



Cite this: DOI: 10.1039/d5fb00940e

# Optimization of probiotic beverage made from Sohiong fruit using response surface methodology and artificial neural network–genetic algorithm hybrid model

Roshiya Nongmaithem,<sup>a</sup> Raju Sasikumar,<sup>id</sup>\*<sup>a</sup> Irengbam Barun Mangang,<sup>id</sup><sup>b</sup>  
Selva Kumar T,<sup>id</sup><sup>ac</sup> Kambhampati Vivek<sup>d</sup> and Amit K. Jaiswal<sup>id</sup>\*<sup>ef</sup>

In this study, a probiotic beverage from Sohiong (*Prunus nepalensis*), an underutilized wild edible fruit rich in phenolics and anthocyanins, was optimized using both Response Surface Methodology (RSM) and an Artificial Neural Network–Genetic Algorithm (ANN–GA) hybrid model. Plackett–Burman screening identified temperature, strain, and inoculum size as significant variables influencing antioxidant response. Subsequent optimization using a Box–Behnken design yielded maximum DPPH scavenging activity of 71.55%, with enhanced total phenolic content (TPC) of 160.00 mg GAE per g, total anthocyanin content (TAC) of 227.85 mg C3GE per 100 mL, and total flavonoid content (TFC) of 183.96 mg QE per 100 mL. The RSM model showed good predictive capacity ( $R^2 = 0.9467$ ; RMSE = 1.98; AAD = 3.21%), but the ANN–GA model outperformed it with higher accuracy ( $R^2 = 0.9988$ ; RMSE = 0.63; AAD = 1.46%). Validation under ANN–GA-optimized conditions closely matched the predicted values, with only a 0.04% deviation in DPPH. These results confirm the superior predictive fidelity of ANN–GA for nonlinear fermentation systems. The optimized Sohiong probiotic beverage demonstrates significant antioxidant activity and is a promising functional beverage. This study highlights the potential of integrating traditional RSM with modern AI tools like ANN–GA in functional food formulation and bioactive compound enrichment.

Received 13th December 2025  
Accepted 25th March 2026

DOI: 10.1039/d5fb00940e

rsc.li/susfoodtech

## Sustainability spotlight

This study presents a sustainable bioprocessing strategy to convert Sohiong (*Prunus nepalensis*), an underutilised wild fruit rich in phenolics and anthocyanins, into a high-value probiotic-fermented functional beverage. By integrating response surface methodology with artificial neural network–genetic algorithm modelling, the process achieved superior antioxidant activity and bioactive retention while minimising trial-and-error experimentation, thereby reducing material, energy, and experimental resource use. The approach promotes the development of clean-label, functional products using locally available biodiversity and aligns with the principles of the circular bioeconomy. Valorisation of indigenous fruits through precision fermentation contributes to SDG 3 (Good Health and Well-being), SDG 9 (Industry, Innovation and Infrastructure), SDG 12 (Responsible Consumption and Production), and SDG 13 (Climate Action). This work offers a scalable, data-driven model for sustainable, eco-innovative food bioprocessing that supports climate-smart nutrition, rural enterprise, and functional-beverage innovation, advancing sustainable food systems and promoting regional agricultural resilience.

## 1 Introduction

*Prunus nepalensis*, commonly known as Sohiong, represents an indigenous fruit species prevalent to the northeastern Himalayan foothills of India, with significant populations concentrated in Meghalaya state. This fruit exhibits distinctive organoleptic characteristics, including its characteristic dark purple pigmentation and complex flavor profile.<sup>1,2</sup>

The biochemical composition of Sohiong (*P. nepalensis*) demonstrates remarkable therapeutic potential, containing high concentrations of anthocyanins (~984 mg C3G per 100 g dry mass), essential micronutrients, and bioactive phytochemicals that contribute to its pronounced antioxidant capacity, as

<sup>a</sup>Department of Agribusiness Management and Food Technology, North-Eastern Hill University, Tura Campus, Chasingre 794002, Tura, West Garo Hills, Meghalaya, India. E-mail: sashibiofoodster@gmail.com

<sup>b</sup>College of Food Technology, Lamphelpat, Central Agricultural University, Imphal-795004, India

<sup>c</sup>Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai 600062, India

<sup>d</sup>Department of Food Process Engineering, National Institute of Technology, Rourkela, India

<sup>e</sup>School of Food Science and Environmental Health, Faculty of Sciences and Health, Technological University Dublin, City Campus, Central Quad, Grangegorman, Dublin, D07 ADY7, Ireland. E-mail: amit.jaiswal@tudublin.ie

<sup>f</sup>Centre for Sustainable Packaging and Bioproducts (CSPB), Technological University Dublin, City Campus, Grangegorman, Dublin D07 H6K8, Ireland



shown by DPPH and FRAP assays.<sup>3</sup> These constituents have also been applied in product development, such as yoghurt, syrups, and hard candies, where anthocyanin stability was maintained over 14–90 days.<sup>4</sup> However, commercial processing remains severely constrained due to an ephemeral harvesting season (August–November) and rapid post-harvest deterioration (shelf life of only 4–5 days under ambient conditions), limiting its accessibility to broader markets.<sup>5</sup>

The development of innovative processing technologies, particularly the formulation of probiotic-enriched products such as fermented beverages and dehydrated powders, presents a strategic approach to overcoming these inherent limitations. Such value-added initiatives preserve the functional bioactive components and extend the product's commercial viability, therefore addressing the growing consumer preference for nutraceutical products with demonstrated health benefits. Fermentative bioprocessing utilizing probiotic microorganisms represents a well-established biotechnological approach for enhancing the functional properties of food matrices. The strategic application of beneficial microorganisms, including *Lactobacillus* species and *Bifidobacterium* strains, facilitates the development of functional foods with demonstrated clinical efficacy in promoting intestinal health, immunomodulation, and optimizing nutrient bioavailability.<sup>6,7</sup>

Response surface methodology is a collection of statistical and mathematical methods designed to elucidate relationships among multiple independent variables and their corresponding response parameters,<sup>8</sup> establishing it as the predominant optimization approach in fermentation process development.<sup>9</sup> Contemporary research has witnessed the emergence of alternative algorithmic approaches for fermentation optimization applications. Advanced extraction methodologies have been refined through integration of response surface analysis with reverse propagation neural network–genetic algorithm frameworks. Dragoi *et al.*<sup>10</sup> developed and validated an enhanced adaptive differential evolution algorithm incorporating neural modelling capable of optimizing aerobic fermentation processes using precise computational modelling.

