Intelligent Systems Research and Applications Journal
ISSN: 3120 - 0826

Intelligent Systems Research and Applications Journal

www.israj.org
eISSN: 3120 - 0826

ISRAJ
Intelligent Systems Research and Applications Journal

A Reinforcement Learning-Based Framework for Multi-Criteria Decision-Making

Laxminarayan Sahoo Author[1,*], Sumanta Lal Ghosh [1,]

[1] Department of Computer and Information Science, Raiganj University, Raiganj-733134, India

ARTICLE INFO

Article history:
Received 15 January 2026
Received in revised form 1 March 2026
Accepted 6 March 2026
Available online 8 March 2026

Keywords:

Multi-Criteria decision-Making; Reinforcement Learning; Learning-to-Rank; Preference Aggregation.

ABSTRACT

Multi-Criteria Decision-Making (MCDM) methods have been applied to various ranking problems in finance, engineering, and management; however, many traditional methods rely on data normalization and fixed weights assigned to the criteria. Often, such data normalization has led to distortion of ordinal information and inconsistent results. In this study, we propose an RL-based framework for solving MCDM problems when the decision data are ranked. The ranking task was then reformulated as a learning-to-rank optimization problem, where the preference scores are assigned to alternatives and learned by maximizing agreement with a benchmark ranking. The proposed approach was first illustrated with a simple example and then applied to the real-world evaluation of banks using the CAMELS criteria. The results showed that the RL-based approach achieved greater consistency with benchmark rankings than normalization-based MCDM methods, thereby extending the applicability of rank-preserving approaches.

1. Introduction

Multi-Criteria Decision-Making (MCDM) techniques are widely used to support decisions involving multiple criteria. Such decision-making situations often arise in engineering, economics, finance, management, medicine, and policy-making [6,7]. In most practical situations, decision-makers must assess and rank alternatives that vary in performance across multiple criteria, a task that cannot be easily accomplished through intuitive or single-criterion assessment. In the banking and financial sector, MCDM methods are widely used to evaluate bank performance, assess financial soundness, and assess risk susceptibility. Banks operate in complex environments and are judged by criteria that reflect capital adequacy, asset quality, management efficiency, profitability, liquidity, and market risk sensitivity. Among the most widely accepted models for such assessments is the CAMELS rating method [19,20], which combines six major criteria of bank performance: Capital Adequacy, Asset Quality, Management Capability, Earnings Ability, Liquidity, and Sensitivity to Market Risk. Because of its holistic approach, the CAMELS model has been widely employed by regulators and researchers for comparative bank assessments. Although the CAMELS system is

[*] *Corresponding author.*
E-mail address: lxsahoo@gmail.com

https://doi.org/10.59543/ydm6xr12

widely used, translating its various indicators into a single, accurate ranking remains difficult. This problem becomes more complicated when the data used in the evaluation process are presented as ranks or ordinal data rather than numerical data [3,10,16,20,34]. Rank data is usually preferred because it is less sensitive to extreme values, makes all values comparable, and better reflects expert opinion. However, most traditional MCDM models cannot handle ordinal data and rely on normalization techniques [1,9], which can be detrimental to ranking. Some popular MCDM models developed in the literature include MOORA, RAM, FUCA, and CURLI [34].

The models differ in how they normalize data, assign weights to criteria, and process information. Normalization-based models, such as MOORA and RAM, are useful when criteria are measured on different scales. However, recent studies have observed that when all criteria are measured on the same scale, normalization can distort rank distances and yield inconsistent or misleading results [35]. This problem has been identified in recent comparative studies [12]. In an experiment assessing the Vietnamese banking system using the CAMELS framework, it was found that normalization-based approaches produced rankings that were highly inconsistent with the CAMELS benchmark, whereas rank-preserving approaches such as FUCA and CURLI, which do not use normalization, were much more consistent with it. This shows that data structure is an important factor in choosing an MCDM approach [30]. Although FUCA and CURLI work well on rank-based data, they still rely on rigid aggregation strategies and cannot learn the relative importance of criteria adaptively [31]. Moreover, in most traditional MCDM approaches, criterion weights are pre-specified and typically derived from expert knowledge or subjective assumptions, although objective weighting methods have also been developed [14], including structured methods such as the Best-Worst approach [23]. The choice of weights can greatly affect the ranking outcome, and even slight variations can cause significant changes in the final ranking [8]. In recent years, machine learning techniques have received considerable attention as alternatives or supplements to conventional decision-making approaches. Among these methods, reinforcement learning [28] has been identified as a robust paradigm for optimizing and decision-making through learning from feedback. Reinforcement learning enables an agent to learn optimal actions by interacting with a given environment and maximizing a reward signal, and has been successfully applied in various settings [22].

The learning-based approach of reinforcement learning makes it very attractive for complex decision tasks where modelling is challenging, particularly in high-dimensional and representation-learning settings [4]. The use of reinforcement learning to solve ranking problems in information retrieval and recommendation systems is known as learning to rank [29]. Nevertheless, the use of reinforcement learning to solve MCDM problems is still relatively scarce. In contrast to traditional MCDM approaches, reinforcement learning does not require defining aggregation rules or weights in advance. Rather, it can learn how to aggregate the criteria by optimizing a performance criterion, for instance, consistency with respect to a reference ranking. Motivated by the shortcomings of current MCDM approaches and recent results on rank-based decision matrices, this paper introduces a reinforcement learning-based approach to multi-criteria decision-making. The proposed approach models the MCDM problem as a learning-to-rank task, in which a policy assigns preference scores to alternatives based on their criterion values. The policy is learned to maximize the consistency of ranking with a reference ranking using Spearman’s rank correlation coefficient as the reward function.

