PIONEER
ACADEMIC PUBLISHING LIMITED

Journal of Progress in **Engineering and Physical Science**

Volume 4, Number 5 **October 2025**

CONTENTS

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.pioneerpublisher.com/jpeps
Volume 4 Number 5 October 2025

Algorithmic Reinforcement and the Co-Evolution of User Preferences Through a Mechanistic Analysis of Diversity Loss in Recommender Systems

K. Nowak[1]

[1] Warsaw University of Technology, Poland
Correspondence: K. Nowak, Warsaw University of Technology, Poland.

doi:10.56397/JPEPS.2025.10.01

Abstract

Recommender systems have evolved into adaptive infrastructures that mediate human attention, learning, and decision-making across digital environments. This paper presents a mechanistic analysis of how algorithmic reinforcement processes co-evolve with user preferences, producing a progressive reduction in informational diversity. By conceptualizing recommendation as a coupled dynamical system, the study explains how reinforcement learning architectures internalize behavioral feedback and transform transient user interactions into long-term preference structures. The analysis identifies a recursive mechanism in which both algorithmic policies and user cognition adapt toward equilibrium states that maximize predictability and engagement at the expense of novelty. Empirical findings and theoretical models from recent reinforcement learning research are synthesized to elucidate the dynamics of diversity loss as an emergent property of co-adaptation. The paper proposes a mechanistic framework that integrates stochastic exploration, entropy regularization, and temporal reward shaping to sustain informational variety in reinforcement-driven ecosystems. This approach reconceptualizes recommender systems as co-evolutionary environments where the preservation of diversity is a structural necessity for epistemic resilience, cognitive openness, and sustainable engagement.

Keywords: algorithmic reinforcement, recommender systems, diversity loss, informational entropy, reinforcement learning

1. Introduction

Recommender systems shape much of contemporary digital experience by filtering, ranking, and suggesting information according to individual behavioral traces. These systems translate human activity into quantifiable signals that inform future algorithmic decisions, producing a recursive cycle of interaction between user and model. Algorithmic reinforcement emerges when these cycles continuously adjust recommendations based on observed responses, embedding the history of user engagement into the architecture of the system itself. The mechanism transforms transient preferences into persistent behavioral patterns, aligning exposure to content that maximizes predicted satisfaction or attention. Over repeated iterations, the recommender's optimization objectives begin to steer the evolution of user tastes, creating a co-dependent dynamic in which algorithms both reflect and construct preference landscapes.

As this interaction intensifies, a structural narrowing of informational diversity occurs. The user's choice space becomes progressively confined to familiar categories because the system interprets consistency as preference certainty. Engagement-driven optimization accentuates this contraction, since content that diverges from prior patterns yields lower immediate reward signals. The loss of diversity manifests not only in the reduction of novel or unexpected items but also in the attenuation of cognitive variety, where users encounter fewer distinct perspectives or genres. Such homogenization represents a form of algorithmic path dependence, where early behavioral signals disproportionately influence long-term exposure trajectories.

A mechanistic understanding of this process requires examining the interplay between learning algorithms, feedback design, and user cognition. Reinforcement learning agents operate through the maximization of cumulative reward functions, while users respond through evolving attention and interest parameters. Each click, skip, or linger constitutes a data point in a continuous adaptive system. Over time, the optimization of engagement metrics induces emergent properties that are not explicitly programmed, such as polarization or preference ossification. The recommender thus becomes an active participant in shaping the informational environment, operating within a feedback architecture that couples computational efficiency with human behavioral plasticity.

2. Algorithmic Reinforcement Dynamics

At the algorithmic level, reinforcement learning has become central to the design of modern recommender systems because it provides a structured paradigm for modeling sequential decision-making under uncertainty. In this setting, each recommendation constitutes an *action*, and each user response—whether a click, dwell, or skip—serves as a *reward* that updates the system's policy. The recursive adaptation of the algorithm produces a continuous feedback loop in which the model both learns from and reshapes user behavior. This process has been extensively documented in reinforcement learning–based recommender surveys such as Afsar et al. (2022), which describe how engagement-maximizing policies tend to overemphasize short-term gains while diminishing exposure diversity.

The core of algorithmic reinforcement lies in the definition of its reward structure. Conventional designs prioritize immediate engagement signals, often ignoring the long-term diversity of user experience. Such optimization biases can induce *mode collapse*, a state in which the system repeatedly selects from a narrow set of high-reward items. In studies by Lin et al. (2023), actor-critic architectures used in deep reinforcement learning recommender systems exhibit strong tendencies toward reward overfitting, where the model reinforces repetitive exposure to previously popular content. These findings illustrate how algorithmic reinforcement transforms localized behavioral regularities into systemic recommendation biases.

A deeper mechanistic understanding requires examining how different reinforcement learning formulations affect diversity outcomes. Policy-gradient algorithms adjust parameters by estimating gradients of expected returns, while value-based methods such as Deep Q-Networks assign numerical value to discrete recommendation options. Although both methods optimize long-term user satisfaction in theory, in practice they gravitate toward strategies that replicate familiar interaction patterns. Zou et al. (2019) show that optimizing for cumulative engagement without temporal discounting leads to reinforcement of repetitive behaviors rather than exploration of new content spaces. This effect demonstrates that algorithmic reinforcement, when guided by narrowly defined reward signals, tends to amplify preexisting behavioral biases rather than diversify them.

Architectural features within deep models further intensify this convergence. Embedding-based recommendation systems learn latent representations that capture correlations among users, items, and contextual factors. When reinforcement updates operate within these compressed manifolds, the model gradually collapses onto regions of the latent space that correspond to high-probability, high-reward interactions. Studies such as Yu et al. (2024) propose modified reward structures that integrate diversity and novelty metrics directly into the reinforcement objective, demonstrating measurable improvements in content variety without substantial losses in predictive accuracy.

Algorithmic reinforcement dynamics therefore encompass not only a computational optimization process but a co-evolutionary mechanism linking machine learning parameters to human behavioral tendencies. Each iteration of feedback stabilizes mutual expectations between system and user, reducing uncertainty while constraining variability. The emergent equilibrium reflects a convergence of algorithmic learning and human preference formation. Diversity loss thus becomes a structural byproduct of adaptive optimization—a form of informational entropy reduction within the human–algorithm ecosystem, as also noted by Zhao et al. (2025) in their comprehensive analysis of fairness and diversity trade-offs in recommendation algorithms.

In practical implementations, deep reinforcement learning architectures such as actor–critic models and policy-gradient algorithms demonstrate strong sensitivity to the structure of historical data. These architectures optimize policies by evaluating the expected return of each action sequence, yet their dependence on previously observed reward distributions produces a narrowing of exploration space. The agent learns to exploit the most predictable regions of user behavior because those regions offer the highest cumulative reward signal. When applied to recommender systems, this process intensifies the repetition of familiar content and suppresses the discovery of less represented items. Lin et al. (2023) identify this as a manifestation of overfitting within the reinforcement framework, where the model's learned policy becomes confined to historical engagement patterns. Such confinement reduces informational entropy in the recommendation process, creating deterministic cycles of exposure that reinforce prior user trajectories.

The mechanism of overfitting in reinforcement recommenders originates in the structure of the reward gradient. Actor–critic models update both policy and value networks based on observed feedback. When user interactions are concentrated in narrow behavioral clusters, the agent's reward landscape becomes steep around these dominant actions. This leads to a disproportionate update magnitude for frequently rewarded items and minimal parameter adjustment for novel or uncertain content. The result is a rapid convergence toward exploitation. Empirical evidence suggests that such convergence diminishes system robustness by amplifying homogeneity in user experiences and weakening long-term engagement stability.

To mitigate this, recent research explores *multi-objective reinforcement learning* frameworks that balance accuracy with novelty and diversity. Stamenkovic et al. (2022) propose auxiliary loss functions that integrate diversity-aware reward terms into the training process. These formulations optimize for both relevance and entropy, allowing the system to maintain recommendation precision while sustaining exposure to new information. By incorporating diversity constraints into the reinforcement objective, the model learns to navigate a broader state–action space, countering the collapse toward high-reward attractors. This adjustment transforms the algorithm's learning trajectory from a purely exploitative regime toward a balanced adaptive system that values informational variety as part of its optimization landscape.

3. Mechanistic Co-Evolution of User Preferences

User preferences evolve through an intricate process of continuous interaction with algorithmic environments. Each recommendation presented to a user represents not only a predictive output of the system but also a behavioral input that shapes subsequent algorithmic decisions. This recursive interaction establishes a coupled feedback mechanism in which human cognition and computational inference co-regulate one another. The user's attention, emotion, and memory interact with the algorithm's reward and policy structures, producing a cycle of mutual adaptation. Empirical research by Möller et al. (2020) shows that personalized recommendation processes in news and media contexts systematically narrow users' informational exposure. The mechanism is not the direct result of deliberate design but arises from the inherent feedback structure of the system itself, where repetitive exposure subtly shifts user expectations of relevance.

The co-evolution of user preferences can be understood as an emergent property of coupled dynamical systems. In such systems, the algorithm's policy update function and the user's preference formation process operate as interacting equations that continuously influence each other's parameters.

Table 1. A Comparative Summary of Algorithmic and Human Adaptation Mechanisms in the Co-Evolution Process

Mechanistic Dimension	**Algorithmic Adaptation (System)**	**User Adaptation (Human)**	**Emergent Effect**
Learning Rule	Gradient-based policy update	Cognitive reinforcement through exposure	Feedback amplification
Objective Function	Maximization of engagement reward	Minimization of cognitive effort	Mutual predictability
Memory Dynamics	Weighted by recent interactions	Shaped by repetition and familiarity	Preference consolidation
Exploration Tendency	Decreases over optimization cycles	Decreases as novelty perception declines	Reduced informational entropy
Adaptation Speed	High (batch or real-time updates)	Moderate (habit formation)	Temporal coupling and equilibrium

Each round of interaction—comprising recommendation, evaluation, and behavioral response—modifies the internal state of both components. The recommender updates its policy based on perceived engagement value, while the user's perceptual schema adapts to the regularities of exposure. Over successive iterations, the joint system begins to self-organize into stable attractor states characterized by high predictability and low diversity. These attractors represent equilibrium configurations where the mutual reinforcement between algorithm and user produces minimal uncertainty.

Cognitively, this process manifests through phenomena such as confirmation bias and habituation. When an algorithm repeatedly presents content consistent with established preferences, the user's attentional filters become attuned to those patterns. Novel or dissonant information receives reduced cognitive processing, diminishing the likelihood of behavioral signals that could trigger algorithmic exploration. The recommender interprets this decline in engagement as evidence of irrelevance, which leads to further suppression of novel content. Over time, this creates a form of perceptual canalization, where the boundaries of curiosity and interest contract around the dominant themes reinforced by the system. Research on adaptive personalization suggests that such preference consolidation results from reinforcement contingencies at both the neural and computational levels, as attention and prediction error jointly stabilize around familiar stimuli.

At the system level, co-evolution produces measurable consequences for informational diversity and social epistemology. When aggregated across millions of users, individual preference reinforcement can amplify macro-scale polarization. Distinct clusters of users evolve along separate attractor pathways, each governed by different feedback parameters within the same underlying algorithmic architecture. The emergent segmentation of audiences reflects the self-organizing properties of the coupled human–machine ecosystem. This pattern has been observed in large-scale simulations of user–algorithm interactions, where even neutral recommendation policies lead to polarized equilibria when user adaptation dynamics are included in the model. Studies such as Zou et al. (2019) provide evidence that feedback reinforcement on engagement-driven platforms accentuates divergence between groups while reducing diversity within them.

The mechanistic view of co-evolution reveals that diversity loss is not a secondary artifact but an intrinsic outcome of adaptive feedback. The user's evolving cognitive model and the recommender's learning algorithm share the same optimization direction toward predictability and efficiency. Both agents, human and computational, minimize uncertainty by aligning their internal states. The mutual drive for coherence and relevance transforms the open space of potential experiences into a constrained domain of familiar patterns. Sustaining diversity within this framework requires interventions that intentionally introduce stochasticity or novelty into the recommendation process. The challenge lies in designing adaptive systems that respect user engagement while maintaining informational entropy at levels that support exploration and cognitive flexibility.

4. Diversity Loss and Systemic Implications

Diversity loss in recommender systems emerges as a structural byproduct of optimization processes that favor stability, efficiency, and engagement. When the algorithm repeatedly selects high-reward actions associated with known user preferences, the variance of recommendations gradually decreases. This contraction of the recommendation space leads to a reduction in informational entropy, a measurable decline in the variety and novelty of suggested items. The mechanism mirrors biological systems that overexploit successful survival strategies at the expense of genetic variation. In recommender environments, algorithmic exploitation accelerates convergence toward predictable outcomes. The outcome is a narrowing of exposure that not only limits user discovery but also reshapes collective cultural and informational ecosystems.

The loss of diversity results from the dominance of short-term reward signals embedded within algorithmic objectives. Engagement metrics such as click-through rate or watch duration are often treated as proxies for satisfaction, yet they capture only the surface dynamics of user interaction. Zhao et al. (2025) show that systems optimized solely for engagement tend to produce homogeneous recommendation distributions even when trained on diverse datasets. This occurs because the gradient-based optimization process amplifies frequent feedback patterns, suppressing low-frequency but potentially meaningful signals. Once these patterns dominate the policy space, the algorithm enters a cycle of self-confirmation in which diversity loss becomes an emergent equilibrium rather than a correctable anomaly.

Efforts to reintroduce diversity often rely on post-processing strategies or fairness constraints. Antikacioglu and Ravi (2017) demonstrate that network flow optimization can redistribute exposure across items and categories with minimal sacrifice to predictive performance. Such interventions, while effective in controlled evaluations, operate downstream of the reinforcement loop. They address the visible outcomes of diversity loss without altering its underlying causal mechanisms. Since the root of the phenomenon lies in the co-adaptive relationship between user feedback and algorithmic reward, external rebalancing techniques cannot produce long-term systemic change. The optimization pressure that drives convergence remains active, gradually eroding post-hoc diversity adjustments.

Sustainable mitigation of diversity loss requires integrating diversity directly into the algorithmic reward structure.

Table 2. Taxonomy of Diversity Intervention Strategies in Recommender Systems

Approach Type	**Mechanism of Action**	**Representative Studies**	**Expected Outcome**
Post-processing (Re-ranking)	Adjusts item distribution after recommendation	Antikacioglu & Ravi, 2017	Short-term diversity increase, minimal accuracy loss
Reward Engineering	Embeds diversity into reinforcement objective	Yu et al., 2024	Balanced novelty–relevance optimization
Regularization Techniques	Adds entropy or variance penalties to model loss	Zhao et al., 2025	Prevents over-concentration in latent space
Stochastic	Introduces probabilistic	Zou et al., 2019	Sustains long-term

Exploration	decision noise		user curiosity
Multi-agent Interaction	Models diversity as emergent from multi-user dynamics	Simulated studies in social recommender contexts	Systemic resilience and exposure variety

Yu et al. (2024) propose multi-objective reinforcement learning frameworks that simultaneously optimize for relevance, novelty, and fairness. These models redefine success by assigning explicit value to informational variety. By embedding entropy-related terms into the reward function, the agent learns to associate exploration with positive long-term return. Such approaches extend the conceptual scope of recommendation from accuracy maximization to ecological balance within the informational environment. The resulting policies exhibit greater resilience to preference homogeneity and more stable long-term engagement patterns.

At a systemic scale, diversity loss alters the informational topology of digital ecosystems. The contraction of exposure pathways reduces cross-domain connectivity, weakening the circulation of ideas and diminishing opportunities for serendipitous discovery. When aggregated across populations, these effects manifest as macro-level informational silos. The system's optimization dynamics thereby influence social cognition and collective knowledge formation. Empirical analyses of content recommendation in news and entertainment contexts reveal that diminished diversity correlates with higher polarization and reduced epistemic breadth. The algorithmic pursuit of engagement thus produces a feedback loop that reconfigures cultural attention landscapes into fragmented clusters.

The implications extend beyond user satisfaction or content fairness. Diversity loss reshapes the cognitive ecology of digital societies by altering how individuals encounter novelty and form judgments about relevance. A mechanistic understanding of this process highlights the inseparability of algorithmic design and human behavior in shaping informational diversity. To counteract systemic homogenization, recommender systems must evolve toward self-regulating architectures that sustain variability as an intrinsic property of learning rather than an externally imposed constraint. Embedding stochastic exploration, contextual randomization, and dynamic novelty thresholds into reinforcement learning policies can maintain informational diversity while preserving personalization quality. These adjustments mark a transition from reactive correction to adaptive equilibrium, positioning diversity as a foundational element of algorithmic intelligence rather than a peripheral metric of fairness.

5. Toward a Mechanistic Framework of Co-Evolution

A mechanistic framework of co-evolution between users and recommender systems begins with the recognition that both entities operate as adaptive agents interacting through feedback. The algorithm functions as a learning agent whose policy evolves through exposure to user behavior, while the user adapts cognitively and affectively in response to the system's recommendations. Each interaction modifies both agents' internal states, creating a continuous exchange of information and influence. When modeled mathematically, this relationship can be expressed as a set of coupled dynamical equations, where the gradient of algorithmic policy updates is conditioned on user response distributions, and the evolution of user preferences depends on the temporal structure of exposure. These coupled equations describe a nonlinear system capable of exhibiting emergent phenomena such as stability, oscillation, and collapse, depending on the balance between exploration and exploitation.

The process can be conceptualized as an autocatalytic loop. Algorithmic outputs act as catalysts that accelerate specific forms of user learning, and user actions feed back into the system as new data that reinforce algorithmic adaptation. Over successive cycles, both agents co-adjust toward equilibrium configurations in which mutual predictability is maximized. Within this equilibrium, diversity loss corresponds to an entropic reduction, where the system's state space contracts into narrow basins of attraction. Zhao et al. (2025) describe this contraction as the systemic manifestation of fairness-diversity trade-offs, arising not from explicit bias but from the inherent optimization structure of feedback-driven models. The coupled system thus transitions from an exploratory regime, rich in potential states, to an exploitative equilibrium characterized by low variability and high stability.

From a theoretical perspective, this dynamic mirrors adaptive processes in biological and ecological systems. In evolutionary biology, diversity is maintained through mutation and recombination that introduce stochasticity into reproductive processes. A similar mechanism is required in recommender ecosystems to avoid informational monocultures. Controlled stochastic exploration introduces variability into algorithmic decision-making without compromising predictive efficiency. Research in multi-objective reinforcement learning has demonstrated that integrating randomness as a structural parameter improves long-term adaptability by preventing convergence to narrow attractor basins. Adaptive regularization schemes, inspired by mutation in population dynamics, can function as a diversity-preserving mechanism by continuously perturbing the policy landscape. Such perturbations enable the system to escape local optima and sustain the flow of novel experiences for users.

The mechanistic framework also requires modeling the temporal interdependence between preference formation and recommendation generation. Zou et al. (2019) show that long-term user engagement can be optimized through reinforcement models that discount immediate rewards in favor of delayed satisfaction. Introducing temporal discount factors alters the geometry of the co-evolutionary system, elongating the feedback horizon and promoting sustained diversity in exposure. The inclusion of time-sensitive reward shaping transforms the coupled system from a reactive to an anticipatory mode, where the algorithm seeks trajectories that maintain user curiosity across extended interactions.

A mechanistic framework of co-evolution therefore integrates principles of dynamical systems theory, evolutionary computation, and behavioral modeling into a unified analytic structure. It conceptualizes the recommender ecosystem as a self-organizing field of interactions rather than a one-directional flow from data to prediction. Diversity loss becomes interpretable as a phase transition, in which the feedback architecture of user–algorithm interaction passes from a state of high entropy and flexibility to one of low entropy and rigidity. To sustain equilibrium between adaptability and stability, recommender systems must be designed to maintain dynamic diversity as an endogenous property. Such systems would not merely mitigate bias or increase novelty but would continually regenerate informational variation through self-regulated exploration. This orientation represents a shift from static optimization toward co-evolutionary intelligence, where algorithmic and human learning processes evolve in tandem to preserve both engagement and epistemic openness.

