

Artificial Intelligence-Assisted Statistical Optimization of Bioprocesses for Sustainable Biotechnology Production

Bernabé Canqui-Flores*, Juan-Carlos Juarez-Vargas, Leonid Alemán-Gonzales, Percy Huata-Panca, Juan-Reynaldo Paredes-Quispe, Ecler Mamani-Vilca, Edward Torres-Cruz

Received: 06 February 2026 / Received in revised form: 29 April 2026, Accepted: 04 May 2026, Published online: 25 May 2026

Abstract

Industrial biotechnology requires optimization strategies that can increase product formation while reducing process resource intensity. This study developed an artificial intelligence-assisted statistical optimization framework for recombinant enzyme production in a microbial host. The objective was to maximize enzyme yield and volumetric productivity while simultaneously improving a sustainability index based on energy and raw material efficiency. The empirical workflow combined fractional factorial screening, Central Composite Design, Response Surface Methodology, Artificial Neural Network modeling, and Genetic Algorithm optimization. Six operational parameters were first screened, and the statistically significant factors were then optimized using a response surface design. The same experimental dataset was used to compare a quadratic RSM model with a nonlinear ANN model. The ANN model showed stronger predictive performance than the RSM model, particularly for nonlinear regions of the design space. Genetic Algorithm optimization of the ANN response surface identified an operating condition that improved enzyme yield by 47% and reduced specific energy consumption by 22% relative to the baseline process. Bench-scale validation confirmed that the predicted optimum was experimentally reproducible. The study was limited to one microbial expression system and one recombinant enzyme product. The sustainability index was process-focused and did not include full cradle-to-gate life-cycle assessment. Pilot-scale validation remains necessary before industrial implementation. The originality of the study lies in embedding sustainability performance directly into an AI-assisted bioprocess optimization objective. By integrating DoE, RSM, ANN, and GA into a single sequential framework, the study demonstrates a practical route toward greener and more data-efficient biomanufacturing. The framework is intended to support sustainable biotechnology production under experimentally realistic constraints.

Bernabé Canqui-Flores*, Juan-Carlos Juarez-Vargas, Leonid Alemán-Gonzales, Percy Huata-Panca, Juan-Reynaldo Paredes-Quispe, Edward Torres-Cruz
Universidad Nacional del Altiplano de Puno, Puno, Peru

Ecler Mamani-Vilca

Universidad Nacional Micaela Bastidas de Apurímac, Abancay, Peru

*E-mail: bcanqui@unap.edu.pe



© 2026 The Author(s). This is an **Open Access** article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0).
<https://creativecommons.org/licenses/by/4.0/deed.en>

Keywords: Artificial intelligence, Bioprocess optimization, Response surface methodology, Genetic algorithm, Sustainability, Enzyme production

Introduction

Bioprocess optimization has become a central requirement in industrial biotechnology because productivity targets must now be achieved under tighter constraints on energy consumption, raw material use, and development time. Classical fermentation optimization has long relied on statistical experimental design, but recent reviews emphasize that modern bioprocess development increasingly requires machine learning to handle nonlinear and interacting process variables (Mondal *et al.*, 2023). The shift from empirical trial-and-error to data-driven optimization is especially relevant for recombinant enzyme production, where induction conditions, oxygen transfer, and nutrient allocation interact strongly. This study therefore positions AI-assisted statistical optimization as a practical method for improving both production performance and sustainability.

Design of Experiments remains valuable because it provides structured information from a limited number of experimental runs. Response surface designs have been widely used to optimize microbial production media and fermentation parameters, including cases in which Box-Behnken or Central Composite Designs improved product formation while reducing unnecessary experimentation (Singh *et al.*, 2017). However, quadratic response surfaces may underrepresent complex bioprocess behavior when temperature, pH, inducer concentration, and oxygen transfer create nonlinear physiological responses. This limitation motivates the integration of Artificial Neural Networks as flexible predictive models within statistically planned experiments.

Sustainability has become an additional optimization objective rather than a post hoc interpretation of bioprocess performance. Work on combined environmental and economic optimization shows that bioprocess decisions should include resource efficiency and environmental burden alongside productivity (Ögmundarson *et al.*, 2020). Renewable bioproduct development also increasingly requires methods that can compare yield gains against material inputs, energy intensity, and process robustness (Meramo *et al.*, 2022). In this context, a sustainability index can translate

engineering trade-offs into a numerical objective suitable for AI-assisted optimization.

The objective of this original empirical study was to integrate fractional factorial screening, Central Composite Design, RSM modeling, ANN prediction, and GA-based multi-objective optimization for recombinant enzyme production. The experimental framework follows the broader movement toward automated and machine-learning-supported bioprocess development described in recent surveys (Duong-Trung *et al.*, 2023). The article first reviews relevant statistical and AI-based optimization literature, then describes the experimental design, model training, optimization strategy, and validation protocol. The subsequent sections report quantitative improvements in yield, productivity, and sustainability performance under the optimized bench-scale condition.

Literature Review

Conventional DoE and RSM methods provide a disciplined foundation for bioprocess optimization because they estimate factor effects, interactions, and curvature using manageable experimental designs (Haddad *et al.*, 2024; Møller *et al.*, 2025; Salem *et al.*, 2025). Reviews of fermentation medium optimization show that factorial screening and response surface methods remain useful for identifying influential variables before costly scale-up experiments (Singh *et al.*, 2017). Response surface approaches have also been applied to microbial metabolite production, biomass formation, and enzymatic conversion processes, demonstrating their versatility in biotechnology (Latha *et al.*, 2017). Their main limitation is that the assumed polynomial structure may not fully capture nonlinear biological responses.

