

Research Article

Optimization of Process Parameters for the Production of Soy Protein by Ultrafiltration Using ANN

Hima John ¹, Saroj Kumar Giri ², A. Subeesh ², Punit Chandra ²
and R. Pandiselvam ³

¹ICAR-National Dairy Research Institute, Karnal 132001, Haryana, India

²ICAR-Central Institute of Agricultural Engineering, Bhopal 462038, Madhya Pradesh, India

³Physiology, Biochemistry and Post-Harvest Technology Division, ICAR-Central Plantation Crops Research Institute (CPCRI), Kasaragod 671124, Kerala, India

Correspondence should be addressed to Hima John; himajohn013@gmail.com and R. Pandiselvam; r.pandiselvam@icar.gov.in

Received 12 August 2023; Revised 30 December 2023; Accepted 2 January 2024; Published 10 January 2024

Academic Editor: R. Mahendran

Copyright © 2024 Hima John et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The growing popularity of soy proteins among vegans and vegetarians, owing to their high protein content and widespread availability, has led to scientific studies on its various extraction methods mainly on ultrafiltration. This research employed artificial neural network (ANN) and Box-Behnken design (BBD) methodologies to predict the process parameters of ultrafiltration for the preparation of soy protein. Using BBD, the optimum process parameters of ultrafiltration were identified via the desirability function approach. The optimized permeate flux was 11.13 litres per hour (LPH) and 85.52% protein content in retentate. The identified ideal process parameters for ultrafiltration to achieve maximal protein retention encompassed a 10 kDa membrane module, a transmembrane pressure of 117 kPa (17 PSI), a volume concentration ratio of 3.5, diafiltration set at 1, and a flow rate of 65% of the pump capacity, exhibiting an absolute percent error value of 2.81. Employing these refined process parameters, the predicted value for protein retentate stood at 80.49%. The predictive accuracy of the model achieved an impressive 99.61% for protein retention. The ANN model effectively predicted the optimal ultrafiltration conditions, resulting in maximal protein retention and a protein content accuracy of 96.41% and 99.61%, respectively.

1. Practical Applications

Ultrafiltration process optimization for the production of various high-quality protein concentrates and isolates using artificial neural network can be done with high accuracy.

2. Introduction

Pressure-driven membrane separation processes, particularly ultrafiltration (UF), have gained widespread adoption for protein concentration and purification [1]. This technique emerges as a promising alternative to traditional acid precipitation methods due to its minimal pH shock, ambient temperature conditions, and superior membrane selectivity during purification and fractionation [2, 3]. Efficient protein

recovery hinges upon suitable membrane modules and operating parameters, influencing the effectiveness of UF in producing various protein forms such as concentrates and isolates. Aguero et al. [4] reported higher protein yield, better protein quality, and other functional attributes after using a membrane separation process. While UF exhibits significant potential, membrane fouling remains a primary concern impacting its performance, necessitating prefiltration strategies. As membrane separation methods represent physical separation techniques, there is an anticipated surge in their applications in the near future. A number of researchers [5–8] have identified the scope of membrane-based techniques, especially ultrafiltration and microfiltration for concentrating and purifying soy proteins. The relatively lower quality of commercially available soy proteins

derived from conventional methods underscores the opportunity for utilizing appropriate ultrafiltration membrane modules to enhance their quality.

