IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 June 2024, accepted 11 July 2024, date of publication 16 July 2024, date of current version 29 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3429359

RESEARCH ARTICLE

A Multi-Channel Advertising Budget Allocation Using Reinforcement Learning and an Improved Differential Evolution Algorithm

MENGFAN LI[1], JIAN ZHANG[2], ROOHALLAH ALIZADEHSANI[3], (Member, IEEE), AND PAWEŁ PŁAWIAK[4,5]

[1]Department of Arts, University of Melbourne, Melbourne, VIC 3000, Australia
[2]School of Law, University of Malaya, Kuala Lumpur 50603, Malaysia
[3]Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Waurn Ponds, VIC 3216, Australia
[4]Department of Computer Science, Faculty of Computer Science and Telecommunications, Cracow University of Technology, 31-155 Kraków, Poland
[5]Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, 44-100 Gliwice, Poland

Corresponding authors: Jian Zhang (s2100678@siswa.um.edu.my) and Paweł Pławiak (pawel.plawiak@pk.edu.pl)

ABSTRACT Budget allocation across multiple advertising channels involves periodically dividing a fixed total budget among various channels. Yet, the challenge of making sequential decisions to optimize long-term benefits rather than short-term gains is often overlooked. Additionally, more apparent connections must be made between actions taken on one advertising channel and the outcomes on others. Furthermore, budget limitations narrow down the range of potential optimal strategies that can be pursued. In response to these challenges, this study unveils a pioneering multi-channel advertising budget allocation approach that leverages a reinforcement learning (RL) Q-learning framework enriched with an advanced Differential Evolution (DE) algorithm to refine the Q-learning methodology. The RL element makes informed sequential decisions, adeptly adjusting strategies to favor long-term rewards by assimilating environmental feedback. Complementing this, the enhanced DE algorithm introduces an inventive clustering-based mutation technique, exploiting key groupings within the DE population to generate novel and practical solutions. The model is further bolstered by a discretization tactic aimed at simplifying the model by streamlining costs. The proposed methodology is rigorously validated using two extensive datasets: the Chinese Internet Company Advertising Dataset (CICAD) and CRITEO-UPLIFT v2, employing metrics like Area Under the Cost Curve (AUCC) and Expected Outcome Metric (EOM) as measures of performance. The empirical results affirm the superiority of the model, showcasing its exceptional performance with significant scores (AUCC = 0.750 and EOM = 0.736 for CICAD; AUCC = 0.813 and EOM = 0.829 for CRITEO-UPLIFT v2), thereby illustrating the model's proficiency in navigating the multifaceted challenges associated with multi-channel budget allocation and establishing a new benchmark in the field.

INDEX TERMS Budget allocation with constraint, marketing, social networks, reinforcement learning, differential evolution.

I. INTRODUCTION

The rising variety of digital media platforms, including websites, mobile apps, and social networks, has transformed the marketing landscape. Marketers are increasingly leveraging these diverse channels to strengthen customer relationships. However, they face the significant challenge of optimally distributing their advertising budget across these platforms to achieve maximum impact, a dilemma known as Budget Allocation with Constraints [1]. This challenge is gaining attention in advertising and marketing due to its critical role in strategic planning and substantial financial stakes. For example, in the early months of 2019 alone, search advertising spending in the US topped 28 billion USD,

The associate editor coordinating the review of this manuscript and approving it for publication was Daniela Cristina Momete.

making up roughly half of the total online advertising expenditure. This underscores the importance and economic scale of devising effective budget allocation strategies in digital advertising [2].

Much of the current research is focused on budget allocation to a single channel [3], [4]. While single-channel strategies may appear more straightforward for budgeting, they overlook crucial complexities. A critical area that needs attention is underestimating the synergistic potential of multi-channel integration. Today's marketing landscape features a variety of channels, each with unique benefits. By strategically leveraging these channels together, the overall impact of an advertising campaign can be significantly amplified. Each channel has the potential to complement and enhance the effectiveness of the others. Consequently, the narrow approach of allocating budgets to individual channels without considering the synergy between them may result in suboptimal outcomes [5].

Research on multi-channel budget optimization is limited [6], [7], [8]. However, they face several challenges in allocating budgets effectively across multiple channels. One key issue is the need for sequential decision-making considering the entire advertising lifecycle. This requires considering at least two factors: the carryover effects within a single channel and the influence on other channels, aiming to maximize long-term benefits rather than just immediate gains. Specifically, spending in one channel may also enhance performance in others. Additionally, the inherent budget limitations pose the question of how to distribute available funds wisely to optimize overall benefits. The RL approach, such as Q-learning [9], offers a promising approach to address these challenges by learning optimal strategies through trial and error, factoring in the immediate and long-term impacts of budget allocations across channels. By continuously updating its strategy based on real-world feedback, RL can dynamically adjust budget distributions to meet changing market conditions and channel interactions, thus potentially maximizing the cumulative benefits within budget constraints [10].

Meta-heuristic algorithms provide a robust and sophisticated approach to optimizing budget allocation [11], [12], [13], [14], significantly when enhancing the Q-learning algorithm [15]. They are particularly adept at navigating the complex and layered landscapes of budgeting decisions, where traditional optimization techniques may fall short [16], [17]. Their strength lies in effectively identifying near-optimal solutions across vast and intricate decision spaces, a critical advantage when dealing with the nuanced and often non-linear nature of financial planning. The DE algorithm stands out within this meta-heuristic family for its unique and potent strategy [14], [18]. It uses a straightforward yet impactful method that relies on the differential between various solution vectors to probe and capitalize on the search domain [19]. This tactic enables the DE algorithm to adjust its search dynamically, canvassing a wide range of potential solutions [20]. This adaptability is invaluable in budget optimization, where many variables and conditions are unpredictable. The DE algorithm operates through three key phases: mutation, crossover, and selection. Mutation involves creating a new solution by scaling and combining differences among current candidate solutions. Crossover then merges this new mutation vector with an existing one, introducing variation and new possibilities. Finally, selection is choosing the optimal solutions among the latest candidates and the existing pool based on their efficacy. Notably, the mutation step is critical in generating viable and promising solutions. It injects diversity and innovation into the solution pool, ensuring that the search process remains dynamic and can uncover superior budget allocation strategies [21].

This article introduces a comprehensive approach to optimizing multi-channel advertising budgets by integrating RL and the DE algorithm. The proposed model is designed to empower marketers to efficiently allocate their limited budgets across various channels, enhancing the overall impact of their advertising efforts. The methodology is structured around three core components:

- Discretization of cost space: The initial phase of the model addresses the challenge of estimating the potential value (Q value) of each budget allocation within the entire budget range. To simplify this complex task, The cost space is segmented into finite sub-intervals, with the average cost of each interval representing a distinct action. This transformation converts the Q function into a more manageable step function. By establishing thresholds for budget allocation within each channel, the model ensures that surpassing these thresholds leads to significant increases in the Q value. This strategic discretization streamlines the action space and captures the intricate spending patterns across different channels, facilitating more precise budget allocation decisions.
- Analysis of inter-channel effects: The model delves into the intricate dynamics between various advertising channels to understand how an advertisement in one channel can influence subsequent advertisements in the same channel and others. An impact factor is incorporated into the value function to quantify the effects of actions in one channel on the performance of others. This in-depth analysis of synergistic and competitive interactions between channels enables the model to recommend budget allocations that enhance advertising effectiveness.
- Optimization: With the foundation of Q functions for each channel, the model advances to the optimization stage, where the focus shifts to strategically distributing the budget to maximize expected rewards. By employing an enhanced DE algorithm, the model identifies the optimal sub-interval for budget allocation for each channel based on Q values. This optimization is conducted within the constraints of the overall budget and the stipulation that only one budget level can be chosen for each channel. This critical step ensures that the finite budget is allocated to leverage the discretized

cost space and the inter-channel insights, aiming for the highest possible return on investment (ROI).

The effectiveness of the approach was assessed through empirical evaluations using two datasets: CICAD and CRITEO-UPLIFT v2. Performance was measured using key metrics like the AUCC and the EOM. The proposed model demonstrated significant strength, achieving an AUCC of 0.734, an EOM of 0.760 with the CICAD dataset, an impressive AUCC of 0.801, and an EOM of 0.801 for the CRITEO-UPLIFT v2 dataset. These results significantly surpass those of existing models in the field.

This study makes three principal contributions:

- Enhanced Q-learning reward function for multi-channel interactions: The intricate challenges of interactions among various decision-making channels were acknowledged. In response, the Q-learning reward function was meticulously revamped and augmented, enabling effective navigation and accommodation of these complexities.
- Innovative discretization technique for complexity reduction: A unique technique was unveiled to curb systemic complexity. This method paves the way for a more streamlined and optimized system representation by transforming all cost data points into a sequence of subintervals.
- Advanced mutation strategy within the DE framework: The study pioneers a clustering-based mutation technique, exploiting key groupings within the DE population to generate novel and practical solutions.

The remainder of the article follows this structure: Section II provides a brief literature review. Section III introduces the basics, while Section IV delves into the proposed approach in more detail. Section V contains the experimental results, along with relevant analysis and evaluations. Finally, Section VI concludes the paper.

II. RELATED WORK

Budget allocation is a critical issue in economics, attracting substantial interest from researchers. Given marketing's dynamic nature, budget allocation strategies are developed to optimize specific objectives, like increasing sales, while adhering to certain limitations. This section will review the relevant literature on budget allocation, categorized into two main areas: single-channel and multi-channel approaches.

A. SINGLE CHANNEL

Recently, research has extensively explored budget allocation strategies, particularly in single channels. Zhou et al. [22] put forward a groundbreaking approach that melds machine learning (ML) and operations research (OR) more intimately. Their methodology introduces a pivotal decision factor that connects ML forecasts with OR optimization, effectively mitigating the inaccuracies typically associated with conventional bifurcated approaches. The problem arises in ensuring the decision factor accurately reflects the complexities of real-world marketing scenarios and effectively bridges the gap between ML predictions and OR solutions. Jannink et al. [23] presented a cutting-edge strategy for optimizing budget distributions within breeding programs, harnessing the power of Bayesian optimization coupled with stochastic simulations. When applied to a clonal crop breeding framework, their approach underscored the importance of equitable budget distributions across various stages and unearthed intricate budgetary decision dynamics. The challenge lies in adapting this strategy to other domains where the parameters and constraints differ significantly from breeding programs. Zhang et al. [24] unveiled a technique for fine-tuning expansive pre-trained language models, ingeniously allocating parameter updates in alignment with the significance of weight matrices. Employing singular value decomposition for incremental updates, their strategy enhances performance by judiciously managing parameter budgets, especially in scenarios with many downstream tasks. The difficulty here is determining the importance of weight matrices and how they impact the model's overall performance. Ai et al. [25] introduced the Large-Scale Budget-Constrained Causal Forest (LBCF) algorithm, ingeniously crafted to navigate the Budget-Constrained Treatment Selection (BTS) challenge across vast datasets. Their methodology marked significant progress, particularly noticeable on extensive video platforms. However, the challenge is ensuring the algorithm's scalability and adaptability to different data types and constraints. Hao et al. [26] conceptualized a dynamic knapsack model to elevate sequential advertising strategies within the e-commerce domain. Their innovative dual-level optimization framework, augmented by a technique to streamline the action space, demonstrated unparalleled proficiency in maximizing cumulative revenue compared to existing strategies. The primary issue is the complexity of implementing and adjusting the model in a fast-paced e-commerce environment. Betlei et al. [27] revealed an innovative approach to Uplift Modeling, dedicated to prioritizing individuals based on the potential benefits of interventions like medical prescriptions or guidance. The foundation of their methodology is the AUUC-max learning objective, meticulously honed for AUUC and validated through empirical datasets. A significant problem is ensuring the model's predictions accurately reflect individual responses to treatments in diverse populations. Du et al. [28], tackling the crucial challenge of user retention on user-centric online platforms, proposed a novel framework optimized for heterogeneous treatment effects. They advanced two algorithms, with the latter ingeniously integrating estimation with optimization, specifically designed for enhanced targeted effects. The difficulty lies in accurately estimating and optimizing for heterogeneous treatment effects across a diverse user base. Künzel et al. [29] pioneered a meta-algorithm, the X-learner, proficient in deducing conditional average treatment effects (CATE). Relying on foundational algorithms like random forests (RFs) and Bayesian additive regression trees (BARTs), its effectiveness was confirmed through simulations and real-world applications, with their insights encapsulated in a user-friendly software package. The problem is ensuring the meta-algorithm's adaptability and accuracy across varied applications and datasets. Athey and Wager [30] introduced a forward-thinking methodology to derive treatment assignment protocols from observational data, considering vital factors such as budget constraints, equity, and simplicity. Their approach, applicable to binary and continuous treatments, comes with solid assurances regarding policy effectiveness. The challenge is in applying this methodology in practice, considering the complexities of real-world data and the need for fairness and simplicity in treatment assignments. Xiao et al. [31] crafted an innovative solution for sequential incentive marketing, meticulously calibrating incentive distribution within budgetary limits. Their sophisticated learning algorithm, which integrates bisection search with model-centric planning, is framed as a constrained Markov decision process (CMDP). Its efficacy has been demonstrated through empirical validation. The critical issue is the algorithm's ability to adapt to changing market dynamics and consumer behaviors while staying within budgetary constraints.

Single-channel methods have made significant strides in budget allocation but often need to capture the synergistic potential that multi-channel integration can offer. The current marketing landscape, with its diverse channels each offering unique benefits, suggests that a more holistic approach could significantly amplify the impact of advertising campaigns. Paying attention to channel strategies may lead to less-than-optimal outcomes by neglecting the interplay between channels.