Notably, the approach does not require data normalization or pre-specified criterion weights. To assess the efficiency of the proposed approach, a practical example from the existing literature that evaluates 30 banks in Vietnam is used. The proposed approach is then applied to the dataset, and its performance is compared with existing MCDM approaches, including MOORA, RAM, FUCA, and CURLI. The numerical results show that the ranking error is high for normalization-based methods, whereas rank-preserving methods perform better. Most importantly, the proposed reinforcement learning method performs perfectly on the CAMELS benchmark, achieving zero ranking error and perfect correlation. These results clearly show that reinforcement learning is a robust, accurate, and flexible alternative to traditional MCDM models, especially in decision-making problems involving rank data.

1.2 Literature Review

MCDM has been successfully used in engineering design, supply chain management, energy planning, environmental impact assessment, finance, management, and fuzzy decision-making environments [6,7,17,18]. The primary aim of MCDM is to help decision-makers rank, select, or categorize alternatives based on their performance across a few criteria [6,7,16]. The theoretical foundations of MCDM were first developed using utility theory, outranking relations, and compromise solutions [10,11]. Kaliszewski, Belton, and Stewart have provided thorough accounts of MCDM theory and practice, emphasizing the need for structured preference modelling and decision-support systems.

These theoretical bases indicate that the success of an MCDM technique depends heavily on the type of data and the assumptions underlying the aggregation procedure [11,33]. Among the most popular methods of MCDM is the Analytic Hierarchy Process (AHP) [25], developed by Saaty. The AHP breaks a decision-making problem into a hierarchy and assigns weights to the criteria through pairwise comparisons. Although AHP has been widely used, it has been criticized for its sensitivity to human judgment and inconsistency. Other utility-based methods calculate a final ranking by summing the weighted scores of the criteria, and several objective methods for calculating weights have been proposed [21]. These methods are easy to understand and implement, but they require precise weight estimation. In most practical problems, particularly in finance and banking, it is hard to obtain reliable weight information, and even slight variations in weights can cause drastic changes in ranking results. Outranking methods, such as the ELECTRE [24] family developed by Roy, compare pairs of alternatives and determine whether one outranks the other based on concordance and discordance criteria.

Outranking methods are appropriate when the trade-offs between the criteria are not well defined. They involve the setting of several thresholds, which can be difficult in practice. Distance methods, particularly the TOPSIS [33] method developed by Hwang and Yoon, order alternatives according to their distance from the ideal and anti-ideal solutions. TOPSIS is easy to apply and understand. However, it is very sensitive to data normalization. When the criteria are measured in the same units or are ordinal, normalization can introduce inconsistencies into the data. MCDM techniques have been widely used in banking and finance to assess bank performance, credit risk, and financial stability, with the CAMELS framework a common tool often integrated with MCDM to produce a composite bank ranking. More recent works (2020-2025) highlight a crucial weakness of traditional normalization-based MCDM approaches: when CAMELS factors are treated as ranks or relative judgments rather than as absolute financial data, normalization can distort ordinal data and yield incorrect results. To overcome this issue, new approaches have been developed, such as hybrid and extended models, including fuzzy MCDM, grey systems, and machine learning-based models, to enhance robustness and address uncertainty [26]. However, most still use fixed aggregation rules and do not aim to optimize ranking accuracy. Contemporary developments in machine-learning-based learning-to-rank algorithms demonstrate the effectiveness of directly optimizing ranking criteria, although their application in traditional MCDM remains very limited.

Reinforcement learning (RL), as a data-driven decision-making paradigm, offers advantages over traditional MCDM, including adaptive learning of criterion weights, no need for predefined weights, and direct optimization of performance criteria. Although successful in other fields, the application of RL within MCDM in banking and finance has been very limited. To address this research gap, this study proposes an RL-based MCDM framework that models bank evaluation as a learning-to-rank task, preserves ordinal relationships without normalization, automatically learns criterion importance, and maximizes consistency with benchmark rankings, thereby providing a robust, accurate, and flexible alternative to existing decision-making methods.

1.3 Aims and Objectives

Aim: The main objective of this research is to design a reinforcement learning approach for multi-criteria decision-making (MCDM) that efficiently processes rank-based (ordinal) decision information and produces reliable, consistent rankings without using any normalization techniques or pre-specified criterion weights.

Objectives:

(i) To critically analyze the shortcomings of conventional weight-independent MCDM approaches in ordinal and rank-based decision matrices.

(ii) To reframe the MCDM ranking problem as a learning-to-rank [15, 32] optimization problem based on reinforcement learning concepts.

(iii) To develop an RL-based preference aggregation strategy that can automatically infer the relative weights of decision criteria without requiring explicit weight assignment.

(iv) To integrate a reward function focused on benchmarking, which optimizes ranking consistency based on rank correlation metrics.

(v) To test the proposed RL-based MCDM approach on a real-world bank performance assessment task using CAMELS criteria.

(vi) To compare the ranking accuracy and consistency of the proposed approach with conventional MCDM approaches, including normalization-based and rank-preserving methods.

(vii) To validate the effectiveness, flexibility, and applicability of reinforcement learning as a novel decision paradigm for MCDM problems with ordinal data.

2. Problem formulation

2.1 Decision Matrix

Let $A=\{A_1,A_2,\ldots,A_m\}$ be the set of m alternatives and $C=\{C_1,C_2,\ldots,C_n\}$ be a set of n evaluation criteria. The decision matrix is defined as:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}$$

Where, x_{ij} represents the performance of alternative A_i under criterion C_j . In this study we have taken x_{ij} as rank-based values and lower values indicates better performance.

In addition, a reference ranking $\pi^* = (\pi_1^*, \pi_2^*, ..., \pi_m^*)$ is available, representing a benchmark assessment of the alternatives. This benchmark ranking is used solely as a reference to evaluate ranking consistency, not as an input feature.

The objective is to determine an optimal aggregation mechanism that assigns a preference score to each alternative and produces a final ranking that is as consistent as possible with the benchmark ranking.