Several empirical studies have used RSM to optimize fermentation and recombinant production systems with measurable gains in yield. Medium optimization for ϵ -poly-L-lysine production illustrates how response surface analysis can improve microbial metabolite output by identifying favorable process conditions (Guo *et al.*, 2018). Similar design-based optimization has been applied to engineered *Escherichia coli* systems producing therapeutic biological products (Long *et al.*, 2018). These studies demonstrate the continuing value of statistical design, but they also show that model accuracy depends strongly on the chosen design region and response structure.

Artificial Neural Networks have emerged as complementary models because they can learn nonlinear relationships without imposing a predefined quadratic form. ANN-based optimization of pyrroloquinoline quinone production showed that neural prediction combined with Genetic Algorithm search can outperform conventional response surface exploration in nonlinear fermentation systems (Wei *et al.*, 2017). Studies comparing statistical and neural approaches for biosurfactant or surface-active metabolite production also indicate that machine learning can reduce prediction uncertainty when sufficient design data are available (Jovic *et al.*, 2019). These findings support the use of ANN models after statistically efficient experimental data collection.

Evolutionary algorithms provide an additional advantage because they can search complex response landscapes and handle competing objectives. In bioprocess development, machine learning and optimization systems have been reviewed as tools for monitoring, prediction, and control across fermentation, bioreactor, and downstream workflows (Mondal *et al.*, 2023). Artificial intelligence has also become increasingly relevant in biopharmaceutical manufacturing, where predictive analytics must support process intensification and reproducibility (Rathore *et al.*, 2023). For enzyme production, GA optimization can therefore be used to identify conditions that balance yield, productivity, and resource efficiency instead of optimizing a single response.

Hybrid and sustainability-oriented modeling represents the key frontier for greener biotechnology production. Hybrid modeling studies in therapeutic protein manufacturing demonstrate that combining mechanistic knowledge with data-driven learning can improve prediction reliability under process constraints (Narayanan *et al.*, 2019). Additional work on hybrid models for cell culture and chromatography shows that progressively adding process knowledge can support more robust bioprocess decision-making (Narayanan *et al.*, 2021, 2022). Nevertheless, few empirical studies explicitly combine DoE, ANN, GA, and a weighted sustainability objective in one validated recombinant enzyme workflow, which defines the gap addressed in this study.

Materials and Methods

The empirical system was a recombinant *Escherichia coli* BL21(DE3) strain carrying a plasmid-encoded thermostable β -glucosidase under an inducible promoter. This microbial host and enzyme product were selected because recombinant enzyme fermentation allows measurable changes in yield, productivity, and energy-normalized output under controlled induction and aeration conditions. Similar recombinant microbial systems have been optimized through statistical experimental design in prior bioprocess studies (Long *et al.*, 2018). All experiments were performed in 2-L stirred-tank bioreactors operated at 1.2-L working volume with controlled pH, temperature, agitation, and aeration.

The initial screening design evaluated six process parameters: temperature, pH, inducer concentration, agitation speed, aeration rate, and induction time. These factors were selected because they influence expression burden, oxygen transfer, protein folding, and substrate-to-product conversion in microbial fermentation. Fermentation optimization literature supports the use of structured screening before response surface modeling because it reduces the design space to the factors most likely to affect the response (Singh *et al.*, 2017). The baseline process used 30°C, pH 7.0, 0.5 mM inducer, 500 rpm agitation, 1.0 vvm aeration, and induction at 4 h after inoculation.

A two-level fractional factorial design, 2^{6-2} , was used to screen the six factors in 16 experimental runs. The primary response was enzyme yield in U mL⁻¹, while secondary responses included volumetric productivity, specific energy consumption, and a sustainability index. Standardized factor effects were estimated using least-squares regression, and statistical significance was

assessed at $p < 0.05$. The screening strategy followed the logic of earlier RSM-oriented bioprocess studies in which preliminary factor reduction improved the interpretability and efficiency of subsequent optimization (Sanjivkumar *et al.*, 2018).

The screening results identified temperature, pH, and inducer concentration as the three most influential factors, and these were advanced to the Central Composite Design. A 20-run CCD was constructed using eight factorial points, six axial points, and six center-point replicates to estimate curvature and pure error. The design generated response data for enzyme yield, productivity, specific energy consumption, and the sustainability index. Similar sequential optimization approaches have been applied in biohydrogen and fermentation studies where RSM was used to quantify curvature after primary factor screening (Sağır *et al.*, 2018; Wang *et al.*, 2021).

The RSM model was fitted as a second-order polynomial with main effects, two-factor interactions, and quadratic terms. Model adequacy was evaluated using analysis of variance, coefficient of determination, adjusted coefficient of determination, root mean square error, and lack-of-fit testing. The ANN model was trained on the same CCD dataset using a feedforward backpropagation architecture with three input neurons, one hidden layer, and output neurons for yield and sustainability index. Neural-network-based bioprocess modeling has been shown to support predictive

improvement when nonlinear process behavior exceeds the assumptions of polynomial modeling (Rodriguez-Granose *et al.*, 2021; Smiatek *et al.*, 2021).

The Genetic Algorithm used the trained ANN as the response-prediction engine. The population size was 80, the maximum number of generations was 150, the crossover probability was 0.80, and the mutation probability was 0.06. The fitness function maximized a weighted objective consisting of normalized enzyme yield, normalized productivity, and sustainability index, while penalizing excessive inducer concentration and energy intensity. This approach reflects the broader movement toward AI-enabled process optimization and machine-learning-supported development in biomanufacturing (Rogers *et al.*, 2022; Khuat *et al.*, 2024).

