electronics

MDPI

Article

An Explainable Student Performance Prediction Method Based on Dual-Level Progressive Classification Belief Rule Base

Jiahao Mai †, Fanxu Wei †, Wei He *, Haolan Huang and Hailong Zhu

The School of Computer Science and Information Engineering, Harbin Normal University, Harbin 150025, China; maijiahao@stu.hrbnu.edu.cn (J.M.); 2024300692@stu.hrbnu.edu.cn (F.W.); 2024300685@stu.hrbnu.edu.cn (H.H.); zhuhailong@hrbnu.edu.cn (H.Z.)

* Correspondence: hewei@hrbnu.edu.cn

† These authors contributed equally to this work.

Abstract: Explainable artificial intelligence (XAI) is crucial in education for making educational technologies more transparent and trustworthy. In the domain of student performance prediction, both the results and the processes need to be recognized by experts, making the requirement for explainability very high. The belief rule base (BRB) is a hybrid-driven method for modeling complex systems that integrates expert knowledge with transparent reasoning processes, thus providing good explainability. However, class imbalances in student grades often lead models to ignore minority samples, resulting in inaccurate assessments. Additionally, BRB models face the challenge of losing explainability during the optimization process. Therefore, an explainable student performance prediction method based on dual-level progressive classification BRB (DLBRB-i) has been proposed. Principal component regression (PCR) is used to select key features, and models are constructed based on selected metrics. The BRB's first layer classifies data broadly, while the second layer refines these classifications for accuracy. By incorporating explainability constraints into the population-based covariance matrix adaptation evolution strategy (P-CMA-ES) optimization process, the explainability of the model is ensured effectively. Finally, empirical analysis using real datasets validates the diagnostic accuracy and explainability of the DLBRB-i model.

Keywords: explainable artificial intelligence; student performance prediction; belief rule base; evidential reasoning; class imbalance

check for updates

Citation: Mai, J.; Wei, F.; He, W.; Huang, H.; Zhu, H. An Explainable Student Performance Prediction Method Based on Dual-Level Progressive Classification Belief Rule Base. *Electronics* **2024**, *13*, 4358. https://doi.org/10.3390/electronics13224358

Academic Editor: Alberto Fernandez Hilario

Received: 11 October 2024
Revised: 1 November 2024
Accepted: 4 November 2024
Published: 6 November 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Student performance prediction is of great significance in the field of education. From the perspective of educational theory, successful prediction of student performance can positively influence the formulation of students' achievement goals [1], enhance students' personal expectancy value [2], and encourage the construction of students' self-efficacy [3]. It generates positive driving forces for students' futures from multiple perspectives.

With artificial intelligence technology bringing about new changes in the field of education, it provides opportunities for predicting students' academic performance and improving their grades. In actual educational scenarios, explainable prediction models are the key to improving education quality [4]. They help teachers understand students' learning patterns and solve problems early [5]. It contributes to the comprehensive improvement and development of subsequent education and learning plans. By adopting advanced data analysis and prediction technologies, the education department is gradually achieving accurate prediction of students' academic performance, thus enabling the formulation of more effective teaching strategies and support plans. Therefore, constructing an accurate and explainable student performance prediction model will enhance the efficiency and success rate of education [6]. Such a model should not only make accurate predictions but also provide valuable insights in the decision-making process.

At the same time, in most educational scenarios, there is usually a class imbalance problem in student performance data [7]. That is, the number of some grade categories is lower than that of other categories. This imbalance can cause the prediction model to be biased towards the majority class, thus reducing the recognition accuracy of the minority class.

The DLBRB—i student performance prediction model proposed in this study provides the following contributions:

- Six explainability criteria are innovatively proposed to enhance the explainability of the model.
- The accuracy of prediction in the case of student class imbalance is improved through a dual-layer progressive classification design.

The remainder of this study is structured as follows: Section 2 is related work, introducing common methods and existing problems in the field of educational prediction, Section 3 summarizes the problems in building student performance prediction models, Section 4 describes the implementation process of the DLBRB-i model, Section 5 verifies the model's effectiveness through case studies, Section 6 is the discussion section, and Section 7 provides a conclusion of the research content.

2. Related Work

Student performance prediction models have made new contributions to three classic educational theories: achievement goal theory, expectancy value theory, and self-efficacy theory.

Achievement goal theory emphasizes the types of goals that students pursue in the learning process, such as mastery goals and performance goals [8]. Mastery goals are related to students' sustained interest, while performance goals are related to academic performance. Mastery goals and performance goals have a significant impact on students' motivation, cognitive participation, and academic performance. Through student prediction models, for those students whose predictions show that they cannot reach their own goals, performance prediction can help them adjust their learning strategies and choose more appropriate learning methods to improve their grades. Performance prediction can also help those students with better grades to maintain their motivation and further challenge themselves by setting higher personal goals instead of merely being satisfied with their existing grade levels.

Expectancy value theory emphasizes the influence of students' expectations of success and the values of tasks on their learning motivation and behavior [9]. This theory holds that students' academic achievements are affected not only by their beliefs in their abilities and expectations of success but also by their evaluations of the values of learning tasks. Performance prediction can help teachers identify those students who need additional support and can be used to adjust teaching strategies to instill correct values in students. At the same time, it can set reasonable expectations for students and help them recognize the importance of learning, thereby enhancing their learning attitude.

Self-efficacy theory mainly focuses on an individual's belief in their ability to complete a task in a specific situation [10]. The concept of self-efficacy emphasizes an individual's expectation of the effect of their own behavior, and this expectation affects their motivation, emotion, and behavior. Self-efficacy theory holds that an individual's self-efficacy sense is the main determinant of behavior change. Student performance prediction can be used as a tool to promote self-efficacy, enhancing students' self-efficacy through multiple dimensions such as positive feedback, goal setting, emotion management, and behavior selection, thereby promoting their academic achievements and personal development.

Current educational prediction models mainly use three methods: the physical model-driven method, data-driven method, and hybrid model method [11].

Physical model-driven approaches to educational prediction are constructed upon a foundation of educational theories and empirical rules. These models leverage expert knowledge and experiences to define their components and interrelationships, ensuring a

theoretical underpinning that guides their structure. Jiao et al. [12] proposed an AI-based predictive model that enhances the quality of online education, particularly in engineering education, thereby improving students' learning outcomes and satisfaction. Liu et al. [13] proposed the WResNeXtGMJ model, which significantly improves the prediction accuracy of students' learning adaptability and academic scores by combining deep learning algorithms and educational data mining techniques. P et al. [14] proposed a machine learning-based model that not only predicts students' academic scores with high accuracy but also identifies the most meaningful features affecting the learning process. Timotheou et al. [15] revealed the various impacts of digital technology in education. Their research emphasizes the positive role of digital technology in education and points out various factors affecting education, providing information for policymakers, educators, and researchers to understand the role of digital technology in education and to guide effective and efficient change. The advantage of this model lies in its theoretical completeness and explainability. However, it may oversimplify the real situation and ignore the interaction of various educational characteristics in actual student performance prediction.

Data-driven methods rely more on learning features, patterns, and trends from a large amount of historical data rather than relying on prior knowledge. Mi [16] proposed a data-driven student learning student performance prediction model based on the radial basis function (RBF) neural network. This model can effectively predict students' learning academic scores by analyzing basic learning situation data. Tenison et al. [17] proposed a model that combines cognitive theory and data-driven methods. This model provides a new perspective on analyzing students' search behavior in simulated digital environments and understanding and evaluating students' digital information literacy. Huang et al. [18] proposed a data-driven course scheduling optimization model. This model integrates data on students' academic records and course arrangements, using advanced optimization techniques such as mixed-integer programming (MIP), to provide a systematic method for higher education institutions to improve graduation rates and optimize educational resource allocation. Castro-Faix et al. [19] proposed a data-driven genetics learning process revision model. This revision model helps educators to more accurately understand students' progress in genetics learning and provides a methodology to test and improve the learning process to better fit students' actual learning experiences and needs. Xue et al. [20] present a data-driven, multidimensional assessment model for English listening and speaking courses in higher education. It combines factor analysis with neural network modeling to build a thorough evaluation system. The model helps teachers to promptly assist students who need help, based on various assessors' feedback, thus improving academic performance. A significant advantage of data-driven models is that they can discover hidden patterns and correlations from a large amount of historical data. However, the challenge with this approach is the need for a large amount of high-quality data for reliable modeling, and the models lack explainability. Academic score datasets are often small samples, and predictive models need to have stronger explainability to be recognized by educational experts.

Hybrid models combine the advantages of physical models and data-driven models. On the one hand, they use educational theory to guide model design; on the other hand, they also use actual data to correct the model, making it closer to the actual situation. Karthikeyan et al. [21] developed a hybrid educational data mining model (HEDM) for assessing student academic score. They highlighted the benefits of combining theory with data in educational data mining. The HEDM model improves the prediction of student outcomes, enhancing educational quality and student success. Nuankaew et al. [22] conducted a systematic literature review, exploring hybrid learning models from the perspectives of educational data mining and learning analytics, emphasizing the potential of educational data mining and learning analytics in improving educational quality and learning outcomes. Raes et al. [23] proposed the concept of a hybrid learning environment, which combined on-site and remote learning elements to adapt to the educational needs during the COVID-19 pandemic. The research results show that hybrid learning models can

provide flexible learning methods, which are crucial for promoting student participation and learning outcomes. Wang et al. [24] conducted a systematic review by reviewing journal articles in the Web of Science (WoS) database, providing a detailed overview of the current status and development trends of hybrid learning in physical education. Ray et al. [25] combined deep learning methods and rule-based methods to propose a student emotion scoring model based on rule extraction. Hybrid models balance the interpretive power of theory and the insights of data, improving prediction accuracy by combining the two.

To address the challenges of class imbalance and explainability in student performance prediction [26], a hybrid model based on the belief rule base (BRB) has been proposed. This model employs IF-THEN rules and belief distribution mechanisms to handle uncertainty and complexity in educational data. It has been widely applied in various fields [27], including fault diagnosis and safety assessment, and has demonstrated excellent nonlinear modeling capabilities in student performance prediction. The BRB model integrates professional knowledge [28], providing explainability and aiding in understanding the decision-making process. Driven by the evidence reasoning (ER) algorithm [29], the BRB model offers transparent reasoning and combines probability distribution with trust measurement, providing a clear probability interpretation for its output. This model has been designed to enhance transparency and credibility in practical applications, particularly in scenarios requiring high decision-making transparency [30].

In educational student performance prediction models, Chen et al. [31] proposed an affective learning assessment model based on evidential reasoning and belief rule base. This method addresses the issues present in existing affective learning assessment models due to uncertainties during data collection and errors in model parameters. Liu et al. [29] introduced a student performance prediction method based on automatically constructing belief rule bases. This approach focuses on modeling in situations where expert knowledge is lacking. They demonstrated the effectiveness of the BRB method in predicting student performance. However, they overlooked two key issues in student performance prediction models based on BRB. The first is the challenge of class imbalance, where the majority of students' grades cluster around an average range, potentially leading to the neglect of minority class samples during model training. This can result in inaccurate predictions across all academic score ranges. The second issue is the loss of explainability during model optimization due to the stochastic nature of optimization algorithms, which can produce different model outputs under the same initial conditions. To address these issues, a student performance prediction method based on an explainable dual-level progressive classification belief rule base has been proposed. This method includes a two-layer BRB structure for coarse and detailed classification, an optimization algorithm with explainability constraints (P-CMA-ES), and its application in actual case studies, demonstrating explainable and accurate prediction effects.

3. Problem Formulation

In the construction of a student performance prediction model based on the DLBRB-i, three critical issues are addressed:

Issue one: How to address the problem of class imbalance in data distribution.

Class imbalance may lead to the model's neglect of minority classes during prediction, thereby restricting the model's applicability. This issue is tackled by implementing a two-tiered BRB model that conducts a preliminary screening to identify groups of closely related and potentially confusable data points. Following this, a fine-tuning classification is applied

to accurately predict the class affiliations of each confusable data group. The modeling process is as follows:

$$\begin{gathered} Layer1: \\ \{x_1, x_2, ..., x_{m'}\} = F_{selecion}(\{x_1, x_2, ..., x_m\}, \rho) \\ y = B(x_1, x_2, ..., x_{m'}, C, \psi) \\ Layer2: \\ Sub_n : \{x_1^n, x_2^n, ..., x_{m'}^n\} = F_{selecion}(\{x_1^n, x_2^n, ..., x_{m'}^n\}, \rho) \\ y^n = B(x_1^n, x_2^n, ..., x_{m'}^n, C, \psi) \end{gathered} \tag{1}$$

where $F_{selecion}(\cdot)$ represents the process of screening the student performance prediction features, ρ represents the parameter used in feature selection, $\{x_1, x_2, ..., x_m\}$ represents a set of all m student educational attributes, and $\{x_1, x_2, ..., x_{m'}\}$ represents the set of m' selected student educational attributes. The modeling process of the $Layer1$ and $Layer2$ models is described. $B(\cdot)$ represents the BRB prediction model, y represents the output of the first layer model, y^n represents the output of each sub-BRB in the second layer, C represents the explainability criterions to be considered in the modeling process, and ψ represents the model parameters.

Issue 2: How to set reasonable explainability criteria for the model.

The randomness of the optimization algorithm during the model's training process introduces uncertainty in the model's behavior, which diminishes explainability. This diminishment makes it challenging for the student performance prediction model to be trusted by educational experts. Consequently, it is imperative that the model, during its initial construction, rule reasoning, and parameter optimization, is developed in accordance with established explainability criteria.

$$Criteria \ : \ \{C_1, \ C_2, \ ..., \ C_n\} \tag{2}$$

where n is the number of explainability criteria.

Issue 3: How to design a model optimization process that ensures explainability.

