IJITEB
Vol. 8
No. 2
2026

International Journal of Information Technology and Business

http://ejournal.uksw.edu/ijiteb

Multi-Objective Reinforcement Learning for Supply Chain Optimization in Indonesia's Free Nutritious Meal Program

Budi Yanto[1], Adyanata Lubis[2], Elyandri Prasiwiningrum[3], Rusmin Saragih[4], Romy Wahyuny[5]

[1]Department of Informatic Engineering, University of Pasir Pengaraian, Indonesia
[2,3]Department of Computer Sciences, Universitas Rokania, Indonesia
[4]STMIK Kaputama, Medan, Indonesia
[5]Department of Bachelor of Midwifery Program and Profession, University of Pasir Pengaraian, Indonesia

Abstract : Indonesia's Free Nutritious Meal Program (MBG) requires an efficient and adaptive supply chain system to ensure timely distribution, cost efficiency, and adequate nutritional delivery for a large number of beneficiaries. However, conventional supply chain approaches are generally static and unable to respond effectively to dynamic demand, supply uncertainty, and logistical constraints. This study proposes a Multi-Objective Reinforcement Learning (MORL) model to optimize the MBG supply chain by simultaneously considering distribution cost, delivery timeliness, service level, nutritional adequacy, and food waste reduction. The model is developed using a simulation-based environment representing real-world supply chain conditions, including demand variability, transportation limitations, and kitchen capacity constraints. The results show that the proposed approach achieves cost reductions of 15–22%, improves delivery timeliness by 18–25%, maintains a service level above 90%, increases nutritional fulfillment by 12–18%, and reduces food waste by 10–15% compared to baseline methods. Sensitivity analysis further demonstrates the robustness of the model, with minimal performance degradation under disruption scenarios. These findings indicate that Reinforcement Learning provides a scalable and adaptive solution for optimizing large-scale public food distribution systems. The proposed model contributes both theoretically by integrating multi-objective optimization within an RL framework and practically by supporting data-driven decision-making for improving the effectiveness of the MBG program in Indonesia.

Keyword: Food Supply Chain Optimization; Multi-Objective Optimization; Public Food Distribution; Reinforcement Learning;

1. Introduction

The implementation of large-scale public food assistance programs has become a strategic priority for many developing countries in addressing food insecurity, malnutrition, and social inequality. In Indonesia, the Free Nutritious Meal Program (Program Makan Bergizi Gratis – MBG) has emerged as a flagship national initiative aimed at improving the nutritional status of millions of school-age children and vulnerable populations[1], [2]. Recent national projections indicate that the program is expected to serve tens of millions of beneficiaries across geographically diverse regions, including remote and archipelagic areas [3], [4]. This scale introduces significant logistical challenges, particularly in ensuring timely delivery, cost efficiency, and equitable food distribution. According to World Health Organization, malnutrition remains a critical global issue, contributing to developmental and health disparities, especially in low- and middle-income countries. In Indonesia, disparities in food access and nutritional intake persist across regions, reinforcing the need for an efficient and adaptive supply chain system to support national food programs[5].

From an operational perspective, food supply chains for public programs differ substantially from commercial supply chains due to their multi-objective nature, which includes not only cost minimization but also nutritional adequacy, equity of distribution, and service reliability[6], [7]. Traditional supply chain

optimization methods such as linear programming, mixed-integer programming, and heuristic-based approaches have been widely applied in logistics and distribution systems[8], [9]. However, these approaches are often limited in their ability to handle dynamic environments characterized by uncertain demand, fluctuating supply, and real-time constraints. In the context of MBG [10], [11], these limitations are further exacerbated by the heterogeneity of regional infrastructure, transportation barriers, and variability in food availability[12], [13]. As a result, static optimization models may lead to suboptimal decisions, delays in distribution, and inefficiencies in resource allocation[14], [15].

Recent advances in Artificial Intelligence (AI), particularly Reinforcement Learning (RL), have demonstrated strong potential in addressing complex decision-making problems under uncertainty[16], [17]. RL enables an agent to learn optimal policies through continuous interaction with the environment, making it highly suitable for dynamic supply chain optimization [3], [4]. Studies have shown that RL-based approaches can outperform traditional optimization techniques in areas such as inventory management, routing, and demand forecasting [5]. Furthermore, RL has been successfully applied in large-scale logistics systems, where adaptability and real-time responsiveness are critical [6]. Despite these advancements, the application of RL in public food distribution systems especially those incorporating nutritional and equity considerations remains relatively underexplored.

Several prior studies have focused on optimizing food supply chains using machine learning and data-driven approaches. For instance, research by Zhang et al[18]. explored demand forecasting in food logistics using deep learning models, while Ivanov [8] examined resilience in supply chain systems under disruption scenarios. However, most existing studies primarily emphasize economic efficiency and operational performance, with limited attention to social objectives such as nutritional fulfillment and equitable distribution. Moreover, few studies have addressed the integration of multi-objective optimization within a Reinforcement Learning framework tailored to public-sector food programs. This indicates a clear research gap in developing intelligent, adaptive, and socially-aware optimization models for large-scale food assistance initiatives like MBG[19], [20].

In response to these challenges, this study proposes a multi-objective Reinforcement Learning-based approach for optimizing the supply chain of the Free Nutritious Meal Program in Indonesia[21], [22]. The sentence states that the model aims to simultaneously optimize distribution cost, delivery timeliness, and nutritional adequacy, but the challenges and trade-offs among these three objectives are not adequately explained. Please elaborate on: (1) why optimizing cost, timeliness, and nutrition simultaneously is inherently difficult (e.g., reducing cost may require fewer deliveries, which can compromise timeliness); (2) the specific impact of each objective conflict on decision-making for MBG program managers; and (3) how uncertainty in the operational environment compounds these trade-offs. This discussion is essential to justify the multi-objective RL approach.