6. Conclusion

Algorithmic reinforcement and the co-evolution of user preferences form an adaptive system characterized by interdependence, feedback sensitivity, and emergent order. The recommender algorithm learns from behavioral traces while simultaneously constructing the conditions under which those behaviors occur. This duality transforms recommendation from a predictive problem into an ecological process, where patterns of attention, curiosity, and exposure evolve through continuous mutual influence. Diversity collapse within such a system represents a structural manifestation of its learning dynamics rather than an incidental outcome. When optimization objectives prioritize short-term engagement, the system’s adaptive capacity narrows. The loss of informational diversity signifies a reduction in the system’s entropy, a contraction of the possible trajectories through which users and algorithms can co-adapt.

A mechanistic analysis of this phenomenon reveals the intrinsic coupling between algorithmic design and cognitive evolution. Reinforcement learning models define reward functions that encode implicit value systems, shaping how users encounter information and interpret relevance. Behavioral feedback loops translate individual actions into population-level signals that guide model updates. Over repeated iterations, this alignment between machine inference and human response leads to a systemic equilibrium that privileges predictability. Diversity loss thus emerges as a property of stability within the coupled human–algorithm environment. It reflects the success of optimization in achieving coherence while failing to preserve the variability required for exploration and discovery.

Integrating diversity into the core architecture of recommender systems requires a redefinition of optimization itself. Diversity cannot remain an auxiliary metric applied after relevance is maximized. It must be treated as an essential dimension of value alongside accuracy, satisfaction, and fairness. Multi-objective reinforcement learning frameworks demonstrate that diversity can coexist with engagement when treated as part of the reward structure. Systems designed with adaptive stochasticity, entropy regularization, and temporally extended objectives can sustain variation without sacrificing efficiency. Such architectures transform recommendation from a static feedback engine into a dynamic learning ecosystem capable of self-regulation and renewal.

The ethical implications of this transformation extend beyond algorithmic performance. Diversity preservation concerns the integrity of collective cognition and the resilience of digital culture. Recommender systems now act as infrastructure for knowledge formation, shaping what societies learn and how they perceive the world. A loss of diversity within these infrastructures narrows the horizon of public imagination and reduces exposure to alternative perspectives. Mechanistic awareness of co-evolution offers a pathway toward corrective design, emphasizing equilibrium over dominance and plurality over optimization singularity.

Recommender systems that internalize diversity as a structural principle will not simply deliver content more equitably; they will cultivate environments in which users remain open to uncertainty and novelty. Such systems embody an epistemic ethic grounded in balance, where engagement does not preclude exploration and personalization does not erase variation. The co-evolution of human and algorithmic intelligence then becomes a process of shared adaptation that sustains cognitive richness, cultural multiplicity, and long-term informational resilience.

References

Afsar, M. M., Crump, T., & Far, B. (2022). *Reinforcement learning based recommender systems: A survey.* ACM Computing Surveys.

Chen, M., Wang, Y., Xu, C., Le, Y., & Sharma, M. (2021). Values of user exploration in recommender systems. In *Proceedings of the 29th ACM International Conference on Multimedia* (ACM MM).

Hazrati, N. (2023). *Long-term impact of recommender systems on the evolution of users' choice behaviour.* Free University of Bozen-Bolzano Research Repository.

Hazrati, N., & Ricci, F. (2022). Recommender systems effect on the evolution of users' choices distribution. *Information Processing & Management, 59*(6), 103037.

Kunaver, M., & Požrl, T. (2017). Diversity in recommender systems — A survey. *Knowledge-Based Systems, 123*, 154–162.

Lin, Y., Liu, Y., Lin, F., Zou, L., Wu, P., & Zeng, W. (2023). A survey on reinforcement learning for recommender systems. *IEEE Transactions on Neural Networks and Learning Systems*.

Ma, X., Li, M., & Liu, X. (2024). Advancements in recommender systems: A comprehensive analysis based on data, algorithms, and evaluation. arXiv preprint arXiv:2407.18937.

Stamenkovic, D., Karatzoglou, A., & Arapakis, I. (2022). Choosing the best of both worlds: Diverse and novel recommendations through multi-objective reinforcement learning. In *Proceedings of the 16th ACM Conference on Recommender Systems* (RecSys).

Yu, J., Lyu, S., & Chen, P. (2024). Deep Reinforcement Learning for Boosting Individual and Aggregate Diversity in Product Recommendation Systems. In *Proceedings of the 18th ACM Conference on Recommender Systems (RecSys 2024).* Association for Computing Machinery.

Zhao, Y., Wang, Y., Liu, Y., & Cheng, X. (2025). Fairness and diversity in recommender systems: A survey. *Proceedings of the ACM on Recommender Systems*.

Zou, L., Xia, L., Ding, Z., Song, J., Liu, W., & Yin, D. (2019). Reinforcement learning to optimize long-term user engagement in recommender systems. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD).*

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.pioneerpublisher.com/jpeps
Volume 4 Number 5 October 2025

The Application of IoT-Empowered Intelligent Control System for High-Speed Dispersing Equipment in Textile Printing and Dyeing Additives Production

Dongmei Shi[1]

[1] Wuxi Lianda Chemical Co., Ltd., Jiangsu 214201, China
Correspondence: Dongmei Shi, Wuxi Lianda Chemical Co., Ltd., Jiangsu 214201, China.

doi:10.56397/JPEPS.2025.10.02

Abstract

This paper focuses on the textile printing and dyeing additives production sector, targeting the high-speed dispersing equipment, a crucial production device. A smart control system based on Internet of Things (IoT) technology has been developed and applied. By installing high-precision sensors on the equipment to real-time monitor 12 key parameters such as rotational speed and temperature, and leveraging advanced AI algorithms to dynamically optimize the stirring process, the system has been proven effective. It can reduce labor costs and reliance on skilled workers. Based on these achievements, a universal intelligent equipment upgrade framework for the textile printing and dyeing additives industry is proposed, aiming to provide references for other enterprises' intelligent transformation within the industry. This research not only offers strong support for Wuxi Lianda Chemical's production efficiency improvement and cost control, but also paves a new way for the sustainable development and intelligent upgrade of the whole textile printing and dyeing additives industry.

Keywords: Internet of Things (IoT), textile printing and dyeing additives, high-speed dispersing equipment, intelligent control system, production efficiency, energy consumption reduction, product quality, intelligent upgrade, universal upgrade framework, AI algorithm optimization, real-time data acquisition, dynamic parameter adjustment, green and sustainable development

1. Introduction

1.1 Development Status and Challenges of the Textile Printing and Dyeing Additives Industry

As an important branch of the chemical industry, the textile printing and dyeing additives industry has seen a continuous growth in market demand along with the ongoing development of the global textile industry. In recent years, the stricter environmental policies and consumers' increasing demand for high-quality textiles have driven the expansion of the industry scale, with a particularly prominent demand for efficient, environmentally-friendly and multifunctional textile printing and dyeing additives. However, traditional production processes mainly rely on manual operation and experience, which have many limitations. They include low production efficiency that cannot meet the needs of large-scale production, high energy consumption that increases production costs, unstable product quality that cannot guarantee consistency, and the rising labor costs that add to the production burden of enterprises.

1.2 Development and Application Trends of IoT Technology in the Industrial Field

Meanwhile, the rapid development of IoT technology has brought new opportunities to traditional manufacturing. IoT, through various information-sensing devices such as sensors, Radio Frequency Identification (RFID) technology, Global Positioning System (GPS), etc., can real-time collect data of any objects or processes that need to be monitored, connected and interacted with. It then sends the data to the data processing center via a data transmission network for intelligent processing. Its basic architecture consists of the perception layer, transmission layer and application layer, which are responsible for data collection, transmission and processing, respectively. The application of IoT technology in intelligent manufacturing has achieved remarkable results. For example, by installing sensors on production equipment to real-time collect equipment operation data, combined with big data analysis and artificial intelligence algorithms, predictive maintenance of equipment can be realized, improving the equipment operation efficiency and reliability (Xiong, X., Zhang, X., Jiang, W., Liu, T., Liu, Y., & Liu, L., 2024). In addition, IoT technology is also widely used in production process optimization, quality control, supply chain management and other fields, significantly enhancing the intelligent level of manufacturing.

1.3 Research Purpose

This research aims to empower traditional textile printing and dyeing additives production with IoT technology, and develop a smart control system for high-speed dispersing equipment based on IoT. By real-time collecting key parameters and combining AI algorithms to optimize the stirring process, significant improvements in production efficiency, energy consumption reduction and product quality stability are expected. Based on the actual operation data of Wuxi Lianda Chemical, the role of this system in reducing labor costs and stabilizing product consistency will be analyzed. Meanwhile, a universal intelligent equipment upgrade framework for the textile printing and dyeing additives industry will be proposed to provide references for the industry's intelligent upgrade.

2. Technical Principle and Current Status of "High-Speed Dispersing Equipment for Softener Production"

2.1 Working Principle of High-Speed Dispersing Equipment

High-speed dispersing equipment is one of the indispensable devices in the production of textile printing and dyeing additives, especially playing a key role in the softener production process. Its core function is to fully mix raw materials through the high-speed rotation of stirring blades, ensuring uniform distribution of components, thereby improving product quality and production efficiency. The equipment is usually driven by a motor, with the rotational speed controlled by a frequency converter to meet the needs of different production stages. The design and layout of the stirring blades are crucial to the mixing effect. Reasonable blade shape and angle can generate effective turbulence to accelerate the dispersion process of raw materials. Meanwhile, the temperature control inside the equipment is also an important factor affecting product quality. Through the built – in heating or cooling system, the reaction process is ensured to proceed at an appropriate temperature, thus guaranteeing the stability and consistency of the product.

2.2 Current Operation Status and Problems of Traditional High-Speed Dispersing Equipment

Despite the wide application of high-speed dispersing equipment in the production of textile printing and dyeing additives, there are still many problems in the operation of traditional equipment. Firstly, the parameter settings of traditional equipment mostly rely on manual experience and fixed patterns, lacking real-time monitoring and dynamic adjustment capabilities. This leads to the inability to flexibly adjust stirring speed and temperature according to raw material characteristics and production needs during the production process, thereby affecting production efficiency and product quality. Secondly, traditional equipment has high energy consumption. Due to the lack of precise control means, it often needs to run for a long time to achieve the ideal mixing effect, which not only increases production costs but also causes unnecessary environmental burden. In addition, the stability of product quality is difficult to guarantee. Due to the instability of manual operation and equipment performance fluctuations, the consistency of the product is poor, and batch-to-batch differences are prone to occur. Finally, the maintenance cost of traditional equipment is high. Due to the lack of real-time monitoring and fault early warning functions, equipment failures are often difficult to detect in advance, leading to increased repair costs and extended production downtime.

2.3 Application Potential of IoT Technology in High-Speed Dispersing Equipment

The rapid development of IoT technology provides new ideas and methods to solve the problems existing in traditional high-speed dispersing equipment. By installing various sensors such as rotational speed sensors, temperature sensors, pressure sensors, etc. on the equipment, key parameters during the equipment operation process can be real-time collected. These data are transmitted to the central control system via a wireless network, combined with advanced data analysis and artificial intelligence algorithms to realize dynamic optimization of the stirring process. For example, based on the real-time collected rotational speed and temperature data, the system can automatically adjust the stirring speed and heating power to ensure that the production process is always in the best state. In addition, IoT technology can also realize remote monitoring and fault early warning of equipment. By real-time analyzing equipment operation data, potential failures can be detected in advance to reduce downtime and maintenance costs. Through the application of IoT technology, not only can production efficiency be significantly improved and energy consumption reduced, but also the stability and consistency of product quality can be enhanced, providing strong support for the intelligent upgrade of the textile printing and dyeing additives industry.

3. Development of Intelligent Control System for High-Speed Dispersing Equipment Based on IoT

3.1 System Architecture Design

To achieve intelligent control of high-speed dispersing equipment, a complete IoT system architecture has been designed. The architecture consists of the perception layer, transmission layer, data processing layer and application layer. In the perception layer, a variety of high-precision sensors, including rotational speed sensors, temperature sensors, pressure sensors, etc., have been installed to real-time collect key parameters during equipment operation.

3.2 Research on Intelligent Control Algorithm

The intelligent control algorithm is the core part of the entire system. A deep learning algorithm has been adopted to optimize the stirring process. Specifically, the Long Short-Term Memory (LSTM) model has been chosen, as it can effectively deal with time-series data and capture the dynamic relationships between parameters. Production data from Wuxi Lianda Chemical over the past year, including rotational speed, temperature, pressure, energy consumption and product quality indicators, etc., have been used for model training. The training dataset contains over 10,000 samples, each with 12 feature parameters (Liu, Z., 2022). After multiple rounds of training and validation, the LSTM model constructed has achieved an accuracy rate of 92.5% and a recall rate of 90.3% on the test set.

Table 1.

Project	**Description**
Core part	Intelligent control algorithm
Algorithm type	Deep learning algorithm
Specific model	Long Short-Term Memory (LSTM)
Training result	Test set accuracy rate: 92.5% Test set recall rate: 90.3%

3.3 System Integration and Testing

After the completion of system architecture design and algorithm development, system integration and testing have been carried out. First, hardware integration of sensors, wireless communication modules, edge-computing devices and cloud-based servers has been conducted to ensure the compatibility and communication stability between various components. Then, field tests have been performed on Wuxi Lianda Chemical’s thousand-ton-class production line. The test results show that with the application of the intelligent control system, the production efficiency of softeners has been increased by 40%, and the unit energy consumption has been reduced by 20%. During the testing process, the stability and reliability of the system have also been assessed. The system has been continuously operated for 30 days without any failures, and the stability of data collection and transmission has reached 99.5%. This indicates that the intelligent control system not only can significantly enhance production efficiency and product quality, but also possesses high stability and reliability, meeting the demands of industrial production.

4. Application Practice of Intelligent Control System in Wuxi Lianda Chemical

4.1 Overview of Wuxi Lianda Chemical's Thousand-Ton-Class Production Line

The thousand-ton-class production line of Wuxi Lianda Chemical Co., Ltd. is one of the company's important production facilities, mainly producing textile printing and dyeing additives, including softeners, detergents and refining agents. Since its commissioning in 2015, the production line has been using traditional production processes and equipment. Before the implementation of the intelligent control system, the production efficiency of the production line was relatively low. In terms of energy consumption, the average energy consumption per batch was 150kWh (Huang, J., & Qiu, Y., 2025), and labor costs accounted for 30% of the total production costs. The consistency of product quality was poor, with a product pass rate of 90% (Liu, Z., 2025). These current situations indicate that there is a significant room for improvement in efficiency, energy consumption and quality of the production line.

Table 2.

Project	Description
Traditional production mode	Production efficiency: Low Energy consumption: 150kWh per batch Labor cost proportion: 30% of total production costs Product quality consistency: Poor Product pass rate: 90%

4.2 Implementation Process of Intelligent Control System

In order to improve production efficiency, reduce energy consumption and enhance product quality, Wuxi Lianda Chemical decided to introduce an intelligent control system based on IoT. The implementation process includes system installation and debugging, personnel training and technical support, as well as production line transformation and optimization. High-precision sensors, including rotational speed sensors, temperature sensors and pressure sensors, have been installed on the high-speed dispersing equipment to real-time collect key parameters during the equipment operation process. At the same time, industrial-grade wireless communication modules have been installed to ensure stable data transmission to the data processing layer. Edge-computing devices have been deployed beside the production line for data preprocessing and preliminary analysis. The preprocessed data is transmitted to the cloud-based server for further analysis and storage. The intelligent control system has been developed in the cloud, integrating advanced AI algorithms and user-friendly interface design. Production line operators have been trained on the system, including the use of sensors, operation of the data collection system, and operation of the intelligent control system interface. A technical support team has been provided to be responsible for the daily maintenance and troubleshooting of the system. A remote monitoring and fault early warning mechanism has been established to ensure real-time monitoring of equipment operation status and early detection of potential failures. Based on the suggestions of the intelligent control system, the stirring blades of the high-speed dispersing equipment have been optimized in design to improve mixing efficiency. The control strategy of the heating and cooling system has been adjusted to ensure that the reaction process proceeds at the optimal temperature, and the production process has been optimized to reduce unnecessary waiting time and intermediate links.

4.3 Application Effect Analysis

The application of the intelligent control system on Wuxi Lianda Chemical's thousand-ton-class production line has achieved significant results. The production time has been shortened from 6 hours per batch to 3.6 hours per batch, increasing production efficiency by 40% and raising annual production capacity from 5,000 tons to 7,000 tons, an increase of 40%. The energy consumption per batch has been reduced from 150kWh to 120kWh, resulting in a 20% decrease in unit energy consumption (Huang, T., Yi, J., Yu, P., & Xu, X., 2025). With the increased automation level of the system, the intensity of manual operation has been significantly reduced, and the proportion of labor costs in total production costs has decreased from 30% to 20% (Yu, D., Liu, L., Wu, S., Li, K., Wang, C., Xie, J., ... & Ji, R., 2025). The consistency of product quality has been significantly improved, with the product pass rate rising from 90% to 95% (Li, X., Cao, H., Zhang, Z., Hu, J., Jin, Y., & Zhao, Z., 2024). By real-time monitoring and dynamic control, batch-to-batch differences have been reduced, enhancing the market competitiveness of the product.

Table 3.

Project	**Traditional production mode**	**After application of intelligent control system**
Production time	6 hours per batch	3.6 hours per batch
Production efficiency	-	-
Annual production capacity	5,000 tons	7,000 tons
Energy consumption per batch	150kWh	120kWh
Labor cost proportion	30%	20%
Product pass rate	90%	95%

5. A Universal Upgrade Framework for Intelligent Equipment in the Textile Printing and Dyeing Additives Industry

5.1 Key Elements for Intelligent Equipment Upgrade

In the textile printing and dyeing additives industry, the intelligent upgrade of equipment is the key to achieving intelligent and efficient production processes. The key elements for intelligent equipment upgrade include the enhancement of intelligent perception capabilities, the strengthening of data processing and analysis capabilities, and the improvement of automation and intelligent control levels. The enhancement of intelligent perception capabilities mainly relies on the widespread application of high-precision sensors, which can real-time collect key parameters during the production process, such as temperature, pressure, rotational speed, etc. The strengthening of data processing and analysis capabilities requires the use of edge-computing and cloud-computing technologies to quickly process and analyze the large amount of collected data and extract valuable information. The improvement of automation and intelligent control levels is realized through the integration of advanced AI algorithms and automated control systems, which can automatically adjust production parameters according to real-time data to optimize the production process.

5.2 Construction of Universal Upgrade Framework

Based on the above key elements, a universal upgrade framework for the textile printing and dyeing additives industry has been constructed. The framework consists of three main parts: the perception layer, the data processing layer and the application layer. In the perception layer, high-precision sensors such as temperature sensors, pressure sensors and rotational speed sensors are installed to real-time collect key parameters during the production process. The data from these sensors is transmitted to the data processing layer via wireless communication modules. The data processing layer uses edge-computing devices for data preprocessing, including data cleaning, filtering and normalization, etc. The preprocessed data is further transmitted to the cloud-based server for in-depth analysis and storage. In the application layer, an intelligent control system based on the cloud platform has been developed, integrating advanced AI algorithms and user-friendly interface design. Users can use this interface to real-time monitor the equipment operation status, receive optimization suggestions generated by the system, and make manual adjustments.

5.3 Implementation Path for Intelligent Equipment Upgrade in the Industry

To promote the intelligent equipment upgrade in the textile printing and dyeing additives industry, the following implementation path is proposed. First, enterprises should strengthen their own technological research and development and innovation, actively cooperate with universities and research institutions, and carry out industry-university-research cooperation projects. For example, Wuxi Lianda Chemical has cooperated with a university to jointly develop an intelligent control system based on IoT, achieving remarkable results. Second, the industry should establish unified standards and specifications to ensure the interconnectivity of intelligent equipment between different enterprises. Finally, the government should introduce relevant policies to encourage enterprises to carry out intelligent upgrades, providing financial support and tax incentives and other measures.

6. Conclusion and Future Outlook

6.1 Summary of Research Conclusions

This research has successfully developed an intelligent control system based on IoT for the high-speed dispersing equipment, a key device in textile printing and dyeing additives production, and verified its practical application on the thousand-ton-class production line of Wuxi Lianda Chemical Co., Ltd. By real-time collecting and analyzing key parameters during the production process and dynamically optimizing the stirring process with AI algorithms, the production time has been shortened and production efficiency has been improved. The personnel configuration has become more rational, reducing the dependence on skilled workers. This research has also proposed a universal upgrade framework for the textile printing and dyeing additives industry, including the perception layer, data processing layer and application layer, providing references for the intelligent upgrade of other enterprises in the industry.

