Optimization studies on batch extraction of phenolic compounds from *Azadirachta indica* using genetic algorithm and machine learning techniques

Sunita S. Patil [a], Umesh B. Deshannavar [b,c,*], Shambala N. Gadekar-Shinde [d], Amith H. Gadagi [e], Santosh A. Kadapure [b]

[a] *Department of Chemical Engineering, Dr. D.Y. Patil Institute of Engineering, Management and Research, Pune, Maharashtra, India*
[b] *Department of Chemical Engineering, KLE Dr. M. S. Sheshgiri College of Engineering and Technology, Belagavi, Karnataka, India*
[c] *Dr. J. J. Magdum College of Engineering, Jaysingpur, India*
[d] *Department of Chemical Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India*
[e] *Department of Mechanical Engineering, KLE Dr. M. S. Sheshgiri College of Engineering and Technology, Belagavi, Karnataka, India*

ARTICLE INFO

ABSTRACT

Phenolic compounds play a crucial role as secondary metabolites due to their substantial biological activity and medicinal value. These compounds are present in various parts of plant species. This study focused on solid-liquid batch extraction to recover total phenolic compounds from *Azadirachta indica* leaves. The experimental design was based on the Taguchi L_{16} array, considering four independent factors: extraction time, temperature, particle size, and solid-to-solvent ratio. Among these factors, the particle size exerted the maximum influence. Particle size inversely affects the yield of total phenolic content (TPC), while temperature, time, and solid-to-liquid ratio have a direct impact. The process factors concerned were investigated both experimentally and through machine learning techniques. Support vector regression (SVR) and random forest method (RFM) algorithms were utilized for predicting TPC, while a genetic algorithm (GA) was employed to derive optimal process parameters. The GA predicts the optimal extraction factors, yielding the maximum TPC. During this study, these factors were the following: particle size of 0.15 mm, extraction time of 40 min, solid-to-liquid ratio of 1:25 g/mL, and a temperature of 55 °C, with a predicted value of 23.039 mg GAE/g of plant material. Notably, in this study, the SVR values of TPC yield closely matched the experimental values for the training and test data set when compared with the random forest method values.

1. Introduction

The extraction of phenolic compounds from natural sources has garnered increasing attention in recent years due to their distinctive pharmaceutical properties. These compounds are increasingly being utilized as antioxidants, antimicrobials, anti-inflammatories, and antiviral agents in the pharmaceutical industry. Phenolic compounds are extracted from fruits, vegetables, herbal plants, and various edible and non-edible plant species, finding application in medical, food, and cosmetic industries.

Table 1
Process factors and their respective levels in the Taguchi design.

Factor	Description	Unit	L1	L2	L3	L4
A	Particle size	mm	0.150	0.212	0.425	0.600
B	Time	min	15	30	45	60
C	Solid-to-solvent ratio	g/mL	1:20	1:30	1:40	1:50
D	Temperature	°C	25	35	45	55

L1: Level 1, L2: Level 2, L3: Level 3, L4: Level 4.

Interestingly, fruit waste products such as peels, seeds, stalks, and leaves often exhibit higher phenolic compound percentages compared to the edible fleshy parts. For example, grapes (*Vitaceae*) skin [1], papaya (*Carica papaya*) leaves [2], rambutan (*Nephelium lappaceum*) peel [3], avocado (*Persea Americana*) pulp [4], pomegranate (*Punica granatum*) peel [5], and mango (*Mangifera indica*) peel [6]. Phenolic compounds are also extracted from various vegetable and plant parts, including rose (*Pingyin rose*) [7], arrowleaf dock (*Rumex dentatus*) [8], fenugreek *(Trigonella foenum-graecum)* seeds [9], bay (*Laurus Nobilis* L) leaves [10], yarrow dust (*Achillea millefolium*) [11], green tea (*Camellia sinensis*) leaves [12], oak (*Quercus bark*) [13], mushroom waste (Shiitake) (*Lentinula edodes*) [14], bugleweed (*Ajuga ciliate*) leaves [15], and brinjal (Solanum melongena) [16]. Phenolic compound extraction can be achieved through both conventional and advanced methods. Traditional techniques encompass Soxhlet extraction, maceration, percolation, and solid-liquid batch extraction. Advanced methodologies involve ultrasound-assisted extraction, microwave-assisted extraction, supercritical fluid extraction, and pressurized liquid extraction [17]. Extraction parameters such as particle size, temperature, extraction time, solid-to-liquid ratio, and solvent type profoundly influence the extraction process. Particularly in solid-liquid batch extraction, the choice of solvent significantly impacts process efficiency. Different solvents possess varying solubility characteristics towards distinct bioactive compounds. Polar solvents, including methanol, ethanol, n-propanol, and acetone, dissolve polar solutes, while nonpolar solvents like hexane, toluene, and benzene are effective for dissolving nonpolar solutes [18]. During solvent selection, crucial physical properties such as vapor pressure, boiling point, viscosity, surface tension, and polarity should be considered. Given that phenolic compounds exhibit polarity, the preliminary study employs polar solvents such as methanol, ethanol, and acetone.

Neem (*Azadirachta indica*) represents a valuable natural resource abundant in essential chemical constituents. In India, neem has been revered for its medicinal uses since ancient times. All parts of the neem tree, namely leaves, bark, twigs, trunk, and fruit, harbor substantial quantities of phenolic compounds [19]. Traditional extraction methods have been harnessed to extract phenolic compounds from neem leaves, revealing antioxidant and antimicrobial activities [20–22]. The total phenolic content (TPC) has been extracted from neem leaves using conventional stirred batch extraction, altering one variable at a time [23]. However, this conventional optimization approach proves labor-intensive, costly, and time-consuming. Modern optimization techniques such as particle swarm optimization (PSO), genetic algorithm (GA), and ant colony optimization (ACO), which are cost-effective and efficient in terms of time and labor, can be employed [24].

This study endeavors to apply batch extraction methodology utilizing the Taguchi L$_{16}$ experimental design for extracting phenolic compounds from neem leaves. Additionally, modern optimization tools and machine learning techniques are integrated. The objectives encompass identifying optimal process parameters for maximizing TPC yield using a GA, experimental exploration of extraction factors, and employing machine learning techniques such as support vector regression (SVR) and random forest method (RFM) for predicting (TPC).

2. Materials and methods

2.1. Plant material and chemicals

Healthy and mature leaves of *Azadirachta indica* were collected from the college campus in Pune, India. The leaves underwent thorough cleaning with tap water followed by distilled water to remove dirt. After rinsing, the leaves were dried in a hot air oven at 40–45 °C until a constant weight was achieved. The dried material was ground using a Jaipan grinder, then sieved through different mesh number screens, and stored in plastic bags at 5 °C. Analytical grade gallic acid, Folin-Ciocalteu reagent, anhydrous sodium carbonate, and selected solvents were sourced from "S. D. Fine-Chem Ltd., Mumbai, India."

2.2. Design of experiments for batch extraction

Numerous experimental factors such as temperature, pressure, particle size, solvent, solid-to-liquid ratio, time, and frequency can influence extraction efficiency [25]. The experimental design for the four selected factors and their respective levels is based on Taguchi's method. The independent factors considered are extraction time, temperature, particle size, and solid-to-solvent ratio as detailed in Table 1. These independent factors and their levels were screened during a preliminary study using a conventional one-factor-at-a-time approach (Figs. S1–S4, Supplementary data). The ranges for the selected factors were: particle size (0.09, 0.15, 0.212, 0.425, 0.60, and 1.0 mm), temperature (15, 25, 35, 45, 55, and 60 °C), extraction time (15, 30, 45, 60, 75, and 90 min), and solid-to-liquid ratio (1:10, 1:20, 1:30, 1:40, 1:50, and 1:60 g/mL). After each experimental run, the extract underwent filtration using Whatman no. 42 filter paper at atmospheric pressure and then cooled to room temperature for determination of phenolic content. The ranges of these factors were subsequently employed to design the Taguchi experimental design, which is presented in Table 2. The

Table 2
Taguchi design and results for batch extraction.

Run	Particle size (mm)	Time (min)	Solid-to-solvent ratio (g/mL)	Temperature (°C)	TPC yield (mg GAE/g plant material)	S/N[a] ratio (dB)
1	0.150	15	1:20	25	5.109 ± 0.17	14.1667
2	0.150	30	1:30	35	10.105 ± 1.10	20.090
3	0.150	45	1:40	45	11.923 ± 1.03	21.527
4	0.150	60	1:50	55	17.268 ± 0.92	24.745
5	0.212	15	1:30	45	7.389 ± 1.00	17.372
6	0.212	30	1:20	55	8.882 ± 1.10	18.970
7	0.212	45	1:50	25	7.359 ± 0.15	17.336
8	0.212	60	1:40	35	9.915 ± 0.38	19.925
9	0.425	15	1:40	55	6.760 ± 0.14	16.598
10	0.425	30	1:50	45	7.752 ± 0.75	17.788
11	0.425	45	1:20	35	6.191 ± 0.26	15.835
12	0.425	60	1:30	25	4.535 ± 0.27	13.131
13	0.600	15	1:50	35	3.519 ± 0.52	10.928
14	0.600	30	1:40	25	3.488 ± 0.42	10.851
15	0.600	45	1:30	55	6.053 ± 0.55	15.639
16	0.600	60	1:20	45	4.102 ± 0.95	12.259

[a] S/N ratio: Signal-to-noise ratio.

