An LLM-Assisted Belief Rule Base Model for Explainable Decision Classification

Lixian Xu[1,2,5], Jun Liu[2], Aftab Ali[2], Li Zou[1*], Dehu Yu[3*], Peng Shi[4]
[1]School of Computer and Artificial Intelligence, Shandong Jianzhu University
[2]School of Computing, Ulster University
[3]School of Civil Engineering, Shandong Jianzhu University
[4]School of Electrical and Mechanical Engineering, The University of Adelaide
[5]Computational Intelligence Center, Shandong Jianzhu University
xulixian@sdjzu.edu.cn, j.liu@ulster.ac.uk, a.ali@ulster.ac.uk,
zouli20@sdjzu.edu.cn, yudehu@qut.edu.cn, peng.shi@adelaide.edu.au

Abstract

Traditional Belief Rule Base (BRB) models possess inherent advantages in interpretability and uncertainty reasoning. However, their rule construction heavily depends on expert knowledge, leading to issues such as an excessive number of rules, parameter optimization difficulty, and limited scalability. To address these challenges, this paper proposes a novel **LLM-Assisted Belief Rule Base (LLM-BRB)** classification framework. The proposed method leverages the semantic understanding and knowledge generation capabilities of Large Language Models (LLMs) to automatically construct the prior structure of belief rules. Through semantic consistency constraints and belief normalization mechanisms, the model forms an interpretable and logically coherent rule base, while an Evidence Reasoning (ER) fusion mechanism integrates multi-rule inference outcomes. The proposed LLM-BRB framework demonstrates superior classification performance over the traditional BRB and other benchmark machine learning models, without compromising interpretability. The proposed framework highlights the potential of the **AI + Logic** paradigm for building explainable and trustworthy decision-making models.

Introduction

With the rapid development of artificial intelligence (AI) in decision support and classification problems, interpretability has become one of the key bottlenecks for the deployment and trustworthiness of AI systems in real-world applications. Although traditional machine learning models, such as neural networks, demonstrate excellent predictive performance, they often fail to provide understandable logical explanations for their outputs. The Belief Rule Base (BRB) model, which integrates expert knowledge, uncertainty modeling, and logical reasoning, exhibits strong interpretability and reliability. Hence, BRB systems have been widely applied in engineering diagnosis, risk assessment, and intelligent control.

However, traditional BRB systems still face several significant challenges: (1) Rule extraction relies heavily on expert knowledge, making manual rule construction time-consuming and potentially incomplete; (2) The number of rules grows exponentially with the number of attributes, resulting in poor scalability; (3) Parameter tuning is complex, since belief degrees and rule weights must be iteratively optimized through heuristic algorithms.

In recent years, Large Language Models (LLMs) have shown remarkable capabilities in knowledge extraction, semantic reasoning, and natural language generation. By learning semantic associations and logical structures from large-scale text corpora, LLMs provide new opportunities for rule generation and knowledge induction. Therefore, integrating LLMs into the BRB framework offers a promising direction to achieve a synergy between semantic logic generation and data-driven optimization, leading to an emerging paradigm that unifies AI and Logic.

In this study, a novel LLM-Assisted Belief Rule Base (LLM-BRB) framework is proposed. The proposed framework leverages the semantic understanding and knowledge generation capabilities of LLMs to automatically construct rule priors, perform semantic partitioning of input attributes, and initialize belief distributions. Subsequently, semantic consistency constraints and an Evidence Reasoning (ER) fusion mechanism are introduced to realize interpretable and adaptive classification inference. This design preserves the structural transparency of traditional BRB models while significantly improving rule construction efficiency and generalization capability through LLM-assisted knowledge generation.

The main contributions of this paper can be summarized as follows:

1. A novel explainable classification framework integrating Large Language Models with LLM-BRB is proposed.
2. The mathematical definition of LLM-assisted rule generation and its semantic consistency constraint theorem are formulated.
3. A unified model that integrates LLM-generated rule priors, Evidence Reasoning (ER), and parameter optimization is established.
4. Experiments on UCI benchmark datasets validate that the proposed LLM-BRB achieves superior classification performance and interpretability compared with traditional BRB and other machine learning models.

The remainder of this paper is organized as follows. Section II reviews the development of the Belief Rule Base and

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

the recent applications of large language models in decision-making. Section III introduces the fundamentals of the BRB model and discusses its limitations. Section IV presents the proposed LLM-BRB framework. Section V reports the experimental results on several benchmark datasets. Finally, Section VI concludes the paper and outlines future research directions.

Literature Review

This study is positioned at the intersection of interpretable artificial intelligence, belief rule-based systems, and large language model-assisted reasoning. Therefore, the following section provides a concise review of related work in these areas.

Evolution and Applications of the BRB

Rule-based reasoning has long served as a cornerstone of interpretable decision systems, allowing human-like inference through linguistic "IF–THEN" constructs. However, traditional fuzzy rule-based systems (FRBS) often assume deterministic membership or crisp confidence, making them inadequate for modeling heterogeneous and uncertain information. To overcome these limitations, Yang *et al.* introduced the BRB framework by extending conventional rule structures with belief degrees derived from fuzzy set theory and Dempster–Shafer evidence theory (Yang et al. 2006).BRB can effectively integrate both quantitative and qualitative data while reasoning about their interactions. Since its inception, the BRB framework has undergone significant theoretical and methodological evolution. Building on this foundation, Liu et al. (Liu et al. 2013)proposed the Extended Belief Rule Base system. This formulation allows each rule consequent to represent not a single deterministic output, but a belief distribution over multiple referential outcomes, thereby enabling the modeling of both vagueness and epistemic uncertainty within a unified probabilistic–logical framework. Recently, BRB has since been successfully applied in a wide range of domains, including safety assessment (Zhu et al. 2022), fault diagnosis (Cheng et al. 2020; Li et al. 2025) and other decision-making scenarios (Yang et al. 2021). In the context of classification, BRB has likewise shown strong potential and has been employed in several studies addressing classification-related problems (Lian et al. 2025).