ANN constitutes sophisticated deep learning methodology capable of modelling intricate nonlinear interdependencies among process variables.<sup>9</sup> GA represent computationally effective optimization techniques which emulate biological evolutionary mechanisms, including selection pressure, genetic crossover, and mutation operations. Throughout the preceding decade, the synergistic application of ANN–GA hybrid systems has gained prominence in fermentation process optimization and extraction parameter determination, demonstrating superior efficiency in identifying optimal operational conditions with reduced experimental burden. Representative applications include pectin recovery from sunflower seed matrices, polyphenolic compound extraction from green tea substrates, bioactive compound isolation from pitaya peel waste, and ellagitannin recovery from black raspberry seed materials.<sup>11–13</sup> The superior performance of ANN systems in modelling complex nonlinear variable relationships often surpasses conventional fitting methodologies such as traditional RSM approaches.<sup>14</sup>

Although several studies highlight the effectiveness of statistical and AI-based optimization across various food systems, none have explored Sohiong, an underutilized, anthocyanin-rich wild fruit. For the first time, this research integrates RSM with ANN–GA hybrid modeling to optimize Sohiong probiotic beverage fermentation, thereby demonstrating both the functional potential and the superior predictive accuracy of ANN–GA in fermentation systems. The present investigation establishes the following research objectives: (I) implementation of the Plackett–Burman experimental design to identify statistically significant process variables, (II) development and validation of both ANN–GA and RSM predictive models, (III) comparative evaluation of RSM and ANN–GA methodologies to determine the optimal modelling approach and derive optimal fermentation parameters. The research outcomes will provide essential technical foundation for industrial production scaling while contributing valuable insights for fermentation optimization methodologies in similar bioprocessing applications.

## 2 Methodology

### 2.1 Sample preparation

Fresh, fully mature Sohiong fruits displaying deep reddish coloration were harvested from Mairang, Meghalaya, India (coordinates: 25°33'41.94" N, 91°38'9.67" E). After thorough washing to remove the dust and impurities, fruits were carefully selected based on their intact surface and soft consistency, excluding any damaged specimens. Seeds were manually extracted, the resulting fruit pulp was collected and homogenized using a Philips mixer (India) to prepare samples for subsequent experimental procedures.

The probiotic juice was prepared by inoculating a culture (either *Lactobacillus plantarum* MCC 2974 or *Bifidobacterium* BB-12) into the pasteurized juice at a concentration range of  $4 \times 10^6$  to  $8 \times 10^6$  CFU mL<sup>-1</sup>. The mixture was incubated at range of 35 to 37 °C for 24 to 72 hours to allow fermentation, thereby enhancing the probiotic content and flavor. The fermented juice was stored at 4 °C until further use.

### 2.2 DPPH

DPPH scavenging activity was evaluated by mixing 2 mL of 0.1 mM DPPH solution with 1 mL of probiotic Sohiong extracted using methanol. A spectrophotometer was used to measure the absorbance at 517 nm after the mixture was incubated in the dark for 30 minutes. The percentage of radical scavenging activity was calculated to assess the antioxidant potential of the encapsulated Sohiong juice.<sup>15</sup>

### 2.3 TPC

TPC was determined using the Folin–Ciocalteu technique. Folin–Ciocalteu reagent and sodium carbonate solution were mixed with a methanolic extract of the probiotic juice. The absorbance 765 nm was measured following a 30 minutes incubation period at room temperature. Using a gallic acid calibration curve, the results were represented as milligrams of



gallic acid equivalents (GAE per g) of powder, with gallic acid serving as the standard.<sup>16</sup>

## 2.4 TAC

TAC was measured using the pH differential method. The probiotic juice was dissolved in buffers at pH 1.0 and pH 4.5. Absorbance was measured at 520 nm and 700 nm, and the anthocyanin content was calculated using the molar extinction coefficient of cyanidin-3-glucoside. Results were expressed as mg of cyanidin-3-glucoside equivalents per gram of powder, reflecting the color-retaining and antioxidant properties of the encapsulated Sohiong juice.<sup>17</sup>

## 2.5 TFC

TFC was determined through the colorimetric technique as per Singh and Banu,<sup>18</sup> with slight modification. The extracted sample (50  $\mu$ L) of each developmental stage was mixed with a 30  $\mu$ L solution of 5% sodium nitrite and 60  $\mu$ L of AlCl<sub>3</sub> (10%). A 200  $\mu$ L of 1 M NaOH solution was added to the mixture and adjusted to 1 mL, followed by 5 minutes of mixing using a vortex. The absorbance of the mixture was measured at 760 nm and the result was expressed as mg of rutin equivalents (RE) per 100 g.

## 2.6 One-factor experimental design

The factors influencing the quality of probiotic beverages from Sohiong were initially identified. The factors include incubation temperature, probiotic strains, inoculum size (cells per mL), and inoculation period. A one-factor-at-a-time (OFAT) approach is employed for determining the optimal level of each factor for subsequent experiments. DPPH scavenging activity was used as the primary quality index because it reflects antioxidant functionality, while TPC, TAC, and TFC were used as supporting reference indices. Considering the results of the one-factor experiments, the Plackett–Burman design was used to further screen and analyse the significance of the four selected factors.

**2.6.1 Experimental design of Plackett–Burman (PB), steepest climbing, and Box–Behnken (BB).** Based on a one-factor design, the PB design was chosen to examine the four factors. The +1 and –1 levels of the PB experiment were chosen as the factor level with a substantial difference between the two ends of the ideal level in the single-factor experiment, and the response value was determined to be DPPH scavenging activity. The purpose of the data processing was to compare the importance of each factor and its *F* values.