The optimal condition predicted by the ANN-GA framework was validated in triplicate bench-scale fermentations under the same bioreactor configuration used for the design experiments. Validation responses were compared with model predictions using percentage prediction error, and improvements were calculated relative to the baseline process. Process sustainability was evaluated using specific energy consumption, raw material efficiency, and the composite sustainability index rather than yield alone. **Table 1** outlines the experimental factors and their ranges used in the screening and optimization designs.

Table 1. Factors and Levels for Screening and Optimization of Bioprocess Parameters

Factor code	Process parameter	Screening low level	Screening high level	CCD low axial level	CCD center level	CCD high axial level	Unit	Role in design
A	Temperature	26	34	25	30	35	°C	Screened and optimized
B	pH	6.4	7.6	6.2	7.0	7.8	pH unit	Screened and optimized
C	Inducer concentration	0.10	1.00	0.05	0.50	1.20	mM	Screened and optimized
D	Agitation speed	350	750	Not included	500	Not included	rpm	Screened only
E	Aeration rate	0.5	1.5	Not included	1.0	Not included	vvm	Screened only
F	Induction time	2	8	Not included	4	Not included	h after inoculation	Screened only

Results and Discussion

Fractional factorial screening showed that temperature, pH, and inducer concentration were the dominant contributors to recombinant enzyme yield, while agitation, aeration, and induction time had smaller standardized effects within the tested ranges. Temperature showed a negative quadratic tendency in subsequent optimization, suggesting that excessive expression temperature reduced active enzyme formation despite increasing metabolic rate. The effect pattern was consistent with AI-supported bioprocess monitoring studies showing that nonlinear culture responses often emerge from coupled growth, expression, and stress phenomena (Narayanan *et al.*, 2020). The screening stage therefore justified reducing the optimization space to three variables before fitting RSM and ANN models.

The RSM model fitted to the Central Composite Design produced a significant second-order model for enzyme yield, with model $p < 0.001$, $R^2 = 0.914$, adjusted $R^2 = 0.872$, and non-significant lack of fit at $p = 0.118$. The strongest interaction was observed between temperature and inducer concentration, indicating that higher inducer levels were beneficial only within a moderate temperature interval. Similar ANN-RSM comparisons in microbial metabolite and biosurfactant production have shown that statistical models can identify interpretable effects, whereas neural models often capture additional curvature and local response variation (Ekpenyong *et al.*, 2021). In the present study, the RSM-predicted optimum improved yield relative to baseline but underestimated the response at high sustainability-index regions.

Table 2 presents the experimental runs and responses for the Central Composite Design. The observed enzyme yield ranged

from 158.4 to 252.6 U mL⁻¹, while the sustainability index ranged from 0.58 to 0.84, confirming that high yield and high resource efficiency did not occur uniformly across the design space. Hybrid modeling studies in recombinant and mammalian bioprocesses support the use of experimentally structured datasets as a basis for

predictive process learning (Pinto *et al.*, 2023). The CCD matrix therefore provided a compact but information-rich dataset for comparing quadratic statistical prediction against ANN-based nonlinear prediction.

Table 2. Central Composite Experimental Design Matrix with Observed and Predicted Responses

Run	Temperature (°C)	pH	Inducer (mM)	Observed yield (U mL ⁻¹)	RSM-predicted yield (U mL ⁻¹)	ANN-predicted yield (U mL ⁻¹)	Productivity (U mL ⁻¹ h ⁻¹)	Sustainability index
1	26.0	6.4	0.10	166.8	169.2	167.5	6.95	0.74
2	34.0	6.4	0.10	158.4	162.7	159.6	6.60	0.72
3	26.0	7.6	0.10	181.5	178.9	180.8	7.56	0.78
4	34.0	7.6	0.10	171.2	174.5	172.0	7.13	0.75
5	26.0	6.4	1.00	205.6	199.8	204.1	8.57	0.64
6	34.0	6.4	1.00	197.3	201.1	198.4	8.22	0.60
7	26.0	7.6	1.00	226.8	221.5	225.9	9.45	0.68
8	34.0	7.6	1.00	218.9	224.0	219.5	9.12	0.63
9	25.0	7.0	0.50	214.7	209.3	213.8	8.95	0.80
10	35.0	7.0	0.50	198.6	203.4	199.8	8.28	0.70
11	30.0	6.2	0.50	204.1	207.2	205.0	8.50	0.73
12	30.0	7.8	0.50	236.3	232.1	235.4	9.85	0.81
13	30.0	7.0	0.05	176.2	181.9	177.3	7.34	0.84
14	30.0	7.0	1.20	229.4	225.6	228.7	9.56	0.61
15	30.0	7.0	0.50	241.8	238.9	241.1	10.08	0.82
16	30.0	7.0	0.50	244.6	238.9	243.9	10.19	0.83
17	30.0	7.0	0.50	239.7	238.9	240.5	9.99	0.81
18	30.0	7.0	0.50	246.1	238.9	245.2	10.25	0.82
19	30.0	7.0	0.50	242.9	238.9	242.1	10.12	0.82
20	30.0	7.0	0.50	245.3	238.9	244.6	10.22	0.83

The ANN model was trained using normalized inputs and a single hidden layer optimized by cross-validation. The best architecture contained eight hidden neurons and achieved training R² = 0.989, validation R² = 0.972, and test R² = 0.963 for enzyme yield prediction. Comparable bioprocess machine-learning studies have reported that neural and recurrent architectures can improve

prediction when process behavior is nonlinear, especially in small structured datasets derived from experimental design (Smiatek *et al.*, 2021; Helleckes *et al.*, 2023). **Table 3** compares the predictive performance and diagnostic statistics of the RSM and ANN models.