B. MULTI-CHANNEL

Budget optimization across multiple advertising channels has yet to be extensively explored in academic research. Shen et al. [32] introduced a novel cross-channel advertising budget allocation framework to optimize budget distribution across various channels to maximize overall conversions. This framework stands out by addressing the competition among advertisers and employing an iterative algorithm with an entropy constraint for rapid convergence and straightforward implementation in large-scale online advertising systems. However, the challenge here is effectively modeling the competitive dynamics among advertisers to ensure that the framework's global optimization strategy leads to actual improvements in conversion rates across the board. Zhou et al. [22] proposed a novel method for solving marketing resource allocation problems that bridge the gap between machine learning (ML) and operations research (OR). Traditionally, these domains have worked in isolation, with ML predicting model parameters that OR then use for optimization, often leading to compounded errors. By introducing a decision factor and a custom loss function, their approach directly informed OR solutions through simple sorting or comparison operations on this factor, allowing for direct heterogeneous causal learning and unbiased estimations upon loss convergence. Deng et al. [7] investigated how to maximize total conversion in digital advertising by navigating through ROI and budget constraints across multiple channels. They found that optimizing per-channel budgets instead of ROIs leads to optimal global conversion. To facilitate this, they introduced an efficient learning algorithm that helps set these budgets, closely approximating the global optimal conversion, even with limited information about the ad auctions, thus mirroring real-world advertising scenarios. They address the inherent uncertainty and need for control in digital advertising, where advertisers cannot predict or directly influence the outcome of individual ad auctions across channels. Kou et al. [33] created the vector evaluation genetic algorithm (VEGA), crafting budget allocation rules for simulation optimization issues. They recast VEGA's selection dilemma with an optimal computing budget allocation method, deriving an asymptotically optimal rule and a practical approximation. These rules, tested against existing ones, enhance VEGA's efficacy in multi-objective problems by optimizing the search process in stochastic settings. Luzon et al. [34] introduced a method for optimizing budget allocation in social network advertising campaigns by dynamically targeting budget distribution over time based on an effectiveness function. This function relates advertising spend to user exposure, allowing for optimal campaign duration and segment exposure within a given budget. A significant limitation of this method is its reliance on accurate effectiveness function estimation, which can be challenging in dynamic and unpredictable market environments. Li et al. [35] proposed improving differential privacy (DP) k-means clustering by ensuring convergence and usability through optimal privacy budget allocation. This approach reformulates budget allocation as a combinatorial optimization problem and uses genetic algorithms for efficient solution finding. A key challenge is the NP-hard nature of selecting an optimal privacy budget strategy, complicating the balance between privacy protection and data usability. Zhang et al. [6] developed a sophisticated hierarchical framework for online advertising budget allocation. This framework effectively connects decisions across several levels of hierarchy. At the broadest level, it allows advertisers to distribute their budgets across different markets, similar to the approach in this study. However, Zhang et al.'s method relies heavily on lower-level decisions, setting it apart from the proposed approach. The primary issue here is ensuring that the decisions made at each level of the hierarchy are coherent and lead to the overall strategic goal of budget optimization despite the complexity of interactions between different levels. Wang et al. [5] offered a cutting-edge solution to the complex challenges of allocating budgets in multi-channel advertising. Their Q-MCKP algorithm combines the advantages of Reinforcement Learning with the Multi-Choice Knapsack Problem, adeptly facilitating sequential decision-making, delineating inter-channel actions, and managing budget constraints effectively. The challenge tackled by Wang et al. is the dynamic nature of budget allocation in

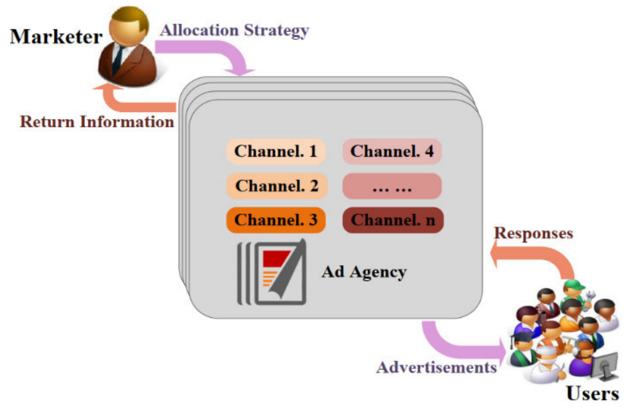


FIGURE 1. The procedure of conducting advertisements across multiple channels (inspired by [36]).

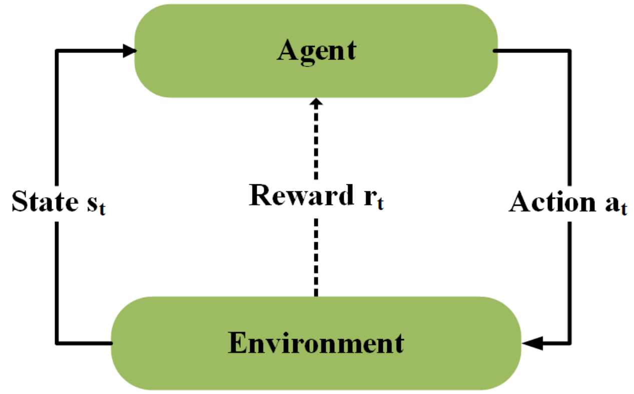


FIGURE 2. Overview of the MDP framework utilized in RL (inspired by [38]).

a multi-channel environment, where decisions need to be continuously adapted based on the evolving landscape of channel performance and budget constraints.

Existing multi-channel methods often need help efficiently allocating budgets across diverse channels due to the need for decision-making that spans the entire lifecycle of advertising campaigns. These methods must contend with the nuances of carryover effects and the influence between channels to maximize long-term benefits beyond immediate gains. Additionally, the limitations of fixed budgets necessitate a thoughtful distribution of resources to achieve the best overall outcomes. RL is proposed as a promising solution in response to these challenges. RL is particularly effective at dealing with these complexities by continuously refining strategies through feedback. This allows adjustments to be made in response to changing market conditions and channel interactions, ultimately optimizing the interplay among channels while adhering to budget constraints.

III. PREREQUISITES

This section provides some background, including the problem statement, reinforcement learning, and differential evolution.

A. PROBLEM STATEMENT

As depicted in Figure 1, within the given context, a marketing professional must judiciously distribute her advertising budget across multiple channels, given that the total budget is pre-determined. The initial step involves the marketer formulating a budget allocation plan for an advertising agency. Following this, the agency will distribute advertisements to diverse users via pertinent channels per the devised strategy. Subsequently, the marketer will receive feedback on each channel's performance, omitting user-specific details. This will be followed by the marketer receiving the results from each channel sans any user-related information. Each channel accrues a cost when information is submitted, necessitating the marketers to have the competency to allocate budgets based on the marketing channels' conditions.

This process involves a series of decision-making steps. Suppose at a given time t, the state of channel i is represented by s, and a marketer is allocated a fixed budget B to promote an advertisement across a collection of n channels. Let $x_t = (x_1, x_2, \ldots, x_n)$ symbolize the budget division, where x_i is the budget portion assigned to channel i. Let $R_i(s_{i,t}, x_{i,t})$ denote the projected gains from channel i when $x_{i,t}$ is allocated. The aim is to tackle the optimization issue of budget distribution:

$$\max_x \sum\nolimits_{i=1}^{n} R_i(s_{i,t}, x_{i,t}) \; s.t \sum\nolimits_{i=1}^{n} x_{i,t} \le B \tag{1}$$

Essentially, the objective is to skillfully allocate budget B across n channels to optimize the total expected returns per time. Before any allocation, it is crucial to understand the expected returns for each sub-budget based on historical data. However, the sequential nature of the advertising process necessitates a series of advertising actions over time, where sub-budgets for each channel are selected to maximize aggregate benefits. In an ideal sequential advertising approach, each allocation strategy would be devised to optimize the expected benefits. The complexity of the sequential advertising approach stems from the restriction that information about future events is only available after an unavoidable delay. To overcome this obstacle, the RL is employed.

B. REINFORCEMENT LEARNING

RL is a machine learning paradigm in which an agent learns to make decisions by interacting with an environment. The ultimate goal of RL is for the agent to acquire a policy, which is a mapping from states to actions that maximize the cumulative reward obtained over time. The agent explores the environment, takes action, receives feedback through rewards, and uses this feedback to refine its policy and enhance its decision-making abilities [37].

In RL, algorithms leverage the framework of a Markov Decision Process (MDP) to determine optimal policies (See Figure 2). An MDP is characterized by a quintuple (S, A, P, ρ_0, r), where S denotes the state space, and A

represents the action space [39]. These sets define the conceivable states and available actions, respectively. The transition dynamics distribution, denoted as $P : S \times A \times S \rightarrow R$, encapsulates the probabilities of transitioning from one state to another upon executing a specific action. The initial state distribution, $\rho_0 : S \rightarrow R$, indicates each state's probability of commencing the process. Additionally, the reward function, $r : S \times A \rightarrow R$, outlines the immediate reward the agent receives upon executing a particular action in a specified state. Throughout the interaction, at each timestep t, the agent is confronted with a state s_t and decides by selecting an action a_t based on the policy π: S × A → [0, 1]. Subsequently, the environment responds by providing a reward $r(s_t, a_t)$ based on the current state and chosen action, leading to a transition to the subsequent state s_{t+1}, as dictated by the dynamics of the environment. The discounted return at timestep t can be expressed as $R_t = \sum_{k=t}^{\infty} \gamma^{k-t} r(s_k, a_k)$, where γ denotes the discount factor. Q-values, which represent the anticipated outcomes of policy π upon executing action a within state s, signify the quality of state-action interactions. This can be computed as shown in Equation 2 [40]:

$$Q^{\pi}(s, a) = E[R_t | s_t = s, a_t = a, \pi] \tag{2}$$

The optimal action-value function, represented as the highest anticipated reward among all strategies after observing state s and performing action a, is calculated as shown in Equation 8 [40]:

$$Q^{*}(s, a) = max_{\pi} E[R_t | s_t = s, a_t = a, \pi] \tag{3}$$

This function embodies the Bellman equation, which posits that the maximum anticipated outcome for a specific maneuver is the sum of the rewards from the current maneuver and the maximum expected outcome from subsequent maneuvers in the next instance. This concept is illustrated in Equation 4 [40]:

$$Q^{*}(s, a) = E[r + \gamma max_{a'} Q^{*}(s', a') | s_t = s, a_t = a] \tag{4}$$

The calculation of the optimal action-value function is conducted incrementally using the Bellman equation, as illustrated in Equation 5 [40]:

$$Q_{i+1}(s, a) = E[r + \gamma max_{a'} Q_i(s', a') | s_t = s, a_t = a] \tag{5}$$

The RL algorithm, called Q-learning, is one of the methods used to estimate optimal value functions online. It commences with initial Q-value estimates for each state and proceeds to update these estimates at each time step according to the following formula [40]:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha\left(r_{t+1} + \gamma max_{a'} Q\left(s_{t+1}, a'\right) - Q(s_t, a_t)\right) \tag{6}$$

In the particular scenario, the attribute vector associated with a specific channel at a given time is used to represent the state of that channel. Its cost and revenue (in terms of registrations) can be regarded as the action and reward value, respectively. Therefore, in situations where channel i is in state $s_{i,t}$, the expected gains from different sub-budgets $R_i(s_{i,t}, x_{i,t})$, can be calculated using the value function $Q_i(s_{i,t}, a_{i,t})$, where $a_{i,t}$ represents the sub-budget $x_{i,t}$.

C. DIFFERENTIAL EVOLUTION

Differential evolution (DE) [41] has gained substantial recognition for its impressive performance in tackling many optimization problems, solidifying its position as a powerful population-based approach. DE functions via three core operations: mutation, crossover, and selection. The algorithm starts with an initial population, typically generated from a uniform distribution, forming the basis for the subsequent evolutionary process. In DE, the mutation operation holds significant importance. It generates a mutated vector, injecting diversity and exploration into the population. During mutation, fresh candidate solutions arise by perturbing existing population members. This perturbation combines information from multiple members to craft a new candidate solution primarily using vector arithmetic. Specifically, it involves multiplying the difference between two randomly selected members by a scaling factor and adding it to a base member. The mutation operation is central to preserving population diversity and promoting exploration. DE effectively navigates the optimization landscape by introducing novel solutions and surmounting local optima. The quality and diversity of the mutation vector substantially impact the performance of DE and its ability to converge toward an optimal solution.

Here is the mutation operator responsible for creating a mutated vector [42]:

$$\vec{v}_{i,g} = \vec{x}_{r_1,g} + F(\vec{x}_{r_2,g} - \vec{x}_{r_3,g}) \tag{7}$$

where $\vec{x}_{r_1,g}$, $\vec{x}_{r_2,g}$, and $\vec{x}_{r_3}$ denote three distinct candidate solutions randomly chosen from the current population, while F represents a scaling factor. In the crossover phase, the mutant and target vectors are combined, typically using the Binomial crossover technique [42]:

$$u_{i,j,g} = \begin{cases} v_{i,j,g} & \text{if rand } (0, 1) \leq \text{CR or j} = j_{rand} \\ x_{i,j,g} & \text{otherwise} \end{cases} \tag{8}$$

where CR represents the crossover rate, j_{rand} is a randomly selected integer from the set $\{1, 2, \ldots, D\}$, where D represents the dimension of the candidate solution. After the crossover, the selection operation chooses the better solution from the target and trial vectors.

IV. THE PROPOSED MODEL

This article presents a novel method for optimizing budgets across multiple advertising channels, leveraging RL and DE. The model is built on three foundational elements: Discretization of cost space, analysis of inter-channel effects, and Optimization.

A. DISCRETIZATION OF COST SPACE

As previously indicated, the primary task involves solving the Q function for each channel. Nonetheless, when dealing with a specific state, accurately approximating the Q value for each sub-budget within the range of $[0, B]$ is challenging. Hence, it is a logical approach to partition the cost space into finite sub-intervals, wherein the average cost of each interval represents an action. This strategic maneuver significantly trims down the expanse of the action space, thereby rendering the Q function as a step function.

For a given channel i, $Q_i(s_{i,t}, a_{i,t})$ stands as the Q function encompassing action $a_{i,t}$ and state $s_{i,t}$,. Additionally, a set of thresholds, denoted as $\omega_{i,j}$, is considered, where crossing a threshold results in a sudden increment within $Q_i\left(s_{i,t}, a_{i,t}\right)$. Formally, the definition of $Q_i\left(s_{i,t}, a_{i,t}\right)$ can be articulated as follows:

$$Q_i\left(s_{i,t}, a_{i,t}\right) = \begin{cases} Q_{i1}\left(s_{i,t}\right) & \omega_{i,0} < a_{i,t} < \omega_{i,1} \\ Q_{i2}\left(s_{i,t}\right) & \omega_{i,1} < a_{i,t} < \omega_{i,2} \\ Q_{i3}\left(s_{i,t}\right) & \omega_{i,2} < a_{i,t} < \omega_{i,3} \\ & . \\ & . \\ & . \\ Q_{in_i}\left(s_{i,t}\right) & \omega_{i,n_i-1} < a_{i,t} < \omega_{i,n_i} \end{cases} \tag{9}$$

In this context, a set of j_i+1 thresholds exist, denoted by $\{\omega_i\}_{j=0,1,\ldots,n_i}$, and a set of n_i Q functions, represented as $\{Q_{ij}\left(s_{i,t}\right)\}_{j=1,1,\ldots,n_i}$.