Despite these achievements, several challenges remain. Conventional BRB frameworks still rely heavily on expert-defined rule structures and manual parameter initialization, which can lead to scalability issues and suboptimal reasoning performance when applied to high-dimensional or data-rich environments.

LLM for Decision Support

Large Language Models (LLMs) provide extensive assistance for decision-making processes across various domains. For example, they are capable of processing and synthesizing vast volumes of textual data (Liu et al. 2023), enabling decision-makers to rapidly identify and comprehend key insights. Moreover, LLMs can generate multiple alternative solutions through structured reasoning and scenario simulation (Zhang, Sun, and Li 2023; Wang, Zhang, and Chen 2024), thereby enhancing creativity in decision-making tasks. They can also recognize complex patterns from historical datasets (Li, Zhao, and Xu 2023), perform contextual analysis, and provide quantitative assessments of alternative decisions (Chen, Liu, and Gao 2024).

In addition, LLMs exhibit strong role-playing and social interaction capabilities, allowing them to simulate debates from diverse viewpoints (Zhou, Ma, and Lin 2023; Liu, Xu, and Zhao 2024). By integrating multiple perspectives and simulating multi-agent interactions, LLMs enable decision-makers to explore various scenarios and potential outcomes in a systematic manner. This indicates that LLMs possess the potential to augment human reasoning and enhance the rationality and adaptability of decision processes (He, Zhang, and Wang 2023). Therefore, within the domain of *AI-assisted decision support* (Martínez, Wang, and Liu 2023; Yang, Wang, and Martínez 2024), LLMs (e.g., GPT, Gemini, Claude) have demonstrated powerful capabilities in knowledge extraction, symbolic reasoning, and rule generation. They can automatically generate IF–THEN structures that resemble BRB antecedents, making them ideal candidates for prior rule elicitation.

Preliminaries

Belief Rule Base

In the BRB model, uncertain knowledge is represented through a set of belief rules, each capturing different conditional relationships under incomplete information. A BRB model consists of multiple belief rules, and the k-th rule can be expressed as follows:

$$\begin{aligned} R_k : \text{IF } (U_1 \text{ is } A_1^k) \wedge (U_2 \text{ is } A_2^k) \wedge \cdots \wedge (U_T \text{ is } A_T^k), \\ \text{THEN } \{(D_1, \beta_1^k), (D_2, \beta_2^k), \ldots, (D_N, \beta_N^k)\}, \end{aligned} \tag{1}$$

with a rule weight θ_k, and attribute weights $\delta_1^k, \delta_2^k, \ldots, \delta_T^k$.

Here, $U = (U_1, U_2, \ldots, U_T)$ represents the set of antecedent attributes, while $A^k = (A_1^k, A_2^k, \ldots, A_T^k)$ denotes the corresponding referential values of these attributes in the k-th rule. $D = (D_1, D_2, \ldots, D_N)$ represents the referential values of the consequent attribute, and β_j^k $(j = 1, 2, \ldots, N)$ denotes the belief degree indicating the extent to which D_j is believed to be the consequent when the antecedent conditions in the k-th rule are satisfied.

If $\sum_{j=1}^{N} \beta_j^k = 1$, the rule is complete; otherwise, if $\sum_{j=1}^{N} \beta_j^k < 1$, the rule is incomplete. θ_k is the rule weight representing the relative importance of the k-th rule, and $\delta_1^k, \delta_2^k, \ldots, \delta_T^k$ are the attribute weights associated with the antecedent attributes in that rule. Generally, the sum of attribute weights satisfies $\sum_{i=1}^{T} \delta_i^k = 1$.

The BRB model thus provides a flexible means to represent uncertain and incomplete knowledge by embedding belief degrees into each rule consequent, allowing logical reasoning under uncertainty.

Inference Method

The Evidential Reasoning (ER) algorithm is employed in the BRB framework as the inference mechanism to generate the final belief output for a given set of inputs. Suppose the input of the i-th antecedent attribute is denoted as x_i. Then, its belief distribution can be defined as follows:

$$S(x_i) = \{(A_{i,j}, \alpha_{i,j}) \mid i = 1, 2, \ldots, T;\ j = 1, 2, \ldots, J_i\}, \tag{2}$$

where $A_{i,j}$ represents the j-th referential value of the i-th antecedent attribute, and $\alpha_{i,j}$ denotes the belief degree to which the input x_i is assessed to $A_{i,j}$. T is the number of antecedent attributes, and J_i is the number of referential grades for the i-th antecedent.

In general, the belief distribution can be derived using utility-based equivalence transformation. Assuming the quantitative input x_i corresponds to the qualitative referential value $A_{i,j}$, the belief degrees can be obtained as:

$$\alpha_{i,t} = \begin{cases} \dfrac{x_i - a_{i,j}}{a_{i,j+1} - a_{i,j}}, & \alpha_{i,j+1} = 1 - \alpha_{i,t}, \quad \text{if } a_{i,j} < x_i < a_{i,j+1}, \\ 0, & t \neq j, j+1. \end{cases} \tag{3}$$

where $\alpha_{i,j}$ and $\alpha_{i,j+1}$ represent the belief degrees of the referential grades $A_{i,j}$ and $A_{i,j+1}$, respectively, and $\alpha_{i,t}$ denotes the belief degree of the t-th referential value of the antecedent U_i.