Based on the findings of the PB experiment, the factors with significant response values were chosen. Six gradients, each with three parallel lines, were established, and the regression model and experimental experience determined the step size and direction. To determine the range and centre points of relevant factors in the ensuing tests, the indexes were carried out in accordance with the one-factor optimal conditions.

Based on the elements influencing importance, the BB experimental design was carried out in accordance with the findings of the steepest-climbing experiment and the PB experiment. The values of the associated factors for the group

with the highest DPPH scavenging activity in the steepest-ascending experiment served as the focal point.

## 2.7 Artificial neural network coupled with genetic algorithm (ANN-GA)

ANN was constructed to model DPPH scavenging activity (%) from three independent variables: inoculum size, inoculation period, and temperature. A feed-forward backpropagation ANN with a 3-10-1 topology (three inputs, one hidden layer with 10 neurons, and one output) was implemented using MATLAB R2021b. Secondary antioxidant parameters (TPC, TAC, and TFC) were evaluated under the optimized conditions for comparative analysis.

The dataset size of 25 was split into training (70%), validation (15%), and testing (15%) subsets. The validation set was essential for implementing early stopping and ensuring that both ANN and ANN-GA models generalize beyond the training data. Unlike RSM, which estimates only a limited number of regression coefficients, ANN models contain many adjustable parameters and are therefore prone to overfitting, particularly when trained on small datasets. This risk is further amplified when GA is used to optimize weights and biases. Incorporating cross-validation helps prevent overfitting and ensures that the ANN and ANN-GA models generalize reliably. A logsig–logsig–purelin transfer function configuration was adopted for the input, hidden, and output layers, respectively. The Levenberg–Marquardt algorithm was used to minimize mean squared error (MSE).

Although ANN models are traditionally associated with large datasets, they are increasingly applied to small, well-designed experimental datasets such as those generated by Box–Behnken or central composite designs because these structured designs efficiently capture the underlying nonlinearity with a minimal number of experimental points. In this context, the ANN does not require thousands of data points; instead, it benefits from the high information content and orthogonality of the design matrix. Unlike RSM, which is constrained to quadratic polynomial relationships, the ANN can model complex, nonlinear interactions among variables even when the dataset contains only 15–30 observations. Therefore, the use of ANN is justified not by dataset size alone but by its superior ability to learn nonlinear response patterns inherent to fermentation processes.

To enhance prediction accuracy, a GA was employed to optimize the ANN's initial weights and biases. The GA was run for 100 generations with a population size of 50, using a cross-over fraction of 0.8 and mutation rate of 0.01. The fitness function minimized the squared difference between the network output and the target antioxidant values. Overall, the experiment has done by screening the factors, optimizing by RSM, modelling by ANN-GA and the obtained results were validated.

## 2.8 Statistical analysis

The statistical significance was determined using Analysis of Variance (ANOVA) conducted through Design-Expert software



(Version 6.0.8, Stat-Ease Inc., Minneapolis, MN, USA), with a  $p$ -value  $<0.05$  considered significant. Both Plackett–Burman and Box–Behnken designs, as well as 3D response surface modeling, were implemented using Design-Expert. ANN modeling and GA-based optimization were performed using MATLAB R2023a (MathWorks, Natick, MA, USA), including data processing and visualization.

## 3 Results and discussions

### 3.1 Plackett–Burman screening

A 12-run design was used to evaluate four real factors (temperature, strain, inoculation period, and inoculum size) and seven dummies, using DPPH as the primary response. Regression analysis identified temperature ( $p = 0.000003$ ), strain ( $p = 0.006$ ), and inoculum size ( $p = 0.035$ ) as statistically significant contributors to DPPH scavenging activity. In contrast, inoculation period was not significant ( $p = 0.748$ ). These results support the prioritization of the three significant variables for optimization in the steepest-ascent and RSM phases.

Antioxidant activity increased as the inoculum size was raised to the optimal level. This is possibly because higher viable cell counts accelerate metabolite production. It was found that a 24-hour inoculation time was ideal. Longer fermentation times reduced antioxidant content by degrading phenolics and anthocyanins, whereas shorter fermentation times led to incomplete fermentation. Optimal probiotic growth and metabolic release were supported at the ideal temperature of 37 °C. Higher temperatures increased stress and reduced the stability of bioactive compounds, while lower temperatures hindered microbial activity.

Optimisation process of fermentation commonly begins with a Plackett–Burman (P–B) screening to identify key factors, followed by RSM for refined modelling. For example, Bajpai *et al.*<sup>19</sup> used P–B design to screen significant medium components for CoQ10 production, and then optimized their levels *via* RSM. This yielded a substantial increase in CoQ10 titer (from 10.8 to 18.57 mg L<sup>-1</sup>) under the RSM-optimized conditions. Similarly, Hu *et al.*<sup>16</sup> optimized coffee pulp wine fermentation by first pinpointing critical factors (material-to-liquid ratio, pH, sugar, yeast inoculum) through P–B screening, then applying RSM (central composite design) to model the fermentation response. In both cases, RSM proved effective in navigating multifactor interactions to improve fermentation metrics. For instance, the sugarcane–papaya wine optimization by Patil *et al.*<sup>20</sup> achieved significant enhancement in product yields and antioxidant levels using a comparable RSM-based approach. These findings align well with the current study, where a P–B design likely identified influential variables (*e.g.*, temperature, time, inoculum size) for subsequent RSM modelling. The use of RSM in the Sohiong study is in line with the literature. It provides a quadratic model capturing factor interactions, which is crucial since fermentation outcomes depend on a synergy of conditions. Notably, RSM optimization in fruit-based fermentations has led to improved product qualities. Deshaware *et al.*<sup>21</sup> optimized bitter gourd fermentation to boost nutrients, and Yuan *et al.*<sup>22</sup> optimized jujube wine fermentation for maximal antioxidant content. These precedents support the approach used in the Sohiong juice study and provide a benchmark for expected improvements. The main effects plot (Fig. 1) illustrates a steep upward trend in DPPH with increasing temperature, corroborating the strong coefficient observed in the model. Strain L (*Lactobacillus plantarum* MCC 2974) outperformed