A significant challenge arises in achieving an accurate division of the cost for each channel within the range of [0, B], necessitating the precise determination of the threshold ωl. For most marketers, daily advertising strategies across each channel tend to be similar, leading to the accumulation of data within a limited number of intervals. Therefore, it is aimed at the discretized sub-intervals comprehensively capturing the relevant details while avoiding an overly restricted number of intervals. This method ensures a satisfactory and productive scope for exploration. In other words, the immediate earnings within identical sub-intervals should exhibit similarity, and the data distribution across each interval must aim for the most uniform dispersion possible. These insights introduced a new assessment metric called the Dispersive Coefficient (DC) [43]. This index helps determine whether the discretization of an interval is necessary. Given a scenario where mj data instances exist within an interval denoted as $B_j = \{(c_1, r_1), (c_2, r_2), \ldots, (c_{m_j}, r_{m_j})\}$ (where c_j and r_j represent cost and revenue, respectively), DC for this interval is defined as follows:

$$\mathrm{DC}\left(B_j\right) = \sqrt{\sum\nolimits_{i=1}^{m_j} \frac{(r_i - \bar{r})^2 + (c_i - \bar{c})^2}{m_j}} + \mu\sqrt{m_j} \tag{10}$$

where $\bar{r}$ and $\bar{c}$ represent the mean values of the costs and revenues derived from the m_j data, while μ is a balancing parameter. The initial expression gauges the extent of scattering within an interval. In contrast, the subsequent expression is formulated to regulate the volume of data encompassed within the interval and to avert excessive concentration within any given interval. DC is a crucial tool for assessing the need for discretization. It helps optimize the number of intervals by analyzing the similarity of earnings and uniformity of data distribution within each sub-interval. This, in turn, assists in devising a more informed and effective advertising strategy across different channels.

With this understanding of the DC, the approach for discretizing costs is presented. This methodology starts by considering a single interval $[0, B]$ and then divides this interval into, on average, L/10 sub-intervals, thereby establishing the initial collection of segmented sub-intervals. In each subsequent cycle, a binary query is sent to each sub-interval, and the set of sub-intervals is adjusted based on the outcome of the DC function. As additional iterations progress and the collection of sub-intervals is further refined, the typical result is the acquisition of the finalized set of sub-intervals, denoted as φ. Additionally, θ represents a parameter responsible for governing the quality of the solution. The pseudocode outlining the proposed approach is depicted in Algorithm 1.

B. ANALYSIS OF INTER-CHANNEL EFFECTS

In the investigation, the entire lifecycle of advertisements (ads) spanning channels i and j was examined. The objective was to decipher these diverse channels' intricate dynamics and reciprocal influences. When an advertisement, denoted as A_{ik}, is showcased on channel i at time instance t_k, its ramifications are twofold. On one hand, it produces lingering effects on succeeding ads within the same channel i. Concurrently, it also influences ads broadcasted on channel j. Given this dual impact, it is paramount to incorporate both dimensions when quantifying the anticipated rewards linked to A_{ik}.

The discounted return, denoted as Rt, represents the aggregated rewards exclusively from channel i and warrants modifications. Within the framework of channel i, it becomes essential to embed a key construct: the inclusion of an impact factor $\acute{Q}_t$ (s, a) within the value function. This factor contemplates ramifications on auxiliary channels when an action a is operationalized in a given state s. The structural representation of the impact factor is delineated as follows:

$$\acute{Q}_t\,(\mathrm{s}, \mathrm{a}) = \sum\nolimits_{j \in I \setminus \{i\}} \sum\nolimits_{k=0}^{\infty} \varepsilon_{i,j,t+k+1} \cdot \gamma_i^k \cdot r_{j,t+k+1} \tag{11}$$

$$\varepsilon_{i,j,t+k+1} = \frac{a_{i,t+k+1}}{a_{j,t+k+1}} \cdot d, \quad d < 1 \tag{12}$$

where I denotes the set of channels, γ_i represents the discount factor for channel i, $\varepsilon_{i,j,t+k+1}$ acts as a coefficient multiplied by a value signifying the ratio of the cost of channel i relative to the cost of channel j at time $t + k + 1$. Additionally, d is a positive scalar value less than one, and $r_{j,t+k+1}$ signifies the immediate reward obtained from channel j at time $t + k + 1$. Hence, the updated Q_i^* for channel i can be derived using the

Algorithm 1 Pseudo-Code of Discretization Approach for Costs

```
Input: L: maximum number of intervals, B: budget, P: maximum number of iterations
Output: φ: {(ω_j, ω_{j+1}]}_{j=0,1,...,l} (l < L)
l = 0, φ = Null
for j = 0 to ((L/10) − 1) do
  ω_j = (j * B)/(L/10)
  ω_{j+1} = ((j + 1) * B)/(L/10)
  φ = φ ∪ (ω_j, ω_{j+1}]
  l = l + 1
 while l < L & p < P do
  for every (ω_j, ω_{j+1}] ∈ φ do
    dc = DC(B_j)
    if dc > θ then:
      ω' = (ω_j + ω_{j+1})/2
      φ = φ − (ω_j, ω_{j+1}]
      φ = φ ∪ (ω_j, ω']
      φ = φ ∪ (ω', ω_{j+1}]
      l = l + 1
 p = p + 1
```

following equation:

$$Q_i^*(s,a) = E_\pi\left[\sum_{k=0}^{\infty}\gamma_i^k r_{t+k+1} + \sum_{j\in I\setminus\{i\}}\sum_{k=0}^{\infty}\varepsilon_{i,j,t+k+1}\cdot\gamma_i^k r_{j,t+k+1} \mid s_t = s, a_t = a\right]$$
$$= E_\pi\left[r_{t+1} + \sum_{j\in I\setminus\{i\}}(\varepsilon_{i,j,t+1}.r_{j,t+1}) + \gamma_i Q_i^*(s_{t+1}, a') \mid s_t = s, a_t = a\right] \quad (13)$$

The value of channel i at time t is modified as follows:

$$Q_i(s_t, a_t) \leftarrow Q_i(s_t, a_t) + \alpha\left(r_{t+1} + \sum_{j\in I\setminus\{i\}}(\varepsilon_{i,j,t+1}.r_{j,t+1}) + \gamma_i \max_{a'} Q_i(s_{t+1}, a') - Q_i(s_t, a_t)\right) \quad (14)$$

C. OPTIMIZATION

With the Q function of each channel now in possession, the next step is budget allocation across channels to maximize anticipated rewards at time t. To achieve this, an enhanced DE algorithm is used to determine the appropriate interval j (where the average cost within this interval is considered an action) for allocating the budget to each channel i. This decision is encoded through a binary variable, denoted as y_{ij}, which assumes a value of 1 when the budget corresponding to threshold level j for channel i is assigned.

$$\max_{y_{i,j}\in\{0,1\}} \sum_{i=1}^{n}\sum_{j=1}^{J_i} Q_{i,j}(s_{i,t})y_{i,j} \quad (15)$$

$$st.\ \sum_{i=1}^{n}\sum_{j=1}^{J_i} a_{i,j}y_{i,j} \le B,\ \sum_{j=1}^{j_i} y_{i,j} \le 1,\quad \forall i \in \{1, 2, \ldots, n\} \quad (16)$$

The initial inequality represents the budget limitation, while the subsequent condition indicates that, for any given channel, only one budget level can be selected at most. Moreover, $a_{i,j}$, serving as an actionable parameter, denotes the mean cost within the j-th interval of channel i. J_i represents the number of intervals associated with channel i.

1) THE IMPROVE DE

An improved DE algorithm is used to optimize Equation 15. The mutation operator is refined, drawing inspiration from the improvements suggested by [44]. A potential region in the search space is identified using k-means clustering on the existing population. This clustering divides the population into k distinct regions in the search space, with the value of k randomly chosen between $[2, \sqrt{N}]$. The cluster that shows the lowest mean objective function value is subsequently recognized as the region of interest.

A novel mutation operator based on clustering is employed, characterized as:

$$\overrightarrow{v^{clu}{}_l} = \overrightarrow{win_g} + F(\overrightarrow{x_{r_1,g}} - \overrightarrow{x_{r_2,g}}) \quad (17)$$

where $\vec{x}_{r_1,g}$ and $\vec{x}_{r_2,g}$ denote two randomly selected candidate solutions, while $\overrightarrow{win}_g$ signifies the optimal candidate solution within the identified promising region. However, it is crucial to highlight that $\overrightarrow{win}_g$ might not always represent the most optimal solution in the current population. The mutation procedure, influenced by clustering, is executed repeatedly for M instances. Following this, the population undergoes updates in alignment with the Generic Population-Based Algorithm (GPBA) [45], adhering to the subsequent guidelines:

- Selection: k candidate instances are randomly produced and regarded as the initial starting points for the k-means algorithm.
- Generation: M candidate instances are produced through mutation based on clustering and are designated as the set v^{clu}.
- Replacement: M candidate instances are chosen randomly from the existing population and designated as set B.
- Update: the top-performing M candidate solutions are chosen from $v^{clu} \cup B$ to create set B'. In conclusion, the fresh population is acquired by combining $(P - B) \cup B'$.

Algorithm 2 Pseudo-Code of the Improved DE Algorithm

Input: D: dimensionality of candidate solution, $MaxFES$: maximum number of function evaluations, F: scaling factor, CR: crossover probability
Initialize the population P as follows: $P = (\vec{x}_{r_1}, \vec{x}_{r_1}, \ldots, \vec{x}_{NP})$, where each $\vec{x}_{r_i}$ is randomly generated
Compute the objective function value for each solution in the population P
while $FES < MaxFES$ **do**
  for $i = 1$ *to* NP **do**
    Select individuals $\vec{x}_{r_1,g}, \vec{x}_{r_2,g}, \vec{x}_{r_3,g}$ from P randomly ($\vec{x}_{r_1,g} \neq \vec{x}_{r_2,g} \neq \vec{x}_{r_3,g}$)
    $\vec{v}_{i,g} = \vec{x}_{r_1,g} + F(\vec{x}_{r_2,g} - \vec{x}_{r_3,g})$
    Choose j_{rand} as a random number within the range [0, 1]
    for $j = 1$ *to* D **do**
      if $rand(0, 1) \leq CR$ **or** $j == j_{\text{rand}}$ **then**
        $u_{\text{i,j,g}} = v_{\text{i,j,g}}$
      else:
        $u_{\text{i,j,g}} = x_{\text{i,j,g}}$
    if $f(u_{i,g}) < f(x_{i,g})$ **then**
      $x_{\text{i,g+1}} = u_{\text{i,g}}$
    else:
      $x_{\text{i,g+1}} = x_{\text{i,g}}$
Choose a value for k randomly from the interval $[2, \sqrt{N}]$
Divide the population P into k clusters
Calculate the mean objective function value for all clusters
Determine the value of $\overrightarrow{win}_g$ as the best solution within the winning cluster
for $i = 1$ *to* M **do**
  Choose $\vec{x}_{r_1,g}$ and $\vec{x}_{r_2,g}$ from P randomly ($\vec{x}_{r_1,g} \neq \vec{x}_{r_2,g}$)
  $\overrightarrow{v^{clu}}_{\iota} = \overrightarrow{win}_g + F(\vec{x}_{r_1,g} - \vec{x}_{r_2,g})$
Randomly select M candidate solutions from the population P and designate them as set B
Select the top-performing M solutions from $v^{clu} \cup B$ and designate them as set B'
Generate a new population as $(P - B) \cup B'$

TABLE 1. Summary and features of CICAD spanning 540 days across channels A, B, C, D, and E.

Feature	Description
Day_of_week	The specific day within the week
Cost_i	Cost incurred i day(s) prior ($i = 1$ *to* 7)
Exposure_i	Count of exposures from i day(s) prior ($i = 1$ *to* 7)
Click_i	Count of clicks recorded i day(s) prior ($i = 1$ *to* 7)
Registration_i	Total number of registrations i day(s) prior ($i = 1$ *to* 7)

The pseudo-code representation of the improved DE algorithm can be found in Algorithm 2.

V. EXPERIMENTS

This section describes the used datasets and highlights their characteristics. The benchmarks are defined, and the criteria and assessment methods used to evaluate the performance of the models are explained. The section concludes with a presentation of settings and the results, emphasizing the key findings from the analysis and model evaluations and discussing their relevance in the context of the research objectives.

A. DATASET

This paper presents two types of datasets:

- Chinese Internet Company Advertising Dataset (CICAD) [5]: The dataset is sourced from an authentic internet enterprise based in China, emphasizing user engagement via advertising endeavors. The dataset spans 540 days and captures advertising activities across five distinct channels: A, B, C, D, and E. Their respective record counts are 503, 452, 520, 467, and 503. Each data entry encapsulates the advertising date, the incurred cost, the aggregate of clicks generated, the sum of exposures (often termed impressions), and the total registrations accrued. Multiple time-oriented features were devised to provide a nuanced understanding of the channel dynamics at various temporal junctions. These intricacies are delineated in Table 1. In representing the state of each channel, 29 unique features are deployed. The actions associated with each channel are extrapolated using the discretization approach described in Algorithm 1, with the reward being a function of the registrations achieved.
- CRITEO-UPLIFT v2: The dataset provided by the AdTech firm Criteo was featured in the AdKDD'18 workshop [46]. It stems from a randomized control trial (RCT) in which a selected group of users was intentionally withheld from advertising exposures. The dataset encompasses 12 attributes, including a binary treatment marker and two response tags about visits and conversions. In the context of cost-unaware uplift modeling, the incremental visit serves as the primary predictive target.