At the initial stage, the parameters of the DLBRB-i model are often set by domain experts. Given the complex interrelations among various features of student academic score data, these initial parameters may not be fully accurate or optimal, potentially affecting the model's predictive academic score. The specific process proposed in this paper, which incorporates explainability constraints, is described as follows:

$$v_{best} = O(x, y, C, \Omega) \tag{3}$$

where $O(\cdot)$ represents the optimization function, Ω represents the parameter set in optimization, and v_{best} represents the optimal parameters of the model. The explainability criterion c introduced in the optimization process serves as a guideline for explainability in the optimization process.

4. The Student Performance Prediction Model Based on DLBRB-i

In Section 4.1, the conceptual framework of the DLBRB-i model is shown. In Section 4.3, the implementation process of the DLBRB-i model is outlined. In Section 4.4, the construction of the DLBRB-i model is described. In Section 4.5, the reasoning process of the DLBRB-i model is detailed. Finally, in Section 4.6, the optimization process of the DLBRB-i model is presented.

4.1. Conceptual Framework

The conceptual framework of the DLBRB-i model proposed in this study mainly includes three core components: educational theory, empirical work, and algorithms. These parts are interrelated and interact with each other, forming the foundation of the entire model. The specific descriptions are shown in Figure 1.

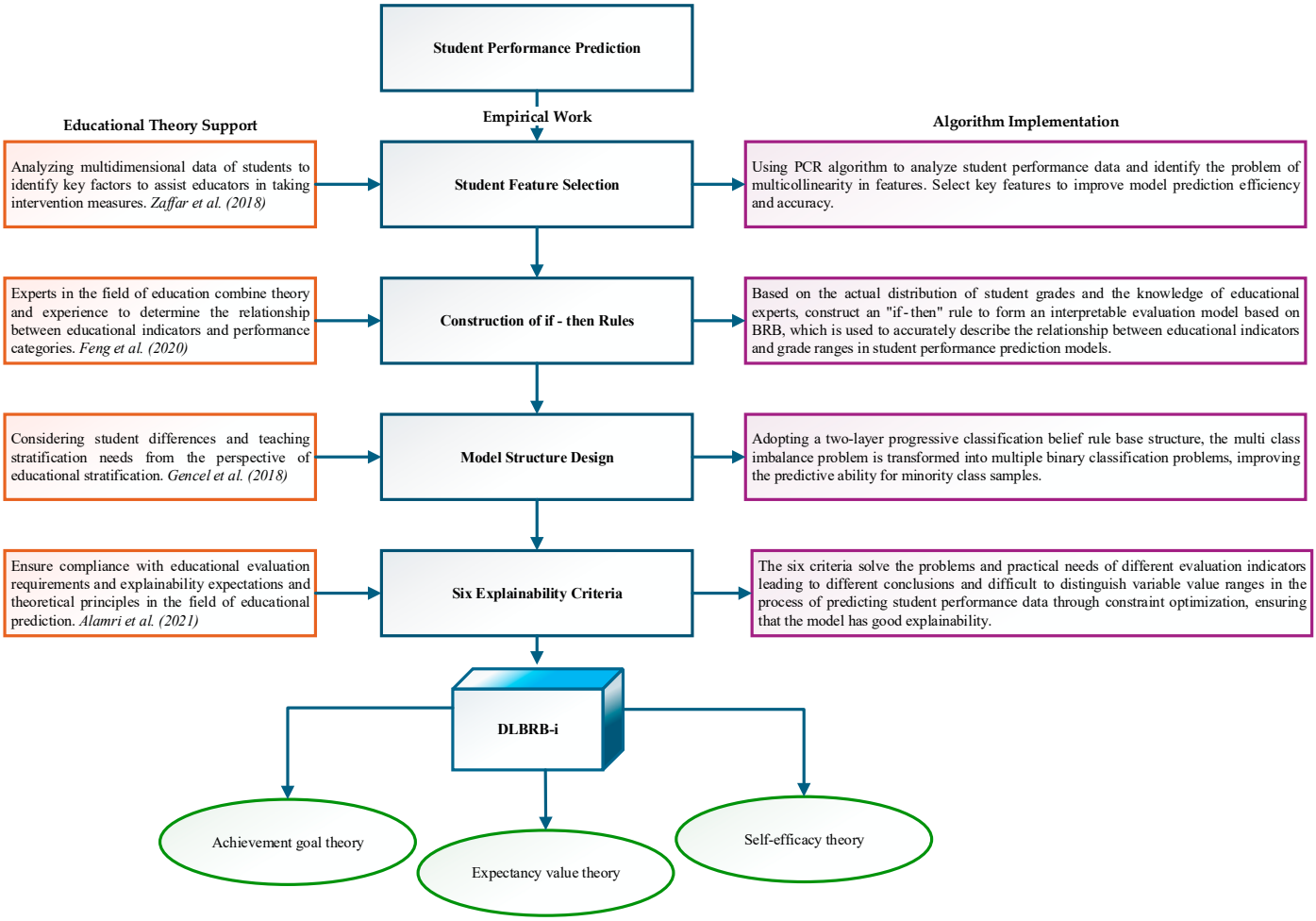

Figure 1. Conceptual framework [32–35].

4.2. Assumptions

The effectiveness of expert knowledge: The paper assumes that the knowledge and experience provided by experts and teachers about student performance is accurate and relevant, and can be used to develop "if-then" rules that effectively reflect the actual situation.

Validity of data: Assuming that the collected student data information is true and reliable.

4.3. Implementation Process of the DLBRB-i Model

The implementation process of the model can be succinctly summarized as follows:

1. Educational features are selected using the PCR algorithm to reduce the model's spatial complexity and enhance prediction efficiency [36].
2. The DLBRB-i model, once the features are selected, is inputted into a dual-layer BRB structure that performs two distinct classifications [37], mitigating the negative impact of class imbalance on student performance prediction.
3. The ER algorithm is applied during the reasoning process to yield transparent and reliable explainable outputs [38].
4. The model's parameters are optimized using a P-CMA-ES algorithm that incorporates explainability constraints, ensuring the model remains explainable throughout the optimization process.

The model offers three main advantages: it increases efficiency by reducing the number of features required, it exhibits robustness by maintaining the good predictive academic scores despite imbalanced data distributions, and it ensures explainability through the use of transparent reasoning methods and optimization techniques that consider explainability,

making the model's predictions explainable and traceable. The modeling process of this model is depicted in Figure 2.

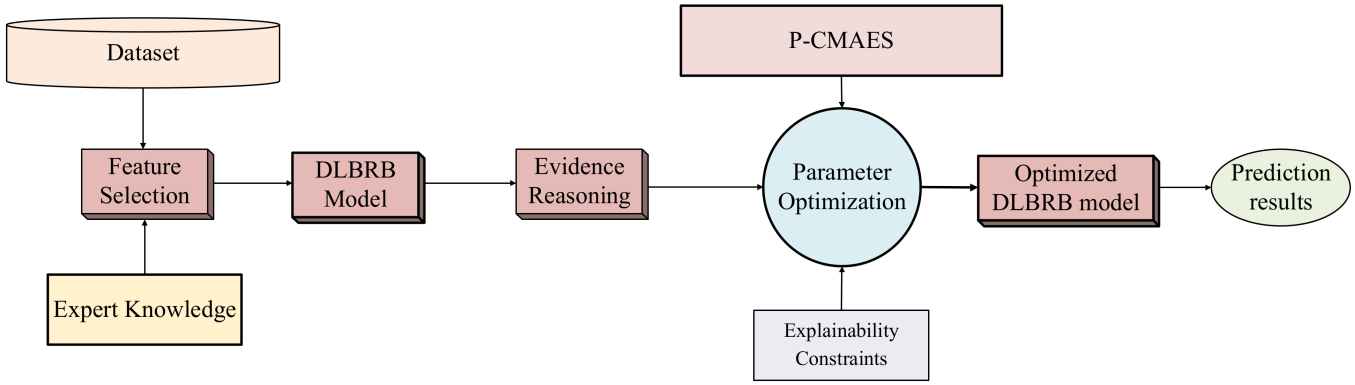

Figure 2. DLBRB-i construction and optimization.

4.4. Specific Implementation Process of DLBRB-i

4.4.1. PCR Feature Selection Process

Principal component analysis, an efficient method for reducing data dimensionality, condenses information from high-dimensional initial variables into a lower-dimensional space. However, an excessive number of variables can result in multicollinearity issues.

To resolve multicollinearity, principal component regression (PCR) was developed as an improvement over principal component analysis. PCR involves conducting multiple regression analyses on the extracted principal components [39], which eliminates multicollinearity and preserves the original information's meaning, leading to clearer and more reliable regression outcomes.

In the domain of student performance prediction, there are correlations between different indicators. Utilizing all types of indicators for predicting student grades increases model complexity and can reduce both efficiency and accuracy. Thus, it is essential to select key indicators while ensuring accuracy and reducing data dimensionality. In PCR, the principal components serve as linear combinations of the original information, reflecting the original feature information to the greatest extent [40]. As the principal components are orthogonal and uncorrelated, they address the issue of feature multicollinearity. The detailed process is as follows:

1. Assume the input data are:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{4}$$

2. To eliminate the impact of data in different dimensions, the data are standardized as follows:

$$\overline{x}_{ij} = \frac{x_{ij} - \frac{1}{n}\sum_{i=1}^{n} x_{ij}}{s_j} (i = 1, 2, \cdots, nj = 1, 2, \cdots, m) \tag{5}$$

$\overline{x}_{ij}$ represents the standardized data of the ith indicator of the jth sample, and s_j is the standard deviation of the jth indicator.

3. Construct the covariance matrix C:

$$C = \frac{1}{n-1} X^T X \tag{6}$$

4. Compute the eigenvalues $\lambda_1, \lambda_1, \cdots, \lambda_m$ and eigenvectors $u_1, u_2, \cdots, u_m$ of the covariance matrix.
5. Sort the eigenvalues and compute the contribution rate q_j for each eigenvalue and the cumulative contribution rate p_c for the first c principal components. Typically, select the number of principal components such that the cumulative contribution rate reaches 90%, but the number can also be determined based on the actual needs of the study while ensuring the cumulative contribution rate is met.

$$q_j = \frac{\lambda_j}{\sum_{i=1}^{m} \lambda_i}, j = 1, 2, \cdots, m \tag{7}$$

$$p_c = \frac{\sum_{k=1}^{c} \lambda_k}{\sum_{k=1}^{m} \lambda_k} c = 1, 2, \cdots m \tag{8}$$

6. Calculate the loadings of each principal component with respect to the explained variable and derive the regression results.

$$h_{ij} = \sqrt{\lambda_i} u_{ij} i = 1, 2, \cdots, m \tag{9}$$

$$g_i = h_{i1}x_1 + h_{i2}x_2 + \cdots + h_{im}x_m \tag{10}$$

7. The calculation process h_{ij} is the variable loading from g_i to X_j. Based on the cumulative contribution rate, the first c principal components are selected for linear regression, and the output of the regression model is:

$$\begin{array}{l} y = d + p_1 g_1 + p_1 g_1 + \cdots + p_c g_c \\ d + p_1(h_{11}x_1 + h_{12}x_2 + \cdots + h_{1m}x_m) + \\ p_2(h_{21}x_1 + h_{22}x_2 + L + h_{2m}x_m) + \cdots + \\ p_c(h_{c1}x_1 + h_{c2}x_2 + L + h_{cm}x_m) \\ = d + (p_1 h_{11} + p_2 h_{21} + L + p_c h_{c1})x_1 + \\ (p_1 h_{12} + p_2 h_{22} + \cdots + p_c h_{c2})x_2 + \cdots + \\ (p_1 h_{1m} + p_2 h_{2m} + \cdots + p_c h_{cm})x_m \end{array} \tag{11}$$

4.4.2. Model Construction

The formulation of "if-then" rules involve decision making with the participation of education field experts, including teachers. These experts are based on their long-term observation and experience of the educational process and student performance. Incorporating support from educational theory [41], potential relationships between various educational indicators and grade categories were identified, thereby defining the conditional part of the rule.

For the determination of belief degrees, experts subjectively evaluate the likelihood of students achieving certain grades under different conditions, which is then quantified. For example, experts might judge that under the condition of "good performance in lab work and excellent scores in midterm exams", the belief degree that a student will achieve an "excellent" grade might be high. This belief degree is then fine-tuned through statistical analysis of the actual grade distribution of students who met these conditions, ensuring its rationality and accuracy. At the same time, there will be extensive discussions and consultations among different education experts to ensure consistency and distinguishability of the belief degrees.

The DLBRB-i model, constructed on a set of "if-then" rules, comprises an explainable evaluation model based on BRB. The relationship between the educational indicators and the grade ranges in the student performance prediction model is as follows:

$$\begin{aligned} &R_k: \text{ if } x_1^k \text{ is } A_1^k \wedge x_2^k \text{ is } A_2^k \wedge \cdots \wedge x_M^k \text{ is } A_M^k, \\ &\text{then } \{(D_1, \beta_1^k), (D_2, \beta_2^k), \cdots, \left(D_N, \beta_N^k\right)\} \ (\textstyle\sum_{i=1}^{N} \beta_i^k \leq 1), \\ &\text{with a rule weight } \theta_k, \\ &\text{and attribute weights } \delta_1^k, \delta_2^k, \cdots, \delta_M^k, \\ &k = 1, \cdots, L. \end{aligned} \tag{12}$$

where R_k denotes the kth rule, $x_1^k, x_2^k, \cdots, x_M^k$ denote the M antecedents used in the kth rule, $A_1^k, A_2^k, \cdots, A_M^k$ denote the corresponding references, and $\delta_1^k, \delta_2^k, \cdots, \delta_M^k$ denote the attribute weights. $D_1, D_2, \cdots, D_N$ are the N consequents in the kth rule and $\beta_1^k, \beta_2^k, \cdots, \beta_N^k$ are the related belief degrees. If $\sum_{i=1}^{N} \beta_i^k \leq 1$, the belief distribution is complete; otherwise, it is incomplete, which means the distribution contains ignorance. θ_k is the rule weight of the kth rule. L is the number of rules.