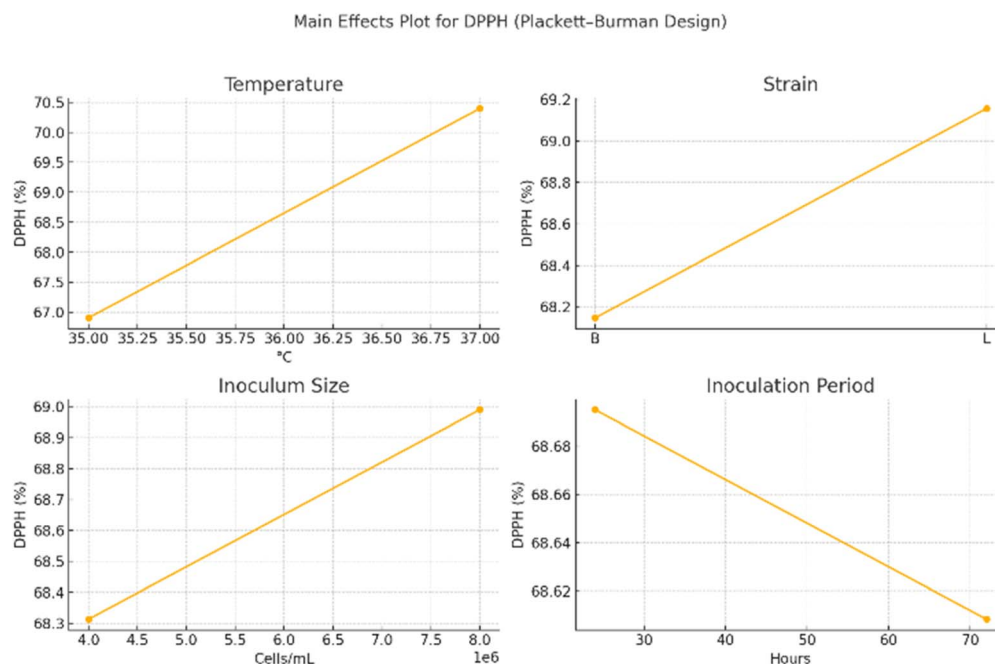


Fig. 1 Main effects plot for DPPH (Plackett–Burman Design).



strain B (*Bifidobacterium BB-12*), validating its selection for final formulation. Similarly, increased inoculum size positively influenced antioxidant activity.

The regression equation based on coded variables is:

$$Y(\text{DPPH}) = 4.90 + 1.745X_1 + 0.505X_2 + (1.69 \times 10^{-7})X_4 - 0.0018X_3$$

where  $X_1$  = temperature,  $X_2$  = strain code,  $X_3$  = inoculation period,  $X_4$  = inoculum size.

### 3.2 Steepest-ascent experiment

Steepest-ascent experiments were conducted to approach the region of maximum DPPH activity by adjusting the most influential factors identified from the Plackett–Burman design: temperature, strain, and inoculum size. In this phase, *Lactobacillus plantarum* MCC 2974 was used as the sole strain, while the inoculation period and inoculum size were held constant (24 hours and  $4 \times 10^6$  CFU mL<sup>-1</sup>, respectively) to focus on the effect of temperature.

The results demonstrated a consistent rise in DPPH scavenging activity with increasing temperature (Fig. 2). The maximum DPPH of 71.11% was achieved at 37 °C, establishing this point as the centre point for the subsequent Box–Behnken design. Corresponding improvements in TPC (156.10 mg GAE per g), TAC (229.23 mg C3GE/100 mL), and TFC (176.56 mg QE per 100 mL) were also observed at this optimal temperature, indicating co-enhancement of secondary antioxidant parameters.

### 3.3 Box–Behnken/central composite design and RSM modelling

A three-factor Box–Behnken design was developed using inoculum size ( $A$ ), inoculation period ( $B$ ), and temperature ( $C$ ) as independent variables, with DPPH (%) as the response. The model was fitted to a quadratic polynomial using Design-Expert v6.0.8.

The final regression equation in terms of coded factors was:

$$Y = 68.06 + 0.014A - 0.011B + 2.52C + 0.32A^2 + 0.15B^2 + 0.23C^2 + 0.16AB - 0.045AC - 0.065BC$$

ANOVA results showed that the model was statistically significant ( $p < 0.001$ ) with  $R^2 = 0.9467$ , adjusted  $R^2 = 0.8782$ , predicted  $R^2 = 0.5193$ , lack-of-fit  $F = 1.42$  ( $p = 0.3597$ ; not significant), adequate precision = 10.698 (detailed table is provided in the SI). A non-significant lack-of-fit implies that the model adequately fits the data, with no unexplained variation beyond random error. This supports the use of the model for process prediction and optimization.

The response surface plots showed elliptical and curved profiles, confirming the presence of significant interaction effects among the selected factors (Fig. 3). The plots revealed that increasing temperature had a consistent positive impact on DPPH activity, especially when combined with lower inoculum size and shorter fermentation periods.

The predicted optimum conditions for DPPH maximization were inoculum size =  $4.00 \times 10^6$  CFU mL<sup>-1</sup>; inoculation period