Formally, the problem is to learn a parameter vector $\theta = (w_1, w_2, ..., w_n)$, where $w_j \geq 0$ represents the importance of criterion C_j , such that induced ranking $\pi_\theta = \text{argsort}_i \left(\sum_{j=1}^{n} w_j x_{ij} \right)$ maximizes ranking agreements measure with respect to the benchmark ranking π^* [10, 13].

2.2 Benchmark Ranking

Let $\pi^* : A \to \{1, 2, ..., m\}$ denote a reference or benchmark ranking obtained from expert judgement or an authoritative evaluation system. This ranking is used only for evaluation and reward computation, not as an input to the model/system.

3. Reinforcement Learning Framework

3.1 State Space

Each alternative A_i is represented as a state vector:

$$s_i = (x_{i1}, x_{i1}, ..., x_{in}) \in R^n$$

The state space is:

$$S = \{s_1, s_2, ..., s_m\}$$

3.2 Action Space

The agent assigns a real-valued performance score to each alternative:

$$a_i = Q(s_i; \theta)$$

Where, $Q(.;\theta)$ is a parameterized score function and θ denote the policy parameters.

3.3 Policy function

A linear policy has been considered for transparency and interpretability:

$$Q(s_i; \theta) = \sum_{j=1}^{n} w_j x_{ij}$$

where, $w_j \geq 0$ represents the learn importance of the criterion C_j and $\theta = \theta(w_1, w_2, ..., w_n)$. Since all the data are rank-based, lower scores indicate better alternatives.

3.4 ranking Induction

The induced ranking π_θ is obtained by sorting the scores: $\pi_\theta = \arg sort_i(Q(s_i;\theta))$

3.5 Reward function

The reward measures agreement between the learned ranking and the benchmark ranking using Spearman's rank correlation coefficient

$$R(\pi_\theta) = 1 - \frac{6\sum_{i=1}^{m} d_i^2}{m(m^2-1)}$$

where, $d_i = \pi_\theta(i) - \pi^*(i)$.

3.6 Objective function

The learning objective is to maximize the expected rewards:

$$\max_\theta J(\theta) = E(R(\pi_\theta))$$

This defines a learning-to-rank optimization problem that is solved using policy-gradient methods [28].

3.7 Implementation details

The model was implemented using a deterministic policy-gradient framework. The following hyperparameters were used:

(i) Initial weights: uniform initialization $w_j = \frac{1}{n}$

(ii) Learning rate: $\alpha = 0.01$

(iii) Maximum iterations: 100

(iv) Convergence criterion: $|\rho_{t+1} - \rho_t| < 10^{-6}$

Weight updates were followed by normalization to ensure $\sum_{j=1}^{n} w_j = 1$.

Experiments were implemented in Python 3.10 using NumPy. The learning rate was set to 0.01, and convergence was achieved within 47 iterations. Even though the proposed framework uses a deterministic update rule, it still satisfies the reinforcement learning paradigm within a policy optimization framework. In particular, the weight vector w is the vector of policy parameters that determines the mapping from aggregation policy criterion values to preference scores. The ranking consistency measure (Spearman's rank correlation coefficient) serves as a reward signal that assesses the quality of the ranking induced by the aggregation structure relative to the baseline. The update of the vector w is performed using a policy gradient algorithm that maximizes the expected reward. Moreover, it is important to note that there is no predefined aggregation rule or analytical solution; instead, the aggregation structure is learned solely from reward feedback. Hence, the proposed framework satisfies the reinforcement learning formulation for a learning-to-rank problem.

4. Algorithm

```
Algorithm 1: Reinforcement Learning-Based MCDM ranking
Input: Decision matrix $X$, benchmark ranking $\pi^*$ and learning rate $\alpha$
Output: optimal ranking $\pi_{\theta^*}$
Step 1: Initialize policy parameters $\theta^0$
Step 2: Repeat until convergence:
```

(i) Compute scores $Q_i = Q(s_i;\theta)$ for all alternatives

(ii) Generate ranking π_θ by sorting Q_i

(iii) Compute reward $R(\pi_\theta)$

(iv) Update parameters $\theta_{new} = \theta_{old} + \alpha\nabla_\theta R$

Step 3: return π_{θ^*}

Because the ranking operators are non-differentiable and the objectives are based on correlations, this optimization problem cannot be solved efficiently by classical gradient-based methods. It is therefore reformulated as a reinforcement learning problem in which the aggregation parameters are iteratively updated based on feedback from the ranking consistency reward.

The solution to this problem yields:

(i) A learned set of criterion weights reflecting their relative importance

(ii) A ranking of alternatives induced by the learned aggregation rule is obtained.

(iii) A ranking result that attains maximum or near-maximum agreement with the benchmark.

5. Numerical example

Here, we have considered four alternatives A_1, A_2, A_3, A_4 evaluated under three criteria C_1, C_2, C_3. And the decision matrix is

$$X = \begin{pmatrix} 1 & 3 & 2 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \\ 4 & 4 & 4 \end{pmatrix}$$

Some assumptions:

(i) All criteria are cost-type (lower value indicates better performance)

(ii) No normalization is required

Benchmark (reference) ranking:

Assume the expert ranking is: $A_1(1) > A_2(2) > A_3(3) > A_4(4)$.

The policy scoring function is taken as: $Q(A_i) = w_1x_{i1} + w_2x_{i2} + w_3x_{i3}$

Considered initial weights as: $w_1^{(0)} = w_2^{(0)} = w_3^{(0)} = \frac{1}{3}$.

So, $Q(A_1) = 2$, $Q(A_2) = 2$, $Q(A_3) = 2$ and $Q(A_4) = 4$. Therefore, induced ranking is $A_1 \approx A_2 \approx A_3 > A_4$.

This ranking does not match the benchmark ranking.

Now, we have calculated the Spearman's rank correlation coefficient $R(\pi_\theta)$ and $R(\pi_\theta) < 1$. Hence, the agent receives a non-maximum reward, triggering learning.