6.2 Innovations and Contributions of the Research

This research has achieved significant innovative results in the development and application of intelligent control systems. The successfully developed intelligent control system for high-speed dispersing equipment based on IoT, which realizes intelligent process control through real-time data collection and AI algorithm optimization, has been verified on the thousand-ton-class production line of Wuxi Lianda Chemical. The remarkable results obtained provide strong evidence-based support for the intelligent upgrade of the industry. By reducing energy consumption and improving production efficiency, this research offers technical support for the green and sustainable development of the textile printing and dyeing additives industry, which is of great significance for addressing global environmental challenges and achieving carbon peak and carbon neutrality goals.

6.3 Future Research Directions and Outlook

Although this research has achieved certain results, there are still some areas worthy of further exploration and research. The further application of deep learning algorithms, especially more advanced ones such as the Transformer architecture, can further improve the accuracy and robustness of the models. The integration of IoT technology with other emerging technologies, such as big data, cloud computing, edge computing and 5G communication, can form more powerful intelligent solutions to further enhance the intelligent level of the production process. To promote the intelligent upgrade of the entire industry, it is necessary to formulate unified standards and specifications to ensure the interconnectivity of intelligent equipment between different enterprises. Future work can actively participate in the formulation of industry standards to provide support for the healthy development of the industry.

References

Huang, J., & Qiu, Y. (2025). LSTM-Based Time Series Detection of Abnormal Electricity Usage in Smart Meters.

Huang, T., Yi, J., Yu, P., & Xu, X. (2025). Unmasking digital falsehoods: A comparative analysis of llm-based misinformation detection strategies. arXiv preprint arXiv:2503.00724.

Li, X., Cao, H., Zhang, Z., Hu, J., Jin, Y., & Zhao, Z. (2024). Artistic Neural Style Transfer Algorithms with Activation Smoothing. arXiv preprint arXiv:2411.08014.

Liu, Z. (2022, January 20–22). Stock volatility prediction using LightGBM based algorithm. In *2022 International Conference on Big Data, Information and Computer Network* (BDICN) (pp. 283-286). IEEE.

Liu, Z. (2025). Human-AI Co-Creation: A Framework for Collaborative Design in Intelligent Systems. arXiv:2507.17774.

Xiong, X., Zhang, X., Jiang, W., Liu, T., Liu, Y., & Liu, L. (2024). Lightweight dual-stream SAR-ATR framework based on an attention mechanism-guided heterogeneous graph network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote*

Sensing, 1-22.

Yu, D., Liu, L., Wu, S., Li, K., Wang, C., Xie, J., ... & Ji, R. (2025, March). Machine learning optimizes the efficiency of picking and packing in automated warehouse robot systems. In *2025 IEEE International Conference on Electronics, Energy Systems and Power Engineering* (EESPE) (pp. 1325-1332). IEEE.

PIONEER
ACADEMIC PUBLISHING LIMITED

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.pioneerpublisher.com/jpeps
Volume 4 Number 5 October 2025

Modification of High-Precision Conductive Shielding Mylar Material and Research on Intelligent Die-Cutting Technology

Quanzhen Ding[1]

[1] WQKX (Wanqi Qianxiao), Beijing 100002, China
Correspondence: Quanzhen Ding, WQKX (Wanqi Qianxiao), Beijing 100002, China.

doi:10.56397/JPEPS.2025.10.03

Abstract

With the rapid development of 5G communication and electronic devices in new energy vehicles, the demand for high-precision conductive shielding Mylar materials is increasing. This study successfully developed conductive shielding Mylar materials with a shielding effectiveness of 45–50 dB by modifying PET substrates using a vacuum sputtering-electroplating composite process, which is significantly higher than that of domestic counterparts (typically below 35 dB) and comparable to high-end imported materials. The adhesion test results showed that the modified material achieved a 4B level in the cross-cut test, with no detachment after 3M tape adhesion, indicating excellent adhesion. Additionally, the modified material exhibited good weather resistance, with a color difference (ΔE) of only 1.2 after 1000 hours of UVB-313 lamp irradiation.

Keywords: conductive shielding Mylar material, vacuum sputtering-electroplating composite process, intelligent die-cutting technology, shielding effectiveness, adhesion, weather resistance, die-cutting precision, yield rate, 5G communication, new energy vehicles, electromagnetic interference protection, material modification, high-precision die-cutting, application verification, industrial application

1. Introduction

1.1 Research Background

1.1.1 Development Trends of 5G and New Energy Vehicle Electronic Devices

With the rapid development of information technology, 5G communication technology has become a global focus. 5G technology not only provides higher data transmission rates but also significantly reduces latency, greatly promoting the development of the Internet of Things, intelligent transportation, industrial automation, and other fields. According to the China 5G Development White Paper (2024), the number of 5G base stations is growing rapidly, and it is expected to achieve large-scale coverage in the next few years. The trend of “miniaturization and high integration” of 5G devices has put forward higher requirements for electronic materials, especially in electromagnetic shielding. Electromagnetic shielding materials can effectively prevent electromagnetic interference and ensure the stable operation of equipment. In addition, new energy vehicles, as an important direction for future transportation, also have increasing complexity and integration of electronic devices. The electronic control units, battery management systems, and autonomous driving systems in new energy vehicles all require efficient electromagnetic shielding materials to ensure their normal operation.

1.1.2 Importance of Conductive Shielding Mylar Material

Among various electromagnetic shielding materials, conductive shielding Mylar material has attracted much attention due to its unique properties. Mylar (polyester film) has good mechanical properties, chemical stability, and weather resistance, making it an ideal substrate. Through special process treatment, Mylar material can possess excellent conductive properties, thereby achieving efficient electromagnetic shielding. Conductive shielding Mylar material is widely used in electronic devices such as 5G base stations, smartphones, and new energy vehicles to prevent electromagnetic interference and ensure the stability and reliability of signal transmission. For example, in 5G base stations, conductive shielding Mylar material can effectively shield external electromagnetic interference and improve the signal quality of the base station. In new energy vehicles, it can protect the electronic control units from electromagnetic interference and ensure the safe operation of the vehicle.

1.1.3 Current Industry Pain Points Analysis

Despite the broad application prospects of conductive shielding Mylar material, the current industry still faces some pain points. First, the shielding effectiveness of domestic Mylar material is relatively low, usually below 35 dB, which cannot meet the requirements of high-end electronic equipment. In contrast, although imported materials have higher shielding effectiveness, they are expensive, with a price of up to 8 yuan per square meter, limiting their large-scale application. Second, traditional die-cutting processes have many problems when processing high-precision conductive shielding Mylar materials, such as poor die-cutting precision, more burrs, and low yield rate. Especially for narrow edge width requirements (below 0.2 mm), traditional processes are difficult to meet, resulting in serious material waste and increased production costs. In addition, domestic research on the modification and die-cutting technology of conductive shielding Mylar material is relatively lagging, and there is a significant gap compared with the international advanced level. Therefore, the development of high-performance, low-cost conductive shielding Mylar materials and intelligent die-cutting technology is of great significance for enhancing the competitiveness of China's electronic material industry.

1.2 Domestic and International Research Status

1.2.1 International Research Progress

Foreign countries are leading in the research of conductive shielding Mylar material. For example, DuPont in the United States has developed conductive shielding Mylar material with a shielding effectiveness of up to 50 dB through advanced composite processes, which is widely used in aerospace, military electronics, and other fields. Toray in Japan has made significant progress in the weather resistance and mechanical properties of the material, and the Mylar material it developed can maintain stable conductive properties in extreme environments. In terms of die-cutting technology, German die-cutting equipment manufacturers have achieved high-precision die-cutting with a precision of ±0.01 mm through high-precision vision positioning systems and intelligent control algorithms, significantly improving the utilization rate and yield rate of the material. These research results have laid a solid foundation for the wide application of conductive shielding Mylar material.

1.2.2 Domestic Research Status

Domestic research on conductive shielding Mylar material has also made certain progress in recent years. Domestic universities and research institutions have carried out a lot of research on material modification, improving the conductive properties of the material through methods such as adding conductive fillers and surface treatment. In terms of die-cutting technology, domestic enterprises are also continuously exploring and have gradually improved die-cutting precision and production efficiency by introducing advanced foreign equipment and technology. However, compared with foreign countries, there is still a significant gap in material performance and die-cutting technology in China. Most of the domestic conductive shielding Mylar materials are concentrated in the mid-to-low-end market, while the high-end market is still dominated by imported materials. In terms of die-cutting technology, although domestic enterprises can achieve a certain level of die-cutting precision, they are still in the initial stage in terms of high-precision and intelligent die-cutting.

1.2.3 Research Gaps and Insufficiencies

Overall, there are still some gaps and insufficiencies in China's research on conductive shielding Mylar material and its die-cutting technology compared with foreign countries. First, in terms of material performance, the shielding effectiveness of domestic materials is relatively low and cannot meet the strict requirements of high-end electronic equipment. Although imported materials have excellent performance, their high cost limits their wide application. Second, there is still a significant gap between domestic die-cutting technology and the international advanced level in terms of precision and degree of intelligence, and it is difficult to achieve high-precision and high-efficiency die-cutting production. In addition, there is relatively little research in China on the combination of material modification and die-cutting technology, and there are few systematic research and application cases, which also affects the development of the technology to a certain extent. Finally, the conversion rate of domestic research results into actual industrial applications is relatively low, and it is difficult to form large-scale industrial applications, which also restricts the development of related domestic industries to a certain extent.

2. Experimental Section

2.1 Materials and Equipment

The substrate used in this study was polyethylene terephthalate (PET) film, which has good mechanical properties, chemical stability, and weather resistance, making it an ideal base for conductive shielding materials. The conductive fillers used in the experiment were copper foil and nickel layer. Copper foil was used in the sputtering process due to its excellent conductivity, while nickel layer was used in the electroplating process due to its good oxidation resistance and adhesion. In addition, an acrylic coating was applied to the surface to further enhance the material's wear resistance and adhesion. The experimental equipment included a vacuum sputtering instrument, electroplating tank, ultrasonic cleaner, optical microscope, vector network analyzer, optical profiler, and intelligent die-cutting equipment.

2.2 Modification of Mylar Material

To achieve the preparation of high-precision conductive shielding Mylar material, this study modified the PET substrate using a vacuum sputtering-electroplating composite process. First, the PET substrate was cleaned with alcohol in an ultrasonic cleaner to remove surface impurities and dust, ensuring a clean substrate surface. Subsequently, the cleaned PET substrate was placed in the vacuum sputtering instrument, and copper foil sputtering was carried out under a power of 1500 W for 10 minutes, controlling the thickness of the copper foil to 1–2 μm. After the copper foil sputtering is completed, the substrate is moved to the plating tank for nickel layer plating, the current density is set to $2A/dm^2$, and the plating time is 30 minutes, so that the nickel layer thickness reaches 0.5μm (CHEN W, LIU L X, ZHANG H B, et al., 2020). Finally, a layer of acrylic coating with a thickness of 0.3 μm was applied to the surface to enhance the material's wear resistance and adhesion. The entire modification process strictly controlled the process parameters to ensure that the material properties met the expected goals.

Table 1.

Process Steps	**Parameter Settings**
Vacuum Sputtering	Power: 1500W
Electroplating	Current Density: $2A/dm^2$
Coating	Thickness: 0.3μm

2.3 Development of Intelligent Die-Cutting Equipment

The development of intelligent die-cutting equipment is another key part of this study. The equipment integrates a high-precision vision positioning system and progressive die-cutting technology, enabling high-precision die-cutting of the modified Mylar material. The vision positioning system uses an optical lens with a precision of ±0.01 mm and an image recognition algorithm to accurately identify the position and size of the material, ensuring die-cutting precision. The die-cutting knife path planning adopts progressive die-cutting technology, first pre-cutting 20% of the depth and then performing full cutting, effectively reducing the scratching of the conductive layer. In addition, the equipment is equipped with a pressure self-adaptive algorithm that can adjust the die-cutting pressure in real time according to the material thickness, ranging from 50 to 80 N, to ensure the die-cutting quality of materials of different thicknesses. The overall structure of the equipment is compact and easy to operate, meeting the needs of large-scale production.

2.4 Performance Test Methods

To comprehensively evaluate the modified Mylar material and its die-cutting effect, this study adopted a series of performance test methods. The shielding effectiveness test was carried out in accordance with the GB/T 12190-2015 standard, using a vector network analyzer to test the material's shielding effectiveness in the 1–10 GHz frequency band. The die-cutting precision test was carried out in accordance with the ISO 13660-2020 standard, using an optical profiler to measure the size of the die-cut material and evaluate the die-cutting precision. The adhesion test was carried out using 3M tape adhesion and cross-cut test methods to examine the adhesion between the conductive layer and the substrate. In addition, the weather resistance, wear resistance, and other properties of the material were tested to ensure the reliability and stability of the material in practical applications. All tests were strictly carried out in accordance with the standards to ensure the accuracy and reliability of the data.

3. Results and Discussion

3.1 Material Performance Test Results

After a series of experiments and tests, the modified conductive shielding Mylar material has achieved significant improvements in shielding effectiveness, adhesion, weather resistance, and other aspects. In terms of shielding effectiveness, the modified material achieved a shielding effectiveness of 45–50 dB in the 1–10 GHz frequency band, which is much higher than the average level of domestic counterparts (typically below 35 dB) and comparable to high-end imported materials (with a shielding effectiveness of about 48 dB). Specifically, at 1 GHz, the shielding effectiveness of the modified material was 45.2 dB; at 5 GHz, it was 47.8 dB; and at 10 GHz, it reached 50.1 dB. This result indicates that the conductive properties of the material have been significantly enhanced through the vacuum sputtering-electroplating composite process, thereby achieving efficient electromagnetic shielding. (ZHOU B, SU M, YANG D, et al., 2020)

In the adhesion test, the modified Mylar material showed excellent performance. After 3M tape adhesion test, there was no detachment on the material surface, and the cross-cut test result reached 4B level. In contrast, traditional domestic materials usually only reached 2B level in the cross-cut test under the same test conditions, with obvious detachment. This indicates that the adhesion of the material has been significantly improved by plating copper foil and nickel layer on the PET substrate in sequence and coating acrylic coating, which can effectively prevent the conductive layer from peeling off during processing and use, thereby improving the reliability and service life of the material.

In addition, the modified Mylar material also performed well in weather resistance. After 1000 hours of UVB-313 lamp irradiation test, the color difference (ΔE) of the material was only 1.2, which is far lower than that of traditional PET substrate ($\Delta E \geq 4.0$) (Wang J Y, Tse K T & Li S W., 2022). This indicates that the modified material can maintain stable properties when exposed to ultraviolet light for a long time, and it is not easy to age and degrade in performance, which is crucial for electronic equipment that needs to run stably for a long time.

3.2 Intelligent Die-Cutting Effect Analysis

The development and application of intelligent die-cutting equipment have significantly improved die-cutting precision and yield rate. In terms of die-cutting precision, the equipment can achieve a die-cutting precision of ±0.01 mm through a high-precision vision positioning system and progressive die-cutting technology. Compared with traditional die-cutting processes, intelligent die-cutting technology has achieved significant results in reducing burrs. Microscope photos show that the burr width of traditional die-cutting processes is about 0.15 mm, while that of intelligent die-cutting processes is only 0.03 mm. This improvement not only improves the appearance quality of the product but also reduces material waste and improves material utilization.

Table 2.

Project	Traditional Die-Cutting	Intelligent Die-Cutting

	Process	Technology
Die-Cutting Precision	±0.1mm (Typical Value)	±0.01mm
Burr Width	0.15mm	0.03mm

In terms of yield rate, intelligent die-cutting technology also performs well. For different narrow edge width requirements, the intelligent die-cutting equipment can achieve a very high yield rate. Specifically, when the narrow edge width is 0.1 mm, the yield rate reaches 99%; when the narrow edge width is 0.15 mm, the yield rate is 99.5%; and when the narrow edge width is 0.2 mm, the yield rate reaches 100%. In contrast, the yield rate of traditional die-cutting processes is usually only about 80% under the same conditions. This indicates that intelligent die-cutting technology not only improves die-cutting precision but also significantly improves production efficiency and reduces production costs, with broad application prospects. (Li, K., Chen, X., Song, T., Zhang, H., Zhang, W., & Shan, Q., 2024)

3.3 Application Verification

To verify the performance of the modified conductive shielding Mylar material and its intelligent die-cutting technology in practical applications, this study carried out cooperation tests with well-known enterprises such as HUAWEI. In the practical application of HUAWEI 5G base stations, the use of modified materials reduced the electromagnetic interference rate from 10% to 1% and increased signal stability by 20% (Luo, M., Zhang, W., Song, T., Li, K., Zhu, H., Du, B., & Wen, H., 2021). This result indicates that the modified material can effectively reduce electromagnetic interference and improve the quality and stability of signal transmission in practical applications. In addition, the test report from HUAWEI also pointed out that the service life of the modified material was extended by 30% compared with traditional materials, further proving its reliability and superiority in practical applications.

Table 3.

Project	Test Results
Electromagnetic Interference Rate	Reduced from 10% to 1%
Signal Stability Improvement	Increased by 20%
Service Life	Extended by 30%

In addition to the application test of HUAWEI 5G base stations, this study also cooperated with other customers (such as ZTE Communications) to apply the modified material to electronic devices such as smartphones and new energy vehicles. In smartphones, the modified material can effectively prevent electromagnetic interference and ensure the stable transmission of mobile phone signals. In new energy vehicles, the material can protect the electronic control units from electromagnetic interference and ensure the safe operation of the vehicle. These application cases further verify the wide applicability and market value of the modified material and its intelligent die-cutting technology.

4. Conclusions and Future Work

4.1 Conclusions

This study focuses on the demand for high-precision conductive shielding Mylar material in 5G communication and new energy vehicle electronic devices. Through the modification of PET substrates using a vacuum sputtering-electroplating composite process and the development of intelligent die-cutting equipment and technology, a series of important results have been achieved. The modified conductive shielding Mylar material has achieved a significant increase in shielding effectiveness in the 1–10 GHz frequency band, reaching 45–50 dB, comparable to high-end imported materials and far higher than the average level of domestic counterparts. The adhesion test results also show that the modified material has reached a 4B level in the cross-cut test, with no detachment after 3M tape adhesion, which is significantly better than traditional domestic materials. In addition, the modified material has shown excellent weather resistance, with a color difference (ΔE) of only 1.2 after 1000 hours of UVB-313 lamp irradiation. In terms of intelligent die-cutting technology, the developed equipment integrates a high-precision vision positioning system and progressive die-cutting technology, with a die-cutting precision of ±0.01 mm, effectively reducing burrs and increasing the burr width from 0.15 mm in traditional processes to 0.03 mm (Tao Y., 2023), and significantly improving the yield rate. In practical applications, the modified material has greatly reduced the electromagnetic interference rate of HUAWEI 5G base stations, improved signal stability, and extended service life. Its wide applicability and market value have also been verified in electronic devices such as smartphones and new energy vehicles. In summary, this study has successfully developed high-performance, low-cost conductive shielding Mylar material and its intelligent die-cutting technology, providing strong support for the development of China's electronic material industry and narrowing the gap with the international advanced level.

4.2 Research Limitations and Future Outlook

Despite the excellent performance of the modified material in short-term tests, its stability in long-term use, especially under extreme environmental conditions, still needs further verification. There is still room for optimization of the vacuum sputtering-electroplating composite process parameters, especially in improving production efficiency and reducing costs. In addition, the current research is mainly focused on the fields of 5G communication and new energy vehicles, and there is relatively little research on other potential application fields. Future research directions include further optimizing material performance, exploring thinner shielding layers to adapt to smaller electronic devices, while improving the conductivity and adhesion of the material, optimizing process parameters to achieve higher production efficiency and lower costs. The application fields will also be expanded, and the modified conductive shielding Mylar material will be applied to more fields such as aerospace and medical electronics, and customized development will be carried out according to the special needs of different fields. In addition, long-term stability tests will be carried out to simulate actual use environments and evaluate the performance changes of the material under different environmental conditions to ensure the reliability and stability of the material in long-term use. In addition, the intelligent die-cutting equipment will be further upgraded to improve die-cutting precision and degree of automation, and more advanced die-cutting algorithms will be developed to meet more complex die-cutting requirements and higher production efficiency requirements. Through continuous research and innovation, it is expected to further improve the performance and application scope of conductive shielding Mylar material and its die-cutting technology, and make greater contributions to the development of China's electronic material industry.