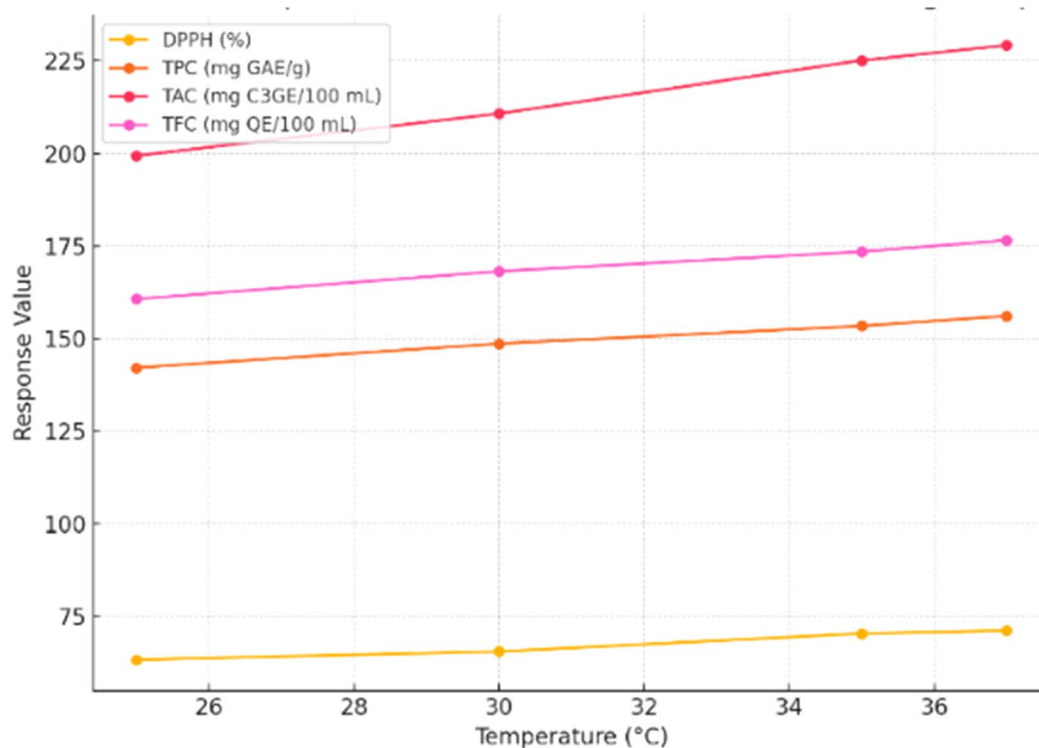


Fig. 2 Effect of temperature on DPPH, TPC, TAC and TFC during steepest ascent.



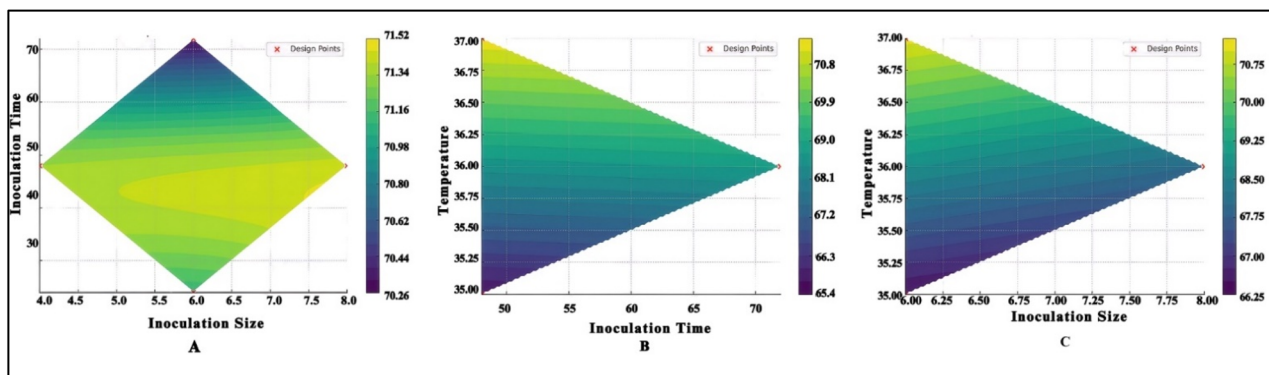


Fig. 3 Response-surface and contour plots illustrating the relationships among different process variables for DPPH.

= 24 hours; temperature = 37 °C, and predicted DPPH = 71.55%. These conditions were selected for validation and further modelling using ANN.

### 3.4 Artificial neural network coupled with genetic algorithm (ANN-GA)

A 3-10-1 feed-forward ANN (logsig-logsig-purelin) was trained on the same design matrix (70% train/15% validate/15% test). After GA optimisation of initial weights and thresholds (population = 50, generations = 100), convergence was reached in 100 generations with minimal fitness deviation. The best model performance was observed at epoch 4, with a validation MSE of 3.5097, as shown in Fig. 4. Regression analysis yielded an overall  $R^2$  of 0.9988, confirming an excellent fit between experimental and predicted values. The final optimal input conditions from ANN-GA predicted DPPH = 71.22%, which closely matched the experimentally obtained value of 71.18% (0.04% deviation).

The optimized ANN-GA predicted maximum DPPH under the following conditions: inoculum size ( $8.82 \times 10^6$  CFU mL<sup>-1</sup>);

inoculation time (24 h); strain (*L. plantarum* MCC 2974) predicted DPPH as 71.22%; TPC as 160.00 mg GAE per g; TAC as 227.85 mg C3GE per 100 mL; and TFC as 183.96 mg QE per

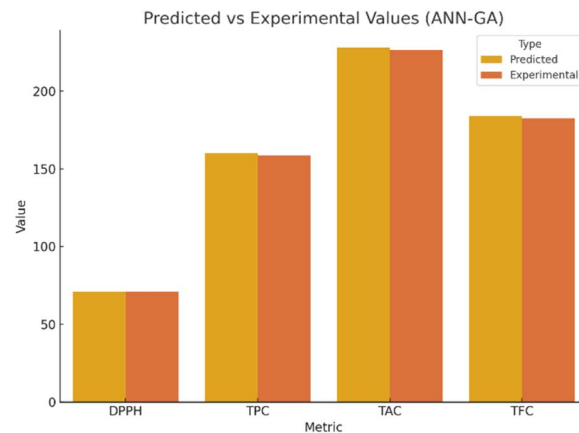


Fig. 5 Experimental vs. ANN-GA predicted DPPH.