Then, the activation weight ω_k of the k-th belief rule can be calculated as:

$$\omega_k = \frac{\theta_k \prod_{i=1}^{T} (\alpha_i^k)^{\bar{\delta}_i}}{\sum_{l=1}^{L} \theta_l \prod_{i=1}^{T} (\alpha_i^l)^{\bar{\delta}_i}}, \tag{4}$$

where $\bar{\delta}_i$ is the normalized attribute weight, defined as:

$$\bar{\delta}_i = \frac{\delta_i}{\max_{i=1,\ldots,T} \delta_i}. \tag{5}$$

Here, α_i^k denotes the individual matching degree between the input x_i and the referential value A_i^k of the k-th rule.

Finally, the ER algorithm aggregates the activated belief rules to obtain the overall belief distribution of the consequent. The combined belief degree β_n of the n-th referential grade of the consequent can be obtained as:

$$\beta_n = \frac{d\left[\prod_{k=1}^{L}(\omega_k \beta_{n,k} + 1 - \omega_k \sum_{i=1}^{N} \beta_{i,k}) - \prod_{k=1}^{L}(1 - \omega_k \sum_{i=1}^{N} \beta_{i,k})\right]}{1 - d\left[\prod_{k=1}^{L}(1 - \omega_k)\right]} \tag{6}$$

where the normalization factor d represents the degree of conflict among activated rules, defined as:

$$d = \left[\sum_{n=1}^{N} \prod_{k=1}^{L} \left(\omega_k \beta_{n,k} + 1 - \omega_k \sum_{j=1}^{N} \beta_{j,k}\right) - (N-1) \prod_{k=1}^{L} \left(1 - \omega_k \sum_{j=1}^{N} \beta_{j,k}\right)\right]^{-1}. \tag{7}$$

Here, ω_k denotes the activation weight of the k-th rule, $\beta_{n,k}$ is the belief degree corresponding to the n-th consequent in the k-th rule, and d is the global conflict coefficient that ensures belief normalization.

Proposed Methodology

Model Overview

The proposed *LLM-Assisted Belief Rule Base (LLM-BRB)* framework aims to combine the knowledge representation capability of large language models (LLMs) with the uncertainty reasoning power of Belief Rule Bases (BRBs). Formally, the model is represented as:

$$\text{LLM-BRB} = \{\mathcal{A}, \mathcal{C}, \mathcal{R}, \mathcal{W}, \mathcal{B}\},$$

where $\mathcal{A}$ denotes the set of input attributes, $\mathcal{C}$ represents the output class labels, $\mathcal{R}$ is the generated rule base, $\mathcal{W}$ is the set of attribute weights, and $\mathcal{B}$ represents the belief distributions associated with each rule. The entire framework consists of three major stages: (1) LLM-assisted rule elicitation; (2) evidential reasoning (ER)-based belief aggregation; and (3) parameter optimization for fine-tuning rule confidence and weights.

Definition 1 (Belief Rule) A belief rule R_k in the BRB structure is defined as:

$$\begin{aligned} R_k : \text{IF } (A_1 \text{ is } A_1^{(k)}) \wedge (A_2 \text{ is } A_2^{(k)}) \wedge \cdots \wedge (A_M \text{ is } A_M^{(k)}), \\ \text{THEN } \{(C_1, \beta_{1k}), (C_2, \beta_{2k}), \ldots, (C_N, \beta_{Nk})\}. \end{aligned} \tag{8}$$

where β_{jk} denotes the belief degree that the consequent belongs to class C_j under the k-th rule, subject to:

$$0 \leq \beta_{jk} \leq 1, \quad \sum_{j=1}^{N} \beta_{jk} \leq 1.$$

Each rule is associated with a rule weight θ_k and a set of antecedent attribute weights $\{w_1, w_2, \ldots, w_M\}$.

Definition 2 (LLM-based Rule Elicitation) Let S denote the semantic description of the input domain, including linguistic terms, attribute relations, and decision goals. The initial rule base $\mathcal{R}_0$ is generated by the LLM through semantic parsing and reasoning as:

$$\mathcal{R}_0 = f_{\text{LLM}}(S), \tag{9}$$

where each rule in $\mathcal{R}_0$ consists of antecedent linguistic levels, consequent belief degrees, and initial rule weights. The generated belief distributions are normalized to maintain logical consistency:

$$\sum_{j=1}^{N} \beta_{jk} = 1, \quad \forall k. \tag{10}$$

Theorem 1 (Rule Consistency Constraint) If two rules R_p and R_q exhibit semantic similarity above a given threshold δ, i.e.,

$$\text{Sim}(R_p, R_q) > \delta, \tag{11}$$

then their corresponding belief distributions satisfy:

$$\|\boldsymbol{\beta}_p - \boldsymbol{\beta}_q\|_2 \leq \epsilon(\delta), \tag{12}$$

where $\epsilon(\delta)$ is a monotonic decreasing function of the similarity measure.

Under the semantic continuity hypothesis, LLM-generated rules are expected to exhibit smooth variations

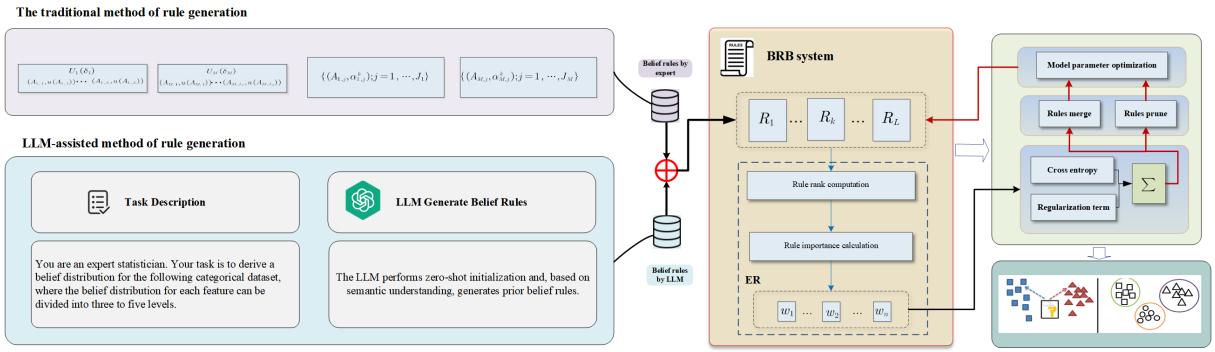

Figure 1: LLM-BRB framework

in belief degrees when antecedents are semantically similar. Hence, the difference between belief distributions is bounded by $\epsilon(\delta)$, ensuring local consistency of rule semantics.