Numerous studies have explored the UF process for the preparation of various soy protein fractions under varied conditions [9–11]. Prior research has demonstrated the efficacy of BBD in optimizing soy protein extraction [12, 13]. Similarly, ANN has emerged as a powerful tool for understanding complex systems due to its accuracy and computational efficiency [14, 15]. Its applications in membrane separation processes, including microfiltration and ultrafiltration, have showcased predictive capabilities for flux, rejection, and separation efficiency in various domains [16–19]. In dairy processing, ANNs have been used as an efficient method for modelling and simulating the ultrafiltration of milk in cross-flow mode [20–22]. Park et al. [23] used ANN for predicting the fouling behaviour of ultrafiltration in pilot scale operations. ANN was used to model the flux decline during the ultrafiltration of whey by Gaudio et al. [24]. Commercially available soy proteins are prepared by conventional acid precipitation methods. The overall quality and functional properties of soy proteins are adversely affected by pH shock and other processing conditions of acid precipitation methods [25], and these lead to the scope of using suitable ultrafiltration membrane modules for producing high-quality soy proteins. Less number of investigations has been carried out to maintain the low molecular weight bioactive components also with soy proteins using low molecular weight cutoff hollow fibre ultrafiltration membrane modules.

Recent studies have reported the applications of ANN and RSM techniques as different process optimization tools in food processing [26–28]. Despite these advancements, there is limited literature on comparing the optimization of soy protein production via ultrafiltration using ANN and BBD. This study seeks to bridge this gap by employing both methods, thereby offering a comprehensive comparison of their optimization results for soy protein ultrafiltration.

3. Materials and Methods

3.1. Extraction of Protein from Defatted Soy Flour. For isolation of protein, soybean (variety JS335) was procured from the research farm of ICAR-Central Institute of Agricultural Engineering, Bhopal, Madhya Pradesh (77024°10′E longitude and 23018°35′N latitude), and defatting was done using Sox plus (Soxtron, Tulin-6 number) apparatus [29]. The subsequent extraction of protein from the defatted soy flour took place in purified water at pH 9 (adjusted with 0.2 M NaOH) and at 50°C with a solid/liquid ratio 1/10, using a mechanical stirrer (Jyoti, model JSI-555, India) for one and half hours. Solid-liquid separation was performed in a centrifuge (Remi instruments-model K-70, India) at 10,000 g for 20 minutes at 15°C temperature. The resultant supernatant was then utilized as the feed for the subsequent ultrafiltration process following the methodology outlined by John and Sinha [10].

3.2. Production of Soy Protein Isolate Using Ultrafiltration. Following the centrifugation process, the supernatant, constituting the protein extract, underwent prefiltration to eliminate particles prone to causing fouling in the subsequent ultrafiltration phase. A Millipore microfiltration unit employing a cellulose nitrate membrane, 47 mm in diameter with a 5.0 μ pore size, was utilized for this purpose. The ultrafiltration study was conducted utilizing a laboratory-scale GE Healthcare hollow fibre module, boasting a surface area of 650 cm². The schematic depiction of the soy protein preparation process from defatted soy flour (DSF) via ultrafiltration is delineated in Figure 1. Each trial utilized 75 g of defatted soy flour (DSF), resulting in the acquisition of 550 ml of extract. Postcentrifugation, the supernatant was subjected to prefiltration prior to its entry into the ultrafiltration unit. Approximately 375 ml of retentate, the component of interest, was collected from the process, while the permeate was discarded, indicating successful concentration of the target compounds.

3.2.1. Permeate Flux. Flux is defined as the permeate flow per unit membrane area in unit time. It is expressed in litres per square meter of membrane surface area per hour.

$$\text{Permeate flux, } J(\text{LMH}) = \frac{\text{permeate flow, } Q_p}{\text{surface area of membrane, } A} \quad (1)$$

Surface area of hollow fibre cartridge is 650 cm².

3.2.2. Protein Rejection. It is the percentage of protein removed from the feed stream by the membrane. If one is interested in the retentate part, it is the percentage of protein retained in the retentate part.

$$\text{Rejection, } R = 1 - \frac{\text{protein concentration in permeate}}{\text{protein concentration in retentate}} \quad (2)$$