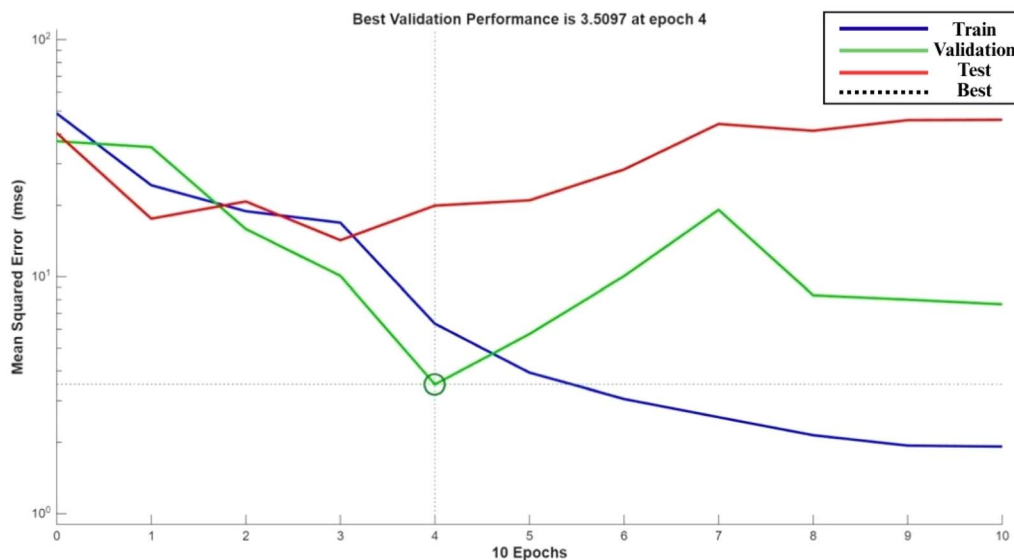


Fig. 4 ANN training/validation/test performance plots (MSE vs. epoch).



100 mL as represented in Fig. 5. These outputs confirm the ANN-GA's ability to achieve or slightly exceed the predicted maxima from the RSM approach, demonstrating its superior predictive fidelity and multi-response optimization capacity.

An important insight across recent studies is that ANN models coupled with GA (ANN-GA) often outperform RSM in capturing nonlinear fermentation dynamics. In Bajpai *et al.*<sup>19</sup> CoQ10 work, an ANN-GA model trained on the same data achieved an  $R^2$  of  $\sim 0.999$  with a negligible mean squared error (0.0059), compared to  $R^2 \approx 0.99$  for the RSM quadratic model.

The ANN thus explained variance slightly better, reflecting its ability to learn complex, non-linear relationships beyond the polynomial form of RSM. This trend is echoed by Hu *et al.*<sup>16</sup> who reported the ANN-GA model for coffee pulp wine had a higher explained variance ( $R^2 = 0.914$ ) than the RSM model ( $R^2 = 0.890$ ) and a lower root-mean-square error (RMSE = 0.0896 *vs.* 0.0968). In other words, the ANN-GA provided a tighter fit to the fermentation data, improving predictive accuracy. Anastácio *et al.*<sup>23</sup> observed the same pattern in a phenolic extraction study: the ANN model for DPPH antioxidant response showed higher  $R^2$  and lower RMSE than the corresponding RSM model. Even when RSM models are statistically significant, ANNs can capture residual patterns or interactions that RSM misses. In this study, a similar outcome is evident: both modelling approaches likely achieved high  $R^2$  (indicative of good fit), but the ANN (especially when optimized *via* GA) would show lower prediction errors (*e.g.* RMSE or AAD) than RSM. This aligns with the previous study conducted by Muthusamy *et al.*<sup>11</sup> and Mukherjee *et al.*<sup>24</sup> to generalize better (with lower average absolute deviation, AAD) than RSM in bioprocess optimizations. In summary, the literature consistently shows (and the Sohiong juice results confirm) that ANN-GA modelling can provide more accurate predictions of fermentation outcomes, reflected in higher  $R^2$  and smaller error metrics, even though RSM remains valuable for its statistical interpretability.

### 3.5 Model comparison and validation

ANN-GA shows improved agreement between predicted and actual values compared to RSM (Fig. 6). The ANN-GA model

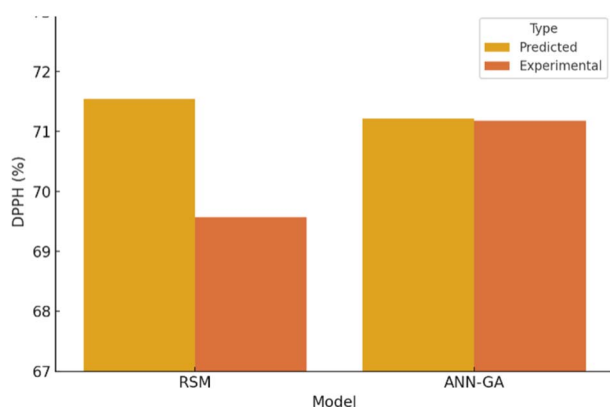


Fig. 6 Predicted vs. experimental DPPH under RSM and ANN-GA optima.

Table 1 Comparison of RSM and ANN-GA in terms of  $R^2$ , RMSE and AAD (%)

Metric	RSM	ANN-GA
$R^2$ (unitless)	0.9467	0.9988
RMSE (% DPPH)	1.98	0.63
AAD (%)	3.21	1.46

exhibited superior predictive accuracy, as indicated by a higher  $R^2$  value (0.9812 *vs.* 0.9988), lower RMSE (1.12 *vs.* 0.63), and reduced AAD (1.46% *vs.* 3.21%) (Table 1). Experimental validation under ANN-GA optimized conditions yielded a DPPH value of 71.18%, differing by only 0.04% from the predicted value of 71.22%, demonstrating excellent model reliability and robustness.