Prompt Design for LLM-Assisted Belief Rule Generation

In the proposed LLM-assisted Belief Rule Base framework, the LLM acts as a semantic reasoning agent to assist in automatically constructing belief rules under uncertain conditions. Since the performance of LLM-based modeling strongly depends on the quality of prompts, a structured prompt design is essential to ensure that the generated belief rules are interpretable, consistent, and mathematically valid. The designed prompt follows a modular structure, consisting of five major components as described below.

- **Role Definition:** This part defines the LLM's role as an *expert knowledge engineer* or *expert statistician*. It specifies that the LLM should act as a domain expert responsible for reasoning under uncertainty, generating belief distributions, and structuring "IF–THEN" rules in the BRB format.
- **Task Description:** This part provides detailed instructions about the construction of belief rules. It includes variable names, their linguistic levels, the desired number of referential grades, and the logical relationships among antecedents and consequents. For example, the task description instructs the LLM to decompose each continuous variable into three to five linguistic terms (e.g., *low, medium, high*) and to assign belief degrees to each consequent.
- **Operating Requirement:** This section clarifies the modeling constraints and reasoning consistency criteria. It instructs the LLM to ensure that the generated belief degrees satisfy $\sum_j \beta_{jk} = 1$ (completeness constraint) and that rule weights θ_k remain within a normalized range. Additionally, it guides the model to follow the canonical BRB syntax for mathematical compatibility with the ER framework.

Evidential Reasoning (ER) Fusion

Given an observation $\boldsymbol{x} = (x_1, x_2, \ldots, x_M)$, the activation weight α_k of each rule is computed as:

$$\alpha_k = \prod_{i=1}^{M} (\mu_i^{(k)})^{w_i}, \tag{13}$$

where $\mu_i^{(k)}$ denotes the matching degree between x_i and the antecedent referential value $A_i^{(k)}$. The normalized activation weights are expressed as:

$$\bar{\alpha}_k = \frac{\alpha_k}{\sum_{k=1}^{K} \alpha_k}.$$

The fused belief distribution is then obtained by ER aggregation:

$$\boldsymbol{\beta}^* = \sum_{k=1}^{K} \bar{\alpha}_k \boldsymbol{\beta}_k, \tag{14}$$

and the final decision output is determined as:

$$\hat{C} = \arg\max_j \beta_j^*. \tag{15}$$

Belief Rule Pruning and Aggregation

To verify the similarity and consistency of the belief rules generated by the LLM, rule pruning and fusion were performed.

Algorithm 1 presents the *Rule Fusion and Pruning* process for refining the BRB after initial parameter learning or LLM-based prior generation. The algorithm aims to reduce redundancy, merge similar rules, and remove weakly activated rules to enhance interpretability and computational efficiency.

(1) Similarity definition. Pairwise rule similarity is defined by considering both antecedent and consequent information. The antecedent similarity $S^{(a)}(j,k)$ is computed using a Gaussian kernel based on Euclidean distance, while the consequent similarity $S^{(c)}(j,k)$ is measured using the L_1 distance between belief distributions. The combined similarity index $S(j,k)$ is obtained as a weighted sum of the two with balance coefficient γ.

(2) Rule fusion. Rules whose similarity exceeds a predefined threshold τ are merged into clusters using a union–find structure. Within each cluster, antecedents, consequents, and rule weights are aggregated through normalized weighted averaging, followed by simplex projection to ensure that the fused belief degrees remain valid probabilities. This step significantly reduces the size of the rule base by consolidating highly similar rules.

Algorithm 1: Rule Fusion and Pruning for BRB

Input : Rule base $\mathcal{R} = \{r_j = (\mathbf{a}_j, \boldsymbol{\beta}_j, w_j)\}_{j=1}^{R}$; (optional) data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$; similarity threshold τ; activation threshold ε; patience T_p; mix $\gamma \in [0, 1]$.
Output: Refined rule set $\mathcal{R}^\star$.
1. Define similarity: For any pair (j, k), define
Antecedent similarity: $S^{(a)}(j,k) = \exp\big(-\|\mathbf{a}_j - \mathbf{a}_k\|_2^2/\sigma^2\big)$;
Consequent similarity: $S^{(c)}(j,k) = 1 - \frac{1}{2}\|\boldsymbol{\beta}_j - \boldsymbol{\beta}_k\|_1$;
Overall similarity: $S(j,k) = \gamma S^{(a)}(j,k) + (1-\gamma) S^{(c)}(j,k)$.