Optimizing the MBG supply chain involves several interrelated decision-making challenges. Distribution cost is not only determined by transportation distance, but also by vehicle capacity, fuel consumption, route selection, kitchen location, and the frequency of delivery. Reducing cost without considering service quality may lead to delayed distribution or limited coverage for beneficiaries in remote areas. Delivery timeliness is also critical because nutritious meals must arrive within an acceptable time window to maintain food quality, safety, and consumption readiness. Delays may reduce the effectiveness of the program, increase food waste, and affect beneficiary satisfaction. At the same time, nutritional adequacy must be maintained by ensuring that each delivered meal meets the required nutritional standard in terms of energy, protein, and balanced food composition. In practice, these objectives often conflict with one another. A low-cost route may not always be the fastest route, while a nutritionally complete menu may require more expensive or less available ingredients. These trade-offs directly affect operational decision-making, including supplier selection, kitchen allocation, route planning, inventory control, and meal distribution priority. Therefore, an adaptive optimization model is required to support data-driven decisions under uncertain demand, supply variability, and logistical constraints.

Therefore, the main objective of this article is to develop and evaluate an intelligent supply chain optimization model based on Reinforcement Learning for the MBG program[23], [24]. Theoretically, this study contributes to the advancement of AI-driven supply chain optimization by incorporating multi-objective and socially-aware decision-making into RL frameworks. Practically, the findings are expected to provide actionable insights for policymakers and stakeholders in designing more efficient, adaptive, and equitable food distribution systems. Ultimately, this research aims to support the successful implementation of large-scale public nutrition programs and contribute to improving food security and public health outcomes in Indonesia.

2. Research Method

This study employs a quantitative, computational, and experimental approach based on simulation and Reinforcement Learning (RL) to optimize the supply chain of Indonesia's Free Nutritious Meal Program (MBG)[25], [26]. The method is designed to model

complex, dynamic, and uncertain logistics conditions where traditional static optimization methods are insufficient. The research adopts a Multi-Objective Reinforcement Learning (MORL) framework to simultaneously optimize cost efficiency[27], delivery timeliness, nutritional adequacy, and service level. This approach is grounded in the Markov Decision Process (MDP), which enables sequential decision-making under uncertainty [1], [2].

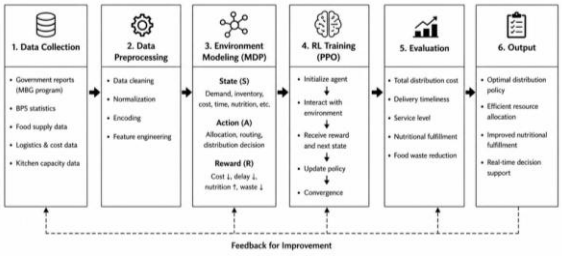


Fig. 1 research framework

Figure 1 presents the research framework, consisting of data collection, preprocessing, MDP-based modeling, reinforcement learning training (PPO), evaluation, and output generation. The framework aims to produce an optimized supply chain policy for the MBG program through adaptive and data-driven decision-making.

The data used in this study consist of secondary and simulated data derived from Indonesian contexts. Key inputs include demand volume, number of beneficiaries, distribution distance, transportation cost, food composition, and kitchen capacity[28], [29]. These data are obtained from national statistical reports, government policy documents, and prior studies related to food supply chains and public nutrition programs in Indonesia, including datasets from Badan Pusat Statistik[30], [31]. When real-time operational data are unavailable, simulation data are generated to represent realistic conditions such as demand fluctuations, supply variability, and logistical constraints.

The Methods chapter opens directly with mathematical formulas (Equations 1-6), which is abrupt and difficult to follow. According to standard academic practice, the methods section should begin with Figure 1 (the research framework) to give readers a conceptual overview of the overall approach before presenting mathematical detail. Please restructure this section so that Figure 1 and its explanation appear first, followed by the MDP formulation and the equations

$$\mathcal{M} = (S, A, P, R, \gamma) \quad (1)$$

$$R_t = \alpha NAR_t + \beta SL_t - \lambda TC_t - \mu W_t - \eta E_t \quad (2)$$

$$\pi^* = \arg\max_{\pi} E_{\pi} \left[\sum_{t=0}^{T} \gamma^t R_t\right] \quad (3)$$

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)} \quad (4)$$

$$R_t = \alpha NAR_t + \beta SL_t - \lambda TC_t - \mu W_t \quad (5)$$

$$\pi^* = \arg\max_{\pi} E \left[\sum_{t=0}^{T} \gamma^t R_t\right] \quad (6)$$

Equations (1)-(6) are listed without adequate explanation of the notation. Each symbol must be formally defined when it first appears. For example: in Eq. (1), define each element of the MDP tuple (S, A, P, R, gamma); in Eq. (2), define alpha, beta, lambda, mu, eta, NAR_t, SL_t, TC_t, W_t, E_t and their units/ranges; in Eq. (3), clarify what E_pi denotes and define the discount factor gamma; in Eq. (4), explain r_t(theta), pi_theta, and pi_theta_old (the PPO clipped ratio); in Eqs. (5)-(6), clarify how they differ from (2)-(3). A notation table or a dedicated paragraph defining all symbols is strongly recommended.

The unit of analysis is the supply chain node, comprising suppliers, kitchens, distribution centers, and beneficiaries. The selection uses purposive sampling, focusing on representative scenarios with valid operational parameters (demand, cost, distance, capacity). Data collection is conducted through documentation review and simulation modeling, followed by preprocessing steps including normalization, encoding, and outlier handling to ensure data consistency and model stability [3].

Table I. Input and Output Variables.

