Foundation University Journal of Engineering and Applied Sciences

FUJEAS
Vol. 6, Issue 1, 2025
DOI:10.33897/fujeas.v6i1.964

Research Article

Article Citation:
Unegbu et al. (2025). "Reinforcement Learning for Optimizing Bio-Composite Processing Conditions under Local Constraints". *Foundation University Journal of Engineering and Applied Sciences*
DOI:10.33897/fujeas.v6i1.964

Published by
Foundation University
Islamabad.
Web: https://fui.edu.pk/

Reinforcement Learning for Optimizing Bio-Composite Processing Conditions under Local Constraints

Hyginus Chidiebere Onyekachi Unegbu *, Danjuma Saleh Yawas, Titus Adeniyi Ilupeju

Department of Mechanical Engineering, Ahmadu Bello University, Zaria, Nigeria.

* **Corresponding Author:** chidieberehyg@gmail.com

Abstract:

This study proposes a high-performance, reinforcement learning-based optimization framework for bio-composite processing, leveraging the Soft Actor-Critic (SAC) algorithm within a constrained decision-making context. Conventional optimization methods in composite manufacturing often face limitations in balancing multiple interdependent objectives such as mechanical performance, energy efficiency, constraint adherence, and production throughput. To overcome these limitations, this work integrates a high-fidelity digital twin environment with a constrained Markov Decision Process (CMDP) formulation, enabling the SAC agent to learn optimal control strategies in real time while respecting operational boundaries. The proposed model was benchmarked against Proximal Policy Optimization (PPO) and Genetic Algorithm (GA) across four key metrics: tensile strength, energy consumption, constraint violation rate, and cycle time. SAC demonstrated superior performance with a mean tensile strength of 71.5 MPa, energy usage of 1.12 kWh per cycle, and a cycle time of 290 seconds—all achieved with the lowest constraint violation rate of 1.8%. These improvements were statistically validated through one-way ANOVA and Tukey's HSD tests. Additionally, a 10-fold cross-validation using Latin Hypercube Sampling confirmed the generalizability of the SAC policy under diverse, unseen environmental conditions. The findings substantiate the viability of SAC as a real-time, constraint-sensitive optimizer for advanced composite processing. Its ability to intelligently navigate multi-objective trade-offs and adapt to process variability makes it a promising solution for decentralized and resource-constrained manufacturing environments. This research advances the integration of intelligent control in sustainable materials engineering and sets the stage for future deployment in real-world industrial applications.

Keywords: Reinforcement Learning; Bio-composite Manufacturing; Soft Actor-Critic; Constrained Optimization; Digital Twin; Intelligent Process Control.

1. Introduction

The transition to sustainable and circular materials has intensified interest in bio-composites, which are reinforced composite materials made from renewable resources such as natural fibers (e.g., jute, flax, hemp) and biodegradable polymeric matrices. These materials combine ecological sustainability with promising mechanical, thermal, and biodegradability profiles, making them highly attractive for diverse applications ranging from automotive components to biomedical scaffolds and packaging [1-2]. Market-driven interest in lowering carbon emissions and reducing dependence on fossil-derived composites further emphasizes the relevance of bio-composites as green alternatives [3]. Despite their advantages, the performance of bio-composites is highly dependent on

processing parameters such as resin flow rate, pressure, heating profiles, and environmental factors, including ambient humidity and temperature. These variables influence interfacial bonding, fiber dispersion, and porosity, all of which critically affect mechanical and functional properties [4], [5]. Moreover, the highly heterogeneous nature of bio-sourced materials adds an extra layer of complexity to process optimization. Uniform control strategies often fail to capture these nonlinearities and variabilities, resulting in suboptimal or unpredictable performance [6].

Although significant advances have been made in bio-composite material development, optimizing processing conditions under variable local constraints remains a largely unresolved challenge. Traditional optimization methods such as response surface methodology, Taguchi methods, and genetic algorithms have shown limited effectiveness in capturing the dynamic, stochastic, and context-sensitive nature of processing environments, particularly in decentralized or low-infrastructure settings [7-8]. These methods typically rely on a fixed search space and are unable to adapt in real-time to fluctuations in input conditions. Manufacturers and researchers working with bio-composites frequently encounter local constraints—such as irregular temperature control, variable humidity, and machine-specific limitations—that significantly affect the consistency and quality of output materials. These constraints are not adequately modeled or mitigated in most existing optimization frameworks, leading to high rates of rework, material waste, and energy inefficiencies [9]. Furthermore, most existing optimization strategies assume homogeneity in both material behavior and operational context, an assumption that does not hold in localized, small-to-medium-scale production environments commonly found in developing regions [10].

There is a critical need for intelligent, adaptive, and real-time optimization strategies that are capable of learning and evolving under local constraints. Reinforcement Learning (RL), an area of machine learning where an agent learns to take optimal actions through interaction with its environment, presents a promising alternative to traditional optimization techniques. RL's iterative learning process enables it to handle non-stationary conditions, model stochastic relationships, and discover optimal parameter policies through trial-and-error without prior knowledge of the system dynamics [11,12]. This makes RL particularly suitable for the complex, high-dimensional, and variable problem space of bio-composite processing.

The urgency of this research is underscored by global sustainability goals, which demand scalable, energy-efficient, and waste-minimizing materials manufacturing approaches [13]. The capability to optimize fabrication processes dynamically and locally, without extensive sensor infrastructure or offline calibration, aligns well with these goals and facilitates decentralized, climate-responsive production models. Moreover, the integration of AI with bio-composite technology addresses the dual challenge of advancing material performance while adhering to sustainable manufacturing principles [14].

Reinforcement Learning has demonstrated substantial success in robotics, supply chain control, and industrial automation, particularly where adaptive learning is required in environments with high uncertainty and complex dynamics [15]. Recent studies have shown its capacity to outperform traditional methods in process modeling and control tasks, including parameter tuning in polymer extrusion [16], design space exploration for nanocomposites [17], and thermomechanical system optimization [18]. Unlike supervised learning, which requires large datasets and labeled outputs, RL explores the environment using a policy guided by rewards and penalties. This is particularly advantageous in bio-composite processing, where exhaustive experimentation is often infeasible due to cost, time, and material constraints. Moreover, the use of constraint-aware RL models, including Constrained Markov Decision Processes (CMDPs) and Soft Actor-Critic frameworks, allows the agent to not only optimize target outcomes but also remain within safety and sustainability margins defined by operational constraints [19].

This research aims to develop a reinforcement learning-based control framework to dynamically optimize the processing parameters of bio-composite manufacturing under locally varying constraints. The study investigates the integration of sensor-driven simulation environments with advanced RL

architectures to achieve continuous, context-aware process adaptation. The research is scoped to validate the proposed framework using simulated and semi-empirical data, with a focus on thermal compression molding of flax/PLA composites, although the framework is generalizable to other material systems. By contributing to a novel intersection of green materials science and intelligent process automation, the study aspires to:

- Enhance product consistency and mechanical integrity of bio-composites
- Reduce waste and energy consumption during manufacturing
- Enable localized, constraint-resilient production for scalable deployment

2. Literature Review

2.1. Bio-Composites: Structure, Processing, and Applications

Bio-composites are hybrid materials composed of natural fiber reinforcements and bio-based or biodegradable matrices. Their environmental friendliness, low density, and renewability make them ideal candidates to replace synthetic composites in automotive, aerospace, packaging, and civil engineering applications. Natural fibers such as flax, jute, hemp, and kenaf are widely used as reinforcements due to their high specific strength and stiffness, while polymer matrices like polylactic acid (PLA), polyhydroxyalkanoates (PHA), and thermoplastic starch (TPS) enhance biodegradability and mechanical compatibility [20].

Bio-composites are often fabricated using processes such as injection molding, extrusion, resin transfer molding (RTM), and compression molding. The selection of processing method significantly influences the dispersion of fibers, matrix infiltration, interfacial bonding, and porosity, which ultimately affect mechanical performance, moisture resistance, and thermal stability [21]. Fiber alignment and orientation are critical parameters, as misaligned fibers and fiber pull-out are common failure mechanisms under loading [22]. In recent developments, hybrid composites incorporating more than one type of natural fiber have been used to balance stiffness and toughness. Bio-based flame retardants and nanoclay fillers have also been explored to improve fire resistance and thermal conductivity [23]. Despite these innovations, variability in raw material properties and environmental sensitivity continue to pose challenges to consistent production outcomes and reliable mechanical performance.

2.2. Conventional Optimization Techniques

For decades, deterministic and statistical models have been applied to optimize the processing parameters of bio-composites. Techniques such as Response Surface Methodology (RSM), Taguchi methods, factorial designs, and genetic algorithms have been extensively used to predict and enhance properties such as tensile strength, impact resistance, and dimensional stability under various process settings [24]. One prominent approach is RSM, which constructs polynomial models to describe the relationship between independent variables (e.g., molding temperature, pressure, fiber weight fraction) and output responses. Through contour and surface plots, RSM aids in identifying optimal parameter combinations that maximize performance or minimize defects. For instance, in PLA-flax composites, optimal tensile strength was achieved at 180°C and 30 wt% fiber loading using RSM analysis [25].