As illustrated in Figure 3, the DLBRB-i model, a two-layer BRB model, is designed to enhance the precision and granularity of student performance prediction. This model consists of two hierarchical BRB structures:

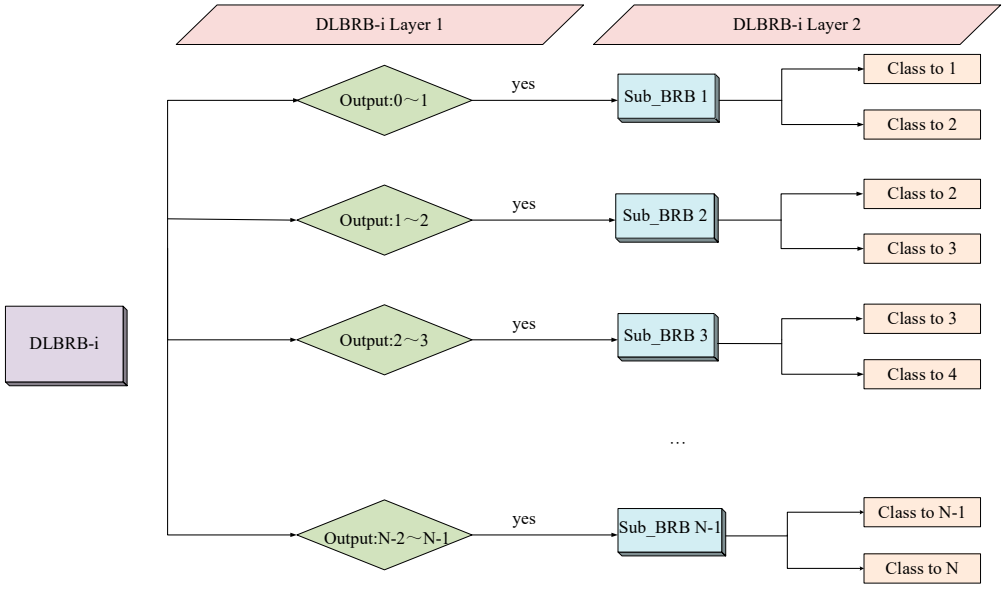

Figure 3. DLBRB-i classification process.

The first layer serves as a coarse classification layer, acting as a multi-class classifier that offers preliminary diagnostic outcomes for a range of student performance categories. Its design ensures comprehensive coverage of potential student performance categories, establishing a holistic student performance assessment framework. The main BRB's output provides a general evaluation of the student performance status, supplying basic information for further detailed analysis.

The second layer is a fine classification layer, made up of several sub-BRBs, each dedicated to two specific student performance categories. This design guarantees that each student performance category is analyzed in detail. Each sub-BRB receives samples within a certain range from the primary classifier. For example, sub-BRB 1 receives samples within the 0–1 range. The output from the main classifier is a preliminary assessment of the student performance status, and sub-BRB 1 further analyzes this to more accurately distinguish between performance level 1 and performance level 2. In this way, each sub-BRB can conduct a more precise diagnosis for specific student performance categories.

From the perspective of practical educational prediction scenarios, this model design demonstrates ample rationality:

(1) Solving the imbalance issue in student grade categories: Student grade data often exhibit category imbalance, where most grades cluster around certain categories while minority categories have fewer samples. A single-layer model may be dominated by majority categories, making it difficult to accurately predict minority categories. In a two-layer model, the first layer performs coarse classification, which can quickly divide the general categories and capture the main distribution trends. The second layer then conducts fine classification based on the results of the first layer, performing binary classification on minority categories and adjacent easily-confused sections. This converts the multi-category imbalance issue into multiple binary classification problems, thereby improving the predictive capability for minority category samples.

(2) Aligning with the hierarchical nature of education: From the perspective of instructional stratification, IE Gencel's [34] research demonstrated that stratified courses have a positive impact on participants' levels of reflective thinking and autonomous learning. The first layer of a two-layer model can perform initial coarse classification to broadly assess the overall learning level of students, similar to initial grouping or stratification in teaching practices. Subsequently, the second layer takes individual differences into account, reflecting the concept of differentiated instruction built upon stratified teaching. This approach better represents the actual situation in the educational process and predicts grades more accurately.

The DLBRB-i model's two-layer design effectively breaks down the general student performance prediction into specific student performance category diagnoses. The primary BRB in the first layer delivers a broad classification of student performance, while the sub-BRBs in the second layer refine these classifications, ensuring accurate and targeted diagnoses. This tiered diagnostic approach not only boosts diagnostic efficiency but also strengthens the comprehension and management of intricate student performance patterns.

4.4.3. Explainability Definition

Cao et al. [42] outlined general explainability criteria, upon which they defined explainability criteria for student performance prediction models. As shown in Figure 4, the six explainability criteria for explainable student performance prediction are precisely defined [43]. The six criteria are as follows:

DLBRB-i model	
Criteria for Explainability	1. Maintain the integrity of evaluation metrics and results 2. Keep variable value ranges distinct 3. Maintain the reasonableness of matching standards 4. Keep the evaluation rules reasonable 5. Maintain the standardization and rationality of information conversion 6. Keep the parameters clear and physical

Figure 4. Six explainability criteria for DLBRB-i.

The integrity of assessment metrics and results should be maintained. In the process of predicting student grades, different evaluation metrics may lead to varying conclusions. Thus, ensuring the consistency and comprehensiveness of the metrics and their outcomes is of great importance [44]. To achieve this, each assessment metric should be assigned at

least two reference values, and at least one confidence rule should be activated to guarantee the quality and reliability of the evaluations.

Variable value ranges should be maintained as distinguishable. When designing models for predicting the student academic score, reference value intervals for each assessment metric should be logically defined, ensuring that no intervals overlap. This way, the meanings represented by different domains can be clearly differentiated, thereby enhancing the explainability of the model to a certain extent.

The normative rationality of the match degree needs to be ensured. In BRB, the sum of the match levels between the reflection values and the current input target evaluation indicator data should be limited to a range between 0 and 1 to guarantee the validity of its practical significance.

The rationality of the evaluation rules must be guaranteed. In designing a model for predicting student grades, logical consistency between the evaluation metrics and the assessment outcomes must be ensured. Based on the authority and reliability of expert knowledge, incorporating expert recommendations to establish initial rules can ensure the rationality of these rules. During the process of model optimization, this rationality should be continued and maintained.

The parameters must possess clear physical meaning. In constructing a model for predicting student grades, the inference process must maintain traceability and explainability [45]. In the BRB model, these parameters include, but are not limited to, rule weights, attribute weights, rule activation levels, and confidence levels. All parameters involved in the evaluation process should have actual physical significance.

The normative rationality of information transformation should be guaranteed. The establishment of this explainability standard aims to require that the inherent characteristics of the original data be preserved intact during the process of information conversion. According to the method proposed by Yang et al. [46], adopting a rule-and-utility-based information transformation strategy can achieve comprehensive and reasonable changes in information while ensuring that the transformation process adheres to norms and maintains the core attributes of the original information.

4.5. Reasoning Process

The DLBRB-i model employs the ER algorithm as its reasoning mechanism, which is a reasoning computational process based on evidence fusion. The explainability of this process is reflected in the causal relationships between steps. When constructing the model, the explainability of the reasoning process refers to the provision of thorough and clear explanations when handling uncertain information. The ER algorithm possesses clear and explicit semantics, capable of fully describing and transforming various forms of uncertain information. During the operation of the model, the weighted belief distributions of evidence are calculated through orthogonal operations and operators. Additionally, the ER algorithm ensures the traceability of the evidence combination process through rigorous probabilistic reasoning, and the reasoning results are presented in the form of belief distributions, thereby ensuring explainability.

Step 1: Initialization, using expert knowledge to formulate initial rules.

Step 2: Calculate the matching degree and unify different forms of information into the same confidence framework. The formula is as follows.

$$\alpha_j^i = \begin{cases} \frac{A_{i(j+1)} - x_i}{A_{i(j+1)} - A_y} & j = k, A_{ij} \le x_i \le A_{i(j+1)} \\ \frac{x_i - A_{i(j+1)}}{A_{i(j+1)} - A_y} & j = k+1 \\ 0 & j = 1, 2, \ldots L, j \ne k, k+1 \end{cases} \tag{13}$$

where a_j^i denotes the match degree under the jth reference value of input attribute x_i.

Step 3: Calculate the activation weight. The formula is as follows:

$$\begin{aligned} w_k &= \frac{\varepsilon_k \prod_{i=1}^{n} (\alpha_j^i)^{\overline{\theta_i}}}{\sum_{l=1}^{L} \varepsilon_l \prod_{i=1}^{n} (\alpha_j^i)^{\overline{\theta_i}}}, \\ \overline{\theta_i} &= \frac{\theta_i}{\max \theta_i} (i = 1, 2, ..., n) \end{aligned} \tag{14}$$

where w_k denotes the activation weight of the kth rule, ε_k represents the weight of the kth rule, and θ_i is the weight of the attribute.

Step 4: Combine different rules based on the ER parsing algorithm. The calculation process of the combined confidence is as follows:

$$\beta_m = \frac{\mu[\prod_{k=1}^{L} (w_k \beta_{m,k} + 1 - w_k \sum_{j=1}^{M} \beta_{j,k}) - \prod_{k=1}^{L} (1 - w_k \sum_{j=1}^{M} \beta_{j,k})]}{1 - \mu[\prod_{k=1}^{L} (1 - w_k)]} \tag{15}$$

$$\mu = [\sum_{n=1}^{N} \prod_{k=1}^{L} (w_k \beta_{m,k} + 1 - w_k \sum_{j=1}^{M} \beta_{j,k}) - (M-1) \prod_{k=1}^{L} (1 - w_k \sum_{j=1}^{M} \beta_{j,k})]^{-1} \tag{16}$$

where β_m represents the confidence level D_m associated with evaluation grade $(0 \leq \beta_m \leq 1)$, and $\beta_{m,k}$ denotes the confidence level corresponding to the mth grade of the kth rule.

Step 5: Finally, output the utility value, the formula is as follows:

$$y_h = \sum_{m=1}^{M} \mu(D_m) \beta_m \tag{17}$$

where y_h represents the output value of the hth layer model, and $\mu(D_m)$ denotes the utility value at reference point β_m.

4.6. Optimization Process

The initial parameters of the DLBRB-i model can be set by experts based on their domain experience; however, given the complexity of the actual system structure, parameter setting is subjective. To reduce the prediction error of the HBRB model, the model parameters are optimized using labeled samples. Figure 5 illustrates the specific optimization process. In this paper, an objective optimization function based on the mean squared error (MSE) is constructed, described as follows:

$$\begin{aligned} &\min MSE(\varepsilon_k, \theta_n, \beta_{m,k}) \\ &s.t. \sum^{M} \beta_{m,k} = 1 \\ &0 \leq \beta_{m,k} \leq 1, k = 1, 2, \cdots, L \end{aligned} \tag{18}$$

The model falls within the category of nonlinear optimization with constraints. Therefore, this paper employs the population-based covariance matrix adaptation evolution strategy (P-CMA-ES) algorithm for optimization, which is a method for solving nonlinear and nonconvex optimization problems. This algorithm is derived from the covariance matrix adaptation evolution strategy (CMA-ES) algorithm but differs in that it reduces algorithm complexity and enhances the effectiveness of the optimization process. P-CMA-ES is an intelligent optimization algorithm with global optimization capabilities, capable of generating an initial population guided by expert knowledge and then iteratively selecting the optimal solution based on the constraint conditions. It has significant advantages in small sample sizes and nonlinear optimization.

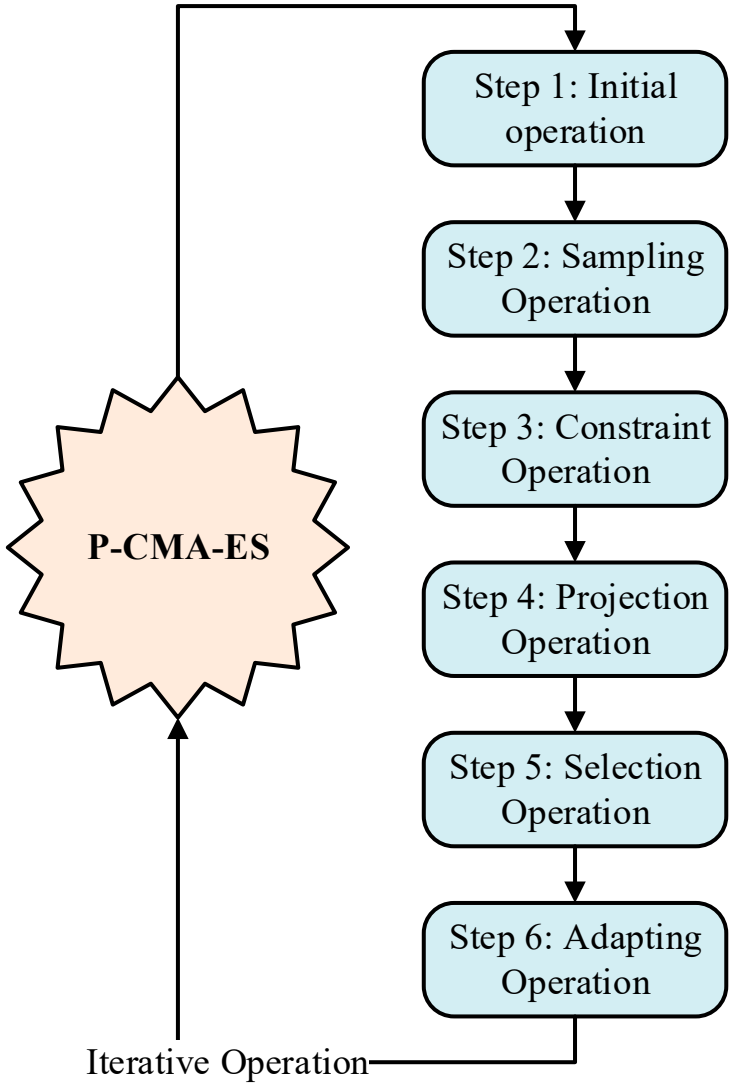

Figure 5. P-CMA-ES optimization process.