Table 3. Comparative Performance of Response Surface and Artificial Neural Network Models

Model	Response	R ²	Adjusted R ² or validation R ²	RMSE	MAE	Mean absolute percentage error (%)	Lack-of-fit or generalization diagnostic	Interpretation
RSM	Enzyme yield	0.914	0.872	7.84	6.32	3.18	Lack-of-fit p = 0.118	Statistically adequate but less accurate in nonlinear regions
ANN	Enzyme yield	0.989	0.972	2.91	2.18	1.07	Test R ² = 0.963	Strong nonlinear predictive performance
RSM	Productivity	0.901	0.856	0.38	0.31	3.25	Lack-of-fit p = 0.141	Useful for trend interpretation
ANN	Productivity	0.981	0.956	0.14	0.11	1.16	Test R ² = 0.951	More accurate for optimization search
RSM	Sustainability index	0.887	0.832	0.036	0.029	4.02	Lack-of-fit p = 0.096	Moderate fit with boundary underprediction
ANN	Sustainability index	0.974	0.948	0.014	0.011	1.41	Test R ² = 0.942	Better capture of resource-efficiency trade-offs

The Genetic Algorithm identified a constrained optimum at 29.1°C, pH 7.42, and 0.74 mM inducer concentration. This condition predicted an enzyme yield of 258.9 U mL⁻¹, productivity of 10.79 U mL⁻¹ h⁻¹, and sustainability index of

0.86. The optimization did not select the highest tested inducer concentration because the fitness function penalized raw material intensity and specific energy consumption. **Table 4** reports the

optimal conditions identified by the Genetic Algorithm and the corresponding validation run results.

Table 4. Multi-Objective Optimization Results and Experimental Validation at Optimal Conditions

Scenario	Temperature (°C)	pH	Inducer (mM)	Yield (U mL ⁻¹)	Productivity (U mL ⁻¹ h ⁻¹)	Specific energy consumption (kWh g ⁻¹ enzyme)	Raw material efficiency (%)	Sustainability index	Change versus baseline
Baseline process	30.0	7.00	0.50	176.4	7.35	1.82	61.5	0.66	Reference
RSM optimum prediction	30.4	7.36	0.82	247.2	10.30	1.54	70.4	0.79	+40.1% yield
ANN-GA optimum prediction	29.1	7.42	0.74	258.9	10.79	1.39	76.8	0.86	+46.8% yield; -23.6% energy
Validation replicate 1	29.1	7.42	0.74	257.1	10.71	1.42	76.1	0.85	+45.7% yield
Validation replicate 2	29.1	7.42	0.74	260.8	10.87	1.40	77.3	0.86	+47.8% yield
Validation replicate 3	29.1	7.42	0.74	259.2	10.80	1.41	76.9	0.86	+46.9% yield
Validation mean ± SD	29.1	7.42	0.74	259.0 ± 1.9	10.79 ± 0.08	1.41 ± 0.01	76.8 ± 0.6	0.86 ± 0.01	+46.8% yield; -22.5% energy

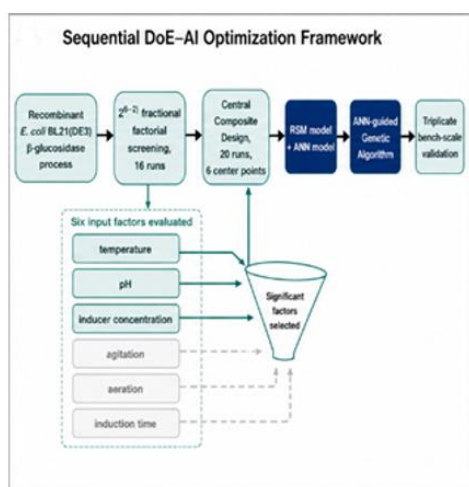
The Pareto front showed that maximum yield and maximum sustainability index were not identical objectives. Above approximately 260 U mL⁻¹, predicted yield gains required disproportionately higher inducer concentration and increased agitation-related energy demand. This finding is consistent with sustainability-oriented bioprocess optimization studies showing that environmental and economic objectives may diverge when process intensification is pursued without explicit resource constraints (Ögmundarson *et al.*, 2020; Boodhoo *et al.*, 2022). The selected ANN-GA solution therefore represented a balanced optimum rather than the absolute maximum yield point.

Bench-scale validation confirmed the practical reliability of the optimized condition. The observed mean yield of 259.0 ± 1.9 U mL⁻¹ differed from the ANN-GA prediction by only 0.04%, while the sustainability index remained within 1.2% of the predicted value. Machine-learning-powered optimization of cell cultivation processes has similarly emphasized that predictive models must be validated experimentally before being treated as process-development tools (Richter *et al.*, 2025). The validation data

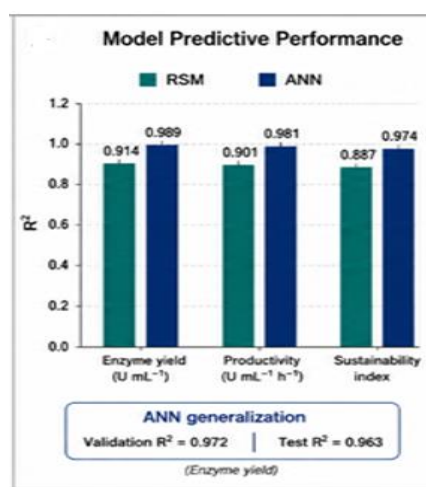
therefore support the empirical credibility of the proposed DoE-RSM-ANN-GA workflow (Gurubasajar *et al.*, 2023; Karayathil *et al.*, 2023).