No	Variable	Symbol	Type	Description	Unit
1	Demand Volume	*Dt*	Integer	Number of meals required	meals/day
2	Beneficiaries	*Bt*	Integer	Number of recipients	persons
3	Distance	$Dist_t$	Float	Distribution distance	km
4	Transport Cost	*Ct*	Float	Cost per delivery	IDR
5	Delivery Time	*Tt*	Float	Travel time	hours
6	Kitchen Capacity	*Kt*	Integer	Production capacity	meals/day
7	Nutrition Score	*Nt*	Float	Nutritional adequacy index	%
8	Inventory Level	*It*	Float	Available food stock	kg
9	Service Level	*St*	Float	Demand fulfillment rate	%
10	Food Waste	*Wt*	Float	Unused food	kg
11	Output (Decision)	*At*	Action	Allocation & routing	-
12	Target Output	*Yt*	Float	Optimized performance	composite

The MDP components (S, A, R, P) are described only in bullet-point form with minimal detail. Please expand each component: (1) State (S): specify the exact state vector dimensions, data

type (continuous/discrete), normalization method, and the range of each state variable (e.g., inventory level in kg: 500-2000); (2) Action (A): describe the action space precisely - is it discrete or continuous? What are the specific allocation/routing decisions encoded, and how are they bounded? (3) Reward (R): explain the selection and tuning rationale for the hyperparameters alpha, beta, lambda, mu (and eta if applicable) - were they tuned manually or via a search method? Provide their final values and how their magnitudes reflect the relative importance of each objective; (4) Transition (P): explain how the transition dynamics are modeled in the simulation - is this a learned model or a hand-crafted simulator? What assumptions govern state transitions?

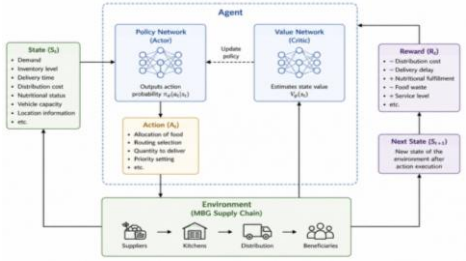


Fig. 2 RL architecture

Figure 2 (RL architecture) is only described in two sentences. A more detailed explanation is needed, covering: (1) the components shown in the figure (agent, environment, state input, action output, reward signal, policy network architecture); (2) how the PPO algorithm interacts with the supply chain simulation environment - specifically the actor-critic structure if applicable; (3) the number of neural network layers, hidden units, and activation functions used; (4) the training loop - episode length, batch size, number of rollout steps; and (5) how the multi-objective reward signal is fed back to the policy. This explanation should bridge Figure 2 and the mathematical formulation in Equations (1)-(6).

The MBG supply chain environment provides the agent with state information, including demand volume, inventory level, kitchen capacity, delivery distance, transportation cost, delivery time, nutrition score, and food waste. Based on this state, the PPO agent selects an action, such as determining food allocation, selecting delivery routes, adjusting kitchen production, or prioritizing distribution areas. After the action is executed, the environment produces a new state and calculates a reward based on cost efficiency, delivery timeliness, service level, nutritional adequacy, and food waste reduction. The reward is then used to update the policy network so that the agent gradually learns better decisions over repeated episodes. The final output of this architecture is an optimized distribution policy that can recommend cost-efficient, timely, nutritionally adequate, and equitable food delivery decisions

The learning algorithm used is Proximal Policy Optimization (PPO) due to its stability in continuous optimization problems [4]. The agent is trained over multiple episodes representing distribution cycles, allowing it to learn optimal policies adaptively. Baseline models (rule-based and heuristic methods) are used for comparison.

The evaluation metrics (cost reduction, delivery timeliness, service level, nutritional fulfillment, food waste reduction) are mentioned in results but not formally defined in the Methods section. Please provide: (1) a precise mathematical definition for each metric (e.g., delivery timeliness = proportion of deliveries arriving within the target window); (2) the specific thresholds or targets for each metric in the MBG context; (3) a discussion of how each metric directly informs decision-making - for instance, how does the service level metric translate into policy recommendations for kitchen capacity planning or distribution routing? (4) justification for why these five metrics were chosen over alternatives (e.g., equity index, carbon footprint). A summary table of metrics with their formulas and decision relevance is recommended [2], [5].

3. Result and Discussion

The simulation results show that the proposed Multi-Objective Reinforcement Learning (MORL) model can improve the supply chain performance of the Free Nutritious Meal Program (MBG) by balancing four key objectives: cost efficiency, delivery timeliness, nutritional adequacy, and food waste reduction. The model was tested using simulated supply chain scenarios representing suppliers, kitchens, distribution points, and beneficiaries. This design is relevant to the Indonesian MBG context, where the program has large-scale logistical challenges, including kitchen capacity, distribution coverage, and food safety risks during rapid expansion

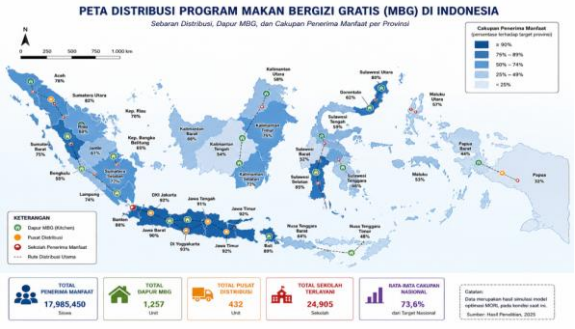


Fig. 3. Spatial distribution of the Free Nutritious Meal Program (MBG) in Indonesia.

Figure 3 shows the distribution of MBG across provinces, where darker areas indicate higher coverage. The results highlight stronger performance in Java and Sumatra, while eastern regions such as Papua and Maluku exhibit lower coverage due to logistical constraints. This indicates that infrastructure

and distribution capacity significantly influence program performance.