Taguchi experimental design encompassed 16 experiments aimed at investigating the impact of each factor and its level on phenolic content extraction. This approach provided optimal conditions for achieving higher total phenolic yield.

2.3. Experimental methods

2.3.1. Preliminary study for solvent selection

A preliminary study was conducted to determine the suitable solvent for phenolic compound extraction. Acetone, ethanol, and methanol were evaluated using batch extraction methodology, with a particle size of 0.425 mm, solid-to-liquid ratio of 1:40 g/mL, and a temperature of 30 °C. The experiments were carried out in a 500 mL stirred glass reactor. Post 60 min, the extracts were filtered using Whatman no. 42 filter paper and subjected to total phenolic content analysis.

2.3.2. Soxhlet extraction

Dried *Azadirachta indica* powdered leaves (size: 0.425 mm), weighing 10 g. were placed in the thimble of a Soxhlet apparatus. For the extraction of TPC, 400 mL of methanol solvent was added to a 500 mL three-neck round bottom flask [26]. The extraction process involved the siphoning of the extracting solvent from the thimble holder back into the flask when it reached overflow level. This method was carried out at the boiling point of methanol, yielding the maximum recoverable TPC.

2.3.3. Solid-liquid batch extraction

Batch extraction experiments were conducted utilizing a 500 mL agitated cylindrical baffled glass reactor equipped with a marine propeller. Temperature control was achieved using a water bath. The study evaluated the impact of the following four independent factors: extraction time (15–60 min), particle size (0.15–0.60 mm), solid-to-solvent ratio (1:20–1:50 g/mL), and temperature (25–55 °C) on TPC yield. Following each experiment (as detailed in Table 2), the extracts were filtered using Whatman no. 42 filter paper and analyzed by following the process described in Section 2.4. Each of the experiments involved was conducted thrice.

2.4. Analysis of total phenolic content

TPC in the plant extract was determined using the Folin-Ciocalteu method [27]. In this approach, 1 mL of filtered extract was mixed with 60 mL of distilled water in a 100 mL volumetric flask. Subsequently, 5 mL of Folin-Ciocalteu reagent was added and thoroughly mixed. Between 1 and 8 min, 15 mL of 20 % (w/v) sodium carbonate solution was added, followed by thorough shaking. The flask was then filled to volume with distilled water and incubated for 30 min at 40 °C. Sample absorbance was measured at 765 nm using a UV spectrophotometer (UV-1900, Shimadzu). The concentration of total phenolic compounds was determined via the calibration plot and recorded as mg/L of gallic acid equivalent. For the calibration plot, the absorbance was measured using a UV-Spectrophotometer for standard gallic acid solutions with concentrations of 50, 100, 150, 200, 250, 500, and 1000 mg GAE/g of plant material (R^2 = 0.9964). Results from three analyses were averaged, with TPC reported as mg GAE/g of plant material.

In the Taguchi method, process performance is assessed through a quality loss function, converted into a signal-to-noise (S/N) ratio. S/N ratio falls into three categories: nominal, the best; larger, the better; and smaller, the best. Given the study's focus on maximizing phenolic compound yield, the "larger is better" approach was selected for the S/N ratio, which was calculated using Equation (1) [28, 29].

$$\frac{S}{N} = -10\log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{R_i^2}\right) \tag{1}$$
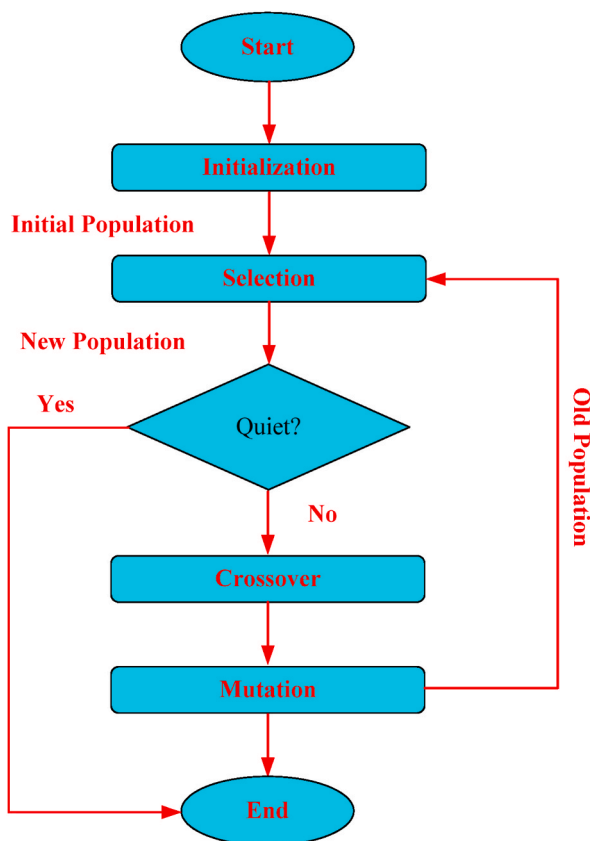
Fig. 1. Flowchart of GA [30].

where n represents the number of repetitions and R_i represents the response (TPC yield) for the ith experiment.

2.5. Ranking of factors

The average TPC yield was calculated for each factor at every level. Delta signifies the difference between the maximum and minimum values of each factor. Factor ranking was determined based on delta values, with factors possessing higher delta values deemed more influential [29].

2.6. Multiple regression analysis and genetic algorithm

2.6.1. Multiple regression analysis

In this study, a multiple regression analysis was conducted to establish an objective function essential for optimizing process parameters: particle size (mm), time (min), solid-to-solvent ratio (g/mL), and temperature (^{o}C). This was done to maximize TPC yield.

2.6.2. Genetic algorithm

GA serves as a heuristic optimization technique inspired by biological reproductive systems. GA is applied to optimize both constrained and unconstrained objective functions. Commencing with a randomized set of populations, the algorithm iteratively refines this population. Each individual in the population is represented by a string of genes known as a chromosome, where each gene signifies a process or design variable. These strings commonly adopt binary bits, alphabets, or real numbers. Each iteration of GA generates offspring through the random pairing of parents, who are selected from the assumed population set based on their fitness levels. This reproductive process begets children for subsequent generations. The algorithm strives to converge upon an optimal set of populations with each iteration. GA aptly handles objective functions characterized by nonlinearity, heuristics, non-differentiability, and discontinuity. The operational course of GA is illustrated by the flowchart (Fig. 1).

With reference to Fig. 1, the GA workflow can be stated as follows:

a. GA generates the random set of the initial population pool using a random number generator.

b. Parent pairing, as generated in Step (a), leads to the creation of a fresh set of populations termed as children. This process, known as crossover, involves selecting parents based on their fitness levels.
c. The pool of children resulting from Step (b) undergoes a mutation process wherein bits in a child's string undergo complementation.
d. Children resulting from Step (b) become parents for the next generation.
e. GA continues iterating until the stipulated stopping criteria are satisfied.

3. Machine learning techniques

3.1. Random forest method

The RFM is extensively applied to regression and classification problems. This supervised machine-learning algorithm hinges on the popular bagging concept. The RFM integrates a multitude of decision trees; predictions from each tree contribute to the final prediction, which is computed by averaging the predictions of all trees. Essential constituents of the random forest include the root node, leaf node, and decision nodes. Nodes emerging after root node splits are recognized as decision nodes, while nodes without further subdivisions are referred to as leaf nodes.

The development of the random forest model involves the following steps:

a. Creation of subsets from the original data, followed by feature and row sampling.
b. Construction of individual decision trees for each data subset.
c. Enhancement of tree growth through node splitting. The node with the lowest impurity serves as the splitting point, and impurity is calculated by Equation (2):

$$\text{Gini Index} = 1 - \sum_{b=1}^{b=B} (P_b)^2 \tag{2}$$

here B signifies individual nodes, and b corresponds to the probability value for each node.

d. Acquisition of the output from each decision tree.
e. Computation of the final prediction as the average prediction from each decision tree is computed as per Equation (3).

$$\overline{\varnothing} = \frac{1}{M} \sum_{m=1}^{m=M} \varnothing_m(n^{'}) \tag{3}$$

where $n^{'}$ signifies an individual sample, M denotes the total tree count, $\varnothing_m$ represents each tree's prediction, and $\overline{\varnothing}$ signifies the average prediction.

3.2. Support vector regression

SVR ranks among the most potent supervised machine learning methods and is a subset of Support Vector Machine (SVM). It excels in handling nonlinear facets of processes owing to its kernel method implementation. SVR effectively estimates real-valued functions and relies on a symmetric loss function for training. This function penalizes both high and low error predictions equally. SVR leverages Vapnik's insensitive approach to establish a flexible tube surrounding the evaluated function. This tube, known as an ε-tube, mitigates errors smaller than a threshold, while points outside the tube are penalized. SVR introduces an ε-insensitive space around the ε-tube function, formulating an optimization problem to determine the tube shape that best approximates the continuous-valued function while balancing prediction error and model complexity. The optimization function, derived from tube geometrical parameters and loss function, is minimized to yield the flattest tube encompassing the maximum training instances. SVR primarily strives to identify the optimal hyperplane such that all data points are at a minimum distance.