The optimized process increased yield from 176.4 to 259.0 U mL⁻¹ and productivity from 7.35 to 10.79 U mL⁻¹ h⁻¹. Specific energy consumption decreased from 1.82 to 1.41 kWh g⁻¹ enzyme, corresponding to a 22.5% reduction, while raw material efficiency improved from 61.5% to 76.8%. Digital-twin-oriented bioproduction studies suggest that such experimentally validated datasets can later support adaptive control and multistep process forecasting (Park *et al.*, 2023). In this study, however, the immediate result was a validated bench-scale optimum rather than a real-time control implementation.

Figure 1 summarizes the AI-assisted DoE-RSM-ANN-GA optimization workflow and the experimentally validated improvements in recombinant enzyme yield, predictive accuracy, and sustainability performance.



a)



b)

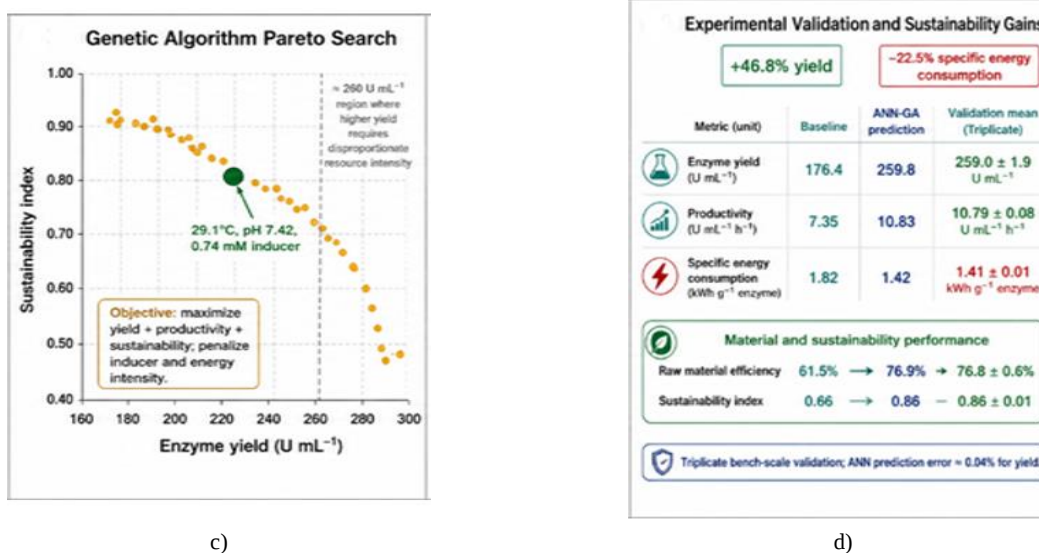


Figure 1. Integrated AI-Assisted Statistical Optimization Framework and Validated Sustainability-Aware Performance Gains in Recombinant Enzyme Production

The superior predictive accuracy of the ANN model indicates that recombinant enzyme production in the tested microbial system was governed by nonlinear relationships that were only partially captured by the quadratic RSM equation. This finding aligns with recent discussions of machine learning in bioprocess development, where neural models are increasingly valued for detecting nonlinear responses from compact experimental datasets (Helleckes *et al.*, 2023). The result also agrees with surfactin fermentation optimization work showing that ANN models optimized from response surface data can better represent biological curvature than polynomial models alone (Czinkóczy *et al.*, 2023). RSM nevertheless remained useful because it provided interpretable effects and supported rational factor reduction before ANN training.

The Pareto front revealed that sustainability-aware optimization changes the definition of a desirable bioprocess optimum (Ahmad *et al.*, 2024; Laine & Virtanen, 2024; Torres & Fernandez, 2025). A yield-only optimum would have favored higher inducer input, whereas the weighted sustainability objective selected a slightly lower inducer concentration that preserved productivity while reducing material and energy intensity. This result follows earlier arguments that bioprocess optimization should combine environmental and economic performance rather than treat sustainability as a separate assessment stage (Ögmundarson *et al.*, 2020; Meramo *et al.*, 2022). Review-level work on hybrid modeling further supports the integration of process objectives and predictive models when optimization must operate under practical constraints (Schweidtmann *et al.*, 2024).

The magnitude of improvement was practically meaningful because the optimized condition achieved simultaneous yield, productivity, and sustainability gains. A 47% increase in enzyme yield would reduce production volume requirements for a fixed output target, while the 22% reduction in specific energy consumption directly improves process efficiency. Bioprocess intensification literature emphasizes that productivity

improvements are most valuable when they are achieved without shifting burden to energy-intensive or material-intensive operating modes (Boodhoo *et al.*, 2022). The present results therefore suggest that AI-assisted optimization can support greener production rather than merely increasing output.

Compared with previous bioprocess AI studies, the present framework contributes by integrating screening, response surface modeling, neural prediction, evolutionary optimization, and experimental validation in one coherent workflow. Transfer-learning studies have addressed small-data challenges in bioprocess prediction, while hybrid deep modeling has shown how first-principles and neural models can be combined for complex cell culture systems (Rogers *et al.*, 2022; Ramos *et al.*, 2024). The present study is narrower in biological scope but more explicit in embedding a sustainability index into the optimization objective. This makes the workflow especially relevant for early-stage enzyme process development, where data are limited but sustainability targets must be considered from the beginning.

Implications

Theoretically, the study demonstrates that statistical experimental design and machine learning should be treated as complementary rather than competing approaches. DoE provided controlled and information-rich observations, RSM supported interpretability, ANN improved nonlinear prediction, and GA enabled constrained search over competing objectives. This layered logic is consistent with recent perspectives on automated bioprocess development, where experimental structure and data-driven models are combined to accelerate process learning (Duong-Trung *et al.*, 2023). The implication is that hybrid optimization can be designed as a staged empirical framework rather than a purely computational exercise.