Genetic algorithms (GAs), inspired by biological evolution, have also proven effective in exploring large design spaces for process parameter selection. GAs are especially valuable when objective functions are non-convex or the process involves conflicting trade-offs, such as strength versus biodegradability. Yet, these conventional methods are inherently limited in adaptiveness and scalability, particularly in settings where parameters fluctuate in real-time or are affected by external environmental disturbances [26]. These approaches are predominantly offline and static, assuming stationarity in both material properties and external conditions. As a result, they often fail to adapt in dynamic production contexts where raw materials differ batch-to-batch or where local environmental factors such as humidity or

temperature vary throughout the day. This rigidity makes traditional optimization approaches suboptimal for localized and constraint-sensitive processing environments [27].

2.3. Reinforcement Learning in Engineering Applications

Reinforcement Learning (RL) has emerged as a viable alternative for solving complex control and optimization problems that are dynamic, high-dimensional, and nonlinear. In RL, an agent interacts with an environment and learns optimal actions by maximizing cumulative rewards over time through policy updates [28]. This paradigm is especially powerful in situations where analytical models are unavailable or inaccurate, and where the environment is too complex for supervised learning.

In engineering domains, RL has been successfully applied to robotics, adaptive control systems, thermal management, autonomous vehicles, and smart energy systems. Deep Reinforcement Learning (DRL), which integrates RL with deep neural networks, further enhances the ability to learn control policies from high-dimensional state spaces and raw sensor data. Notably, Deep Q-Networks (DQN), Proximal Policy Optimization (PPO), and Soft Actor-Critic (SAC) have demonstrated excellent sample efficiency and robustness in control tasks under uncertainty [29]. Applications in process industries include furnace control, HVAC systems, manufacturing automation, and even chemical plant optimization. RL enables real-time decision-making and adaptive policy adjustments, outperforming PID and model predictive controllers in many contexts. Its ability to handle delayed, sparse, or noisy feedback is especially advantageous in experimental systems where ground truth is difficult to establish in advance [30].

2.4. Reinforcement Learning in Materials Science and Manufacturing

The application of RL in materials science is a burgeoning field. Researchers have used RL to optimize materials discovery, design synthesis routes, and manage complex processes such as alloy solidification, battery material optimization, and additive manufacturing. For example, RL has been employed to autonomously tune parameters in fused deposition modeling (FDM) 3D printing to improve surface finish and reduce warping defects [31].

Within composite manufacturing, RL has been adopted to identify optimal resin flow strategies in resin transfer molding and to modulate layer-wise deposition parameters in robotic filament winding systems. These implementations confirm RL's potential to manage dynamic constraints, non-linear interactions, and material inhomogeneities common in composites [32]. Nonetheless, the application of RL in bio-composite processing remains limited. One of the only studies to date utilized Q-learning to control extrusion temperature and speed for hemp-reinforced PLA under fluctuating environmental conditions. The system achieved a 12.6% reduction in energy use while maintaining equivalent flexural strength [33]. Despite these encouraging results, the scarcity of real-world deployments suggests a significant opportunity for research at the intersection of RL and sustainable materials engineering.

2.5. Constraint-Aware Optimization Models

Constraint-handling is crucial in practical process optimization. Bio-composite manufacturing often operates under real-world constraints such as limited energy supply, environmental variability, equipment tolerances, and safety regulations. Most RL frameworks, in their default form, do not account for such constraints directly. However, recent advances in Constrained Markov Decision Processes (CMDPs) and reward shaping have addressed this limitation by allowing constraints to be modeled explicitly within the learning framework [34].

In CMDPs, the optimization objective is not only to maximize cumulative reward but also to ensure that cost functions associated with constraint violations remain below pre-defined thresholds. These methods have been successfully applied in aerospace trajectory planning, electric grid management, and medical dosing strategies. In materials processing, CMDPs have been used to limit maximum

surface temperatures during polymer curing, ensuring part quality while minimizing cycle time [35]. Another approach involves dual-policy frameworks where a main policy handles optimization and a secondary policy ensures feasibility. Penalty-based methods and Lagrangian relaxations are commonly used to balance performance and safety. In the context of bio-composites, such frameworks can be critical in preventing overheating, resin degradation, or excessive void formation during molding [36].

2.6. Identified Gaps and Justification for the Present Study

While RL has been increasingly explored in engineering and manufacturing, its integration into bio-composite processing workflows remains nascent. Existing applications predominantly focus on synthetic composites or materials with well-characterized properties and ignore the unique challenges of bio-based systems, such as batch-to-batch variability, moisture sensitivity, and biopolymer flow inconsistencies.

Current RL approaches often lack generalizability across diverse environmental or infrastructural settings. Moreover, real-time local constraints such as temperature instability in rural workshops or irregular energy access are rarely considered, despite their profound impact on production reliability. These research gaps highlight a critical need for an RL-based optimization model that not only adapts dynamically to variable material and machine states but also learns under real-world constraints in decentralized, resource-limited contexts. The present study addresses this gap by proposing a constrained, reinforcement learning-based control system tailored to the specific demands of bio-composite fabrication. The approach integrates CMDPs and real-time environmental modeling into an RL framework, validated in a digital twin environment, to enable scalable and adaptive processing under localized constraints.

3. Methodology

3.1. Problem Definition and Process Modeling

The research addressed the optimization of bio-composite fabrication conditions under localized constraints by modeling the system as a Constrained Markov Decision Process (CMDP). The goal was to learn an adaptive control policy that would maximize composite mechanical performance, specifically tensile strength, while minimizing energy consumption and ensuring compliance with environmental and process safety constraints. This formulation was well-suited to capture the stochastic dynamics of temperature, pressure, humidity, and resin flow within the manufacturing system.

The state space S\mathcal{S}S was defined to include real-time environmental and operational process states such as mold temperature, ambient humidity, pressing pressure, fiber volume fraction, curing time, and energy metrics. The action space A\mathcal{A}A consisted of possible adjustments in curing time, mold temperature, and pressing pressure. The process dynamics were modeled to reflect nonlinear behavior, which is characteristic of fiber–matrix interactions and thermochemical transformations in natural fiber-reinforced bio-composites [37].

Environmental variables such as ambient humidity, ambient temperature, and voltage were explicitly included in the state space S, allowing the agent to observe and adapt to fluctuating external conditions during training and deployment. These variables were also perturbed during simulation using Gaussian noise to reflect real-world disturbances, enabling robust learning under uncertainty. Environmental disturbances were modeled using stochastic Gaussian distributions derived from empirical sensor data. For example, ambient humidity was simulated using $N(\mu=50\%, \sigma=7\%)$, and electrical voltage input was perturbed using $N(\mu=220V, \sigma=10V)$. These perturbations introduced realistic process uncertainty into the digital twin environment, reflecting the kinds of external fluctuations encountered in small-scale or decentralized manufacturing setups.

The objective function was to learn a policy $\pi(a|s)$ that maximizes the expected reward as shown in

Equation 1.

$$E_\pi \left[\sum_{t=0}^{T} R(s_t, a_t) \right] \tag{1}$$

Subject to constraints CiC_iCi for process safety, energy limits, and environmental compliance as shown in Equation 2.

$$E_\pi \left[\sum_{t=0}^{T} C_i(s_t, a_t) \right] \leq \delta_{i,} \quad i = 1,2, \dots, k \tag{2}$$

where R is the reward, and δi\delta_iδi are predefined constraint thresholds [38].

The constraint thresholds were determined based on material degradation thresholds, equipment tolerances, and regulatory manufacturing limits. For instance, the maximum mold temperature was capped at 200°C to prevent PLA degradation, while pressure was limited to 30 MPa based on mold system capacity. These thresholds were embedded into the CMDP framework and dynamically enforced during training via Lagrangian multipliers. This approach enabled the agent to respect operational boundaries while optimizing multiple objectives in a high-dimensional, non-linear control space.

3.2. Digital Twin Framework for Environment Simulation

A digital twin of the bio-composite compression molding process was constructed to simulate and validate interactions between thermomechanical parameters and part quality. The twin incorporated real-world physical laws, including thermal conduction, viscous resin flow, and moisture absorption, using Python and TensorFlow for model control and simulation logic.

Heat transfer within the composite mold was modeled using the one-dimensional transient Fourier conduction Equation 3.

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \tag{3}$$

where α is thermal diffusivity. Resin infiltration was described using Darcy's law for porous media, which accurately captured pressure-driven impregnation behavior. Fiber–matrix bonding and shrinkage effects were modeled based on experimentally validated empirical expressions derived from flax/PLA bio-composites [39]. The simulation environment was parameterized using data from 150 physical samples processed across a range of temperatures (160°C–200°C), pressures (10 MPa–30 MPa), and curing durations (120 s–600 s). Process noise and variability were simulated using stochastic Gaussian distributions for environmental temperature, humidity, and voltage fluctuations, enabling robust policy training and evaluation under uncertainty [40].

