exercises based on a specific cognitive perspective and employ an empirical response function to express the relationship between learners' performance on exercises and their knowledge status. Psychometrics-based methods are easy to understand, and the predefined parameters and the connections between them have clear actual significance and interpretability. Second, deep learning diagnosis methods appeared with the development of deep learning technology, which collects the learning resource-related information and learners' learning data, then models learners' learning process and their knowledge status through deep neural networks such as recurrent neural networks [16], convolutional neural networks [18], memory neural networks [19], and graph neural networks [21]. With support from deep neural networks' powerful representation learning capabilities, deep learning-based diagnosis methods achieve higher accuracy in learning performance prediction.

The abovementioned methods are well developed for intelligent learning diagnosis and have been applied in different scenarios of intelligent tutoring systems. However, there are still some significant challenges. First, traditional psychometric learning diagnosis models introduce a few specific cognitive parameters to model learners from a single perspective, making them difficult to fit the actual learning situation perfectly. For example, the Deterministic Input, Noisy "And" gate (DINA) model can only diagnose the learner's mastery of knowledge in a relatively rough way [12]. When different learners have not mastered a certain knowledge point, it cannot reflect more detailed differences. Second, due to the black-box characteristics of deep learning methods, the interpretability of deep learning diagnosis is

relatively low. Given an example, the deep knowledge tracing (DKT) model summarizes a learner's state of all knowledge concepts in one hidden state, so that it is difficult to know exactly how much a student has mastered a certain knowledge concept [16].

In this paper, we proposed a novel unified interpretable intelligent learning diagnosis framework that can effectively solve the above problems. First, multichannel psychometric learning diagnosis models are introduced to initialize the learning diagnosis of learners from a multicognitive perspective. Second, the learner representation network and learning resource representation network are constructed separately using the stacked autoencoder (SAE) network. Then, the learner-learning resource response network is designed to extract deep learning features. Subsequently, deep learning features are fused with shallow learning features, and a self-attention mechanism is introduced to weigh the fused features. Finally, a learning performance prediction network is designed through convolutional neural networks. While obtaining better learning performance predictions, the proposed framework provides interpretability in three aspects: cognitive parameters, learner-resource response network, and weights of self-attention mechanism. Our main contributions can be summarized as follows:

(i) Framework: The proposed unified interpretable intelligent learning diagnosis framework, which benefits from the powerful representation learning ability of deep learning and the interpretability of psychometrics, achieves a better performance of learning prediction and provides interpretability from three aspects: cognitive parameters, learner-resource response network, and weights of self-attention mechanism. Furthermore, the proposed framework is a generic learning diagnosis pattern that can incorporate a variety of different psychometric models and deep learning diagnosis methods into this framework.

(ii) Mechanisms: Based on the proposed framework, a 2-channel learning diagnosis mechanism LDM-ID is implemented based on the fusion of IRT and DINA on the one hand, and a 3-channel learning diagnosis mechanism LDM-HMI is implemented based on the fusion of Ho-DINA, MIRT, and IRT on the other hand. The proposed two learning diagnosis mechanisms provide reference examples for developing new learning diagnosis models in different educational scenarios. The codes of these two mechanisms are released at https://github.com/CCNUZFW/LDM-ID-HMI.

(iii) Application: The proposed two mechanisms are compared with 11 state-of-the-art methods in two publicly available online class databases and a self-built offline class database, and the proposed method in this paper obtains better learning prediction accuracy (AUC absolute improves by about 5%), while the stability of learning prediction is also

better (RMSE relatively decreases by about 7%), and is comparable to current deep learning diagnosis methods in terms of prediction time cost. In addition, this paper constructs and open-sources a database of participants in offline courses named CL21 to verify the effectiveness of this method in both online and offline classes. CL21 is released at https://github.com/CCNUZFW/CL21.

The rest of the paper is organized as follows: Section 2 briefly reviews some related work on intelligent learning diagnosis. Section 3 presents the intelligent learning diagnosis model, including the basic assumptions, definitions, and problem formula. Section 4 introduces our proposed unified interpretable intelligent learning diagnosis framework in detail, and Section 5 describes two specific implementation methods under the proposed framework. Then, experiments are conducted on two real-world datasets and one virtual dataset, and the results of the proposed method are compared with several baseline methods in Section 6. Finally, we conclude the paper and identify future research that could be carried out based on our new findings in Section 7.

2. Related Work

This section briefly reviews the main research work related to intelligent learning diagnosis, which can be divided into two categories: (1) learning diagnosis based on psychometrics and (2) learning diagnosis based on deep learning. While the psychometrics-based learning diagnosis methods aim at mining learners' potential cognitive situation, the deep learning-based learning diagnosis methods focus on the deep representation learning of learning data and the accurate prediction of learning performance.

2.1. Learning Diagnosis Based on Psychometrics. This category of learning diagnosis methods is derived from psychometrics [24]. These methods are based on cognitive psychological theories, using modern statistical methods to construct learning diagnosis models and combining estimation and inference to diagnose the cognitive structure and state of learners [25]. Psychometrics-based learning diagnosis methods can be divided into two subcategories: (1) continuous learning diagnosis models and (2) discrete learning diagnosis models.

2.1.1. Continuous Learning Diagnosis Models. Continuous learning diagnosis models which are based on item response theory (IRT) [26] assume learners have a "latent trait" and learners' responses to exercises are determined by the exercise's factors and the learner's potential features jointly. The learner's knowledge state diagnosis result is demonstrated by modeling the relationship among the learner's potential features, the exercise's factor, and the learner's answering response [8]. Typical methods, such as IRT [26] use logical functions to describe the relationship between learner parameters and responses, but the learner characteristic parameters in this model are an expression of the learner's overall mastery level and cannot be refined to every knowledge point. As a result, a multi-dimensional IRT (MIRT) model [10] was developed, which estimates the ability of learners in multiple dimensions at the same time, and considers the relationship between each ability dimension. Furthermore, since MIRT does not consider the hierarchical relationship between multidimensional ability traits, De la Torre et al. proposed a high-order item response model (Ho-IRT) [11], which can directly deal with the potential traits of learners with a hierarchical structure.

2.1.2. Discrete Learning Diagnosis Models. Discrete learning diagnosis models assume that the ability value is discontinuous and the latent knowledge space is composed of K dichotomous variables. Therefore, there are 2^K knowledge mastery states. The learners are divided into these mastery states according to their exercise responses and distinguish their cognitive status of knowledge accordingly. For example, the DINA [12] model uses guess and slip parameters to define the attributes of the exercise, and it employs a knowledge mastery mode vector that takes the value of 0 or 1 to represent the learner's mastery of each knowledge point. It is assumed that the learner must master all the knowledge points corresponding to the exercise to answer correctly based on the noncompensation assumption. Correspondingly, based on compensation assumption, the Deterministic Input, Noisy "Or" gate (DINO) [13] model assumes that the learner can make the correct answer after mastering at least one of the knowledge points. Furthermore, for more fine-grained modeling, the Noisy Input, Deterministic "And" gate (NIDA) [14] model defines the parameters at the knowledge points level of the exercises, and refines the scope of the parameters compared with DINA and DINO models. To be able to report both the macroscopic general ability and the microscopic cognitive state at the same time, the High-order DINA (Ho-DINA) model [15] is proposed, which assumes that the knowledge points are independent of each other and subordinate to higher-order general capabilities. Assuming that the learner's response matrix is determined by the inner product of the learner's knowledge proficiency matrix and the knowledge point vector contained in the exercise, the probabilistic matrix factorization-based learning diagnosis model (PMF-LDM) is established through matrix decomposition [27].

Although psychometric-based learning diagnostic methods can provide explanations of learning diagnosis in terms of cognitive parameters, current methods define targeted cognitive parameters for a specific learning scenario. These learning diagnostic models based on specific cognitive parameters can only interpret learning diagnostic results from a single perspective, so it becomes important to construct a unified learning diagnosis framework that includes multiple cognitive parameters from integrated perspectives and can be applied to multiple learning scenarios.

2.2. Learning Diagnosis Based on Deep Learning. Deep learning methods have powerful representation capabilities and have been widely used in computer vision [28, 29],

speech processing [30, 31], natural language understanding [32], and other fields [33].

Currently, advanced deep learning methods have been widely studied in diagnosis and prediction tasks. For intelligent fault diagnosis, the advent of deep learning has constructed an end-to-end diagnosis procedure [34]. Furthermore, out-of-distribution detection-assisted trustworthy machinery fault diagnosis approach is developed to enhance the reliability and safety of intelligent diagnosis models [35]. For the wheel wear status diagnosis of high-speed trains, a multiplex local temporal fusion architecture is combined with the transformer architecture [36].

Deep learning is also extensively used in the learning diagnosis task of learners in the field of intelligent tutoring systems [37]. Deep learning-based learning diagnosis usually models the learner's learning process based on their answering records sequence through an end-to-end trainable neural network. Classified by the network structures, the learning diagnosis methods based on deep learning could be divided into five categories: (1) the recurrent neural network-based learning diagnosis, (2) the convolutional neural network-based learning diagnosis, (3) the memory network-based learning diagnosis, (4) the graph neural network-based learning diagnosis, and (5) the deep neural network-based learning diagnosis.

2.2.1. The Recurrent Neural Network-Based Learning Diagnosis. The DKT proposed by Piech et al. [16] first applied the recurrent neural network to learning diagnosis, which could break the restriction of empirical response function by simulating the interaction between the learner and the exercise with recurrent neural networks. However, DKT can only simply predict whether a learner has mastered or not mastered a specific knowledge point, and can do nothing about the intermediate state of learner knowledge mastery. Furthermore, Su et al. [17] proposed the exercise-enhanced recurrent neural network (EERNN) that predicts the learner's performance through the learner's answer record and the text information of each exercise. For multi-features, Xiong et al. [38] proposed an extended DKT model, adding more answering features such as answering time, question difficulty, and learners' previous knowledge to assist in predicting learning performance. In terms of the improvement of the model structure, Pandey and Karypis [39] proposed the Self-Attentive Knowledge Tracing (SAKT) model, which is the first to apply the attention mechanism to the learning diagnosis task. Besides, Ghosh et al. [40] used a monotonic attention mechanism to limit the weight of questions that are far away from the learner's current answer. Based on the attention mechanism, Choi et al. [41] further applied the encoder-decoder model to the learning diagnosis task.

2.2.2. The Convolutional Neural Network-Based Learning Diagnosis. Different from models based on recurrent neural networks, to achieve personalized modeling, Shen et al. [18] proposed the convolutional knowledge tracing (CKT)

method, which utilizes the sensitivity of convolutional neural networks for the temporal and spatial information to consider the prior knowledge and learning rate for each learner. Furthermore, Wang et al. [42] combined a convolutional neural network and a recurrent neural network into an integrated model to build convolutional recurrent knowledge tracing (CRKT).

2.2.3. The Memory Network-Based Learning Diagnosis. Since DKT only uses a latent state to represent the learner's knowledge mastery state and cannot simulate the dynamic change process of learners' mastery of different knowledge points, Zhang et al. [19] proposed Dynamic Key-Value Memory Networks (DKVMN), which draws on the idea of memory networks, uses a key matrix for representing the knowledge and a value matrix for indicating the learner's mastery of each knowledge point and predicts learner's performance according to the key matrix and value matrix. Furthermore, Tsutsumi et al. [20] add exercises' difficulty and learners' ability to DKVMN to improve the accuracy and interpretability.

2.2.4. The Graph Neural Network-Based Learning Diagnosis. Inspired by cognitive theory, the structural information between knowledge points is also helpful for learning diagnosis. Therefore, Nakagawa et al. [21] explored the knowledge graph structure and applied it to learning diagnosis with graph neural networks. Besides, Yong et al. [22] considered the exercise-knowledge correlations for learning diagnosis through graph convolutional networks. At the same time, Tong et al. [43] are more focused on using graph neural networks to capture the relationship between different knowledge points, such as similarity or prerequisite relationship. Furthermore, Abdelrahman and Wang [44] used the latent knowledge graph to improve the reading and writing process of the memory matrix of the DKVMN model.