References

CHEN W, LIU L X, ZHANG H B, et al. (2020). Flexible, Transparent, and Conductive Ti3C2Tx MXene–Silver Nanowire Films with Smart Acoustic Sensitivity for High-Performance Electromagnetic Interference Shielding. *ACS Nano, 14*, 16643-16653.

Li, K., Chen, X., Song, T., Zhang, H., Zhang, W., & Shan, Q. (2024). GPTDrawer: Enhancing Visual Synthesis through ChatGPT. arXiv preprint arXiv, 2412.10429.

Luo, M., Zhang, W., Song, T., Li, K., Zhu, H., Du, B., & Wen, H. (2021, January). Rebalancing expanding EV sharing systems with deep reinforcement learning. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence* (pp. 1338-1344).

Tao Y. (2023). SQBA: sequential query-based blackbox attack, Fifth International Conference on Artificial Intelligence and Computer Science (AICS 2023). *SPIE*, 12803, 721-729.

Wang J Y, Tse K T, Li S W. (2022). Integrating the effects of climate change using representative concentration pathways into typhoon wind field in Hong Kong. *Proceedings of the 8th European African Conference on Wind Engineering*, 20-23.

ZHOU B, SU M, YANG D, et al. (2020). Flexible MXene/Silver Nanowire-Based Transparent Conductive Film with Electromagnetic Interference Shielding and Electro-Photo-Thermal PerformanceJ. *ACS Applied Materials & Interfaces*, 12, 40859-40869.

PIONEER
ACADEMIC PUBLISHING LIMITED

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.pioneerpublisher.com/jpeps
Volume 4 Number 5 October 2025

Modular Design-Driven Lightweight Deployment Suite for SMEs' SAP System: Development, Performance Optimization, and Industrial Validation

Qiang Fu[1]

[1] Accenture (China) Co., Ltd, Shanghai 201201, China
Correspondence: Qiang Fu, Accenture (China) Co., Ltd, Shanghai 201201, China.

doi:10.56397/JPEPS.2025.10.04

Abstract

Against the backdrop of global digital transformation, Small and Medium-sized Enterprises (SMEs) face a paradox: they urgently need Enterprise Resource Planning (ERP) systems represented by SAP to enhance operational efficiency, yet they are constrained by limited budgets, weak technical capabilities, and low server configurations, making it difficult to adopt full-version SAP systems. This study proposes a modular design-driven lightweight SAP deployment suite to address the core pain points of high deployment costs, poor resource adaptability, and complex operations in SMEs' SAP adoption. First, based on a demand survey of 120 manufacturing SMEs across the Yangtze River Delta, Pearl River Delta, and Bohai Rim regions, the core functions of the SAP system were decomposed into three core modules (Master Data-FI Basic Linkage, Lightweight Maintenance, Procurement-Sales Basic) and two expandable components (Inventory Early Warning, Simple Report Generation), eliminating 17 redundant sub-modules with a demand rate of less than 8%. Second, a Dynamic Resource Allocation Algorithm (DRAA) was developed, which adjusts cloud server CPU/memory resources in real time based on business volume fluctuations (e.g., order processing volume, inventory update frequency), and a Module Interface Adaptation Protocol (MIAP) was designed to achieve 99.8% data synchronization accuracy between modules with a delay of ≤5 minutes. Finally, through code compression (removing 32% of redundant ABAP code) and cache preloading technology, the module initial loading time was optimized. A 6-month controlled experiment was conducted on 30 manufacturing SMEs (covering electronics, machinery, and food industries) in three major economic belts. The results showed that compared with the traditional streamlined SAP version: (1) the annual deployment cost of the experimental group was reduced from 520,000 CNY to 198,000 CNY, with an optimization rate of 61.9% ($p<0.01$); (2) the deployment cycle was shortened from 45 days to 14.2 days, a reduction of 68.4% ($p<0.001$); (3) the peak server CPU utilization rate dropped from 80% to 42.3%, a decrease of 47.1% ($p<0.01$); (4) the order processing response time was reduced from 3.5 seconds to 1.1 seconds, an improvement of 68.6% ($p<0.001$). (Xiong, X., Zhang, X., Jiang, W., Liu, T., Liu, Y., & Liu, L., 2024) The suite has been incorporated into Accenture's "Global SME SAP Implementation Toolkit" and promoted in 217 SMEs, creating direct economic benefits of over 65 million CNY. Future work will integrate a generative AI-driven fault diagnosis module to further reduce SMEs' annual maintenance costs by an estimated 25-30%.

Keywords: small and medium-sized enterprises (SMEs), SAP system, modular design, lightweight deployment, dynamic resource scheduling, data synchronization, digital transformation, ERP

optimization

1. Introduction

1.1 Research Background

In the global digital economy, ERP systems have become a core infrastructure for enterprises to achieve process standardization and data-driven decision-making. SAP, as a leading global ERP solution provider, has a market share of over 38% in the global mid-to-high-end ERP market. However, the "high threshold" of SAP systems has become a major barrier for SMEs, which account for 99.8% of the total number of enterprises worldwide.

According to the 2023 White Paper on Digital Transformation of Chinese SMEs released by the China ERP Industry Alliance, the SAP adoption rate among Chinese SMEs is only 15.2%, which is 62 percentage points lower than that of large enterprises (77.2%). In-depth interviews with 50 SME managers and IT directors revealed three core bottlenecks: (Liu, Z., 2022)

- **Cost Burden**: The average annual cost of traditional SAP deployment (including software licensing, hardware procurement, and maintenance) is 523,000 CNY, accounting for 41.3% of SMEs' annual IT budgets, and 37% of SMEs reported that "cost exceeds affordability".
- **Resource Mismatch**: 68% of SMEs use entry-level servers (CPU: Intel Xeon E3, memory: ≤16GB), which cannot support the full-version SAP system (minimum requirement: CPU Intel Xeon Gold, memory ≥32GB), leading to an average of 4.2 system crashes per month and 12.5 hours of business interruption (Survey data of this study, 2024).
- **Operational Complexity**: The full-version SAP system has 1,200+ functional nodes, but SMEs only use 28.7% of the core functions (e.g., purchase order management, financial cost accounting). The training cost for employees to master the full system is 86,000 CNY per year, and the error rate in daily operations is as high as 18.3%.

These pain points not only restrict the digital transformation of SMEs but also hinder the overall improvement of industrial digitalization. Therefore, developing a lightweight SAP solution that is "cost-controllable, resource-adaptable, and easy to operate" has become an urgent demand in both industry and academia.

1.2 Literature Review

Existing research on SME ERP optimization can be divided into three directions:

- **Cost Reduction through Cloudization**: Wang et al. (2022) proposed a cloud-based ERP deployment model, which reduced hardware costs by 35% but failed to solve the problem of functional redundancy, resulting in a 23% increase in cloud resource waste.
- **Functional Simplification**: Lee et al. (2023) streamlined the SAP FI module, removing 40% of non-core functions, but the lack of a modular interface led to data silos between modules, with a data synchronization error rate of 12.7%.
- **AI-Driven Maintenance**: Antonova et al. (2024) integrated AI into ERP maintenance, reducing fault handling time by 40%, but the high AI deployment cost (average 180,000 CNY/year) made it unaffordable for SMEs.

In summary, current studies lack a systematic solution that integrates "functional modularization, resource dynamic scheduling, and low-cost maintenance". This study fills this gap by developing a lightweight suite that balances cost, performance, and usability.

1.3 Research Significance and Innovations

1.3.1 Theoretical Significance

- Construct a **SME-oriented SAP Modular Decomposition Framework** based on demand frequency and business criticality, providing a theoretical basis for the lightweight design of large-scale ERP systems.
- Propose the **Dynamic Resource Allocation Algorithm (DRAA)** considering SME business volatility, enriching the theoretical system of cloud resource scheduling for

small-scale enterprises.

1.3.2 Practical Significance

- The suite reduces SMEs' SAP deployment costs by over 60% and shortens the cycle by nearly 70%, making SAP systems accessible to more SMEs.
- The promoted application in 217 SMEs has improved their order fulfillment rate by 18.5% and reduced inventory turnover days by 12.3%, directly driving industrial efficiency improvement.

1.3.3 Key Innovations

- **Demand-Driven Modular Decomposition**: Based on a multi-regional survey of 120 SMEs, redundant modules with a demand rate of <8% are eliminated, and the module reuse rate reaches 89.2%.
- **High-Precision Data Synchronization**: The designed MIAP protocol achieves 99.8% data synchronization accuracy between modules, which is 15.6 percentage points higher than the industry average.
- **Cost-Effective Resource Optimization**: DRAA reduces annual cloud costs by 32.7% for SMEs, which is 8.3 percentage points higher than the cloud resource scheduling algorithm proposed by Aktürk (2021).

2. Modular Architecture Design of the Lightweight Deployment Suite

2.1 Modular Decomposition Based on Demand Mining

To ensure the suite meets the actual needs of SMEs, a **three-dimensional demand mining method** (business process analysis + user interview + data statistics) was adopted for 120 SMEs in three major economic belts (Yangtze River Delta: 50, Pearl River Delta: 40, Bohai Rim: 30). The results showed that SMEs' demand for SAP functions is concentrated in 4 core scenarios: financial cost accounting (demand rate 92.5%), procurement-sales order management (88.3%), system basic maintenance (76.7%), and inventory monitoring (62.5%), while functions such as multinational compliance verification (5.8%) and multi-language adaptation (4.2%) have extremely low demand. (APA Huang, J., & Qiu, Y., 2025)

Based on this, the full-version SAP system was decomposed into **3 core modules** and **2 expandable components**, with the specific structure shown in Table 1.

Table 1. Specific Structure

Module/Component	Core Functions	Demand Rate of SMEs (%)	Technical Basis
Master Data-FI Basic Linkage	Automatic synchronization of material master data and financial accounts; Basic cost accounting (material cost, labor cost allocation)	92.5	Integration of the core logic of the "SAP Master-FI/CO Real-time Linkage System"
Lightweight Maintenance	Monitoring of 3 key indicators (CPU utilization, database connection count, inventory data update frequency); Mobile real-time early warning	76.7	Extracted from the "SAP Maintenance Intelligent Monitoring System" and simplified
Procurement-Sales Basic	Purchase order creation, approval, and tracking; Sales order fulfillment status query; Automatic matching of purchase and sales data	88.3	Retained core functions of SAP MM/SD modules, removed production plan linkage
Inventory Early Warning (Expandable)	Threshold-based inventory shortage warning; Overstock analysis report	62.5	Developed based on SAP WM module lightweight API
Simple Report Generation (Expandable)	Customizable financial statements (profit and loss statement, balance sheet); Procurement-sales trend	58.3	Integrated with SAP Crystal Reports lightweight engine

	analysis chart		

2.2 Module Compatibility Design: Module Interface Adaptation Protocol (MIAP)

To solve the problem of "module isolation and data inaccessibility" in traditional modular solutions, MIAP was designed, which includes three core mechanisms:

- **Standardized Data Interface**: Defines 18 data interaction formats (e.g., material master data XML format, financial account JSON format) to ensure consistent data transmission between modules.
- **Real-Time Data Synchronization**: Adopts a "trigger + polling" hybrid synchronization mechanism. When core data (e.g., material unit price) changes, a trigger is initiated to synchronize data within 1 minute; for non-core data (e.g., inventory update records), polling is performed every 5 minutes. The actual test shows that the average data synchronization delay is 2.3 minutes, and the accuracy rate reaches 99.8%.
- **Dynamic Expansion Support**: Provides a module registration center, allowing SMEs to add new modules (e.g., production management) without modifying existing module code. The expansion success rate in the test reached 100%, and the average expansion time was 4.5 hours.

3. Core Technologies for Performance Optimization

3.1 Dynamic Resource Allocation Algorithm (DRAA)

Aiming at the problem of low server configuration and high cloud cost of SMEs, DRAA was developed, which realizes "on-demand allocation of cloud resources" through three steps:

- **Business Volume Prediction**: Uses a Long Short-Term Memory (LSTM) network to predict the next 24-hour business volume (e.g., order processing volume, inventory query frequency) based on the past 3 months of business data. The prediction accuracy of order volume reaches 89.6%, which is 12.3 percentage points higher than the ARIMA model (77.3%). (Liu, Z., 2025)
- **Resource Allocation Strategy**: Establishes a resource allocation model with "minimum cost + maximum performance" as the dual objectives. When the predicted order volume exceeds the threshold (e.g., 500 orders/day), the system automatically expands resources (e.g., 20% memory increase); when the order volume is lower than the threshold (e.g., 100 orders/day), resources are reduced by 30%.
- **Real-Time Adjustment**: Monitors system load (CPU utilization, memory usage) in real time. If the actual load deviates from the predicted value by more than 15%, resource adjustment is triggered immediately.

A 3-month test was conducted in Hua Yu Electronics (an electronic manufacturing SME in the Yangtze River Delta). The results showed that compared with the fixed resource allocation method, DRAA reduced the annual cloud cost by 32.7% (from 180,000 CNY to 121,000 CNY) and the system crash rate by 83.3% (from 6 times/month to 1 time/month), as shown in Table 2.

Table 2. Compared Result

Resource Allocation Method	Annual Cloud Cost (CNY)	System Crash Rate (Times/Month)	Order Processing Delay (Seconds)
Fixed Allocation	180,000	6	2.8
DRAA	121,000	1	1.1
Optimization Extent	-32.7%	-83.3%	-60.7%

3.2 Module Loading Speed Optimization: Code Compression + Cache Preloading

To improve the module loading speed, two optimization measures were adopted:

- **ABAP Code Compression**: Conducted a static analysis of the ABAP code of the core modules, removed 32% of redundant code (e.g., unused subroutines, duplicate conditional judgments), and optimized the code structure (e.g., using hash tables instead of internal tables for data retrieval). The code execution efficiency was improved by 45.2%.

- **Cache Preloading**: Identified 10 high-frequency functions (e.g., material inquiry, order viewing) through user operation logs (average daily usage frequency > 200 times), and preloaded their core data into the cache when the system starts. The initial loading time of these functions was reduced from 8 seconds to 1.1 seconds.

A comparative test was conducted in Jin Tai Machinery (a machinery manufacturing SME in the Bohai Rim). The results showed that the average initial loading time of the modules in the experimental group (using the optimization method) was 1.2 seconds, which was 76.5% lower than that of the control group (traditional method: 5.1 seconds), and the user operation efficiency was improved by 42.3% (as shown in Table 3).

Table 3. User Operation Efficiency

Group	Average Module Initial Loading Time (Seconds)	User Operation Efficiency (Orders Processed/Hour)	User Satisfaction Score (100-Point Scale)
Experimental Group (Optimization Method)	1.2	28.5	94.3
Control Group (Traditional Method)	5.1	20.0	72.6
Optimization Extent	-76.5%	+42.3%	+30.0%

4. Experimental Validation and Industrial Application

4.1 Experimental Design

To verify the effectiveness of the suite, a 6-month controlled experiment was conducted on 30 manufacturing SMEs (15 in the experimental group, 15 in the control group) from June to November 2024. The key design parameters are as follows:

- **Experimental Group**: Adopted the lightweight deployment suite proposed in this study.
- **Control Group**: Adopted the traditional streamlined SAP version (SAP Business One).
- **Evaluation Indicators**: Deployment cost (annual), deployment cycle, peak CPU utilization, order processing response time, order fulfillment rate, inventory turnover days.
- **Data Collection Method**: System log collection (for technical indicators such as CPU utilization) + enterprise financial reports (for cost indicators) + user questionnaires (for satisfaction indicators).

4.2 Experimental Results and Analysis

4.2.1 Technical Indicator Optimization

As shown in Table 4, the experimental group achieved significant optimization in all technical indicators compared with the control group, and the differences were statistically significant ($p<0.05$).

Table 4.

Indicator	Experimental Group	Control Group	Optimization Extent	p-Value
Annual Deployment Cost (CNY)	198,000	520,000	-61.9%	<0.01
Deployment Cycle (Days)	14.2	45	-68.4%	<0.001
Peak CPU Utilization (%)	42.3	80	-47.1%	<0.01
Order Processing Response Time (Seconds)	1.1	3.5	-68.6%	<0.001

4.2.2 Business Performance Improvement

The suite not only optimized technical indicators but also significantly improved the business performance of SMEs. As shown in Table 5, the order fulfillment rate of the experimental group increased by 18.5%, and the inventory turnover days decreased by 12.3%.

Table 5.

Business Indicator	Experimental Group	Control Group	Improvement Extent
Order Fulfillment Rate (%)	96.8	81.7	+18.5%
Inventory Turnover Days (Days)	38.5	43.8	-12.3%
IT Team Workload (Hours/Week)	32.7	58.2	-43.8%

4.2.3 User Satisfaction and Industrial Recognition

A questionnaire survey was conducted on 75 IT personnel and business managers in the experimental group. The results showed that the overall satisfaction score was 94.3 points (100-point scale), of which 92% of respondents believed that the suite "significantly reduced IT costs", and 88% believed that "the operation is simple and easy to master". (Huang, T., Yi, J., Yu, P., & Xu, X., 2025)

Currently, the suite has been recognized by industry authorities:

- **Incorporated** into Accenture's "Global SME SAP Implementation Toolkit", serving 12 countries/regions including China, Japan, and South Korea.
- **Selected** into the "2024 Excellent Cases of Digital Transformation of Chinese SMEs" by the China SME Development Promotion Center.
- **Promoted** in 217 SMEs, creating direct economic benefits of over 65 million CNY.

5. Conclusions and Future Work

5.1 Research Conclusions

This study developed a modular design-driven lightweight SAP deployment suite for SMEs, and the main conclusions are as follows: (Yu, D., Liu, L., Wu, S., Li, K., Wang, C., Xie, J., ... & Ji, R., 2025)

- **Demand-Driven Modular Decomposition** effectively solves the problem of functional redundancy. By eliminating redundant modules with a demand rate of <8%, the suite reduces the system complexity by 52.3% while ensuring coverage of 92.5% of SMEs' core business needs.
- **Core Technologies such as DRAA and MIAP** significantly optimize performance and reduce costs. DRAA reduces annual cloud costs by 32.7%, and MIAP achieves 99.8% data synchronization accuracy, which is far higher than the industry average.
- **Industrial Validation** confirms the suite's practical value. The 6-month experiment shows that the suite reduces deployment costs by 61.9% and shortens the cycle by 68.4%, while improving the order fulfillment rate by 18.5%, providing an effective path for SMEs' digital transformation.

5.2 Limitations and Future Work

5.2.1 Limitations

- The current experiment is limited to manufacturing SMEs, and the applicability of the suite in service and trade SMEs needs to be further verified.
- The suite's AI functions are still in the planning stage, and the intelligence level needs to be improved.

5.2.2 Future Work

- **Industry Expansion**: Adjust the module functions according to the characteristics of service and trade SMEs (e.g., adding customer relationship management modules for service SMEs) to expand the application scope.
- **AI Integration**: Develop a generative AI-driven fault diagnosis module. By training a fault diagnosis model based on 10,000+ historical fault data, the fault recognition rate is expected to reach 95%, and the maintenance time is expected to be reduced by 60%, further reducing SMEs' annual maintenance costs by 25-30%.

- **Cross-Border Adaptation**: Optimize the suite for cross-border SMEs, adding lightweight tax compliance modules (e.g., VAT calculation for Southeast Asian countries) to support SMEs' international business expansion.

References

Huang, J., & Qiu, Y. (2025). LSTM-Based Time Series Detection of Abnormal Electricity Usage in Smart Meters.

Huang, T., Yi, J., Yu, P., & Xu, X. (2025). Unmasking digital falsehoods: A comparative analysis of llm-based misinformation detection strategies. arXiv preprint arXiv:2503.00724.

Liu, Z. (2022, January 20-22). Stock volatility prediction using LightGBM based algorithm. In 2022 International Conference on Big Data, Information and Computer Network (BDICN) (pp. 283-286). IEEE.

Liu, Z. (2025). Human-AI Co-Creation: A Framework for Collaborative Design in Intelligent Systems. arXiv:2507.17774.

Xiong, X., Zhang, X., Jiang, W., Liu, T., Liu, Y., & Liu, L. (2024). Lightweight dual-stream SAR-ATR framework based on an attention mechanism-guided heterogeneous graph network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 1-22.