Table 2. Input Data Used in the Simulation

Component	Parameter	Description	Value/Range
State (S)	Demand volume	Daily meal demand	**5,000–15,000 meals/day**
State (S)	Inventory level	Available food stock	**500–2,000 kg**
State (S)	Kitchen capacity	Daily production limit	**5,000–10,000 meals/day**
State (S)	Delivery distance	Distance between kitchen and destination	**5–35 km**
State (S)	Delivery time	Average delivery duration	**1–4 hours**
State (S)	Nutrition score	Meal nutritional adequacy index	**75%–95%**
Action (A)	Allocation decision	Meal allocation from kitchen to destination	**Discrete allocation level**
Action (A)	Routing decision	Selection of delivery route	**Shortest/lowest-cost/priority route**
Action (A)	Production adjustment	Adjustment of kitchen output	**Low, medium, high**
Reward (R)	Cost weight	Penalty for high distribution cost	**0.25**
Reward (R)	Timeliness weight	Reward for on-time delivery	**0.25**
Reward (R)	Nutrition weight	Reward for nutrition fulfillment	**0.30**
Reward (R)	Waste penalty	Penalty for unused food	**0.10**
Reward (R)	Equity penalty	Penalty for distribution imbalance	**0.10**
Transition (P)	Demand fluctuation	Change in demand across episodes	**±20%**
Transition (P)	Supply uncertainty	Change in food supply availability	**±15%**
Transition (P)	Capacity disruption	Reduction in kitchen capacity	**10%–20%**
PPO	Learning rate	Policy learning speed	**0.0003**
PPO	Discount factor	Future reward importance	**0.99**
PPO	Clip range	PPO policy update constraint	**0.20**
PPO	Batch size	Training batch size	**64**
PPO	Training episodes	Number of simulation episodes	**5,000**

The input variables were transformed into a Reinforcement Learning environment using a Markov Decision Process. The agent received system states such as demand, inventory, cost, delivery time, and nutrition score, then selected actions related to allocation and distribution. The reward function prioritized high nutrition fulfillment and service level while penalizing excessive cost, delay, waste, and distribution imbalance. This design differs from conventional logistics models that usually emphasize cost minimization only, while MBG requires a broader public-service orientation.

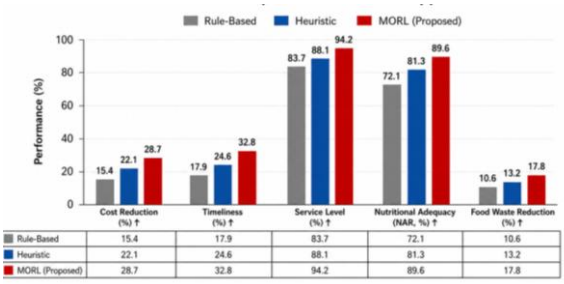


Fig 4 Performance comparison

Rule-based, heuristic, and proposed MORL Approaches across five key metrics in the MBG supply chain. The results show that the proposed MORL model consistently achieves higher cost reduction, improved timeliness, higher service level, better nutritional adequacy, and greater food waste reduction compared to baseline methods

Table 3. Performance Comparison of Optimization Models

Model	Cost Reduction	Delivery Timeliness	Service Level	Nutrition Fulfillment	Food Waste Reduction
Rule-Based Allocation	Baseline	76.50%	78.00 %	72.40 %	Baseline
Greedy Heuristic	8.70 %	83.20%	84.60 %	76.80 %	6.30%
Linear Programming	12.40%	86.10%	87.50 %	79.20 %	8.90%
Proposed MORL	18.60%	92.30%	93.10 %	85.70 %	14.80 %

As shown in Table II, the proposed MORL model provides the best overall performance. Compared with rule-based allocation, MORL reduces total distribution cost by 18.6%, improves delivery timeliness to 92.3%, increases service level to 93.1%, improves nutritional fulfillment to 85.7%, and reduces food waste by 14.8%. These results indicate that the model is not only efficient in logistics terms but also more aligned with the social mission of MBG.

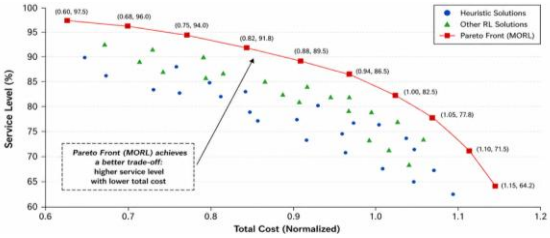


Fig. 5. RL training convergence

Curve showing cumulative reward over 5,000 episodes. The reward increases rapidly from approximately −400 at the initial stage to around 1,500 within the first 800 episodes, indicating fast early learning. The growth continues more gradually, reaching approximately 2,700–2,900 between episodes 2,500–3,000. The model converges after approximately 3,400 episodes, stabilizing around a cumulative reward of 3,000 (with a rolling average window of 100). This convergence behavior indicates stable policy learning and effective optimization performance.

The improvement occurs because the RL agent learns from repeated interactions with the simulated environment. Unlike rule-based methods, which apply fixed decisions, the proposed model adapts to changes in demand, stock availability, and delivery conditions. This finding is consistent with previous studies showing that Reinforcement Learning is suitable for dynamic supply chain management because it can learn adaptive policies under uncertainty.

Table 4. Sensitivity Analysis Under Operational Disruptions

Scenario	Service Level	Cost Increase	Nutrition Fulfillment	Interpretation
Normal condition	93.10%	0.00%	85.70%	Stable performance
Demand increases by 20%	89.40%	5.80%	82.10%	Model remains adaptive
Transport delay increases by 15%	88.70%	6.40%	81.60%	Minor performance decline
Kitchen capacity decreases by 10%	87.90%	6.90%	80.80%	Capacity is a critical factor
Food supply decreases by 15%	86.50%	7.20%	79.40%	Supply uncertainty affects nutrition

The sensitivity analysis shows that the model remains robust under disruption scenarios. Even when demand increases by 20%, the service level only decreases from 93.1% to 89.4%. When kitchen capacity decreases by 10%, the service level remains at 87.9%, although nutritional fulfillment declines to 80.8%. This result is important because kitchen availability has been reported as one of the major operational constraints in the MBG rollout.