Step 1: Define the dimension of the problem and initialize the parameter set.

Step 2: Perform sampling operations and define the expected value. Generate the covariance matrix of the corresponding population by using the normal distribution. The formula is shown below:

$$\eta_k^{g+1} \sim mean^g + \wp^g N(0, C^g)(k = 1 \cdots \lambda) \tag{19}$$

where η_i^{g+1} represents the ith offspring of the $g+1$ generation, $mean^g$ denotes the expected value of the g generation. $\wp^g$ is the update step size, and C^g is the covariance matrix.

Step 3: Based on the projection algorithm, the solution is mapped to the hyperplane and the parameters are constrained. The formula is as follows:

Explainability constraint 1:

$$mean^{(g)} = \begin{cases} EK, & if \quad g = 1 \\ mean^{(g)}, & if \quad g \neq 1 \end{cases} \tag{20}$$

This constraint integrates expert knowledge into the initial population of the model, serving as a guiding mechanism during the optimization process. This not only enhances the effectiveness of the optimization but also brings the starting point of the optimization closer to the optimal solution, thereby accelerating convergence towards the best possible

outcome. Moreover, by leveraging expert insights, this constraint helps to refine the initial conditions, ensuring that the model's optimization trajectory is more aligned with real-world expectations and practical applicability.

Explainability constraint 2:

$$\begin{array}{l} \eta_{lp} \le \eta \le \eta_{up} : \\ \theta_{lp_k} \le \theta_k \le \theta_{up_k}\ k \in \{1, 2, ..., L\}. \\ \delta_{lp_i} \le \delta_i \le \delta_{up_i}\ \mathrm{n} \in \{1, ..., \mathrm{N}\}. \\ \beta_{lp_{k,n}} \le \beta_{k,n} \le \beta_{up_{k,n}}\ i, m \in \{1, 2, ..., T_k\} \end{array} \tag{21}$$

This constraint ensures that the parameters retain their physical significance throughout the optimization process, thereby maintaining the model's explainability.

Explainability constraint 3:

$$\rho(\eta, EK) = \sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2 + \cdots + (x_n - x_n')^2}\rho(x_n, x_n') \le d \tag{22}$$

This constraint ensures that the optimization process of the model is based on the local judgment of the expert knowledge, which further makes the optimized parameters of the model more similar to the expert knowledge.

$$\begin{array}{l} \eta_i^{g+1}(1 + V_n \times (j-1) : V_n \times j) = \\ \eta_i^{g+1}(1 + V_n \times (j-1) : V_n \times j) - P_n^T \times (P_n \times P_n^T)^{-1} \\ \times \eta_i^{g+1}(1 + V_n \times (j-1) : V_n \times j) \times P_n \end{array} \tag{23}$$

$P = [1 \cdots 1]_{1 \times N}$ represents the parameter vectors. $V_n = (1, \cdots, R)$ denotes the number of constrained variables, and $j = (1, \cdots, R+1)$ is the number of constraints.

Step 4: Select the subpopulation that meets the constraints and update the target mean. The formula is described as follows:

$$mean^{g+1} = \sum_{i=1}^{\alpha} \varpi_i \eta_i^{g+1} \tag{24}$$

ϖ_i represents the weight of the optimal solution, and all weights sum to 1.

Step 5: Update the covariance matrix of the sub-population and gradually approach the optimal solution. The process is described as follows:

$$C^{g+1} = (1 - b_1 - b_\alpha)C^g + b_1 J^{g+1}(J^{g+1})^T + b_\alpha \sum_{i=1}^{\alpha} \varpi_i (\frac{\eta_i^{g+1} - mean^g}{\wp^g})(\frac{\eta_i^{g+1} - mean^g}{\wp^g})^T \tag{25}$$

$$J_s^{g+1} = (1 - b_s)J_s^g + \sqrt{b_s(2 - b_s)(\sum_{i=1}^{\alpha} \varpi_i^2)^{-1}} \times \frac{mean^{g+1} - mean^g}{\wp^g} \tag{26}$$

$$\wp^{g+1} = \wp^g \exp(\frac{b_\wp}{d_\wp}(\frac{\parallel J_\wp^{g+1} \parallel}{E \parallel N(0, \mathrm{I}) \parallel} - 1)) \tag{27}$$

where b_1 and b_n represent the learning rates, J^g denotes the evolutionary path at the gth generation, $d_\wp$ is the damping coefficient, $E \parallel N(0, \mathrm{I}) \parallel$ represents the expectation of the $N(0, \mathrm{I})$ normal distribution, and I is the identity matrix.

Step 6: Determine the appropriate parameter set. If it does not meet the conditions for the optimal solution, execute step 1 to update the parameter vector until it meets the conditions.

5. Case Study

In Section 5.1, the dataset utilized in this study is described. Section 5.2 details the construction of the DLBRB-i model on the given dataset. The predictive academic score of the model is validated in Section 5.3. The explainability of the model is confirmed in Section 5.4. Comparative experiments are conducted in Section 5.5 to substantiate the model's merits.

5.1. Dataset Background

The C programming language course, a staple in computer science curricula, serves as the subject for student performance prediction. Such predictions facilitate educators in gauging the impact of their teaching methods and enable students to evaluate their learning progress, thereby informing enhancements in future educational interactions. This research compiles demographic data, scores from different phases of the C language course in the first semester, and educational indicators from students' regions of origin for 356 students across three cohorts—2021, 2022, and 2023—from the Software Engineering program at a teacher training university in Northeast China. To safeguard participant privacy, all data collected have been de-identified, ensuring no disclosure of information that could identify individuals. Student grades are classified into five categories: Failing (F), Poor (P), Middle (M), Good (G), and Excellent (E). The modeling and testing of the model are implemented in MatLab.

5.2. DLBRB-i Student Performance Prediction Modeling

To enhance the accuracy of student performance prediction models, this study employs principal component analysis (PCA) to reduce the dimensionality of objective features. This method not only retains most of the information contained in the original objective features but also yields several uncorrelated principal components. Following the execution of PCA and the selection of principal components, a PCR regression equation is constructed. In the field of education, Unhawa et al. [47] transformed students' cognitive and non-cognitive abilities and traditional academic background lag measures into a set of principal components (PCs) to predict student performance. However, since these components are linear combinations of the original variables and do not correspond to any specific original variable or concept, they lack direct physical meaning. Therefore, this study chose to maintain the practical significance of the features entering the predictive model by selecting the original variables most related to the principal components as key features. This approach to feature selection is more aligned with the practical scenarios of educational predictions.

Specifically, PCA is performed using ten objective features from the student data samples. Table 1 describes the representation of these ten features. After orthogonal transformation, ten linearly independent variables are obtained. Principal component analysis is a statistical method that reduces the dimensionality of variables (i.e., the ten objective features in this paper), resulting in principal components that reflect most of the information of the original variables and are linear combinations of them. In PCA, each component is a weighted sum of all original variables (with different weighting coefficients for each component). The contribution rates of each principal component are shown in Table 1.

Table 1. Feature representation methods.

Symbol Representation	Feature Meaning
X_1	Number of masters
X_2	Experiment
X_3	Group assignment
X_4	Term 2
X_5	Midterm
X_6	Place of origin
X_7	Term 3
X_8	Term 1
X_9	Ethnic group
X_{10}	College entrance examination

As can be observed from Table 2, the cumulative contribution rate of Component 1, Component 2, Component 3, and Component 4 has reached 95.4813%. The remaining components contain less information; therefore, this paper selects Components 1, 2, 3, and 4 for regression fitting analysis. Table 3 presents the composition of these principal components.

Table 2. Principal component analysis results.

Principal Components	Cumulative Contribution Rate
PC_1	57.2828
PC_2	75.4752
PC_3	89.2337
PC_4	95.4813
PC_5	97.1365
PC_6	98.7118
PC_7	99.3277
PC_8	99.7095
PC_9	99.9999
PC_{10}	100.0000

Table 3. Composition of principal components.

Features	PC_1	PC_2	PC_3	PC_4
X_1	−0.1006	0.5033	0.4692	0.0664
X_2	0.3679	0.9023	0.3908	0.5578
X_3	0.4452	0.1062	0.2262	0.4131
X_4	0.1762	0.2748	0.4890	0.2288
X_5	0.5247	0.6157	0.4836	0.0439
X_6	0.2790	−0.2547	0.6676	0.4296
X_7	0.7806	0.3435	0.4462	0.1149
X_8	0.3247	0.1997	0.0708	0.0809
X_9	0.2100	−0.1768	0.4451	−0.1631
X_{10}	0.3822	0.2312	0.4539	−0.3378

It can be seen from Table 3 that the expression formulas of the four components are

$$\begin{aligned} PC_1 = {} & -0.1006 * X_1 + 0.3679 * X_2 + 0.4452 * X_3 + \\ & 0.1762 * X_4 + 0.5247 * X_5 + 0.2790 * X_6 + 0.7806 * X_7 + \\ & 0.3247 * X_8 + 0.2100 * X_9 + 0.3822 * X_{10} \end{aligned} \tag{28}$$

$$\begin{aligned} PC_2 = {} & 0.5033 * X_1 + 0.9023 * X_2 + 0.1062 * X_3 + \\ & 0.2748 * X_4 + 0.6157 * X_5 - 0.2547 * X_6 + 0.3435 * X_7 + \\ & 0.1997 * X_8 - 0.1768 * X_9 + 0.2312 * X_{10} \end{aligned} \tag{29}$$

$$\begin{aligned} PC_3 = {} & 0.4692 * X_1 + 0.3908 * X_2 + 0.2262 * X_3 + \\ & 0.4890 * X_4 + 0.4836 * X_5 + 0.6676 * X_6 + 0.4462 * X_7 + \\ & 0.0708 * X_8 + 0.4451 * X_9 + 0.4539 * X_{10} \end{aligned} \tag{30}$$

$$\begin{aligned} PC_4 = {} & 0.0664 * X_1 + 0.5578 * X_2 + 0.4131 * X_3 + \\ & 0.2288 * X_4 + 0.0439 * X_5 + 0.4296 * X_6 + 0.1149 * X_7 + \\ & 0.0809 * X_8 - 0.1631 * X_9 - 0.3378 * X_{10} \end{aligned} \tag{31}$$

The linear regression method is used for fitting, and the fitting equation is:

$$y = 2.5955 + 0.1613 * PC_1 + 0.3562 * PC_2 - 0.0224 * PC_3 + 0.2469 * PC_4 \tag{32}$$

Equations (28) to (32) are substituted into Equation (32) to derive the PCR regression Equation (33) for predicting student academic performance.

$$\begin{aligned} y = {} & 2.5955 + 0.1689 * X_1 + 0.5097 * X_2 + 0.2066 * X_3 + \\ & 0.1718 * X_4 + 0.3039 * X_5 + 0.0454 * X_6 + 0.2666 * X_7 + \\ & 0.1419 * X_8 + 0.0012 * X_9 + 0.0504 * X_{10} \end{aligned} \tag{33}$$

The PCR results indicate that the features X_2 (Experiment) and X_5(Midterm), which contribute the most, are selected as inputs for the DLBRB-i model. Each attribute is assigned six reference points: Very Low (VL), Low (L), Medium (M), Moderately High (LH), High (H), and Very High (VH). The exact values of these reference points are determined by experts based on the specific characteristics of student grades, as illustrated in Table 4. The precision and complexity of the student performance prediction model are affected by the number of reference points. Furthermore, Table 5 includes reference values for five types of student grade intervals.

Table 4. Input reference value configuration.

Attribute	Attribute Weight	VL	L	M	LH	H	VH
X_2	0.7	100	110	119	132	141	150
X_5	0.3	40	59	73	82	91	101

Table 5. Output reference value configuration.

Reference Points	Failing (F)	Poor (P)	Middle (M)	Good (G)	Excellent (E)
Reference value	0	1	2	3	4

Figures 6 and 7 respectively show the student count by grade ranges and the data distribution of Feature 1, Feature 2, and Output. Figures 6 and 7 illustrate that the dataset has a pronounced class imbalance issue, with the quantity of samples for class 3 notably surpassing those of the other classes. This imbalance may result in biases in the predictions made by conventional machine learning models, potentially skewing towards the majority class.

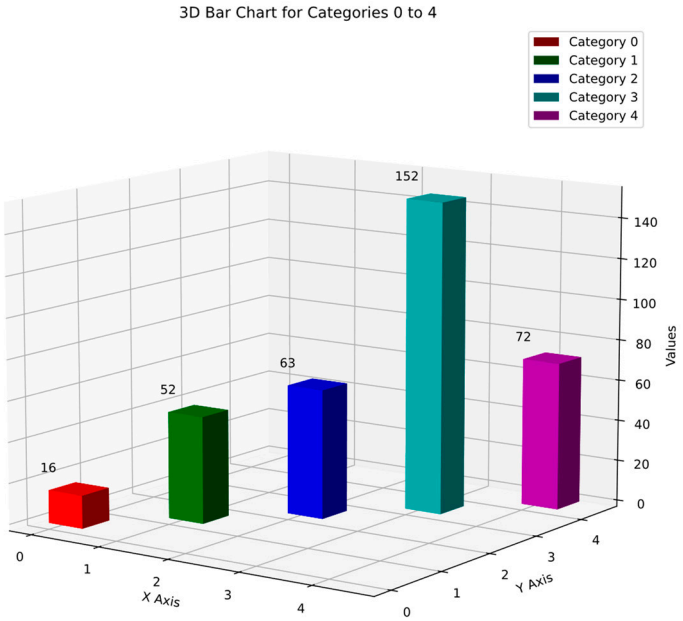

Figure 6. Student count by grade ranges.