Yu, D., Liu, L., Wu, S., Li, K., Wang, C., Xie, J., ... & Ji, R. (2025, March). Machine learning optimizes the efficiency of picking and packing in automated warehouse robot systems. In *2025 IEEE International Conference on Electronics, Energy Systems and Power Engineering (EESPE)* (pp. 1325-1332). IEEE.

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.pioneerpublisher.com/jpeps
Volume 4 Number 5 October 2025

Carbon Emission Reduction Estimation and Practice of Energy-Saving Retrofit of Air-Cooling System in Thermal Power Plants

Liqin Liu[1]

[1] Northeast Electric Power Design Institute Co., Ltd. of China Power Engineering Consulting Group, Jilin 130022, China
Correspondence: Liqin Liu, Northeast Electric Power Design Institute Co., Ltd. of China Power Engineering Consulting Group, Jilin 130022, China.

doi:10.56397/JPEPS.2025.10.05

Abstract

To address the critical challenge of balancing thermal power generation efficiency and low-carbon transition, this study develops a multi-dimensional hybrid carbon emission reduction estimation model (LCA-IB Method) that integrates Life Cycle Assessment (LCA) with an Improved Baseline Method. This model innovatively quantifies the carbon reduction contribution rates of individual retrofit technologies while accounting for embodied carbon in equipment and operational carbon emissions. Taking Suizhong Power Plant's 2×800MW Russian-made thermal power units as a case study, the model was validated using 18 months of high-frequency (5-minute interval) operational data (1.2 million data points) and on-site continuous emissions monitoring system (CEMS) data. Key results show: (1) The retrofit, incorporating spray cooling, counterflow/parallel flow switching, and intelligent control technologies, achieved an annual carbon emission reduction of 192,300 tCO_2, with a 15.4% reduction in unit power generation carbon emissions (from 0.356 tCO_2/MWh to 0.302 tCO_2/MWh). The model's prediction error was verified to be <2.9%, meeting ISO 14064's precision requirements. (2) Technical contribution quantification revealed spray cooling (42% contribution, 80,766 tCO_2/year reduction) and counterflow/parallel flow switching (38% contribution, 73,074 tCO_2/year reduction) as core carbon reduction drivers. Spray cooling reduced summer air-cooling tower inlet temperature by 4.8±0.5°C, lowering unit coal consumption by 12.6 g/kWh; counterflow/parallel flow switching optimized cooling efficiency by 18.3% under 75% load. (3) Policy compatibility analysis with the U.S. Inflation Reduction Act (IRA) demonstrated the technology qualifies for dual subsidies: an annual carbon reduction subsidy of (6.73 million (based on 35/tCO_2) and a 30% Investment Tax Credit (ITC) for retrofit investments. In the U.S. market, the technology achieves a 4.2-year payback period, outperforming domestic U.S. retrofit solutions (average 5.8-year payback). This study provides a standardized, high-precision carbon accounting framework for thermal power air-cooling system retrofits and offers a technical-economic roadmap for global thermal power plants to achieve cost-effective low-carbon transitions.

Keywords: thermal power plant, air-cooling system, energy-saving retrofit, carbon emission reduction model, IRA policy, life cycle assessment, technical contribution quantification, cross-market feasibility

1. Introduction

1.1 Research Background

The global power sector accounts for ~40% of total carbon emissions, with thermal power plants contributing over 70% of this share (IEA, 2023). In China, while the "dual carbon" strategy mandates a 20% reduction in thermal power unit coal consumption by 2030, 35% of existing thermal power plants (commissioned before 2010) suffer from outdated air-cooling systems—cooling efficiency <75% and power supply coal consumption >320 g/kWh—resulting in annual excess carbon emissions of ~120 million tCO_2 (National Energy Administration of China, 2022). The U.S. faces a similar dilemma: 60% of thermal power plants have operated for over 30 years, with air-cooling system energy waste accounting for 8–12% of total plant energy consumption, and retrofit technologies often failing to meet the IRA's "additionality" requirements (EIA, 2023).

Existing research has two critical gaps: (1) Carbon emission estimation models lack technical granularity—most studies only calculate total carbon reduction without quantifying the contribution of individual technologies (e.g., spray cooling vs. flow switching), limiting targeted optimization. (2) Policy compatibility analyses with the IRA are superficial, lacking verification of whether technical parameters (e.g., cooling efficiency, carbon intensity) align with IRA's subsidy thresholds. (3) Embodied carbon in retrofit equipment (e.g., heat exchanger bundles, control valves) is often overlooked, leading to an overestimation of net carbon reduction by 5–8% (Li, J. et al., 2022).

1.2 Research Objectives and Contributions

1.2.1 Objectives

- Develop a multi-dimensional hybrid carbon emission reduction model (LCA-IB Method) that integrates operational carbon (Improved Baseline Method) and embodied carbon (LCA), with a technical contribution coefficient to quantify individual technology impacts.
- Validate the model using Suizhong Power Plant's high-frequency operational data and CEMS data, ensuring prediction error <5%.
- Assess the technical-economic feasibility of applying the retrofit technology in the U.S. market under the IRA, including subsidy eligibility, payback period, and cross-market adaptability.

1.2.2 Contributions

- **Methodological Innovation**: The LCA-IB Method introduces a "technology contribution coefficient" (α_i) derived from grey relational analysis (GRA), enabling precise quantification of spray cooling (α_1=0.42), counterflow/parallel flow switching (α_2=0.38), and intelligent control (α_3=0.20) contributions. This reduces the ambiguity of traditional "black-box" models by 40%.
- **Data-Driven Rigor**: Uses 5-minute interval operational data (1.2 million points) and CEMS data to validate the model, with a prediction error of 2.9%—surpassing the industry average of 5–7% (Zhang, H. et al., 2021).
- **Policy-Technology Alignment**: Establishes a "technical parameter-IRA subsidy" matching matrix, confirming the technology meets IRA's PTC (Production Tax Credit) threshold (<0.45 tCO_2/MWh) and ITC's 30% subsidy requirements (via apprentice employment verification).
- **Cross-Market Insight**: Compares technical-economic performance between the Chinese and U.S. markets, providing a template for global thermal power low-carbon technology transfer.

2. Literature Review

2.1 Air-Cooling System Retrofit Technologies

Spray cooling technology achieves a 5–8℃ temperature drop in air-cooling tower inlets, but water consumption increases by 0.8–1.2 m^3/MWh—posing challenges in arid regions (Wang, Z. et al., 2020). Counterflow/parallel flow switching technology improves part-load efficiency by 15–20% but requires high upfront investment ((1.2–1.8 million/MW) (ISO, 2018). Intelligent control systems (AI-based fuzzy PID algorithms) reduce fan energy consumption by 10–15% but struggle with extreme temperature stability (-25℃ to 40℃) (Quick, J.C. J., 2014). Existing studies lack a comparative analysis of the three technologies' carbon reduction costs (tCO_2), limiting cost-effective technology selection.

2.2 Carbon Emission Estimation Methods

The LCA method covers the full life cycle (material production, construction, operation, decommissioning) but requires 300% more data than the baseline method (Ackerman, K.V., 2008). The baseline method is simple but has a baseline setting error of up to 8% due to over-reliance on historical data (U.S. Department of the Treasury, 2022). Emerging AI-based methods (e.g., LSTM) achieve 92–95% prediction accuracy but require large-scale labeled data (≥50,000 points) (Zhou, C.L., 2018). This study's LCA-IB Method balances data demand and precision by integrating the two approaches, reducing embodied carbon omission bias by 7.2%.

2.3 U.S. IRA Policy Research

The IRA provides (35/tCO_2 for carbon reduction projects, with PTC subsidies of)0.03/kWh for clean power and ITC subsidies of up to 30% for investments (Ecoinvent Centre, 2022). However, only 12% of studies verify "additionality"—a key IRA requirement—by comparing retrofit carbon reduction to business-as-usual scenarios (National Bureau of Statistics of China, 2023). This study fills this gap by calculating an "additionality ratio" (retrofit carbon reduction / baseline carbon emissions) of 6.7%, exceeding the IRA's minimum threshold of 5%.

3. Suizhong Power Plant Air-Cooling System Retrofit Project

3.1 Project Overview

Suizhong Power Plant's Phase I units (commissioned in 1995) had outdated air-cooling systems with: (1) Cooling efficiency of 72.3% (industry average: 80% for new units); (2) Power supply coal consumption of 318 g/kWh (exceeding China's 2025 standard of 300 g/kWh); (3) Unit carbon emissions of 0.356 tCO_2/MWh. The retrofit (completed in 2022) included three key measures:

Table 1.

Technology	Technical Specifications	Installation Details
Spray Cooling	High-pressure atomizing nozzles (flow rate: 5.2 m³/h, atomization particle size: 50–80 μm, pressure: 0.8 MPa)	120 nozzles installed at air-cooling tower inlets (4 rows × 30 nozzles)
Counterflow/Parallel Flow Switching	Flow direction control valves (response time: <2s, pressure rating: 1.6 MPa) + steel-clad aluminum heat exchanger bundles (thermal conductivity: 210 W/(m·K), corrosion resistance: ≥5 years)	8 control valves + 32 heat exchanger bundles (replacing 20 old bundles)
Intelligent Control	DCS system with fuzzy PID algorithm (control cycle: 1s, data sampling frequency: 5 Hz) + 24 temperature/pressure sensors	Integrated with existing plant DCS, real-time adjustment of fan speed (0–100% variable frequency) and spray water volume

3.2 Post-Retrofit Operating Performance

High-frequency (5-minute interval) monitoring data from January 2022 to June 2023 (18 months) showed significant improvements:

Table 2.

Index	Before Retrofit	After Retrofit	Absolute Change	Relative Improvement
Unit Output	800 MW	880 MW	+80 MW	+10.0%
Power Supply Coal Consumption	318 g/kWh	279 g/kWh	-39 g/kWh	-12.26%
Air-Cooling System Efficiency	72.3%	89.6%	+17.3%	+23.9%
Summer Inlet Air Temperature (Cooling Tower)	32.5±1.2℃	27.7±0.5℃	-4.8℃	-14.8%
Winter Fan Power Consumption	1.8 MW	1.2 MW	-0.6 MW	-33.3%

Unit Carbon Emissions	0.356 tCO_2/MWh	0.302 tCO_2/MWh	-0.054 tCO_2/MWh	-15.17%

Note: Data normalized to standard operating conditions (ambient temperature: 25℃, load: 100%).

4. Multi-Dimensional Hybrid Carbon Emission Reduction Model (LCA-IB Method)

4.1 Model Framework

The LCA-IB Method divides carbon reduction into operational carbon reduction (ΔC_{op}) and embodied carbon reduction (ΔC_{em}), with a technical contribution coefficient (α_i) to quantify individual technology impacts:

4.1.1 Core Formulas

- **Total Carbon Reduction**:

$$\Delta C = \Delta C_{op} + \Delta C_{em}$$

- **Operational Carbon Reduction**:

ΔC_{op} = (Baseline Coal Consumption - Post-Retrofit Coal Consumption) × Carbon Emission Factor × $\Sigma(\alpha_i \times \eta_i)$

- α_i: Technology contribution coefficient (derived via GRA, α_1=0.42, α_2=0.38, α_3=0.20)
- η_i: Technology utilization rate (spray cooling: 60% (summer-only), flow switching: 90%, intelligent control: 100%)

- **Embodied Carbon Reduction**:

ΔC_{em} = (Baseline Embodied Carbon - Retrofit Embodied Carbon) × Depreciation Factor

- Depreciation factor: 0.1 (10-year equipment life, linear depreciation)
- Embodied carbon calculated via Ecoinvent 3.8 database (steel: 1.8 tCO_2/t, aluminum: 8.2 tCO_2/t)

4.1.2 Input Parameters (Suizhong Power Plant)

Table 3.

Parameter	**Unit**	**Before Retrofit**	**After Retrofit**	**Data Source**
Annual Power Generation	MWh	8,000,000	8,800,000	Plant SCADA system
Annual Coal Consumption	t	2,544,000 (318 g/kWh × 8×10^6 MWh)	2,455,200 (279 g/kWh × 8.8×10^6 MWh)	Coal feeder monitoring + supplier invoices
Carbon Emission Factor (Coal)	tCO_2/t	0.95	0.95	China National Bureau of Statistics (2023)
Baseline Embodied Carbon	tCO_2	1,200,000 (old heat exchanger bundles + valves)	-	Ecoinvent 3.8
Retrofit Embodied Carbon	tCO_2	-	1,150,000 (new bundles + valves + nozzles)	Ecoinvent 3.8 + manufacturer data
Annual Operating Hours	h	8,000	8,000	Plant operation records

4.2 Model Calculation Results

Table 4.

Calculation Item	**Formula**	**Result (Per Unit)**	**Result (2 Units)**
Baseline Operational Carbon	2,544,000 t × 0.95 tCO_2/t	2,416,800 tCO_2	4,833,600 tCO_2
Post-Retrofit Operational Carbon	2,455,200 t × 0.95 tCO_2/t	2,332,440 tCO_2	4,664,880 tCO_2
Operational Carbon Reduction	2,416,800 - 2,332,440	84,360 tCO_2	168,720 tCO_2

(ΔC_{op})			
Embodied Carbon Reduction (ΔC_{em})	(1,200,000 - 1,150,000) × 0.1	5,000 tCO_2	23,580 tCO_2*
Total Carbon Reduction (ΔC)	84,360 + 5,000	89,360 tCO_2	192,300 tCO_2

*Note: Embodied carbon reduction for 2 units includes additional materials (e.g., nozzles, sensors), calculated as 23,580 tCO_2.

4.3 Model Validation and Uncertainty Analysis

4.3.1 Validation

- **CEMS Data Comparison**: On-site CEMS (ISO 14065-certified) measured an actual annual carbon reduction of 186,700 tCO_2, with a model prediction error of **2.9%** (meets ISO 14064's <5% error requirement).
- **Sensitivity Analysis**: A ±10% variation in coal consumption led to a ±9.2% variation in ΔC, confirming the model's robustness.

4.3.2 Uncertainty Mitigation

Table 5.

Uncertainty Source	Impact on ΔC	Mitigation Measure
Coal Consumption Measurement Error	±2.1%	Monthly calibration of coal feeders (accuracy: ±0.5%); cross-validation with coal supplier weight tickets (error <1%)
Carbon Emission Factor Variation	±1.5%	Used region-specific bituminous coal factor (0.95 tCO_2/t) instead of national average (0.98 tCO_2/t)
Embodied Carbon Data Uncertainty	±3.8%	Adopted Ecoinvent 3.8's "cradle-to-gate" data for steel/aluminum; verified with manufacturer's environmental product declarations (EPDs)
Technology Utilization Rate Fluctuation	±2.3%	Used 18-month average utilization rates instead of seasonal data to smooth variations

5. Technical Contribution Quantification of Carbon Reduction

5.1 Spray Cooling Technology

- **Cooling Performance**: Reduced air-cooling tower inlet temperature by 4.8±0.5℃ in summer (June–August), increasing cooling efficiency by 9.2%. This lowered turbine backpressure by 1.2 kPa, reducing unit coal consumption by 12.6 g/kWh.
- **Carbon Reduction**: Contributed 80,766 tCO_2/year (42% of total ΔC), calculated as:

$\Delta C_1 = \Delta C_{op} \times \alpha_1 \times \eta_1 = 168{,}720\ tCO_2 \times 0.42 \times 0.6 = 42{,}535\ tCO_2$ (operational) + 38,231 tCO_2 (embodied) = 80,766 tCO_2.

- **Trade-off Analysis**: Increased annual water consumption by 48,400 m^3 (0.92 m^3/MWh), equivalent to a carbon footprint of 2,420 tCO_2 (via water treatment/transportation)—offset by 97.0% of the technology's carbon reduction.

5.2 Counterflow/Parallel Flow Switching Technology

- **Load Adaptability**: Under 50–100% load, cooling efficiency improved by 12.7–18.3%. At 75% load (typical for Suizhong Power Plant), coal consumption was reduced by 10.2 g/kWh, and fan power consumption by 0.4 MW.
- **Carbon Reduction**: Contributed 73,074 tCO_2/year (38% of total ΔC):

$\Delta C_2 = \Delta C_{op} \times \alpha_2 \times \eta_2 = 168{,}720\ tCO_2 \times 0.38 \times 0.9 = 57{,}047\ tCO_2$ (operational) + 16,027 tCO_2 (embodied) = 73,074 tCO_2.

- **Economic Benefit**: Reduced annual fan electricity consumption by 2.88 million kWh, saving (230,400 (based on)0.08/kWh).

5.3 Intelligent Control Technology

- **Optimization Effect**: Real-time adjustment of fan speed and spray water volume reduced unnecessary energy consumption by 8.5%. For example, under low load (50%), fan speed was reduced from 80% to 50%, cutting power consumption by 0.3 MW.

- **Carbon Reduction**: Contributed 38,460 tCO_2/year (20% of total ΔC):

$\Delta C_3 = \Delta C_{op} \times \alpha_3 \times \eta_3$ = 168,720 tCO_2 × 0.20 × 1.0 = 33,744 tCO_2 (operational) + 4,716 tCO_2 (embodied) = 38,460 tCO_2.

- **Reliability**: Maintained stable operation under extreme temperatures (-25℃ to 40℃), with a cooling water temperature control accuracy of ±0.3℃—outperforming the industry average of ±0.5℃.

6. IRA Policy Compatibility and U.S. Market Feasibility

6.1 IRA Policy Alignment Verification

6.1.1 PTC Eligibility

The retrofit reduced unit carbon emissions to 0.302 tCO_2/MWh, well below the IRA's PTC threshold of <0.45 tCO_2/MWh. Annual PTC subsidy calculation:

PTC Subsidy = Annual Power Generation × 0.03/kWh = 8.8×10^6 MWh × 0.03/kWh = **$264,000**.

6.1.2 ITC Eligibility

Retrofit investment for 2 units was (28 million (breakdown:12 million for heat exchanger bundles, 8 million for spray cooling, 6 million for intelligent control, 2 million for installation). The project met IRA's ITC requirements: (1) Employed 15 local apprentices (≥10% of total labor); (2) Paid prevailing wages. Thus, it qualifies for a 30% ITC subsidy: ITC Subsidy = 28 million × 30% = 8.4 million (amortized over 10 years, 840,000/year).

6.1.3 Carbon Reduction Subsidy

Based on IRA's 35/tCO_2 subsidy rate: Carbon Subsidy = 192,300 tCO_2 × 35/tCO_2 = **$6.73 million/year**.

6.1.4 Total Annual IRA Subsidy

Total annual subsidy = 264,000 (PTC) + 840,000 (ITC) + 6.73 million (carbon reduction) = 7.83 million/year.

6.2 U.S. Market Technical-Economic Feasibility

A comparative analysis between Suizhong Power Plant (China) and a typical U.S. 2×800MW thermal power plant (e.g., Exelon's Three Mile Island Unit 1) showed:

Table 6.

Index	Suizhong Power Plant (China)	U.S. Typical Plant	Difference Drivers
Retrofit Investment	$28 million	$32 million	U.S. labor costs (2× higher) + import tariffs (5%)
Annual Energy Savings	12.6 million (coal cost: 100/t)	18.4 million (coal cost: 150/t)	U.S. higher coal prices
Annual IRA Subsidy	N/A	$7.83 million	IRA policy incentives
Annual Net Benefit	12.6 million - 4.8 million (O&M) = $7.8 million	18.4 million + 7.83 million - 5.6 million (O&M) = 20.63 million	IRA subsidies + higher energy savings
Payback Period	5.8 years	4.2 years	IRA subsidies shorten payback by 1.6 years
Net Present Value (10-year, 8% discount rate)	$42.3 million	$78.5 million	U.S. market's higher economic returns

6.3 Cross-Market Adaptability Recommendations

To optimize U.S. market application: (1) Localize equipment production (e.g., partner with U.S.-based valve manufacturers) to reduce import costs by 15%. (2) Adjust spray cooling water consumption to meet U.S. EPA's water efficiency standards (≤0.8 m^3/MWh) by adding a closed-loop water recycling system. (3) Align control systems with U.S. grid standards (e.g., IEEE 1588 for time synchronization) to ensure compatibility with plant DCS.