The relationship between dependent and independent parameters in SVR is approximated by Equation (4):

$$A = f(B) = W \, x \varnothing(B) + C \tag{4}$$

Here, A signifies the dependent variable, B denotes the independent variable, W represents the weight matrix, C is a constant, $\varnothing(B)$ signifies nonlinearly mapped high-dimensional space.

The regularized function in SVR is minimized by using Equation (5) by the constraint expression as given by Equation (6).

$$\text{Minimize:} R(f) = \gamma \frac{1}{N} \sum_{J=1}^{J=M} (\delta_j + \delta_j^{*}) + \frac{1}{2} W^2 \tag{5}$$
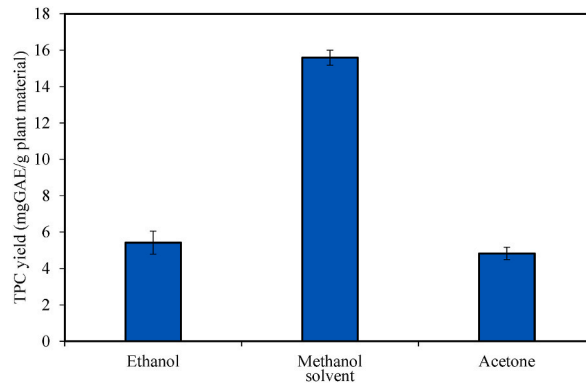
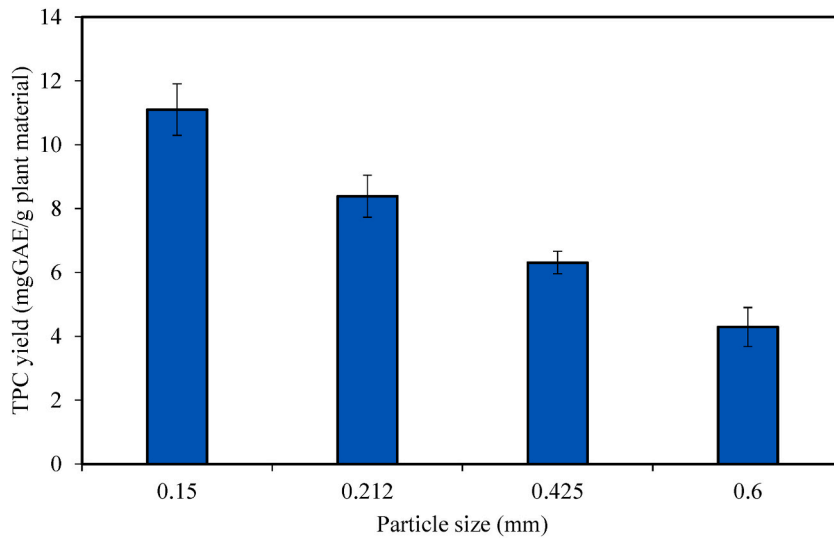**Fig. 2.** Effect of solvent on TPC yield.

Fig. 3. Effect of particle size TPC yield.

$$\text{Constraints:} \begin{cases} A_j - wx\varnothing(B_j) - C \leq (\delta + \delta_j) \\ wx\varnothing(B_j) + C - A_i \leq \delta + \delta_j^* \\ \delta_j, \delta_j \geq 0 \end{cases} \qquad (6)$$

Incorporating the kernel function $K(B, B_j)$ and Lagrangian multipliers (α_j, α_j^*), the solution to the optimization problem can be represented by Equation (7).

$$\text{Maximize:} -\frac{1}{2} \sum_{j=1}^{j=m} 1 \sum_{k=1}^{k=m} \left(\alpha_j - \alpha_j^*\right)\left(\alpha_k - \alpha_k^*\right) K\left(B_j, B_k\right) - \varepsilon \sum_{j=1}^{j=m} \left(\alpha_j - \alpha_j^*\right) + \sum_{j=1}^{j=m} A_j \left(\alpha_j - \alpha_j^*\right) \qquad (7)$$

$$\text{Constraints:} \begin{cases} \sum_{j=1}^{j=m} \left(\alpha_j - \alpha_j^*\right) = 0 \\ \alpha_j, \alpha_j^* \varepsilon (0, \gamma) \end{cases} \qquad (8)$$

The solution to the above optimization problem, as expressed in Equation (7), subjected to the constraint according to Equation (8) yields the final SVR function. as expressed by Equation (9).

$$A = f(B) = \left(\alpha_j - \alpha_j^*\right) x K\left(B, B_j\right) + C \qquad (9)$$
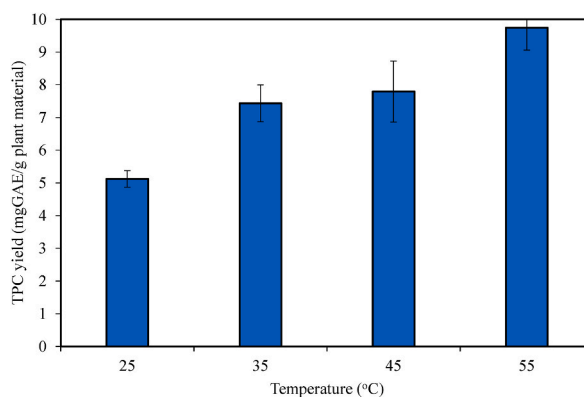
Fig. 4. Effect of temperature on TPC yield.
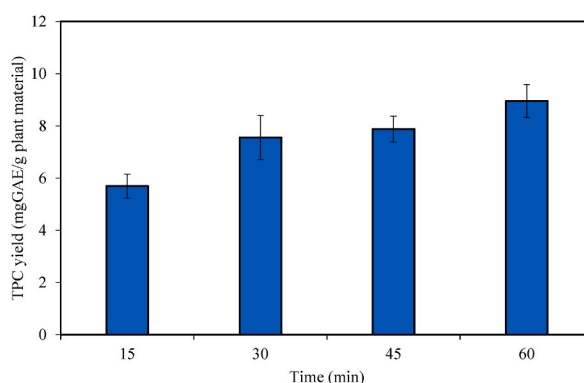

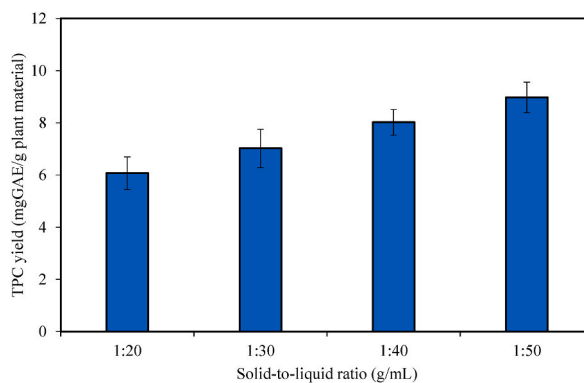**Fig. 5.** Effect of time on TPC yield.

Fig. 6. Effect of solid-to-solvent ratio on TPC yield.

4. Results and discussions

4.1. Preliminary study for solvent selection

An assessment of three chosen solvents—ethanol, methanol, and acetone—was conducted to evaluate their impact on TPC yield. Fig. 2 indicates that solvent efficacy concerning phenolic compounds follows the order: methanol > ethanol > acetone. Variations in phenolic compound count across different solvents may be attributed to their respective solubilities. Given that diffusion serves as the primary mechanism in solid-liquid extraction, smaller solvent molecules permeate the plant matrix more effectively. The superior yield of TPC with methanol can be attributed to its smaller molecular size and physical properties [31]. Consequently, methanol was chosen for subsequent investigations.

Table 3

Ranking of process factors based on TPC yield.

Level	Particle size (mm)	Time (min)	Solid-to-solvent ratio (g/mL)	Temperature (°C)
1	11.101	5.694	8.975	5.123
2	8.386	7.557	8.021	7.432
3	6.309	7.882	7.021	7.792
4	4.290	8.955	6.071	9.741
Delta	6.811	3.261	2.904	4.618
Rank	1	3	4	2

Table 4

ANOVA.

Source	DF[a]	Adj SS[b]	Adj MS[c]	F-Value	P-Value
Regression	4	172.89	43.222	24.95	0
Particle size	1	94.6	94.599	54.6	0
Time	1	20.43	20.43	11.79	0.006
Solid-to-liquid ratio	1	17.46	17.457	10.08	0.009
Temperature	1	40.4	40.402	23.32	0.001
Error	11	19.06	1.733		
Total	15	191.95			

[a] Degree of freedom.

[b] adjusted sums of squares.

[c] adjusted mean squares.

4.2. Effect of factors on TPC yield

The graphical representation of TPC yield acquired from Taguchi L_{16} experimental design runs is presented in Figs. 3–6. The average TPC yield was computed for all four factors at specified levels.

Fig. 3 illustrates the effect of particle size on TPC yield. Notably, decreasing particle size was associated with an increase in TPC yield. The Taguchi optimization study revealed the highest TPC yield of 11.101 mg GAE/g of plant material for a particle size of 0.15 mm. In contrast, a particle size of 0.6 mm yielded 4.29 mg GAE/g of plant material. Diminishing particle size facilitated greater surface interaction between the solid and solvent. The reduced particle size minimized the diffusion path between solute and solvent, consequently augmenting the mass transfer rate [32,33]. The reduction in plant material particle size directly exposed a higher number of plant cells to the solvent [34]. Bucic-Kojic et al. [35] examined the impact of particle size on total polyphenol kinetics in grape seeds and discovered that a smaller particle size range yielded the highest total polyphenol concentration. Our findings align with this prior research. Notably, the literature on neem leaves lacks mention of a lower particle size limit.