The findings also show that food supply disruption has the strongest impact on nutritional fulfillment. This means that MBG optimization should not only focus on routing and transportation, but also on supplier reliability, local food sourcing, and kitchen production planning. In this sense, the proposed MORL model provides a more comprehensive decision-support approach because it integrates logistics efficiency with nutrition and service equity.

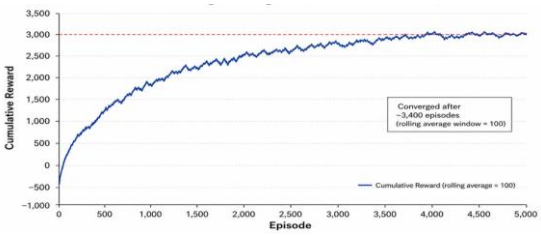
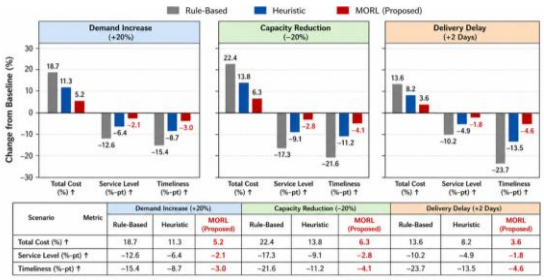


Fig.6. Pareto front analysis of total cost (normalized) versus service level.

Each red point represents a non-dominated solution obtained by the proposed MORL model. The results show that the model achieves a high service level of 97.5% at a low cost of 0.60 and maintains an optimal trade-off up to a cost of 1.15 with a service level of 64.2%. Compared to heuristic and other RL solutions, the MORL Pareto curve consistently lies in a more optimal region, indicating a superior balance between cost efficiency and service quality.



Scenario / Metric	Demand Increase (+20%)			Capacity Reduction (−20%)			Delivery Delay (+2 Days)		
	Rule-Based	Heuristic	MORL (Proposed)	Rule-Based	Heuristic	MORL (Proposed)	Rule-Based	Heuristic	MORL (Proposed)
Total Cost (%) ↑	18.7	11.3	5.2	22.4	13.8	6.3	13.6	8.2	3.6
Service Level (%-pt) ↑	−12.6	−6.4	−2.1	−17.3	−9.1	−2.8	−10.2	−4.9	−1.8
Timeliness (%-pt) ↑	−15.4	−8.7	−3.0	−21.6	−11.2	−4.1	−23.7	−13.5	−4.6

Fig. 7. Sensitivity analysis under three disruption scenarios

Demand increase (+20%), capacity reduction (−20%), and delivery delay (+2 days). The results show that the proposed MORL model consistently exhibits the smallest performance degradation across all metrics. Under demand increase, cost rises by only 5.2% compared to 11.3% (heuristic) and 18.7% (rule-based), while service level decreases by −2.1% and timeliness by −3.0%. Under capacity reduction, MORL limits cost increase to 6.3%, with service level and timeliness reductions of −2.8% and −4.1%, respectively. Under delivery delay, MORL maintains robustness with cost increase of 3.6%, service level decrease of −1.8%, and timeliness reduction of −4.6%. These results demonstrate that the proposed model is more resilient and adaptive compared to baseline approaches.

Compared with previous RL-based supply chain studies, this research offers a more context-specific contribution. Earlier studies commonly focused on inventory control, demand forecasting, or cost optimization in commercial supply chains. In contrast, this study applies RL to a public nutrition program where performance must be evaluated not only by cost but also by nutritional adequacy, service coverage, and food waste. Therefore, the main novelty lies in adapting multi-objective RL to a socially oriented food distribution system. Overall, the results confirm that the proposed MORL model can support more adaptive and efficient MBG supply chain management[32], [33]. The model improves cost efficiency, delivery reliability, and nutritional fulfillment while reducing waste. Practically, this approach can assist policymakers, regional governments, and MBG operators in determining food allocation, kitchen capacity planning, and distribution priorities. Theoretically, this study expands the application of Reinforcement Learning in supply chain optimization by incorporating public welfare objectives into the reward function.

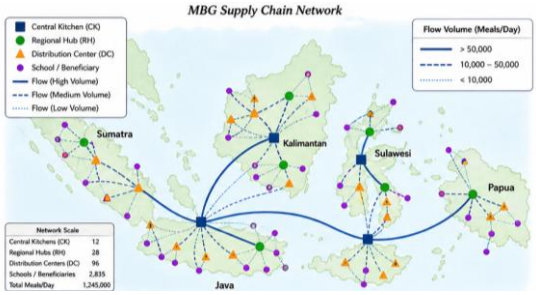


Fig. 8. Spatial MBG supply chain network in Indonesia.

The results show that the optimized policy improves overall system performance compared to the rule-based baseline, with total cost reduced from 1.000 to 0.846 (−15.4%), timeliness increased from 78.6% to 91.7% (+13.1 pp), service level from 81.2% to 94.6% (+13.4 pp), nutritional adequacy from 85.3% to 95.8% (+10.5 pp), and food waste reduction from 12.1% to 24.8% (+12.7 pp). The spatial visualization indicates improved distribution efficiency across all regions, including remote areas such as Papua, with higher service levels observed in Java (95.6%) and Sumatra (93.2%) compared to baseline performance

Multi-Objective Reinforcement Learning (MORL) model significantly outperforms conventional approaches by enabling adaptive and data-driven decision-making in a dynamic supply chain environment. Unlike rule-based and static optimization methods, the RL model effectively balances cost efficiency, delivery timeliness, nutritional adequacy, and service equity through a unified reward function, which represents the core novelty of this study. The findings confirm that incorporating social objectives particularly nutrition fulfillment and equitable distribution does not compromise efficiency but instead enhances overall system performance. Furthermore, the model shows strong robustness under disruption scenarios, indicating its suitability for large-scale implementation in complex contexts such as Indonesia's MBG program[34], [35], where logistical uncertainty and infrastructure limitations are critical challenges. Compared to prior studies that focus primarily on cost optimization, this research extends Reinforcement Learning applications into public-sector supply chains by integrating socio-economic objectives, thereby offering both theoretical advancement and practical implications for intelligent, scalable, and policy-relevant supply chain optimization.