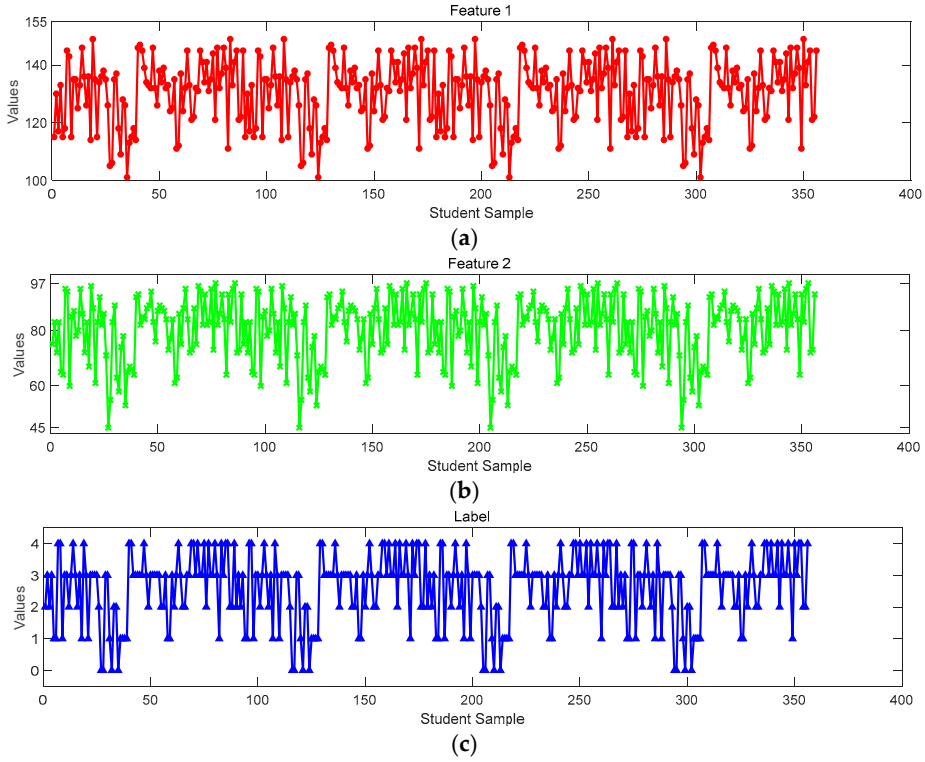

Figure 7. Data distribution of Feature 1 (**a**), Feature 2 (**b**), and output (**c**).

The first layer of the DLBRB-i student performance prediction model classifies all input student academic score data. It processes the entire dataset, predicting all categories, and provides a foundational classification result for subsequent sub-BRBs. Identified are minority categories through preliminary classification, and data groups for each adjacent category are categorized and input into each subsequent second-layer BRB, enhancing the predictive ability for minority categories.

The second layer of the DLBRB-i model conducts more detailed predictions for specific pairs of academic score categories. Following the preliminary classification results from the first layer, each sub-BRB in the second layer focuses on more refined diagnostics for specific pairs of student academic scores. For example, one sub-BRB is dedicated to distinguishing between student academic scores categorized as 0 and 1, while another focuses on distinguishing between categories 1 and 2.

Class imbalance is addressed by transforming the multi-class problem into multiple binary classification problems in the second layer. This helps to alleviate the imbalance and enhances the diagnostic academic score for minority categories.

In summary, the second layer of the DLBRB-i model effectively solves the class imbalance issue and improves prediction accuracy through refined binary classification based on the first layer. This makes the model more robust and reliable in practical applications.

5.3. Experimental Results Analysis

This study evaluates the academic score of the model based on the following four evaluation criteria:

1. ACC

The accuracy (ACC) metric is a comprehensive indicator that provides an overall view of a model's performance across various categories. In the context of the DLBRB-i model, which is designed for predicting student performance, the ACC metric would measure the model's diagnostic precision across all samples, regardless of the specific academic categories they represent. This metric is crucial for evaluating the model's effectiveness in

multiclass diagnostic scenarios within an educational context. Computationally, ACC is derived as follows:

$$ACC = \frac{CANum}{ANum} \times 100 \tag{34}$$

where ACC represents the overall accuracy of the model, $CANum$ denotes the number of samples for which the model produced correct output results, and $ANum$ signifies the total number of samples evaluated.

2. Recall

The recall rate of the student performance prediction model refers to the proportion of positive instances successfully identified by the model among all true positive instances. In other words, the recall rate measures the model's ability to correctly identify all true positive instances. The formula for the recall rate is as follows:

$$Recall = \frac{TP}{TP + FN} \tag{35}$$

In this formula, TP (true positive) represents the number of true positive cases successfully recognized by the model, which is the number of samples correctly judged as positive and actually positive by the model, and FN (false negative) represents the number of true positive cases that the model failed to recognize, which is the number of samples incorrectly judged as negative but actually positive by the model.

3. Precision

Precision is another important metric for evaluating the academic score of student performance prediction model and is typically used to measure the proportion of true positive instances among all instances predicted as positive. In other words, the precision ratio measures the accuracy of the model when predicting positive instances. Precision can be defined as the ratio of the number of true positive instances to the total number of instances predicted as positive. The formula for precision is:

$$Precision = \frac{TP}{TP + FP} \tag{36}$$

In this formula, TP (true positive) represents the number of true positive cases successfully recognized by the model, FN (false negative) represents the number of true positive cases that the model failed to recognize, and FP (false positive) refers to the number of samples in which a model incorrectly predicts negative cases as positive cases.

4. F1 Score

The F1 score is a metric that comprehensively evaluates the academic score of a model by combining precision and recall metrics. The F1 score is the harmonic mean of precision and recall, providing a more comprehensive assessment of the model's academic score in positive instance prediction. The formula for calculating the F1 score is:

$$F1 = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \tag{37}$$

The F1 score ranges between 0 and 1, with values closer to 1 indicating a better academic score in balancing precision and recall. The F1 score is suitable for tasks that require a balance between the model's precision and recall, as it considers the balance between identifying true positive instances and avoiding false positive instances. Through the F1 score, we can more comprehensively evaluate the overall academic score of the model.

In this study, instances are randomly sampled from five different types of student academic score, and the training and test sets are divided from 356 student data in a ratio of 7:3. The experiment is conducted using the attribute reference values and result reference

values set in Section 4.1, with the iteration parameter of the P-CMA-ES optimization algorithm set to 100 times.

The prediction results of the five types of student academic scores in the BRB are shown in Figure 8, with an accuracy rate of 81.31%. The accuracy in the minority class of category 0 is only 25%, while in the minority classes of categories 1 and 2, the accuracy is around 78%. This demonstrates that traditional models find it difficult to make accurate judgments on minority class samples.

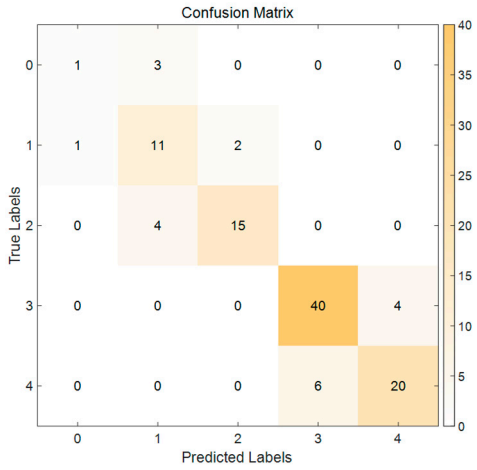

Figure 8. Prediction results of BRB.

As indicated by Figure 9 and the confusion matrix explanation of the four sub-BRBs in Figure 10, the diagnostic accuracy of the model is improved from 81.31% to 98.13% after the refinement by the second layer BRB. The diagnostic precision in the minority classes has been significantly enhanced. This indicates that the method has addressed the issue of poor prediction accuracy for minority classes in ordinary BRBs, improved the prediction accuracy of minority classes, and increased the overall accuracy, demonstrating the effectiveness of the model.

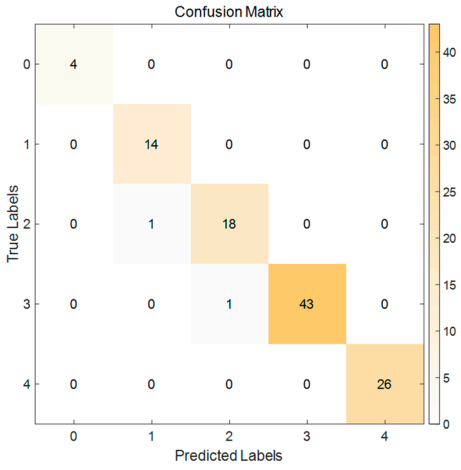

Figure 9. Prediction results of DLBRB-i.

The academic score of the two models is evaluated based on four criteria, as depicted in Table 6. The DLBRB-i model demonstrates the superior academic score across all four metrics, showing a considerable enhancement over the outcomes from the first layer of DLBRB-i. This indicates that the model excels in classification tasks, maintaining high accuracy and recall rates for both minority and majority classes, even when dealing with imbalanced datasets.

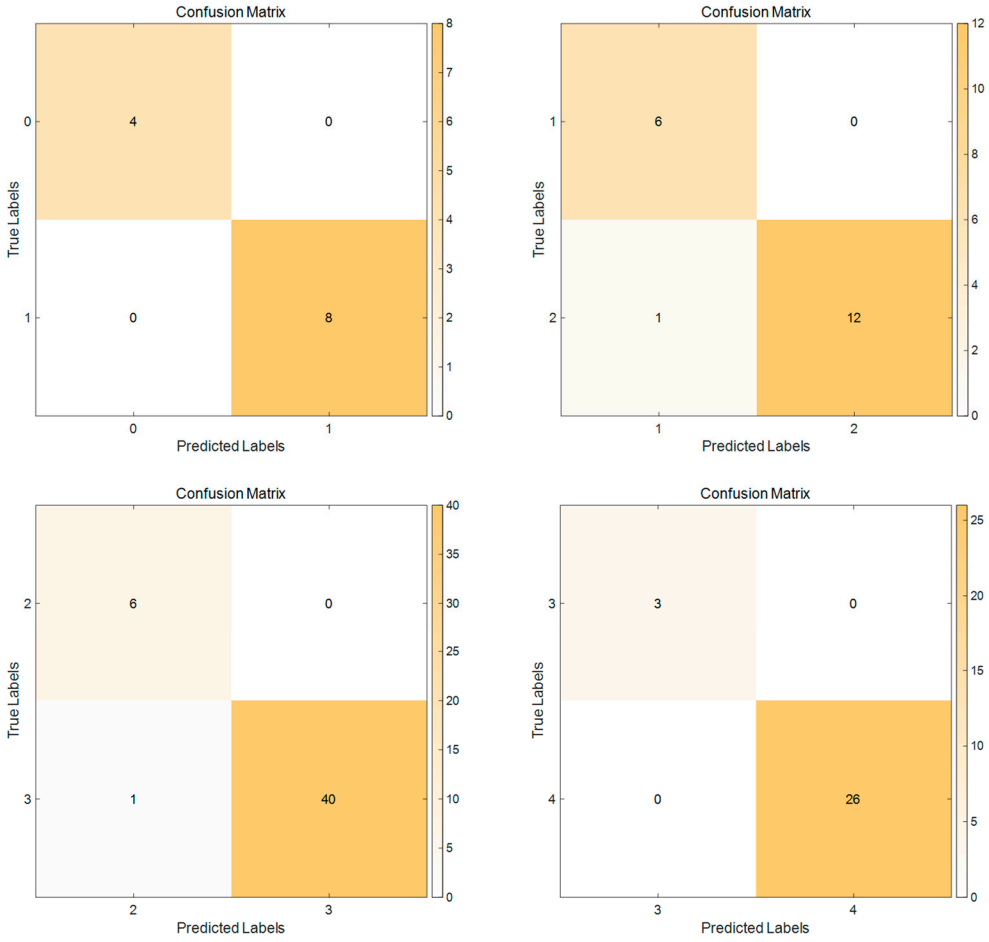

Figure 10. Confusion matrices of the four sub-BRBs.

Table 6. Comparison of four evaluation criteria for two models.

	DLBRB-i	First Layer DLBRB-i	Improvement Ratio
Accuracy	0.98	0.81	20.00%
F1	0.98	0.56	75.00%
Precision	0.98	0.59	66.10%
Recall	0.98	0.66	50.00%

5.4. Model Explainability

To assess the impact of explainability constraints, this study compared the belief degrees of expert knowledge, the optimized DLBRB model, and the DLBRB-i model with explainability constraints. Figure 11 illustrates that after optimizing the standard DLBRB model, significant deviations from expert knowledge belief degrees were observed, indicating a limitation in explainability. The DLBRB-i model, however, closely matches the expert's judgment, proving its success in maintaining explainability. The matching effect has been highlighted with dashed lines in the figure.

Additionally, the DLBRB-i model demonstrates a high consistency with expert knowledge, retaining more expert knowledge elements and showing excellent explainability in both theory and practice. This confirms the significance of explainability constraints in enhancing model transparency and trustworthiness.