As temperature ascended from 25 to 55 °C, the TPC yield exhibited an upward trajectory, with maximum yield attained at 55 °C (Fig. 4). The elevated TPC yield can be attributed to increased solute solubility and diffusivity at higher temperatures, resulting from decreased solution viscosity and consequent acceleration of the extraction process [36]. Temperature elevation induced softening and swelling of plant material [37], intensifying the mass transfer of total phenolics from *Azadirchata indica* leaves and thus elevating the TPC yield. Notably, the study of solid-liquid extraction of TPC from grape marc showed an extraction rate rise in the temperature range of 25–60 °C [38]. Similarly, anthocyanin extraction from milled berries exhibited increased yields with temperature elevation from 6 to 30 °C [32]. However, elevated temperature values led to solvent evaporation and thermal degradation of thermolabile phenolic compounds [39]. Consequently, the optimal temperature was determined as 55 °C due to the maximum TPC yield observed at this point.

Fig. 5 portrays the effect of extraction time on TPC yield, indicating that prolonged time positively influenced TPC yield. The TPC yield escalated from 5.69 to 8.95 mg GAE/g of plant material with extended extraction time. The peak yield of 8.955 mg GAE/g of plant material was achieved after 60 min of extraction. A comparable positive time effect emerged in the batch extraction of phenolic compounds from neem leaves [40]. The extraction process initially exhibited faster extraction rates owing to the higher concentration gradient between solvent and solid material. As the concentration gradient diminished and equilibrium approached, the extraction process decelerated.

The affirmative impact of the solid-to-solvent ratio on TPC yield is evident in Fig. 6. The maximum yield of TPC was achieved at a ratio of 1:50 g/mL. A higher solid-to-solvent ratio was found to amplify the concentration gradient, propelling the transfer of phenolic compounds from neem leaves into the solvent during extraction [40]. Comparable findings were noted in the study of solid-liquid batch extraction of phenolic compounds from neem leaves, where an increase in the solid-to-solvent ratio from 1:10 to 1:50 g/mL led to enhanced extraction of total phenolic compounds [23]. Sivarajan et al. [41] also observed an increase in the percentage extraction yield of phenolic content from two spice varieties with a solid-to-solvent ratio escalation from 1:20 to 1:50 g/mL.

The impact of factors on TPC yield was validated through factor ranking. As per Table 3, particle size exerted the most substantial influence on TPC yield, followed by temperature, time, and solid-to-liquid ratio.

Table 5
GA-optimized values for maximum TPC yield.

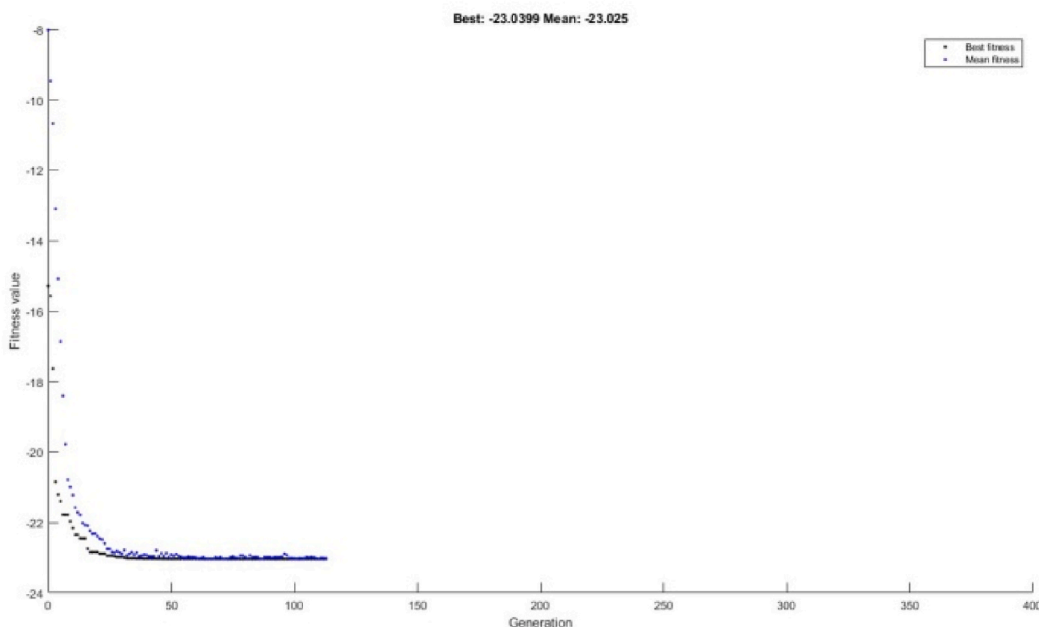| Particle size (mm)	Time (min)	Solid-to-solvent Ratio (g/mL)	Temperature (°C)	TPC yield (mgGAE/g)
0.15	40	1:25	55	23.0399

Fig. 7. Fitness value vs. generation.

4.3. Multiple regression analysis

The regression equation obtained in MINITAB Statistical Software by utilizing the experimental data of the Taguchi L_{16} array is provided by Equation (10). The ANOVA results (Table 4) indicate that all process variables, including particle size, time, solid-to-liquid ratio, and temperature, possess a P-Value of less than 0.05, significantly affecting TPC yield with a 95 % confidence level.

$$\text{TPC Yield} = 21.7 + 47.2*\text{Particle Size} - 0.547*\text{Time} + 329*\text{Solid to Liquid ratio} - 0.796*\text{Temperature} - 0.319*\text{Particle Size}$$
$$*\text{Time} - 84*\text{Particle Size}*\text{Solid to Liquid ratio} - 0.093*\text{Particle Size}*\text{Temperature} + 8.43*\text{Time}$$
$$*\text{Solid to Liquid ratio} + 0.01289*\text{Time}*\text{Temperature} + 0.61*\text{Solid to Liquid ratio}$$
$$*\text{Temperature} - 65.2*\text{Particle Size}^2 - 0.00072*\text{Time}^2 - 9641*\text{Solid to Liquid ratio}^2 + 0.00563*\text{Temperature}^2 \qquad (10)$$

4.4. Genetic algorithm

In this study, the GA is implemented in MATLAB R2022b software, and its specifics are as follows:

- Population size = 50
- Elite count = 2.5
- Population type = double vector
- Cross over fraction = 0.8
- Max Generations = 400
- Max stall generations = 50
- Function tolerance = 1×10^{-6}

The regression equation serves as the objective function for optimization in GA, and the constraints are expressed in Equations (11)–(14). The regression equation is negated and considered an objective function since it involves a maximization problem.

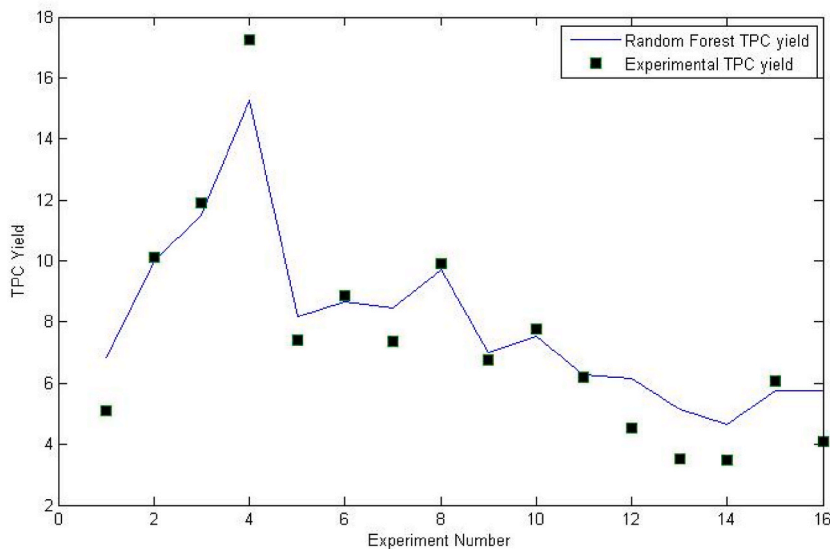$$0.15 \leq \text{Particle size} \leq 0.6 \qquad (11)$$