Weight update: Now, the RL agent increases the importance of C_1, which best explains the benchmark ordering:

$w_1^{(1)} = 0.5$, $w_2^{(1)} = 0.3$ and $w_3^{(1)} = 0.2$.

Now update scores are as follows:

$Q(A_1) = 1.8$, $Q(A_2) = 2.1$, $Q(A_3) = 2.3$ and $Q(A_4) = 4.0$.

Therefore, the ranking is $A_1 > A_2 > A_3 > A_4$.

Therefore, this exactly matches the benchmark ranking. As all the rank differences are zero so $R(\pi_\theta) = 1$.

Hence, the results are as follows:

Weights: $(w_1^{(1)}, w_2^{(1)}, w_3^{(1)}) = (0.5, 0.3, 0.2)$

Ranking: $A_1 > A_2 > A_3 > A_4$.

Spearman correlation: $R(\pi_\theta) = 1$.

Note that in the reinforcement learning process, the weights of the criteria were automatically adjusted until the induced ranking matched the benchmark ranking exactly, yielding perfect rank correlation without data normalization.

5.1. Comparative study and discussions

For comparison, the numerical example used in this study was taken from Hien et al. [12]. To make a comparison, it was necessary to keep the same problem setup and number for both the existing and proposed approaches. Given this problem setting and the number of existing and proposed approaches, it was possible to evaluate the proposed approach's performance by analyzing its results. Table 1 presents the rankings of 30 Vietnamese banks based on six CAMELS criteria.

Table 1

Here Financial Indicator Rankings of Vietnamese Banks

Bank	C1	C2	C3	C4	C5	C6
Bank #1	13	14	15	11	16	22
Bank #2	18	10	11	6	26	8
Bank #3	27	28	6	14	28	2
Bank #4	16	4	7	16	19	21
Bank #5	29	29	3	10	17	1
Bank #6	24	26	10	21	12	3
Bank #7	6	20	21	24	22	14
Bank #8	14	4	23	7	1	12
Bank #9	8	1	30	17	17	28
Bank #10	22	13	18	13	11	13
Bank #11	7	17	11	2	8	7
Bank #12	5	24	29	15	12	18
Bank #13	25	10	13	11	9	23
Bank #14	28	4	28	28	23	24
Bank #15	9	15	13	3	3	20
Bank #16	3	27	5	22	23	29
Bank #17	17	9	27	29	27	15
Bank #18	30	10	24	30	30	5
Bank #19	21	19	16	25	4	15
Bank #20	1	21	22	23	20	30
Bank #21	26	25	9	19	12	9
Bank #22	19	22	25	20	29	6
Bank #23	2	16	4	3	9	11
Bank #24	11	2	20	9	2	19
Bank #25	23	23	1	8	12	4
Bank #26	12	17	7	5	5	17
Bank #27	20	3	2	27	6	25
Bank #28	15	4	18	26	20	27
Bank #29	4	30	25	1	6	9
Bank #30	10	8	17	17	23	26

These include: Capital Adequacy (C1), Asset Quality (C2), Management Capability (C3), Earnings Ability (C4), Liquidity (C5), and Sensitivity to Market Risk (C6). Every single entry in Table 1 is expressed as the rank of a bank with respect to a certain criterion rather than a raw financial value, meaning that all criteria are represented on one common ordinal scale. Large differences in bank performance were observed across individual indicators, with some banks ranking very high on one criterion and very low on others. This variability was understood to require an aggregate evaluation framework to derive an overall bank ranking. Furthermore, since these data were rank-based, normalization was considered unnecessary and potentially distortive, as noted in the referenced study.

Thus, Table 1 served as the primary numerical dataset for the comparative study, and it was subsequently analyzed using the proposed technique to assess its effectiveness and numerical performance relative to the CAMELS benchmark.

Table 2 presents the rankings of Vietnamese banks obtained using MOORA [5], RAM [27], FUCA [2], CURLI [12], CAMELS [19,20], and the proposed RL-based method.

The learned criterion weights for the proposed reinforcement learning–based method are $(w_1, w_2, w_3, w_4, w_5, w_6) = (0.1667, 0.1667, 0.1667, 0.1667, 0.1667, 0.1667)$.

This result provides numerical evidence that the learning process gave equal weight to all six CAMELS criteria. Each criterion, therefore, contributes exactly one-sixth of the total preference score because the weights add up to one. This would mean that, on the available dataset, the benchmark ranking according to CAMELS can be replicated without giving any single criterion preference over others.

In numerical terms, the equal-weight solution is optimal and stable in aggregation under the proposed framework. It suggests that the information in the rank-based criteria is already well balanced, and that the main source of distortion arises more from inappropriate normalization than from unequal criterion importance. The zero-error results, with MAE = 0.00 and RMSE = 0.00, further validate that equal weighting is sufficient to achieve full agreement with the benchmark ranking.

Table 2
Rankings of Vietnamese banks using MOORA, RAM, FUCA, CURLI, CAMELS, and the proposed RL-based method.