B. METRIC

In evaluating the proposed method's performance, metrics were carefully selected to provide comprehensive insights into the effectiveness and efficiency of the models. These metrics include:

- AUCC (Area Under the Cost Curve): The AUCC is a widely recognized metric in existing research, as evidenced by its frequent citation in seminal papers [22], [28]. Its primary function is to assess the ability of uplift models to rank outcomes in scenarios where treatments are binary correctly. The AUCC's significance lies in its focus on the incremental impact of an intervention, making it particularly relevant for this study, which aims to optimize budget allocation for maximum effect. The methodology for evaluating AUCC is thoroughly discussed in key literature [28], providing a robust framework for the analysis.
- EOM (Expected Outcome Metric): Complementing the AUCC, the EOM is another metric extensively used in recent studies [25], designed to quantitatively predict the expected result of a specific budget allocation strategy. Depending on the study's focus, this could be response rate, revenue generation, or cost efficiency. The strength of EOM lies in its empirical basis, allowing for the derivation of unbiased estimates of outcomes from randomized controlled trial (RCT) data. This attribute of EOM makes it exceptionally adaptable and applicable across various scenarios, offering a broader perspective on the potential real-world impact of budget allocation strategies.
- ROAS (Return on Advertising Spend): ROAS is an indispensable metric in advertising analytics, offering a direct measure of the financial effectiveness of advertising expenditures. ROAS calculates the revenue generated per dollar spent on advertising as a fundamental indicator, providing an immediate reflection of campaign profitability. This metric is critically essential for assessing the efficiency of budget utilization within specific campaigns and is widely referenced in marketing literature for its straightforward, quantitative evaluation of advertising success [47], [48]. ROAS supports strategic decision-making in real-time advertising adjustments and budget allocation by highlighting the direct correlation between advertising spending and revenue outcomes [49].
- CLV (Customer Lifetime Value): CLV serves as a strategic metric in understanding the long-term value of customers, projecting the net profit attributed to the entire future relationship with a customer. CLV offers a deep dive into customer profitability that transcends transactional interactions by estimating the total revenue a business can expect from a single customer throughout its relationship [50]. This metric is extensively utilized to tailor marketing strategies, prioritize customer segments, and optimize resource allocation toward high-value customers. The comprehensive nature of CLV, encompassing both current and future potential earnings from customers, makes it an essential tool for businesses aiming to maximize lifetime customer value and foster enduring customer relationships. Its utilization is well-documented in numerous studies, underscoring its ability to effectively influence long-term business growth and profitability strategies [51].

The choice of these four indicators was driven by their proven reliability and relevance in measuring the performance of budget allocation models. AUCC provides a nuanced understanding of model ranking capability within a binary treatment context, while EOM extends the evaluation to the anticipated real-world outcomes of these allocations. Together, they offer a balanced and comprehensive assessment framework, aligning with the study's objectives to optimize budget allocation and understand the practical implications of these optimizations in real-world settings. ROAS and CLV are selected to address the complexities and multi-faceted impacts of advertising campaigns more effectively [49]. ROAS is crucial for providing immediate financial feedback on advertising expenditures, offering a clear and quantifiable measure of how effectively each dollar spent contributes to revenue generation. This metric is handy in fine-tuning advertising strategies and ensuring budget allocations are immediately profitable, thus resolving the problem of traditional metrics that overlook the direct financial outcomes of marketing efforts. Furthermore, CLV is included as a strategic metric that extends beyond the short-term horizon to gauge the long-term value generated from customer relationships. By incorporating CLV, the aim is to capture the immediate returns and the projected lifetime profitability of customers acquired through specific advertising strategies. This approach is essential for developing more sustainable marketing strategies that prioritize and optimize resources toward more valuable customer segments, thus addressing the issue of traditional metrics failing to consider campaigns' broader, long-term impacts on customer retention and value [51].

C. SETTING

All experiments were conducted on a 64-bit Windows operating system with 64 gigabytes (GB) of random-access memory (RAM). This significantly enhanced the system's capacity to handle large datasets and perform demanding computations smoothly. It also featured a one-terabyte (TB) solid-state drive (SSD), ensuring fast data access and substantial storage capacity—essential for efficiently managing large volumes of data. This configuration allowed for the smooth running of multiple applications and services simultaneously, meeting the high computational demands of the research with both efficiency and reliability.

A crucial technique, k-fold cross-validation, was employed to optimize hyperparameters for the proposed and comparative models. This technique is a key factor in ensuring the robustness and generalizability of the model. It involves partitioning the data into distinct subsets and training the model on k-1 of these while using the remaining subset for testing. This process is repeated k times, with each subset used exactly once as the test set. By averaging

TABLE 2. Summary of optimized hyperparameters for RL and DE components.

Algorithm	Hyperparameter	Range of values	Optimized value
RL	Learning rate	0.001 to 0.1	0.01
	Discount factor (γ)	0.8 to 0.99	0.95
	Exploration rate (ε)	0.01 to 0.2	0.1
	Number of episodes	100 to 1000	500
	Batch size	32 to 128	64
DE	Population size	20 to 100	50
	Mutation factor (F)	0.5 to 1.0	0.8
	Crossover probability (CR)	0.7 to 1.0	0.9
	Number of generations	50 to 200	100

the results from all k experiments, a more comprehensive assessment of the model's performance across different data segments is obtained, significantly reducing the likelihood of anomalies due to peculiarities in any single split. This approach enhances confidence in the model's performance.

The k-fold cross-validation method in the proposed model helps identify the most effective hyperparameter settings that balance the trade-off between exploration and exploitation in RL and the evolutionary parameters in DE. Furthermore, validating across multiple subsets effectively mitigates the risk of overfitting, ensuring that both algorithms adapt well to new, unseen data and maintain their efficacy across diverse advertising scenarios. Table 2 details the outcomes of these optimization activities for both the RL and DE in the proposed model. It offers a detailed summary of the selected hyperparameters for the RL and DE components, including the ranges and specific values that delivered optimal performance during the validation stage.

D. RESULT

The proposed model was evaluated against seven established methods: OptiMark [2], MCBB [7], DRP [22], AdCob [32], VEGA [33], DySNAdOpt [34], and GAPBAS [35]. It was also compared with three variants: Proposed w/o RL, Proposed w/o DA, and Proposed w/o DE, each lacking a key component—RL, Discretization Approach (DA), and DE, respectively. Table 3 presents a comprehensive comparison of these models across two datasets using AUCC and EOM metrics.

Among the competing models, GAPBAS emerged as the top performer on both datasets, securing AUCC scores of 0.703 and 0.810 on the CICAD and CRITEO-UPLIFT v2 datasets, respectively, alongside EOM scores of 0.645 and 0.757. This progression from OptiMark to GAPBAS illustrates a consistent enhancement in performance, showcasing the evolution of modeling techniques and their increasingly adept handling of multi-channel budget allocation challenges. The proposed model outshines all competitors with remarkable improvement margins. Compared with GAPBAS, the best of the compared models, the proposed model achieved a 20.2% increase in AUCC for CICAD and a 13.1% increase for CRITEO-UPLIFT v2. Similarly, it showed a 33.6% rise in EOM for CICAD and a 16.5% increase for CRITEO-UPLIFT v2. These significant leaps in performance highlight the synergy achieved by integrating RL, DE, and discretization approaches, adeptly navigating the complexities of multi-channel budget allocation to optimize advertising campaign outcomes.

The comparison with its derivative models reveals the integral value of each component. For instance, the proposed model boosts AUCC by 16.5% over the Proposed w/o RL variant and 13.6% over the Proposed w/o DE variant on CICAD. The EOM also saw significant uplifts of 22.4% and 17.1% over the Proposed w/o RL and Proposed w/o DE variants, respectively. This analysis underscores the critical contribution of RL, DE, and discretization in enhancing the model's performance, with each element playing a pivotal role in fine-tuning budget allocations and maximizing conversion rates.

Generally, the results of this study align with and expand upon recent advancements in multi-channel budget allocation and optimization. Similar investigations, such as those conducted by DRP and AdCob, have explored the complexities of distributing budgets across various channels using different methodologies to enhance marketing campaign effectiveness. Our proposed model builds on these foundational studies by incorporating RL, DA, and DE. These integrations have proven to significantly advance the management of the complex challenges associated with such allocations. The effective synergy of these methods reflects the progressive nature of current research and underscores our model's innovative capacity to outperform existing methods significantly. By situating our findings within the broader context of recent innovations, we highlight the significance and impact of our approach in advancing the field of budget allocation in digital marketing.

In the statistical analysis, paired t-tests were conducted to evaluate the significance of the performance differences between the proposed and existing advanced models, specifically focusing on the OptiMark model as a representative comparison. The analysis revealed extremely low p-values across all metrics, with the CICAD dataset showing p-values of 7.09×10^{-12} for the AUCC and 6.77×10^{-14} for the EOM. Similarly, for the CRITEO-UPLIFT v2 dataset, the p-values were 4.71×10^{-12} for AUCC and 2.56×10^{-14} for EOM. These results indicate a statistically significant superiority of the proposed model over the OptiMark model. The remarkably low p-values across both datasets and performance metrics strongly suggest that the enhancements integrated into the model, such as the advanced DE algorithm and the strategic incorporation of RL techniques, substantially contribute to its improved performance.

Table 4 showcases the performance of various advertising models using ROAS and CLV metrics, enabling a comparative analysis between the proposed model and its competitors across the CICAD and CRITEO-UPLIFT v2 datasets. For

TABLE 3. Performance comparison of the proposed model against other models on the CICAD and CRITEO-UPLIFT v2 datasets using the AUCC and EOM metrics.

Method	CICAD		CRITEO-UPLIFT v2	
	AUCC	EOM	AUCC	EOM
OptiMark [2]	0.631±0.019	0.570±0.006	0.719±0.012	0.682±0.011
MCBB [7]	0.652±0.010	0.582±0.006	0.742±0.026	0.716±0.005
DRP [22]	0.659±0.008	0.594±0.016	0.749±0.015	0.716±0.006
AdCob [32]	0.664±0.014	0.603±0.012	0.781±0.006	0.752±0.001
VEGA [33]	0.687±0.025	0.614±0.018	0.789±0.019	0.768±0.026
DySNAdOpt [34]	0.690±0.010	0.630±0.019	0.806±0.016	0.748±0.015
GAPBAS [35]	0.703±0.016	0.645±0.016	0.810±0.023	0.757±0.012
Proposed w/o RL	0.725±0.016	0.704±0.010	0.785±0.016	0.806±0.010
Proposed w/o DA	0.744±0.018	0.726±0.012	0.803±0.019	0.826±0.011
Proposed w/o DE	0.750±0.019	0.736±0.022	0.813±0.018	0.829±0.018
Proposed	0.845±0.005	0.862±0.010	0.916±0.002	0.882±0.003

TABLE 4. Performance comparison of the proposed model against other models on the CICAD and CRITEO-UPLIFT v2 datasets using the ROAS and CLV metrics.

Method	CICAD		CRITEO-UPLIFT v2	
	ROAS	CLV	ROAS	CLV
OptiMark [2]	1.0±0.1	$100±4	1.4±0.2	$105±3
MCBB [7]	1.1±0.2	$105±2	1.8±0.3	$109±1
DRP [22]	1.3±0.1	$108±3	1.5±0.4	$102±9
AdCob [32]	1.4±0.0	$102±6	1.3±0.2	$110±8
VEGA [33]	1.8±0.2	$115±8	1.9±0.1	$106±4
DySNAdOpt [34]	1.9±0.1	$120±6	1.8±0.3	$107±1
GAPBAS [35]	1.8±0.2	$115±7	1.7±0.6	$118±2
Proposed	2.8±0.1	$138±5	3.8±0.0	$159±6

instance, a ROAS of 1.0 for the OptiMark model on the CICAD dataset indicates a break-even scenario where each dollar invested in advertising returns precisely one dollar in revenue.

This suggests that while the campaign is not losing money, it is also not generating any profit, which might be considered a conservative outcome in terms of investment return. On the other hand, the proposed model shows a ROAS of 2.8 on the same dataset, significantly outperforming OptiMark and other models. This indicates that the proposed model generates approximately $2.80 in revenue for every dollar spent on advertising, reflecting a highly efficient use of the advertising budget that substantially increases profitability. Similarly, the proposed model achieves a remarkable CLV of $138 on the CICAD dataset. It suggests that each customer acquired or retained through the campaign is expected to contribute an average of $138 in net profit over their relationship with the company. This is a considerable increase compared to the $100 CLV reported for OptiMark, underscoring the proposed model's superior capability in generating immediate revenue and fostering valuable long-term customer relationships. Furthermore, the variation in ROAS and CLV across different models and datasets highlights the adaptability and performance of the proposed model under various conditions. For example, while the VEGA model shows a competitive ROAS of 1.8 and a CLV of $115 on the CICAD dataset, it does not perform as well on the CRITEO-UPLIFT v2 dataset with a slightly higher ROAS but a lower CLV. In contrast, the proposed model maintains and enhances its performance on the CRITEO-UPLIFT v2 dataset, with a ROAS of 3.8 and a CLV of $159, demonstrating its robustness and effectiveness across different advertising environments. These metrics are critical in understanding the efficiency and impact of each model. A higher ROAS directly relates to better immediate financial returns, while a higher CLV indicates more profitable and sustainable customer relationships. Therefore, the proposed model's superior performance in these metrics suggests it is more effective at utilizing budgets to generate immediate revenue and create long-term value, making it a compelling choice for advertisers aiming to maximize immediate and future returns from their campaigns.

Table 5 compares computational efficiency metrics for various models applied to the CICAD and CRITEO-UPLIFT v2 datasets. The metrics considered are Runtime, measured in seconds, and GPU Usage, measured in gigabytes (GB). The OptiMark model shows a balanced performance on both datasets with moderate runtime and GPU usage. In contrast, the MCBB and GAPBAS models exhibit higher GPU demands, particularly evident in the latter's substantial consumption on the CRITEO-UPLIFT v2 dataset, marking the highest among all models at 10.25 GB. DRP and the proposed model exhibit similar efficiency on the CICAD dataset regarding runtime, but the proposed model leverages GPU resources more effectively. This indicates an optimization that does not compromise the computational

TABLE 5. Performance computational efficiency of the proposed model against other models on the CICAD and CRITEO-UPLIFT v2 datasets.