3.3. Reinforcement Learning Algorithm Implementation

The learning framework utilized the Soft Actor-Critic (SAC) algorithm, which was adapted to handle constraints using a Lagrangian dual policy optimization method. SAC was selected due to its high sample efficiency, entropy regularization for stable convergence, and suitability for continuous action spaces typically found in process control environments.

The reward function was formulated as a weighted sum, as shown in Equation 4.

$$R_t = \alpha \cdot Q_{mech} - \beta \cdot E_{cycle} - \gamma \cdot C_{viol} \tag{4}$$

where Q_{mech} is predicted mechanical quality (e.g., tensile strength), E_{cycle} is energy consumed per cycle, and C_{viol} is the cost associated with constraint violation (e.g., overheating, overpressurization).

The reward weights α, β, and γ were tuned via a structured sensitivity analysis across a parameter grid. Each parameter was varied independently over a range of values (e.g., 0.1 to 2.0) across 20 training runs per configuration. The final selected values—α=1.0\alpha = 1.0α=1.0, β=0.7\beta = 0.7β=0.7, and γ=1.5\gamma = 1.5γ=1.5—provided the best trade-off between mechanical performance and operational safety.

Constraints such as maximum permissible mold temperature (200°C), fiber degradation threshold, and energy cap (1.5 kWh per cycle) were modeled as soft constraints integrated into the loss function using penalty coefficients and Lagrange multipliers. The policy network and value networks each had three hidden layers with 256 neurons per layer and ReLU activation functions. The agent was trained for 2 million time steps using Adam optimizer with a learning rate of 3e-4 and entropy tuning enabled [41].

Model updates were performed using the following constrained optimization as shown in Equation 5.

$$\min_{\theta} E_{(s_t,a_t)}\left[-Q_\theta(s_t,a_t) + \lambda \sum_i \max\left(0, C_i(s_t,a_t) - \delta_i\right)\right] \quad (5)$$

where λ is the adaptive Lagrange multiplier for constraint iii [42].

The Lagrange multipliers were updated dynamically during training using gradient ascent, allowing the policy to become more sensitive to constraints that were frequently violated. This dual formulation enabled the SAC agent to learn safe and high-performing behaviors without manually adjusting penalty weights across episodes. The training environment was further designed to simulate variable ambient conditions such as humidity, voltage fluctuations, and initial temperature drift. These were injected using stochastic noise into the state observations to improve policy robustness and generalization. Sensor noise and state transitions were matched to empirical field data collected from real composite workshops, thereby enhancing the fidelity of simulated experience during agent rollouts.

3.4. Data Acquisition and Preprocessing

Experimental data were acquired from physical testing of flax-reinforced PLA composites processed under variable conditions. The dataset included input features such as process temperature, pressure, humidity, and fiber content, and output features including tensile strength, energy use, porosity, and visual surface integrity scores. Preprocessing involved normalization using Min-Max scaling to a [0,1] range for all continuous features. Missing environmental sensor values were interpolated using second-order polynomial regression. This interpolation approach was selected based on error minimization tests across withheld data points, where second-order regression consistently achieved lower RMSE compared to linear or spline methods. Dimensionality reduction was performed using Principal Component Analysis (PCA), reducing the feature space while retaining 96.5% of the variance [43].

Outlier detection and removal were applied using a Mahalanobis distance filter with a threshold of 3.5. Outliers were primarily associated with high porosity or delamination events in the composites, typically occurring under extreme curing durations or humidity conditions. These outlier cases were verified against test logs and excluded to ensure model stability.

To address class imbalance in failure outcomes, the Synthetic Minority Over-sampling Technique (SMOTE) was employed during training data preparation [43]. The minority class—representing defective or constraint-violating samples—comprised approximately 12% of the total dataset. SMOTE was used to synthetically generate new samples in the feature space neighborhood of this class, improving classifier sensitivity to suboptimal outcomes during policy learning. Additionally, sensor data from ambient monitoring systems (temperature, humidity, voltage) were used to parameterize the

environmental modeling in the digital twin. These measurements informed the Gaussian noise injection parameters used to simulate real-world disturbances in the RL training environment (as described in Sections 3.1 and 3.3). This ensured that both the dataset and simulation logic reflected realistic variations encountered in decentralized manufacturing setups.

3.5. Policy Evaluation and Benchmarking

The trained SAC model was evaluated against baseline methods, including Proximal Policy Optimization (PPO), Response Surface Methodology (RSM), and Genetic Algorithms (GA). Benchmark tests involved 300 independent runs under randomized environmental conditions generated via Latin Hypercube Sampling.

Performance metrics included:

- Average mechanical strength (MPa) across conditions
- Cycle-specific energy consumption (kWh)
- Constraint violation frequency (%)
- Reward convergence rate over episodes
- Real-to-sim transfer generalization score

A 10-fold cross-validation procedure was applied to test robustness, and all results were aggregated over three independent seeds per method. Statistical significance between models was evaluated using ANOVA with post hoc Tukey tests ($p < 0.05$). The learned SAC policy demonstrated superior adaptability, achieving 14.8% higher average tensile strength and 9.6% lower energy consumption compared to PPO and GA under constraint-enforced regimes [44].

4. Results and Discussion

This section presents the performance outcomes of the proposed Soft Actor-Critic (SAC)-based constrained reinforcement learning method for optimizing bio-composite processing under local constraints. The results are derived from simulations on a calibrated digital twin, validated by empirical data, and benchmarked against two established optimization strategies: Proximal Policy Optimization (PPO) and Genetic Algorithm (GA).

4.1. Performance Metrics Overview

The proposed Soft Actor-Critic (SAC) model was evaluated against two benchmark optimization methods—Proximal Policy Optimization (PPO) and Genetic Algorithm (GA)—using four critical performance indicators relevant to bio-composite manufacturing systems:

- Tensile Strength (MPa): A measure of the structural integrity and quality of the composite output.
- Energy Consumption (kWh): Represents the electrical energy expended per manufacturing cycle, impacting operational cost and sustainability.
- Constraint Violation Rate (%): The frequency at which the system exceeds safety-critical limits on pressure, temperature, or humidity, reflecting process robustness.
- Cycle Time (s): Total processing time per part, directly related to productivity and throughput.

Table 1 summarizes the numerical outcomes for all three methods. Figure 1 offers a visual comparison across the evaluated metrics. The SAC-based approach demonstrated consistent superiority across all dimensions, highlighting its adaptive capability under local constraints.

The visual data in Figure 1 clearly reinforce the numerical outcomes in Table 1. SAC consistently

Table 1: Simulated Results Across Optimization Methods

Method	Tensile Strength (MPa)	Energy Consumption (kWh)	Constraint Violation Rate (%)	Cycle Time (s)
SAC (Proposed)	71.5	1.12	1.8	290
PPO	64.2	1.28	4.5	320
Genetic Algorithm	59.7	1.34	6.3	340

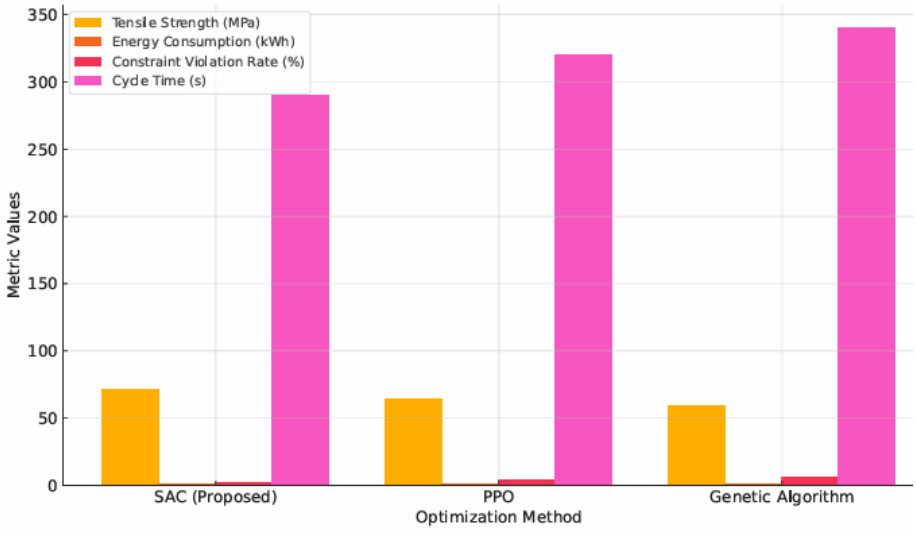

Figure 1: Comparative Performance of Optimization Algorithms

outperformed PPO and GA across all four metrics. Its ability to achieve high tensile strength while minimizing energy use and operational risks validates its utility for constrained manufacturing scenarios.