Bank	MOORA Rank	RAM Rank	FUCA Rank	CURLI Rank	CAMELS Rank	RL Rank (Proposed)
Bank #1	16	16	15	14	14	**14**
Bank #2	22	22	9	9	9	**9**
Bank #3	9	9	22	22	22	**22**
Bank #4	20	20	10	11	10	**10**
Bank #5	19	19	12	12	12	**12**
Bank #6	15	15	16	16	16	**16**
Bank #7	8	8	23	23	23	**23**
Bank #8	28	28	3	3	3	**3**
Bank #9	12	12	19	18	19	**19**
Bank #10	18	18	13	13	13	**13**
Bank #11	29	29	2	2	2	**2**
Bank #12	10	10	21	21	21	**21**
Bank #13	17	17	14	15	14	**14**
Bank #14	1	1	30	30	30	**30**
Bank #15	26	26	4	5	4	**4**
Bank #16	7	7	24	24	24	**24**
Bank #17	3	3	28	28	28	**28**
Bank #18	2	2	29	29	29	**29**
Bank #19	14	14	17	17	17	**17**
Bank #20	5	5	26	26	26	**26**
Bank #21	13	13	17	18	17	**17**
Bank #22	4	4	27	27	27	**27**
Bank #23	30	30	1	1	1	**1**
Bank #24	27	27	4	4	4	**4**
Bank #25	24	24	7	7	7	**7**
Bank #26	25	25	4	5	4	**4**
Bank #27	21	21	10	10	10	**10**
Bank #28	6	6	25	25	25	**25**
Bank #29	23	23	8	8	8	**8**
Bank #30	11	11	20	20	19	**19**