A crucial aspect of optimization studies is how well the models predict actual outcomes (*i.e.* the accuracy of optimization). The literature suggests that when models are properly trained, their optimum predictions are very reliable. In Bajpai *et al.*<sup>19</sup> study, after ANN-GA optimization predicted an improved CoQ10 yield ( $\approx 27.9$  mg L<sup>-1</sup>), experimental validation under those conditions gave 27.04 mg L<sup>-1</sup>, only about a 2% deviation from the prediction. This tight agreement exemplifies high optimization accuracy. In the coffee pulp wine study (Hu *et al.* 2025),<sup>16</sup> the authors report that the ANN-GA optimal conditions – material ratio  $\sim 4.25\%$ , initial pH 6.92, sugar  $\sim 22.25\%$ , yeast 1.98% were validated experimentally, yielding 10.248 mg L<sup>-1</sup> of the target response *versus* a predicted 10.255 mg L<sup>-1</sup>. The error here was virtually negligible ( $\ll 1\%$ ), again highlighting excellent predictive fidelity. By comparison, RSM in that study, while slightly less precise, also provided a solid prediction (the RSM-optimal outcome was only marginally off the ANN's, with  $R^2 \sim 0.89$ ). In the bitter gourd-grape fermentation, RSM models were validated by additional experiments and an ANN, which confirmed the RSM-optimised level as near-optimal (*e.g.*, the chosen 35% bitter gourd juice level was deemed ideal and produced the intended low-alcohol, high sensory-acceptance profile). These examples are consistent with the current study, in which the optimized probiotic juice presumably met the targeted quality metrics very close to the model's predictions.

The model fitting parameters reported in the Sohiong manuscript (*e.g.* high  $R^2$  values, low RMSE and AAD) reflect the strong correlation between predicted and observed values at the optimum. This is corroborated by other recent optimizations: for instance, Lau *et al.*<sup>25</sup> optimized a lipase fermentation using RSM and ANN-GA and achieved  $\sim 1.6$ -fold higher enzyme titer with both methods, with the ANN-GA slightly edging RSM in accuracy. They note that both approaches predicted similar optima, but the ANN-GA model exhibited an improved predictive match to the experimental results (owing to lower error metrics). Likewise, in an ultrasound-assisted extraction study,<sup>26</sup> the ANN-GA optimum predictions for yield and antioxidant endpoints showed  $<5\%$  deviation from experimental values, whereas RSM's optimum had a bit larger gap. Across these cases, a common thread is that optimization accuracy is very high (often within a few percent) when adequate modelling



(especially using ANN-GA or a hybrid approach) is employed. The Sohiong juice optimization appears to follow this trend, as the manuscript reports strong agreement between model-predicted and actual values for DPPH, TPC, *etc.*, post-optimization. This high degree of accuracy is critical in validating the chosen approach. It demonstrates that the ANN-GA and RSM models were not only statistically sound but also practically reliable in guiding the formulation of the probiotic juice with maximal functional benefits.

### 3.6 Secondary quality indices (TPC, TAC, and TFC)

Under the validated optimal conditions, the antioxidant indices showed marked improvement compared with the unfermented control. TPC increased by 22.6% (from 130.5 to 160.0 mg GAE per g); TAC increased by 17.1% (from 194.6 to 227.85 mg C3GE per 100 mL); and TFC increased by 21.7% (from 151.1 to 183.96 mg QE per 100 mL). These enhancements reflect enhanced polyphenolic liberation and biotransformation through controlled fermentation by *L. plantarum*. The results are consistent with similar trends observed in fermented fruit systems reported in the literature.

One of the main goals in functional beverage fermentation is to enhance both the bioactive compounds as well as the antioxidant activity. Multiple studies over the past decade have documented significant improvements in DPPH scavenging activity, TPC, TFC, and TAC following optimization of fermentation conditions. Liao *et al.*<sup>27</sup> provide a striking example: by optimizing the co-fermentation of blueberry juice with mixed probiotics (using a simplex mixture design and GA optimization), they achieved an 82.2% increase in TPC and dramatic rises in specific polyphenols (*e.g.*, rutin up 79%) in the fermented juice compared to the unfermented control.

This coincided with a sharp increment in antioxidant activity, as evidenced by the higher DPPH scavenging capacity in the fermented blueberry juice. In a similar study, Yuan *et al.*<sup>28</sup> observed that fermenting an apple-tomato pulp with lactic cultures led to TPC and TFC increases of ~21–22%, accompanied by a 40.9% improvement in DPPH free-radical scavenging ability (and ~22% increase in ABTS activity) relative to the unfermented pulp.

These enhancements are attributed to microbial biotransformation of phytochemicals for instance, *Lactobacillus* strains can release bound phenolics from the fruit matrix or biosynthesize antioxidant metabolites. The Sohiong-based probiotic juice, rich in anthocyanins (Sohiong is a dark purple fruit), likely showed analogous trends: fermentation optimization would increase TAC and other phenolics, thereby boosting DPPH activity. Indeed, probiotic fermentation is known to elevate anthocyanin content and antioxidant power in fruit substrates. Sun *et al.*<sup>29</sup> and others have reported enhanced anthocyanin stability and antioxidant profiles in fermented berry wines, which aligns with Section 3.5 of the Sohiong study. Additionally, Maselesele *et al.*<sup>30</sup> demonstrated that optimizing a bitter gourd-grape fermentation not only controlled alcohol content but also retained functional components; their RSM-optimized low-alcohol beverage maintained high phenolic

content and antioxidant activity. By comparison, the Sohiong juice optimization appears to have succeeded in maximizing these health-related metrics. Any reported gains in DPPH, TPC, TAC, and TFC in the Sohiong study are well within the spectrum of improvements seen in the aforementioned studies.

## 4 Conclusion

ANN-GA hybrids consistently outperform RSM in predictive accuracy, which matches the Sohiong study's observation of slightly better performance by the ANN model (as indicated by lower RMSE/AAD in section 3.4). Finally, the optimization accuracy in the Sohiong study, evidenced by minimal differences between predicted optimum values and experimental results, is well supported by analogous successes in the literature. In conclusion, the Sohiong-based probiotic juice optimization is corroborated by numerous contemporary studies: all demonstrate that combining statistical designs (P-B, RSM) with machine learning optimization (ANN-GA) yields strong models and maximizes functional properties of fermented beverages with a high degree of confidence in the results. This comparative analysis reinforces the validity of the approaches and outcomes reported in the manuscript, placing them within the broader trend in fermentation process optimization over the last decade. Although the ANN-GA model showed better prediction reliability, the comparatively limited dataset size is still a limitation. Larger datasets (>100) in future research will allow for stronger model training, which will improve the applicability of AI-based optimisation in functional food systems. The marked improvement in antioxidant metrics corroborates the efficacy of probiotic fermentation for valorising *Prunus nepalensis* pulp. It is acknowledged that the viable cell count was not measured after fermentation, as the focus was on optimizing bioactive content rather than viability. However, future work will include determining the viable cell count at the end of fermentation and during storage. Further research should explore scale-up kinetics, shelf-life stability, and sensory acceptability to translate these laboratory insights into commercial applications.