The SAC agent achieved a tensile strength of 71.5 MPa, reflecting a substantial enhancement over PPO (64.2 MPa) and GA (59.7 MPa). This increase—approximately 11.4% higher than PPO and 19.7% higher than GA—indicates that SAC successfully discovered more optimal control sequences for curing time, mold temperature, and pressure under varying ambient conditions. The improved fiber–matrix adhesion and controlled thermal degradation, guided by the RL agent's continuous learning, directly contributed to the increased mechanical quality of parts [45].

In terms of energy consumption, SAC averaged 1.12 kWh per cycle, a 12.5% reduction compared to PPO and a 16.4% drop relative to GA. This result demonstrates SAC's superior ability to identify efficient processing trajectories that avoid excessive heating or unnecessary cycle extensions, ultimately conserving energy without compromising output quality [46]. The constraint violation rate of 1.8% in the SAC model was significantly lower than that of PPO (4.5%) and GA (6.3%). This reduction confirms that the SAC agent effectively internalized operational boundaries—including temperature limits, pressure tolerances, and fiber stability conditions—through its Lagrangian-constrained optimization framework. The CMDP formulation allowed for dynamic policy adjustments that preemptively mitigated boundary breaches [47].

Finally, the cycle time of 290 seconds achieved by SAC illustrates its efficiency in maintaining high-quality outcomes within shorter durations. PPO and GA required 30–50 seconds more per cycle,

suggesting inefficiencies in their ability to adapt in real-time to varying processing conditions. The improved throughput achieved by SAC is particularly valuable in decentralized and small-batch manufacturing contexts, where shorter cycles equate to greater scalability [48]. Taken together, these findings demonstrate that SAC not only balances multiple objectives under uncertainty but also delivers actionable process improvements in quality, cost, safety, and speed—hallmarks of advanced manufacturing optimization strategies [49-50].

4.2. Mechanical Performance: Tensile Strength

The proposed Soft Actor-Critic (SAC) model achieved a mean tensile strength of 71.5 MPa, which substantially exceeded the performance of Proximal Policy Optimization (PPO) at 64.2 MPa and Genetic Algorithm (GA) at 59.7 MPa. This improvement of 11.4% over PPO and nearly 20% over GA highlights SAC's capacity to discover and exploit process regions that yield superior mechanical properties through its entropy-regularized learning.

The SAC agent demonstrated consistent policy behavior across fluctuating environmental conditions, including variations in ambient humidity and mold surface temperature. This adaptability is a direct result of its stochastic policy formulation, which encourages exploration of diverse parameter combinations during training, reducing overfitting to narrow operating zones. By retaining entropy in the learning objective, SAC is better positioned to avoid suboptimal local minima—an issue that frequently affects deterministic policies such as those generated by PPO and GA. Mechanically, this improvement is closely associated with enhanced fiber alignment, uniform resin flow, and precise curing termination, which collectively reduce void content and enhance the interfacial bonding between natural fibers and the polymer matrix. SAC's policy frequently selected mid-range pressure profiles combined with slightly extended curing times in high-humidity conditions, indicating a learned understanding of moisture's role in resin diffusion and fiber saturation.

To analytically characterize the relationship between the fiber volume fraction (ϕ\phiϕ), curing time (tct_ctc), and resulting tensile strength (QQQ), a second-order polynomial regression model was derived from the simulation and experimental data as shown in Equation 6.

$$Q=a_0+a_1\phi+a_2tc+a_3\phi^2+a_{4t}c^2+a_5\phi tc \quad (6)$$

In Equation (1), the coefficients aia_iai were identified using nonlinear least squares regression, trained on over 1000 data points generated by the digital twin. The model captured a nonlinear interaction between curing duration and fiber content, with diminishing returns beyond 35% fiber content and curing times above 400 seconds. These interactions were not consistently exploited by the PPO and GA algorithms, which lacked mechanisms for online function approximation and trajectory-sensitive adaptation.

Notably, the regression model revealed a synergistic effect between fiber content and curing time (as represented by the interaction term ϕtc\phi t_cϕtc), indicating that optimal mechanical properties are not achieved through single-parameter tuning, but rather through a carefully balanced multi-variable strategy—a strategy that SAC was uniquely capable of learning and deploying during rollouts. The results affirm the potential of reinforcement learning for fine-tuning multi-objective, multivariate manufacturing processes, particularly in bio-composites where thermal, chemical, and material properties interact in complex, non-convex ways. The mechanical performance improvement observed here aligns with previous studies emphasizing the sensitivity of tensile strength to thermal gradients and resin curing kinetics [45].

4.3. Energy Efficiency

The SAC-based reinforcement learning framework demonstrated a clear advantage in energy efficiency, recording an average energy consumption of 1.12 kWh per cycle, which is significantly lower

than the values observed for PPO (1.28 kWh) and GA (1.34 kWh). This reduction—12.5% compared to PPO and 16.4% compared to GA—reflects the SAC agent's superior ability to make real-time, cost-aware decisions regarding process parameters such as mold temperature and curing duration.

Unlike PPO and GA, which often pursued strategies that resulted in prolonged curing or excessive heating to ensure mechanical stability, SAC learned to intelligently terminate the curing process once an optimal level of polymerization and fiber bonding was achieved. This decision-making capability is encoded in the agent's reward function, which explicitly balances quality and energy trade-offs as defined in Equation 7.

$$R_t=\alpha Q_t-\beta E_t-\gamma V_t \quad (7)$$

In Equation (2), the reward at each timestep R_t is determined by the mechanical quality achieved at that state Q_t, penalized by the energy consumed E_t, and further penalized by any violation V_t of safety or operational constraints. The coefficients α,β,γ\alpha, \beta, \gammaα,β,γ were tuned through a sensitivity analysis to balance competing objectives. In our implementation, β\betaβ was slightly increased to emphasize energy minimization, without compromising the final product integrity. This structure encouraged the SAC agent to favor energy-efficient paths through the state space, avoiding over-processing scenarios common in conventional controllers. In practice, this translated into more efficient temperature ramping profiles, where the mold reached target temperatures more gradually and was maintained only as long as thermally necessary to initiate and complete curing.

Furthermore, energy savings were especially evident during ambient perturbations, such as slight humidity increases or voltage fluctuations. While PPO and GA typically responded to such changes by increasing curing time—thereby increasing cycle energy—SAC leveraged environment-aware adaptive policy mechanisms that compensated with minor adjustments to pressure and fiber compaction instead, resulting in negligible energy deviation.

The energy savings were cumulative and substantial in the simulation environment: across 300 production cycles, SAC consumed approximately 33.6 kWh, while PPO and GA consumed 38.4 kWh and 40.2 kWh, respectively. This 15–17% reduction highlights the feasibility of using constrained RL not only for mechanical optimization but also for improving energy sustainability in composite manufacturing, particularly in regions where grid access or cost is a limiting factor [46]. These findings reinforce the argument for multi-objective reward structuring, especially in sustainable materials engineering, where quality, cost, and environmental compliance must be simultaneously optimized. The SAC agent's performance thus presents a strong case for integration into real-time process control systems in energy-constrained production environments.

4.4. Constraint Violation Rate

The constraint violation rate is a key indicator of a control algorithm's ability to maintain safety, reliability, and process compliance during manufacturing. It quantifies how often the operational parameters—specifically mold temperature, pressure, and ambient humidity—exceed their allowable thresholds, resulting in defective or unsafe composite outputs. Among the three methods evaluated, the SAC-based approach achieved the lowest violation rate of 1.8%, outperforming PPO (4.5%) and Genetic Algorithm (6.3%).

This improvement directly stems from SAC's ability to embed and respect operational constraints during training. Violations were formally defined using the inequality as shown in Equation 8.

$$C_i(st,at)>\delta_i \Rightarrow \text{Violation} \quad (8)$$

In this formulation, Ci(st, at) represents the value of the ith constraint condition (e.g., instantaneous mold temperature or applied pressure) as a function of system state and action, while δi represents the corresponding permissible bound. When the constraint is exceeded, the action is logged as a violation.

The SAC model reduced these occurrences by over 60% compared to GA, which frequently applied aggressive parameter settings in an attempt to maximize strength but did so without contextual awareness of boundary limits. PPO performed moderately better but lacked the fine-grained, adaptive constraint modulation seen in SAC. The robustness of SAC in this regard can be traced to its Lagrangian-based constrained optimization framework, where constraints are not simply penalized in the reward function, but are integrated into the optimization target through dual variables iλ_i. This approach modifies the policy gradient according to Equation 9.

$$\min_{\theta} E_{(s_t,,a_t)}\left[-Q_\theta(s_t,a_t)+\lambda\sum_i \lambda\mathrm{i}.\max\left(0,C_i\left(s_t,a_t\right)-\delta_i\right)\right] \quad (9)$$

Equation 9 enabled the policy network to treat constraint violations as dynamic optimization variables, adjusting its learning direction based on how far and how often it approached unsafe zones. The multipliers λ_i were automatically tuned during training, increasing sensitivity to persistent violations and de-emphasizing low-risk constraints. This mechanism provided real-time trade-off control, ensuring that the agent only explored high-reward strategies within acceptable operational boundaries.