2.2.5. The Deep Neural Network-Based Learning Diagnosis. Neural Cognitive Diagnosis Model (NeuralCDM) [45] is proposed to project students and exercises to factor vectors and incorporates neural networks to obtain complex learning interactions. Besides, deep matrix factorization-based learning diagnosis (DeepMFLD) [46] acquires the learners' embedding and the exercises' embedding separately by deep matrix decomposition, then fuses them with a deep neural network, and finally predicts the learner's performance by a fully connected layer. Furthermore, item response ranking (IRR) framework [47] introduces pairwise learning into learning diagnosis to sufficiently model the monotonicity between learning interactions.

Although the accuracy of learning diagnosis can be improved by introducing deep learning technology, its black-box and end-to-end properties in learning modeling [23] lead to its inability to provide supporting explanatory information and model intermediate feedback for instructional and learning decisions.

2.3. Research Questions and Motivation. Based on the research of the related work, we summarized the research questions and motivation of this paper as follows:

(i) RQ1: How to build a unified learning diagnosis framework that incorporates multidimensional cognitive parameters and can be applied to a broader range of learning scenarios?

Inspired by the multimodal approach, we fused multiple psychometrics-based learning diagnosis models containing different cognitive parameters into a unified multimodal learning diagnosis model. On the one hand, this allows for multidimensional interpretable cognitive parameters. On the other hand, it also enables the proposed unified learning diagnosis framework to be extended to a more practical and broader range of educational application scenarios.

(ii) RQ2: How to effectively characterize learners and learning resources and appropriately establish effective interactive responses between learner and learning resource features?

Inspired by the representational power of deep learning methods, on the one hand, we constructed a learner network and a learning resource network based on shallow learning features to extract better representational deep learner features and deep learning resource features; on the other hand, we characterized the interactive responses between them by constructing a learner-learning resource response network. Furthermore, the prediction network is constructed by combining shallow and deep learning features to obtain better representativeness as well as to provide interpretable cognitive parameters.

(iii) RQ3: How to effectively improve learning performance predictions while also providing appropriate explanatory information that can be used to improve the acceptability and usability of learning diagnosis?

Inspired by the interpretable AI approach, first, we provided self-contained explanatory information describing learners' cognitive states from multidimensional cognitive parameters; second, we provided explanatory information on the interaction responses between learners and learning resources by constructing learner-learning resource response networks for discovering which learning interaction channels are more effective; and finally, we obtained global explanatory information on the importance of different learning features in terms of decision making due to the attention mechanism for identifying essential learning cues.

3. Intelligent Learning Diagnosis Model

This section introduces some assumptions and definitions, and then details the problem definition.

3.1. Assumptions and Definitions of Intelligent Learning Diagnosis. First, this paper is based on the following assumptions and definitions:

Assumption 1. The obtained online and offline learning data is used with permission.

Assumption 2. The data is authentic and complete. The learners participating in the data collection are required to take the course seriously and use their initiative to complete each exercise, to ensure that there is no halfway exit leading to incomplete data or inconsistency in answering results. And the data can reflect the actual level of learners.

Assumption 3. The learner's answer results are partially independent. On the one hand, it means that the learner's answering results of course exercises will not be affected by other people. At the same time, it will not affect others. On the other hand, it means that the answering results of learners in different exercises will not be affected by each other.

Definition 1. Shallow feature of learning resource.
For the learning resources involved in the course (mainly refer to the exercises performed by learners), it obtains descriptive parameters for them based on different educational interpretation meanings to form a parameter matrix $EC \in R^{e \times d_e}$, where e and d_e represent the number of exercises and the dimension of shallow learning resource feature, respectively.

Definition 2. Shallow feature of learners.
Combining multiple focal points of different cognitive diagnosis models, it describes learners from different angles and forms a parameter matrix $SC \in R^{s \times d_s}$, where s and d_s, respectively, represent the number of learners and the dimension of shallow learner feature.

Definition 3. Deep feature of learning resources and learners.
Based on the two descriptive parameters, it introduces deep learning technology to mine higher-order bottleneck features, which models and characterizes the response between learning resources and learners more effectively.

3.2. Problem Description. According to the above definitions and assumptions, the unified interpretable intelligent learning diagnosis task is described as follows. Given that S learners study the same course and complete the same exercise E corresponding to course C, the multiple psychometric models are integrated to obtain the learner and learning resource parameter matrix, and the learner's answer performance and cognitive state are predicted based on the parameter matrix. Among them, S is a set of learners, E is a set of exercises, K is a set of knowledge points, and the answer record of learner s is $R_s = \{(e_1, r_1), (e_2, r_2), \ldots, (e_t, r_t)\}$. Among them, e_t is the exercises performed by the learner s at time t, r_t is the corresponding answer result, $r_t \in \{0, 1\}$, $r_t = 1$ means the

answer is correct, and $r_t = 0$ means the answer is wrong. q_i is the set of knowledge points involved in exercise i and $q_{ij} = 1$ means that there is a correlation between exercise i and knowledge point j, otherwise, it is irrelevant. The set of answer records of all learners $R = \{R_1, R_2, \ldots, R_s\}$ consists the answer matrix R, and $Q = \{q_1, q_2, \ldots, q_i\}$ consist of the exercise-knowledge point association matrix Q. Then, the problem can be defined as follows:

Definition 4. Intelligent learning diagnosis problem.

Given the exercising logs of each learner and the knowledge points involved in each exercise, our goal is threefold: (1) extract interpretable parameters of learners and learning resources; (2) construct a learner-resource response network and deep-shallow feature fusion mechanism to obtain the explanatory information on the importance of learning features; and (3) predict the learning performance $\widehat{r_{t+1}}$ of the specific learner on time $t + 1$ and offer interpretable information for decision-making.

4. Proposed Unified Interpretable Intelligent Learning Diagnosis Framework

Based on the assumptions and definitions of the above learning diagnosis model, this paper proposes a unified interpretable learning diagnosis framework, which is versatile and can flexibly formulate specific strategies according to different needs to realize the interpretable learning diagnosis and performance prediction of learners. As shown in Figure 1, the proposed framework mainly consists of five components: initial learning diagnosis, learner and learning resources representation network, learner-resource response network, deep-shallow learning feature fusion, and learning performance prediction network. The framework that takes advantage of the powerful representation learning ability of deep learning and the interpretability of psychometrics, achieves a good performance of learning prediction and provides interpretability from three aspects: cognitive parameters, learner-resource response network, and weights of self-attention mechanism.

4.1. Initial Learning Diagnosis.
The main purpose of the initial learning diagnosis module is to construct the multichannel cognitive parameter set.

The cognitive parameter set construction process is shown in Figure 1. Based on the learner's historical learning records and learning resource information, the multichannel psychometric model is used to make a preliminary diagnosis of the learner, and to estimate the parameters of the learning resource at the same time, thereby constructing the learning resource parameter set EC and learner parameter set SC.

Taking into account the interpretability of the learner's cognitive state, it is worth noting that the parameters are highly interpretable with practical significance in cognitive theory and psychometrics, such as the difficulty and discrimination of the exercise, the ability of the learner, which differs from learning parameters through network training in the deep learning-based learning diagnosis methods.

$$EC = \{PM_1(R, Q), PM_2(R, Q), \ldots, PM_n(R, Q)\},$$
$$SC = \{PM_1(R, Q), PM_2(R, Q), \ldots, PM_n(R, Q)\}. \quad (1)$$

PM_n is the nth psychometric model. Then, based on the answer record $R_s = \{(e_1, r_1), (e_2, r_2), \ldots, (e_t, r_t)\}$ of learner s, we obtained the learner parameter vector SC_s and the learning resource parameter vector EC_1 to EC_t related to learner s.

4.2. Learner and Learning Resources Representation Network.
This module consists of two parts: learners representation network and learning resources representation network.

Specifically, given the learner parameter vector SC_s of learner s and the learning resource parameter vector EC_e of exercise e, firstly, in order to eliminate the difference between the ranges of various parameters, the parameter values are discretized and one-hot encoded, and then SC_s is coded as a d_0 dimensional 0-1 vector x^s and EC_e is coded as a d_1 dimensional 0-1 vector x^e.

To better represent learners and learning resources based on cognitive parameters, the learner representation network and learning resource representation network are designed, respectively, which take advantage of auto-encoder to excavate the deep learning features of learners and learning resources. The SAE can copy the input to the output through a hidden layer. It consists of two parts: an encoder that can be represented by the function $h = f(\cdot)$ and a decoder that generates reconstruction information $y = g(\cdot)$, by training the auto-encoder to reproduce the input, h can obtain some useful features, to achieve the purpose of feature learning for the learner's parameter vector. The learners and learning resource auto-encoder designed in this paper can be expressed as follows:

$$h^s = f(x^s) = \sigma(W_1^s \times x^s + b_1^s),$$
$$y^s = g(h^s) = \sigma(W_2^s \times h^s + b_2^s),$$
$$h^e = f(x^e) = \sigma(W_1^e \times x^e + b_1^e),$$
$$y^e = g(h^e) = \sigma(W_2^e \times h^e + b_2^e), \quad (2)$$

W, b, and $\sigma(\cdot)$ are weights, bias, and activation function, respectively. We take the intermediate results h^s and h^e of the coding layer as the feature learning results, which can represent a more valuable part of the information in the original parameter vector.

4.3. Learner-Resource Response Network.
Then, the learner-resource response network is used to simulate the answer process in the actual situation. It takes the deep characteristics of learners and learning resources as the input and outputs the deep characteristics after the interaction between learners and learning resources.

$$f = concat(h^s, h^e),$$
$$f_d = \sigma(W_3 \times f + b_3), \quad (3)$$

where σ is the tanh activation function and h and y are the encoder and the decoder, respectively. In addition, the learner parameter encoder's weight matrix $W_1^s \in R^{d_2 \times d_0}$, the
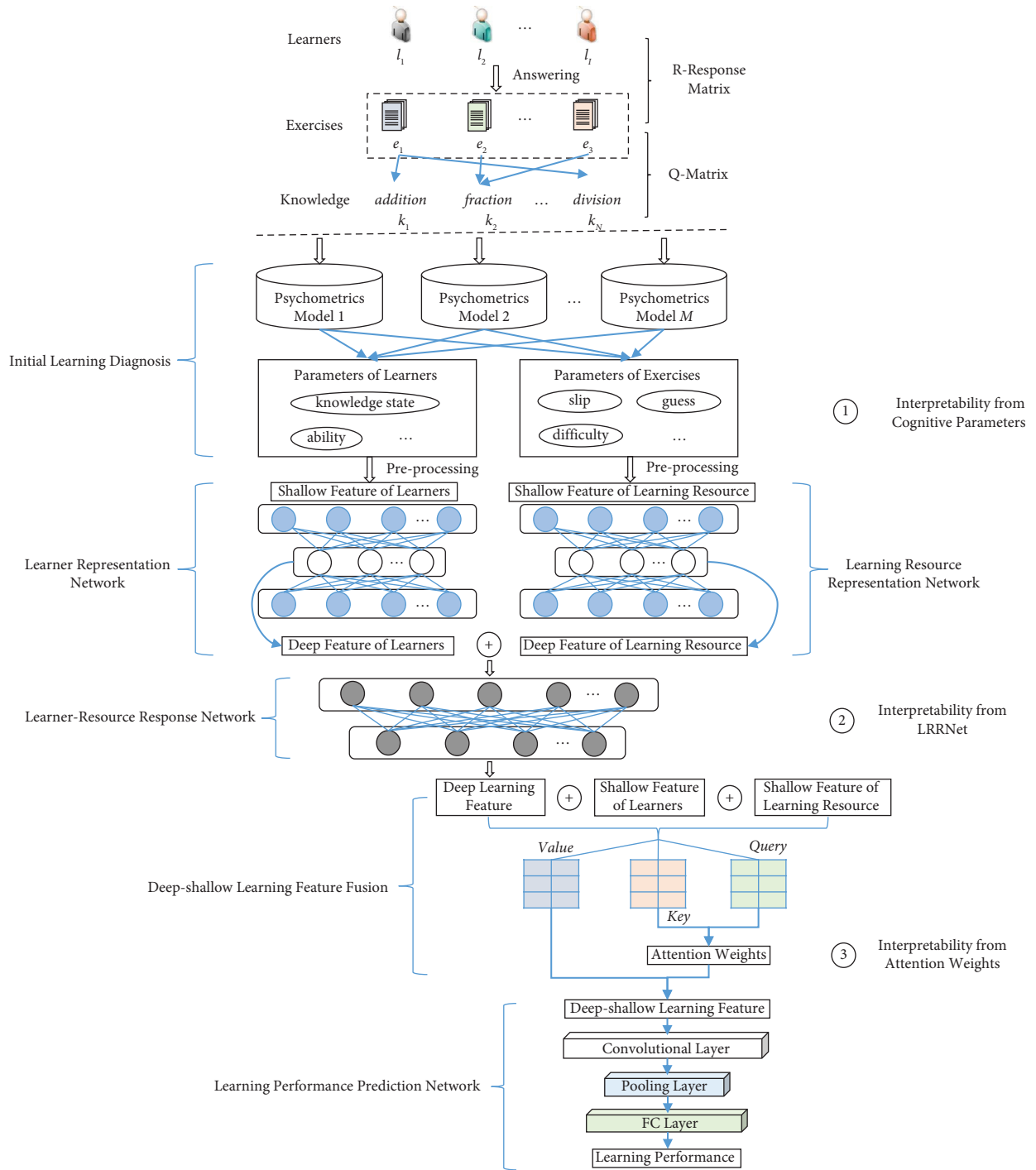
FIGURE 1: The inputs of the proposed unified interpretable intelligent learning diagnosis framework are the learners' response matrix R and the exercise-knowledge point matrix Q. The proposed framework contains five main components: (1) initial learning diagnosis, (2) learner and learning resource representation network, (3) learner-resource response network, (4) deep-shallow learning feature fusion, and (5) learning performance prediction network. The interpretability can be obtained from three aspects: (1) the interpretability from cognitive parameters, (2) the interpretability from LRRNet, and (3) the interpretability from attention weights.