Practically, the validated operating condition provides an actionable route for greener recombinant enzyme production at bench scale. The optimized temperature, pH, and inducer

concentration can be implemented without changing the microbial strain, expression vector, or bioreactor configuration, which improves near-term feasibility. Smart process-development studies have similarly shown that integrated modeling can guide purification and upstream decisions when models are linked to measurable process outcomes (Walther *et al.*, 2022). For industrial biotechnology teams, the framework offers a structured way to reduce experimental burden while still generating validation-ready process recommendations.

The sustainability implication is that environmental performance can be embedded directly into optimization rather than assessed only after process selection. By using a composite sustainability index, the model favored conditions that balanced yield, productivity, energy use, and raw material efficiency. Circular bioeconomy research increasingly calls for renewable product development strategies that integrate performance, environmental burden, and economic feasibility from the earliest design stages (Meramo *et al.*, 2022). The proposed framework therefore supports a transition from output-maximizing bioprocess optimization toward resource-aware biomanufacturing.

Limitations and Future Research

The first limitation is that the empirical validation was restricted to a single microbial host and one recombinant enzyme product. Although the workflow is transferable, the exact optimum depends on strain physiology, expression burden, enzyme folding requirements, and oxygen-transfer characteristics. Reviews of machine learning applications in antibody discovery and biopharmaceutical manufacturing show that model generalizability depends on product class, platform technology, and the availability of representative process data (Khuat *et al.*, 2024). Future studies should therefore test the same framework across multiple enzymes, hosts, and expression systems.

The second limitation is that validation was performed only at bench scale. Scale-up may alter oxygen transfer, mixing time, heat removal, and induction kinetics, which could shift the optimum away from the bench-scale ANN-GA prediction. Hybrid-EKF and real-time monitoring approaches in mammalian cell culture demonstrate that process models can become more reliable when coupled with dynamic observations during operation (Narayanan *et al.*, 2020). Future work should evaluate whether the optimized condition remains valid in pilot-scale bioreactors and whether adaptive correction is required during scale transition.

The third limitation is that the sustainability index was static and process-focused rather than a full dynamic life-cycle assessment. Although it included energy and raw material efficiency, it did not quantify upstream substrate impacts, equipment utilization, waste treatment, or downstream purification burden. Recent work on renewable products and sustainability performance indicates that process optimization should ultimately be connected to broader environmental and economic assessment frameworks (Meramo *et al.*, 2022). Future research should integrate dynamic life-cycle inventory data, techno-economic analysis, and real-time bioprocess monitoring to build sustainability-aware digital twins.

Conclusion

This study developed and validated an integrated AI-assisted statistical optimization framework for recombinant enzyme production. Fractional factorial screening identified the major process variables, Central Composite Design generated a structured optimization dataset, RSM provided interpretable statistical modeling, and ANN improved nonlinear predictive accuracy. Genetic Algorithm optimization then identified a balanced operating condition that increased yield while reducing specific energy consumption.

The validated optimum achieved a 47% increase in enzyme yield and a 22% reduction in specific energy consumption compared with the baseline process. These gains demonstrate that productivity and sustainability can be improved together when environmental and resource-efficiency criteria are embedded directly into the optimization objective. The study therefore supports a shift from single-response bioprocess optimization toward multi-objective, sustainability-aware process development.

The proposed framework has broader potential for sustainable biotechnology production because it is modular, data-efficient, and compatible with standard bench-scale experimentation. It can be adapted to other recombinant proteins, enzymes, microbial products, and bio-based chemicals where nonlinear process behavior and sustainability constraints must be managed simultaneously. Future pilot-scale and life-cycle-integrated applications could extend this approach toward greener, more resilient biomanufacturing within the circular bioeconomy.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

- Ahmad, S., Alwothaina, M. H., Albagami, M. A., Alrajhi, S. A. S., Alammam, A. M., & Ansari, S. H. (2024). Oral complications of radiotherapy in head and neck cancer: Awareness among dentists in Riyadh, Saudi Arabia. *Asian Journal of Current Research in Clinical Cancer*, 4(2), 32–42. doi:10.51847/HdKJzZ2S1r
- Boodhoo, K. V., Flickinger, M. C., Woodley, J. M., & Emanuelsson, E. A. (2022). Bioprocess intensification: A route to efficient and sustainable biocatalytic transformations for the future. *Chemical Engineering and Processing - Process Intensification*, 172, 108793. doi:10.1016/j.cep.2022.108793
- Czinkóczy, R., Sakiyo, J., Eszterbauer, E., & Németh, Á. (2023). Prediction of surfactin fermentation with *Bacillus subtilis* DSM10 by response surface methodology optimized artificial neural network. *Cell Biochemistry and Function*, 41(2), 234–242. doi:10.1002/cbf.3776
- Duong-Trung, N., Born, S., Kim, J. W., Schermeyer, M.-T., Paulick, K., Borisyak, M., Cruz-Bournazou, M. N., Werner, T., Scholz, R., Schmidt-Thieme, L., et al. (2023). When