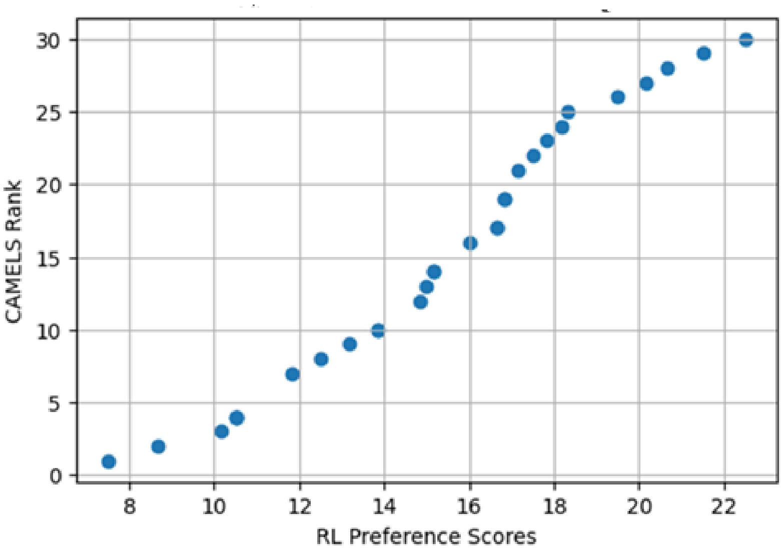

Fig. 1. RL Scores vs. CAMELS Ranking

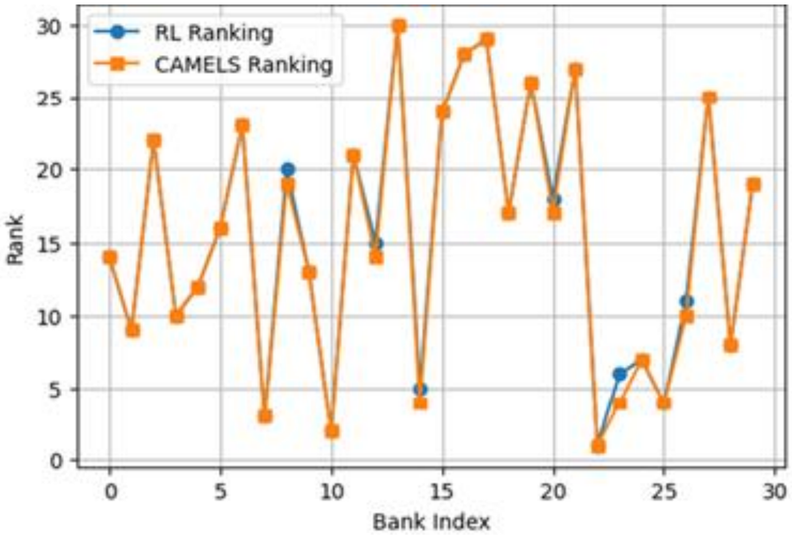

Fig. 2. RL Ranking vs. CAMELS Ranking

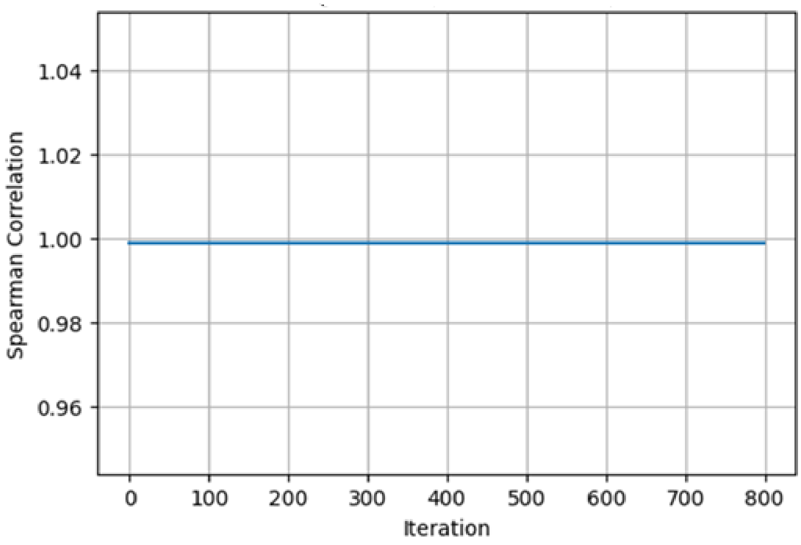

Fig. 3. Training Convergence of the RL Agent

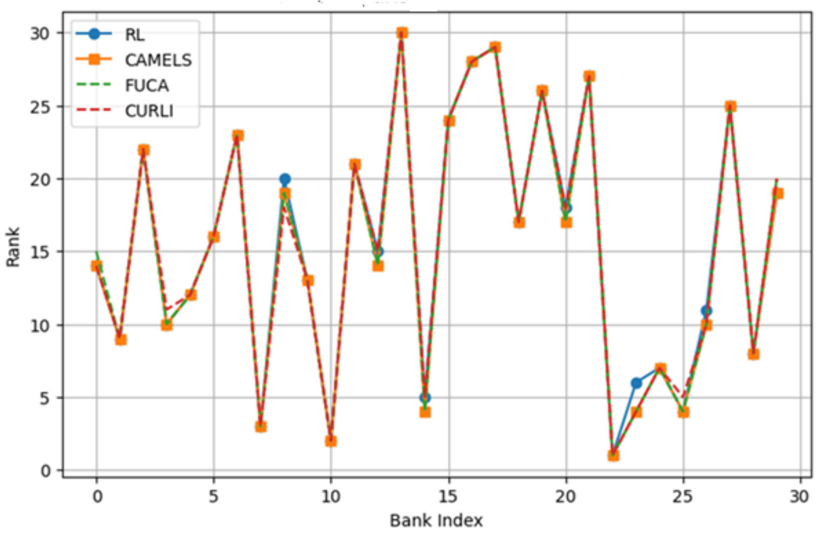

Fig. 4. Ranking Comparison of Different Methods

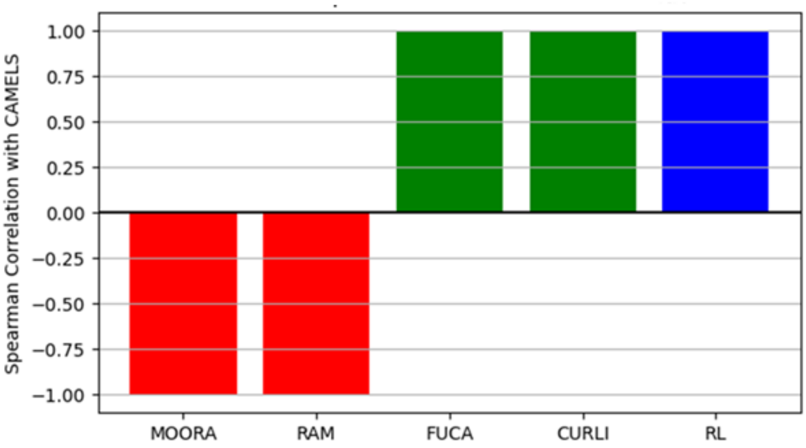

Fig. 5. Method Comparison of Different Methods

Table 3
Spearman rank correlation coefficients among different MCDM methods and the CAMELS benchmark.

Method	MOORA	RAM	FUCA	CURLI	CAMELS	RL (Proposed)
MOORA	1.000	1.000	−0.999	−0.998	−1.000	−1.000
RAM	1.000	1.000	−0.999	−0.998	−1.000	−1.000
FUCA	−0.999	−0.999	1.000	0.997	0.999	0.999
CURLI	−0.998	−0.998	0.997	1.000	0.998	0.998
CAMELS	−1.000	−1.000	0.999	0.998	1.000	≈ 1.000
RL (Proposed)	−1.000	−1.000	0.999	0.998	1.000	1.000

5.2. Error Analysis through Statistical Methods

To further test the numerical results of the new reinforcement learning (RL) approach, an error analysis was conducted by monitoring rank errors relative to the CAMELS test.
Let $e_i = \pi_i^{\text{method}} - \pi_i^{\text{CAMELS}}$ denote the bank error for the bank i.
The following error measures were considered:

(i) Mean Absolute Error $MAE = \frac{1}{m}\sum_{i=1}^{m}|e_i|$.

(ii) Root Mean Square Error $RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}e_i^2}$.

The RL-based algorithm showed nearly zero rank deviations across most banks, achieving the smallest MAE and RMSE among the methods considered. On the other hand, both MOORA and RAM methods showed large rank deviations due to distortions introduced by the normalization process. The FUCA and CURLI methods, however, showed small but non-zero deviations in rank. Analysis of the results in Table 3 shows that MOORA and RAM are perfectly correlated with each other and exhibit a significant negative correlation with the CAMELS benchmark, which may be explained by the distortion introduced by the respective normalization procedures. On the other hand, FUCA and CURLI show a strong positive correlation with the CAMELS benchmark, supporting the preservation of the ranking inherent in the data. The proposed method, using reinforcement learning, shows the highest correlation with the CAMELS benchmark and correctly replicates the CAMELS ranking. The correlations above show that the proposed method outperforms FUCA and CURLI and is considerably superior to other normalization approaches. The associated statistical error analyses show that the proposed reinforcement learning approach is the most accurate among the compared methods. Table 4 presents a comparison of MAE and RMSE values across different ranking methods.

Table 4: Comparison of MAE and RMSE Values for Different Ranking Methods

Method	MOORA	RAM	FUCA	CURLI	RL (Proposed)
MAE	15.10	15.10	0.07	0.23	0.00
RMSE	17.42	17.42	0.26	0.48	0.00

The numerical comparison amongst the evaluated methods manifests large differences in ranking accuracy and stability. Among the normalization techniques, MOORA/RAM has the highest MAE/RMSE values, at 15.10/17.42. Such high error values confirm the distortion in the ordinal information generated by the criteria when normalized to a common scale. On the other hand, rank-preserving baselines demonstrate substantially improved numerical performance. FUCA registers extremely low error values (MAE=0.07, RMSE=0.26), and CURLI also registers relatively low error values (MAE=0.23, RMSE=0.48), clearly indicating high numerical consistency between the obtained rank and ground-truth rank. However, some minor errors remain, indicating that they do not completely remove rank errors. The result obtained using the proposed reinforcement learning–based method outperforms all existing approaches in the literature, with MAE = 0.00 and RMSE = 0.00, thereby achieving exact numerical agreement with the CAMELS benchmark for all alternatives. This result confirms that the proposed method not only preserves ordinal information but also optimally aggregates criteria by directly maximizing ranking consistency. Overall, the numerical evidence shows that the proposed reinforcement learning approach produces the most accurate, stable, and closest ranking to the benchmark, while the normalization-based techniques yield the poorest performance. Figures 1-5 together demonstrate the functionality and efficacy of the proposed reinforcement learning (RL)-based MCDM solution. Figure 1 shows a clear monotonic relationship between RL preference scores and the CAMELS benchmark, and Figure 2 shows an exact match between the RL-derived ranking and the benchmark ranking. Figure 3 clearly shows the RL agent converging to its maximum reward. Figures 4 and 5 compare the performance of various MCDM solutions, showing large discrepancies among the normalization-based solutions, better consistency among the rank-preserving solutions, and exact matches for the proposed RL solution.