bias vector $b_1^s \in R^{d_2}$, the decoder's weight matrix $W_2^s \in R^{d_0 \times d_2}$, bias vector $b_2^s \in R^{d_0}$, the learning resource parameter encoder's weight matrix $W_1^e \in R^{d_3 \times d_1}$, bias vector $b_1^e \in R^{d_3}$, decoder's weight matrix $W_2^e \in R^{d_1 \times d_3}$, and the bias vector $b_2^e \in R^{d_1}$ are the parameters of the learner

representation network and learning resource representation network. The weight matrix $W_3 \in R^{(d_2+d_3) \times d_4}$ and the bias vector $b_3 \in R^{d_2+d_3}$ are the parameters of the learner-resource response network. f_d is the deep feature of learners and learning resources.

4.4. Deep-Shallow Learning Feature Fusion. To improve the performance and interpretability of the model, we regard the original cognitive parameter vector as shallow features and the deep characterization features as deep features, and then we combine the two features as the overall features for diagnosis and prediction.

$$f = \text{concatenate}\left(f_d, SC_s, EC_e\right), \qquad (4)$$

where the dimensions of f are d_5, in which $d_5 = d_4 + d_e + d_s$.

In terms of interpretability for the prediction process, we focus on the reasons why the model makes such predictions. Therefore the prediction network in this paper employs the self-attention mechanism to process the fusion features of the learner and the learning resource, in order to fully mine the correlation and importance information between the features of each dimension, so as to provide interpretability from the feature level for the final diagnosis and prediction results, in other words, we can know exactly which features have a more significant impact on the final prediction result.

A convolutional layer is used with a convolution kernel size of 1 to obtain the query vector matrix Query, the key vector matrix Key, and the value vector matrix value through f as follows:

$$\begin{aligned} \text{Query} &= \text{Conv}\left(f\right), \\ \text{Key} &= \text{Conv}\left(f\right), \\ \text{Value} &= \text{Conv}\left(f\right). \end{aligned} \qquad (5)$$

The weight of each dimension of data in the feature is calculated by calculating the dot product between the query vector matrix and the key vector matrix, and the weight indicates the degree of relevance between the task to be queried and each input data.

$$\text{similarity}\left(\text{Query}, \text{Key}_i\right) = \text{Query} \bullet \text{Key}_i. \qquad (6)$$

On the one hand, numerical conversion of the degree of relevance can be carried out to normalize and organize the original calculated scores into a probability distribution with the sum of weights being 1. On the other hand, it can also highlight the weights of essential elements.

$$a_i = \text{SoftMax}\left(\text{similarity}_i\right) = \frac{e^{\text{similarity}_i}}{\sum_{j=1}^{d_s} e^{\text{similarity}_j}}. \qquad (7)$$

The input data is weighted and summed to obtain the fused feature data.

$$f_a = \sum_{i=1}^{d_5} a_i \cdot \text{Value}_i. \qquad (8)$$

4.5. Learning Performance Prediction Network. With the features acquired from the feature extraction process, we constructed a learner performance prediction network based on a convolutional neural network, used learner feature vectors and learning resource feature vectors to predict the probability of learners answering a specific exercise

correctly, and diagnosed the knowledge mastery of learners and analyze the learning resources through the feature information of learners and learning resources.

The convolutional layer is used to extract the spatial information of the deep-shallow fusion feature, the convolution kernel size is set to 3, and the step size is set to 1. Then, the ReLU activation function is added to reduce the dependence between parameters, overfitting is avoided, and the maximum pooling layer is used to reduce the feature dimension, then the predicted value p of the learner's answer is obtained under a specific learning task accordingly. It can be expressed as follows:

$$\begin{aligned} f_c &= \text{Conv}\left(f_a\right), \\ f_{re} &= \text{Relu}\left(f_c\right), \\ f_p &= \text{MaxPool}\left(f_{re}\right), \\ p &= \sigma\left(W_4 \times f_p + b_4\right), \end{aligned} \qquad (9)$$

where f_c, f_{re}, and f_p are the intermediate variables of the prediction network.

To learn the various parameters of the model, the cross-entropy loss function is used between the student's real answer value r_t and the predicted answer value p_t as the objective function L of the model as follows:

$$L = -\sum_{i=1}^{N} r_t \log p_t + (1 - r_t) \log\left(1 - p_t\right). \qquad (10)$$

4.6. Interpretability of the Proposed Framework. The interpretability of our proposed framework includes three aspects, which correspond to the three modules of the framework:

4.6.1. Interpretability from Cognitive Parameters. Based on the learner's historical learning records and learning resource information, the multichannel psychometric model is used to make a preliminary diagnosis of the learner, and to estimate the parameters of the learning resource at the same time, thereby constructing the learning resource parameter set *EC* and learner parameter set *SC*. Taking into account the interpretability of the learner's cognitive state, it is worth noting that the parameters are highly interpretable with practical significance in cognitive theory and psychometrics, such as the difficulty and discrimination of the exercise and the ability of the learner, which differs from learning parameters through network training in the deep learning-based learning diagnosis methods.

4.6.2. Interpretability from Learner-Resource Response Networks. In this module, we first characterize the learners and learning resources, and then construct the learner-resource response network. The learner-resource response network is used to simulate the complex learning activities in the process of learners' answers, that is, the interactive response between learners and learning resources, which is closer to the actual situation and has a better interpretability of the learning process.
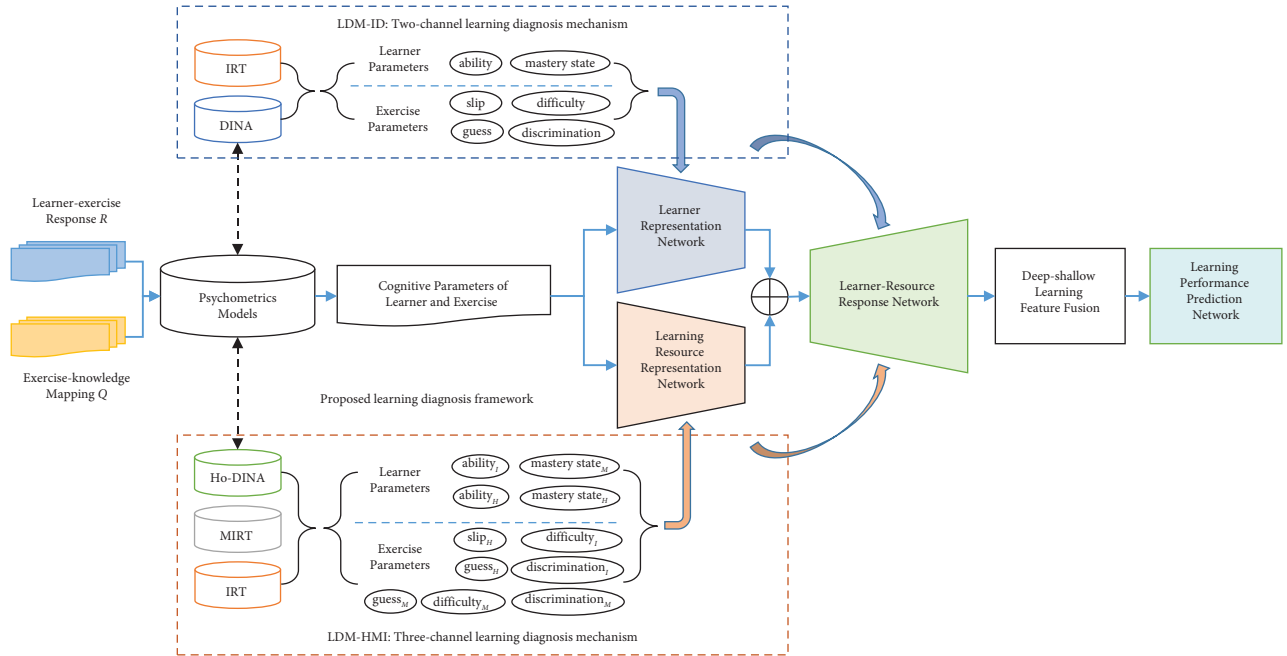
FIGURE 2: Based on the proposed unified interpretable learning diagnosis framework, two learning diagnosis mechanisms were developed. LDM-ID was based on IRT and DINA, which extracted two categories of learner cognitive parameters and four categories of exercise cognitive parameters. LDM-HMI was based on Ho-DINA, MIRT, and IRT, which extracted four categories of learner cognitive parameters and seven categories of exercise cognitive parameters, respectively.

4.6.3. Interpretability from Weights of Self-Attention Mechanism. In terms of interpretability for the prediction process, we focused on the reasons why the model makes such predictions, therefore the prediction network in this paper employs the self-attention mechanism to process the fusion features of the learner and the learning resource, in order to fully mine the correlation and importance information between the features of each dimension, so as to provide interpretability from the feature level for the final diagnosis and prediction results, in other words, we can know exactly which features have a greater impact on the final prediction result.

5. Multichannel Intelligent Learning Diagnosis Mechanisms

Within the proposed unified interpretable intelligent learning diagnosis framework in Section 4, two specific multichannel intelligent learning diagnosis mechanisms are implemented in this section. As shown in Figure 2, the first implementation is a two-channel learning diagnosis mechanism based on the fusion of IRT and DINA (LDM-ID), and the other is a three-channel learning diagnosis mechanism based on the fusion of Ho-DINA, MIRT, and IRT (LDM-HMI).

5.1. Two-Channel Learning Diagnosis Mechanism Based on the Fusion of IRT and DINA (LDM-ID). Generally, LDM-ID applies the IRT and DINA models for initial learning diagnosis, in which the DINA model uses a mastery vector to represent the master of the learner on each knowledge

component. The dimension of the mastery vector is equal to the number of knowledge, and the slip rate and guess rate are employed to represent the exercise parameters. The IRT model uses only one value to represent the overall ability of the learner and describes the exercises with difficulty and discrimination. The information on the parameters used to describe the learners and learning resources in these two models is integrated to obtain the original input parameter sets *EC* and *SC*, and then it performs subsequent learning diagnostics.

5.1.1. Initial Diagnosis Models of LDM-ID. The process of learning diagnosis from the initial parameter set has been described in detail in Section 4.