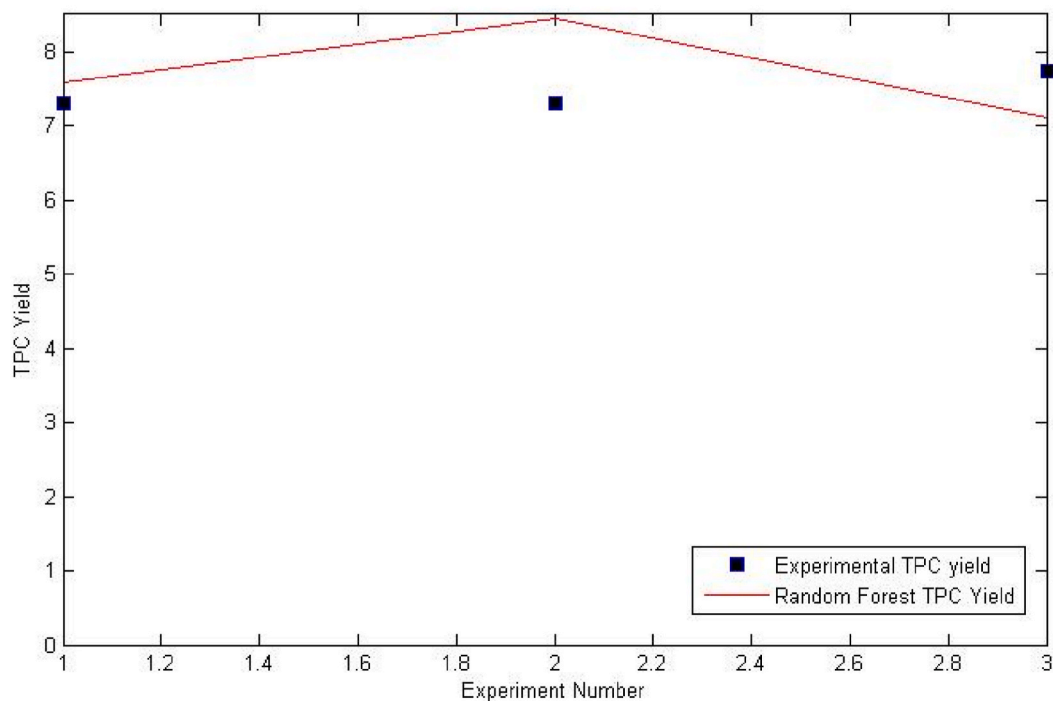**Fig. 8.** Comparison plot of experimental and random forest values of TPC yield for training data set.

Fig. 9. Comparison plot of experimental and random forest values of TPC yield for the test data set.

$15 \leq$ Time ≤ 60 (12)

$1{:}20 \leq$ Solid-to-liquid ratio $\leq 1{:}50$ (13)

$25 \leq$ Temperature ≤ 55 (14)

The GA-optimized values of particle size, time, solid-to-liquid ratio, temperature, and their corresponding maximum TPC yield are provided in Table 5.

The variation of fitness values across the generations is depicted in Fig. 7, showing stabilization at lower Generation values.

Upon substitution into the regression equation, the GA-optimized parameters yielded a TPC value of 23.03992 mg GAE/g, which closely aligns with the GA TPC yield value of 23.03990 mg GAE/g.

10
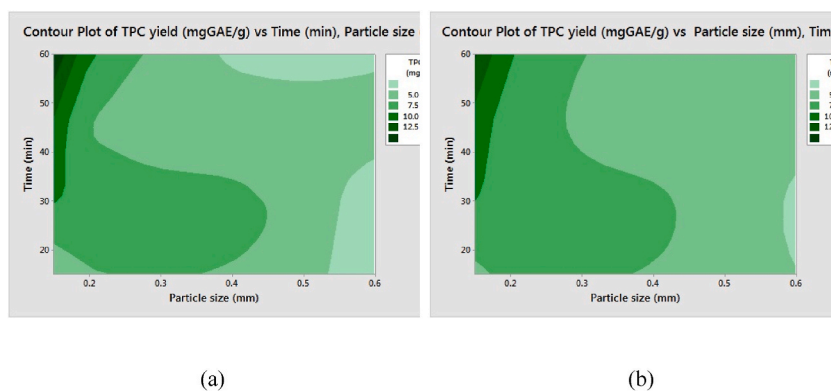
Fig. 10. Contour plot of experimental TPC yield vs. particle size and time: (a) experimental (b) random forest.
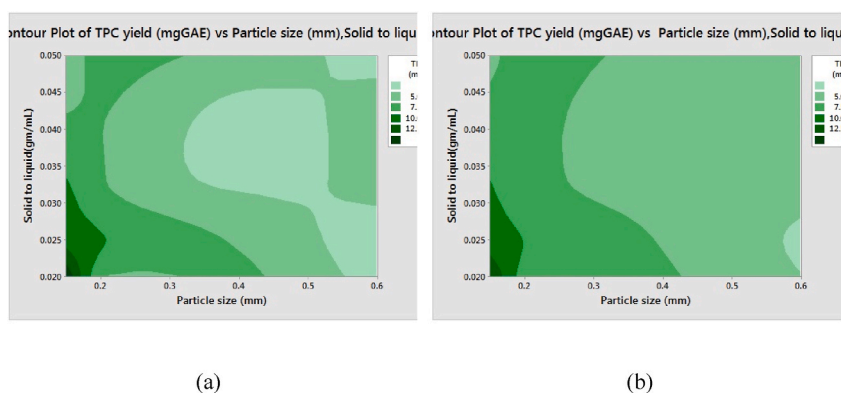

Fig. 11. Contour plot of experimental TPC yield vs. particle size and solid-to-solvent ratio: (a) experimental (b) random forest.
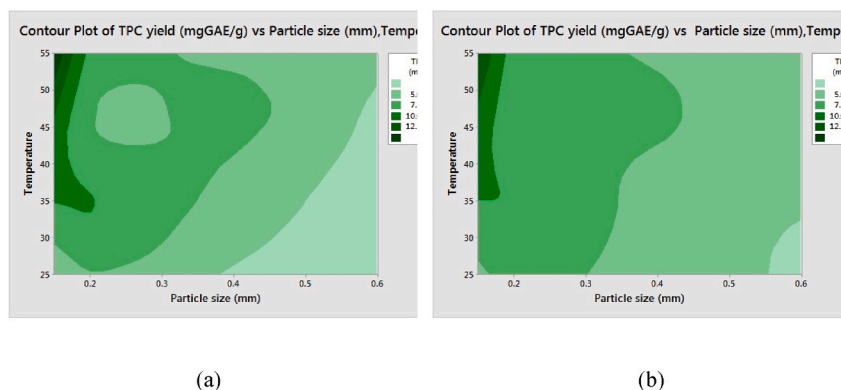

Fig. 12. Contour plot of experimental TPC yield vs. particle size and temperature: (a) experimental (b) random forest.

4.5. Random forest method

The random forest algorithm is implemented in Python's Jupyter Notebook. A total of 100 decision trees were employed in this study. The experimental data from Taguchi's L_{16} array were utilized to train the random forest machine learning model. Comparison plots of experimental and random forest values of TPC yield for the training and test data sets are shown in Figs. 8 and 9, respectively.

It is evident from Figs. 8 and 9 that the random forest values of TPC yield closely align with the experimental values for both the training and test data sets. The average and maximum errors in the SVR random forest estimated values of TPC yield for the training data set are 0.16 % and 0.289 %, respectively. The maximum and average errors between the random forest–predicted TPC yield values for the test data set are 3.1 % and 2.98 %, respectively.
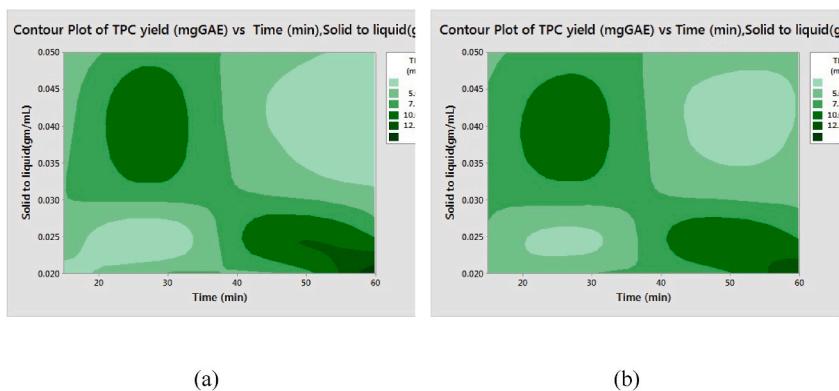
(a)

(b)

Fig. 13. Contour plot of experimental TPC yield vs. time and solid-to-solvent ratio: (a) experimental (b) random forest.
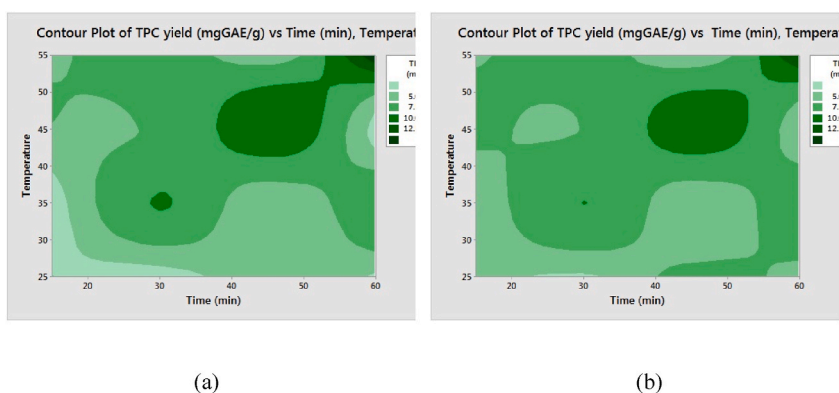

(a)

(b)

Fig. 14. Contour plot of experimental TPC yield vs. time and temperature: (a) experimental (b) random forest.
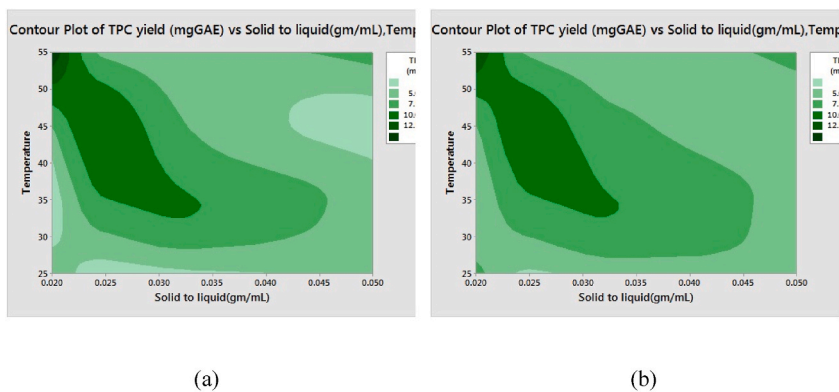

(a)

(b)

Fig. 15. Contour plot of experimental TPC yield vs. solid-to-solvent ratio and temperature: (a) experimental (b) random forest.