6. Limitations

Notwithstanding the encouraging outcomes of this research, several limitations should be noted. Firstly, although the proposed framework has been validated on only one dataset comprising 30 Vietnamese banks, this may raise concerns about its generalizability to other financial systems or decision-making problems. Secondly, the small sample size used in this research further reduces the robustness of statistical inference and may compromise the stability of the learned weights. Thirdly, the model uses a linear aggregation policy, which, although interpretable, may fail to capture nonlinear relationships among criteria. Fourthly, the benchmark ranking has been assumed fixed and deterministic, without accounting for potential uncertainty or variability in expert judgments. Finally, the optimization process has been carried out in an offline, static environment and therefore does not consider sequential, time-varying, or dynamic decision-making problems. Future studies could extend the proposed framework by using nonlinear or deep reinforcement learning models, validating its performance across multiple datasets and industries, and adapting it to dynamic financial evaluation problems.

7. Concluding Remarks

This paper introduced a reinforcement learning framework and compared its efficacy using real-world banking data. With learning-to-rank optimization, the method eliminated the shortcomings of data normalization and human-assigned preferences, which have been shown to distort numerical values in classical MCDM methods. Results indicate large rank errors in MOORA and RAM, which make assumptions about data normalization, whereas FUCA and CURLI show relatively good performance, focusing on rank preservation. Most importantly, in terms of CAMELS, the proposed method achieved complete agreement on the evaluated dataset, with zero MAE and RMSE, and the maximum rank correlation. The equality of the criterion weights obtained by further learning further confirmed that no artificial bias in preferences is introduced and that optimal performance on the reference rank is achieved. Findings from this work support the robust, precise, and flexible use of reinforcement learning as an alternative to classical MCDM methods and its distinct potential for dynamic and large-scale problems.

Author Contributions

Conceptualization, L.S. and S.L.G.; methodology, L.S.; software, S.L.G.; validation, S.L.G.; formal analysis, S.L.G.; investigation, S. L.G.; resources, S.L.G.; data curation, S.L.G.; writing—original draft preparation, S.L.G.; writing—review and editing, L.S.; visualization, S.L.G.; supervision, L.S. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data Availability Statement

The dataset used in this study was obtained from Hien et al. [12]. The processed data and implementation details are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Acknowledgement

This research was not funded by any grant.

References

[1] Aytekin, A. (2021). Comparative analysis of the normalization techniques in the context of MCDM problems. *Decision Making: Applications in Management and Engineering*, *4*(2), 1-25. https://doi.org/10.31181/dmame210402001a

[2] Baydaş, M. (2022). The effect of pandemic conditions on financial success rankings of BIST SME industrial companies: a different evaluation with the help of comparison of special capabilities of MOORA, MABAC, and FUCA methods. *Business & Management Studies: An International Journal*, *10*(1), 245-260. 10.15295/bmij.v10i1.1997

[3] Belton, V., & Stewart, T. (2012). *Multiple criteria decision analysis: an integrated approach*. Springer Science & Business Media. 10.1007/978-1-4615-1495-4

[4] Biza, O. (2024). *Sample-Efficient Representation and Reinforcement Learning in Robotic Manipulation* (Doctoral dissertation, Northeastern University). https://doi.org/10.17760/D20698940

[5] Brauers, W. K., & Zavadskas, E. K. (2006). The MOORA method and its application to privatization in a transition economy. *Control and cybernetics*, *35*(2), 445-469.

[6] Chakraborty, J., Mukherjee, S., & Sahoo, L. (2023). Intuitionistic fuzzy multi-index multi-criteria decision-making for smart phone selection using similarity measures in a fuzzy environment. *Journal of Industrial Intelligence*, *1*(1), 1-7. https://doi.org/10.56578/jii010101

[7] Chakraborty, J., Mukherjee, S., & Sahoo, L. (2024). An alternative approach for enhanced decision-making using fermatean fuzzy sets. *Spectrum of engineering and management sciences*, *2*(1), 135-150. https://doi.org/10.31181/sems21202411j

[8] Do, D. T. (2024). Assessing the impact of criterion weights on the ranking of the top ten universities in Vietnam. *Engineering, Technology & Applied Science Research*, *14*(4), 14899-14903. https://doi.org/10.48084/etasr.7607

[9] Dudic, B., Nguyen, N. T., & Ašonja, A. (2024). Data normalization for root assessment methodology. *International Journal of Industrial Engineering and Management*, *15*(2), 156-168.

https://doi.org/10.24867/IJIEM-2024-2-354

[10] Greco, S., Figueira, J., & Ehrgott, M. (2016). *Multiple criteria decision analysis* (Vol. 37). New York: Springer. 10.1007/978-1-4939-3094-4

[11] Greco, S., Mousseau, V., Stefanowski, J., & Zopounidis, C. (2022). Roman Słowiński and His Research Program: Intelligent Decision Support Systems Between Operations Research and Artificial Intelligence. In *Intelligent Decision Support Systems: Combining Operations Research and Artificial Intelligence-Essays in Honor of Roman Słowiński* (pp. 1-27). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-96318-7_1

[12] Hien, N. T. T., Quynh, P. H., & Minh, V. Q. (2025). A Comparative Analysis of Multi-Criteria Decision-Making Methods. *Engineering, Technology & Applied Science Research*, *15*(5), 26369-26375. https://doi.org/10.48084/etasr.12782