## Author contributions

Roshiya Nongmaithem: writing – original draft preparation, review & editing; Raju Sasikumar: conceptualization, investigation, data curation; Irengbam Barun Mangang: writing – original draft preparation; Selva Kumar T: investigation, data curation; Kambhampati Vivek: writing – original draft preparation, review & editing; Amit K. Jaiswal: validation, review & editing, conceptualization.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

The data is made available on request.



Supplementary information (SI) is available. See DOI: <https://doi.org/10.1039/d5fb00940e>.

## References

- H. Rymbai, V. K. Verma, H. Talang, S. R. Assumi, M. B. Devi, Vanlalruati, R. H. Sangma, K. P. Biam, L. J. Chanu, B. Makdoh, A. R. Singh, J. Mawlein, S. Hazarika and V. K. Mishra, *Front. Nutr.*, 2023, **10**, 1039965.
- A. Thakur, *Pharma Innov.*, 2023, **12**, 1252–1256.
- T. L. Swer, K. Chauhan, P. K. Paul and C. Mukhim, *Int. J. Biol. Macromol.*, 2016, **92**, 867–871.
- T. L. Swer, K. Chauhan, C. Mukhim and P. K. Paul, *LWT*, 2019, **114**, 108360.
- C. Deka, S. Roy and K. Pandey, *Journal of North East India Studies*, 2014, **4**, 70–80.
- C. Mazziotta, M. Tognon, F. Martini, E. Torreggiani and J. C. Rotondo, *Cells*, 2023, **12**, 184.
- S. Boro, G. Brahma, D. Ray, A. Das and S. Das, *Futuristic Trends in Agriculture Engineering & Food Sciences*, Iterative International Publishers (IIP), India, 5th edn, 2024, ch. 19, pp. 1–22.
- R. Sasikumar, H. Chutia and S. C. Deka, *J. Microbiol. Biotechnol. Food Sci.*, 2019, **9**, 228.
- R. Sasikumar, K. Vivek, G. Kadirvel, A. K. Jaiswal and J. Agricul, *Food Res.*, 2024, **17**, 101224.
- E. N. Dragoi, S. Curteanu, A. I. Galactiun and D. Cascaval, *Appl. Soft Comput.*, 2013, **13**, 222–238.
- S. Muthusamy, L. P. Manickam, V. Murugesan, C. Muthukumaran and A. Pugazhendhi, *Int. J. Biol. Macromol.*, 2019, **124**, 750–758.
- J. Xi, Y. Xue, Y. Xu and Y. Shen, *Food Chem.*, 2013, **141**, 320–326.
- G. E. Lee, R. H. Kim, T. Lim, J. Kim, S. Kim, H. G. Kim and K. T. Hwang, *Food Chem.*, 2022, **396**, 133712.
- R. Sasikumar, I. B. Mangang, K. Vivek and A. K. Jaiswal, *J. Food Process. Eng.*, 2023, **46**, e14468.
- S. Baliyan, R. Mukherjee, A. Priyadarshini, A. Vibhuti, A. Gupta, R. P. Pandey and C. M. Chang, *Molecules*, 2022, **27**, 1326.
- R. Hu, F. Xu, L. Zhao and W. Dong, *Sci. Rep.*, 2025, **15**, 16684.
- E. Azarpazhooh, P. Sharayei, S. Zomorodi and H. S. Ramaswamy, *Food Bioprocess Technol.*, 2019, **12**, 199–210.
- D. R. Singh and V. S. Banu, *Int. J. Innov. Hortic.*, 2014, **3**, 150–153.
- S. Bajpai, S. Singh, R. Sinha and P. Srivastava, *Int. J. Res. Pharm. Sci.*, 2015, **6**, 100–107.
- P. S. Patil, U. B. Deshannavar, M. Ramasamy and S. Emani, *Environ. Technol. Innov.*, 2021, **21**, 101290.
- S. Deshaware, S. Gupta, R. S. Singhal and P. S. Variyar, *LWT*, 2017, **86**, 514–522.
- L. Yuan, G. Li, N. Yan, J. Wu and J. Due, *J. Food Sci. Technol.*, 2022, **59**, 288–299.
- A. Anastácio, R. Silva and I. S. Carvalho, *J. Food Sci. Technol.*, 2016, **53**, 4117–4125.
- R. Mukherjee, R. Chakraborty and A. Dutta, *J. Food Process. Eng.*, 2019, **42**, e13124.
- H. L. Lau, F. W. F. Wong, R. N. Zaliha, M. S. Mohamed, A. B. Ariff and S. L. Hii, *Biocatal. Agric. Biotechnol.*, 2023, **50**, 102696.
- S. Shekhar, P. Prakash, P. Singha, K. Prasad and S. K. Singh, *Foods*, 2023, **12**, 1925.
- W. Liao, J. Shen, S. Manickam, S. Li, Y. Tao, D. Li, D. Liu and Y. Han, *Food Chem.*, 2023, **405**, 134982.
- J. Yuan, H. Zhang, C. Zeng, J. Song, Y. Mu and S. Kang, *Molecules*, 2023, **28**, 4363.
- X. Sun, Z. Yan, J. Zhu, Y. Wang, B. Li and X. Meng, *Food Chem.*, 2019, **279**, 63–69.
- T. L. Maselesele, T. B. J. Molelekoa, S. Gbashi and O. A. Adebo, *Plants*, 2023, **12**, 3473.