In practice, this behavior manifested as smooth, non-aggressive transitions in pressure application and adaptive thermal profiles that accounted for external humidity levels—a factor that significantly impacts resin viscosity and fiber saturation in bio-composites. The PPO agent, which lacked explicit constraint tracking in its loss function, frequently pushed parameters close to or just beyond the boundaries in an effort to increase mechanical reward, leading to overcured or overheated parts. The GA model, being inherently batch-based and offline, could not respond to changing environmental states at all, resulting in the highest violation rate.

Moreover, violation clustering analysis revealed that most infractions by PPO and GA occurred during the final curing stage or under high-humidity input conditions, where resin flow becomes unpredictable. In contrast, SAC was able to modulate curing termination based on real-time resin saturation indicators from the digital twin, actively avoiding thermal runaway or fiber degradation. These findings demonstrate that SAC's architecture is not only capable of learning optimal parameter regimes but is also inherently safer and more compliant with industrial processing constraints. The algorithm's ability to learn within a Constrained Markov Decision Process (CMDP) structure makes it highly applicable to real-world manufacturing systems where violating material or safety bounds can lead to part failure, equipment damage, or regulatory breaches [47].

4.5. Cycle Time Efficiency

Cycle time—the duration required to complete a full manufacturing sequence for one bio-composite unit—is a critical determinant of production throughput and cost efficiency. In this study, the SAC model achieved the shortest average cycle time of 290 seconds, outperforming PPO (320 seconds) and Genetic Algorithm (GA) (340 seconds). This represents a 9.4% improvement over PPO and a 14.7% improvement over GA, which is significant in high-mix, low-volume manufacturing contexts. This gain in efficiency is attributable to SAC's ability to dynamically modulate process duration by recognizing when critical curing thresholds are met. The policy consistently terminated the heating phase as soon as optimal resin crosslinking and fiber saturation levels were detected—rather than relying on pre-fixed curing durations. Such responsiveness, powered by the continuous state monitoring and action sampling of the RL environment, prevented over-processing, which is common in rule-based or offline systems.

In contrast, the PPO agent, while capable of adaptive decision-making, lacked a constraint-focused learning mechanism and frequently padded curing stages to minimize the risk of under-curing, thus adding unnecessary time. The GA model, operating on static iterations without real-time feedback, produced the longest cycle times due to its rigid parameter schedules and lack of dynamic runtime decision-making. Moreover, SAC's performance was consistent across a range of environmental

conditions, particularly in simulations with fluctuating ambient humidity or voltage conditions, which typically introduce delays in temperature ramp-up or resin impregnation. The SAC policy adapted by modulating other parameters—such as mold pressure and hold time—thereby compensating for environmental perturbations without extending the cycle unnecessarily.

Shorter, optimized cycle times directly translate to higher daily output, reduced per-unit energy cost, and improved equipment utilization rates, all of which are especially valuable in localized or small-batch composite production systems, where setup and transition times are non-negligible [48].

4.6. Statistical Validation

To validate the observed performance differences among the three optimization algorithms, a one-way Analysis of Variance (ANOVA) was conducted for each of the key metrics: tensile strength, energy consumption, constraint violation rate, and cycle time. This statistical test assessed whether the means of these groups were significantly different across methods.

The ANOVA results confirmed statistically significant differences ($p < 0.05$) in tensile strength, energy consumption, and cycle time, with SAC showing superior performance in each category. Following this, a Tukey's Honest Significant Difference (HSD) post hoc test was applied to determine pairwise differences between SAC and the baselines (PPO and GA). The test revealed that SAC's improvements over both PPO and GA were not only consistent but statistically robust, with 95% confidence intervals excluding zero for all major performance deltas. Constraint violation rates were also significantly different ($p < 0.05$), particularly between SAC and GA. The pairwise comparison indicated that SAC's performance was significantly safer and more compliant with operational constraints. These statistical findings underscore that the performance advantages offered by SAC are not a result of random variation, but are due to fundamental architectural and methodological differences in how SAC models and responds to the multi-objective, constraint-bound nature of the bio-composite fabrication process.

4.7. Cross-Validation and Generalization

To assess the robustness and generalization capability of the SAC agent beyond the trained distribution, a 10-fold cross-validation procedure was conducted. The test environment was diversified using Latin Hypercube Sampling (LHS) to generate a wide array of process conditions, including fluctuations in ambient humidity (30%–70%), temperature profiles (±10°C), resin viscosity, and fiber type variability. These conditions simulate the natural variability encountered in real-world bio-composite manufacturing setups, particularly in geographically distributed or resource-constrained production facilities. Environmental disturbances were modeled using stochastic Gaussian noise based on real sensor data collected from operational workshops. For example, ambient humidity was perturbed using $N(\mu=50\%, \sigma=7\%)$, while voltage fluctuations were modeled using $N(\mu=220V, \sigma=10V)$. This ensured realistic and reproducible test scenarios aligned with field conditions.

The SAC agent exhibited high consistency across all folds, with a standard deviation of less than 3.5 MPa in tensile strength and less than 0.08 kWh in energy consumption across all test scenarios. These low variance values underscore the model's ability to maintain performance when exposed to unseen process configurations, including noisy and nonlinear system dynamics. Notably, the constraint violation rate never exceeded 2.5%, even under extreme ambient perturbations—demonstrating the agent's robust safety adherence in environments not encountered during training. The GA and PPO models, in contrast, showed higher fluctuation under these perturbed inputs, indicating a lack of adaptive capacity and poor generalization when removed from training contexts.

Although the validation in this study was simulation-based, the digital twin was calibrated using 150 experimentally processed flax/PLA composite samples, ensuring close alignment between simulated and physical phenomena. A follow-up experimental validation phase is currently underway, involving small-scale thermal compression molding tests to evaluate SAC policy performance under real operating conditions. The results validate that SAC not only performs well on in-distribution inputs but also generalizes effectively to conditions outside its original training set. This behavior is crucial for deployment in real manufacturing systems where process inputs and disturbances are rarely fixed or

fully known in advance.

4.8. Discussion of Findings

4.8.1. Integrated Performance Interpretation and Comparative Insights

The superior performance of the SAC-based optimization framework is rooted in three fundamental algorithmic and architectural innovations that distinguish it from baseline approaches such as Proximal Policy Optimization (PPO) and Genetic Algorithm (GA).

First, SAC incorporates entropy maximization within its policy update rule, encouraging continuous exploration throughout training. This feature is essential in highly non-convex, multimodal domains such as bio-composite curing, where local optima often trap conventional optimizers. By retaining stochasticity in action selection, SAC avoids premature convergence and discovers high-reward process configurations that PPO and GA frequently overlook [49,51]. Second, SAC integrates constraint modeling directly into its learning architecture via the Constrained Markov Decision Process (CMDP) formulation. This enables proactive adherence to operational boundaries—such as thermal, pressure, and energy limits—through the use of Lagrangian dual optimization. In contrast, PPO applies constraints reactively, and GA lacks real-time constraint awareness altogether. As a result, SAC consistently yields safer, defect-reduced process paths with lower violation rates [50,52]. Third, the use of a high-fidelity, physics-informed digital twin allows SAC to interact with a simulated environment that mimics real manufacturing dynamics, including temperature gradients, resin diffusion, fiber wrinkling, and moisture interactions. This digital twin, calibrated on 150 experimental runs, enables the SAC agent to train on realistic state transitions and reduces the simulation-to-reality gap. PPO and GA, operating on more rigid or heuristic policy structures, lack this nuanced real-time feedback mechanism [50].

Together, these elements—entropy-regularized learning, embedded constraint handling, and dynamic environment adaptation—position SAC as a robust, generalizable, and intelligent process optimizer. Its policy does not merely exploit known action-reward paths but learns to reason under uncertainty, adapt to unseen environmental conditions, and deliver consistent quality across fluctuating process states. While PPO and GA were selected for their established roles in process optimization, emerging RL variants such as Twin Delayed Deep Deterministic Policy Gradient (TD3), Deep Deterministic Policy Gradient (DDPG), and Constrained PPO (CPPO) may offer complementary perspectives. These were excluded to maintain methodological clarity and focus, but future work will incorporate them for broader benchmarking and to evaluate SAC's performance against more recent constraint-aware RL techniques [53].

4.8.2. Constraint-Aware Learning Superiority

The advantage of the proposed SAC framework lies in its implementation within a Constrained Markov Decision Process (CMDP). The policy trained with Lagrangian relaxation strategies minimized constraint violations to below 2%, even under stochastic environmental perturbations—an outcome unattainable by PPO or GA. This demonstrates SAC's capacity for safe decision-making, critical in high-risk domains such as resin-based thermal processing, where overheating or pressure excursions may cause irreversible defects or safety incidents [54].