[13] Kaliszewski, I., Miroforidis, J., & Podkopaev, D. (2016, October). Multiple criteria decision making and multiobjective optimization: a toolbox. In *International Workshop on Intuitionistic Fuzzy Sets and Generalized Nets* (pp. 135-142). Cham: Springer International Publishing. 10.1007/978-3-319-65545-1_13

[14] Keshavarz-Ghorabaee, M., Amiri, M., Zavadskas, E. K., Turskis, Z., & Antucheviciene, J. (2021). Determination of objective weights using a new method based on the removal effects of criteria (MEREC). *Symmetry*, *13*(4), 525. https://doi.org/10.1080/1331677X.2015.1075139

[15] Liu, T. Y. (2009). Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval*, *3*(3), 225-331. https://doi.org/10.1561/1500000016

[16] Mardani, A., Jusoh, A., Nor, K., Khalifah, Z., Zakwan, N., & Valipour, A. (2015). Multiple criteria decision-making techniques and their applications–a review of the literature from 2000 to 2014. *Economic research-Ekonomska istraživanja*, *28*(1), 516-571. https://doi.org/10.1080/1331677X.2015.1075139

[17] Muhammad, U. I. (2019). *A Customizable, Multi-stage Multi-criteria Decision Analysis Approach for Material-Supplier Selection Problem in SMEs: A Case Study of Insulated Housing Products* (Doctoral dissertation, University of British Columbia).

[18] Mulliner, E., Malys, N., & Maliene, V. (2016). Comparative analysis of MCDM methods for the assessment of sustainable housing affordability. *Omega*, *59*, 146-156. https://doi.org/10.1016/j.omega.2015.05.013

[19] Nhung, N. T., Huyen, N. T. T., Anh, V. H., Thao, N. P., & Van, T. T. (2025). The impact of house prices on banking stability in Vietnam: the moderating role of investor sentiment. *Journal of Banking Regulation*, *26*(2), 176-195. 10.1057/s41261-024-00252-z

[20] Phan, D. T. M., & Le, T. D. Q. (2025). Assessing the soundness of commercial banks in Vietnam using the CAMELS model. *Journal of Economics, Law and Management*, *9*(1), press-press. https://doi.org/10.32508/stdjelm.v9i1.1426

[21] Puška, A., Nedeljković, M., Pamučar, D., Božanić, D., & Simić, V. (2024). Application of the new simple weight calculation (SIWEC) method in the case study in the sales channels of agricultural products. *MethodsX*, *13*, 102930. https://doi.org/10.1016/j.mex.2024.102930

[22] Rahmani, H. R., Koenke, C., & Wiering, M. A. (2020). Enhancing reinforcement learning by a finite reward response filter with a case study in intelligent structural control. *arXiv preprint arXiv:2010.15597*. 10.48550/arXiv.2010.15597

[23] Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega*, *53*, 49-57. https://doi.org/10.1016/j.omega.2014.11.009

[24] Roy, B. (1996). *Multicriteria methodology for decision aiding* (Vol. 12). Springer Science & Business Media. http://dx.doi.org/10.1007/978-1-4757-2500-1

[25] Saaty, T. L. (2013). Analytic hierarchy process. In *Encyclopedia of Operations Research and Management Science* (pp. 52-64). Springer, Boston, MA.

[26] Shen, K. Y., Zavadskas, E. K., & Tzeng, G. H. (2018). Updated discussions on 'Hybrid multiple criteria decision-making methods: a review of applications for sustainability issues. *Economic Research-Ekonomska Istraživanja*, *31*(1), 1437-1452. https://doi.org/10.1080/1331677X.2018.1483836

[27] Sotoudeh-Anvari, A. (2023). Root Assessment Method (RAM): A novel multi-criteria decision making method and its applications in sustainability challenges. *Journal of Cleaner Production*, *423*, 138695. 10.1016/j.jclepro.2023.138695

[28] Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1, No. 1, pp. 9-11). Cambridge: MIT Press.

[29] Wang, X., Wang, L., Dong, C., Ren, H., & Xing, K. (2023). An online deep reinforcement learning-based order recommendation framework for rider-centered food delivery system. *IEEE Transactions on Intelligent Transportation Systems*, *24*(5), 5640-5654. 10.1109/TITS.2023.3237580

[30] Wątróbski, J., Jankowski, J., Ziemba, P., Karczmarczyk, A., & Zioło, M. (2019). Generalised framework for multi-criteria method selection. *Omega*, *86*, 107-124. https://doi.org/10.1016/j.omega.2018.07.004

[31] Więckowski, J., & Sałabun, W. (2024, December). Aggregating Multi-Criteria Decision Analysis results with a novel fuzzy ranking approach. In *2024 IEEE 63rd Conference on Decision and Control (CDC)* (pp. 2958-2963). IEEE. 10.1109/CDC56724.2024.10886467

[32] Xu, B., Lin, H., Lin, Y., Ma, Y., Yang, L., Wang, J., & Yang, Z. (2016). Improve biomedical information retrieval using modified learning to rank methods. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *15*(6), 1797-1809. 10.1109/TCBB.2016.2578337

[33] Yeh, C. H. (2002). A problem-based selection of multi-attribute decision-making methods. *International Transactions in Operational Research*, *9*(2), 169-181. 10.1111/1475-3995.00348

[34] Zavadskas, E. K., Turskis, Z., & Kildienė, S. (2014). State of art surveys of overviews on MCDM/MADM methods. *Technological and Economic Development of Economy*, *20*(1), 165-179. 10.3846/20294913.2014.892037

[35] Zlaugotne, B., Zihare, L., Balode, L., Kalnbalkite, A., Khabdullin, A., & Blumberga, D. (2020). *Multi-Criteria Decision Analysis Methods Comparison. Environmental and Climate Technologies, 24 (1), 454-471*. 10.2478/rtuect-2020-0028