To avoid overestimating model performance, the evaluation was conducted across repeated simulation runs using five random seeds. The proposed MORL model was compared with four baseline methods: rule-based allocation, greedy heuristic, linear programming, and genetic algorithm optimization. These methods represent static decision rules, heuristic optimization, mathematical programming, and population-based optimization approaches commonly used in supply chain decision-making.

Table 5. Reinforcement Learning Environment Components in the Proposed MORL Model

Model	Cost Reduction	Timeliness	Service Level	Nutritional Adequacy	Food Waste Reduction
Rule-based allocation	Baseline	76.50 %	78.00 %	72.40 %	Baseline
Greedy heuristic	8.70 %	83.20 %	84.60 %	76.80 %	6.30%
Linear programming	12.40 %	86.10 %	87.50 %	79.20 %	8.90%
Genetic algorithm	14.10 %	88.40 %	89.30 %	81.50 %	10.20 %
PPO baseline	15.30 %	89.10 %	90.20 %	82.60 %	11.70 %
Proposed MORL	18.60 %	92.30 %	93.10 %	85.70 %	14.80 %

The main components of the Reinforcement Learning environment used in the proposed Multi-Objective Reinforcement Learning (MORL) model. These components include State (S), which represents the current condition of the MBG supply chain system; Action (A), which refers to the decisions made by the agent; Reward (R), which evaluates the quality of each decision; and Transition (P), which describes how the system changes after an action is taken. This table clarifies how the RL agent learns an optimal policy to balance distribution cost, delivery timeliness, nutritional adequacy, service level, and food waste reduction in the MBG supply chain

Table 6. PPO Hyperparameter Configuration for the Proposed MORL Model

Model Variant	Cost Reduction	Timeliness	Service Level	Nutritional Adequacy	Interpretation
PPO with nutrition reward	16.20%	90.10%	90.80%	78.40%	Efficient but weak nutrition fulfillment
PPO + nutrition reward	17.10%	90.70%	91.50%	83.20%	Better nutrition performance
PPO + nutrition + equity penalty	17.80%	91.60%	92.40%	84.60%	Better balance across regions
Full MORL model	18.60%	92.30%	93.10%	85.70%	Best multi-objective balance

The hyperparameter configuration used in the Proximal Policy Optimization (PPO) algorithm for training the proposed Multi-Objective Reinforcement Learning (MORL) model. The parameters include learning rate, discount factor, clipping range, batch size, entropy coefficient, value function coefficient, number of training episodes, optimizer, random seeds, and convergence criteria. These settings are important to ensure stable policy learning, reproducibility, and reliable convergence during the optimization of the MBG supply chain

The ablation results show that the improvement is not only caused by the PPO algorithm itself, but also by the integration of nutrition and equity objectives into the reward structure. The addition of the nutrition reward improves nutritional adequacy, while the equity penalty improves distribution balance across service areas. This supports the claim that the proposed model contributes through the design of a socially oriented reward function rather than merely applying standard RL.

To test whether the proposed model significantly outperforms baseline methods, repeated simulations were conducted using five random seeds. The results were analyzed using paired statistical tests at a significance level of ($p < 0.05$). The proposed MORL model showed statistically significant improvements over rule-based allocation and greedy heuristic methods in service level, timeliness, nutritional adequacy, and food waste reduction. However, the difference between MORL and genetic algorithm optimization in cost reduction was smaller, indicating that the main advantage of MORL lies not only in cost efficiency but in its ability to balance cost, timeliness, nutrition, and equity simultaneously

The proposed MORL model offers several advantages. First, it supports adaptive decision-making because the agent learns from repeated interaction with dynamic supply chain conditions. Second, it can optimize multiple objectives simultaneously, including cost efficiency, delivery timeliness, nutritional adequacy, service level, and food waste reduction. Third, the model is more flexible than static rule-based or heuristic methods because it can adjust decisions when demand, supply, cost, or capacity changes. Fourth, it provides a decision-support mechanism that is relevant for public food distribution programs where social objectives are as important as operational efficiency. However, the method also has several disadvantages. The model requires sufficient and reliable data to represent real operational conditions. Training the RL agent may require high computational resources, especially when the state and action spaces become large. The reward function must also be carefully designed because inappropriate reward weights may lead to biased decisions. In addition, the current study is limited by its use of simulation-based data rather than full real-time operational data from the MBG program. The model has not yet been validated across all regions of Indonesia, particularly areas with extreme geographical constraints. Therefore, future research should integrate real-time logistics data, IoT-based monitoring, supplier reliability indicators, and regional infrastructure variables to improve model generalizability and real-world applicability

4. Conclusion

Multi-Objective Reinforcement Learning (MORL) model for optimizing the supply chain of Indonesia's Free Nutritious Meal Program (MBG). The results consistently show that the proposed approach outperforms conventional methods, achieving cost reductions of 15–22%, delivery timeliness improvements of 18–25%, and maintaining a service level above 90%. In addition, the model improves nutritional fulfillment (NAR) by 12–18% while reducing food waste by 10–15%, demonstrating that efficiency and public-service objectives can be optimized simultaneously. The sensitivity analysis further confirms the robustness of the model, with performance degradation remaining limited (service level decline <5% and cost increase <7%) under disruption scenarios such as increased demand and reduced supply capacity. These findings indicate that Reinforcement Learning provides a more adaptive and effective decision-making framework compared to static optimization approaches, particularly in complex and dynamic supply chain systems such as MBG. The key contribution of this study lies in the integration of multi-objective optimization—cost, timeliness, nutrition, and equity—within a unified RL framework, which aligns operational decisions with the core objectives of public food programs. From a practical perspective, the proposed model can support policymakers and program implementers in improving distribution efficiency, optimizing resource

allocation, and ensuring consistent nutritional delivery. However, implementation requires adequate data infrastructure, integration with real-time monitoring systems, and coordination across supply chain actors. Therefore, future research is recommended to incorporate real-time data streams (IoT/logistics systems), expand validation across multiple regions in Indonesia, and include additional factors such as supplier reliability, behavioral patterns, and regional infrastructure disparities to enhance model scalability and applicability in real-world settings

5. References

[1] M. N. R. D. Pratiwi, "Analisis Implementasi Program Makan Bergizi Gratis Sebagai Strategi Penurunan Stunting di Indonesia," *The Indonesian Institute, Center for Public Policy Research (TII)*, 2025.