The response function of the IRT model is as follows:

$$P(s_{I\_\theta}) = e_{I\_\text{gues}} + \frac{1 - e_{I\_\text{gues}}}{1 + e^{-De_{I\_\text{disc}}(s_{I\_\theta} - e_{I\_\text{diff}})}}, \tag{11}$$

where s_{I_θ} is the learner's learning ability parameter. The parameter e_{I_guess} is generally called the "guessing parameter", that is, when the learner's ability value is very low (for example, close to negative infinity), it is still possible to give the right answer to this problem. e_{I_diff} is the problem difficulty parameter and e_{I_disc} is the problem discrimination parameter.

The response function of the DINA model is as follows:

$$P(s_{D\_\alpha}) = e_{D\_\text{gues}}^{(1-\eta)} \cdot \left(1 - e_{D_{slip}}\right)^{\eta}, \tag{12}$$

where s_{D_α} is the learner's mastery vector for each piece of knowledge, η is the learner's ideal answer situation under the corresponding mastery state, e_{D_slip} and e_{D_gues} are the guess

(i) **Input:** Student exercise response matrix R and question-knowledge correlation matrix Q
(ii) **Output:** Student exercise score prediction and interpretable parameters
(1) Initialize parameters matrix EC and SC:
(2) $EC^{(0)} = \left\{ e_{I\_\mathrm{diff}}^{(0)}, e_{I\_\mathrm{disc}}^{(0)}, e_{D\_\mathrm{gues}}^{(0)}, e_{D\_\mathrm{slip}}^{(0)} \right\}$, $SC^{(0)} = \left\{ s_{I\_\theta}^{(0)}, s_{D\_\alpha}^{(0)} \right\}$
(3) **while** not converge or not reach iterations **do**
(4) $L(R) = \prod_{u=1}^{U} L(R_u) = \prod_{u=1}^{U} \sum_{l=1}^{L} P(R_u \mid \alpha_l) P(\alpha_l)$
(5) $L(R \mid \delta, \pi) = \prod_{k=1}^{N} \prod_{j=1}^{M} p(\pi_k \mid \delta_j)^{r_{jk}} [1 - p(\pi_k \mid \delta_j)]^{1-r_{jk}}$
(6) //update parameters in DINA
(7) $e_{D\_\mathrm{slip}} = \widehat{e}_{D\_\mathrm{slip}}, e_{D\_\mathrm{gues}} = \widehat{e}_{D\_\mathrm{gues}} \leftarrow \partial \log L(R)/\partial e_{D\_\mathrm{slip}} = 0, \partial \log L(R)/\partial e_{D\_\mathrm{gues}} = 0$
(8) //update parameters in IRT
(9) $e_{I\_\mathrm{disc}} = \widehat{e}_{I\_\mathrm{disc}}, e_{I\_\mathrm{diff}} = \widehat{e}_{I\_\mathrm{diff}}, s_{I\_\theta} = \widehat{s}_{I\_\theta} \leftarrow \partial \log L(R)/\partial e_{I\_\mathrm{disc}} = 0, \partial \log L(R)/\partial e_{I\_\mathrm{diff}} = 0, \partial \log L(R)/\partial s_{I\_\theta} = 0$
(10) **end while**
(11) **for** student s and exercise e **do**
(12) //extract student features and exercise features via auto-encoder
(13) $h^s = f(SC_s) = \sigma(W_1^s \times SC_s + b_1^s)$
(14) $y^s = g(h^s) = \sigma(W_2^s \times h^s + b_2^s)$
(15) $h^e = f(EC_e) = \sigma(W_1^e \times EC_e + b_1^e)$
(16) $y^e = g(h^e) = \sigma(W_2^e \times h^e + b_2^e)$
(17) $h^s(\cdot), y^s(\cdot) = \mathrm{argmin}_{h^s(\cdot), y^s(\cdot)} \|h^s - SC_s\|^2, h^e(\cdot), y^e(\cdot) = \mathrm{argmin}_{h^e(\cdot), y^e(\cdot)} \|h^e - EC_e\|^2$
(18) $((SC_s, h^s), (EC_e, h^e)) \longrightarrow$ (exercise, stude nt), $f = \mathrm{concat}$ (exercise, stude nt)
(19) //calculate keys queries and values in attention function
(20)   **for** l in $\mathrm{len}(f)$ **do**
(21)   Query = $\mathrm{Conv}(f)$Key = $\mathrm{Conv}(f)$Value = $\mathrm{Conv}(f)$
(22) $f_a = \sum_{j=1}^{d_s} e^{\mathrm{similarity}_i}/\sum_{j=1}^{d_s} e^{\mathrm{similarity}_j} \cdot \mathrm{Value}_i$
(23) $\mathrm{similarity}(\mathrm{Query}, \mathrm{Key}_i) = \mathrm{Query} \bullet \mathrm{Key}_i$
(24)   **end for**
(25) **end for**
(26) $f_p = \mathrm{MaxPool}(\mathrm{Relu}(\mathrm{Conv}(f_a)))$
(27) $p = \sigma(W_3 \times f_p + b_3)$
(28) **return** p, EC, SC, attention weights

ALGORITHM 1: Two-channel learning diagnosis mechanism based on the fusion of IRT and DINA (LDM-ID).

and slip parameters of the exercises, respectively, and the slip parameter indicates that the learner has mastered all the knowledge required by the exercises he answered, but answered incorrectly. The guessing parameter is the probability that the learner has not mastered all the knowledge needed to answer the exercises or even not mastered one knowledge, but answered correctly. The specific parameter estimation process is shown in steps 1–7 of algorithm 1.

Accordingly, a parameter set of learners and a parameter set of learning resources are constructed based on the learner's learning record data and learning resource data as follows:

$$EC = \left\{ e_{I\_\mathrm{diff}}, e_{I\_\mathrm{disc}}, e_{D\_\mathrm{gues}}, e_{D\_\mathrm{slip}} \right\},$$
$$SC = \left\{ s_{I\_\theta}, s_{D\_\alpha} \right\}. \tag{13}$$

5.1.2. Representation and Prediction Network of LDM-ID. We modeled learners and learning resources separately and constructed a performance prediction network to complete the learning diagnosis task as described in Section 4 of the paper. The process is described in steps 12–28 of Algorithm 1.

5.2. Three-Channel Learning Diagnosis Mechanism Based on the Fusion of Ho-DINA, MIRT, and IRT (LDM-HMI). LDM-HMI is an implementation of a three-channel learning diagnosis mechanism under our proposed intelligent learning diagnostic framework, which employs the Ho-DINA model, MIRT model, and IRT model for initial learning diagnosis. Based on the knowledge state vector of learners in the DINA model, the Ho-DINA model constructs high-order potential abilities for learners to describe learners further. Compared with the one-dimensional ability value of the IRT model, the MIRT model uses a multidimensional ability vector to represent learners' knowledge states. Similarly, the parameter information used to describe learners and learning resources in the three models is integrated to obtain the original input parameter set, which can be used for subsequent learning diagnosis.

5.2.1. Initial Diagnosis Models of LDM-HMI. The Ho-DINA model assumes that the cognitive knowledge α_k is independent (partially independent) under a given ability s_{H_θ}. Cognitive knowledge has the following relationship as follows:

(i) **Input:** Student exercise response matrix R and question-knowledge correlation matrix Q
(ii) **Output:** Student exercise score prediction and interpretable parameters
(1) Initialize parameters matrix EC and SC:
(2) $EC^{(0)} = \left\{ e_{I\_\text{diff}}^{(0)}, e_{I\_\text{disc}}^{(0)}, e_{H\_\text{slip}}^{(0)}, e_{H\_\text{gues}}^{(0)}, e_{M\_\text{disc}}^{(0)}, e_{M\_\text{gues}}^{(0)}, e_{M\_\text{diff}}^{(0)} \right\}, SC^{(0)} = \left\{ s_{I\_\theta}^{(0)}, s_{H\_\theta}^{(0)}, s_{M\_\alpha}^{(0)}, s_{H\_\alpha}^{(0)} \right\}$
(3) **while** not converge or not reach iterations **do**
(4) $p(\widehat{\theta}\leftarrow\theta) = \min\left\{ p(\alpha\,|\,\theta,\widehat{\lambda})\,p(\widehat{\theta})/ \quad p(\alpha\,|\,\theta,\widehat{\lambda})\,p(\theta), 1 \right\}, \qquad p(\widehat{\alpha}\leftarrow\alpha) = \min\left\{ L(s, g; \quad \widehat{\alpha})\,p(\widehat{\alpha}\,|\,\theta\theta,\widehat{\lambda})/L(s, g; \alpha)\,p(\alpha\,|\,\widehat{\theta},\widehat{\lambda}), 1 \right\},$
    $p(\widehat{s}, \widehat{g}\leftarrow s, g) = \min\left\{ L(\widehat{s}, \widehat{g}; \widehat{\alpha})\,p(\widehat{s})\,p(\widehat{g})/L(s, g; \widehat{\alpha})\,p(s)\,p(g), 1 \right\}$
(5) $L_1(R) = \prod_{k=1}^{N}\prod_{j=1}^{M} p\,(\pi_k\,|\,\pi)\,p(\pi_k\,|\,\delta_j)^{r_{kj}}[1 - p(\pi_k\,|\,\delta_j)]^{1-r_{kj}}, \; L_2(R) = \prod_{i=1}^{I}\prod_{j=1}^{J} e^{r_{ij}(a_j\,\theta_i + d_j)}/1 + e^{r_{ij}(a_j\theta_i+d_j)}$
(6) //update parameters in Ho-DINA
(7) $e_{H\_\text{slip}} = \widehat{s}, e_{H\_\text{gues}} = \widehat{g}, s_{H\_\theta} = \widehat{\theta}, s_{H\_\alpha} = \widehat{H\_\alpha}$
(8) //update parameters in IRT
(9) $e_{I\_\text{disc}} = \widehat{e}_{I\_\text{disc}}, e_{I\_\text{diff}} = \widehat{e}_{I\_\text{diff}}, s_{I\_\theta} = \widehat{s}_{I\_\theta}\leftarrow\partial\log L_1(R)/\partial e_{I\_\text{disc}} = 0, \partial\log L_1(R)/\partial e_{I\_}\text{diff} = 0, \partial\log L_1(R)/\partial s_{I\_\theta} = 0$
(10) //update parameters in MIRT
(11) $e_{M\_\text{gues}} = \widehat{e}_{M\_\text{gues}}, e_{M\_\text{diff}} = \widehat{e}_{M\_\text{diff}}, e_{M\_\text{disc}} = \widehat{e}_{M\_\text{disc}}, s_{M\_\alpha} = \widehat{s}_{M\_\alpha}\leftarrow\partial\log L_2(R)/\partial e_{M\_\text{gues}} = 0,$
    $\partial\log L_2(R)/\partial e_{M\_\text{diff}} = 0, \partial\log L_2(R)/\partial e_{M\_\text{disc}} = 0, \partial\log L_2(R)/\partial s_{M\_\alpha} = 0$
(12) **end while**
(13) $EC^{(\text{final})} = \left\{ e_{I\_\text{diff}}^{(\text{final})}, e_{I\_\text{disc}}^{(\text{final})}, e_{H\_\text{slip}}^{(\text{final})}, e_{H\_\text{gues}}^{(\text{final})}, e_{M\_}\text{disc}^{(\text{final})}, e_{M\_\text{gues}}^{(\text{final})}, e_{M\_\text{diff}}^{(\text{final})} \right\}$
(14) $SC^{(\text{final})} = \left\{ s_{I\_\theta}^{(\text{final})}, s_{H\_\theta}^{(\text{final})}, s_{M\_\alpha}^{(\text{final})}, s_{H\_\alpha}^{(\text{final})} \right\}$
(15) **for** student s and exercise e **do**
(16) $h^s = f(SC_s) = \sigma(W_1^s \times SC_s + b_1^s)$
(17) $y^s = g(h^s) = \sigma(W_2^s \times h^s + b_2^s)$
(18) $h^e = f(EC_e) = \sigma(W_1^e \times EC_e + b_1^e)$
(19) $y^e = g(h^e) = \sigma(W_2^e \times h^e + b_2^e)$
(20) $h^s(\cdot), y^s(\cdot) = \text{argmin}_{h^s(\cdot), y^s(\cdot)}\|h^s - SC_s\|^2, h^e(\cdot), y^e(\cdot) = \text{argmin}_{h^e(\cdot), y^e(\cdot)}\|h^e - EC_e\|^2$
(21) $((SC_s, h^s), (EC_e, h^e)) \longrightarrow (\text{exercise, student}), f = \text{concat}(\text{exercise, student})$
(22)   **for** l in $\text{len}(f)$ **do**
(23) $\text{Query} = \text{Conv}(f)\text{Key} = \text{Conv}(f)\text{Value} = \text{Conv}(f)$
(24) $f_a = \sum_{i=1}^{d_s} e^{\text{similarity}_i}/\sum_{j=1}^{d_5}\sum_{j=1}^{d_5} e^{\text{similarity}_j} \cdot \text{Value}_i$
(25) $\text{similarity}(\text{Query}, \text{Key}_i) = \text{Query}\bullet\text{Key}_i$
(26)   **end for**
(27) **end for**
(28) $f_p = \text{MaxPool}(\text{Relu}(\text{Conv}(f_a)))$
(29) $p = \sigma(W_3 \times f_p + b_3)$
(30) **return** p, EC, SC, attention weights

ALGORITHM 2: Three-channel learning diagnosis mechanism based on the fusion of Ho-DINA, MIRT, and IRT.

$$P(\alpha\,|\,s_{H\_\theta}) = \prod_{k=1}^{K} P(\alpha_k\,|\,s_{H\_\theta}), P(\alpha_k\,|\,s_{H\_\theta}) = \frac{e^{\lambda_{0k}+\lambda_{1k}s_{H\_\theta}}}{1 + e^{\lambda_{0k}+\lambda_{1k}s_{H\_\theta}}}, \tag{14}$$

where λ_k represents the loading of knowledge k on the ability s_{H_θ}, that is, under the DINA model, it is assumed that knowledge is locally independent and subordinate to a higher-order ability.