Figs. 10(b) to 15(b) reveal that the contours of random forest TPC yield for various cases do not align perfectly with the experimental contours as shown in Figs. 10(a) to 15(a).

4.6. Support vector regression

The SVM algorithm is implemented in Python's Jupyter Notebook, considering the following SVR model hyperparameters:

- Penalty Parameter (C): 10,000
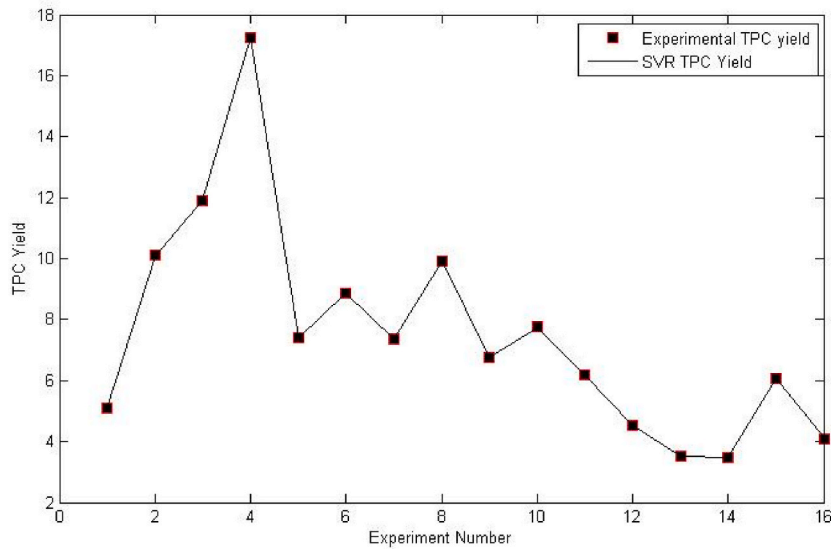- Kernel: Radial Basis Function

Fig. 16. Comparison plot of experimental and SVR values of TPC yield for training data set.
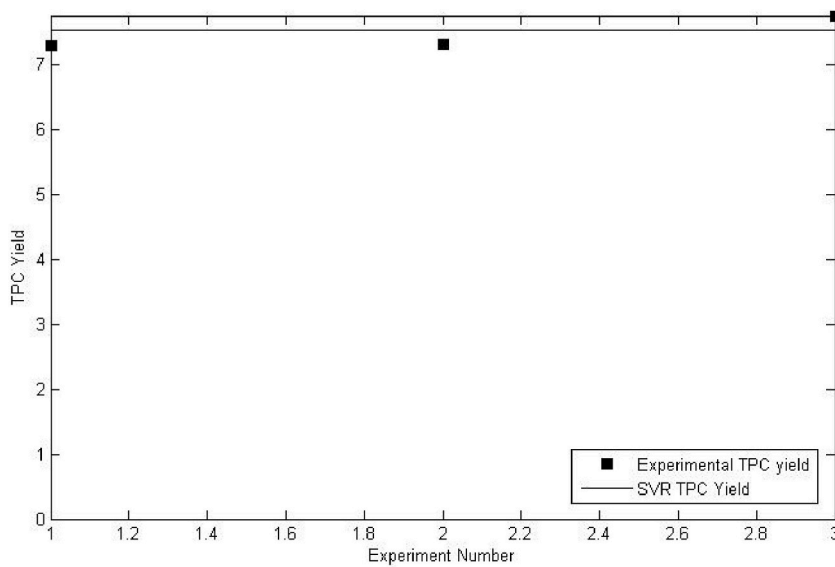

Fig. 17. Comparison plot of experimental and SVR values of TPC yield for the test data set.

• Gamma: 0.5
• Epsilon: 0.07

The experimental data from Taguchi's L_{16} array were used to train the SVR machine-learning model. A comparison of the experimental and SVR values of TPC yield for both the training and test data sets is depicted in Figs. 16 and 17, respectively.

Figs. 16 and 17 illustrate that the SVR TPC yield values closely align with the experimental values for both training and test data sets. The average and maximum errors in the SVR-predicted TPC yield values for the training data set are 0.16 % and 0.289 %, respectively. Likewise, the maximum and average errors between the SVR-estimated TPC yield values for the test data set are 3.1 % and 2.98 %, respectively.

From the SVR contour in Fig. 18 (b), the maximum TPC is obtained when the particle size is minimized and time is maximized. The SVR contour in Fig. 19 (b) indicates that the minimum particle size with the minimum solid-to-liquid ratio yields maximum TPC. In Fig. 20 (b), the SVR contour suggests that the combination of minimum particle size and maximum temperature results in maximum TPC. The plot in Fig. 21 (b) shows that the maximum TPC is achieved with a combination of maximum time and minimum solid-to-liquid ratio. According to the SVR contour in Fig. 22 (b), the maximum TPC can be attained by maintaining both the maximum time
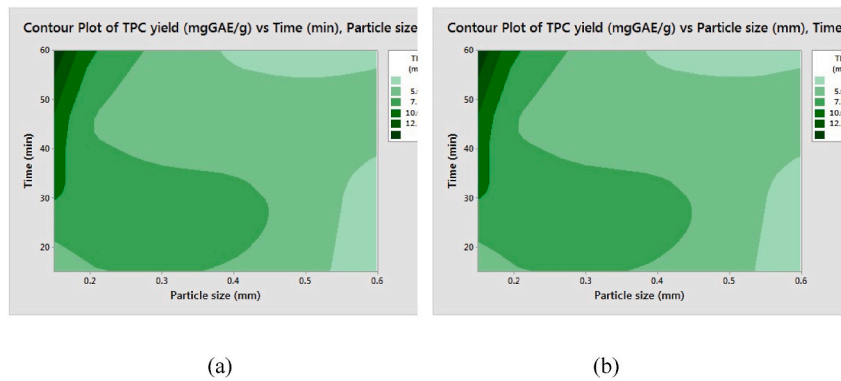
(a)   (b)

Fig. 18. Contour plot of experimental TPC yield vs. particle size and time: (a) experimental (b) SVR.
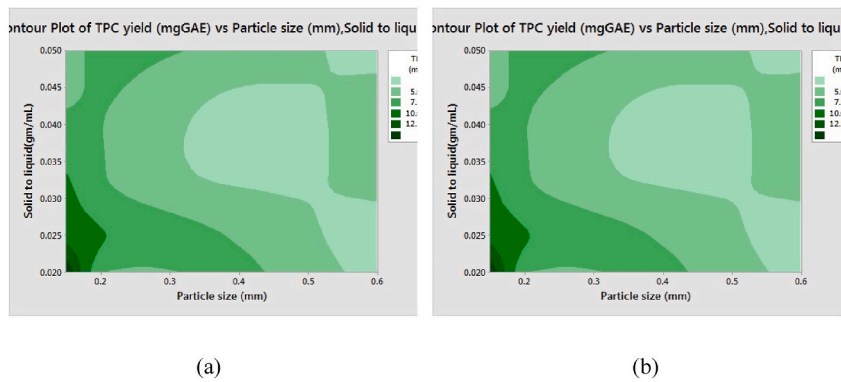

(a)   (b)

Fig. 19. Contour plot of experimental TPC yield vs. particle size and solid-to-solvent ratio: (a) experimental (b) SVR.
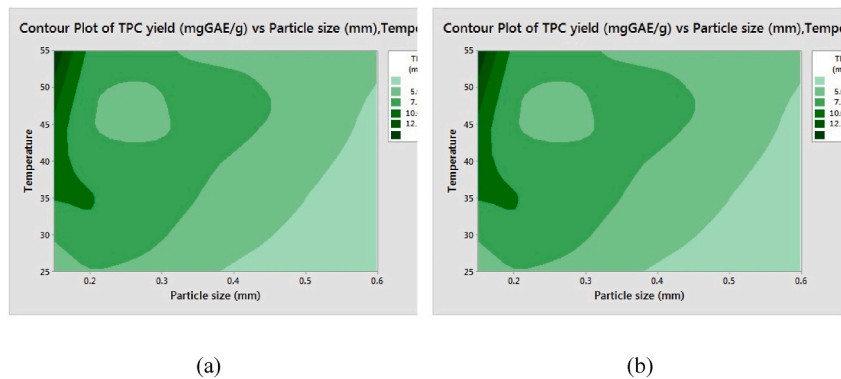

(a)   (b)

Fig. 20. Contour plot of experimental TPC yield vs. particle size and temperature: (a) experimental (b) SVR.

and temperature. In Fig. 23 (b), the SVR contour demonstrates that the combination of maximum temperature and minimum solid-to-liquid ratio produces maximum TPC. Overall, Figs. 18(b) to 23(b) illustrate that the SVR TPC yield contours for different cases match the experimental contours shown in Figs. 18(a) to 23(a).