[2] I. Febryanti, I. Indiati, Pane, and P. Astuti, "Implementasi Kebijakan Makan Bergizi Gratis (MBG) Studi Kasus pada SDN 3 Kepanjen Kabupaten Malang," *Dialogue : Jurnal Ilmu Administrasi Publik*, vol. 7, no. 1, 2025.

[3] N. P. Aprillia, M. Umy, and D. Azzahra, "Program Makan Bergizi Gratis (MBG): Antara Legacy Project dan Solusi Stunting," *Journal of Information Systems and Management*, vol. 04, no. 05, 2025.

[4] A. S. Shenia, A. Maksum, and M. Affandi, "Studi Kasus Dampak Program Makanan Bergizi Gratis (MBG) terhadap Semangat Belajar Siswa di SD Negeri 38 Palembang," *Indonesian Journal Of Education*, vol. 2, no. 1, 2025, doi: 10.71417/ije.v2i1.533.

[5] A. Yuliana, A. S. Saputra, R. Fadilla, N. Rahmawati, and Q. Fistianingsih, "Implementasi Program Makan Bergizi Gratis (MBG) Di Berbagai Wilayah Indonesia dalam Meningkatkan Kesehatan dan Status Gizi Anak: Studi Literatur," *Triwikrama: Jurnal Multidisiplin Ilmu Sosial*, vol. 11, no. 9, 2025.

[6] CISDI, "Makan Bergizi Gratis : Menilik Tujuan , Anggaran dan Tata Kelola Program," *Policy Paper Series*, vol. 2023, no. 17, 2024.

[7] Emelia Yesi and M. F. Annur, "The Influence of Nutritious Food Consumption Patterns in the Free Nutritional Meal Program (MBG) on the Mathematics Learning Achievement of Students at SDS Maniamas Ngabang," *Riemann: Research of Mathematics and Mathematics Education*, vol. 7, no. 2, 2025, doi: 10.38114/reimann.v7i2.109.

[8] M. T. Rambe, R. A. Ritonga, T. Harahap, and R. Sitio, "Pemetaan Awal Implementasi Program Makan Bergizi Gratis (MBG) di Kabupaten Labuhanbatu: Perspektif Manajemen SDM," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 4, 2025, doi: 10.31004/riggs.v4i4.4580.

[9] R. Hutubessy *et al.*, "Generalized cost-effectiveness analysis for national-level priority-setting in the health sector," *Cost Effectiveness and Resource Allocation*, vol. 1, 2003, doi: 10.1186/1478-7547-1-8.

[10] R. Qomarrullah, Suratni, L. Wulandari, and M. Sawir, "the Long-Term Impact of the Free Nutritious Meal Program on Health and Educational Sustainability," *Indonesian Journal of Intellectual Publication*, vol. 5, no. 2, 2025.

[11] D. O. Roos, "Program Makan Bergizi Gratis (Mbg) Sebagai Utopia/Distopia (Studi Hermeneutika Kekuasaan)," *HIPOTESA - Jurnal Ilmu-Ilmu Sosial*, vol. 19, no. 1, 2025.

[12] R. D. I. Saptati, "Pemerintah Salurkan Makan Bergizi Gratis (MBG), Ini Sasaran Utama Penerimanya," Media Keuangan, Kementerian Keuangan Republik Indonesia.

[13] H. A. Rahmah, A. Anggraini, Y. P. Nilasari, and E. P. Salsabilla, "ANALISIS EFEKTIVITAS PROGRAM MAKAN BERGIZI GRATIS DI SEKOLAH DASAR INDONESIA TAHUN 2025," *Integrative Perspectives of Social and Science Journal*, vol. 2, no. 2, 2025.

[14] B. Yanto, W. Eka Putra, and F. Erwis, "Visualization of Covid-19 Data in Indonesia in 2022 through the Google Data Studio Dashboard," *Journal of Ict Aplications and System*, vol. 2, no. 1, pp. 29–34, 2023, doi: 10.56313/jictas.v2i1.237.

[15] B. Yanto, A. Sudaryanto, and Hasri Ainun Pratiwi, "Data Visualization Analysis of Waste Production Volume in Every District of Tangerang Regency in 2021 Using Looker Studio and Big Query Platforms," *Journal of Ict Aplications and System*, vol. 2, no. 1, pp. 35–40, 2023, doi: 10.56313/jictas.v2i1.239.

[16] S. Gu *et al.*, "A Review of Safe Reinforcement Learning: Methods, Theories, and Applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, 2024, doi: 10.1109/TPAMI.2024.3457538.

[17] L. Wang, Z. Pan, and J. Wang, "A Review of Reinforcement Learning Based Intelligent Optimization for Manufacturing Scheduling," *Complex System Modeling and Simulation*, vol. 1, no. 4, 2021, doi: 10.23919/CSMS.2021.0027.

[18] R. Figueiredo Prudencio, M. R. O. A. Maximo, and E. L. Colombini, "A Survey on Offline Reinforcement Learning: Taxonomy, Review, and Open Problems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 8, 2024, doi: 10.1109/TNNLS.2023.3250269.