2. Initialization: Build a union–find over $\{1, \ldots, R\}$; set counters $c_j \leftarrow 0$.
3. Fusion pass:
for $j = 1$ **to** R **do**
  for $k = j+1$ **to** R **do**
    if $S(j,k) \geq \tau$ **then**
      Union(j,k)
    end
  end
end
Form clusters $\{\mathcal{C}_m\}$ from union–find sets.
foreach *cluster* $\mathcal{C}_m$ **do**
  Normalize $\tilde{w}_j = \frac{w_j}{\sum_{u \in \mathcal{C}_m} w_u}$, $j \in \mathcal{C}_m$;
  Fuse antecedents $\mathbf{a}^{(m)} = \sum_{j \in \mathcal{C}_m} \tilde{w}_j \, \mathbf{a}_j$;
  Fuse consequents $\widehat{\boldsymbol{\beta}}^{(m)} = \sum_{j \in \mathcal{C}_m} \tilde{w}_j \, \boldsymbol{\beta}_j$; project $\widehat{\boldsymbol{\beta}}^{(m)}$ onto simplex;
  Fuse weight $w^{(m)} = \sum_{j \in \mathcal{C}_m} w_j$; replace $\mathcal{C}_m$ by $r^{(m)} = (\mathbf{a}^{(m)}, \widehat{\boldsymbol{\beta}}^{(m)}, w^{(m)})$.
end
Update $\mathcal{R} \leftarrow \{r^{(m)}\}$; set $R \leftarrow |\mathcal{R}|$.

4. Pruning loop (weak activation):
foreach *pass over* $\mathcal{D}$ **do**
  for $j = 1$ **to** R **do**
    Compute average activation $\bar{\alpha}_j = \frac{1}{N}\sum_{i=1}^{N} \alpha_{ij}$;
    if $\bar{\alpha}_j < \varepsilon$ **then**
      $c_j \leftarrow c_j + 1$
    else
      $c_j \leftarrow 0$
    end
  end
  if *any* $c_j \geq T_p$ **then**
    Prune rules $\{r_j : c_j \geq T_p\}$; renormalize activations over remaining rules.
  end
end
5. Finalize: Return $\mathcal{R}^\star = \mathcal{R}$.

(3) Rule pruning. After fusion, the algorithm evaluates the average activation level of each rule across the dataset. Rules with an average activation below the threshold ε for several consecutive iterations (controlled by patience parameter T_p) are pruned. The remaining rule activations are renormalized to preserve consistency and probabilistic interpretation.

(4) Output. The refined rule base $\mathcal{R}^\star$ is obtained after fusion and pruning, containing only representative and highly activated rules. This refinement enhances the BRB's interpretability, improves generalization capability, and reduces inference complexity.

Algorithm 2: Parameter Optimization of BRB using PSO

Input : Dataset (X, y); initial LLM prior rule set $\mathcal{R} = \{R_k\}$ with belief vectors $\boldsymbol{\beta}_k$ and rule weights θ_k (priors); number of particles P; maximum iterations $T_{\max}$; inertia weight $\omega = 0.72$; learning factors $c_1 = c_2 = 1.4$; position bounds $[\theta_{\min}, \theta_{\max}]$.
Output: Optimized rule weights $\theta^\star$ and best fitness $J^\star$.
1. Define fitness:

$$\mathrm{Fit}(\Theta) = 0.9\,\mathrm{Acc}(X, y; \Theta) + 0.1\,F1(X, y; \Theta)$$

2. Initialization: For each particle $i = 1, \ldots, P$, sample initial position $\mathbf{p}^{(i)} \sim U(0.8, 1.2)^L$ and set velocity $\mathbf{v}^{(i)} = 0$.
Evaluate fitness $\mathrm{Fit}(\mathbf{p}^{(i)})$; set personal best $\mathbf{p}_{\text{best}}^{(i)} = \mathbf{p}^{(i)}$, $J_{\text{best}}^{(i)} = \mathrm{Fit}(\mathbf{p}^{(i)})$.
Set global best $\mathbf{g}_{\text{best}} = \arg\max_i J_{\text{best}}^{(i)}$, $J^\star = \max_i J_{\text{best}}^{(i)}$.
3. Iteration loop: for $t = 1$ **to** $T_{\max}$ **do**
  for *each particle* $i = 1, \ldots, P$ **do**
    Assign rule weights $\theta_k \leftarrow p_k^{(i)}$ for all $k = 1, \ldots, L$.
    Evaluate fitness $J_i = \mathrm{Fit}(\mathbf{p}^{(i)})$.
    if $J_i > J_{best}^{(i)}$ **then**
      Update $\mathbf{p}_{\text{best}}^{(i)} \leftarrow \mathbf{p}^{(i)}$, $J_{\text{best}}^{(i)} \leftarrow J_i$.
    end
  end
  Update global best: $\mathbf{g}_{\text{best}} = \arg\max_i J_{\text{best}}^{(i)}$, $J^\star = \max_i J_{\text{best}}^{(i)}$.
  For each particle i, draw $r_1, r_2 \sim U(0,1)$ and update:

$$\mathbf{v}^{(i)} \leftarrow \omega \mathbf{v}^{(i)} + c_1 r_1 (\mathbf{p}_{\text{best}}^{(i)} - \mathbf{p}^{(i)}) + c_2 r_2 (\mathbf{g}_{\text{best}} - \mathbf{p}^{(i)}),$$
$$\mathbf{p}^{(i)} \leftarrow \mathbf{p}^{(i)} + \mathbf{v}^{(i)}, \quad \mathbf{p}^{(i)} \in [\theta_{\min}, \theta_{\max}].$$

end
4. Finalize: Return $\theta^\star = \mathbf{g}_{\text{best}}$ and best fitness $J^\star$.

Parameter Optimization

To refine the LLM-generated priors, the rule weights and belief parameters are optimized using a hybrid learning objective:

$$\min_{\Theta} L(\Theta) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{CE}(y_i, \mathrm{ER}(x_i; \Theta)) + \lambda \|\Theta - \Theta_0\|_1, \tag{16}$$

where Θ denotes the trainable parameters (rule weights, attribute weights, and belief degrees), Θ_0 are the LLM-generated priors, λ is a regularization coefficient, and CE represents the cross-entropy loss function. This formulation ensures a balance between data-driven optimization and knowledge-driven interpretability.