3.3. Estimation of Protein. Protein estimation was done as per the method described by Ranganna [30]. About 0.5 g of the sample was weighed into the digestion tubes, and two heaped spatulas of digestion were added to each tube. 10 ml of concentrated H₂SO₄ was also added, and samples digested until the contents of the tubes were sea green in colour. The digested sample was transferred into the distillation chamber, and 20 ml of 4% boric acid was kept in the collecting conical flask. The boric acid turned from reddish pink to green as it collected the ammonia in the distillation chamber. Then, the green-coloured boric acid was titrated against 0.1 N HCl until its colour turned to pink. Then, the protein content of soy proteins was calculated using the given formula:

$$\text{Percent protein content} = \frac{(V_A - V_B) \times 0.1 \times 14.001 \times 6.25}{W} \quad (3)$$

where V_A is the sample titre value, V_B is the blank titre value, and W is the sample weight.

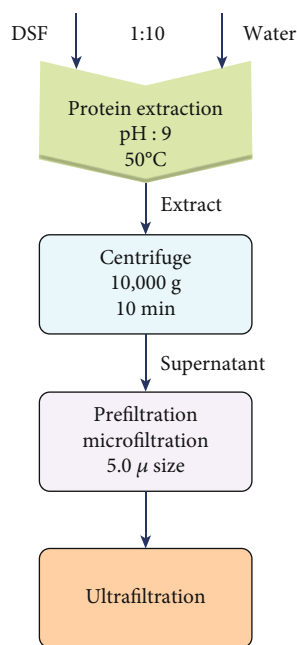


FIGURE 1: Preparation of soy protein from defatted soy flour by ultrafiltration.

3.4. Optimization of Process Parameters of Ultrafiltration for the Preparation of Soy Protein. Membrane modules, transmembrane pressure, diafiltration and volume concentration ratio, and flow rate were selected as the independent variables, and permeate flux, protein content of retentate, and protein retention percentage were the dependent parameters (response variables). Designing of experiments, fitting mathematical models, and optimization of variables were done with Design-Expert software 7.0.0 (trial version) using response surface methodology (BBD). Optimized process parameters were used for producing soy proteins by ultrafiltration in present study. Table 1 outlines the independent parameters alongside the corresponding levels designated for the experiments and subsequent analysis.

Responses can be represented as a function of independent variables.

$$\text{Responses, } Y = f(X_1, X_2, X_3, X_4, X_5). \quad (4)$$

Experimental data were analyzed to optimize the process parameters with respect to the responses. Regression analysis and analysis of variance were conducted to fit the model and to know the statistical significance of the selected model terms. Model adequacy was determined using model analysis, R^2 value, and lack of fit test. Model is adequate if the lack of fit is nonsignificant. R^2 value represents the ratio of the explained variation to total variation [31], and if R^2 is more than 80 percent, the model can be considered for further analysis [32]. Furthermore, the model's performance was assessed using various statistics including root mean square error (RMSE), sum of squares error (SSE), percent error, and chi-square.

TABLE 1: Independent parameters and levels chosen.

Independent parameters	Levels		
	-1	0	1
Ultrafiltration membrane modules (kDa) (X_1)	1	10	30
Transmembrane pressure (kPa) (X_2)	83	117	152
Volume concentration ratio (VCR) (X_3)	3	3.5	4
Diafiltration (batch) (X_4)	0	1	2
Flow rate (% pump capacity) (X_5)	55	65	75

Numerical optimization technique was used to optimize the responses simultaneously. The desired goal for each dependent and independent parameter was chosen. All independent parameters were kept in range, and permeate flux and protein retention percentage was maximized. Optimum solution was obtained based on the combined desirability value. Dependent variables or responses were assigned equal importance. Response surfaces were generated in Design-Expert software to understand the effect of independent parameters on dependent parameters or responses.

For conducting the experiments, 75 grams of defatted soy flour (average particle size 330μ) was taken and 550 ml extract was obtained after centrifugation. Total solid content of extract was recorded as $7 \pm 1\%$.

3.5. ANN Modelling. The BBD and the responses were used to develop the ANN model using Python 3.9.4. The basic ANN structure is given in Figure 2.