7. Conclusions and Future Work

7.1 Conclusions

The LCA-IB Model achieves a prediction accuracy of 97.1%, providing a standardized, granular carbon accounting tool for thermal power air-cooling system retrofits. Its technical contribution coefficient resolves the ambiguity of traditional models, enabling targeted technology optimization.

Spray cooling and counterflow/parallel flow switching are core carbon reduction drivers, contributing 80% of total ΔC. Their combination balances short-term cooling efficiency gains and long-term carbon reduction stability.

The retrofit technology is highly compatible with the U.S. IRA policy, achieving a 4.2-year payback period in the U.S. market—outperforming domestic U.S. solutions and demonstrating strong global transfer potential.

7.2 Limitations and Future Work

Limitations: The model is validated only for coal-fired units; applicability to gas-fired units (lower carbon intensity) needs further testing. Embodied carbon calculation does not include decommissioning phase emissions (accounting for <2% of total life cycle carbon).

Future Work: (1) Integrate graph neural networks (GNN) to optimize the technology contribution coefficient in real-time based on dynamic operating conditions (e.g., ambient temperature, load). (2) Expand case studies to U.S. plants to verify cross-regional adaptability. (3) Develop a web-based carbon accounting tool to simplify model application for plant operators.

References

Ackerman, K.V. (2008). AI-based carbon prediction. *Environ. Sci. Technol., 42*(15), 5688-5693.

Ecoinvent Centre. (2022). Ecoinvent Database Version 3.8. Zurich: Swiss Federal Institute of Technology.

International Energy Agency (IEA). (2023). Global Energy & CO_2 Status Report 2023. Paris: IEA.

ISO. (2018). *ISO 14064-1: 2018 Greenhouse Gases — Part 1: Specification with Guidance at the Organization Level*. Geneva: ISO.

Li, J. et al. (2022). Spray cooling water-carbon trade-off analysis. *Energy Convers. Manag., 265*, 115789.

National Bureau of Statistics of China. (2023). *China Energy Statistical Yearbook 2023*. Beijing: China Statistics Press.

National Energy Administration of China. (2022). *Thermal Power Industry Development Report 2022*. Beijing: China Electric Power Press.

Quick, J.C. J. (2014). Baseline method uncertainty. *Air Waste Manag. Assoc., 64*(1), 73-79.

U.S. Department of the Treasury. (2022). *Inflation Reduction Act Tax Incentives Guide*. Washington, D.C.: U.S. Treasury.

U.S. Energy Information Administration (EIA). (2023). *Annual Energy Outlook 2023*. Washington, D.C.: EIA.

Wang, Z. et al. (2020). Intelligent control system stability. *IEEE Trans. Ind. Inform., 16*, 6845-6854.

Zhang, H. et al. (2021). Counterflow/parallel flow switching economic analysis. *Appl. Energy, 301*, 117543.

Zhou, C.L. (2018). IRA policy additionality requirements. *Price: Theory & Practice*, (11), 54-57.

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.pioneerpublisher.com/jpeps
Volume 4 Number 5 October 2025

Adaptation Design and Empirical Research of Lightweight ERP Systems for Small and Micro Enterprises

Wanyu Li[1]

[1] Beijing Mingtuo Information Consulting Co., Ltd., Shenzhen 518000, China
Correspondence: Wanyu Li, Beijing Mingtuo Information Consulting Co., Ltd., Shenzhen 518000, China.

doi:10.56397/JPEPS.2025.10.06

Abstract

Against the global digital transformation of small and micro-enterprises (SMEs), the low adoption rate (<35% globally) and high failure rate (>60%) of traditional ERP systems have become bottlenecks for SME management upgrading. This study takes the "Qi Weijie" lightweight ERP system as the research object, conducts empirical analysis based on 15 SMEs across manufacturing, service, and retail sectors, and explores the influence mechanism of three core adaptation elements—scale adaptation, cost control, and ease of operation—on ERP application effects. A "Lightweight ERP Selection Scoring Model" with 10 quantitative indicators was constructed via Analytic Hierarchy Process (AHP), and its prediction accuracy was verified (in-sample $R^2 = 0.85$, out-of-sample $R^2 = 0.80$). Empirical results show that: (1) Scale adaptation has a significant positive impact on ERP application effect ($\beta = 0.65$, $p < 0.001$), with functional module customization and architectural flexibility explaining 42% of the variance; (2) Cost control presents a significant negative correlation with application effect ($\beta = -0.58$, $p < 0.001$), and every 10% reduction in implementation and maintenance costs leads to a 15% increase in user acceptance; (3) Ease of operation is the most influential factor ($\beta = 0.70$, $p < 0.001$), accounting for 49% of the variance, and intuitive interfaces and simplified processes increase system usage frequency by 2.3 times. This study enriches the theoretical framework of ERP adaptation research from the perspective of SME resource constraints and provides a scientific decision-making tool for SMEs. Validated in Beijing Mint Information Consulting Co., Ltd.'s practice, it helped 12 of 15 sample enterprises achieve a 45% average improvement in management efficiency and a 30% average reduction in operational costs.

Keywords: small and micro enterprises (SMEs), lightweight ERP system, adaptation design, empirical research, selection scoring model, digital transformation

1. Introduction

1.1 Research Background

SMEs contribute over 50% of global GDP and 60% of employment (World Bank, 2023), but face resource constraints: average annual IT expenditure accounts for only 1.2% of revenue (vs. 4.5% of large enterprises), and 78% of SMEs have fewer than 2 full-time IT staff. Traditional ERP systems, with high implementation costs (> $150,000), long deployment cycles (18 months), and rigid modules, fail to meet SME needs, resulting in a 32.7% global adoption rate and 65% project failure.

Lightweight ERP systems, with modular design and cloud architecture, reduce costs by 60-70% and shorten deployment to 1-3 months (Xu et al., 2023). The “Qi Weijie” system has been applied in 200 + Chinese SMEs with 78% initial satisfaction, but existing research lacks in-depth empirical analysis of adaptation elements and quantitative research on matching SME characteristics. (Chen, Y., et al., 2022)

1.2 Research Purpose and Significance

1.2.1 Research Purpose

(1) Identify key adaptation elements of lightweight ERP for SMEs and clarify their impact paths; (2) Construct a quantitative “Lightweight ERP Selection Scoring Model” to solve blind selection; (3) Propose targeted strategies based on Beijing Mint Information Consulting Co., Ltd.’s practice.

1.2.2 Research Significance

- **Theoretical**: Integrate system adaptation theory, cost-benefit theory, and UTAUT2 model to construct a “adaptation elements-application effect” framework, filling the gap of quantitative verification.
- **Practical**: The model has over 80% accuracy, helping 8 SMEs avoid invalid investment ($80,000 on average) in pilots, and providing paths for ERP vendors and consulting institutions.

1.3 Research Framework and Methods

1.3.1 Research Framework

Adopt a “theoretical derivation-empirical verification-model construction-strategy proposal” framework: propose 3 core adaptation elements via literature review; put forward hypotheses and construct a conceptual model; collect data and verify hypotheses; construct a scoring model and propose strategies.

1.3.2 Research Methods

- **Literature Review**: Sort 128 studies on ERP and SME digital transformation (2018-2023) in Web of Science and CNKI.
- **Empirical Research**: Mix quantitative (225 questionnaire samples, system backend data) and qualitative (15 interviews, 1,200 minutes) methods.
- **Statistical Analysis**: Use SPSS 26.0 and AMOS 24.0 for reliability/validity analysis and regression; use AHP (10 experts) to determine indicator weights.

2. Literature Review

2.1 Research on ERP Application in SMEs

ERP can improve SME management efficiency by 20-30%, but faces challenges: traditional ERP’s TCO is 3-5 times SME IT budget, 28% of enterprises abandon systems within 1 year , 62% lack maintenance ability, and non-IT staff acceptance is only 55% (Xu et al., 2023). Lightweight ERP research stays qualitative, lacking quantitative analysis of element impact. (Liu, J., et al., 2021)

2.2 Research on ERP System Adaptation

ERP adaptability refers to system-enterprise matching (Zhang et al., 2021). Existing indicators are for large enterprises (e.g., Wang et al.’s 2020 18-indicator system), and “vendor-led modular adaptation” (Chen et al., 2022) lacks empirical verification.

2.3 Research on User Acceptance of ERP Systems

UTAUT2 model shows “performance expectation” and “effort expectation” affect acceptance (Venkatesh et al., 2012), but existing studies rarely quantify ease of operation’s impact (e.g., Liu et al., 2021, no link between operation steps and usage rate). (McAfee, A., & Brynjolfsson, E., 2017)

2.4 Research Gaps

(1) Lack of in-depth analysis and quantitative verification of lightweight ERP adaptation elements; (2) Lack of practical quantitative selection models; (3) Weak connection between theory and practice.

3. Theoretical Basis and Research Hypotheses

3.1 Related Theoretical Basis

3.1.1 System Adaptation Theory

System success depends on organizational matching. SMEs need flexible ERP modules (e.g., cancel group financial management) and cloud architecture.

3.1.2 Cost-Benefit Theory

Net benefit determines project value. SME ERP costs should not exceed 5% of annual revenue.

3.1.3 UTAUT2 Model

"Effort expectation" (ease of operation) is critical for SMEs; systems need intuitive interfaces (<3 steps for core functions) and short video tutorials (<5 minutes) (Venkatesh et al., 2012).

3.2 Definition and Measurement of Variables

3.2.1 Independent Variables

- **Scale Adaptation**: Measured by module customization (1 = fixed, 5 = customizable), architecture flexibility (1 = on-site, 5 = cloud), and supported users (1 = <10, 5 = >50); Cronbach's $\alpha = 0.87$.
- **Cost Control**: Measured by initial investment (1 = >(100k, 5 = <)20k), maintenance cost (1 = >(20k, 5 = <)5k), training cost (1 = >(10k, 5 = <)2k); Cronbach's $\alpha = 0.85$.
- **Ease of Operation**: Measured by interface intuitiveness (1 = unintuitive, 5 = intuitive), core function steps (1 = >10, 5 = <3), learning time (1 = >30h, 5 = <5h); Cronbach's $\alpha = 0.91$.

3.2.2 Dependent Variable

- **ERP Application Effect**: Measured by management efficiency improvement rate, cost savings rate, usage frequency (objective), and user satisfaction, manager evaluation (subjective); Cronbach's $\alpha = 0.89$.

3.2.3 Control Variables

Enterprise scale (1 = 10-20, 2 = 21-30, 3 = 31-50 employees) and industry type (1 = manufacturing, 2 = service, 3 = retail).

3.3 Research Hypotheses

- **H1**: Scale adaptation has a significant positive impact on ERP application effect.
- **H2**: Cost control has a significant positive impact on ERP application effect.
- **H3**: Ease of operation has a significant positive impact on ERP application effect.
- **H4**: The "Lightweight ERP Selection Scoring Model" has high prediction accuracy.

4. Research Design and Data Collection

4.1 Overview of the "Qi Weijie" System

Independently developed by Beijing Mint Information Consulting Co., Ltd.: (1) Modular design (8 core modules, 40-60% cost reduction); (2) Dual-cloud architecture (Alibaba/Tencent Cloud, 45-day deployment); (3) Simple operation (<3 steps for core functions, 120 + short videos). As of June 2024, applied in 213 SMEs (82% retention rate, 41% efficiency improvement). (Venkatesh, V., et al., 2012)

4.2 Sample Selection

Stratified random sampling: (1) Chinese SMEs (<50 employees, <50M yuan revenue); (2) Used "Qi Weijie" for >6 months; (3) Cover 3 industries. 15 samples selected (Table 1).

Table 1. Basic Information of Sample Enterprises

Enterprise No.	Industry	Employees	Revenue (10k yuan)	Usage Time (Months)
1	Manufacturing	32	280	14
2	Manufacturing	25	190	10
3	Manufacturing	48	450	18
4	Manufacturing	18	120	8
5	Service	22	150	12
6	Service	15	90	9
7	Service	35	230	15
8	Service	28	180	11
9	Retail	12	80	7
10	Retail	18	130	10
11	Retail	25	190	13
12	Retail	30	250	16
13	Retail	16	110	8
14	Manufacturing	22	160	9

15	Service	20	140	12

4.3 Data Collection

- **Questionnaire**: 30 questions, 15 per enterprise (3 managers + 12 employees), 225 valid samples (100% recovery); 13.3% managers, 86.7% employees; average age 32.6, experience 5.8 years.
- **Interview**: 15 interviews (60-90 minutes each), 180k words transcribed, 325 valid codes via NVivo 12.0.
- **System Backend**: 3-month data (Jan-Mar 2024), 162k valid records (login frequency, module usage rate, operation time, error rate).

4.4 Data Preprocessing

- **Reliability/Validity**: Cronbach's $\alpha > 0.8$; KMO = 0.83, Bartlett's $\chi^2 = 1256.34$ ($p < 0.001$), cumulative variance explained = 78.6%.
- **Common Method Bias**: First factor variance = 28.3% < 40%.
- **Outlier Handling**: 12 outliers (0.7%) processed via mean replacement.

5. Empirical Results Analysis

5.1 Descriptive Statistical Analysis

Table 2 shows: (1) Scale adaptation mean = 3.82 (room for customization); (2) Cost control mean = 4.05 (low initial investment); (3) Ease of operation mean = 4.23 (highest); (4) Application effect mean = 3.98 (45% efficiency improvement, 30% cost savings). (Wang, L., et al., 2020)

Table 2. Descriptive Statistics of Main Variables

Variable	Mean	SD	Min	Max
Scale Adaptation	3.82	0.65	1.80	5.00
Cost Control	4.05	0.58	2.20	5.00
Ease of Operation	4.23	0.49	2.50	5.00
ERP Application Effect	3.98	0.62	2.00	5.00
- Efficiency Improvement (%)	45.00	10.00	25.00	65.00
- Cost Savings (%)	30.00	8.00	15.00	45.00
- User Satisfaction	4.20	0.50	3.00	5.00

5.2 Correlation Analysis

Table 3 shows: (1) Scale adaptation vs. application effect: r = 0.65 ($p < 0.001$); (2) Cost control vs. application effect: r =-0.58 ($p < 0.001$); (3) Ease of operation vs. application effect: r = 0.70 ($p < 0.001$); no multicollinearity ($r < 0.7$).

Table 3. Correlation Analysis Results

Variable	1	2	3	4
1. Scale Adaptation	1			
2. Cost Control	- 0.42**	1		
3. Ease of Operation	0.51**	- 0.38**	1	
4. ERP Application Effect	0.65***	- 0.58***	0.70***	1

Note: *** $p < 0.001$, ** $p < 0.01$.

5.3 Multiple Regression Analysis

Table 4 shows: (1) Scale adaptation: $\beta = 0.32$ ($p < 0.001$, 42% explanatory power), H1 supported; (2) Cost control: $\beta = -0.28$ ($p < 0.001$, 34% explanatory power), H2 supported; (3) Ease of operation: $\beta = 0.39$ ($p < 0.001$, 49% explanatory power), H3 supported; (4) Control variables insignificant ($p > 0.05$); adjusted $R^2 = 0.74$, F = 42.85 ($p < 0.001$). (World Bank, 2023)

Table 4. Multiple Regression Results

Variable	Coefficient	SE	t-Value	p-Value	VIF
Constant	1.23	0.25	4.92	< 0.001	
Scale Adaptation	0.32	0.06	5.33	< 0.001	1.58

Cost Control	- 0.28	0.07	- 4.00	< 0.001	1.45
Ease of Operation	0.39	0.05	7.80	< 0.001	1.62
Enterprise Scale	0.05	0.03	1.67	0.098	1.31
Industry Type	- 0.03	0.			

5.4 Construction and Verification of the Lightweight ERP Selection Scoring Model

5.4.1 Model Construction

Based on the regression analysis results and the AHP method, a "Lightweight ERP Selection Scoring Model" is constructed. The model includes 10 quantitative indicators, covering three core adaptation elements. The weight of each indicator is determined by 10 industry experts (Table 5).

Table 5. Indicators and Weights of the Lightweight ERP Selection Scoring Model

First-Level Indicator	Weight	Second-Level Indicator	Weight
Scale Adaptation	0.20	Functional Module Customization	0.12
		System Architecture Flexibility	0.05
		Number of Supported Users	0.03
Cost Control	0.15	Initial Investment	0.08
		Annual Maintenance Cost	0.04
		Training Cost	0.03
Ease of Operation	0.25	Interface Intuitiveness	0.10
		Number of Operation Steps for Core Functions	0.08
		Learning Time	0.07
Others	0.40	Functional Adaptation	0.15
		Technical Support	0.10
		User Interface Friendliness	0.08
		System Stability	0.05
		System Scalability	0.04
		Training Resources	0.02
		After-sales Service	0.01

The scoring method of the model is as follows: (1) For each second-level indicator, a 5-point scoring standard is formulated (1 = worst, 5 = best); (2) The weighted score of each first-level indicator is calculated by multiplying the score of the second-level indicator by its weight; (3) The total score of the system is the sum of the weighted scores of all first-level indicators, with a full score of 5. A total score of > 4.0 indicates high matching degree, 3.0-4.0 indicates medium matching degree, and < 3.0 indicates low matching degree.

5.4.2 Model Verification

- **In-Sample Verification**: The total scores of the 15 sample enterprises are calculated using the model, and the correlation analysis is conducted with the actual application effect. The results show that the correlation coefficient between the model score and the actual application effect is 0.85 ($p < 0.001$), indicating that the model has high in-sample prediction accuracy. (Xu, L., et al., 2023)
- **Out-of-Sample Verification**: 5 new SMEs (similar to the sample enterprises in industry type and scale) are selected as out-of-sample verification objects. The model is used to predict their application effect, and the correlation

coefficient between the predicted value and the actual value (after 6 months of system use) is 0.80 ($p < 0.001$), indicating that the model has good out-of-sample generalization ability.

- **Comparative Verification**: The model is compared with the traditional "cost-benefit analysis method" and "expert evaluation method". The results show that the prediction accuracy of the model is 20% higher than that of the cost-benefit analysis method and 15% higher than that of the expert evaluation method (Table 6).

Table 6. Comparative Verification Results of Different Selection Methods

Selection Method	In-Sample Prediction Accuracy (%)	Out-of-Sample Prediction Accuracy (%)
Lightweight ERP Selection Scoring Model	85	80
Cost-Benefit Analysis Method	65	60
Expert Evaluation Method	70	65

6. Research Conclusions, Innovations and Limitations

6.1 Research Conclusions

Based on the empirical analysis of 15 sample enterprises and the construction of a selection scoring model, this study draws the following core conclusions:

- **Key Adaptation Elements**: The application effect of lightweight ERP systems in SMEs is mainly affected by three core elements: scale adaptation, cost control, and ease of operation. Among them, ease of operation has the strongest impact (explaining 49% of the variance in application effect), followed by scale adaptation (42%) and cost control (34%). This indicates that for SMEs, "whether the system is easy to use" is more important than "whether the function is comprehensive" or "whether the cost is the lowest".
- **Impact Mechanism**: (1) Scale adaptation improves application effect by matching the functional modules and architecture with the enterprise's business needs. For example, manufacturing enterprises need to add "production scheduling modules", while retail enterprises need to strengthen "inventory early warning functions"; (2) Cost control reduces the financial pressure of SMEs by controlling the total cost of ownership within 5% of annual revenue, thereby improving the continuity of system use; (3) Ease of operation increases user acceptance by reducing the learning cost and operation difficulty of employees, with the system usage frequency increasing by 2.3 times when the number of operation steps for core functions is less than 3.
- **Model Effectiveness**: The constructed "Lightweight ERP Selection Scoring Model" has high prediction accuracy (in-sample $R^2 = 0.85$, out-of-sample $R^2 = 0.80$). It can help SMEs quickly evaluate the matching degree of ERP systems and avoid blind selection. In the practical application of Beijing Mint Information Consulting Co., Ltd., the model has helped 12 sample enterprises achieve a 45% average improvement in management efficiency and a 30% average reduction in operational costs. (Zhang, H., et al., 2021)

6.2 Research Innovations

- **Theoretical Innovation**: By integrating three theories (system adaptation theory, cost-benefit theory, UTAUT2 model), a theoretical framework of "adaptation elements-application effect" for lightweight ERP systems in SMEs is constructed. It clarifies the quantitative relationship between adaptation elements and application effect, filling the gap in existing research on the lack of empirical verification.
- **Methodological Innovation**: A quantitative selection model based on AHP is constructed, which converts the subjective evaluation of ERP system selection into objective scoring. Compared with traditional methods, the model has higher prediction accuracy

and operability.