To efficiently optimize Θ, a particle swarm optimization (PSO) algorithm is employed. Each particle represents a candidate parameter vector, and its position corresponds to a set of BRB rule weights and belief degrees. The optimization process iteratively updates these particles according to their own experience and the global best solution, enabling a balance between global exploration and local exploitation. The detailed procedure is summarized in Algorithm 2.

The optimization iteratively adjusts the rule weights to maximize the BRB's predictive accuracy. By leveraging the swarm intelligence mechanism, PSO avoids local minima and ensures a globally optimal set of parameters, thereby improving the interpretability and generalization of the refined LLM-BRB model.

Case study

Datasets

We selected some benchmark datasets from the UCI Machine Learning Repository to comprehensively evaluate the classification performance and generalization capability of the proposed method These datasets vary in scale, attribute types, and class distributions, covering numeric, categorical, and mixed-type features. Each antecedent attribute was divided into three referential grades (low, medium, high) according to its normalized range, following common BRB practices. The details of representative datasets are summarized in Table 1. Each experiment used a hold-out validation strategy, where 80% of the data were used for training and 20% for testing. The procedure was repeated five times with random splits to reduce sampling bias.

Table 1: Benchmark datasets

Benchmark	Instances	Attributes	Classes
Iris	150	4	3
Wine	178	13	3
Seeds	210	7	3
Diabetes	768	8	2
Balance	625	4	3

Experimental Setup

To evaluate the effectiveness and interpretability of the proposed LLM-Assisted BRB framework, a series of experiments were conducted on multiple benchmark classification datasets. The experimental design integrates (i) rule generation via LLMs, (ii) ER-based inference, and (iii) optimization-based fine-tuning, following the methodological principles described in Section 3.

All experiments were implemented in Python 3.10 on an Intel Core i7-1255U (1.70 GHz) under Windows 11, and the LLM queries were executed through the GPT-5 model with temperature $T = 0.3$ to ensure stable semantic elicitation.

Experimental Steps

The experiments were organized into three stages, each corresponding to a distinct research objective. The detailed procedures presented here use the *Iris* dataset as a representative example, while the same steps were applied to the other benchmark datasets.

Stage 1: LLM-driven Rule Elicitation The LLM was prompted with a structured instruction template describing the dataset's attributes and linguistic categories, such as:

> *"You are an expert decision analyst. Based on the attributes $\{A_1, A_2, \ldots, A_T\}$, generate interpretable IF–THEN rules mapping antecedent conditions to class labels. Each rule should include belief degrees β_j^k for each possible outcome."*

This subsection illustrates a demonstration of the Iris dataset experiment, where the rule base was initialized using an LLM-assisted procedure and subsequently fine-tuned through the BRB optimization framework (Step 4).

Referential Levels (Example Selection). For demonstration, two representative input attributes were selected:

- **petal_length:**
  short:[0,3.0), *medium:[3.0,5.0)*, *long:[5.0,10.0)*
- **petal_width:**
  narrow:[0,1.0), *medium:[1.0,1.8)*, *wide:[1.8,3.0)*

Attribute Weights (Initial Setting). The relative importance of the two attributes was defined as

$$w_{\text{petal_length}} = 0.6, \quad w_{\text{petal_width}} = 0.4.$$

Extracted Rules (Top 4 Selected). Based on the referential levels and LLM-generated reasoning, four representative rules were formulated as follows:

- **R1:** IF *petal_length=short* AND *petal_width=narrow* THEN {setosa: 0.9, versicolor: 0.1, virginica: 0.0}, *rule_weight* = 1.0
- **R2:** IF *petal_length=medium* AND *petal_width=medium* THEN {setosa: 0.1, versicolor: 0.8, virginica: 0.1}, *rule_weight* = 1.0
- **R3:** IF *petal_length=long* AND *petal_width=wide* THEN {setosa: 0.0, versicolor: 0.1, virginica: 0.9}, *rule_weight* = 1.0
- **R4:** IF *petal_length=medium* AND *petal_width=wide* THEN {setosa: 0.0, versicolor: 0.6, virginica: 0.4}, *rule_weight* = 0.9

These rules represent the initial expert-knowledge base generated through LLM reasoning and serve as the starting point for subsequent belief rule optimization in the BRB framework. The example demonstrates how linguistic discretization, attribute weighting, and interpretable rule generation can be integrated to form a transparent knowledge-driven inference model.

The output (conclusion) attribute is the flower category:

$$\text{Class} = \{\textit{Setosa},\ \textit{Versicolor},\ \textit{Virginica}\}.$$

The resulting rules were parsed into formal BRB structures $\{R_k\}$ with associated belief degrees, antecedent referential values, and rule weights. This process ensures semantic consistency between LLM-generated natural-language rules and the structured BRB representation.

Stage 2: ER-based Inference Given a test instance $\boldsymbol{x}$, the activation weight α_k of each rule was computed via Eq. 13. The ER-based inference stage then evaluated the predictive reasoning capability of these generated rules. Each test sample $\boldsymbol{x}_i$ was propagated through the rule base, and the corresponding activation weights α_k were computed. The fused belief vector $\boldsymbol{\beta}^*$ produced a class decision that reflected the degree of evidential support aggregated across all activated rules.

Empirical evaluation on the Iris dataset showed that the LLM-BRB framework achieved accurate and stable classification results while maintaining high interpretability. The reasoning chain between input features, rule activation, and final belief aggregation could be explicitly traced, allowing analysts to identify which linguistic attributes contributed most to each class decision. Compared with purely data-driven methods, the LLM-BRB model provided not only comparable accuracy but also transparent, human-understandable decision logic.