The response function of the MIRT model is as follows:

$$P(s_{M\_\alpha}) = e_{M\_\text{gues}} + \frac{1 - e_{M\_\text{gues}}}{1 + e^{-D \cdot e_{M\_\text{disc}}(s_{M\_\alpha}+e_{M\_\text{diff}})}}, \tag{15}$$

where D is a constant; s_{M_α} is the learner's multidimensional ability vector, and each dimension is defined on a specific knowledge point; e_{M_disc} is the multidimensional discrimination vector of exercise, which is also defined in the dimension of the knowledge point; e_{M_gues} and e_{M_diff} are, respectively, guessing degree and difficulty parameters in the dimension of the question. The parameter estimation process is described in steps 1 to 18 in algorithm 2. The IRT

model is the same as in Section 5.1. Based on this, we constructed a parameter set of learners and learning resources as follows:

$$EC = \left\{ e_{I\_\text{diff}}, e_{I\_\text{disc}}, e_{H\_\text{slip}}, e_{H\_\text{gues}}, e_{M\_\text{disc}}, e_{M\_\text{gues}}, e_{M\_\text{diff}} \right\},$$
$$SC = \left\{ s_{I\_\theta}, s_{H\_\theta}, s_{M\_\alpha}, s_{H\_\alpha} \right\}. \tag{16}$$

5.2.2. Representation and Prediction Network of LDM-HMI. The learners and learning resources are modeled separately, and the performance prediction network is constructed to complete the learning diagnosis task in the same way as described in Section 4. The detailed steps are shown in steps 19–33 in algorithm 2

6. Experiment Evaluation

In this section, we evaluated the learning performance prediction and interpretability attained by the proposed LDM-ID and LDM-HMI. We also provided some potential

applications of the proposed method in the field of intelligent tutoring systems. The details of the experiments will be discussed below, while the codes for the experiments in this paper are released at https://github.com/CCNUZFW/LDM-ID-HMI.

6.1. Experimental Datasets. To verify the effectiveness of the proposed intelligent learning diagnosis model, three datasets were employed to carry out the experiments. These datasets include two real-world datasets Math11, CL212, and a virtual dataset Synthetic-53, of which Math1 and Synthetic-5 are both open datasets that are widely used in intelligent learning diagnosis tasks and are of moderate size, containing information such as learner-response matrices and question-knowledge point matrices that are commonly used in online intelligent learning diagnosis. CL21 is a smaller dataset of real data collected in four actual university offline classrooms and can be used to validate the applicability of the LDM framework in small offline classrooms. The details of experimental datasets are as follows:

 (i) **Math1:** This dataset comes from several objective and subjective test questions, and answer records are collected by Wu et al. [48] from a college mathematics final test. For the consistency between datasets, this experiment selects the records of the learners' answers to the objective questions, which contains 4209 learners' answers on 15 exercises with 11 knowledge points.

 (ii) **CL21:** This dataset is collected by one of our authors, Dr. Chunyan Zeng, who is serving as the instructor of a C programming language class at the Hubei University of Technology in 2021. The dataset is collected through an intelligent tutoring system, named Chaoxing. This dataset includes the answer records of 93 learners on 36 exercises with 12 knowledge points.

 (iii) **Synthetic-5:** This is a virtual dataset that contains the answer records of 2,000 learners on 50 questions with 5 knowledge points.

Table 1 shows a summary of these three datasets. The Q matrices were labeled by domain experts, and the historical response matrices R were collected by the intelligent tutoring systems. With the availability of the above data, the experiments were carried out with an NVIDIA GeForce GTX 2080Ti GPU in the environment of Python 3.6 and Pytorch 1.9.0.

6.2. The Methods to Be Compared. Three proposals need to be evaluated: the effectiveness of learner and learning resource representation networks, the effectiveness of deep and shallow learning feature fusion, and the effectiveness of attention mechanisms. In addition, the proposed models are compared with 11 state-of-the-art methods such as item response theory (IRT) [26], multidimensional item response theory (MIRT) [10], deterministic input noise and door

model (DINA) [12], Ho-DINA [15], PMF-LDM [27], neural cognitive diagnosis model (NeuralCDM) [45], DeepMFLD [46], and IRR [47], which are listed as follows:

 (i) **IRT** [26]: The IRT model is a psychometrics learning diagnosis model that uses logic functions to model the relationship between the learner's learning state, the distinction, and difficulty of test questions, and the learner's response.

 (ii) **MIRT** [10]: The MIRT model can estimate learners' abilities in multiple dimensions at the same time, and consider the relationship between different dimensions of ability. Therefore, MIRT can estimate multidimensional abilities more effectively.

 (iii) **DINA** [12]: The DINA model is also a psychometrics learning diagnosis model. DINA describes the learner as a multidimensional knowledge point mastering vector and judges the learner's answering results according to the question-knowledge correlation.

 (iv) **Ho-DINA** [15]: Based on the DINA model, the Ho-DINA model assumes that the mastery state of knowledge points is also related to the learner's general ability, which is the learner's high-level ability.

 (v) **PMF-LDM** [27]: PMF-LDM is an intelligent learning diagnosis method based on probability matrix decomposition, which assumes that the learner's response matrix is determined by the inner product of the learner's knowledge proficiency matrix and the knowledge point vector contained in the exercise, and uses the random gradient descent method to constantly diagnose the learning performance.

 (vi) **NeuralCDM** [45]: NeuralCDM is a deep learning-based learning diagnosis method that combines neural networks to learn the complex answering process, projecting learners and exercises as latent space vectors, and using multiple neural layers to model its interaction process to predict learners answering results.

 (vii) **DeepMFLD** [46]: DeepMFLD is a matrix decomposition learning diagnosis method based on deep learning. It obtains the learners' embedding and the exercises' embedding separately by deep matrix decomposition, then fuses them with a deep neural network, and finally predicts the learner's performance by a fully connected layer.

 (viii) **IRR-IRT, IRR-MIRT, IRR-DINA, and IRR-NeuralCDM** [47]: The IRR is a general framework, which introduces pairwise learning into learning diagnosis. Four specific methods based on the IRR framework are used as the baseline: IRR-IRT, IRR-MIRT, IRR-DINA, and IRR-NCDM. The first three methods enhance the traditional psychometric diagnostic models IRT, MIRT, and

TABLE 1: Summary of experimental datasets.

Dataset	Answering records	Number of learners	Knowledge points	Number of exercises
Math1	63135	4209	11	15
Synthetic-5	100000	2000	5	50
CL21	3348	93	12	36

DINA based on the IRR framework, while the NeuralCDM is also used to integrate with the IRR framework to form IRR-NeuralCDM.

6.3. Evaluation Metrics. Since the diagnosis results of the learners' knowledge mastery state cannot be judged objectively and accurately, the diagnosis results are usually used to predict the probability that the learner will answer correctly and compare it with the true value to evaluate the performance of the model. This paper uses AUC (area under curve) and RMSE (root mean square error) as the evaluation indicators of the experimental results. AUC is the area under the indicator curve. The AUC value of 0.5 represents the randomly obtainable score. The higher the AUC score, the more accurate the prediction result. RMSE is used to measure the deviation between the estimated value and the true value. The smaller the value, the closer the mining result is to the true value. The calculation formula is as follows:

$$RMSE(y_i, \widehat{y}_i) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (y_i - \widehat{y}_i)^2}, \qquad (17)$$

where m represents the size of the test data set, y_i is the true value of the student's answer, and $\widehat{y}_i$ is the predicted value of the student's answer.

6.4. Experimental Results and Analysis. In our experiments, a 5-fold cross-validation is performed, and the average AUC and RMSE of five times experiments are used as the evaluation metrics. Furthermore, 80% samples of each dataset are employed as the training set, and 20% of the data are left for the test to check whether the proposed method is generalized well.

6.4.1. Comparison with the State-of-the-Art Methods. For the LDM-ID and LDM-HMI methods, detailed settings are as follows: Sigmoid is used as the activation function of the fully connected layer, and the CNN in the prediction network consists of a convolutional layer, a ReLU activation layer, and a maximum pooling layer, where the convolutional kernel size is 3. The model is parameterized by the Adam optimizer, the learning rate is set to 0.001, and the dropout ratio is set to 0.2. In the learner representation network and learning resource representation network, the learner potential vector and learning resource potential vector dimensions are set to 128 and 64, respectively. In addition, in the baseline models, 3PL-IRT is used for comparison, and the capability latitude value is set to 3 in the MIRT model.

As shown in Table 2, in terms of the accuracy of learning performance prediction, among the traditional psychometric-based learning diagnostic models, MIRT outperformed IRT, DINA, Ho-DINA, and PMF-LDM overall on the three datasets. The performance of NeuralCDM, DeepMFLD, IRR, and the proposed methods LDM-ID and LDM-HMI generally outperformed the psychometric-based learning diagnostic models, which indicates that the deep network-based learning diagnostic methods outperform the traditional psychometric-based learning diagnostic methods. Usually, deep networks have more complex structures and more parameters, and the information they can characterize will be richer, while psychometric-based methods usually have simpler structures, fewer parameters, and limited characterization capabilities.

As shown in Figure 3, the LDM-ID and LDM-HMI methods proposed in this paper outperformed all baseline models in terms of the accuracy of learning performance prediction. On the three datasets, their AUC improved by about 11.45% and RMSE decreased by about 0.07 compared to MIRT on average, their AUC improved by about 8.57% and RMSE decreased by about 0.06 compared to NeuralCDM on average. On the one hand, the cognitive parameters in this paper provide more effective input information for the diagnosis and prediction tasks from the source. On the other hand, this paper constructs a more capable network of learners and learning resources and explores the potential characteristics of learners and learning resources more effectively.

Furthermore, among the two implementations of the LDM framework proposed in this paper, LDM-HMI performs better on the larger-scale datasets Math1 and Synthetic-5, with an AUC about 1.39% higher and an RMSE about 0.007 lower than that of LDM-ID over the two datasets, while on the smaller-scale dataset CL21, LDM-ID performs better, with an AUC about 0.53% higher and an RMSE about 0.002 lower than that of LDM-HMI. This indicates that the LDM-ID method may be more suitable for small data volume scenarios. In other words, under the LDM framework of this paper, the specific implementation can be flexibly adjusted according to the requirements to achieve better adaptation to the actual scenarios.