5. Conclusions

The present study on solid-liquid batch extraction aimed to extract TPC from neem leaves using the Taguchi L_{16} experimental design. For this process, methanol was identified as the optimal solvent among the options considered. The ranking of process factors indicated that the particle size of neem leaves holds the greatest influence in the said process. The GA predicted the optimum process factors: 0.15 mm particle size, 40 min extraction time, 1:25 g/mL solid-to-solvent ratio, and 55 °C temperature. Under these predicted
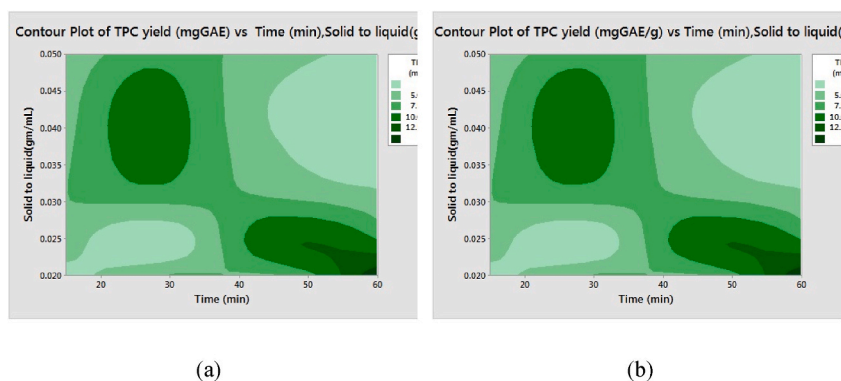
Fig. 21. Contour plot of experimental TPC yield vs. time and solid-to-solvent ratio: (a) experimental (b) SVR.
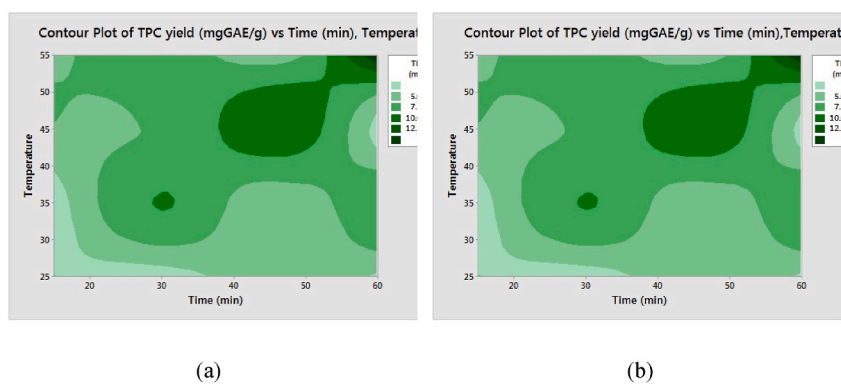

Fig. 22. Contour plot of experimental TPC yield vs. time and temperature: (a) experimental (b) SVR.
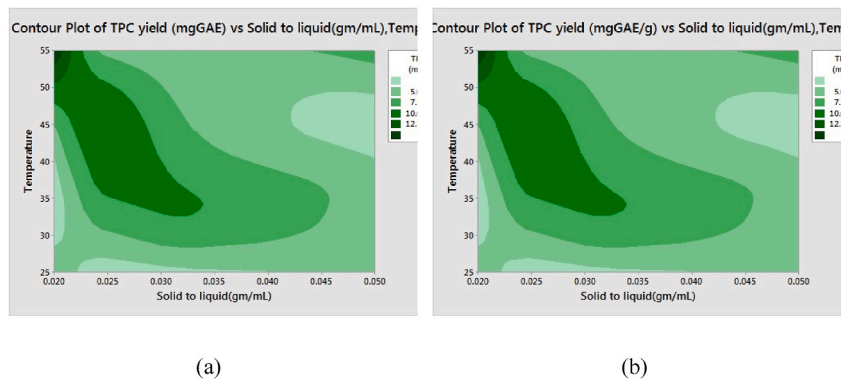

Fig. 23. Contour plot of experimental TPC yield vs. solid-to-solvent ratio and temperature: (a) experimental (b) SVR.

extraction conditions, the TPC yielded 23.039 mg GAE/g. The employed machine learning techniques demonstrated that SVR values of TPC yield exhibited a strong correlation with experimental values for both training and test data sets, as compared to the values from the RFM.

Funding

Data availability

Data included in article/supp. material/referenced in article.

CRediT authorship contribution statement

Sunita S. Patil: Investigation, Writing – original draft, Data curation, Methodology. **Umesh B. Deshannavar:** Conceptualization, Writing – review & editing, Supervision. **Shambala N. Gadekar-Shinde:** Conceptualization. **Amith H. Gadagi:** Software, Validation. **Santosh A. Kadapure:** Conceptualization, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Dr. Umesh B. Deshannavar reports was provided by DR MS Sheshgiri College of Engineering and Technology. Dr. Umesh B. Deshannavar reports a relationship with DR MS Sheshgiri College of Engineering and Technology that includes: employment.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e21991.