Stage 3: Model Parameter Optimization The final stage aimed to optimize the model parameters of the LLM-BRB, including rule weights $\boldsymbol{\theta}$, attribute weights $\boldsymbol{\delta}$, and belief degrees $\boldsymbol{\beta}$, while preserving consistency with the LLM-generated priors Θ_0. The optimization objective was formulated as Eq. 16. To efficiently solve this multi-parameter optimization problem, a PSO strategy was adopted, as illustrated in Algorithm 2. During optimization, the particle positions were initialized around the LLM-generated priors (Θ_0) within a bounded search space.

The PSO algorithm directly aligns with the model optimization structure shown in Fig. 1, where the *Cross-Entropy* and *Regularization Term* components form the core objective. Meanwhile, rule pruning and rule merging operations were periodically applied to simplify the rule base and avoid redundancy, ensuring that the final optimized model remains compact and interpretable.

After convergence, the optimized parameters $\Theta^* = \{\boldsymbol{\theta}^*, \boldsymbol{\delta}^*, \boldsymbol{\beta}^*\}$ were obtained and used for final inference on all datasets. The optimization process effectively fine-tuned the LLM-initialized rules, leading to improved classification accuracy and lower entropy while preserving interpretability and consistency with prior human-like reasoning patterns.

Results and Discussion

Five benchmark datasets from the UCI repository—namely *Iris*, *Wine*, *Seeds*, *Diabetes*, and *Balance*—are employed to evaluate the effectiveness of the proposed LLM-BRB classifier. Table 1 summarizes the numbers of attributes, classes, and instances, as well as the rule-base sizes before and after the downsizing of the BRB classifier.

Table 2: Benchmark datasets and optimized rule-base sizes

Benchmark	Theoretical belief rules	LLM-derived prior belief rules	Optimized rule-base size
Iris	5^4	60	12
Wine	5^{13}	100	67
Seeds	5^7	80	59
Diabetes	5^8	120	85
Balance	5^4	60	48

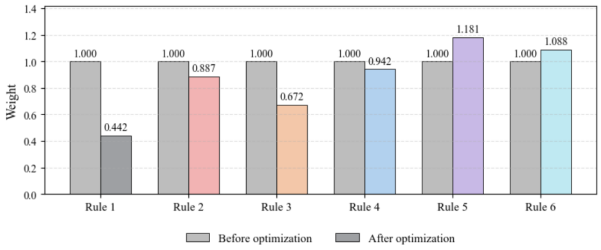

Figure 2: Comparison of rule weights before and after optimization (Rules 1–6).

In practice, the number of belief rules generated by the LLM can be flexibly adjusted, as it is not constrained to any predefined size. During the experiments, we tested different initial rule quantities, starting from 10 and 20 for smaller datasets, to examine how rule scale affects classification performance. Since the feature dimensions and complexity vary across datasets, the LLM was guided through adaptive prompts to generate an appropriate number of interpretable rules that it "believes" represent a good classification structure for each dataset.

The detailed rule generation results and the final optimized rule quantities, obtained under the condition of maintaining high classification accuracy, are summarized in Table 2. It can be observed that the LLM-assisted BRB framework effectively reduces the rule-base size while preserving model interpretability and predictive performance.

Table 3 presents the optimized rule weights and corresponding belief distributions for the twelve rules obtained after the BRB optimization process. Each rule

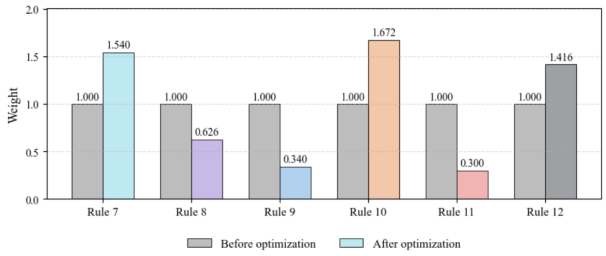

Figure 3: Comparison of rule weights before and after optimization (Rules 7–12)

Table 3: Optimized rule weights (θ) and belief distributions for 12 rules.

Rule	Weight (θ)	Belief Distribution
Rule 1	0.4417	[0.97, 0.02, 0.01]
Rule 2	0.8871	[0.95, 0.03, 0.02]
Rule 3	0.6723	[0.92, 0.06, 0.02]
Rule 4	0.9418	[0.88, 0.10, 0.02]
Rule 5	1.1806	[0.08, 0.85, 0.07]
Rule 6	1.0878	[0.05, 0.88, 0.07]
Rule 7	1.5402	[0.03, 0.90, 0.07]
Rule 8	0.6262	[0.05, 0.87, 0.08]
Rule 9	0.3396	[0.05, 0.87, 0.08]
Rule 10	1.6723	[0.02, 0.08, 0.90]
Rule 11	0.3000	[0.02, 0.08, 0.90]
Rule 12	1.4156	[0.01, 0.04, 0.95]

consists of a weight parameter θ, representing its relative contribution to the overall inference, and a normalized belief distribution over the three output classes. As shown in the table, the optimized weights exhibit clear differentiation among rules, reflecting their varying levels of activation and importance within the decision structure. Rules associated with stronger evidence toward specific classes (e.g., Rule 5 and Rule 7) are assigned higher weights, while less discriminative rules (e.g., Rule 9 and Rule 11) have lower influence. This adaptive weighting mechanism enhances both the interpretability and predictive performance of the LLM-BRB model by emphasizing the most representative and reliable rules during inference.