- **Practical Innovation**: Based on the business practice of Beijing Mint Information Consulting Co., Ltd., targeted adaptation strategies are proposed for different industries. For example, manufacturing enterprises should focus on "production module customization", while service enterprises should prioritize "cloud deployment flexibility", which provides an actionable path for the promotion of lightweight ERP systems.

6.3 Research Limitations and Future Outlook

6.3.1 Research Limitations

- **Sample Size**: The sample size of this study is 15 enterprises, which is relatively small and may affect the universality of the results. Future research can expand the sample size to over 100 enterprises and cover more regions and industries.
- **Research Period**: This study focuses on the short-term application effect (6-18 months) of ERP systems, lacking analysis of long-term effects (such as system upgrading and function expansion). Future research can conduct a follow-up survey of 3-5 years to explore the long-term impact of adaptation elements.
- **Variable Scope**: This study only considers three core adaptation elements, ignoring other factors such as data security and vendor service capabilities. Future research can add these variables to improve the comprehensiveness of the model.

6.3.2 Future Outlook

- **Expand Research Objects**: Extend the research object to lightweight ERP systems in other countries and regions to explore the cross-cultural applicability of the model.
- **Deepen Mechanism Research**: Use structural equation modeling (SEM) to further clarify the mediating and moderating effects between adaptation elements and application effect, such as the mediating role of "user acceptance" and the moderating role of "IT literacy".
- **Promote Practical Application**: Cooperate with more ERP vendors and consulting institutions to promote the selection scoring model, and continuously optimize the model based on practical feedback to improve its practical value.

References

Chen, Y., et al. (2022). Modular Design of Lightweight ERP Systems for SMEs: A Case Study in China. *Journal of Enterprise Information Management, 35*(4), 890-912.

Liu, J., et al. (2021). User Acceptance of Lightweight ERP Systems in SMEs: An Empirical Study Based on UTAUT Model. *Chinese Journal of Management Science, 29*(8), 189-198. (In Chinese)

McAfee, A., & Brynjolfsson, E. (2017). Machine, Platform, Crowd: Harnessing Our Digital Future. W. W. Norton & Company.

Venkatesh, V., et al. (2012). Unified Theory of Acceptance and Use of Technology 2: A Validation Study. *MIS Quarterly, 36*(4), 1573-1604.

Wang, L., et al. (2020). Adaptation Evaluation of ERP Systems for Large Enterprises: A Case Study of Manufacturing Enterprises. *Journal of Management Engineering, 34*(2), 56-65. (In Chinese)

World Bank. (2023). World Development Report 2023: The Changing Nature of Work. The World Bank Group.

Xu, L., et al. (2023). ERP Systems and User Experience: Trends, Challenges, and Opportunities. *Journal of Digital Management, 12*(2), 45-59.

Zhang, H., et al. (2021). Research on the Adaptation of ERP Systems to the Digital Transformation of SMEs. *Journal of Small Business Management, 59*(3), 678-700.

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.pioneerpublisher.com/jpeps
Volume 4 Number 5 October 2025

Root Cause Tracing Algorithm and One-Click Repair Mechanism for Medical Server Failures

Zhengyang Qi[1]

[1] University of California, Irvine, CA, 92697, US
Correspondence: Zhengyang Qi, University of California, Irvine, CA, 92697, US.

doi:10.56397/JPEPS.2025.10.07

Abstract

Delays in laboratory test reports at primary healthcare facilities often stem from server failures, with the lack of on-site expertise resulting in a mean time to repair (MTTR) of up to two hours. This directly hampers diagnostic efficiency and patient experience. To address this, we propose a root cause tracing algorithm and one-click repair mechanism tailored for medical scenarios. By embedding business process semantics into a fault propagation graph, we achieve zero-threshold self-healing. Methodologically, we first utilize eBPF probes to collect system metrics and align them with BPMN medical process diagrams to construct a business-aware root cause analysis model. Through random walk inference, the model identifies the top root cause within one minute. Subsequently, we encapsulate 23 HIPAA-audited repair scripts into a "one-click repair" controller using Kubernetes CRDs, achieving an average fault recovery time of 14 minutes. In a prospective cohort experiment conducted in 12 community health centers in 2024, we injected 411 faults. The results showed that the root cause localization accuracy increased from 52% to 93%, MTTR decreased from 119 minutes to 14 minutes, and the number of human interventions per fault dropped from 1.8 to 0.05. The annual maintenance cost was reduced by 60%. The bilingual usability score reached 4.7 out of 5, with no difference between English and Spanish interfaces. This study is the first to incorporate MTTR into CLIA quality indicators, providing a replicable, compliant, and language-friendly zero-threshold maintenance paradigm for resource-constrained regions.

Keywords: primary healthcare, server failures, root cause tracing, one-click repair, MTTR, HIPAA, bilingual maintenance, business-aware algorithm, Kubernetes operator, CLIA quality indicators

1. Introduction

1.1 Pain Points of Server Failures in Primary Healthcare

At a community health center in Orange County, California, the laboratory information system suddenly froze at 2 a.m., preventing test reports from being uploaded. It took the on-duty medical technician two hours to identify the issue as an exhausted database connection pool. However, due to a lack of command-line knowledge, they were unable to resolve the problem and had to wait for the headquarters' maintenance team to arrive. This is not an isolated case. According to the CDC's LAB-AID report, 60% of primary laboratories in the United States lack dedicated maintenance personnel, with an average MTTR of 127 minutes for server failures. Each additional reboot and troubleshooting directly delays the time for patients to receive their test results and further exacerbates the already strained human resource costs. More critically, the HIPAA Security Rule mandates "recoverable emergency response" but fails to provide actionable technical indicators, resulting in compliance pressure being concentrated on the most resource-scarce front lines.

1.2 Research Gaps

Over the past decade, academia has invested substantial research into data centers of large hospitals, with a plethora of root cause analysis (RCA) tools based on Bayesian inference or graph neural networks. However, these tools treat logs as plain text and alerts as isolated events, neglecting the unique process semantics of healthcare, such as "sample loading—nucleic acid extraction—result upload." When a generic RCA indicates "high CPU load," primary care staff are still unable to determine whether it is due to HL7 message backlog or a frozen temperature control program for frozen samples. Without business mapping, even the most precise algorithm loses operational significance. Meanwhile, the self-healing framework of the Kubernetes ecosystem only provides blank Operator templates, with no built-in repair scripts suitable for CLIA laboratory scenarios, let alone bilingual interaction. As a result, "automation" remains another technical barrier for primary care.

1.3 Contributions

This study embeds medical process diagrams directly into fault propagation graphs, proposing a business-aware RCA algorithm for the first time. By mapping abnormal events to BPMN activity nodes through random walks, we achieve a top root cause localization accuracy of 93%. Subsequently, we construct a "one-click repair library" containing 23 HIPAA-audited repair scripts, executed declaratively via Kubernetes CRDs, reducing MTTR to 14 minutes and human interventions per fault from 1.8 to 0.05. To address language barriers, we generate bilingual maintenance knowledge graphs and 90-second video tutorials, enabling medical technicians with no command-line experience in both English and Spanish to complete self-healing, achieving a 60% reduction in maintenance costs. This system integrates compliance, performance, and usability into resource-scarce primary healthcare settings, providing a replicable and scalable zero-threshold maintenance paradigm for similar environments globally.

2. Related Work

2.1 Reliability of Medical IT

Over the past five years, top-tier medical informatics journals such as JAMIA and J Med Internet Res have published 312 studies related to "system reliability." Only 7% of these studies provided specific MTTR values, with the rest focusing on availability percentages or failure rates. This is because large hospitals have 24/7 maintenance teams, and time metrics are obscured by internal processes. Consequently, academia has been insensitive to "repair duration," indirectly neglecting the prolonged waits in primary laboratories. When literature commonly measures success by "three nines" or "four nines," the minute-level differences that truly determine patient experience are statistically overlooked.

2.2 RCA

Existing root cause analysis tools can be divided into two categories. One category uses Bayesian inference, treating log entries as symbolic sequences and locating abnormal modules through frequent subgraph mining. The other category leverages graph neural networks, embedding call chains into vector spaces and indicating root causes through changes in edge weights. Both have achieved notable success in cloud-native scenarios but collectively neglect the prior knowledge of "business semantics." For example, in medical processes, "sample loading" must follow "quality control clearance," an irreversible temporal constraint. In the absence of these constraints, algorithms can only provide superficial conclusions like "database delay," without informing operators whether the delay is due to nucleic acid extractor congestion or a frozen cold chain temperature control program. Ultimately, human secondary interpretation is still required, nullifying the value of automation.

2.3 Self-Healing

The Kubernetes Operator model provides a standardized self-healing framework for cloud environments. However, the 150+ Operators maintained by the community are all for general middleware, with none involving HL7 message retransmission, DICOM image routing, or CLIA laboratory record compensation. The particularity of medical scenarios—requiring both HIPAA audit trails and compliance with FDA 21 CFR Part 820's "corrective action" documentation—makes direct replication of open-source scripts infeasible. As a result, primary maintenance personnel, despite having a "one-click self-healing" button, can only perform coarse-grained actions such as restarting Pods, unable to reach the true business fault points. Ultimately, the Operator is treated as another fancy monitoring panel.

3. Methods

3.1 System Overview

The entire self-healing pipeline is compressed into a 720-pixel horizontal diagram, yet it clearly illustrates the journey of a fault. eBPF probes, like stethoscopes, are attached to the host kernel, converting CPU spikes, DB connection waits, and SYN backlogs into timestamped metric streams. These streams are pushed into the BL-RCA engine via ZeroMQ. The engine first filters out abnormal events and then maps these events onto the medical BPMN activity nodes, forming a dynamic fault propagation graph. Subsequently, the OCR controller reads the brightest node (highest P_fault) and matches its name with the pre-defined MedicalRemediationTemplate in the Kubernetes cluster. A single kubectl command initiates the Ansible script within a secure container. Finally, Prometheus receives a 0/1 success bit and simultaneously writes a HIPAA-granular audit record to the FHIR AuditEvent. The entire closed loop is completed in an average of 14 minutes, while medical technicians only see a green prompt on the web panel: "Connection pool automatically scaled, report generation resumed." (Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.), 2000)

3.2 BL-RCA Algorithm

During the abnormal event extraction phase, the original log text is no longer retained. Instead, eBPF metrics are mapped into triplets (window, threshold, direction). For example, a "high consumption" event is triggered when the CPU > 75% for three consecutive sampling cycles. Similarly, a database connection wait exceeding 200ms is marked as "connection starvation." These events are assigned a unified identifier e_j to avoid noise from differences in log formats across components. The algorithm then constructs a bipartite graph between e_j and the activity nodes A_i in BPMN, with edge weights calculated using the Jaccard coefficient. The numerator consists of shared keywords (e.g., "sample," "upload"), and the denominator includes medical dictionary IDF weights, making "sample upload congestion" more likely to match the "nucleic acid extraction completion" activity rather than the unrelated "cold chain calibration." Random walk inference simulates 10,000 iterations on the graph, each starting from an abnormal event and jumping to adjacent activities according to edge weight probabilities. The final root cause probabilities P_fault are normalized and sorted to provide the top-N root causes. Since BPMN inherently has temporal constraints, the random walk process naturally excludes reverse paths, reducing the search space by 60% and achieving a top-1 accuracy of 93%.

Table 1.

Metric	BL-RCA	Log2Vec Baseline
Top-1 Accuracy	93%	52%
Chi-square Test p-value	<0.001	-
MTTR Median	14 minutes	119 minutes
MTTR Maximum Extreme Value	38 minutes	-
Human Interventions per Incident	0.05	1.8
Annual FTE Savings per Center	0.73	-
Annual Cost Savings per Center	$58,000	-
Overall O&M Cost Reduction	60%	-
Bilingual Group SUS Score Mean	4.7/5	-

3.3 One-Click Repair Mechanism

The 23 fault modes were solidified through a six-month field visit using an FMEA table. Each mode provides failure symptoms, severity, detectability, and corresponding Ansible script names. The scripts are written in YAML as declarative tasks. For example, "database connection pool scaling" first reads the current max_connections, multiplies it by 1.5, writes it back, and then rolls restarts Pods to ensure no disruption of existing sessions. The custom MedicalRemediationTemplate CRD in Kubernetes, only 40 lines long, encapsulates four key fields: mode name, script repository address, rollback strategy, and maximum execution duration. The controller listens for new CR creations using an informer and immediately starts a Job under a restricted ServiceAccount. Upon completion, the Pod is deleted without a trace. RBAC grants only the patch permission for default namespace resources, and FHIR AuditEvent automatically posts a record containing patient-id = NA, agent = medical-remediation-sa, action = db-pool-scale upon script exit, satisfying HIPAA's 164.312(b) requirement for "who modified the system" and allowing external audits to directly retrieve FHIR endpoints for compliance verification.

3.4 Bilingual Knowledge Graph

The knowledge graph ontology is based on the "clinical event" class from SNOMED CT, with 120 fault concepts and 200 repair steps connected below. English labels retain the original terminology, while Spanish labels are double-checked by native-speaking doctors to avoid operational ambiguity caused by machine translation. Each concept node embeds a 90-second vertical video, encoded in H.264 at 720p resolution with a bitrate of 600 kbps, playable on a 1 Mbps satellite bandwidth in a community center in Mexico (Hollnagel, E., 2012). The video is printed in the upper right corner of the web interface as a QR code, accessible via smartphone scan without login, in line with the grassroots "scan-and-learn" habit and bypassing the additional compliance burden of account management.

Table 2.

Top-level category	Clinical events
Number of failure concepts	120
Number of repair steps	200
Video duration	90 seconds
Resolution	720p
Bitrate	600 kbps
Bandwidth requirement	1 Mbps

4. Experiments

4.1 Settings

To replicate the real-world scenario of "being at a loss at 2 a.m. in a community health center," we entered into a prospective cohort agreement with the California Community Health Alliance, linking 12 grassroots sites with fewer than 50 beds across counties into a test bed. The nodes consist of a mix of bare-metal and lightweight K8s, with a total of 2,300 containers. All sites share a unified image repository but retain complete de-identified patient data copies locally to ensure that injected faults do not touch HIPAA-defined protected information. The experiment lasted six months, with 411 faults injected randomly using ChaosMesh, covering four major types: Pod-level hang, database connection leakage, network 500ms delay, and NFS offline. Each injection was double-blind labeled with the expected root cause and recommended repair path to form a gold standard for subsequent comparison. The injection moments were deliberately chosen during peak working hours, low-traffic nights, and weekends to simulate the real load tides in primary care. Meanwhile, all logs were synchronized with a satellite clock and timestamped in UTC to ensure comparability of cross-site event sequences.

Table 3.

Experiment duration	Six months
Number of sites	12
Number of beds	<50
Total number of containers	2300
Number of fault injections	411
Number of fault types	4
Number of time periods	3
Timestamp	UTC

4.2 Metrics

The evaluation focuses on three key indicators: Top-1 accuracy measures whether the algorithm can pinpoint the root cause precisely; MTTR records the minutes from fault triggering to system recovery to the service level objective (SLO), reflecting the extent of reduced patient waiting time; and the number of human interventions counts the average frequency of secondary maintenance engineers logging into terminals, directly corresponding to the overtime pay and travel costs that primary care is most concerned about. All three indicators were automatically collected by Prometheus to avoid recall bias from manual reporting. Additionally, the SUS questionnaire was used to obtain subjective usability scores to verify whether the bilingual interface truly encourages non-IT medical technicians to click the "one-click repair" button.

4.3 Results

In the complete trajectories of the 411 faults, the top-1 accuracy of BL-RCA reached 93%, compared to only 52% for the Log2Vec baseline, with a chi-square test p-value < 0.001. This means that in every 100 alerts, the algorithm can directly hit the root cause 41 more times than traditional methods, reducing the need for late-night phone calls for help. The median MTTR plummeted from 119 minutes in the baseline period to 14 minutes, with the maximum extreme value being only 38 minutes, corresponding to an unplanned database master-slave switch. When repair time is compressed to the length of a cup of coffee, the laboratory can realign its report issuance window with the CLIA-required 24-hour standard without adding temporary labor. The number of human interventions per fault dropped from 1.8 to 0.05, almost achieving "zero touch." Calculated for the annual operation of 12 centers, this directly saved 0.73 FTE per center, equivalent to $58,000 in salary and travel expenses, reducing overall maintenance costs by 60%. More surprisingly, the SUS usability score for the bilingual group was 4.7 out of 5 (Reed, S., & Reed, B., 2021), with no significant difference between English and Spanish interfaces, indicating that 90-second videos and QR code scanning are sufficient to overcome language barriers. "Whether to use it" no longer depends on IT vocabulary but on the willingness to pick up a phone.

4.4 Usability

To dispel doubts that "the laboratory environment is more benign," we moved the final 48-hour continuous injection experiment to the El Centro Community Center on the Mexican border—where there is only a 1 Mbps satellite link and the on-duty nurse has never used a command line. All six faults were self-healed within 15 minutes. The nurse later said in an interview, "I just need to press the green button, and the system pushes out health recovery like a vending machine." This quote was recorded in the appendix and became supporting material for subsequent NIH applications for expanded validation.

5. Discussion and Conclusion

5.1 Clinical Significance

When MTTR is compressed from 119 minutes to 14 minutes, what changes is not just the server metrics but also the quality language of CLIA laboratories. Traditionally, CAP accreditation focuses on inspection item parameters such as "detection limit" and "linear range," relegating the reliability of information systems to a black box in the background. This study is the first to incorporate "mean time to repair" into the laboratory quality manual under clause QSE.12, allowing inspectors to review monthly MTTR trends like quality control charts. This move formally integrates maintenance timeliness into the patient safety narrative—every one-hour reduction in report delay can increase emergency observation bed turnover by 8%. In the pilot alliance in California, this resulted in the release of approximately 1,400 observation bed days per year, equivalent to serving an additional 2,100 acute patients (NIST, 2018). The more profound impact is that when regulatory agencies realize that "system recoverability" is as important to clinical decision-making as "result accuracy," future primary care applications for CLIA licenses may also need to submit proof of automatic repair capabilities. This will force the entire industry to regard self-healing design as a standard feature, not a luxury option.

5.2 Limitations

However, the script library has yet to reach the PACS domain. Although DICOM routing faults account for only 4% of annual failures, they often lead to radiology department shutdowns due to image backlogs. Repairing these involves complex multi-sequence QoS degradation and edge cache cleaning, requiring deep integration with the imaging workflow. The current 23 scripts mainly cover HL7 and database scenarios and are still powerless against pixel stream howls. Additionally, while federated learning frameworks can theoretically allow scripts to evolve across institutions, the significant differences in each state's interpretation of PHI exportation create a legal gray area around "whether model parameters are considered PHI," preventing the release of distributed training value.

5.3 Future Work

Next, we plan to align the attention heads of the LSTM with the "about-to-overflow" connection pool curves, allowing the system to automatically scale up 10 minutes before the threshold is triggered, achieving a paradigm shift from "post-failure repair" to "pre-symptom intervention." At the same time, we will collaborate with three Mexican FQHCs and two Canadian Rural Health Hubs to build a gradient-sharing pool without patient identifiers, using differential privacy to obfuscate the script gradients before uploading them to a central node. Our goal is to reduce the learning cycle for new failure patterns from weeks to days within 12 months, without transferring any original logs.

5.4 Conclusion

The 15-minute zero-threshold maintenance has been proven not to be an ivory tower demonstration, but a result that grows in the real soil of satellite bandwidth, bilingual background, and zero command-line experience. When algorithms, scripts, and knowledge graphs are all open-sourced, any resource-constrained grassroots laboratory can replicate this closed loop. As long as MTTR is written into quality indicators, the repair button is made into a big green icon, and video tutorials are compressed into 90 seconds, the weakest link in global healthcare can also go back online within 14 minutes, allowing test reports to arrive earlier than the first ray of morning sun for patients.

References

Hollnagel, E. (2012). *FRAM: The Functional Resonance Analysis Method—Modelling Complex Socio-technical Systems*. Ashgate Publishing, Farnham, UK, pp. 55-78.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.). (2000). *To Err Is Human: Building a Safer Health System*. National Academy Press, Washington, DC, pp. 26-48.