Analyses of the belief distributions in the three sub-BRBs, as depicted in Figures 12–15, reveal that the DLBRB-i model's belief distribution is highly consistent with expert knowledge,

further emphasizing the model's explainability strength. In contrast, the DLBRB model, lacking explainability constraints, shows deviations from the initial expert knowledge-based settings.

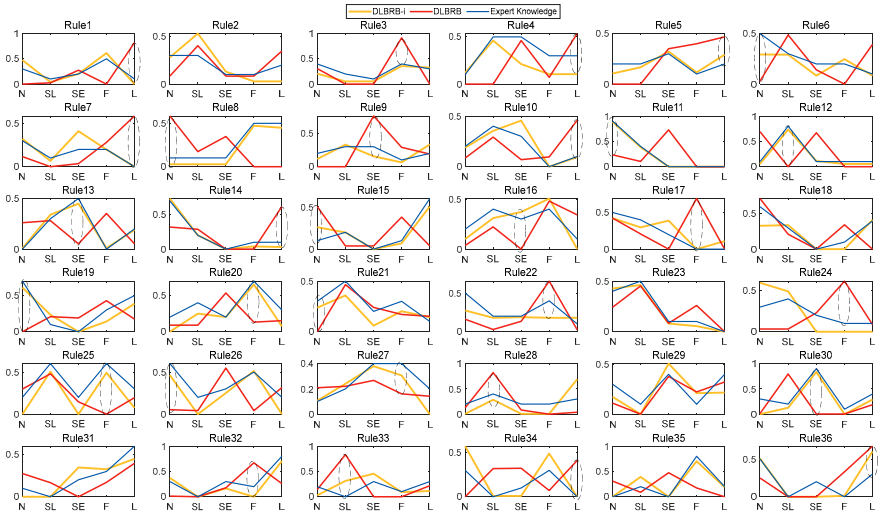

Figure 11. Main BRB belief distribution comparison.

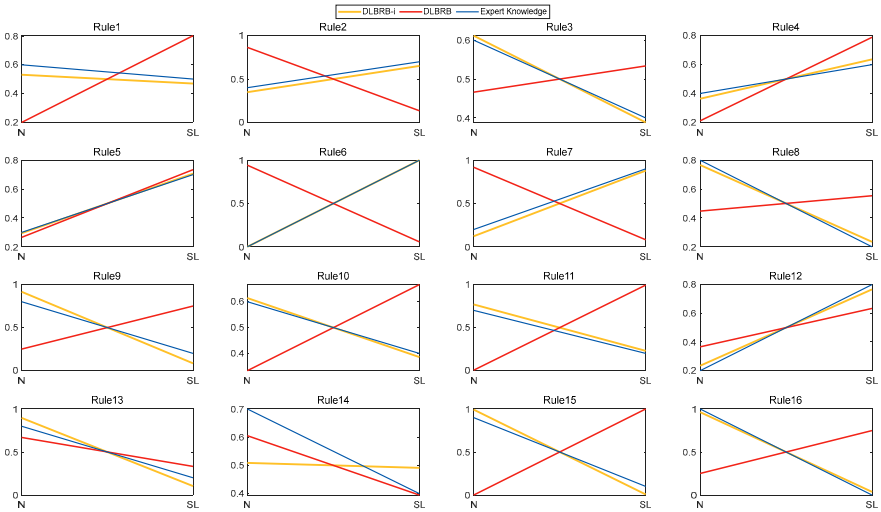

Figure 12. Sub-BRB 0-1 belief distribution comparison.

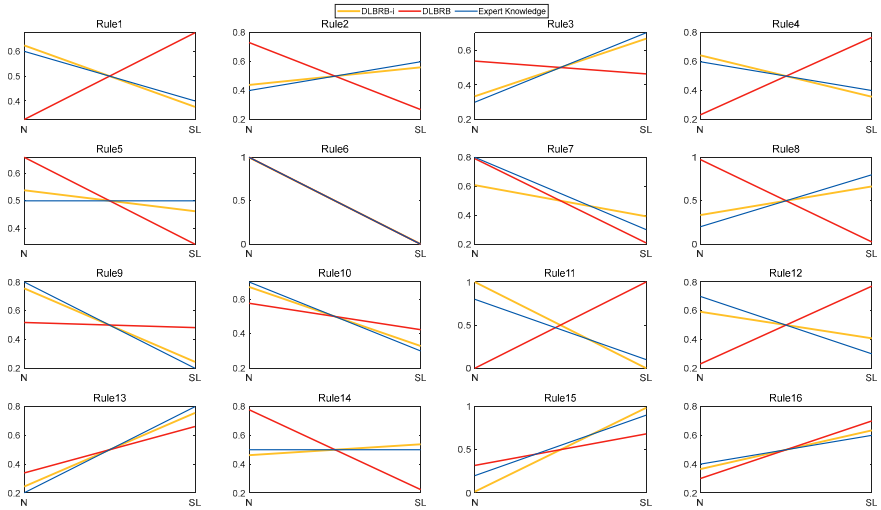

Figure 13. Sub-BRB 1-2 belief distribution comparison.

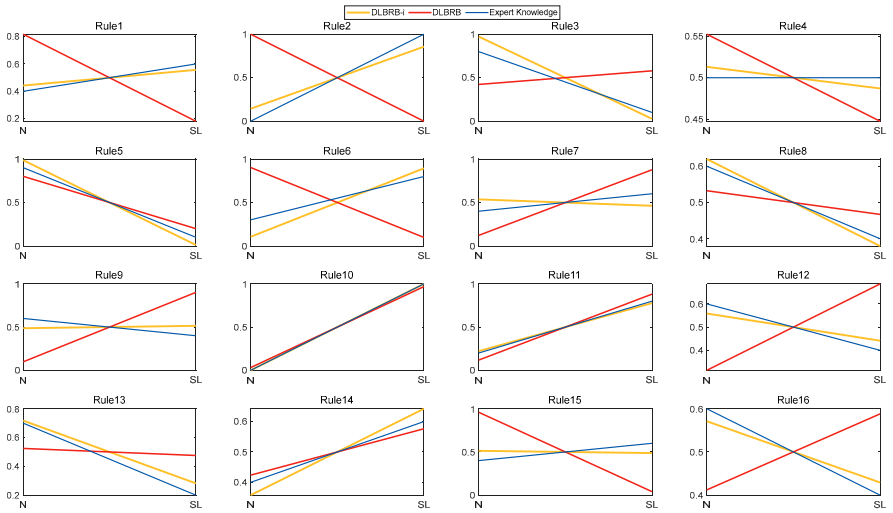

Figure 14. Sub-BRB 2-3 belief distribution comparison.

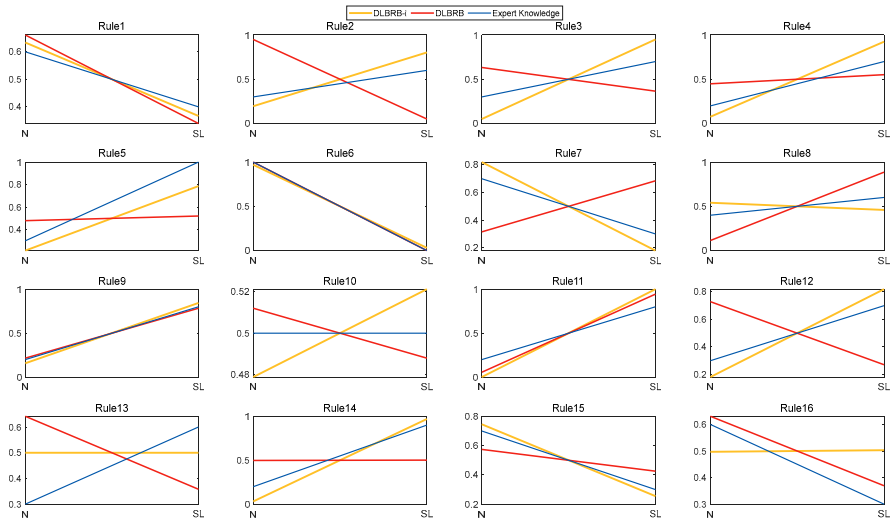

Figure 15. Sub-BRB 3-4 belief distribution comparison.

The DLBRB-i model's output belief distribution aligns with the initial distribution set by experts, ensuring the credibility and reliability of the model's output. This integration of explainability constraints and two-level classification in the DLBRB-i model not only improves explainability but also ensures efficiency and accuracy in complex classification tasks, indicating its potential in fields requiring trusted and transparent decision making.

5.5. Comparative Experiment

As shown in Figures 16–21, the prediction results of six methods were compared with the true values in the dataset. DLBRB-i demonstrated a higher overlap with the actual values. Table 7 shows the parameter settings for the comparison models. Figures 22–25 demonstrate the comparison results of accuracy, F1 score, precision, and recall for six models. DLBRB-i outperforms the other five models across all four evaluation criteria, and DLBRB followed closely behind. This suggests that DLBRB-i excels in predictive accuracy. It also captures the data's true patterns more precisely. The slightly lower academic score of DLBRB suggests that it may require additional tuning or more complex feature engineering to improve its academic score when dealing with imbalanced data or complex patterns. It also implies that DLBRB might be more sensitive to noisy or outlier data, leading to prediction biases.

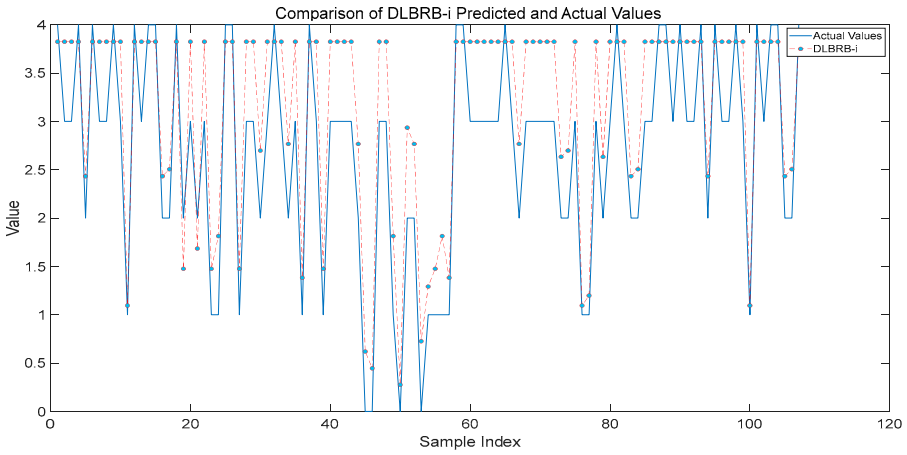

Figure 16. Comparison of DLBRB-i predicted values and actual values.

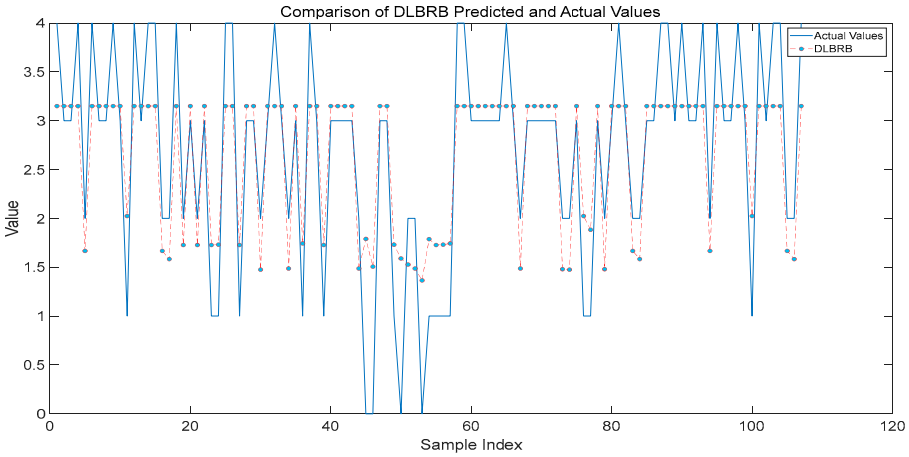

Figure 17. Comparison of DLBRB predicted values and actual values.

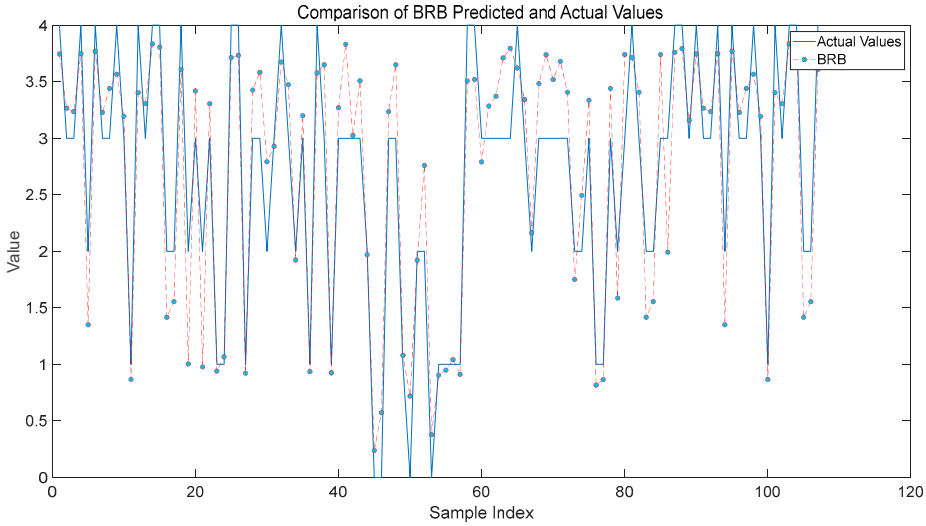

Figure 18. Comparison of BRB predicted values and actual values.