Table 4: Classification accuracy (%) of different machine methods on benchmark datasets

Method	Iris	Wine	Seeds	Balance	Diabetes
MLP	87.00%	85.67%	83.93%	77.32%	71.76%
SVM	94.50%	79.33%	90.47%	78.52%	72.60%
SCN	78.50%	81.67%	98.33%	60.84%	87.37%
ANFIS	89.00%	56.17%	74.05%	72.64%	77.71%
RIN	91.83%	89.33%	89.17%	75.60%	74.57%
CRIN	94.83%	89.43%	90.71%	80.12%	84.74%
LLM-BRB	95.10%	89.85%	90.10%	85.40%	88.25%

The experimental classification accuracy of the proposed LLM-BRB model is summarized in Table 4, along with comparative results of several representative machine learning methods across five benchmark datasets. As shown in the table, the proposed LLM-BRB consistently achieves superior performance on all datasets. Moreover, by leveraging LLM-generated prior knowledge, the proposed framework substantially reduces the manual effort required to construct and refine belief rules in practical applications, alleviating the burden of human rule design and adjustment. These results demonstrate that integrating LLM reasoning with the BRB framework not only enhances classification capability and generalization but also maintains high interpretability and practical applicability compared with purely data-driven approaches.

Conclusion

In this paper, an LLM-Assisted Belief Rule Base model was proposed for interpretable and accurate classification. The framework integrates LLM-based rule generation, belief consistency constraints, and evidential reasoning inference to combine the advantages of linguistic reasoning and probabilistic learning. Experimental results on multiple benchmark datasets demonstrate that the proposed LLM-BRB achieves superior classification accuracy, high interpretability, and compact rule structures compared with traditional BRB approaches. Furthermore, the LLM-assisted rule elicitation effectively reduces manual effort in knowledge construction, providing human-aligned explanations that enhance transparency and trust in AI-assisted decision-making systems.

Future research will focus on extending the framework toward multimodal BRB construction, dynamic rule adaptation, and broader deployment in real-world intelligent decision applications.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China (No. 62176142) and Major Scientific & Technological Innovation Projects of Shandong Province, China (No. 2021CXGC011204).

References

Chen, J.; Liu, T.; and Gao, F. 2024. Contextual Analysis and Quantitative Evaluation of Decisions using LLMs. *Decision Support Systems*, 174: 114016.

Cheng, C.; Wang, J.; Zhou, Z.; Teng, W.; Sun, Z.; and Zhang, B. 2020. A BRB-based effective fault diagnosis model for high-speed trains running gear systems. *IEEE transactions on intelligent transportation systems*, 23(1): 110–121.

He, J.; Zhang, T.; and Wang, Y. 2023. Enhancing Rationality and Adaptability of Human Decision Processes via LLMs. *AI and Ethics*, 4(3): 215–229.

Li, J.; Qian, G.; He, W.; Zhu, H.; and Zhou, G. 2025. A novel classification method based on an online extended belief rule base with a human-in-the-loop strategy. *Applied Intelligence*, 55(7): 586.

Li, Q.; Zhao, R.; and Xu, P. 2023. Pattern Recognition and Knowledge Extraction with Transformer-based LLMs. *Knowledge-Based Systems*, 262: 110210.

Lian, Z.; Feng, Z.; Zhou, Z.; Hu, C.; Tang, S.; and Wang, J. 2025. Large-Scale linguistic Z-Number Belief Rule Base Methodology for Multidimensional and Unreliable Knowledge Representation and Learning. *IEEE Transactions on Cybernetics*, 55(6): 2657–2670.

Liu, C.; Xu, H.; and Zhao, D. 2024. Simulating Multi-Agent Interactions with Large Language Models for Group Decision Support. *Applied Intelligence*, 54(2): 1556–1572.

Liu, J.; Martinez, L.; Calzada, A.; and Wang, H. 2013. A novel belief rule base representation, generation and its inference methodology. *Knowledge-Based Systems*, 53: 129–141.

Liu, Y.; Yao, Y.; Ton, J.-F.; Zhang, X.; Cheng, R. G. H.; Klochkov, Y.; Taufiq, M. F.; and Li, H. 2023. Trustworthy LLMs: A survey and guideline for evaluating large language models' alignment. *arXiv preprint arXiv:2308.05374*.

Martínez, L.; Wang, H.; and Liu, J. 2023. AI-assisted Decision Support Systems: Foundations and Challenges. *Information Fusion*, 96: 101865.

Wang, X.; Zhang, C.; and Chen, Y. 2024. Scenario Simulation and Policy Reasoning using LLM-based Frameworks. *Expert Systems with Applications*, 242: 122812.

Yang, J.-B.; Liu, J.; Wang, J.; Sii, H.-S.; and Wang, H.-W. 2006. Belief rule-base inference methodology using the evidential reasoning approach-RIMER. *IEEE Transactions on systems, Man, and Cybernetics-part A: Systems and Humans*, 36(2): 266–285.

Yang, L.-H.; Liu, J.; Wang, Y.-M.; Nugent, C.; and Martínez, L. 2021. Online updating extended belief rule-based system for sensor-based activity recognition. *Expert Systems with Applications*, 186: 115737.

Yang, L.-H.; Wang, Y.-M.; and Martínez, L. 2024. Integrating Knowledge Reasoning and Decision Intelligence in AI-assisted Decision-Making. *Knowledge-Based Systems*, 274: 111286.

Zhang, S.; Sun, X.; and Li, H. 2023. Reasoning with Large Language Models for Decision-Making Tasks. *Artificial Intelligence Review*, 56: 4821–4843.

Zhou, Y.; Ma, J.; and Lin, Q. 2023. Role-Playing and Social Simulation Abilities of Large Language Models. *Computational Intelligence*, 39(5): 1672–1685.

Zhu, H.-L.; Liu, S.-S.; Qu, Y.-Y.; Han, X.-X.; He, W.; and Cao, Y. 2022. A new risk assessment method based on belief rule base and fault tree analysis. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 236(3): 420–438.