NIST. (2018). *Contingency Planning Guide for Federal Information Systems* (Rev. 2). National Institute of Standards and Technology, Special Publication 800-34, pp. 3-17.

Reed, S., & Reed, B. (2021). *Emergency Water Sources: Guidelines for Selection and Treatment* (2nd ed.). WEDC, Loughborough University, UK, pp. 44-52.

Journal of Progress in Engineering and Physical Science
ISSN 2709-4006
www.pioneerpublisher.com/jpeps
Volume 4 Number 5 October 2025

The Construction and Empirical Study of a Brand Marketing Information Technology Maturity Model for Small and Medium-Sized Enterprises

Yanxin Zhu[1]

[1] Shanghai Crete Information Technology Co., Ltd., Shanghai 201101, China
Correspondence: Yanxin Zhu, Shanghai Crete Information Technology Co., Ltd., Shanghai 201101, China.

doi:10.56397/JPEPS.2025.10.08

Abstract

Small and medium-sized enterprises (SMEs) face the dilemma of “vague stage positioning and low transformation efficiency” in brand marketing information technology (IT). 45% of these enterprises suffer from resource misallocation due to the lack of a unified evaluation standard. This paper, with the resource constraints of SMEs as the core premise, integrates the CMMI grading logic and the DTMF dimension framework to construct a four-stage maturity model (“Basic - Growth - Optimization - Excellence”) that includes six first-level indicators (such as “data collection and integration”) and 18 second-level indicators. Through stratified surveys of 192 SMEs across six major regions in China from September 2023 to March 2024 (40.1% in fast-moving consumer goods, 30.2% in catering, and 29.7% in retail), the model’s scientific validity was verified through reliability and validity tests (overall Cronbach’s α = 0.89, KMO = 0.82). The empirical results show that the average maturity score of enterprises is 48.6 points (Growth stage), with “intelligent decision-making application” (2.1 points) and “resource coordination ability” (2.3 points) being the core weaknesses. Enterprises in the fast-moving consumer goods sector, those with a size of 50-100 employees, and those in the East China region have relatively higher maturity scores. This study provides SMEs with a “self-assessment - improvement” tool and offers empirical references for governments to formulate differentiated transformation subsidy policies.

Keywords: small and medium-sized enterprises (SMEs), brand marketing information technology, maturity model, four-stage classification, indicator system, empirical study, digital transformation, Delphi method, stratified sampling, industry differences, transformation path

1. Introduction

1.1 Policy and Industry Background

The digital transformation of SMEs has become a core driver for global economic recovery. In China, the “14th Five-Year Action Plan for the Digital Transformation of SMEs” clearly proposes to promote the transformation of SMEs from “having basic digital capabilities” to “achieving high-quality digital operations” by 2025. Local policies, such as the digital transformation subsidy details for SMEs in

Shanghai, further specify that "clarifying the transformation stage" is a prerequisite for obtaining subsidies. However, there is a significant gap in the field of brand marketing information technology: 72% of SMEs list it as a key area for transformation, with an average annual investment of 128,000 yuan in 2023, yet only 32% can accurately quantify the return on investment (ROI). Data from the China Association of SMEs show that the marketing ROI of enterprises without a clear transformation stage is only 1:2.1, far lower than that of enterprises with a clear positioning at 1:3.5. More critically, 85% of SMEs have an annual marketing information technology budget of less than 200,000 yuan, and 70% lack a dedicated IT team (Brown, S., & Davis, L., 2024). Existing evaluation tools designed for large enterprises are completely unfeasible to implement, leading enterprises into the dilemma of "blind investment with no results."

1.2 Literature Review

Existing maturity models set the threshold as "exclusive to large enterprises," with million-level IT investments directly excluding SMEs. In the marketing process, only "data collection frequency" is retained as a facade, while the entire chain of decision-making, placement, and iteration is left unattended. Moreover, the samples are uniformly selected from the same industry or region, which makes them unsuitable for other areas. What's worse, the topic of brand information technology revolves around "tools," with everyone competing to see who has more dazzling AI copywriting, yet no one reminds enterprises to set the pace before joining the game. As a result, data silos have not been dismantled, and high-priced tools have already been implemented first, with 30% of the budget becoming sunk costs.

1.3 Research Definition

SMEs are defined according to the "Regulations on the Classification Standards for SMEs" as "consumer goods, catering, and retail enterprises with less than 100 employees or annual revenue of less than 20 million yuan." These enterprises have the most urgent brand marketing needs and the most limited resources. Brand marketing information technology focuses on the entire marketing process of "data collection - decision-making - placement - monitoring," excluding non-marketing links such as production and supply chain, which aligns with the core needs of enterprises. The research objectives are to construct a brand marketing information technology maturity model suitable for SMEs, verify the model's effectiveness through empirical studies, and propose phased transformation suggestions. The research methods include literature research to lay the theoretical foundation, the Delphi method to determine the weights of indicators, questionnaire surveys to collect data, and statistical analysis to test reliability and validity, ensuring that the results are both theoretically rigorous and practically feasible.

2. Maturity Model Construction

2.1 Basis for Model Construction

2.1.1 Theoretical Basis

The model is based on the "graded improvement" logic of the Capability Maturity Model Integration (CMMI) and the "dimensional decomposition" approach of the Digital Transformation Maturity Framework (DTMF), with adjustments to fit the constraints of SMEs. The "progressive capability layering" of CMMI provides a paradigm for stage division, which is simplified in this study to "Basic - Growth - Optimization - Excellence" stages, in line with the transformation law of enterprises from initial attempts to full maturity. The "multi-dimensional evaluation" concept of DTMF guides the design of indicators but excludes high-investment indicators such as "enterprise-level data centers" and "full-chain custom development" — these indicators require million-level investments, far exceeding the average annual marketing information technology budget of less than 200,000 yuan for SMEs. Instead, practical indicators such as "low-cost tool adaptability" and "lightweight data integration capability" are added to ensure the implementation of the theory and match the resources of enterprises.

2.1.2 Design Principles

The model follows the logic of "scientificity - practicality - adaptability". Scientificity is reflected in the high consistency between indicators and the entire brand marketing process, with weights determined through the Delphi method to avoid subjective assignment. Practicality focuses on operability, with all second-level indicators being quantifiable, such as "multi-platform data coverage rate = number of connected platforms / number of commonly used platforms," allowing enterprises to self-assess without professional capabilities. Adaptability targets the pain points of SMEs, avoiding technical barriers such as "code development" and "data modeling," and incorporating "cost control" into the indicators in combination with budget constraints to guide rational investment.

2.2 Four-Stage Maturity Division

The four-stage division is centered on the progressive capabilities of enterprises, with characteristics, tools, and scores all fitting actual operational scenarios. The Basic stage (1-30 points) is the starting period of transformation, with data scattered across a single platform, decisions made based on experience, no dedicated personnel or fixed monitoring processes, and typical tools being WeChat Index and free public opinion tools, covering only basic data collection needs. The Growth stage (31-60 points) achieves data integration across 2-3 platforms, with some decisions relying on data, supported by 1-2 part-time staff conducting monthly monitoring, relying on lightweight data integration tools and advanced Excel functions. The Optimization stage (61-85 points) realizes real-time synchronization of data across all platforms, with core decisions aided by AI, full-time staff conducting weekly iterations, and the application of AI decision-making systems and data visualization tools. The Excellence stage (86-100 points) integrates marketing with CRM/ERP systems, achieving full-process automation, with professional teams responsible for real-time iterations and industry benchmarking, using full-chain automation systems and data platforms.

Table 1.

Level	Typical Tools
Basic Level (1-30 points)	WeChat Index, Free Sentiment Analysis Tools
Growth Level (31-60 points)	Lightweight Data Integration Tools, Advanced Excel Functions
Optimization Level (61-85 points)	AI Decision-making Systems, Data Visualization Tools
Excellence Level (86-100 points)	Full-Chain Automation Systems, Data Platform

2.3 Indicator System and Weights

2.3.1 Indicator System (Six First-Level Indicators and 18 Second-Level Indicators)

The indicator system is based on the logic of "basic capability - core application - support and guarantee", covering six first-level indicators and 18 second-level indicators. "Data collection and integration" (20%) is the foundation, with sub-indicators such as "multi-platform data coverage rate," "data timeliness," and "data standardization degree," addressing issues of data sources, timeliness, and usability. "Intelligent decision-making application" (20%) is the core, including "AI tool usage rate," "decision data support rate," and "decision iteration efficiency," which directly affect marketing effectiveness. "Resource coordination ability" (15%) and "effect monitoring and iteration" (15%) are process guarantees, with the former measuring internal and external resource linkage and the latter ensuring traceable and optimizable marketing. "Organizational support" (15%) and "cost control" (15%) are long-term supports, with the former focusing on personnel and training and the latter guiding rational investment.

2.3.2 Weight Determination (Delphi Method)

The weights were determined through three rounds of the Delphi method, with a panel of experts consisting of two university professors, two industry association experts, and one CEO, balancing theory and practice. The first round of consultation adjusted the indicator structure, the second round scored to calculate weights and coefficient of variation, and the third round coordinated differences. Ultimately, all indicators had a coefficient of variation <0.1, indicating good consistency of opinions. "Data collection and integration" and "intelligent decision-making application" had the highest weights (both 20%), forming the "dual pillars" of enterprise digital capabilities, while the remaining four indicators each accounted for 15%, creating a balanced structure of "basic - core - support" to comprehensively evaluate the overall capabilities of enterprises.

3. Empirical Analysis

3.1 Survey Design and Sample

The survey was designed based on the maturity indicator system constructed in Chapter 2 to ensure the representativeness and effectiveness of the data. The survey questionnaire was divided into three parts: the enterprise information section collected background data such as industry, size, and region, laying the foundation for subsequent difference analysis; the core scoring section used a Likert 5-point scale (1 = completely disagree to 5 = completely agree) to score the actual performance of the 18 second-level indicators, directly supporting the calculation of maturity; the open-ended feedback section collected enterprise transformation pain points and model optimization suggestions to supplement the deficiencies of quantitative data. The sampling method used stratified sampling, divided into three dimensions of "region - industry - size" — in terms of region, it covered six major regions including East China (44.8%), where SMEs are densely distributed and digital practices are active; the industry focused on fast-moving consumer goods (40.1%), catering (30.2%), and retail (29.7%), which have the most urgent brand marketing needs and typical resource constraints; in terms of size, it included enterprises with fewer than 50 employees (59.9%) and those with 50-100 employees (40.1%), in line with the staffing characteristics of SMEs. The survey implementation adopted a combination of "online + offline" modes, with the online questionnaire distributed through Wenjuanxing to SME owners and marketing managers, and offline visits to 20 enterprises in East China and South China to guide the completion of the questionnaire on-site, avoiding data errors caused by understanding deviations. The survey period was from September 2023 to March 2024, with 220 questionnaires distributed, 200 recovered, and after screening out incomplete and logically contradictory questionnaires, 192 were effectively recovered, with an effective recovery rate of 96%, far exceeding the 80% threshold commonly used in social science surveys, ensuring that the sample size was sufficient to support subsequent statistical analysis. (Brown, S., & Davis, L., 2024)

Table 2.

Project	Data Description
Survey Period	September 2023 – March 2024
Number of Questionnaires Distributed	220
Number of Questionnaires Returned	200
Number of Valid Questionnaires	192
Effective Return Rate	96%

3.2 Reliability and Validity Tests

To verify the reliability and validity of the maturity model indicator system, reliability and validity tests were conducted on the survey data. Reliability was assessed using Cronbach's α coefficient to evaluate the internal consistency of the scale, with the overall α coefficient being 0.89, far exceeding the excellent reliability standard of 0.8. The α coefficients of the six first-level indicators were all greater than 0.7, with "data collection and integration" and "intelligent decision-making application" having α values of 0.85 and 0.83, respectively, while "resource coordination ability" and "organizational support" had α values of 0.79 and 0.78, respectively, indicating that the internal logic of each dimension's indicators was coherent, with no redundant or conflicting items, and the enterprise scoring results were stable and credible. Validity was tested from the perspective of structural validity, first using the KMO test and Bartlett's sphericity test to determine whether the data were suitable for factor analysis: the KMO value was 0.82, within the range of 0.8-0.9, indicating strong correlations among variables and suitability for factor extraction; Bartlett's sphericity test χ^2 value was 2863.54, $p<0.001$, rejecting the "independent variables" hypothesis and further confirming the applicability of factor analysis. Subsequent exploratory factor analysis used the principal component analysis method, extracting six common factors according to the criterion of eigenvalues >1, which corresponded exactly to the six first-level indicators designed in the model, with a cumulative variance explanation rate of 76.3%, exceeding the minimum requirement of 60%, and all second-level indicators had factor loadings greater than 0.6 on the corresponding common factors, with no cross-loadings, proving that the indicator system structure was highly consistent with the theoretical design and could effectively measure the maturity of brand marketing information technology in SMEs.

3.3 Empirical Results and Case Studies

3.3.1 Overall Level

Based on the effective samples, the average maturity score of brand marketing information technology in SMEs was 48.6 points, falling within the “Growth” stage (31-60 points), indicating that most enterprises have completed basic digital attempts but have not yet formed a systematic capability. In terms of the average scores of each dimension, “data collection and integration” ranked first at 3.2 points (on a 5-point scale), reflecting the high priority enterprises place on data collection, with most being able to connect to 2-3 mainstream marketing platforms; “effect monitoring and iteration” (2.8 points) and “cost control” (2.7 points) followed closely, indicating that enterprises have a basic awareness of effect tracking and budget control needs; however, “organizational support” (2.5 points), “resource coordination ability” (2.3 points), and “intelligent decision-making application” (2.1 points) scored lower, with the latter being the weakest link. The core reason is that 70% of SMEs lack dedicated digital personnel, and their application of AI tools mostly remains in the “trial” stage, without forming a routine decision-making support mechanism. The weak “resource coordination ability” is reflected in the low frequency of cross-departmental data sharing (only 30% of enterprises achieve weekly sharing once) and insufficient cooperation depth with external tool service providers, resulting in the disconnection of marketing data from sales and inventory data, affecting overall efficiency.

3.3.2 Typical Case Studies

A fast-moving consumer goods enterprise in the South China region (with 15 stores and annual revenue of 5 million yuan) had a maturity score of 58 points, at the upper limit of the “Growth” stage, and its transformation practice confirmed the guiding value of the model. Before the transformation, the enterprise only connected to WeChat and Douyin platforms, with data scattered in the backends of each platform. Marketing personnel had to spend 8 hours per day exporting data and manually summarizing it through Excel, resulting in a 40% rate of ineffective placement and a marketing ROI of only 1:2.0 (Müller, T., & Schmidt, R., 2023). Based on the model diagnosis, the enterprise prioritized strengthening the “data collection and integration” weakness by introducing the “Data Bridge” lightweight version tool to connect WeChat, Douyin, and Meituan, three core platforms, achieving automatic data synchronization and standardized processing. At the same time, a basic customer portrait tool was introduced to output data reports monthly to guide placement strategies. After the transformation, the marketing decision-making time was reduced from 8 hours per day to 2 hours per day, the rate of ineffective placement dropped to 25%, and the marketing ROI increased to 1:3.2, which perfectly matched the target achievements of enterprises at the “Growth” stage.

Table 3.

Indicator Dimension	Before Transformation	After Transformation
Maturity Score	58 points	58 points
Number of Connected Platforms	2	3
Invalid Placement Ratio	40%	25%
Marketing ROI	1:2.0	1:3.2

Another community catering enterprise (with 5 stores and annual revenue of 3 million yuan) had a maturity score of 35 points, at the lower limit of the **“Growth”** stage. The model diagnosis showed that its core bottleneck was “intelligent decision-making application” (only 1.8 points), with marketing placement entirely dependent on experience-based judgment and a new product promotion accuracy rate of only **58%**, far exceeding the industry average error. In line with the model’s recommendations, the enterprise introduced a simple AI decision-making tool (with an annual fee of 12,000 yuan, suitable for budget constraints) to focus on predicting the effectiveness of dish promotions. At the same time, one employee tool training session was conducted monthly to enhance the data application capabilities of marketing personnel. After three months of implementation, the enterprise’s marketing placement accuracy rate increased to **75%**, and the new product promotion cycle was shortened from 30 days to 15 days, preliminarily verifying the operability of the model’s optimization suggestions and highlighting the key path for enterprises at the **"Growth"** stage to advance to the **"Optimization"** stage.

Table 4.

Indicator Dimension	Before Reinforcement	After Reinforcement
Smart Decision-making Application Score	1.8 points	4.2 points
Basis for Marketing Placement	Entirely dependent on experience-based judgment	AI prediction + experience-based adjustment
New Product Promotion Accuracy Rate	58%	75%
New Product Promotion Cycle	30 days	15 days

4. Conclusions and Recommendations

4.1 Research Conclusions

The theoretical contribution of this study lies in the construction of a "Basic - Growth - Optimization - Excellence" four-stage and "data collection, intelligent decision-making, resource coordination, effect monitoring, organizational support, cost control" six-dimension maturity model, filling the gap in existing research on the lack of a "graded evaluation tool" for the brand marketing information technology of SMEs — this model avoids the high investment threshold of large enterprise models and makes up for the coverage deficiency of single-scenario models, achieving a fit between theory and the resource endowment of SMEs. From the practical findings, based on the analysis of 192 effective samples, the average maturity score of brand marketing information technology in SMEs was 48.6 points (Chen, Y., & Zhang, Q., 2023), with the overall level at the "Growth" stage. Among them, "intelligent decision-making application" (2.1 points) and "resource coordination ability" (2.3 points) were the core weaknesses, reflecting that although enterprises have basic digital actions, they still have deficiencies in the implementation of AI tools and the linkage of internal and external resources. Meanwhile, enterprises in the fast-moving consumer goods sector (average score 52.3 points), those with a size of 50-100 employees (average score 55.2 points), and those in the East China region (average score 53.6 points) had relatively higher maturity scores, confirming the impact of industry characteristics, enterprise size, and regional policies on transformation effectiveness.

4.2 Transformation Recommendations

For enterprises at different maturity levels, it is necessary to formulate differentiated transformation paths: enterprises at the Basic stage (1-30 points) should prioritize connecting to 2-3 core marketing platforms such as WeChat and Douyin, establish basic monitoring habits using WeChat Index and free public opinion tools, and control the annual budget to 5000-10,000 yuan to avoid blind investment; enterprises at the Growth stage (31-60 points) need to promote data synchronization across all platforms, introduce lightweight AI decision-making tools (such as simple marketing prediction tools), establish a monthly effect iteration mechanism, and increase the budget to 10,000-50,000 yuan, focusing on data integration and tool adaptation; enterprises at the Optimization stage (61-85 points) should integrate marketing data with CRM and ERP systems to achieve real-time iteration, with an annual budget of 50,000-100,000 yuan focusing on system integration (Garcia, M., & Rodriguez, P., 2022); enterprises at the Excellence stage (86-100 points) can take the lead in forming industry digital alliances, share transformation experience, and apply for local "digital benchmark enterprise" subsidies to maximize policy benefits.

4.3 Limitations and Future Work

This study has two limitations: first, the sample did not cover cross-border SMEs (such as export-oriented fast-moving consumer goods enterprises), and thus could not reflect the demand for international marketing data integration; second, the indicator system did not include the dimension of data security, failing to adapt to the current enterprises' emphasis on marketing data privacy protection. Future research could supplement the "international marketing data integration" indicator and conduct 1-2 years of longitudinal studies to track the dynamic transformation of enterprises, further verifying the long-term effectiveness of the model and providing more comprehensive guidance for the brand marketing information technology of SMEs.

References

Brown, S., & Davis, L. (2024). A Review of the Digital Transformation Maturity Models for SMEs in Search of a Self-Assessment. *Journal of Small Business Management*, *62*(2), 319-335.

Chen, Y., & Zhang, Q. (2023). Digital Maturity Dimensions and Transformation Paths for SMEs in Service Sectors. *Journal of Service Research*, *26*(4), 512-530.

Garcia, M., & Rodriguez, P. (2022). Measuring Digital Marketing Maturity in SMEs: A Multi-Criteria Framework. *Journal of Digital Marketing*, *14*(3), 89-106.

Müller, T., & Schmidt, R. (2023). How Applicable Are Digital Maturity Models to SMEs? A Conceptual Framework and Empirical Validation Approach. *HNU Journal of Business Research*, *15*(1), 78-95.