Method	CICAD		CRITEO-UPLIFT v2	
	Runtime (s)	GPU usage (GB)	Runtime (s)	GPU usage (GB)
OptiMark [2]	2148	7.63	3502	8.13
MCBB [7]	2563	9.65	4500	8.74
DRP [22]	2478	7.52	3150	8.47
AdCob [32]	4586	6.58	5203	7.12
VEGA [33]	3963	10.20	4520	8.75
DySNAdOpt [34]	2569	6.58	4520	7.54
GAPBAS [35]	3560	9.69	3790	10.25
Proposed	3269	8.23	3587	7.52

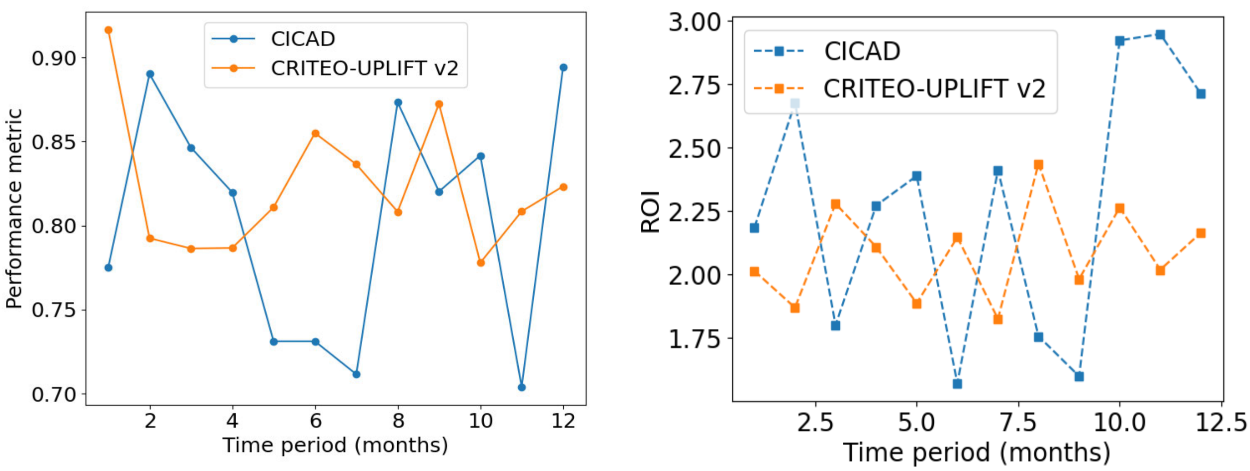


FIGURE 3. Dynamic performance and economic viability of the proposed model over 12 months on CICAD and CRITEO-UPLIFT v2 datasets, illustrating resilience and efficiency in online advertising environments.

time for resource usage. AdCob, while the least efficient in terms of runtime, particularly on the CRITEO-UPLIFT v2 dataset, compensates with the lowest GPU usage, suggesting a trade-off between time and resource consumption. VEGA and DySNAdOpt show contrasting resource allocations, with VEGA consuming the most GPU resources across both datasets, which could imply a more complex but potentially robust computational process. DySNAdOpt, despite a higher runtime on the CRITEO-UPLIFT v2 dataset, manages to keep the GPU usage to a minimum, mirroring the efficiency of AdCob in this regard.

Figure 3 offers an in-depth examination of the proposed model's performance and economic impact over 12 months, utilizing the CICAD and CRITEO-UPLIFT v2 datasets. Figure 3. a represents the model's performance over time. This portrayal vividly illustrates the model's adeptness at adjusting to the ever-evolving landscape of market conditions, advertising trends, and the complex feedback loops inherent in the online advertising ecosystem. The performance across both datasets showcases periodic peaks, suggesting the model's capacity to maintain high-efficiency levels amidst fluctuating market dynamics. This resilience is likely a testament to the model's advanced learning mechanisms, which harness the strengths of RL and DE algorithms. These algorithms work in tandem to continually refine and optimize budget allocation strategies, ensuring the model's sustained high performance despite shifting market trends.

Figure 3. b, focusing on the ROI, further accentuates the proposed model's effectiveness. The ROI trajectories for both datasets display variability, mirroring the complex relationship between advertising expenditure and the resulting conversions or sales. Notably, the ROI consistently exceeds the breakeven point across both datasets, underscoring the model's ability to generate substantial value from advertising investments, thus ensuring positive economic outcomes.

The proposed model's distinct advantage is highlighted through this dual analysis. Its unparalleled performance is attributed to several innovative features, including integrating a clustering-based mutation strategy within the DE algorithm. This approach allows the model to explore and identify optimal budget allocations through novel and practical solutions. Additionally, the model employs a discretization strategy that simplifies the intricate process of budget allocation, enhancing its applicability and effectiveness in real-world scenarios. The model's adaptability, as evidenced

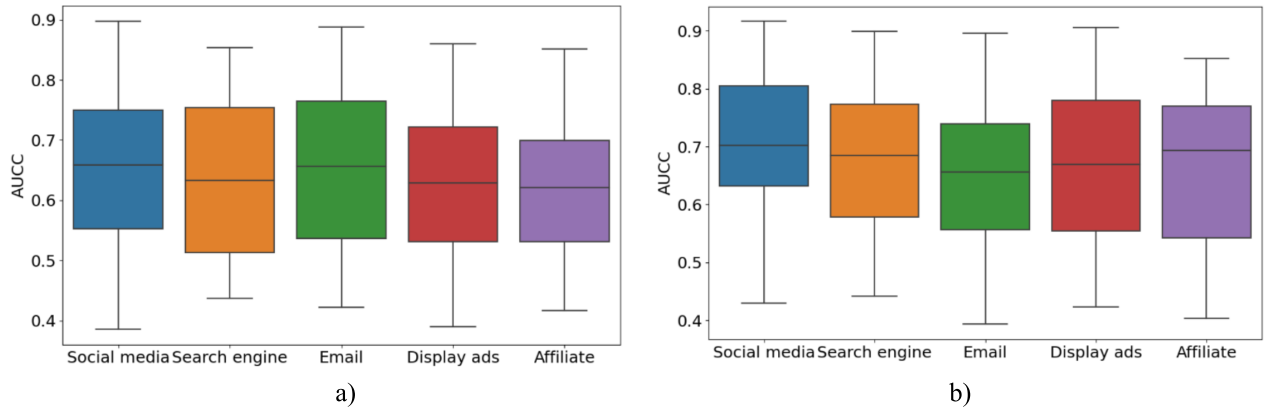


FIGURE 4. Impact of budget reallocation on multi-channel advertising performance. a) original channel performances, b) modified performances with an increased social media budget.

by its sustained performance over an extended period, indicates its proficiency not just in static conditions but also in adapting to dynamic market changes— a crucial attribute in the rapidly changing online advertising landscape. The economic analysis demonstrated through consistent ROI gains confirms the model's practical value, offering advertisers an optimized solution that maximizes campaign impact and ensures favorable financial returns.

Figure 4 skillfully illustrates the dynamics among various advertising channels through two sets of box plots, representing the performance metrics of five primary channels: social media, search engines, email, display ads, and affiliate marketing. Figure 4. a provides a snapshot of these channels' performances before any budgetary changes. Each plot highlights the median performance at its center, while the box around it marks the interquartile range (IQR), capturing the central 50% of the data. The whiskers extending from the boxes shed light on the broader spectrum of outcomes, showcasing the variability beyond the median and quartile data.

Initially, the performances across these channels are notably consistent, with only minimal differences in median values and data spread. However, this consistency is disrupted by a deliberate 50% budget increase for social media, leading to significant shifts in the performance metrics of the other channels (Figure 4.b). This adjustment brings to the forefront the adaptive nature of budget allocation in multi-channel advertising strategies, demonstrating how modifications in one area can profoundly influence the success of others. It emphasizes the interconnectedness of these channels, suggesting that strategic budget reallocations can have far-reaching effects across the advertising landscape.

The changes observed in the performance metrics post-budget adjustment highlight the dynamic and intertwined nature of multi-channel advertising. This complexity demands a thorough comprehension of how these channels interact and influence one another. Advertisers are, therefore, encouraged to take a more expansive view of their budgeting strategies, considering not only the immediate effects but also the broader implications of their financial decisions on all channels.

TABLE 6. Model performance under normal and adversarial conditions across CICAD and CRITEO-UPLIFT v2 datasets.

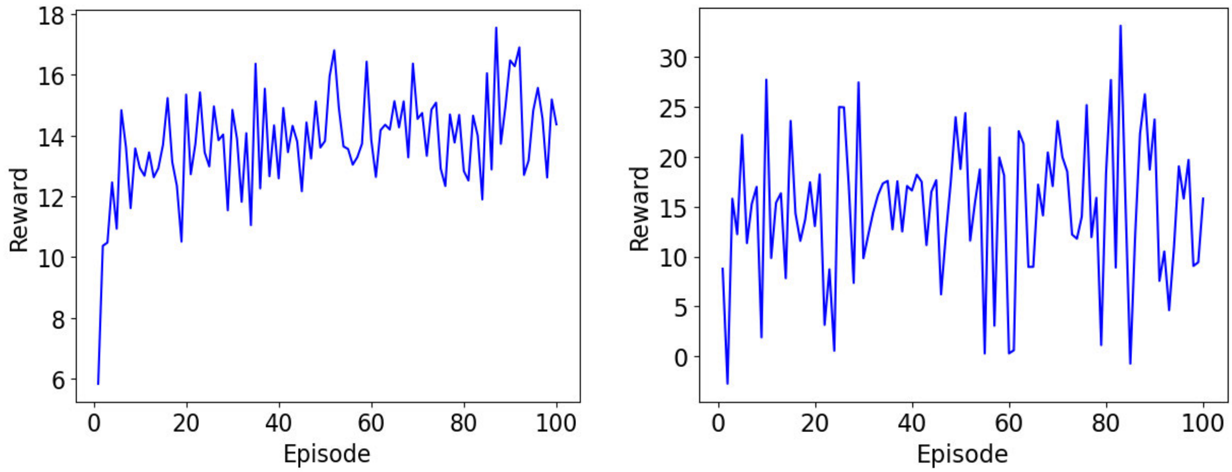
Method	CICAD		CRITEO-UPLIFT v2	
	AUCC	EOM	AUCC	EOM
Normal	0.845	0.862	0.916	0.882
Adversarial	0.820	0.834	0.893	0.874

Table 6 showcases the model's resilience and adaptability by juxtaposing its performance across standard and adversarial scenarios, employing two distinct datasets: the CICAD and CRITEO-UPLIFT v2.

When evaluated on the CICAD dataset under conventional conditions, the model exhibits commendable strength, achieving an AUCC of 0.845 and an EOM of 0.862. However, introducing adversarial examples leads to a modest reduction in its effectiveness, with the AUCC falling to 0.820 and the EOM to 0.834. This decline highlights the model's susceptibility to adversarial tactics but simultaneously underscores its durability, as it continues to perform with notable competence in the face of such disruptions.

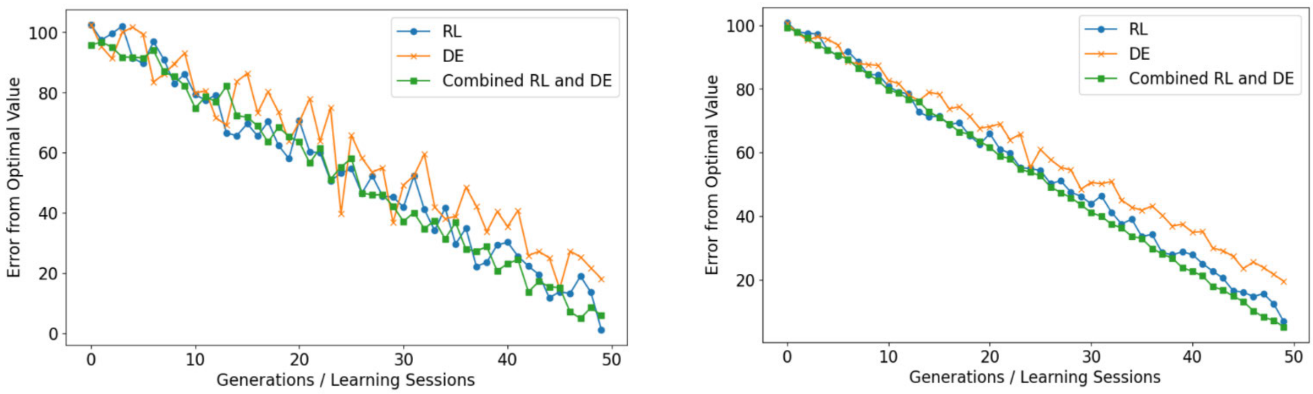
The performance trajectory of the CRITEO-UPLIFT v2 dataset is similar. Under normal testing conditions, the model initially posts robust scores—0.916 in AUCC and 0.882 in EOM. However, when confronted with adversarial examples, its metrics experience a minor decrement, with the AUCC descending to 0.893 and the EOM to 0.874. This consistent performance pattern, even under adversarial pressure, reaffirms the model's robust nature and ability to maintain a high level of effectiveness across various testing landscapes.

Figure 5 depicts the progression of rewards an agent accumulates over a series of episodes for two distinct datasets. In the CICAD dataset, a discernible upward trend suggests a learning curve where the agent gradually improves performance as it gains experience through repeated interactions



a) b)

FIGURE 5. **Reward trajectories over successive episodes for the CICAD dataset (a) and the CRITEO-UPLIFT v2 dataset (b).**



a) b)

FIGURE 6. **Comparative performance analysis of RL, DE, and their integration in minimizing error across generations for a) CICAD, b) CRITEO-UPLIFT v2 datasets.**

with the environment. The rewards start at a lower baseline and exhibit a relatively steady ascent, punctuated by occasional dips that could signify the agent's exploratory actions or environmental changes. In contrast, the CRITEO-UPLIFT v2 dataset displays a more volatile reward path, with higher peaks and a broader range of values. This may indicate a more complex or less predictable environment where the agent's actions yield more fluctuating results. This could signify a more sophisticated or less stable set of dynamics in the advertising system, where the agent must adapt to a broader array of factors to maximize the reward.