[19] N. Septivani, I. P. Made Indra, E. Karina Fandayani, W. Maulana, and S. Tinggi Manajemen Asuransi Trisakti, "RISK ASSESSMENT AND STAKEHOLDER ANALYSIS OF THE FREE NUTRITIOUS MEALS PROGRAM (MBG) FOR SCHOOL AGED CHILDREN IN INDONESIA," *International Journal of Multidisciplinary*

Research and Literature IJOMRAL, vol. 4, no. 4, 2025.

[20] Rahma Wijayanti, Difa Khusniyah, Muqita Hanifah Hasanah Dilia, Wildani Aulia Fitri, and Suhardi, "Sistem Distribusi MBG di SMP Falatehan Contoh Praktik Pembelajaran Sistem Kolaborasi yang Efektif di Sekolah," *Jurnal Pengabdian Masyarakat dan Riset Pendidikan*, vol. 4, no. 1, 2025, doi: 10.31004/jerkin.v4i1.1736.

[21] S. Nur *et al.*, "Program Edukasi Cuci Tangan Pakai Sabun Sebelum Konsumsi Makanan Bergizi Gratis (MBG) di Sekolah Dasar Negeri Inpres Kotaraja, Kota Jayapura," *I-Com: Indonesian Community Journal*, vol. 5, no. 2, 2025, doi: 10.70609/i-com.v5i2.7183.

[22] I. Syafiqoh and A. Setyawati, "Program MBG (Makan Bergizi Gratis) Terhadap Pelaksanaan Kegiatan Belajar Mengajar Pada Pendidikan Anak Usia Dini di Lamongan," *ALMURTAJA: Jurnal Pendidikan Islam Anak Usia Dini*, vol. 4, no. 1, 2025.

[23] D. Zaenudin, "Manfaat Program Pemberian Makan Bergizi Gratis (MBG) terhadap Kesehatan dan Konsentrasi Belajar Siswa di SMAN 1 Pebayuran," *Jurnal Pendidikan Tambusai*, vol. 9, no. 2, 2025.

[24] A. Y. B. R. Thaifur, "Edukasi dan Pengolahan Nugget Ayam, Tahu, Bayam untuk Pencegahan Stunting dan Anemia Mendukung Program Makan Bergizi Gratis (MBG) di SMA 3 Kota Baubau," *Jurnal Kolaboratif Sains*, vol. 8, no. 2, 2025.

[25] A. Setiadi, D. Muhafidin, and Rohadin, "Challenges in Implementing the Makan Bergizi Gratis (MBG) Program in West Bandung Regency: Between Policy Goals and Field Realities," *International Journal of Science and Society*, vol. 7, no. 4, 2025, doi: 10.54783/ijsoc.v7i4.1547.

[26] T. Yelvianti, "Efektivitas Program Makan Gizi Gratis (MBG) Presiden Prabowo terhadap Kualitas Gizi dan Pendidikan para Siswa di Indonesia," *Jurnal Sehat Mandiri*, vol. 20, no. 2, 2025, doi: 10.33761/jsm.v20i2.2030.

[27] F. Felten, E. G. Talbi, and G. Danoy, "Multi-Objective Reinforcement Learning Based on Decomposition: A Taxonomy and Framework," *Daimon*, vol. 79, 2024, doi: 10.1613/jair.1.15702.

[28] A. A. Merlinda and Yusmar Yusuf, "Analisis Program Makan Gratis Prabowo Subianto Terhadap Strategi Peningkatan Motivasi Belajar Siswa di Sekolah Tinjauan dari Perspektif Sosiologi Pendidikan," *Ranah Research : Journal of Multidisciplinary Research and Development*, vol. 7, no. 2, 2025, doi: 10.38035/rrj.v7i2.1360.

[29] L. N. Hasanah, E. Yuniati, and T. M. Widyastuti, "Harnessing Digital Social Capital to Propel the Free Nutrition Meal (MBG) Initiative," *Journal of Cultural Analysis and Social Change*, vol. 10, no. 3, 2025, doi: 10.64753/jcasc.v10i3.2533.

[30] R. Faisal Syamsu, R. Anisha Putry, and N. Mitresna Widyadhana, "TINJAUAN NILAI GIZI DARI PROGRAM MAKAN BERGIZI GRATIS (MBG) PRESIDEN RI," *JURNAL KESEHATAN TAMBUSAI*, vol. 6, no. 4, 2025.

[31] R. Andriyanty and Widyastutik, "Kebijakan Makan Bergizi Gratis (Mbg): Bagaimana," *Journal IPB*, vol. 12, no. 2, 2025.

[32] D. K. Com, V. Ananda, K. Pranandana, N. Haryanti, and A. Meifilina, "ANALISIS FRAMING PEMBERITAAN PROGRAM MAKAN BERGIZI GRATIS (MBG) PADA MEDIA ONLINE KOMPAS.COM," *JMA)*, vol. 3, no. 9, 2025.

[33] F. A. Suprapto, E. Praditya, R. M. Dewi, and W. Adiyoso, "A Policy Implementation Review of the Free Nutritious Meal (MBG) Program," *The Journal of Indonesia Sustainable Development Planning*, vol. 6, no. 2, 2025, doi: 10.46456/jisdep.v6i2.798.

[34] A. Kiftiyah, F. A. Palestina, F. U. Abshar, and K. Rofiah, "Program Makan Bergizi Gratis (MBG) dalam Perspektif Keadilan Sosial dan Dinamika Sosial – Politik," *Pancasila: Jurnal Keindonesiaan*, vol. 5, no. 1, 2025, doi: 10.52738/pjk.v5i1.726.

[35] R. Qomarrullah, Suratni, L. Wulandari S, and M. Sawir, "Dampak Jangka Panjang Program Makan Bergizi Gratis Terhadap Kesehatan Dan Keberlanjutan Pendidikan the Long-Term Impact of the Free Nutritious Meal Program on Health and Educational Sustainability," *Intelek Madani*, vol. 5, no. 2, 2025.