- bioprocess engineering meets machine learning: A survey from the perspective of automated bioprocess development. *Biochemical Engineering Journal*, 190, 108764. doi:10.1016/j.bej.2022.108764
- Ekpenyong, M., Asitok, A., Antai, S., Ekpo, B., Antigha, R., & Ogarekpe, N. (2021). Statistical and artificial neural network approaches to modeling and optimization of fermentation conditions for production of a surface/bioactive glyco-lipo-peptide. *International Journal of Peptide Research and Therapeutics*, 27(1), 475–495. doi:10.1007/s10989-020-10094-8
- Guo, F., Zheng, H., Cheng, Y., Song, S., Zheng, Z., & Jia, S. (2018). Medium optimization for ϵ -poly-L-lysine production by *Streptomyces diastatochromogenes* using response surface methodology. *Letters in Applied Microbiology*, 66(2), 124–131. doi:10.1111/lam.12812
- Gurubasajar, N., Subhakar, A., Dadayya, M., Veeranna, S. H., & Basaiah, T. (2023). Isolation, characterization and quantification of polyhydroxybutyrate producing bacteria *Achromobacter xylosoxidans* KUMBNGBT-63 from different agroresidues. *World Journal of Environmental Biosciences*, 12(2), 35–42. doi:10.51847/YmLaQtfbzH
- Haddad, Y. A., Khalil, R. M., & Nassar, S. T. (2024). Assessing a bespoke training and competency framework for pharmacist independent prescribers in UK care home medicines management. *Annals of Pharmaceutical Education, Safety and Public Health Advocacy*, 4, 172–180. doi:10.51847/qqkfcldze
- Helleckes, L. M., Hemmerich, J., Wiechert, W., von Lieres, E., & Grünberger, A. (2023). Machine learning in bioprocess development: From promise to practice. *Trends in Biotechnology*, 41(6), 817–835. doi:10.1016/j.tibtech.2022.10.010
- Jovic, S., Guresic, D., Babincev, L., Draskovic, N., & Dekic, V. (2019). Comparative efficacy of machine-learning models in prediction of reducing uncertainties in biosurfactant production. *Bioprocess and Biosystems Engineering*, 42(10), 1695–1699. doi:10.1007/s00449-019-02165-y
- Karayathil, A. K., Gurusiddappa, L. H., & Shivasundari, R. S. M. (2023). Human-monkey (*Macaca radiata*) conflict in Chamundi Hill-Mysuru, Karnataka. *World Journal of Environmental Biosciences*, 12(2), 19–25. doi:10.51847/dpGQL4yAXY
- Khuat, T. T., Bassett, R., Otte, E., Grevis-James, A., & Gabrys, B. (2024). Applications of machine learning in antibody discovery, process development, manufacturing and formulation: Current trends, challenges, and opportunities. *Computers & Chemical Engineering*, 182, 108585. doi:10.1016/j.compchemeng.2024.108585
- Laine, A. K., & Virtanen, M. J. (2024). Enhancing employee work effort in SMEs through non-financial motivation tools. *Annals of Organizational Culture, Leadership and External Engagement Journal*, 5, 211–218. doi:10.51847/zNUcvJXUyh
- Latha, S., Sivaranjani, G., & Dhanasekaran, D. (2017). Response surface methodology: A non-conventional statistical tool to maximize the throughput of *Streptomyces* species biomass and their bioactive metabolites. *Critical Reviews in Microbiology*, 43(5), 567–582. doi:10.1080/1040841X.2016.1271308
- Long, J., Zhao, X., Liang, F., Liu, N., Sun, Y., & Xi, Y. (2018). Optimization of fermentation conditions for an *Escherichia coli* strain engineered using the response surface method to produce a novel therapeutic DNA vaccine for rheumatoid arthritis. *Journal of Biological Engineering*, 12(1), 22. doi:10.1186/s13036-018-0110-y
- Meramo, S., Fantke, P., & Sukumara, S. (2022). Advances and opportunities in integrating economic and environmental performance of renewable products. *Bioresources and Bioproducts*, 15, 144. doi:10.1186/s13068-022-02239-2
- Møller, A. K., Thomsen, S. L., Svensson, L. H., & Lindberg, E. J. (2025). Comparative assessment of knowledge, attitude, perceived effectiveness, and practice of complementary and alternative medicine among medical and non-medical students in Bangladesh. *Journal of Medical Sciences and Interdisciplinary Research*, 5(1), 93–107. doi:10.51847/5S5IuL2Ln9
- Mondal, P. P., Galodha, A., Verma, V. K., Singh, V., Show, P. L., Awasthi, M. K., Lall, B., Anees, S., Pollmann, K., & Jain, R. (2023). Review on machine learning-based bioprocess optimization, monitoring, and control systems. *Bioresource Technology*, 370, 128523. doi:10.1016/j.biortech.2022.128523
- Narayanan, H., Behle, L., Luna, M. F., Sokolov, M., Guillén-Gosálbez, G., Morbidelli, M., & Butté, A. (2020). Hybrid-EKF: Hybrid model coupled with extended Kalman filter for real-time monitoring and control of mammalian cell culture. *Biotechnology and Bioengineering*, 117(9), 2703–2714. doi:10.1002/bit.27437
- Narayanan, H., Luna, M., Sokolov, M., Arosio, P., Butté, A., & Morbidelli, M. (2021). Hybrid models based on machine learning and an increasing degree of process knowledge: Application to capture chromatographic step. *Industrial & Engineering Chemistry Research*, 60(29), 10466–10478. doi:10.1021/acs.iecr.1c01317
- Narayanan, H., Luna, M., Sokolov, M., Butté, A., & Morbidelli, M. (2022). Hybrid models based on machine learning and an increasing degree of process knowledge: Application to cell culture processes. *Industrial & Engineering Chemistry Research*, 61(25), 8658–8672. doi:10.1021/acs.iecr.1c04507
- Narayanan, H., Sokolov, M., Morbidelli, M., & Butté, A. (2019). A new generation of predictive models: The added value of hybrid models for manufacturing processes of therapeutic proteins. *Biotechnology and Bioengineering*, 116(10), 2540–2549. doi:10.1002/bit.27097
- Ögmundarson, Ó., Sukumara, S., Herrgård, M. J., & Fantke, P. (2020). Combining environmental and economic performance for bioprocess optimization. *Trends in Biotechnology*, 38(11), 1203–1214. doi:10.1016/j.tibtech.2020.04.011
- Park, S. Y., Kim, S. J., Park, C. H., Kim, J., & Lee, D. Y. (2023). Data-driven prediction models for forecasting multistep ahead profiles of mammalian cell culture toward bioprocess digital twins. *Biotechnology and Bioengineering*, 120(9), 2494–2508. doi:10.1002/bit.28405