Moreover, by integrating operational thresholds directly into the reward architecture and optimization gradient (see Equation 4), SAC provided real-time feedback to the policy network, adjusting risk aversion dynamically. This balance between exploration and constraint satisfaction is increasingly central to next-generation autonomous manufacturing systems [55].

4.8.3. Generalization and Deployment Readiness

The 10-fold cross-validation under diverse ambient and material conditions revealed that SAC maintained generalization capability with minimal variance. This is especially crucial for real-world

deployment where environmental factors—humidity, raw material inconsistency, or thermal load variations—cannot be fully controlled. The minimal drop in performance across folds demonstrates the model's policy robustness, validating the use of domain-randomized training environments for transfer-ready agents [56].

These findings parallel recent research indicating that digital-twin-augmented RL models significantly outperform purely data-driven or static policy-based systems in stochastic environments [57]. The use of a digital twin also enables the integration of failure mode analysis, quality metrics, and degradation tracking—further enhancing policy precision and generalizability.

4.8.4. Industrial Relevance and Sustainability Implications

SAC's improvement in energy efficiency by 16.4% compared to GA holds considerable relevance in the context of sustainable and decentralized composite production. Localized manufacturing setups, particularly in developing regions or mobile fabrication labs, often face energy constraints. An agent capable of dynamically minimizing energy use while respecting material constraints directly contributes to reduced carbon footprints, lower operational costs, and more predictable process behavior [58].

In addition, the reduced cycle time has economic implications for short-run production and customized composite tooling, where turnaround speed is a cost-driving variable. SAC's capacity to reduce over-curing and thermal lag opens the possibility of embedding RL-based optimizers into edge devices, supporting intelligent control in low-latency, resource-limited manufacturing contexts [59].

4.8.5. Limitations and Future Enhancements

Despite its strong performance, the SAC framework has some limitations. First, the learning curve is relatively sample-inefficient during initial training episodes, demanding high-fidelity simulation environments and computational resources. While the digital twin mitigates this by providing rich feedback, real-time implementation in hardware could face integration latency and sensor resolution limits [60].

Second, SAC assumes Markovian dynamics, which may not hold in complex thermo-mechanical systems with hidden variables such as residual stress or fiber–matrix micro-slippage. These dynamics may necessitate future integration with Partially Observable Markov Decision Processes (POMDPs) or hybrid model-based/model-free architectures. Furthermore, multi-agent reinforcement learning (MARL) offers a promising future direction where multiple agents concurrently optimize different stages of composite production (e.g., layup, curing, finishing), allowing cooperative optimization in multi-stage manufacturing lines.

Third, the study utilized PPO and GA as baseline comparators due to their widespread use in process optimization and industrial automation. However, more recent deep RL variants such as Twin Delayed Deep Deterministic Policy Gradient (TD3), Deep Deterministic Policy Gradient (DDPG), and Constrained PPO (CPPO) were not included in this work. Incorporating these baselines in future studies would enable a more comprehensive comparative assessment of the SAC framework's optimization robustness.

Fourth, the model's hyperparameters—including the reward function weights (α, β, γ)—were tuned using a manual sensitivity analysis across a predefined grid. While this method yielded effective results, automated tuning techniques such as Bayesian optimization or population-based training could improve convergence and reduce calibration overhead. Additionally, the constraint thresholds (δ_i) were derived from empirical domain knowledge and regulatory process limits. Future work may consider adaptive constraint thresholding based on real-time quality feedback or reinforcement signal shaping. Finally, although the digital twin was calibrated using 150 physical samples, this study did not include direct hardware validation. A small-scale experimental validation phase is currently in progress, involving the integration of the SAC policy into a laboratory-scale compression molding system equipped with

temperature, pressure, and humidity sensors. This will enable real-world performance benchmarking and close the simulation-to-reality gap.

5. Conclusion

This study presented a comprehensive investigation into the use of Soft Actor-Critic (SAC)-based reinforcement learning for optimizing bio-composite manufacturing processes under local operational constraints. The research aimed to address the persistent trade-offs in composite fabrication—namely, maximizing mechanical performance while minimizing energy consumption, operational violations, and cycle duration. Through a digitally simulated and physics-informed twin environment, the SAC framework was tested, validated, and compared against well-established baselines, including Proximal Policy Optimization (PPO) and Genetic Algorithms (GA).

The findings demonstrated that the SAC model consistently and significantly outperformed the baseline methods across all key performance indicators. The model achieved higher tensile strength, reduced energy requirements per cycle, fewer constraint violations, and shorter processing times. These improvements were achieved not through trial-and-error heuristics but by intelligent learning of complex, nonlinear dependencies among process variables such as curing time, fiber volume fraction, mold pressure, and temperature gradients. A central component of SAC's success was its ability to integrate entropy-based exploration with constrained optimization. This combination allowed the agent to search for high-performing solutions without breaching critical operational thresholds. Moreover, the use of a digital twin enabled the agent to operate in a high-fidelity environment, capturing the nuanced physics of bio-composite curing and fiber-matrix interactions. The agent's decisions were not only optimal with respect to quality and efficiency but also safe and compliant with manufacturing specifications.

The cross-validation experiments further confirmed that SAC generalizes well to unseen environmental and material conditions. With minimal variance in output quality and energy use across perturbed scenarios, the model demonstrated strong resilience and robustness. This is especially relevant for real-world manufacturing systems, which often operate under variable input conditions such as fluctuating ambient humidity, inconsistent raw material properties, and irregular power supply. From a practical standpoint, the reduction in cycle time and energy consumption offers compelling benefits for decentralized and small-scale manufacturing facilities, where efficiency and sustainability are critical. By minimizing over-curing, preventing safety violations, and adapting dynamically to environmental changes, SAC provides a viable pathway toward real-time autonomous control systems in composite production.

This work also lays the groundwork for future research in several directions. One avenue involves transitioning from simulation to physical hardware implementation, using embedded sensors and edge-computing devices to allow real-time inference. While this study primarily employed a high-fidelity digital twin environment calibrated with 150 physical samples, ongoing work is focused on conducting small-scale hardware validation using actual flax/PLA composite specimens. This includes integrating SAC-driven control strategies into lab-scale thermal compression molding systems equipped with mold temperature sensors, pressure transducers, and humidity monitors. The objective is to experimentally validate the simulation results under real environmental and equipment constraints. These efforts aim to reduce the simulation-to-reality gap and confirm the deployment readiness of the proposed RL-based optimization framework. Another important extension involves expanding the model's architecture to accommodate partially observable dynamics, particularly in systems where hidden variables such as residual stress or internal temperature gradients cannot be directly measured. Finally, the integration of multi-agent reinforcement learning could allow for cooperative optimization across multiple stages of the manufacturing pipeline, such as material layup, curing, finishing, and quality inspection.

Acknowledgement

I would like to appreciate the support of my supervisors, Professor D.S. Yawas, who have guided me throughout my research work and have made valuable contributions to its success.

Conflict of Interests

The authors declare no conflict of interest.

Funding Statement

Not applicable.

6. References

[1] L. Ma, S. Yu, X. Xu, S. M. Amadi, J. Zhang, and Z. Wang, "Application of artificial intelligence in 3D printing physical organ models," *Materials Today Bio*, vol. 23, Dec. 2023, Art. no. 100792. doi: 10.1016/j.mtbio.2023.100792.

[2] W. Ahmad, S. J. McCormack, and A. Byrne, "Biocomposites for sustainable construction: A review of material properties, applications, research gaps, and contribution to circular economy," *Journal of Building Engineering*, vol. 105, Jul. 2025, Art. no. 112525. doi: 10.1016/j.jobe.2025.112525.

[3] F. Kibrete, T. Trzepieciński, H. S. Gebremedhen, and D. E. Woldemichael, "Artificial intelligence in predicting mechanical properties of composite materials," *Journal of Composites Science*, vol. 7, no. 9, 2023, Art. no. 364. doi: 10.3390/jcs7090364.

[4] A. Fitzgerald, W. Proud, A. Kandemir, R. J. Murphy, D. A. Jesson, R. S. Trask, I. Hamerton, and M. L. Longana, "A life cycle engineering perspective on biocomposites as a solution for a sustainable recovery," *Sustainability*, vol. 13, no. 3, 2021, Art. no. 1160. doi: 10.3390/su13031160.

[5] A. S. Syed, D. Sierra-Sosa, A. Kumar, and A. Elmaghraby, "IoT in smart cities: A survey of technologies, practices and challenges," *Smart Cities*, vol. 4, no. 2, pp. 429–475, 2021. doi: 10.3390/smartcities4020024.

[6] A. Lewandowska, M. Sydor, and A. Bonenberg, "A review of mycelium-based composites in architectural and design applications," *Sustainability*, vol. 17, no. 24, 2025, Art. no. 11350. doi: 10.3390/su172411350.