In both cases, the variability in the reward trajectories underscores the stochastic nature of reinforcement learning tasks and the importance of exploration. The agent's performance does not improve linearly but evolves through trial and error. The graphs suggest that the agent is successfully refining its policy over time, which is essential for achieving long-term benefits in dynamic systems such as multi-channel advertising budget allocation.

Figure 6 presents a detailed performance comparison of RL, DE, and their combined approach to minimizing error values over time. The graphs illustrate how these algorithms reduce the distance from an optimal value across generations or learning sessions within the context of the CICAD and CRITEO-UPLIFT v2 datasets.

The first graph demonstrates a scenario where individual algorithms and their integration fluctuate as they progress. The RL algorithm shows a gradual improvement trend, reflecting its learning and adjustment capabilities as it receives environmental feedback. However, its path is not strictly linear, suggesting some variability in learning efficiency or the challenges presented by the dataset at different stages. The DE algorithm, known for its robust global search capabilities, exhibits a more volatile reduction in error. This indicates its exploratory nature, where it might escape local minima but encounter periods of less efficient search, leading to the observed fluctuations. Despite these fluctuations, DE decreases the error significantly,

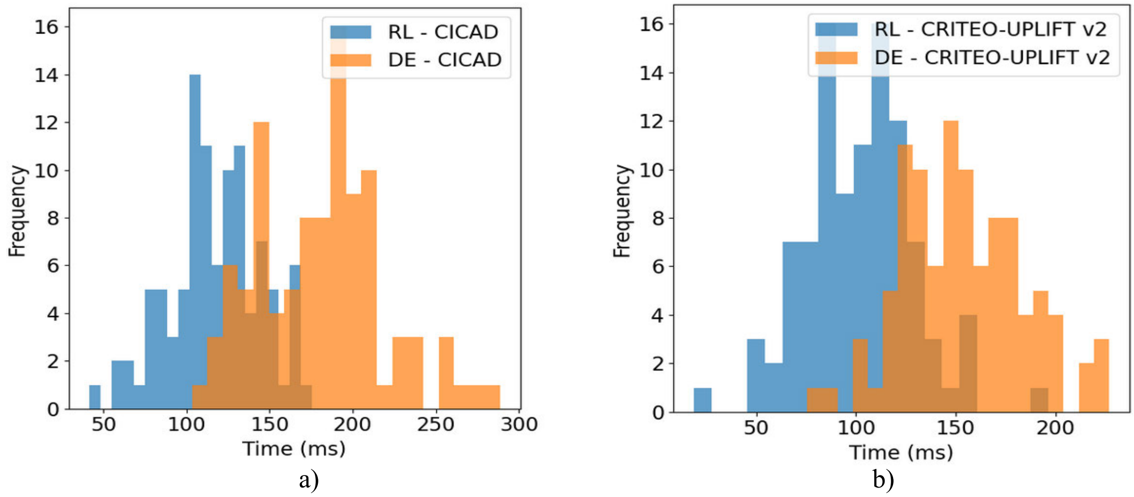


FIGURE 7. Comparative analysis of decision-making times for RL and DE algorithms across a) CICAD, b) CRITEO-UPLIFT v2 datasets.

highlighting its effectiveness in navigating complex problem spaces. The combined approach generally shows a smoother and more consistent decline in error compared to the individual performances. This smoother curve implies that integrating RL and DE harnesses the strengths of both the strategic exploration of DE and the adaptive learning of RL. By mitigating the weaknesses of each algorithm through integration, the combined approach achieves a more stable and effective reduction in error, reaching lower error levels more consistently.

In the second graph, the three approaches' performance is more aligned, suggesting a scenario where the distinct advantages of each method are less pronounced. However, the integrated approach maintains a slight edge, first reaching the lowest error values. This sustained performance superiority in different datasets and under varying conditions underscores the robustness of the combined approach, reinforcing the value of integrating RL and DE to handle diverse and challenging optimization tasks in dynamic environments.

Figure 7 shows the distribution of decision-making times for the RL and DE algorithms applied to CICAD and CRITEO-UPLIFT v2 datasets. Each histogram provides insights into how quickly each algorithm processes data and makes decisions, which is crucial for performance in real-time bidding (RTB) environments.

For the CICAD dataset, the RL algorithm typically completes decision-making faster than the DE algorithm. Most RL decision times are concentrated around 100 ms, with a notable peak slightly below this value. This suggests that RL is generally more efficient due to its ability to incrementally learn and adapt based on past experiences without requiring extensive new computations for each decision. DE shows a broader spread of decision times, with a significant amount of decisions taking longer, up to about 250 ms. This distribution indicates that DE's approach, which involves evaluating multiple solution variations to evolve the optimal decision, inherently requires more computation time, leading to slower decision-making processes.

In the CRITEO-UPLIFT v2 dataset, the decision-making times for both RL and DE are generally faster than the CICAD dataset, which might indicate that the characteristics or complexity of this dataset allow quicker processing. RL again shows faster processing times, with most decisions around 50-100 ms. This faster performance underscores RL's suitability for environments where decisions must be made quickly, such as in RTB. DE's times are again slower on average than RL's, with a broader spread and a peak around 150 ms. While quicker than in the CICAD dataset, the times still reflect DE's more computationally intensive nature.

Across both datasets, RL consistently demonstrates quicker decision-making capabilities than DE. This characteristic is advantageous in RTB environments, where faster decision times can lead to better bidding opportunities and outcomes. DE, while slower, may still be valuable in scenarios where the decision quality from exploring diverse solutions outweighs the need for speed. However, its slower decision-making process might only limit its effectiveness in high-speed environments if optimizations are made to reduce computation times.

1) ANALYSIS OF THE DE ALGORITHM

In a subsequent experiment, the advanced DE algorithm was compared with a range of renowned metaheuristic optimization algorithms, and the efficacy of eight distinct algorithms was rigorously evaluated: Salp Swarm Algorithm (SSA) [52], Bat Algorithm (BA) [53], Firefly Algorithm (FA) [54], Artificial Bee Colony (ABC) [55], and the standard DE. The detailed results of this stringent comparison are presented in Table 7, highlighting the algorithmic effectiveness on both the CICAD and CRITEO-UPLIFT v2 datasets. On the CICAD dataset, the enhanced DE algorithm achieved a commendable

TABLE 7. Performance comparison of the enhanced DE algorithm against prominent metaheuristic optimization algorithms on the CICAD and CRITEO-UPLIFT v2 datasets.

Method	CICAD		CRITEO-UPLIFT v2	
	AUCC	EOM	AUCC	EOM
SSA	0.630±0.015	0.710±0.014	0.701±0.016	0.650±0.023
BA	0.662±0.013	0.680±0.010	0.711±0.008	0.678±0.011
FA	0.684±0.006	0.704±0.013	0.737±0.001	0.691±0.003
ABC	0.692±0.011	0.720±0.017	0.721±0.019	0.707±0.016
DE	0.713±0.008	0.751±0.012	0.764±0.009	0.734±0.013

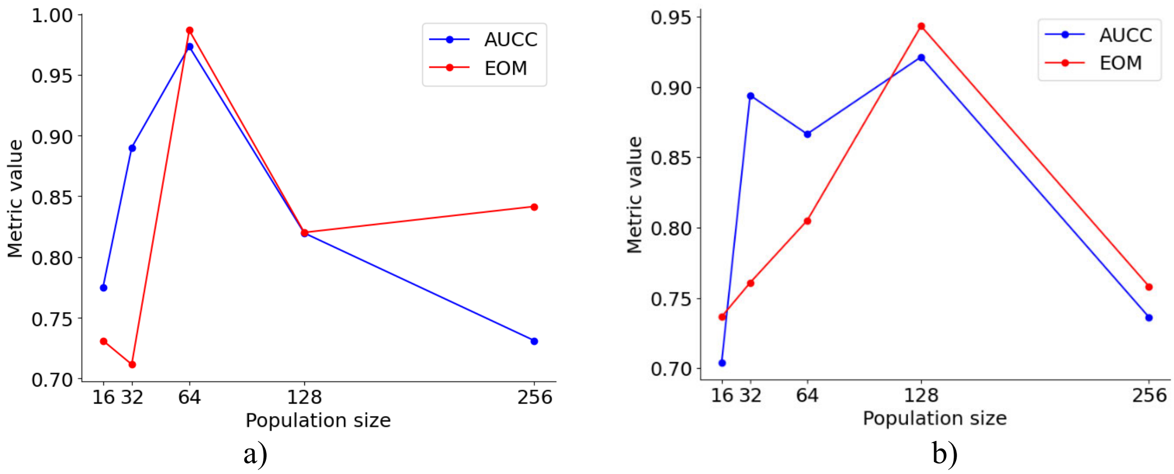


FIGURE 8. Performance variation of the DE algorithm with different population sizes. The plot showcases the efficacy of the model for the datasets CICAD and CRITEO-UPLIFT v2, highlighting optimal population size values of 64 and 128, respectively. a) CICAD, b) CRITEO-UPLIFT v2.

21% error reduction compared to the conventional DE. This marked improvement emphasizes the superior efficiency of the method relative to the traditional DE approach.

Notably, the DE algorithm surpassed other metaheuristic algorithms like ABC, FA, and BA, consolidating its stance as a robust optimization tool in this domain. Similarly, for the CRITEO-UPLIFT v2 dataset, the DE algorithm realized an outstanding 19.27% reduction in error rates. This pronounced performance boost reiterates the versatility and resilience of the DE algorithm across varied datasets, positioning it as an optimal solution for diverse optimization tasks.

A series of experiments were embarked upon to delve deeper into the nuances of the DE algorithm, focusing on exploring the influence of population size on the algorithm's effectiveness. The objective was to identify the optimal population size to enhance the algorithm's performance, particularly in metrics such as the AUCC and the EOM. To achieve this, the population size was varied across a range of values—16, 32, 64, 128, and 256—and the resulting impact on performance was monitored.

The findings, illustrated in Figure 8, shed light on some fascinating patterns. For the CICAD dataset, the peak performance, as measured by AUCC and EOM, was attained with a population size of 64. Similarly, for the CRITEO-UPLIFT v2 dataset, the algorithm reached its optimal performance with a population size of 128. Interestingly, increasing the population size beyond these optimal points did not consistently lead to better results. A notable decrease in performance metrics was observed for both datasets as the population size was further expanded from optimal to 256. This pattern implies that while a larger population size can initially contribute to the algorithm's improved performance, a limit exists beyond which further increases may lead to diminishing returns.

In the context of the DE algorithm, the scaling factor is a crucial parameter governing the amplification of the differential vector between two candidate solutions. This amplification occurs when producing a trial solution during the mutation operation. Typically, the scaling factor in DE lies between 0 and 2. A scaling factor near 0 results in minimal perturbation of the differential vector, leading to finer solution space exploration. This fine-tuned approach can be beneficial for ensuring stability and convergence, especially during the initial optimization phases. Conversely, a scaling factor approaching 2 causes a more pronounced perturbation, facilitating broader solution space exploration. While this can aid in promoting diversity and escaping local optima, it may also render the algorithm less stable, potentially leading to convergence challenges. To assess the influence of the scaling factor, experiments were executed with varying initial values from the set $\{0, 0.1, 0.2, 0.8, 1, 1.2, 1.5, 1.8, 2\}$, as depicted in Figure 9. The model's performance, evaluated using diverse metrics, peaks at scaling factor values of 0.8 for CICAD and 1.2 for CRITEO-UPLIFT v2.

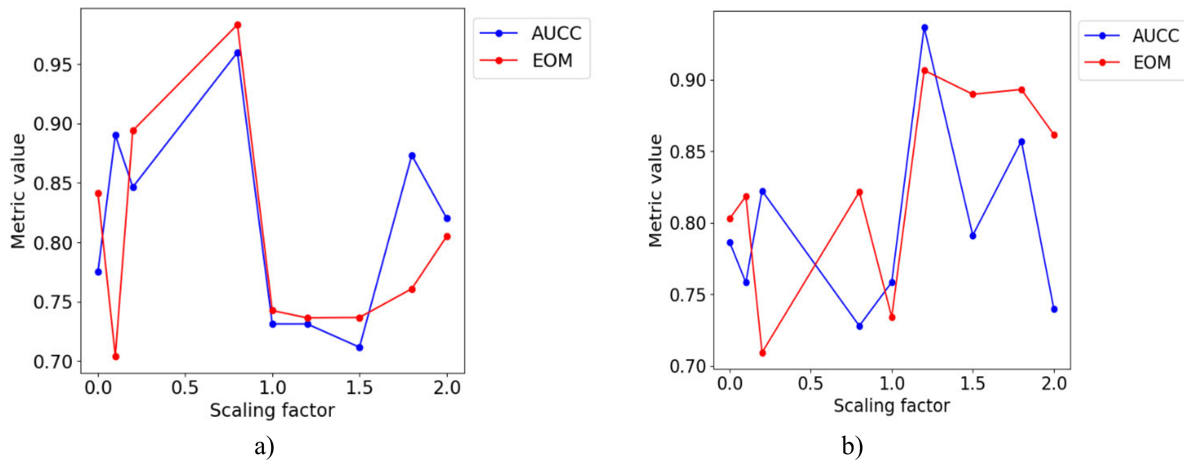


FIGURE 9. Performance variation of the DE algorithm with different scaling factors. The plot showcases the efficacy of the model for the datasets CICAD and CRITEO-UPLIFT v2, highlighting optimal scaling factor values of 0.8 and 1.2, respectively. a) CICAD, b) CRITEO-UPLIFT v2.