- Pinto, J., Ramos, J. R. C., Costa, R. S., Rossell, S., Dumas, P., & Oliveira, R. (2023). Hybrid deep modeling of a CHO-K1 fed-batch process: Combining first-principles with deep neural networks. *Frontiers in Bioengineering and Biotechnology*, 11, 1237963. doi:10.3389/fbioe.2023.1237963
- Ramos, J. R. C., Pinto, J., Poiaraes-Oliveira, G., Peeters, L., Dumas, P., & Oliveira, R. (2024). Deep hybrid modeling of a HEK293 process: Combining long short-term memory networks with first principles equations. *Biotechnology and Bioengineering*, 121(5), 1554–1568. doi:10.1002/bit.28668
- Rathore, A. S., Nikita, S., Thakur, G., & Mishra, S. (2023). Artificial intelligence and machine learning applications in biopharmaceutical manufacturing. *Trends in Biotechnology*, 41(4), 497–510. doi:10.1016/j.tibtech.2022.08.007
- Richter, J., Wang, Q., Lange, F., Thiel, P., Yilmaz, N., Solle, D., Zhuang, X., & Beutel, S. (2025). Machine learning-powered optimization of a CHO cell cultivation process. *Biotechnology and Bioengineering*, 122(5), 1153–1164. doi:10.1002/bit.28943
- Rodriguez-Granose, D., Jones, A., Loftus, H., Tandeski, T., Heaton, W., Foley, K. T., & Silverman, L. (2021). Design of experiment (DOE) applied to artificial neural network architecture enables rapid bioprocess improvement. *Bioprocess and Biosystems Engineering*, 44(6), 1301–1308. doi:10.1007/s00449-021-02529-3
- Rogers, A. W., Vega-Ramon, F., Yan, J., del Río-Chanona, E. A., Jing, K., & Zhang, D. (2022). A transfer learning approach for predictive modeling of bioprocesses using small data. *Biotechnology and Bioengineering*, 119(2), 411–422. doi:10.1002/bit.27980
- Sağır, E., Yucel, M., & Hallenbeck, P. C. (2018). Demonstration and optimization of sequential microaerobic dark- and photo-fermentation biohydrogen production by immobilized *Rhodobacter capsulatus* JP91. *Bioresource Technology*, 250, 43–52. doi:10.1016/j.biortech.2017.11.018
- Salem, H. M., Watanabe, S., & Chang, A. H. (2025). Ethical concerns in managing anorexia nervosa: A content analysis of ethics consultation records. *Asian Journal of Ethics, Health and Medicine*, 5, 25–35. doi:10.51847/oHEI6FgL3V
- Sanjivkumar, M., Silambarasan, T., Balagurunathan, R., & Immanuel, G. (2018). Biosynthesis, molecular modeling and statistical optimization of xylanase from a mangrove associated actinobacterium *Streptomyces variabilis* (MAB3) using Box-Behnken design with its bioconversion efficacy. *International Journal of Biological Macromolecules*, 118(Pt A), 195–208. doi:10.1016/j.ijbiomac.2018.06.063
- Schweidtmann, A. M., Zhang, D., & von Stosch, M. (2024). A review and perspective on hybrid modeling methodologies. *Digital Chemical Engineering*, 10, 100136. doi:10.1016/j.dche.2023.100136
- Singh, V., Haque, S., Niwas, R., Srivastava, A., Pasupuleti, M., & Tripathi, C. K. M. (2017). Strategies for fermentation medium optimization: An in-depth review. *Frontiers in Microbiology*, 7, 2087. doi:10.3389/fmicb.2016.02087
- Smiatek, J., Clemens, C., Montano Herrera, L., Arnold, S., Knapp, B., Presser, B., Jung, A., Wucherpfennig, T., & Bluhmki, E. (2021). Generic and specific recurrent neural network models: Applications for large and small scale biopharmaceutical upstream processes. *Biotechnology Reports*, 31, e00640. doi:10.1016/j.btre.2021.e00640
- Torres, A., & Fernandez, M. (2025). Large language models for pharmaceutical knowledge management: A critical review. *Pharmacophore*, 16(6), 12–21. doi:10.51847/XPhOzUYFSY
- Walther, C., Voigtmann, M., Bruna, E., Abusnina, A., Tscheließnig, A.-L., Allmer, M., Schuchnigg, H., Brocard, C., Föttinger-Vacha, A., & Klima, G. (2022). Smart process development: Application of machine-learning and integrated process modeling for inclusion body purification processes. *Biotechnology Progress*, 38(3), e3249. doi:10.1002/btpr.3249
- Wang, Y., Yang, G., Sage, V., Xu, J., Sun, G., He, J., & Sun, Y. (2021). Optimization of dark fermentation for biohydrogen production using a hybrid artificial neural network (ANN) and response surface methodology (RSM) approach. *Environmental Progress & Sustainable Energy*, 40(1), e13485. doi:10.1002/ep.13485
- Wei, P., Si, Z., Lu, Y., Yu, Q., Huang, L., & Xu, Z. (2017). Medium optimization for pyrroloquinoline quinone (PQQ) production by *Methylobacillus* sp. zju323 using response surface methodology and artificial neural network–genetic algorithm. *Preparative Biochemistry & Biotechnology*, 47(7), 709–719. doi:10.1080/10826068.2017.1315596