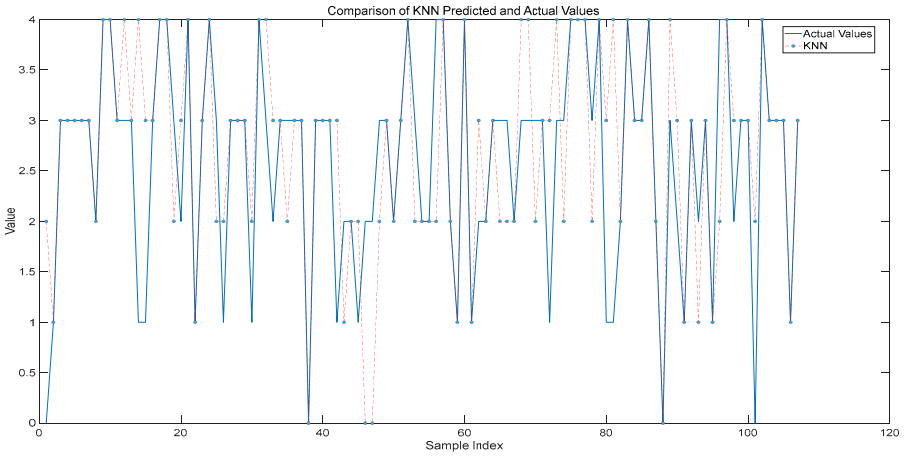

Figure 19. Comparison of KNN predicted values and actual values.

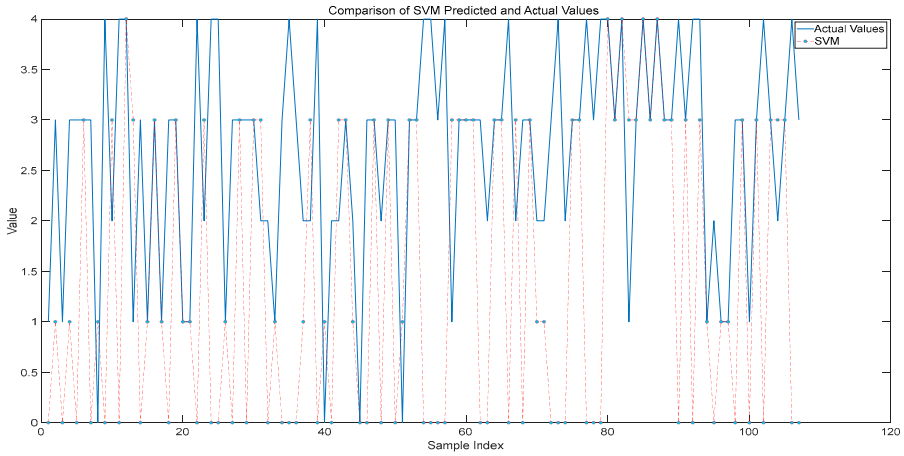

Figure 20. Comparison of SVM predicted values and actual values.

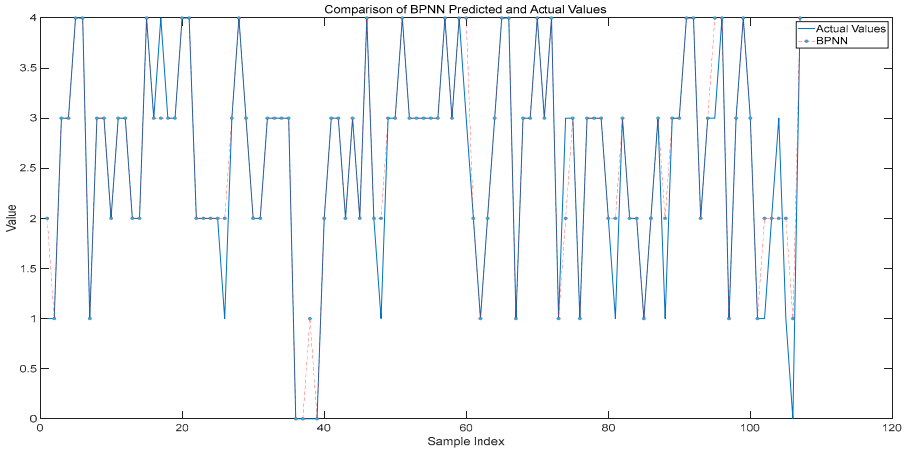

Figure 21. Comparison of BPNN predicted values and actual values.

Table 7. Parameter settings of comparison models.

Models	Parameter Settings
DLBRB&BRB	DLBRB and BRB use the same parameter settings as DLBRB-i.
KNN	Number of nearest neighbors = 5; Weight function = 'uniform'; Nearest neighbor search algorithm = 'auto'; Parameter in distance measurement = 2
SVM	Regularization parameter= 1.0; Kernel function type = 'rbf'; Degree of polynomial kernel = 3; Coefficient of kernel function = 'scale'; Constant term in kernel function = 0.0
BP	Number of neurons in hidden layers = 100; Activation function = 'relu'; Optimization algorithm = 'adam'; L2 regularization parameter = 0.0001; Batch size = 'auto'; Learning rate update strategy = 'constant'; Maximum number of iterations = 200; Momentum parameter = 0.9

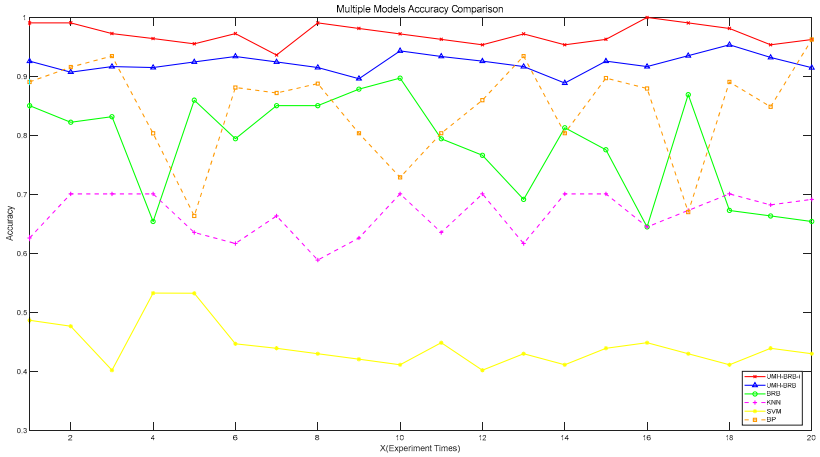

Figure 22. Multi-model accuracy comparison.

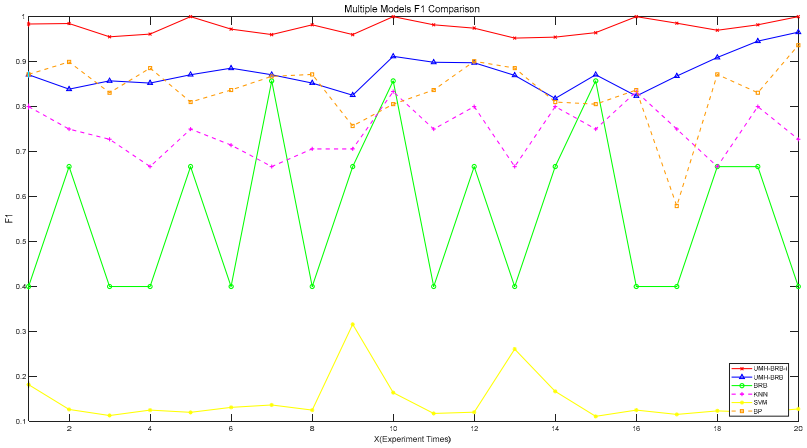

Figure 23. Multi-model F1 comparison.

In comparison, other methods like BPNN, SVM, and KNN did not perform as well on this dataset but may have their advantages on other specific types of datasets or problems.

Table 8 illustrates that in comparative experiments with other machine learning models like BPNN and SVM, DLBRB-i showed a superior academic score across all evaluated metrics, with particularly notable results in recall rate, outperforming other traditional algorithms. This is crucial for identifying minority class samples, which are often hard

to recognize accurately. The results further confirm the suitability and excellence of the DLBRB-i model in dealing with imbalanced datasets.

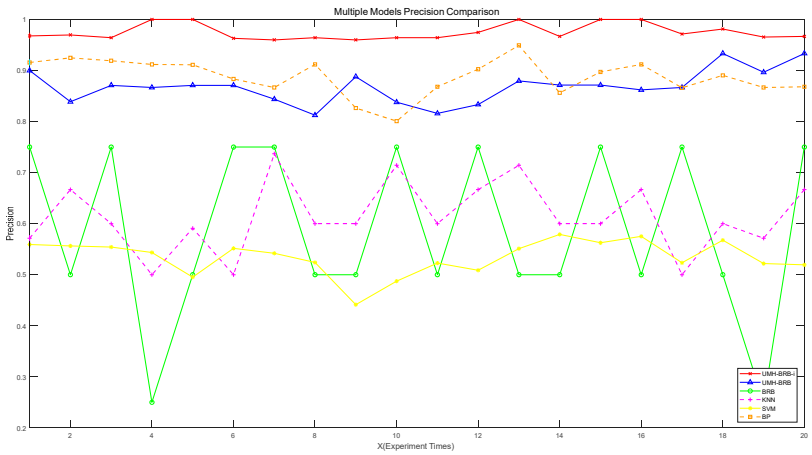

Figure 24. Multi-model precision comparison.

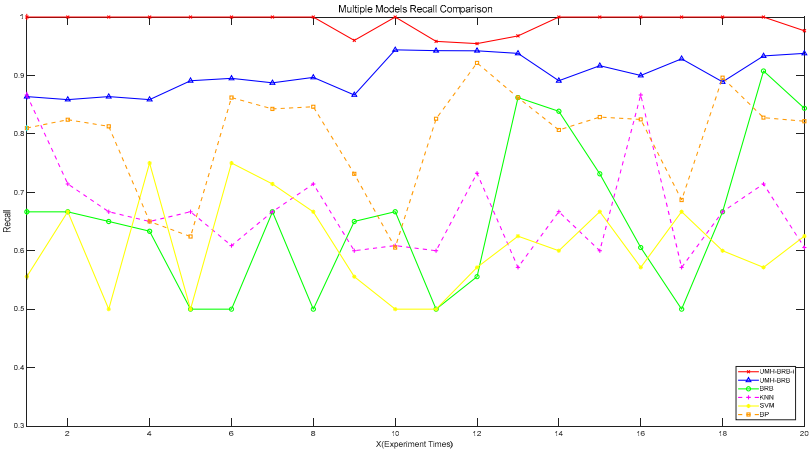

Figure 25. Multi-model recall comparison.

Table 8. Comparison of results for six models.

Model Name	DLBRB-i	DLBRB	BRB	KNN [48]	SVM [49]	BP [50]
Accuracy	0.97	0.92	0.81	0.67	0.44	0.85
F1	0.98	0.88	0.56	0.74	0.15	0.84
Precision	0.98	0.87	0.59	0.61	0.53	0.89
Recall	0.99	0.90	0.66	0.67	0.61	0.80

The hierarchical structure and explainability constraints of the DLBRB-i model not only enhance predictive accuracy but also provide significant advantages in decision support and explainability. These features suggest that the model can adapt to evolving data environments and demands, ensuring ongoing and reliable decision support for users. The comparative experiments also demonstrate the potential and value of the DLBRB-i model in student performance prediction. As research progresses, the DLBRB-i model is anticipated to play a more significant role in a wider range of complex and challenging applications.

In the case analysis of student performance prediction, the explainable dual-layer belief rule base (DLBRB-i) model has shown distinct advantages, particularly in the following areas:

Improved Predictive Accuracy and Applicability:

The DLBRB-i model exhibits greater accuracy and scalability in predicting outcomes compared to the traditional belief rule base (BRB), especially in the presence of class-imbalanced data. This design enables the DLBRB-i to adapt more flexibly to categories with fewer samples, enhancing the predictive academic score for these classes.

Enhanced Model Explainability:

The DLBRB-i model is characterized by its superior explainability. Unlike "black box" models that are driven by data, DLBRB-i offers a clear and transparent reasoning process, allowing users to understand how predictions are derived. This transparency strengthens trust in the model and aids in its optimization and troubleshooting.

Efficient Management of Model Complexity:

The DLBRB-i model employs a layered structure that effectively tackles class imbalance and the challenge of model complexity. This approach prevents the exponential growth of computational complexity that typically occurs with an increase in the number of rules, addressing the "rule explosion" issue. The model maintains predictive accuracy without incurring high computational costs.

In conclusion, the DLBRB-i model has demonstrated significant benefits in dealing with imbalanced datasets and improving prediction accuracy within the case study. It not only performs well in predicting student grades but also provides users with the means to deeply comprehend the model's predictive processes due to its layered structure and high explainability, suggesting its wide-ranging potential in educational applications.

6. Discussion

6.1. Discussion of DLBRB-i

The proposed student performance prediction model, DLBRB-i, significantly outperforms traditional models in terms of explainability and accuracy, especially in the identification of minority class samples. This improvement is attributed to its innovative dual-level architecture and the incorporation of explainability constraints, which ensure that the model's reasoning process aligns closely with expert knowledge. Through comparative analysis, it is evident that the DLBRB-i model not only excels in predictive performance but also in maintaining the clarity of the reasoning process, making it a valuable tool for educators to gauge the impact of their teaching methods and for students to evaluate their own progress.

The DLBRB-i model's enhanced performance can be understood within the context of previous studies that have emphasized the need for explainable AI in educational data mining. While earlier models often lacked the ability to provide convincing explanations for their predictions, DLBRB-i bridges the gap by combining the strengths of theoretical rule-based modeling and empirical data-driven approaches. The model's ability to handle class imbalance and its adherence to the principles of explainability set it apart as a powerful aid in educational analytics.