As shown in Table 3, the LDM-ID and LDM-HMI proposed in this paper are comparable to the deep learning diagnostic method NeuralCDM, DeepMFLD, and IRR in terms of time loss. The prediction time loss of the proposed method on the medium-sized dataset Math1 is comparable to that of the traditional psychometric methods. On the small dataset CL21 and large dataset Synthetic-5, the time loss of the proposed method is one order of magnitude higher compared to the traditional psychometric methods. Overall, the prediction time of the proposed method is about 0.2 ms on a single sample, which is sufficient for practical educational application scenarios.

TABLE 2: Comparison with the state-of-the-art methods in terms of learning performance prediction.

Model	CL21		Math1		Synthetic-5	
	AUC	RMSE	AUC	RMSE	AUC	RMSE
IRT [26]	73.12%	0.4365	68.31%	0.5164	78.52%	0.4824
MIRT [10]	73.24%	0.4352	69.83%	0.4946	79.80%	0.4733
DINA [12]	69.92%	0.4859	64.46%	0.5297	73.80%	0.4968
Ho-DINA [15]	70.37%	0.4723	66.58%	0.5215	72.45%	0.4988
PMF-LDM [27]	63.74%	0.6515	68.03%	0.4884	62.40%	0.6038
NeuralCDM [45]	76.88%	0.4274	72.90%	0.4533	79.77%	0.4756
DeepMFLD [46]	80.40%	0.4226	77.74%	0.4356	76.68%	0.4897
IRR-IRT [47]	71.23%	0.4395	74.10%	0.4505	72.65%	0.4996
IRR-MIRT [47]	81.29%	0.4213	80.72%	0.4231	81.35%	0.4073
IRR-DINA [47]	75.51%	0.4271	81.40%	0.4234	81.03%	0.4087
IRR-NeuralCDM [47]	75.02%	0.4279	78.56%	0.4278	75.85%	0.4917
Proposed LDM-ID	**86.44%**	**0.4105**	81.35%	0.4171	88.02%	0.3667
Proposed LDM-HMI	85.91%	0.4122	**81.89%**	**0.4163**	**90.25%**	**0.3528**

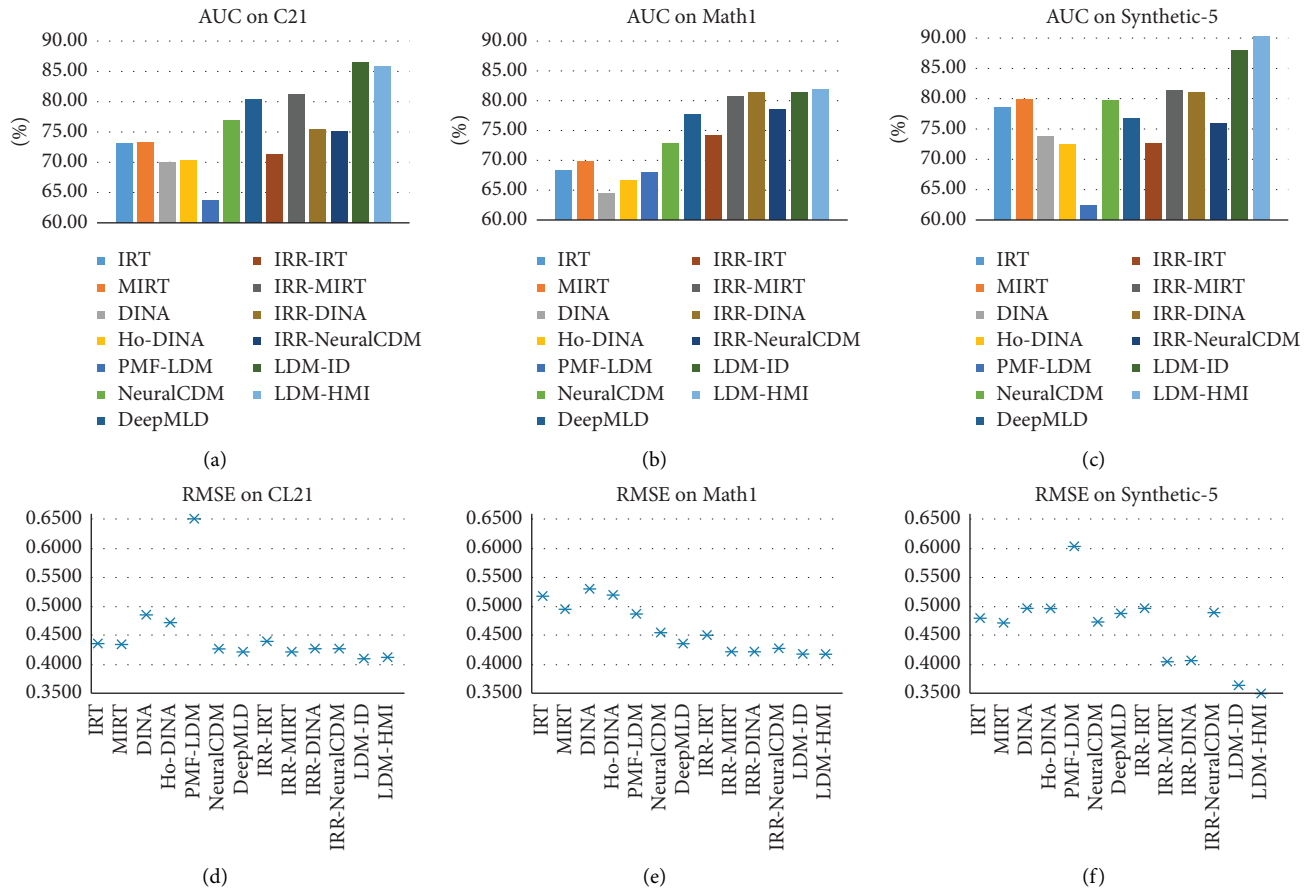The significance of bold values means the best learning performance prediction on each dataset.

FIGURE 3: The proposed methods were compared with 11 state-of-the-art methods. In this paper, 11 state-of-the-art learning diagnosis methods were replicated and experiments were conducted on two publicly available datasets and one self-built dataset, and evaluated using two metrics, AUC and RMSE. A larger AUC indicates higher prediction accuracy, and a smaller RMSE indicates better stability of the methods.

6.4.2. Effectiveness of Learner and Learning Resource Representation Networks. To verify the effectiveness of the learner and learning resource characterization network in the interpretable intelligent learning diagnosis framework proposed in this paper, experiments are conducted to compare the auto-encoder-based characterization network in this paper with the fully connected deep neural network in the traditional approach.

As shown in Table 4, in the LDM-ID and LDM-HMI methods, compared with the fully connected deep neural network, the auto-encoder-based representational network improves the AUC of the model by about 0.85% and 0.9% on

TABLE 3: Comparison with traditional methods in terms of prediction time cost.

Model	Time cost for prediction		
	CL21	Math1	Synthetic-5
IRT [26]	16 ms	203 ms	141 ms
MIRT [10]	90 ms	385 ms	597 ms
DINA [12]	16 ms	31 ms	47 ms
Ho-DINA [15]	57 ms	198 ms	124 ms
PMF-LDM [27]	15 ms	45 ms	61 ms
NeuralCDM [45]	156 ms	216 ms	3359 ms
DeepMFLD [46]	270 ms	352 ms	3822 ms
IRR-IRT [47]	192 ms	401 ms	3591 ms
IRR-MIRT [47]	229 ms	435 ms	3603 ms
IRR-DINA [47]	223 ms	462 ms	3609 ms
IRR-NeuralCDM [47]	218 ms	316 ms	3583 ms
Proposed LDM-ID	187 ms	277 ms	3397 ms
Proposed LDM-HMI	203 ms	303 ms	3374 ms

average over the three data sets, respectively, and reduces the RMSE by about 0.002 and 0.003, respectively. It proves that the auto-encoder-based learner and learning resource representation network is more effective. Among them, the encoder learns to retain as much relevant information in the potential space and discard irrelevant parts, and the decoder learns the potential space information to reconstruct the input, which achieves an effective denoising effect and contributes to a more effective potential feature representation of learners and learning resources.

6.4.3. Effectiveness of Deep and Shallow Learning Feature Fusion.

To verify the effectiveness of the fusion of deep and shallow features in the interpretable intelligent learning diagnosis framework proposed in this paper, experiments are conducted in three forms: using only shallow features, using only deep features, and fusing deep and shallow features.

As shown in Table 5, the fusion of deep and shallow features has the most significant effect on the model effectiveness on the average over the three datasets, with the AUC of the fusion of deep and shallow features improving by about 1.46% and 1.73% and the RMSE decreasing by about 0.011 and 0.016 in the LDM-ID and LDM-HMI approaches relative to the use of only shallow features, relative to the use of only deep features, the AUC of deep and shallow feature fusion is increased by about 0.42% and 0.72%, and the RMSE is reduced by about 0.003 and 0.011. Therefore, the fusion of deep and shallow features helps to improve the performance of the model, because the deep and shallow features provide different valid information from the high-dimensional space and low-dimensional space, respectively, to achieve better prediction results. Also, the fusion of deep and shallow features has better interpretation compared to using only deep features.

6.4.4. Effectiveness of Attention Mechanism.

Based on the deep and shallow feature fusion, considering that each dimension of deep features and shallow features may have different effects on the performance of model prediction, giving different degrees of attention to different dimensional features may produce better results. Therefore, this paper introduces an attention mechanism in the prediction network to further process the deep and shallow fused features, and conducts experimental validation.

As shown in Table 6, after introducing the attention mechanism, the AUC is improved by about 0.92% and 1.42% and the RMSE is reduced by about 0.004 and 0.005 on average in the three datasets in the LDM-ID and LDM-HMI methods, which fully illustrates that the introduction of the attention mechanism can improve the performance of the model. The attention mechanism assigns different weights to the features of different dimensions, which not only helps to improve the accuracy of the prediction results but also can provide explanatory information for the prediction results. It can help us to know the potential connection between the features and the output.

In addition, this paper also verifies the effect of the joint action of the auto-encoder-based learner and learning resource representation network and attention mechanism based on deep and shallow feature fusion, which is conducted in three cases, namely, introducing the auto-encoder-based learner and learning resource representation network alone, introducing the attention mechanism alone, and introducing the auto-encoder-based learner and learning resource representation network and attention mechanism simultaneously. The experiments were conducted.

As shown in Table 7, the model has the best performance when the auto-encoder-based learner and learning resource representation network and the attention mechanism are introduced simultaneously. The AUC improves by 1.02% and 1.82%, and the RMSE decreases by 0.003 and 0.004 in the LDM-ID and LDM-HMI methods on average of the three datasets compared to the auto-encoder-based learner and learning resource representation network alone. Compared with the attention mechanism alone, the AUC of the model increased by 0.96% and 1.29%, and the RMSE decreased by 0.001 and 0.001. This indicates that the model performs best when the auto-encoder-based learner and learning resource representation network and the attention mechanism are combined, which can effectively represent the learners and learning resources while assigning different weights to the features of different dimensions.

6.4.5. Interpretable Analysis and Educational Applications.

The interpretable intelligent learning diagnostic framework proposed in this paper has good interpretability, which comes from three main aspects: cognitive parameters, learner-resource response network, and weighting information of attention mechanisms.

1) Interpretability of cognitive parameters: In the initial part of the model, this paper conducts a preliminary diagnostic evaluation of learners based on different cognitive diagnostic model, and obtains some cognitive parameters that have practical significance in the educational field, such as the multidimensional ability values of learners in the MIRT model, the binary vector of learners' knowledge

TABLE 4: Learning performance prediction experiments for learner and learning resource representation networks.

Model	Network structure	CL21		Math 1		Synthetic-5	
		AUC	RMSE	AUC	RMSE	AUC	RMSE
LDM-ID	DNN	84.78%	0.4179	79.26%	0.4223	86.15%	0.3691
	SAE	**85.32%**	**0.4157**	**80.47%**	**0.4184**	**86.96%**	**0.3682**
LDM-HMI	DNN	81.56%	0.4219	79.74%	0.4215	88.60%	0.3579
	SAE	**82.65%**	**0.4201**	**81.02%**	**0.4168**	**88.93%**	**0.3562**

The significance of bold values means the best learning performance prediction on each dataset.

TABLE 5: Learning performance prediction experiments with a fusion of deep and shallow features.