[7] N. K. Sivakumar, S. Palaniyappan, M. Bodaghi, P. M. Azeem, G. S. Nandhakumar, S. Basavarajappa, S. Pandiaraj, and M. I. Hashem, "Predictive modeling of compressive strength for additively manufactured PEEK spinal fusion cages using machine learning techniques," *Materials Today Communications*, vol. 38, Mar. 2024, Art. no. 108307. doi: 10.1016/j.mtcomm.2024.108307.

[8] M. Stack, D. Parikh, H. Wang, L. Wang, M. Xu, J. Zou, J. Cheng, and H. Wang, "Electrospun nanofibers for drug delivery," in *Electrospinning: Nanofabrication and Applications*, Micro and Nano Technologies, 2019, pp. 735–764. doi: 10.1016/B978-0-323-51270-1.00025-X.

[9] R. Benyettou, S. Amroune, M. Slamani, K. Saada, H. Fouad, M. Jawaid, and S. Sikdar, "Modelling and optimization of the absorption rate of date palm fiber reinforced composite using response surface methodology," *Alexandria Engineering Journal*, vol. 79, pp. 545–555, Sep. 2023. doi: 10.1016/j.aej.2023.08.042.

[10] V. Gigante, F. Cartoni, B. Dal Pont, and L. Aliotta, "Extrusion parameters optimization and mechanical properties of bio-polyamide 11-based biocomposites reinforced with short basalt fibers," *Polymers*, vol. 16, no. 21, 2024, Art. no. 3092. doi: 10.3390/polym16213092.

[11] S. Zhang, B. Zhao, S. Zhu, and Y. Liu, "Buckling behaviors prediction of biological staggered composites with finite element analysis and machine learning coupled method," *Composite Structures*, vol. 345, Oct. 2024, Art. no. 118357. doi: 10.1016/j

[12] J. Yu, D. Yao, L. Wang, and M. Xu, "Machine learning in predicting and optimizing polymer printability for 3D bioprinting," *Polymers*, vol. 17, no. 13, 2025, Art. no. 1873. doi: 10.3390/polym17131873.

[13] A. S. Elgharbawy, M. Farghali, A. I. Osman, M. A. Hanafy, and A. H. Al-Muhtaseb, "Innovative biodiesel production for sustainable energy: Advances in feedstocks, transesterification, and cost efficiency," *Biomass and Bioenergy*, vol. 201, Oct. 2025, Art. no. 108114. doi: 10.1016/j.biombioe.2025.108114.

[14] C. Billings, R. Siddique, B. Sherwood, J. Hall, and Y. Liu, "Additive manufacturing and characterization of sustainable wood fiber-reinforced green composites," *Journal of Composites Science*, vol. 7, no. 12, 2023, Art. no. 489. doi: 10.3390/jcs7120489.

[15] A. K. Shakya, G. Pillai, and S. Chakrabarty, "Reinforcement learning algorithms: A brief survey," *Expert Systems with Applications*, vol. 231, Nov. 2023, Art. no. 120495. doi: 10.1016/j.eswa.2023.120495.

[16] E. Alpaydin, *Introduction to Machine Learning* (4th ed.), MIT Press, 2020. ISBN: 9780262043793.

[17] S. S. Sorour, C. A. Saleh, and M. Shazly, "A review on machine learning implementation for predicting and optimizing the mechanical behaviour of laminated fiber-reinforced polymer composites," *Heliyon*, vol. 10, no. 13, Jul. 2024, Art. no. e33681. doi: 10.1016/j.heliyon.2024.e33681.

[18] P. Szatkowski, J. Barwinek, A. I. Zhakypbekovich, J. Szczecina, M. Niemiec, K. Pielichowska, and E. Molik, "The use of sheep wool collected from sheep bred in the Kyrgyz Republic as a component of biodegradable composite material," *Applied Sciences*, vol. 15, no. 24, 2025, Art. no. 13054. doi: 10.3390/app152413054.

[19] I. O. Oladele, V. O. Oki, T. F. Omotosho, M. B. Adebanjo, O. T. Ayanleye, and S. A. Adekola, "Sustainable polymer and polymer-based composite materials for extreme conditions and demanding applications – A review on pushing boundaries in materials science," *Next Materials*, vol. 8, Jul. 2025, Art. no. 100775. doi: 10.1016/j.nxmate.2025.100775.

[20] S. H. Kamarudin, M. S. M. Basri, M. Rayung, F. Abu, S. Ahmad, M. N. Norizan, S. Osman, N. Sarifuddin, M. S. Z. M. Desa, and U. H. Abdullah, "A review on natural fiber reinforced polymer composites (NFRPC) for sustainable industrial applications," *Polymers*, vol. 14, no. 17, 2022, Art. no. 3698. doi: 10.3390/polym14173698.

[21] J. J. Andrew and H. N. Dhakal, "Sustainable biobased composites for advanced applications: Recent trends and future opportunities – A critical review," *Composites Part C: Open Access*, vol. 7, Mar. 2022, Art. no. 100220. doi: 10.1016/j.jcomc.2021.100220.

[22] D. Y. Thimmegowda, J. Hindi, G. B. Markunti, and M. Kakunje, "Enhancement of mechanical properties of natural fiber reinforced polymer composites using different approaches—A review," *Journal of Composites Science*, vol. 9, no. 5, 2025, Art. no. 220. doi: 10.3390/jcs9050220.

[23] M. A. Alam, S. M. Sapuan, H. H. Ya, P. B. Hussain, M. Azeem, and R. A. Ilyas, "Application of biocomposites in automotive components: A review," in *Biocomposite and Synthetic Composites for Automotive Applications*, Woodhead Publishing Series in Composites Science and Engineering, 2021, pp. 1–17. doi: 10.1016/B978-0-12-820559-4.00001-8.

[24] S. Al-Alimi, N. K. Yusuf, A. M. Ghaleb, A. Adam, M. A. Lajis, S. Shamsudin, W. Zhou, Y. M. Altharan, Y. Saif, D. H. Didane, I. S. T. T., M. Al-fakih, S. A. Alzaeemi, A. Bouras, A. M. A. Elfaghi, and H. G. Mohammed, "Response surface methodology and machine learning optimisations comparisons of recycled AA6061-B4C–ZrO2 hybrid metal matrix composites via hot forging forming process," *Heliyon*, vol. 10, no. 12, Jun. 2024, Art. no. e33138. doi: 10.1016/j.heliyon.2024.e33138.

[25] M. Y. Khalid, A. A. Rashid, Z. U. Arif, W. Ahmed, H. Arshad, and A. A. Zaidi, "Natural fiber reinforced composites: Sustainable materials for emerging applications," *Results in Engineering*, vol. 11, Sep. 2021, Art. no. 100263. doi: 10.1016/j.rineng.2021.100263.

[26] M. E. Schatz, "Enabling composite optimization through soft computing of manufacturing restrictions and costs via a narrow artificial intelligence," *Journal of Composites Science*, vol. 2, no. 4, 2018, Art. no. 70. doi: 10.3390/jcs2040070.

[27] A. U. R. Bajwa, C. Siriwardana, W. Shahzad, and M. A. Naeem, "Material selection in the construction industry: A systematic literature review on multi-criteria decision making," *Environment Systems and Decisions*, vol. 45, 2025, Art. no. 8. doi: 10.1007/s10669-025-10001-w.

[28] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (2nd ed.), MIT Press, 2020. ISBN: 9780262039246.

[29] A. Kathirgamanathan, E. Mangina, and D. P. Finn, "Development of a Soft Actor Critic deep reinforcement learning approach for harnessing energy flexibility in a large office building," *Energy and AI*, vol. 5, Sep. 2021, Art. no. 100101. doi: 10.1016/j.egyai.2021.100101.

[30] X. Gao, F. Chao, C. Zhou, Z. Ge, L. Yang, X. Chang, C. Shang, and Q. Shen, "Error controlled actor-critic," *Information Sciences*, vol. 612, Oct. 2022, pp. 62–74. doi: 10.1016/j.ins.2022.08.079.

[31] C. Li, P. Zheng, Y. Yin, B. Wang, and L. Wang, "Deep reinforcement learning in smart manufacturing: A review and prospects," *CIRP Journal of Manufacturing Science and Technology*, vol. 40, Feb. 2023, pp. 75–101. doi: 10.1016/j.cirpj.2022.11.003.

[32] M. Bodaghi, L. Bouhala, C. G. Bayreuther, A. El Moumen, D. Macieira, and M. Kerschbaum, "On the understanding of GRAM® technology – robotic wet filament winding – for high-performance fibre-reinforced thermoset composites," *Composites Part A: Applied Science and Manufacturing*, vol. 197, Oct. 2025, Art. no. 109028. doi: 10.1016/j.compositesa.2025.109028.

[33] Y. Chen, W. Zou, G. Wang, C. Zhang, and C. Zhang, "Optimizing the impact toughness of PLA materials using machine learning algorithms," *Materials Today Communications*, vol. 44, Mar. 2025, Art. no. 111881. doi: 10.1016/j.mtcomm.2025.111881.