6.2. Limitation of DLBRB-i

Although this model uses a layered structure to alleviate complexity, as the data scale and problem complexity increase, the computational complexity may still face challenges. When dealing with large-scale datasets with many result categories, the computational cost of DLBRB-i may increase dramatically, resulting in longer running times and increased resource consumption.

Further research will be conducted on how to optimize the algorithm design of the two-layer BRB structure, exploring more intelligent rule generation and screening mechanisms. During the first layer of coarse classification, it is possible to more accurately divide large categories with similar characteristics while avoiding the expansion of the number of rules caused by over-segmentation. During the second layer of fine classification, it is possible to study how to adaptively adjust the structure and parameters of the sub-BRB according

to the distribution characteristics of different category samples, to improve classification efficiency and accuracy while controlling the growth of the number of rules.

7. Conclusions

In summary, the DLBRB-i model represents a significant advancement in the field of educational data mining by providing a highly accurate yet explainable framework for predicting student academic scores. The model's architecture facilitates a deep understanding of the reasoning behind its predictions, which is critical for building trust among educators and students alike. Its success in addressing class imbalance issues and maintaining high levels of explainability during the optimization process underscores its value in promoting sustainable educational development.

While the DLBRB-i model showcases impressive results, there are opportunities for future enhancement. Specifically, the model's performance on larger and more complex datasets needs verification, and strategies for increasing robustness and generalization are warranted. Additionally, exploring the model's adaptability to various educational contexts, cultures, and stages will be essential for its widespread adoption. Integrating the latest AI technologies with the DLBRB-i model could lead to groundbreaking improvements in the precision and efficiency of academic score prediction, potentially revolutionizing educational practices. Ultimately, the DLBRB-i model offers a promising pathway toward more transparent and trustworthy educational analytics, enhancing the overall educational experience.

Author Contributions: Conceptualization, J.M. and F.W.; methodology, J.M.; software, F.W.; validation, J.M., F.W. and W.H.; formal analysis, J.M.; investigation, H.Z.; resources, F.W.; data curation, J.M.; writing—original draft preparation, J.M. and F.W.; writing—review and editing, J.M. and F.W.; visualization, F.W.; supervision, W.H.; project administration, H.H.; funding acquisition, W.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the Open Foundation of Key Laboratory of the Ministry of Education on Application of Artificial Intelligence in Equipment under Grant No. AAIE-2023-0102, in part by the Shandong Provincial Natural Science Foundation under Grant No. ZR2023QF010, and in part by the National Natural Science Foundation of China, Grant Nos. 62227814, 62203461, and 62203365.

Data Availability Statement: The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

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

References

1. Senko, C.; Hulleman, C.S.; Harackiewicz, J.M. Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. *Educ. Psychol.* **2011**, *46*, 26–47. [CrossRef]
2. Loh, E.K.Y. What we know about expectancy-value theory, and how it helps to design a sustained motivating learning environment. *System* **2019**, *86*, 102119. [CrossRef]
3. Schunk, D.H.; DiBenedetto, M.K. Self-efficacy theory in education. In *Handbook of Motivation at School*; Routledge: London, UK, 2016; pp. 34–54.
4. Ang, K.L.-M.; Ge, F.L.; Seng, K.P. Big Educational Data & Analytics: Survey, Architecture and Challenges. *IEEE Access* **2020**, *8*, 116392–116414. [CrossRef]
5. Waheed, H.; Hassan, S.-U.; Aljohani, N.R.; Hardman, J.; Alelyani, S.; Nawaz, R. Predicting Academic Performance of Students from VLE Big Data Using Deep Learning Models. *Comput. Hum. Behav.* **2020**, *104*, 106189. [CrossRef]
6. Al-Tameemi, G.; Xue, J.; Hadi Ali, I.; Ajit, S. A Hybrid Machine Learning Approach for Predicting Student Performance Using Multi-class Educational Datasets. *Procedia Comput. Sci.* **2024**, *238*, 888–895. [CrossRef]
7. Bujang, S.D.A.; Selamat, A.; Krejcar, O. Imbalanced classification methods for student grade prediction: A systematic literature review. *IEEE Access* **2022**, *11*, 1970–1989. [CrossRef]
8. Zusho, A.; Clayton, K. Culturalizing achievement goal theory and research. *Educ. Psychol.* **2011**, *46*, 239–260. [CrossRef]
9. Wigfield, A.; Eccles, J.S. Expectancy–value theory of achievement motivation. *Contemp. Educ. Psychol.* **2000**, *25*, 68–81. [CrossRef]
10. Bandura, A.; Adams, N.E. Analysis of self-efficacy theory of behavioral change. *Cogn. Ther. Res.* **1977**, *1*, 287–310. [CrossRef]

11. Xiao, W.; Ji, P.; Hu, J. A survey on educational data mining methods used for predicting students' performance. *Eng. Rep.* **2022**, *4*, e12482. [CrossRef]
12. Jiao, P.; Ouyang, F.; Zhang, Q.; Alavi, A.H. Artificial intelligence-enabled prediction model of student academic performance in online engineering education. *Artif. Intell. Rev.* **2022**, *55*, 6321–6344. [CrossRef]
13. Liu, L.; Wan, L. Innovative models for enhanced student adaptability and performance in educational environments. *PLoS ONE* **2024**, *19*, e0307221. [CrossRef] [PubMed]
14. Sheela Rani, P.; Dhivya, S.; Dharshini Priya, M.; Dharmila Chowdary, A. Predicting Student's Performance Using Machine Learning Algorithm. *Int. J. Adv. Res. Sci. Commun. Technol.* **2021**, 52581–59429. [CrossRef]
15. Timotheou, S.; Miliou, O.; Dimitriadis, Y.; Villagrá Sobrino, S.; Giannoutsou, N.; Cachia, R.; Martínez Monés, A.; Ioannou, A. Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review. *Educ. Inf. Technol.* **2023**, *28*, 6695–6726. [CrossRef] [PubMed]
16. Mi, C. Data-Driven Student Learning Performance Prediction based on RBF Neural Network. *Int. J. Adv. Res. Sci. Commun. Technol.* **2019**, *15*, 1560–1569. [CrossRef]
17. Tenison, C.; Sparks, J.R. Combining cognitive theory and data driven approaches to examine students' search behaviors in simulated digital environments. *Large-Scale Assess. Educ.* **2023**, *11*, 28. [CrossRef]
18. Chen, M.; Huang, X.; Chen, H.; Su, X.; Yur-Austin, J. Data driven course scheduling to ensure timely graduation. *Int. J. Prod. Res.* **2023**, *61*, 336–361. [CrossRef]
19. Castro-Faix, M.; Duncan, R.G.; Choi, J. Data-driven refinements of a genetics learning progression. *J. Res. Sci. Teach.* **2020**, *58*, 3–59. [CrossRef]
20. Xue, S.; Xue, X.; Son, Y.J.; Jiang, Y.; Zhou, H.; Chen, S. A data-driven multidimensional assessment model for English listening and speaking courses in higher education. *Front. Educ.* **2023**, *8*, 1198709. [CrossRef]
21. Karthikeyan, V.G.; Thangaraj, P.; Karthik, S. Towards developing hybrid educational data mining model for efficient and accurate student performance evaluation. *Soft Comput.* **2020**, *24*, 18477–18487. [CrossRef]
22. Nuankaew, P.; Jeefoo, P.; Nasa-Ngium, P.; Nuankaew, W.S. Hybrid Learning and Blended Learning in the Perspective of Educational Data Mining and Learning Analytics: A Systematic Literature Review. *Int. J. Eng. Trends Technol.* **2023**, *71*, 115–132. [CrossRef]
23. Raes, A. Exploring Student and Teacher Experiences in Hybrid Learning Environments: Does Presence Matter? *Postdigit. Sci. Educ.* **2022**, *4*, 138–159. [CrossRef]
24. Wang, C.; Omar Dev, R.D.; Soh, K.G.; Mohd Nasirudddin, N.J.; Yuan, Y.; Ji, X. Blended learning in physical education: A systematic review. *Front. Public Health* **2023**, *11*, 1073423. [CrossRef] [PubMed]
25. Ray, P.; Chakrabarti, A. A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis. *Appl. Comput. Inform.* **2022**, *18*, 163–178. [CrossRef]
26. Feng, Z.C.; Zhou, Z.J.; Hu, C.H.; Chang, L.L.; Hu, G.Y.; Zhao, F.J. A new belief rule base model with attribute reliability. *IEEE Trans. Fuzzy Syst.* **2019**, *27*, 903–916. [CrossRef]
27. Yang, L.H.; Ren, T.Y.; Ye, F.F.; Peter, N.; Wang, Y.-M.; Lu, H. An ensemble extended belief rule base decision model for imbalanced classification problems. *Knowl.-Based Syst.* **2022**, *242*, 108410. [CrossRef]
28. Yang, L.H.; Liu, J.; Ye, F.F.; Wang, J. Highly explainable cumulative belief rule-based system with effective rule-base modeling and inference scheme. *Knowl.-Based Syst.* **2021**, *240*, 107805. [CrossRef]
29. Liu, M.; He, W.; Zhou, G. A New Student performance prediction Method Based on Belief Rule Base with Automated Construction. *Mathematics* **2024**, *12*, 2418. [CrossRef]
30. Chen, X.; Qian, J.; Li, C. A Student performance prediction Model Based on Hierarchical Belief Rule Base with Explainability (HBRB-I). *Mathematics* **2024**, *12*, 2296. [CrossRef]
31. Chen, H.; Zhou, G.; Zhang, X. Learning Emotion Assessment Method Based on Belief Rule Base and Evidential Reasoning. *Mathematics* **2023**, *11*, 1152. [CrossRef]
32. Zaffar, M.; Hashmani, M.A.; Savita, K.S. A study of feature selection algorithms for predicting students academic performance. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*. [CrossRef]
33. Feng, Y. An evaluation method of PE classroom teaching quality in colleges and universities based on grey system theory. *J. Intell. Fuzzy Syst.* **2020**, *38*, 6911–6915. [CrossRef]
34. Gencel, I.E.; Saracaloglu, A.S. The effect of layered curriculum on reflective thinking and on self-directed learning readiness of prospective teachers. *Int. J. Progress. Educ.* **2018**, *14*, 8–20. [CrossRef]
35. Alamri, R.; Alharbi, B. Explainable student performance prediction models: A systematic review. *IEEE Access* **2021**, *9*, 33132–33143. [CrossRef]
36. Li, Y.; Han, P.; He, W.; Zhang, G.; Wei, H. A new prediction system based on self-growth belief rule base with explainability constraints. *Comput. Mater. Contin.* **2023**, *75*, 3761–3780. [CrossRef]
37. Beaulac, C.; Rosenthal, J.S. Predicting university students' academic success and major using random forests. *Res. High. Educ.* **2019**, *60*, 1048–1064. [CrossRef]
38. Jafari, M.J.; Tarokh, M.J.; Soleimani, P. An explainable machine learning framework for customer churn prediction: A case study in the telecommunications industry. *J. Ind. Eng. Manag. Stud.* **2023**, *10*, 141–157. [CrossRef]

39. Elhaik, E. Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Sci. Rep.* **2022**, *12*, 14683. [CrossRef] [PubMed]
40. Greenacre, M.; Groenen, P.J.F.; Hastie, T.; Iodice d'Enza, A.; Markos, A.; Tuzhilina, E. Principal component analysis. *Nat. Rev. Methods Primers* **2022**, *2*, 100. [CrossRef]
41. Resch, K.; Schrittesser, I. Using the Service-Learning approach to bridge the gap between theory and practice in teacher education. *Int. J. Incl. Educ.* **2023**, *27*, 1118–1132. [CrossRef]
42. Cao, Y.; Zhou, Z.; Hu, C.; He, W.; Tang, S. On the Explainability of Belief Rule-Based Expert Systems. *IEEE Trans. Fuzzy Syst.* **2020**, *29*, 3489–3503. [CrossRef]
43. Carvalho, D.V.; Pereira, E.M.; Cardoso, J.S. Machine Learning Explainability: A Survey on Methods and Metrics. *Electronics* **2019**, *8*, 832. [CrossRef]
44. Turabieh, H.; Azwari, S.A.; Rokaya, M.; Alosaimi, W.; Alharbi, A.; Alhakami, W.; Alnfiai, M. Enhanced Harris Hawks optimization as a feature selection for the prediction of student performance. *Computing* **2021**, *103*, 1417–1438. [CrossRef]
45. Mihaljevic, B.; Bielza, C.; Larrañaga, P. Bayesian networks for explainable machine learning and optimization. *Neurocomputing* **2021**, *456*, 648–665. [CrossRef]
46. Yang, J.B.; Liu, J.; Wang, J. Belief rule-base inference methodology using the evidential reasoning approach-RIMER. *IEEE Trans. Syst. Man Cybern. -Part A Syst. Hum.* **2006**, *36*, 266–285. [CrossRef]
47. Ninrutsirikun, U.; Imai, H.; Watanapa, B. Principal component clustered factors for determining study performance in computer programming class. *Wirel. Pers. Commun.* **2020**, *115*, 2897–2916. [CrossRef]
48. Ukey, N.; Yang, Z.; Li, B.; Zhang, G.; Hu, Y.; Zhang, W. Survey on Exact kNN Queries over High-Dimensional Data Space. *Sensors* **2023**, *23*, 629. [CrossRef]
49. Rabbani, S.; Safitri, D.; Rahmadhani, N.; Fadillah Sani, A.A.; Anam, M. Comparative Evaluation of SVM Kernels for Sentiment Classification in Fuel Price Increase Analysis. *Indones. J. Mach. Learn. Comput. Sci.* **2023**, *3*, 153–160. [CrossRef]
50. Fei, R.; Guo, Y.; Li, J.; Hu, B.; Yang, L. An Improved BPNN Method Based on Probability Density for Indoor Location. *IEICE Trans. Inf. Syst.* **2023**, *106*, 773–785. [CrossRef]

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