Model	Representation	CL21		Math 1		Synthetic-5	
		AUC	RMSE	AUC	RMSE	AUC	RMSE
LDM-ID	Shallow feature	83.16%	0.4291	77.41%	0.4388	85.23%	0.3749
	Deep feature	84.31%	0.4206	78.69%	0.4265	85.91%	0.3718
	Shallow + deep feature	**84.78%**	**0.4179**	**79.26%**	**0.4223**	**86.15%**	**0.3691**
LDM-HMI	Shallow feature	79.66%	0.4435	78.33%	0.4397	86.73%	0.3664
	Deep feature	81.04%	0.4359	79.21%	0.4342	87.48%	0.3638
	Shallow + deep feature	**81.56%**	**0.4219**	**79.74%**	**0.4215**	**88.60%**	**0.3579**

The significance of bold values means the best learning performance prediction on each dataset.

TABLE 6: Experiments on the prediction of learning performance of attention mechanism.

Model	Attention setting	CL21		Math 1		Synthetic-5	
		AUC	RMSE	AUC	RMSE	AUC	RMSE
LDM-ID	Without attention	84.78%	0.4179	79.26%	0.4223	86.15%	0.3691
	With attention	**85.39%**	**0.4123**	**79.84%**	**0.4192**	**87.71%**	**0.3669**
LDM-HMI	Without attention	81.56%	0.4219	79.74%	0.4215	88.60%	0.3579
	With attention	**84.70%**	**0.4134**	**80.33%**	**0.4185**	**89.14%**	**0.3537**

The significance of bold values means the best learning performance prediction on each dataset.

TABLE 7: Learning performance prediction experiments for the representation network and attention mechanism.

Model	Attention setting	CL21		Math 1		Synthetic-5	
		AUC	RMSE	AUC	RMSE	AUC	RMSE
LDM-ID	SAE	85.32%	0.4157	80.47%	0.4184	86.96%	0.3682
	Attention	85.39%	0.4123	79.84%	0.4192	87.71%	0.3669
	SAE with attention	**86.44%**	**0.4105**	**81.35%**	**0.4171**	**88.02%**	**0.3667**
LDM-HMI	SAE	82.65%	0.4201	81.02%	0.4168	88.93%	0.3562
	Attention	84.70%	0.4134	80.33%	0.4185	89.14%	0.3537
	SAE with attention	**85.91%**	**0.4122**	**81.89%**	**0.4163**	**90.25%**	**0.3528**

The significance of bold values means the best learning performance prediction on each dataset.

mastery in the DINA model, the difficulty, differentiation, the skipping rate and guessing rate of learning resources. These parameters can complement each other and form a comprehensive and multidimensional description of the learners and learning resources, which can be used as input to the subsequent prediction network to provide a better explanation.

For example, for learners, the proposed interpretable intelligent learning diagnostic framework can be combined with multiperspective cognitive diagnostic theories to evaluate learners' learning states, which can not only understand the differences in learners' multidimensional and fine-grained potential cognition status but also obtain the overall ability level of learners. In addition, it can provide refined tutoring information for subsequent teaching practice. Figure 4(a) shows the knowledge mastery of four randomly

selected learners in the CL21 dataset and their overall proficiency level. It shows that learner $S1$ has a low mastery of knowledge points $K8$ and $K10$. Figure 4(b) shows the overall proficiency level of the four learners, and we can see that the overall proficiency level of learner $S4$ is low, indicating that his foundation may be weak and needs to consolidate his basic knowledge. Figure 5(a) visualizes six cognitive parameters of 20 exercises on the CL21 dataset. This information can be effectively applied in various aspects, such as learner profiling and learner proficiency analysis.

Second, for learning resources, this method can provide comprehensive parameter characterization information of learning resources. Based on these parameters, a more accurate learning resource recommendation service can be provided for learners. For example, when learners have
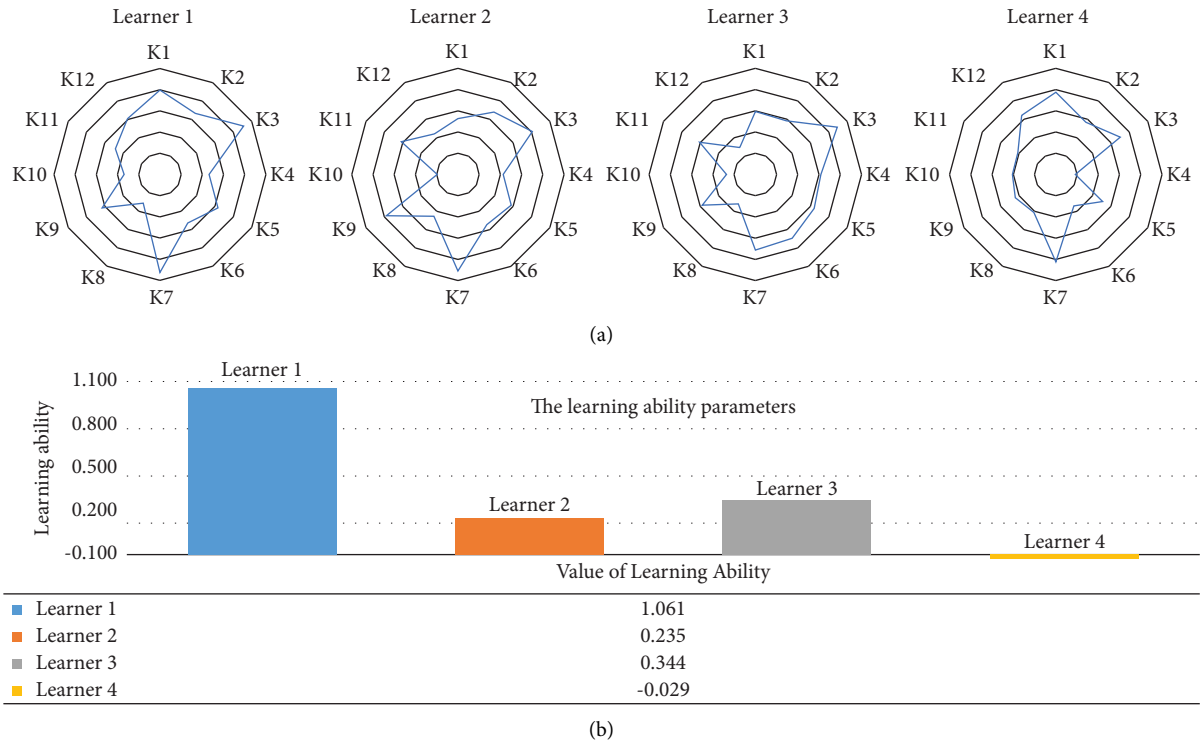
(a)

(b)

FIGURE 4: There are two sets of interpretable cognitive parameters for learners, the knowledge mastery cognitive parameters and the learning ability parameters. For the knowledge mastery cognitive parameter, 0 represents no mastery of the knowledge, and 1 represents full mastery of the knowledge. For the cognitive parameter of learning ability, the range of values is from −1.5 to 1.5, representing low to high learning ability. The higher the learning ability parameters, the higher the probability of answering the exercise correctly. (a) Knowledge Mastery status parameters of learners 1~4 on knowledge points K1~K12. (b) Learning ability parameters of learners 1~4.

(a)

(b)

(c)
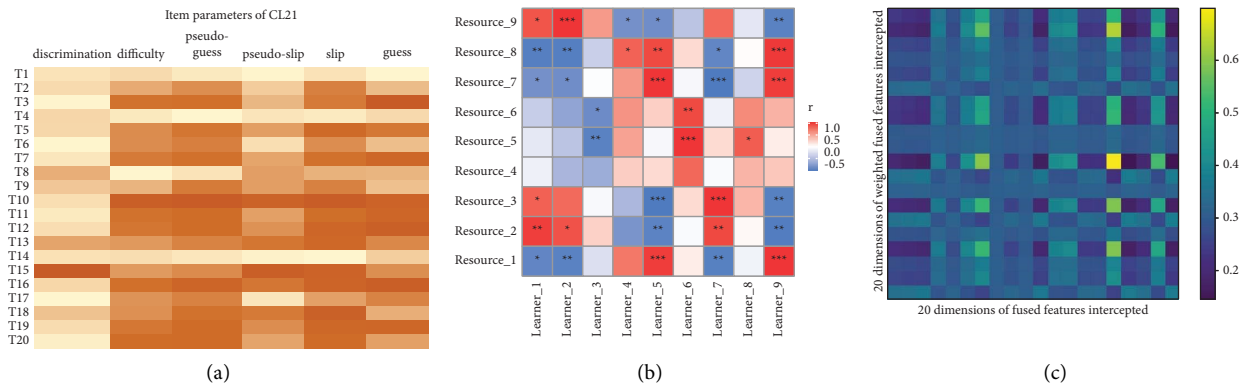
FIGURE 5: As in Figure (a), 6 groups of the relevant cognitive parameters of 20 exercises are intercepted from CL21. The darker the color is, the greater the value is. For Figure (b), the correlation coefficients between learner embedding and learning resource embedding are extracted in the learner-resource response network. For Figure (c), the weight factors of the self-attention mechanism on the learning feature vector are used to speculate the answering results.

a higher mastery of knowledge point *a*, we can recommend more difficult learning resources to help learners master knowledge point a more firmly, or when there is a small difference in the diagnostic results of different learners' mastery of knowledge point *b*, we can select more differentiated learning resources about *b* to provide detailed learning diagnosis for these learners. This rich information about learning resources can assist in practical applications such as educational resource modeling and test question quality analysis.

2) Interpretability of learner-resource response network: After deep characterization of learners and learning resources, this paper constructs a learner-learning resource response network to model the learner-learning resource response process in the actual scenario, simulating the interrelationship between the potential learning state of learners and the response results in the response process, which is more suitable for the actual situation and provides explanations in the modeling method. The heat map of the

Pearson coefficient correlation between the learner characteristics and the learning resource characteristics is constructed, and Figure 5(b) shows that there is a more significant connection between the two. The learner-learning resource response network can be applied to learning resource recommendation, learning interaction analysis, and adaptive learning systems.

3) Interpretability of attention mechanisms: Based on the model's deep and shallow integration features of learners and learning resources, an attention mechanism is introduced to obtain the difference in the degree of influence of features of different dimensions on the actual response results of learners, so that the importance level of features provides retrospective explanatory information for predicting results and provides support for educational cue analysis, educational visualization research, etc.

Second, from the perspective of learning resources, this method can provide comprehensive parameter characterization information for learning resources. Figure 5(c) shows the exercise parameter information of the CL21 data set. Based on these exercise parameters, it is possible to provide learners with more accurate learning resource recommendation services. For example, when the learner has a high degree of mastery of knowledge point a, we can recommend difficult topic resources about knowledge point a to help learners grasp this knowledge point more firmly, or when different learners have a small difference in the diagnosis result of the mastery of knowledge point b, we can make a more detailed learning diagnosis for these learners by selecting exercise resources with a greater degree of discrimination.

7. Conclusion

To address the challenge of balancing diagnostic accuracy and interpretability in traditional learning diagnostic methods, this paper proposed a unified interpretable intelligent learning diagnosis framework that integrates the interpretability of cognitive diagnosis methods and the powerful representation learning ability of deep learning methods, which uses learners' learning records and learning resources related information to perform learning diagnosis and learning performance prediction for learners. Using the diagnostic results of multichannel prior cognitive diagnosis as the basis for embedding representation, a learner-learning resource response network is constructed to simulate the process of learners' answering behaviors. More important influencing factors for prediction results are mined through the self-attention mechanism, making the model more interpretable.

Specifically, in the modeling process, we first constructed cognitive parameter sets to form shallow features based on educational theories from multiple cognitive diagnosis models. Then, we designed a learner representation network and learning resource representation network to mine the deep feature of learners and learning resources, to facilitate the use of shallow features in combination with deep features. In the diagnosis process, a fusion self-attention mechanism based on convolutional neural network architecture is used for learners'

performance prediction, which is helpful in finding the features that have a greater impact on the learner's performance. Experiments on real-world data sets verify the effectiveness of the proposed method, and the analysis of the experimental results can prove to a certain extent the possibility of the method proposed in this paper in real education scenarios.