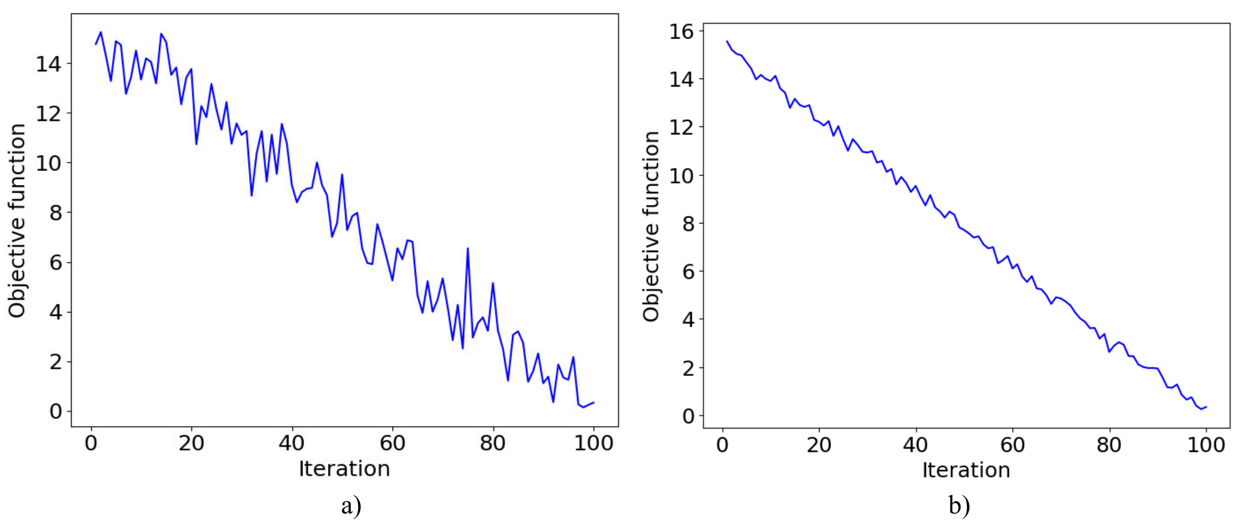


FIGURE 10. Convergence patterns of the objective function values across iterations using the proposed Differential Evolution strategy on two datasets: a) CICAD and b) CRITEO-UPLIFT v2.

Figure 10 presents the convergence trends of the objective function's values across numerous iterations for two distinct datasets within the framework of the proposed DE strategy. Figure 10. a displays the results for the CICAD dataset, showing a marked, albeit irregular, downward trend, indicating an overall successful optimization process with some fluctuations that might represent the exploration phase inherent in the DE approach. Figure 10. b, detailing the CRITEO-UPLIFT v2 dataset, depicts a more consistent and steady decrease in the objective function values, suggesting a smoother optimization pathway and possibly a less complex problem space or a better initial parameter setting for this particular dataset. The variance in convergence patterns between the two datasets underscores the adaptability of the DE strategy to different data characteristics and optimization landscapes. It also reflects the inherent stochastic nature of evolutionary algorithms, where exploration can occasionally lead to temporary increases in the objective function before converging to a more optimal solution.

E. DISCUSSION

This article unveils an innovative strategy for allocating budgets across multiple advertising channels, combining RL Q-learning with an advanced DE algorithm. This integrated model excels in making strategic, sequential decisions and effectively tackles the intricate challenges of distributing budgets across various platforms. Extensive testing on the CICAD and CRITEO-UPLIFT v2 datasets confirms the effectiveness of the proposed approach, which outperforms existing advanced models.

The experiments, including detailed ablation studies, underscore the unique value each component—RL, DA, and DE—brings to the overall efficacy of the model. The remarkable performance against other top-tier models can be attributed to the seamless fusion of RL and DE. RL's ability to adapt and learn from continuous feedback enhances the model's strategic decision-making. Concurrently, the DE algorithm, with its innovative clustering-based mutation, explores and refines potential strategies, leading to the discovery of effective and pragmatic budget allocation solutions.

Every element of the proposed model has been carefully chosen to uniquely tackle the multifaceted challenges inherent in allocating budgets across multiple advertising channels. The RL framework stands out for its exceptional capability in making sequential decisions and its adeptness at learning from varying outcomes. This makes it exceptionally suited to the ever-changing landscapes of advertising, where adaptability and foresight are crucial.

The DE algorithm, enriched with a novel clustering-based mutation approach, plays a pivotal role in efficiently traversing the complex landscape of potential strategies. It promotes exploration and exploitation, uncovering innovative and practical budget allocation solutions. Furthermore, introducing a discretization method streamlines the model's complexity, enhancing its manageability and real-world applicability. This tactic breaks down the vast and intricate budget space into more approachable segments, facilitating easier decision-making and implementation. This strategic simplification improves the model's usability and scalability, enabling it to accommodate a wide range of advertising scenarios and budget sizes without compromising performance. Collectively, these components synergize to create a robust model that meets the demands of multi-channel budget allocation and sets a new standard in the field, offering a blend of precision, adaptability, and practicality.

The theoretical underpinnings of the study hold substantial importance, providing a sophisticated framework that enriches the grasp of multi-channel advertising strategies and extends its benefits to the broader domains of reinforcement learning and evolutionary computation. The findings showcase the potent synergy achievable through the amalgamation of these advanced methodologies, illuminating new pathways for investigating adaptive strategies within complex and ever-evolving systems. This research not only elucidates the intricacies of budget allocation across diverse advertising platforms but also lays down foundational insights that could catalyze innovations in algorithmic approaches to decision-making under uncertainty. The integration of reinforcement learning with evolutionary algorithms exemplifies a novel approach to solving real-world problems characterized by their dynamic and unpredictable nature. Consequently, this study contributes a significant leap forward in the theoretical understanding, promising to inspire future research on harnessing these methodologies to tackle various challenges in various disciplines.

However, the proposed model has limitations:

- Scalability: The scalability challenge is heightened as the diversity and number of advertising channels expand, potentially leading to an exponential increase in computational demands. This is particularly pertinent in today's digital marketing landscape, where new platforms emerge rapidly. To address this, advanced dimensionality reduction techniques could be employed to distill the essential information from vast datasets, thereby reducing computational load. Additionally, leveraging parallel computing frameworks could distribute the workload across multiple processors, significantly enhancing the model's ability to scale up efficiently and handle a broader array of channels without a corresponding spike in computational time or resources.
- Data dependency: The reliance of the model on extensive and high-quality historical data poses a significant limitation, especially in scenarios where data is scarce, outdated, or biased. Such data issues can skew model predictions and diminish performance. To combat this, robust data augmentation techniques, which generate synthetic data points from existing ones, could be implemented to enrich the dataset and improve model robustness. Moreover, exploring unsupervised learning components that do not rely on labeled data could offer valuable insights from unlabeled data, thereby reducing the model's dependency on extensive historical datasets.
- Adaptability to rapid changes: The digital advertising landscape is characterized by its volatility and rapid evolution, with frequent shifts in user preferences, platform algorithms, and market dynamics. The current model might need help promptly adapting to these swift changes, potentially compromising effectiveness. Incorporating faster adaptation mechanisms, such as meta-learning, within the RL component could provide a solution. Meta-learning enables the model to rapidly adjust to new conditions based on previous learning experiences, enhancing its adaptability to sudden market shifts and maintaining its performance in a dynamic environment.
- Technical complexity in integration: The integration of RL with an improved DE algorithm presents a significant technical challenge. Both algorithms have complex parameters and processes that must be meticulously aligned to work in tandem. This complexity makes the model difficult to implement and maintain and requires specialized knowledge and skills, potentially limiting its accessibility and scalability to broader applications where such expertise may not be available. Adopting a modular approach in the model's design can help isolate the functionalities of RL and DE, making each component more manageable and easier to maintain. Encapsulating each algorithm in its module allows their interaction to be controlled more finely, reducing integration complexity. Moreover, implementing knowledge transfer techniques, where insights from domain experts are encoded into the system, can mitigate the need for deep technical expertise in every application scenario.

Developing comprehensive documentation and training programs can also facilitate broader accessibility and usability, helping overcome the barriers posed by the model's complexity.

- Real-time decision-making challenges: Implementing the solution in an RTB environment requires the model to make quick decisions, which might be challenging given the algorithms' complexity. The necessity for rapid response times in RTB can be at odds with the computational demands of both RL and DE, potentially leading to delays that could impact the efficacy of bidding strategies and overall campaign performance. To solve this problem, implementing predictive and preemptive computing strategies will be critical. Leveraging machine learning techniques to predict bidding scenarios and prepare responses in advance can decrease decision-making time during live bidding events. This preemptive setup can be supported by a robust data caching system that stores previously computed results and rapidly retrieves them when similar scenarios arise, further reducing the latency involved in each transaction. To further enhance system responsiveness, a microservices architecture could be adopted. This allows individual components of the RL and DE algorithms to operate independently and scale dynamically according to demand. Each microservice would handle a specific task or process within the larger bidding strategy, operating concurrently across different servers or cloud environments. This improves fault tolerance and system reliability and increases data processing and decision-making efficiency.

VI. CONCLUSION

This study introduced an innovative approach to allocating budgets across multiple advertising channels, which harnessed the capabilities of an RL Q-learning framework combined with an advanced DE algorithm to enhance the Q-learning process. The RL component was adept at making informed sequential decisions, skillfully adapting strategies to prioritize long-term advantages by integrating environmental feedback. The DE algorithm was also augmented with a novel clustering-based mutation technique, leveraging significant clusters within the DE population to create unique and practical solutions. The strategy was refined by implementing a discretization method that simplified the model and made cost management more efficient. The proposed method underwent rigorous validation using two comprehensive datasets: the CICAD and CRITEO-UPLIFT v2. The empirical evidence confirmed the model's dominance, demonstrating outstanding performance with substantial scores (AUCC $= 0.750$ and EOM $= 0.736$ for CICAD; AUCC $= 0.813$ and EOM $= 0.829$ for CRITEO-UPLIFT v2), thus highlighting the model's capability in addressing the complex challenges of multi-channel budget allocation and setting a new standard in the domain.

In future work, the capabilities of the multi-channel advertising budget allocation model are planned to be expanded by incorporating real-time bidding strategies, allowing the model to not only allocate budgets but also dynamically adjust bids in response to changing market conditions. This could significantly enhance the model's reactivity and profitability in the ever-evolving online advertising landscape. Another avenue for future research is integrating a broader range of environmental feedback signals, including consumer behavior metrics and competitor actions, to refine the decision-making process further. By understanding and responding to a broader context, the model could deliver even more sophisticated and nuanced budget allocation strategies, driving higher returns on investment. Finally, future work will include advanced optimization techniques to enhance the efficacy and efficiency of our model through hyperparameter optimization of RL and DE algorithms. Specifically, we plan to explore meta-heuristic algorithms, which are well-regarded for their ability to identify optimal or near-optimal solutions in complex, multi-dimensional spaces. These techniques are particularly effective in navigating environments with multiple local optima, making them ideal for refining our approach to hyperparameter tuning.

REFERENCES

[1] A. Rangaswamy and G. H. Van Bruggen, "Opportunities and challenges in multichannel marketing: An introduction to the special issue," *J. Interact. Marketing*, vol. 19, no. 2, pp. 5–11, May 2005, doi: 10.1002/dir.20037.
[2] Z. Yan, S. Wang, G. Zhou, J. Lin, and P. Jiang, "An end-to-end framework for marketing effectiveness optimization under budget constraint," 2023, *arXiv:2302.04477*.
[3] Q. Deng, K. Hu, and Y. F. Lim, "Cross-channel marketing on e-commerce marketplaces: Impact and strategic budget allocation," *SSRN Electron. J.*, vol. 12, pp. 1–49, Jan. 2023, doi: 10.2139/ssrn.4332631.
[4] M. G. Seok and S. H. Choi, "Practical simulation budget allocation for ranked subset partitioning," *IEEE Access*, vol. 11, pp. 104347–104358, 2023, doi: 10.1109/access.2023.3317283.
[5] X. wang, P. Li, and A. Hawbani, "An efficient budget allocation algorithm for multi-channel advertising," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 886–891, doi: 10.1109/ICPR.2018.8545777.
[6] H. Zhang and Y. Vorobeychik, "Multi-channel marketing with budget complementarities," in *Proc. Int. Conf. Auton. Agents Multiagent Syst.*, 2017, pp. 1232–1240, doi: 10.5555/3091125.3091296.
[7] Y. Deng, N. Golrezaei, P. Jaillet, J. C. N. Liang, and V. Mirrokni, "Multi-channel autobidding with budget and ROI constraints," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 7617–7644. [Online]. Available: https://proceedings.mlr.press/v202/deng23c.html
[8] S. Peterson and O. Yogev, "Method and system for lead budget allocation and optimization on a multi-channel multi-media campaign management and payment platform," U.S. Patent 11 386 449 B2, Jul. 22, 2022.
[9] M. H. Modirrousta, M. A. Shoorehdeli, M. Yari, and A. Ghahremani, "Deep Q-learning recommender algorithm with update policy for a real steam turbine system," *IET Collaborative Intell. Manuf.*, vol. 5, no. 3, pp. 1–11, Sep. 2023, doi: 10.1049/cim2.12081.
[10] A. Kokhahi and D. Li, "GLHAD: A group lasso-based hybrid attack detection and localization framework for multistage manufacturing systems," *J. Comput. Inf. Sci. Eng.*, vol. 24, no. 7, Jul. 2024, Art. no. 071006, doi: 10.1115/1.4063862.
[11] R. Khodadadi, G. Ardeshir, and H. Grailu, "A new method of facial image compression based on meta-heuristic algorithms with variable bit budget allocation," *Signal, Image Video Process.*, vol. 17, no. 8, pp. 3923–3931, Nov. 2023, doi: 10.1007/s11760-023-02622-y.
[12] G. Molnar, "Metaheuristics for problems with limited budget of evaluations," Ph.D. dissertation, Dept. Faculty Elect. Eng. Comput., Univ. Zagreb, Zagreb, Croatia, 2020. [Online]. Available: https://urn.nsk.hr/urn:nbn:hr:168:219792

[13] H. Gharagozlou, J. Mohammadzadeh, A. Bastanfard, and S. S. Ghidary, "RLAS-BIABC: A reinforcement learning-based answer selection using the BERT model boosted by an improved ABC algorithm," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–21, May 2022, doi: 10.1155/2022/7839840.

[14] P. Saeid, M. Pazoki, and M. Zeinolabedini, "Optimization of biomass production from sugar bagasse in anaerobic digestion using genetic algorithm," *Model. Earth Syst. Environ.*, vol. 9, no. 2, pp. 2183–2198, Jun. 2023, doi: 10.1007/s40808-022-01598-x.

[15] J.-J. Ji, Y.-N. Guo, X.-Z. Gao, D.-W. Gong, and Y.-P. Wang, "Q-learning-based hyperheuristic evolutionary algorithm for dynamic task allocation of crowdsensing," *IEEE Trans. Cybern.*, vol. 53, no. 4, pp. 2211–2224, Apr. 2023, doi: 10.1109/TCYB.2021.3112675.