References

[1] M. Gallo, A. Formato, R. Giacco, G. Riccardi, D. Luongo, G. Formato, A. Amoresano, D. Naviglio, Mathematical optimization of the green extraction of polyphenols from grape peels through a cyclic pressurization process, Heliyon 5 (2019), e01526, https://doi.org/10.1016/j.heliyon.2019.e01526.
[2] O. Adeyi, E.O. Oke, B.I. Okolo, A.J. Adeyi, J.A. Otolorin, K. Nwosu-Obieogu, J.A. Adeyanju, G.W. Dzarma, S. Okhale, D. Ogu, P.N. Onu, Process optimization, scale-up studies, economic analysis and risk assessment of phenolic rich bioactive extracts production from Carica papaya L. leaves via heat-assisted extraction technology, Heliyon 8 (2022), e09216, https://doi.org/10.1016/j.heliyon.2022.e09216.
[3] Z. Tingting, Z. Xiuli, W. Kun, S. Liping, Z. Yongliang, A review: extraction, phytochemicals, and biological activities of rambutan (Nephelium lappaceum L) peel extract, Heliyon 8 (2022), e11314, https://doi.org/10.1016/j.heliyon.2022.e11314.
[4] X. Bi, Y. Hemar, M.O. Balaban, X. Liao, The effect of ultrasound on particle size, color, viscosity and polyphenol oxidase activity of diluted avocado puree, Ultrason. Sonochem. 27 (2015) 567–575, https://doi.org/10.1016/j.ultsonch.2015.04.011.
[5] P. Skenderidis, S. Leontopoulos, K. Petrotos, I. Giavasis, Optimization of vacuum microwave-assisted extraction of pomegranate fruits peels by the evaluation of extracts' phenolic content and antioxidant activity, Foods 9 (2020) 1655, https://doi.org/10.3390/foods9111655.
[6] R. Rojas, J.C. Contreras-Esquivel, M.T. Orozco-Esquivel, C. Muñoz, J.A. Aguirre-Joya, C.N. Aguilar, Mango peel as source of antioxidants and pectin: microwave assisted extraction, Waste and Biomass Valorization 6 (2015) 1095–1102, https://doi.org/10.1007/s12649-015-9401-4.
[7] B. Xu, M. Feng, E.S. Tiliwa, W. Yan, B. Wei, C. Zhou, H. Ma, B. Wang, L. Chang, Multi-frequency power ultrasound green extraction of polyphenols from Pingyin rose: optimization using the response surface methodology and exploration of the underlying mechanism, LWT 156 (2022), 113037, https://doi.org/10.1016/j.lwt.2021.113037.
[8] N. Anis, D. Ahmed, Modelling and optimization of polyphenol and antioxidant extraction from Rumex hastatus by green glycerol-water solvent according to response surface methodology, Heliyon 8 (2022), e11992, https://doi.org/10.1016/j.heliyon.2022.e11992.
[9] S. Akbari, N.H. Abdurahman, R.M. Yunus, O.R. Alara, O.O. Abayomi, Extraction, characterization and antioxidant activity of fenugreek (Trigonella-Foenum Graecum) seed oil, Mater. Sci. Energy Technol. 2 (2019) 349–355, https://doi.org/10.1016/j.mset.2018.12.001.
[10] D.B. Muñiz-Márquez, G.C. Martínez-Ávila, J.E. Wong-Paz, R. Belmares-Cerda, R. Rodríguez-Herrera, C.N. Aguilar, Ultrasound-assisted extraction of phenolic compounds from Laurus nobilis L. and their antioxidant activity, Ultrason. Sonochem. 20 (2013) 1149–1154, https://doi.org/10.1016/j.ultsonch.2013.02.008.
[11] M. Milutinović, N. Radovanović, M. Ćorović, S. Šiler-Marinković, M. Rajilić-Stojanović, S. Dimitrijević-Branković, Optimization of microwave-assisted extraction parameters for antioxidants from waste Achillea millefolium dust, Ind. Crops Prod. 77 (2015) 333–341, https://doi.org/10.1016/j.indcrop.2015.09.007.
[12] X. Pan, G. Niu, H. Liu, Microwave-assisted extraction of tea polyphenols and tea caffeine from green tea leaves, Chem. Eng. Process. Process Intensif. 42 (2003) 129–133, https://doi.org/10.1016/S0255-2701(02)00037-5.
[13] M. Bouras, M. Chadni, F.J. Barba, N. Grimi, O. Bals, E. Vorobiev, Optimization of microwave-assisted extraction of polyphenols from Quercus bark, Ind. Crops Prod. 77 (2015) 590–601, https://doi.org/10.1016/j.indcrop.2015.09.018.
[14] W. Xiaokang, J.G. Lyng, N.P. Brunton, L. Cody, J.C. Jacquier, S.M. Harrison, K. Papoutsis, Monitoring the effect of different microwave extraction parameters on the recovery of polyphenols from shiitake mushrooms: comparison with hot-water and organic-solvent extractions, Biotechnol. Reports 27 (2020), e00504, https://doi.org/10.1016/j.btre.2020.e00504.
[15] Y. Zhang, H. Tang, Y. Zheng, J. Li, L. Pan, Optimization of ultrasound-assisted extraction of poly-phenols from Ajuga ciliata Bunge and evaluation of antioxidant activities in vitro, Heliyon 5 (2019), e02733, https://doi.org/10.1016/j.heliyon.2019.e02733.
[16] F. Dranca, M. Oroian, Optimization of ultrasound-assisted extraction of total monomeric anthocyanin (TMA) and total phenolic content (TPC) from eggplant (Solanum melongena L.) peel, Ultrason. Sonochem. 31 (2016) 637–646, https://doi.org/10.1016/j.ultsonch.2015.11.008.
[17] A. Sridhar, M. Ponnuchamy, P.S. Kumar, A. Kapoor, D.V.N. Vo, S. Prabhakar, Techniques and modeling of polyphenol extraction from food: a review, Environ. Chem. Lett. 19 (2021) 3409–3443, https://doi.org/10.1007/s10311-021-01217-8.
[18] D. Jadhav, B.N. R, P.R. Gogate, V.K. Rathod, Extraction of vanillin from vanilla pods: a comparison study of conventional soxhlet and ultrasound assisted extraction, J. Food Eng. 93 (2009) 421–426, https://doi.org/10.1016/j.jfoodeng.2009.02.007.
[19] S. Tandon, N.K. Sand, Qualitative analysis of phenolic constituents from leaves of some plants of family meliaceae, Int. J. Med. Plants Nat. Prod. 2 (2016), https://doi.org/10.20431/2454-7999.0201005.
[20] G. Nahak, R. Sahu, In vitro antioxidative activity of Azadirachta indica and Melia azedarach Leaves by DPPH scavenging assay, Nat. Sci. 8 (2010) 22–28.
[21] M.A. Hossain, W.A.S. Al-Toubi, A.M. Weli, Q.A. Al-Riyami, J.N. Al-Sabahi, Identification and characterization of chemical compounds in different crude extracts from leaves of Omani neem, J. Taibah Univ. Sci. 7 (2013) 181–188, https://doi.org/10.1016/j.jtusci.2013.05.003.
[22] H.S.K. Al-Jadidi, M.A. Hossain, Determination of the total phenols, flavonoids and antimicrobial activity of the crude extracts from locally grown neem stems, Asian Pacific J. Trop. Dis. 6 (2016) 376–379, https://doi.org/10.1016/S2222-1808(15)61051-9.

[23] S. Shewale, V.K. Rathod, Extraction of total phenolic content from Azadirachta indica or (neem) leaves: kinetics study, Prep. Biochem. Biotechnol. 48 (2018) 312–320, https://doi.org/10.1080/10826068.2018.1431784.

[24] G.R. Sinha (Ed.), Modern Optimization Methods for Science, Engineering and Technology, IOP Publishing, 2019, https://doi.org/10.1088/978-0-7503-2404-5.

[25] J. Azmir, I.S.M. Zaidul, M.M. Rahman, K.M. Sharif, A. Mohamed, F. Sahena, M.H.A. Jahurul, K. Ghafoor, N.A.N. Norulaini, A.K.M. Omar, Techniques for extraction of bioactive compounds from plant materials: a review, J. Food Eng. 117 (2013) 426–436, https://doi.org/10.1016/j.jfoodeng.2013.01.014.

[26] J.G. Gujar, S. Chattopadhyay, S.J. Wagh, V.G. Gaikar, Experimental and modeling studies on extraction of catechin hydrate and epicatechin from Indian green tea leaves, Can. J. Chem. Eng. (2010), https://doi.org/10.1002/cjce.20271 n/a-n/a.

[27] V.L. Singleton, R. Orthofer, R.M. Lamuela-Raventós, Analysis of Total Phenols and Other Oxidation Substrates and Antioxidants by Means of Folin-Ciocalteu Reagent, 1999, pp. 152–178, https://doi.org/10.1016/S0076-6879(99)99017-1.

[28] R. Davis, P. John, Application of taguchi-based design of experiments for industrial chemical processes, in: Stat. Approaches with Emphas. Des. Exp. Appl. To Chem. Process., InTech, 2018, https://doi.org/10.5772/intechopen.69501.

[29] E.L.C. Cheah, P.W.S. Heng, L.W. Chan, Optimization of supercritical fluid extraction and pressurized liquid extraction of active principles from Magnolia officinalis using the Taguchi design, Sep. Purif. Technol. 71 (2010) 293–301, https://doi.org/10.1016/j.seppur.2009.12.009.

[30] A.A. Abdul Hamed, M.A. Tawfeek, A.E. Keshk, A genetic algorithm for service flow management with budget constraint in heterogeneous computing, Futur. Comput. Informatics J. 3 (2018) 341–347, https://doi.org/10.1016/j.fcij.2018.10.004.

[31] S.V. Taralkar, S. Chattopadhyay, V.G. Gaikar, Parametric optimization and modeling of batch extraction process for extraction of betulinic acid from leaves of Vitex Negundo Linn, Sep. Sci. Technol. 51 (2016) 641–652, https://doi.org/10.1080/01496395.2015.1105822.

[32] J.E. Cacace, G. Mazza, Mass transfer process during extraction of phenolic compounds from milled berries, J. Food Eng. 59 (2003) 379–389, https://doi.org/10.1016/S0260-8774(02)00497-1.

[33] M. Pinelo, J. Sineiro, M.J. Núñez, Mass transfer during continuous solid-liquid extraction of antioxidants from grape byproducts, J. Food Eng. 77 (2006) 57–63, https://doi.org/10.1016/j.jfoodeng.2005.06.021.

[34] M. Vinatoru, An overview of the ultrasonically assisted extraction of bioactive principles from herbs, Ultrason. Sonochem. 8 (2001) 303–313, https://doi.org/10.1016/S1350-4177(01)00071-2.

[35] A. Bucić-Kojić, M. Planinić, S. Tomas, M. Bilić, D. Velić, Study of solid-liquid extraction kinetics of total polyphenols from grape seeds, J. Food Eng. 81 (2007) 236–242, https://doi.org/10.1016/j.jfoodeng.2006.10.027.

[36] L. Lazar, A.I. Talmaciu, I. Volf, V.I. Popa, Kinetic modeling of the ultrasound-assisted extraction of polyphenols from Picea abies bark, Ultrason. Sonochem. 32 (2016) 191–197, https://doi.org/10.1016/j.ultsonch.2016.03.009.

[37] Y. Tao, Z. Zhang, D.-W. Sun, Kinetic modeling of ultrasound-assisted extraction of phenolic compounds from grape marc: influence of acoustic energy density and temperature, Ultrason. Sonochem. 21 (2014) 1461–1469, https://doi.org/10.1016/j.ultsonch.2014.01.029.

[38] V. Sant' Anna, A. Brandelli, L.D.F. Marczak, I.C. Tessaro, Kinetic modeling of total polyphenol extraction from grape marc and characterization of the extracts, Sep. Purif. Technol. 100 (2012) 82–87, https://doi.org/10.1016/j.seppur.2012.09.004.

[39] A.M. Goula, Ultrasound-assisted extraction of pomegranate seed oil – kinetic modeling, J. Food Eng. 117 (2013) 492–498, https://doi.org/10.1016/j.jfoodeng.2012.10.009.

[40] I. Hismath, W. Mustapha, C. Ho, Optimization of extraction conditions for phenolic compounds from neem (Azadirachta indica) leaves, Int. Food Res. J. 18 (2011).

[41] K. Radha krishnan, M. Sivarajan, S. Babuskin, G. Archana, P. Azhagu Saravana Babu, M. Sukumar, Kinetic modeling of spice extraction from S. aromaticum and C. cassia, J. Food Eng. 117 (2013) 326–332, https://doi.org/10.1016/j.jfoodeng.2013.03.011.