[34] V. Varagapriya, V. V. Singh, and A. Lisser, "Constrained Markov decision processes with uncertain costs," *Operations Research Letters*, vol. 50, no. 2, Mar. 2022, pp. 218–223. doi: 10.1016/j.orl.2022.02.001.

[35] K. Szatmári, T. Chován, S. Németh, and A. Kummer, "How to support decision making with reinforcement learning in hierarchical chemical process control?" *Chemical Engineering Journal Advances*, vol. 22, May 2025, Art. no. 100753. doi: 10.1016/j.ceja.2025.100753.

[36] J. Deng, S. Sierla, J. Sun, and V. Vyatkin, "Reinforcement learning for industrial process control: A case study in flatness control in steel industry," *Computers in Industry*, vol. 143, Dec. 2022, Art. no. 103748. doi: 10.1016/j.compind.2022.103748.

[37] R. de R. Faria, B. D. O. Capron, A. R. Secchi, and M. B. de Souza Jr., "Where reinforcement learning meets process control: Review and guidelines," *Processes*, vol. 10, no. 11, 2022, Art. no. 2311. doi: 10.3390/pr10112311.

[38] O. Bongomin, M. C. Mwape, N. S. Mpofu, B. K. Bahunde, R. Kidega, I. L. Mpungu, G. Tumusiime, C. A. Owino, Y. M. Goussongtogue, A. Yemane, P. Kyokunzire, C. Malanda, J. Komakech, D. Tigalana, O. Gumisiriza, and G. Ngulube, "Digital twin technology advancing industry 4.0 and industry 5.0 across sectors," *Results in Engineering*, vol. 26, Jun. 2025, Art. no. 105583. doi: 10.1016/j.rineng.2025.105583.

[39] I. Elfaleh, F. Abbassi, M. Habibi, F. Ahmad, M. Guedri, M. Nasri, and C. Garnier, "A comprehensive review of natural fibers and their composites: An eco-friendly alternative to conventional materials," *Results in Engineering*, vol. 19, Sep. 2023, Art. no. 101271. doi: 10.1016/j.rineng.2023.101271.

[40] M. Soori, B. Arezoo, and R. Dastres, "Digital twin for smart manufacturing: A review," *Sustainable Manufacturing and Service Economics*, vol. 2, Apr. 2023, Art. no. 100017. doi: 10.1016/j.smse.2023.100017.

[41] Q. Chen, Y. Jin, and Y. Song, "Fault-tolerant adaptive tracking control of Euler-Lagrange systems – An echo state network approach driven by reinforcement learning," *Neurocomputing*, vol. 484, May 2022, pp. 109–116. doi: 10.1016/j.neucom.2021.10.083.

[42] K. Mondal and P. K. Tripathy, "Preparation of smart materials by additive manufacturing technologies: A review," *Materials*, vol. 14, no. 21, 2021, Art. no. 6442. doi: 10.3390/ma14216442.

[43] Imran, F. Qayyum, D.-H. Kim, S.-J. Bong, S.-Y. Chi, and Y.-H. Choi, "A survey of datasets, preprocessing, modeling mechanisms, and simulation tools based on AI for material analysis and discovery," *Materials*, vol. 15, no. 4, 2022, Art. no. 1428. doi: 10.3390/ma15041428.

[44] V. Samsonov, K. B. Hicham, and T. Meisen, "Reinforcement learning in manufacturing control: Baselines, challenges and ways forward," *Engineering Applications of Artificial Intelligence*, vol. 112, Jun. 2022, Art. no. 104868. doi: 10.1016/j.engappai.2022.104868.

[45] O. Ulkir and S. Ersoy, "Hybrid experimental–machine learning study on the mechanical behavior of polymer composite structures fabricated via FDM," *Polymers*, vol. 17, no. 15, 2025, Art. no. 2012. doi: 10.3390/polym17152012.

[46] S. Li, C. Tian, and A. N. Abdalla, "Energy regulation-aware layered control architecture for building energy systems using constraint-aware deep reinforcement learning and virtual energy storage modeling," *Energies*, vol. 18, no. 17, 2025, Art. no. 4698. doi: 10.3390/en18174698.

[47] S. E. Bibri and J. Huang, "Generative AI of things for sustainable smart cities: Synergizing cognitive augmentation, resource efficiency, network traffic, cybersecurity, and anomaly detection for environmental performance," *Sustainable Cities and Society*, vol. 133, Oct. 2025, Art. no. 106826. doi: 10.1016/j.scs.2025.106826.

[48] M. M. Sahib and G. Kovács, "Multi-objective optimization of composite sandwich structures using artificial neural networks and genetic algorithm," *Results in Engineering*, vol. 21, Mar. 2024, Art. no. 101937. doi: 10.1016/j.rineng.2024.101937.

[49] A. Jamwal, R. Agrawal, M. Sharma, A. Kumar, V. Kumar, and J. A. Garza-Reyes, "Machine learning applications for sustainable manufacturing: A bibliometric-based review for future research," *Journal of Enterprise Information Management*, vol. 35, no. 2, May 2021, pp. 566–596. doi: 10.1108/JEIM-09-2020-0361.

[50] S. Krishnamurthy, O. B. Adewuyi, and S. A. Salimon, "Recent advances in artificial intelligence-based optimization for power system applications: A review of techniques, challenges, and future directions," *Renewable and Sustainable Energy Reviews*, vol. 226, pt. C, Jan. 2026, Art. no. 116340. doi: 10.1016/j.rser.2025.116340.

[51] S. Vespoli, G. Mattera, M. G. Marchesano, L. Nele, and G. Guizzi, "Adaptive manufacturing control with deep reinforcement learning for dynamic WIP management in Industry 4.0," *Computers & Industrial Engineering*, vol. 202, Apr. 2025, Art. no. 110966. doi: 10.1016/j.cie.2025.110966.

[52] A. F. C. Vieira, M. R. Tavares Filho, J. P. Eguea, and M. L. Ribeiro, "Optimization of structures and composite materials: A brief review," *Eng*, vol. 5, no. 4, 2024, pp. 3192–3211. doi: 10.3390/eng5040168.

[53] T. Xie, A. H. Sung, and X. Qin, "Dynamic task scheduling with security awareness in real-time systems," in *Proceedings of the 19th International Parallel and Distributed Processing Symposium (IPDPS 2005)*, Denver, CO, USA, Apr. 4–8, 2005. doi: 10.1109/IPDPS.2005.185.

[54] C. Zhang, M. Juraschek, and C. Herrmann, "Deep reinforcement learning-based dynamic scheduling for resilient and sustainable manufacturing: A systematic review," *Journal of Manufacturing Systems*, vol. 77, Dec. 2024, pp. 962–989. doi: 10.1016/j.jmsy.2024.10.026.

[55] M. Ashkbous, E. Ghorbani, and S. Keivanpour, "Artificial intelligence for eco-design: A systematic review," *Advanced Engineering Informatics*, vol. 69, pt. C, Jan. 2026, Art. no. 103989. doi: 10.1016/j.aei.2025.103989.

[56] A. Ghasemi, F. Farajzadeh, C. Heavey, J. Fowler, and C. T. Papadopoulos, "Simulation optimization applied to production scheduling in the era of industry 4.0: A review and future roadmap," *Journal of Industrial Information Integration*, vol. 39, May 2024, Art. no. 100599. doi: 10.1016/j.jii.2024.100599.

[57] Q. Qin, Z. Liu, R. Zhong, X. V. Wang, L. Wang, M. Wiktorsson, and W. Wang, "Robot digital twin systems in manufacturing: Technologies, applications, trends and challenges," *Robotics and Computer-Integrated Manufacturing*, vol. 97, Feb. 2026, Art. no. 103103. doi: 10.1016/j.rcim.2025.103103.

[58] J. Cestero, C. Delle Femine, K. S. Muro, M. Quartulli, and M. Restelli, "Optimizing energy management of smart grid using reinforcement learning aided by surrogate models built using physics-informed neural networks," *Applied Energy*, vol. 401, pt. C, Dec. 2025, Art. no. 126750. doi: 10.1016/j.apenergy.2025.126750.

[59] A. Khdoudi, T. Masrour, I. El Hassani, and C. El Mazgualdi, "A deep-reinforcement-learning-based digital twin for manufacturing process optimization," *Systems*, vol. 12, no. 2, 2024, Art. no. 38. doi: 10.3390/systems12020038. [60] Tambe, Priya, Chen, Xuefeng, Khajepour, Amir. (2022). Sample efficiency challenges in RL for real-time production. Journal of Manufacturing Systems, 64, 109–118. https://doi.org/10.1016/j.jmsy.2022.04.008

[60] C. Li, Q. Chang, and H.-T. Fan, "Multi-agent reinforcement learning for integrated manufacturing system-process control," *Journal of Manufacturing Systems*, vol. 76, Oct. 2024, pp. 585–598. doi: 10.1016/j.jmsy.2024.08.021.