[16] N. Elshaboury, E. M. Abdelkader, and M. Marzouk, "Application of modified invasive weed algorithm for condition-based budget allocation of water distribution networks," in *Design and Construction of Smart Cities*. Cham, Switzerland: Springer, 2021, pp. 121–131, doi: 10.1007/978-3-030-64217-4_15.

[17] S. Vakilian, S. V. Moravvej, and A. Fanian, "Using the artificial bee colony (ABC) algorithm in collaboration with the fog nodes in the Internet of Things three-layer architecture," in *Proc. 29th Iranian Conf. Electr. Eng. (ICEE)*, May 2021, pp. 509–513, doi: 10.1109/ICEE52715.2021.9544399.

[18] M. Roshanzamir, R. Alizadehsani, S. V. Moravvej, J. H. Joloudari, H. Alinejad-Rokny, and J. M. Gorriz, "Enhancing interpretability in machine learning: A focus on genetic network programming, its variants, and applications," in *Proc. Int. Work-Conf. Interplay Between Natural Artif. Comput.* Cham, Switzerland: Springer, 2024, pp. 98–107, doi: 10.1007/978-3-031-61140-7_10.

[19] H. Gharagozlou, J. Mohammadzadeh, A. Bastanfard, and S. S. Ghidary, "Semantic relation extraction: A review of approaches, datasets, and evaluation methods with looking at the methods and datasets in the Persian language," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 22, no. 7, pp. 1–29, Jul. 2023, doi: 10.1145/3592601.

[20] J.-Y. Li, K.-J. Du, Z.-H. Zhan, H. Wang, and J. Zhang, "Distributed differential evolution with adaptive resource allocation," *IEEE Trans. Cybern.*, vol. 53, no. 5, pp. 2791–2804, May 2023, doi: 10.1109/TCYB.2022.3153964.

[21] M. F. Ahmad, N. A. M. Isa, W. H. Lim, and K. M. Ang, "Differential evolution with modified initialization scheme using chaotic oppositional based learning strategy," *Alexandria Eng. J.*, vol. 61, no. 12, pp. 11835–11858, Dec. 2022, doi: 10.1016/j.aej.2022.05.028.

[22] H. Zhou, S. Li, G. Jiang, J. Zheng, and D. Wang, "Direct heterogeneous causal learning for resource allocation problems in marketing," in *Proc. AAAI Conf. Artif. Intell.*, 2023, no. 4, pp. 5446–5454, doi: 10.1609/aaai.v37i4.25677.

[23] J. Jannink, R. Astudillo, and P. Frazier, "Insight into a two-part plant breeding scheme through Bayesian optimization of budget allocations," *Crop Sci.*, vol. 10, Feb. 2024, Art. no. 20230277076, doi: 10.1002/csc2.21124.

[24] Q. Zhang, M. Chen, A. Bukharin, N. Karampatziakis, P. He, Y. Cheng, W. Chen, and T. Zhao, "AdaLoRA: Adaptive budget allocation for parameter-efficient fine-tuning," 2023, *arXiv:2303.10512*.

[25] M. Ai, B. Li, H. Gong, Q. Yu, S. Xue, Y. Zhang, Y. Zhang, and P. Jiang, "LBCF: A large-scale budget-constrained causal forest algorithm," in *Proc. ACM Web Conf.*, Apr. 2022, pp. 2310–2319, doi: 10.1145/3485447.3512103.

[26] X. Hao, Z. Peng, Y. Ma, G. Wang, J. Jin, J. Hao, S. Chen, R. Bai, M. Xie, M. Xu, Z. Zheng, C. Yu, H. Li, J. Xu, and K. Gai, "Dynamic knapsack optimization towards efficient multichannel sequential advertising," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 4060–4070. [Online]. Available: https://proceedings.mlr.press/v119/hao20b.html

[27] A. Betlei, E. Diemert, and M.-R. Amini, "Uplift modeling with generalization guarantees," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2021, pp. 55–65, doi: 10.1145/3447548.3467395.

[28] S. Du, J. Lee, and F. Ghaffarizadeh, "Improve user retention with causal learning," in *Proc. ACM SIGKDD Workshop Causal Discovery*, 2019, pp. 34–49. [Online]. Available: https://proceedings.mlr.press/v104/du19a

[29] S. R. Künzel, J. S. Sekhon, P. J. Bickel, and B. Yu, "Metalearners for estimating heterogeneous treatment effects using machine learning," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 10, pp. 4156–4165, Mar. 2019, doi: 10.1073/pnas.1804597116.

[30] S. Athey and S. Wager, "Policy learning with observational data," *Econometrica*, vol. 89, no. 1, pp. 133–161, 2021, doi: 10.3982/ecta15732.

[31] S. Xiao, L. Guo, Z. Jiang, L. Lv, Y. Chen, J. Zhu, and S. Yang, "Model-based constrained MDP for budget allocation in sequential incentive marketing," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 971–980, doi: 10.1145/3357384.3358031.

[32] G. Shen, S. Sun, D. Gao, S. Li, L. Yang, Y. Shi, and W. Ning, "Cross-channel budget coordination for online advertising system," 2023, *arXiv:2305.06883*.

[33] G. Kou, H. Xiao, M. Cao, and L. H. Lee, "Optimal computing budget allocation for the vector evaluated genetic algorithm in multi-objective simulation optimization," *Automatica*, vol. 129, Jul. 2021, Art. no. 109599, doi: 10.1016/j.automatica.2021.109599.

[34] Y. Luzon, R. Pinchover, and E. Khmelnitsky, "Dynamic budget allocation for social media advertising campaigns: Optimization and learning," *Eur. J. Oper. Res.*, vol. 299, no. 1, pp. 223–234, May 2022, doi: 10.1016/j.ejor.2021.08.019.

[35] Y. Li, X. Song, Y. Tu, and M. Liu, "GAPBAS: Genetic algorithm-based privacy budget allocation strategy in differential privacy K-means clustering algorithm," *Comput. Secur.*, vol. 139, Apr. 2024, Art. no. 103697, doi: 10.1016/j.cose.2023.103697.

[36] A. Nuara. (2020). *Machine Learning Algorithms for the Optimization of Internet Advertising Campaigns*. [Online]. Available: https://hdl.handle.net/10589/171153

[37] A. Mirzaee M. Kasmaee, A. Ataei, S. V. Moravvej, R. Alizadehsani, J. M. Gorriz, Y.-D. Zhang, R.-S. Tan, and U. R. Acharya, "ELRL-MD: A deep learning approach for myocarditis diagnosis using cardiac magnetic resonance images with ensemble and reinforcement learning integration," *Physiological Meas.*, vol. 45, no. 5, May 2024, Art. no. 055011, doi: 10.1088/1361-6579/ad46e2.

[38] R. Amiri, H. Mehrpouyan, L. Fridman, R. K. Mallik, A. Nallanathan, and D. Matolak, "A machine learning approach for power allocation in HetNets considering QoS," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–7, doi: 10.1109/ICC.2018.8422864.

[39] S. Danaei, A. Bostani, S. V. Moravvej, F. Mohammadi, R. Alizadehsani, A. Shoeibi, H. Alinejad-Rokny, and S. Nahavandi, "Myocarditis diagnosis: A method using mutual learning-based ABC and reinforcement learning," in *Proc. IEEE 22nd Int. Symp. Comput. Intell. Informat. 8th IEEE Int. Conf. Recent Achievements Mechatronics, Autom., Comput. Sci. Robot. (CINTI-MACRo)*, Nov. 2022, pp. 000265–000270, doi: 10.1109/CINTI-MACRo57952.2022.10029403.

[40] S. Taherinavid, S. V. Moravvej, Y.-L. Chen, J. Yang, C. S. Ku, and P. L. Yee, "Automatic transportation mode classification using a deep reinforcement learning approach with smartphone sensors," *IEEE Access*, vol. 12, pp. 514–533, 2024, doi: 10.1109/access.2023.3346875.

[41] K. V. Price, "Differential evolution," in *Handbook of Optimization: From Classical to Modern Approach*. Cham, Switzerland: Springer, 2013, pp. 187–214, doi: 10.1007/978-3-642-30504-7_8.

[42] S. V. Moravvej, S. J. Mousavirad, D. Oliva, and F. Mohammadi, "A novel plagiarism detection approach combining BERT-based word embedding, attention-based LSTMs and an improved differential evolution algorithm," 2023, *arXiv:2305.02374*.

[43] I. A. Svendsen and U. Utrevu, "Nearshore mixing and dispersion," *Proc. Roy. Soc. London, Ser. A, Math. Phys. Sci.*, vol. 445, no. 1925, pp. 561–576, 1994, doi: 10.1098/rspa.1994.0078.

[44] S. J. Mousavirad, G. Schaefer, I. Korovin, M. H. Moghadam, M. Saadatmand, and M. Pedram, "An enhanced differential evolution algorithm using a novel clustering-based mutation operator," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2021, pp. 176–181, doi: 10.1109/SMC52423.2021.9658743.

[45] K. Deb, "A population-based algorithm-generator for real-parameter optimization," *Soft Comput.*, vol. 9, no. 4, pp. 236–253, Apr. 2005, doi: 10.1007/s00500-004-0377-4.

[46] E. Diemert, A. Betlei, C. Renaudin, and M. R. Amini, "A large scale benchmark for uplift modeling," in *Proc. KDD*, 2018, pp. 1–7. [Online]. Available: https://hal.science/hal-02515860.

[47] A. Budvytis. (2024). *Modelling Return on Advertising Spend. Subscription-Based B2C Bus. Case Study*. [Online]. Available: https://epublications.vu.lt/object/elaba:192057623

[48] K. V. Sriram, A. A. Poojary, V. Jawa, and G. B. Kamath, "Return on investment and return on ad spend at the action level of AIDA using last touch attribution method on digital advertising platforms," *Int. J. Internet Marketing Advertising*, vol. 17, nos. 1–2, pp. 111–132, 2022, doi: 10.1504/ijima.2022.125145.

[49] R. Almestarihi, A. Y. A. B. Ahmad, R. H. Frangieh, I. A. Abu-AlSondos, K. K. Nser, and A. Ziani, "Measuring the ROI of paid advertising campaigns in digital marketing and its effect on business profitability," *Uncertain Supply Chain Manage.*, vol. 12, no. 2, pp. 1275–1284, 2024, doi: 10.5267/j.uscm.2023.11.009.

[50] K. Gadgil, S. S. Gill, and A. M. Abdelmoniem, "A meta-learning based stacked regression approach for customer lifetime value prediction," *J. Economy Technol.*, vol. 1, pp. 197–207, Nov. 2023, doi: 10.1016/j.ject.2023.09.001.

[51] Y. Yan and N. Resnick, "A high-performance turnkey system for customer lifetime value prediction in retail brands: Forthcoming in quantitative marketing and economics," *Quant. Marketing Econ.*, vol. 22, no. 2, pp. 169–192, Jun. 2024, doi: 10.1007/s11129-023-09272-x.

[52] D. Bairathi and D. Gopalani, "Salp swarm algorithm (SSA) for training feed-forward neural networks," in *Soft Computing for Problem Solving: SocProS*. Cham, Switzerland: Springer, 2019, pp. 521–534, doi: 10.1007/978-981-13-1592-3_41.

[53] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," in *Nature Inspired Cooperative Strategies for Optimization (NICSO)*. Cham, Switzerland: Springer, 2010, pp. 65–74, doi: 10.1007/978-3-642-12538-6_6.

[54] N. F. Johari, A. M. Zain, M. H. Noorfa, and A. Udin, "Firefly algorithm for optimization problem," *Appl. Mech. Mater.*, vol. 421, pp. 512–517, Sep. 2013, doi: 10.4028/www.scientific.net/amm.421.512.

[55] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm," *J. Global Optim.*, vol. 39, no. 3, pp. 459–471, Oct. 2007, doi: 10.1007/s10898-007-9149-x.

MENGFAN LI stands out as a Committed and Innovative Student with the University of Melbourne, where she is making significant strides in machine learning. Her research, fueled by a profound passion for the subject, delves into developing cutting-edge applications with the potential to revolutionize stock market forecasting and social media analytics. Her work is characterized by a meticulous approach to data analysis and algorithm development, aiming to enhance predictive accuracy and provide deeper insights into complex digital landscapes.

JIAN ZHANG an accomplished student with the University of Malaya, is at the forefront of machine learning, driving forward innovations in predictive analytics. His profound expertise has significantly impacted stock market forecasting and social media analysis, improving the approach to data-informed decision-making. Through his groundbreaking work, he is redefining how data is leveraged, offering new perspectives and methodologies that set new standards in accuracy and insight within these dynamic sectors.

ROOHALLAH ALIZADEHSANI (Member, IEEE) received the Bachelor of Science degree in computer engineering-software and the Master of Science degree in computer engineering-software from Sharif University of Technology. His research interests include data mining, machine learning, bioinformatics, heart disease, skin disease, diabetes disease, hepatitis disease, and cancer disease. He is currently a Research Fellow with Deakin University, Australia.

PAWEŁ PŁAWIAK was born in Ostrowiec, Poland, in 1984. He received the B.Eng. and M.Sc. degrees in electronics and telecommunications and the Ph.D. degree (Hons.) in biocybernetics and biomedical engineering from the AGH University of Science and Technology, Kraków, Poland, in 2012 and 2016, respectively, and the D.Sc. degree in technical computer science and telecommunications from the Silesian University of Technology, Gliwice, Poland, in 2020. He is currently the Dean of the Faculty of Computer Science and Telecommunications and an Associate Professor with Cracow University of Technology, Kraków, the Deputy Director of Research with the National Institute of Telecommunications, Warsaw, and an Associate Professor with the Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Gliwice. He has published more than 90 articles in refereed international SCI-IF journals. His research interests include machine learning and computational intelligence (e.g., artificial neural networks, genetic algorithms, fuzzy systems, support vector machines, k-nearest neighbors, and hybrid systems), ensemble learning, deep learning, evolutionary computation, classification, pattern recognition, signal processing and analysis, data analysis and data mining, sensor techniques, medicine, biocybernetics, biomedical engineering, and telecommunications. He is an academic editor and a reviewer of many prestigious and reputed journals.