Our method can obtain the predicted value of the correct probability of the learner's answer, diagnose the learner's overall knowledge point mastery through the characteristic information of the learner and the learning resource, conduct the parameter representation of the learning resource, and provide reference information for the individualized learning of the learner and the intelligent recommendation of the learning resource of the learning platform. Future research can be carried out from the following aspects: (1) in the intelligent learning diagnosis framework of this paper, more learning diagnosis models could be introduced, and different methods and perspectives could be used to model learners and learning resources; (2) it could monitor the learner's learning process more macroscopically and evaluate the learner's learning in periods, and dynamically observe the change process of the learner's knowledge mastery status and learning ability during learning; and (3) it could introduce multimodal data, such as the text, video, and audio information in the learning resources. The learner's behavior information could be used to enrich the input data of the model to obtain more accurate learning evaluation results.

Abbreviations

DKT:	Deep knowledge tracing
DKVMN:	Dynamic key-value memory networks
DINA:	Deterministic input, noisy "and" gate; Ho-DINA, high-order DINA
IRT:	Item response theory
MIRT:	Multidimensional IRT
LDM-ID:	Two-channel learning diagnosis mechanism based on the fusion of IRT and DINA
LDM-HMI:	Three-channel learning diagnosis mechanism based on the fusion of Ho-DINA, MIRT, and IRT
DeepMFLD:	Deep matrix factorization-based learning diagnosis
NeuralCDM:	Neural cognitive diagnosis model
IRR:	Item response ranking
PMF-LDM:	Probabilistic matrix factorization-based learning diagnosis model
SAE:	Stacked auto-encoder.

Data Availability

The data that support this study are available from the corresponding author on request.

Additional Points

Math1: https://staff.ustc.edu.cn/%7Eqiliuql/data/math2015. rar. Synthetic-5: https://github.com/chrispiech/ DeepKnowledgeTracing/tree/master/data/synthetic. CL21:

https://github.com/CCNUZFW/CL21. Codes for LDM-ID and LDM-HMI: https://github.com/CCNUZFW/LDM-ID-HMI

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] T. Hoel and J. Mason, "Standards for smart education – towards a development framework," *Smart Learning Environments*, vol. 5, no. 1, pp. 1–25, 2018.

[2] Z. T. Zhu, M. H. Yu, and P. Riezebos, "A research framework of smart education," *Smart Learning Environments*, vol. 3, no. 1, pp. 1–17, 2016.

[3] W. Gan, Y. Sun, and Y. Sun, "Knowledge structure enhanced graph representation learning model for attentive knowledge tracing," *International Journal of Intelligent Systems*, vol. 37, no. 3, pp. 2012–2045, 2022.

[4] H. Singh and S. J. Miah, "Smart education literature: a theoretical analysis," *Education and Information Technologies*, vol. 25, no. 4, pp. 3299–3328, 2020.

[5] C. M. Chen, Y. L. Hsieh, and S. H. Hsu, "Mining learner profile utilizing association rule for web-based learning diagnosis," *Expert Systems with Applications*, vol. 33, no. 1, pp. 6–22, 2007.

[6] H. C. Chu, G. J. Hwang, and Y. M. Huang, "An enhanced learning diagnosis model based on concept–effect relationships with multiple knowledge levels," *Innovations in Education and Teaching International*, vol. 47, no. 1, pp. 53–67, 2010.

[7] C. Burgos, M. L. Campanario, D. d. l. Peña, J. A. Lara, D. Lizcano, and M. A. Martínez, "Data mining for modeling students' performance: a tutoring action plan to prevent academic dropout," *Computers and Electrical Engineering*, vol. 66, pp. 541–556, 2018.

[8] W. Stout, "Skills diagnosis using IRT-based continuous latent trait models," *Journal of Educational Measurement*, vol. 44, no. 4, pp. 313–324, 2007.

[9] D. J. Hand, W. J. Van der Linden, and R. K. Hambleton, "Handbook of modern item response theory," *Biometrics*, vol. 54, no. 4, p. 1680, 1998.

[10] S. E. Whitely, "Multicomponent latent trait models for ability tests," *Psychometrika*, vol. 45, no. 4, pp. 479–494, 1980.

[11] J. de la Torre and H. Song, "Simultaneous estimation of overall and domain abilities: a higher-order IRT model approach," *Applied Psychological Measurement*, vol. 33, no. 8, pp. 620–639, 2009.

[12] J. De La Torre, "DINA model and parameter estimation: a didactic," *Journal of Educational and Behavioral Statistics*, vol. 34, no. 1, pp. 115–130, 2009.

[13] J. L. Templin and R. A. Henson, "Measurement of psychological disorders using cognitive diagnosis models," *Psychological Methods*, vol. 11, no. 3, pp. 287–305, 2006.

[14] E. Maris, "Estimating multiple classification latent class models," *Psychometrika*, vol. 64, no. 2, pp. 187–212, 1999.

[15] J. De La Torre and J. A. Douglas, "Higher-order latent trait models for cognitive diagnosis," *Psychometrika*, vol. 69, no. 3, pp. 333–353, 2004.

[16] C. Piech, J. Bassen, J. Huang et al., "Deep knowledge tracing," in *Proceedings of the 28th International Conference On Neural Information Processing Systems*, pp. 505–513, Cambridge, MA, USA, December 2015.

[17] Y. Su, Q. Liu, Q. Liu et al., "Exercise-enhanced sequential modeling for student performance prediction," in *Proceedings of the 32 of Thirty Second AAAI Conference On Artificial Intelligence*, pp. 2435–2443, New Orleans, LA, USA, February 2018.

[18] S. Shen, Q. Liu, and E. Chen, "Convolutional knowledge tracing: modeling individualization in student learning process," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1857–1860, Xi'an, China, July 2020.

[19] J. Zhang, X. Shi, I. King, and D. Y. Yeung, "Dynamic key-value memory networks for knowledge tracing," in *Proceedings of the 26th International Conference on World Wide Web*, pp. 765–774, ACM, Perth, Australia, April 2017.

[20] E. Tsutsumi, R. Kinoshita, and M. Ueno, "Deep-IRT with independent student and item networks," in *Proceedings of the 14th International Conference on Educational Data Mining (EDM 2021)*, pp. 510–517, ERIC, Paris, France, June 2021.

[21] H. Nakagawa, Y. Iwasawa, and Y. Matsuo, "Graph-based knowledge tracing: modeling student proficiency using graph neural network," in *Proceedings of the 2019 IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 156–163, IEEE, Thessaloniki, Greece, October 2019.

[22] Y. Yang, J. Shen, and Y. Qu, "GIKT: a graph-based interaction model for knowledge tracing," in *Machine Learning And Knowledge Discovery In DatabasesLecture Notes in Computer Science*, F. Hutter, K. Kersting, J. Lijffijt, and I. Valera, Eds., Springer International Publishing, Midtown Manhattan, NY, USA, 2021.

[23] Y. Lu, D. Wang, Q. Meng, and P. Chen, "Towards interpretable deep learning models for knowledge tracing," in *International Conference on Artificial Intelligence in Education*, pp. 185–190, Springer, Berlin, Germany, 2020.

[24] R. M. Furr, *Psychometrics: An Introduction*, SAGE Publications, Thousand Oaks, CA, USA, 2021.

[25] J. Paulsen and D. S. Valdivia, "Examining cognitive diagnostic modeling in classroom assessment conditions," *The Journal of Experimental Education*, vol. 90, no. 4, pp. 916–933, 2021.

[26] R. K. Hambleton and H. Swaminathan, *Item Response Theory: Principles and Applications*, Springer Science and Business Media, Berlin, Germany, 2013.

[27] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," in *Advances in Neural Information Processing Systems*, pp. 1–8, MIT Press, Cambridge, MA, USA, 2007.

[28] Z. Wang, Z. Wang, C. Zeng, Y. Yu, and X. Wan, "High-quality image compressed sensing and reconstruction with multi-scale dilated convolutional neural network," *Circuits, Systems, and Signal Processing*, vol. 2022, pp. 1–24, 2022.

[29] C. Zeng, J. Ye, Z. Wang, N. Zhao, and M. Wu, "Cascade neural network-based joint sampling and reconstruction for image compressed sensing," *Signal, Image and Video Processing*, vol. 16, no. 1, pp. 47–54, 2022.

[30] Z. Wang, Y. Yang, C. Zeng, S. Kong, S. Feng, and N. Zhao, "Shallow and deep feature fusion for digital audio tampering detection," *EURASIP Journal on Applied Signal Processing*, vol. 2022, no. 1, pp. 69–20, 2022.

[31] C. Zeng, D. Zhu, Z. Wang, M. Wu, W. Xiong, and N. Zhao, "Spatial and temporal learning representation for end-to-end recording device identification," *EURASIP Journal on Applied Signal Processing*, vol. 2021, no. 1, p. 41, 2021.

[32] L. Floridi and M. Chiriatti, "GPT-3: its nature, scope, limits, and consequences," *Minds and Machines*, vol. 30, no. 4, pp. 681–694, 2020.

[33] Z. F. Wang, L. Zhu, Q. S. Min, and C. Y. Zeng, "Double compression detection based on feature fusion," in *Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC)*, pp. 379–384, Ningbo, China, July 2017.

[34] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: a review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, pp. 106587–106639, 2020.

[35] T. Han and Y. F. Li, "Out-of-Distribution etection-assisted trustworthy machinery fault diagnosis approach with uncertainty-aware deep ensembles," *Reliability Engineering and System Safety*, vol. 226, pp. 108648–108714, 2022.

[36] H. Wang, T. Men, and Y. F. Li, "Transformer for high-speed train wheel wear prediction with multiplex local–global temporal fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[37] L. Lyu, Z. Wang, H. Yun, Z. Yang, and Y. Li, "Deep knowledge tracing based on spatial and temporal representation learning for learning performance prediction," *Applied Sciences*, vol. 12, no. 14, pp. 7188–7221, 2022.

[38] X. Xiong, S. Zhao, E. V. Inwegen, and J. Beck, "Going deeper with deep knowledge tracing," in *Proceedings of the International Conference on Educational Data Mining*, pp. 545–550, Raleigh, NC, USA, June 2016.

[39] S. Pandey and G. Karypis, "A self-attentive model for knowledge tracing," pp. 1–6, 2019, https://arxiv.org/abs/1907.06837.

[40] A. Ghosh, N. Heffernan, and A. S. Lan, "Context-aware attentive knowledge tracing," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2330–2339, San Diego, CA, USA, August 2020.

[41] Y. Choi, Y. Lee, and J. Cho, "Towards an appropriate query, key, and value computation for knowledge tracing," in *Proceedings of the Seventh ACM Conference on Learning*, pp. 341–344, New York, NY, USA, August 2020.

[42] W. Wang, T. Liu, L. Chang, T. Gu, and X. Zhao, "Convolutional recurrent neural networks for knowledge tracing," in *Proceedings of the 2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, pp. 287–290, IEEE, Chongqing, China, October 2020.

[43] S. Tong, Q. Liu, W. Huang, Z. Huang, and S. Wang, "Structure-based knowledge tracing: an influence propagation view," in *Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM)*, pp. 541–550, IEEE, Sorrento, Italy, November 2020.

[44] G. Abdelrahman and Q. Wang, "Deep graph memory networks for forgetting-robust knowledge tracing," pp. 1–11, 2022, https://arxiv.org/abs/2108.08105.

[45] F. Wang, Q. Liu, E. Chen et al., "NeuralCD: a general framework for cognitive diagnosis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, pp. 1–16, 2022.

[46] H. J. Xue, X. Dai, J. Zhang, S. Huang, and J. Chen, "Deep matrix factorization models for recommender systems," in *Proceedings of the International Joint Conferences on Artificial Intelligence Organization*, pp. 3203–3209, Melbourne, Australia, August 2017.

[47] S. Tong, Q. Liu, R. Yu et al., "Item response ranking for cognitive diagnosis," in *Proceedings of the International Joint Conferences on Artificial Intelligence Organization*, Montreal, Canada, August 2021.

[48] R. Wu, Q. Liu, Y. Liu et al., "Cognitive modelling for predicting examinee performance," in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pp. 1017–1024, ACM, Buenos Aires, Argentina, July 2015.