ANN was run to fit the regression model. Dependent variables were designated as the target variable, while independent variables served as predictors. Column abbreviations and details were provided under data description to understand the dataset better. The dataset was divided into training and testing sets with an 80:20 ratio to facilitate model training and testing. Essential libraries like "TensorFlow" and "Keras" were installed to implement deep learning ANNs in Python. Tuning of the ANN model involved searching for the best combination for optimal model performance. The "Sequential" module from the Keras library was used to create a sequence of ANN layers. The "Dense" module in Keras helped define each layer, specifying the number of neurons, initialization technique, weights in the network, and activation functions for each neuron. Different model architectures were experimented with by varying the number of layers and neurons to identify the best structure. Among the combinations tested, the model demonstrated superior performance with two hidden layers, each containing five neurons (Figure 2). Hence, two hidden layers with five neurons each and one-input and one-output layers were used for training. The "Dense" module of Keras was used to define each layer where the specification of the number of neurons, the technique to be used to initialize the weights in the network and the activation function for each neuron in that layer, etc. were defined. Specifications like batch size (5, 10, 15, and 20) and epochs (up to 50) were employed during model training. Hyperparameters were fine-tuned using a grid search approach to determine the most effective configuration. The model was trained using the best set of

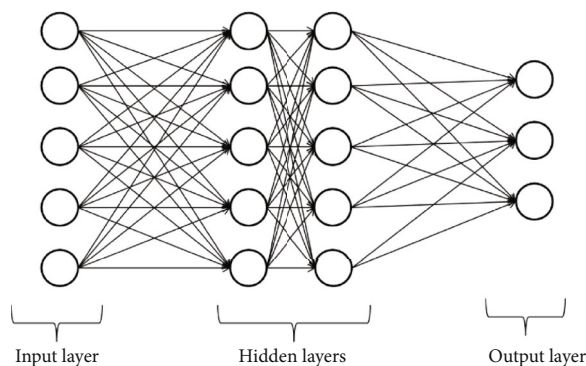


FIGURE 2: Basic ANN structure.

parameters identified earlier, and predictions were made on the testing data. Absolute percentage error was calculated for each row in the testing data. Mean absolute percentage error (MAPE) was computed as the average of all rows, and the ANN model accuracy was derived as $100 - \text{MAPE}$. The model's performance was assessed using various statistics including root mean square error (RMSE), sum of squares error (SSE), percent error, and chi-square.

4. Results and Discussion

4.1. Preparation of Soy Proteins Using Ultrafiltration. Table 2 showcases the proximate analysis results for both the defatted soy flour (DSF) and the soy protein obtained through ultrafiltration. DSF displayed a protein content of 56%, whereas the soy protein derived via ultrafiltration exhibited substantially higher protein content, recording at 88%. Moisture and fat content displayed comparable values between the DSF and the ultrafiltered soy protein. However, a notable disparity was observed in the ash content, with the ultrafiltered soy protein demonstrating a lower ash content in comparison to DSF. This divergence in ash content could potentially be attributed to the characteristics of the ultrafiltration process. Smaller molecular weight of ash particles relative to the selective retention by the ultrafiltration membrane led to a reduced concentration of ash in the resulting soy protein. This outcome aligns with findings from John and Sinha's research [10] and highlights a distinct alteration in the proximate composition, notably the increased protein content and a reduction in ash content, through the ultrafiltration process from DSF to soy protein.

4.2. Optimization of Process Parameters of Ultrafiltration Process Using Response Surface Methodology. Ultrafiltration process parameters were optimized using response surface methodology (Box-Behnken design). With 46 experiments, including 6 central point runs, the study sought to comprehend how independent parameters influenced dependent variables within the ultrafiltration process. Quadratic model was used to fit the data. The statistical significance of each model term was checked by regression analysis and analysis of variance (ANOVA). Table 3 depicts the regression coefficients and significance of each variable on membrane pro-

TABLE 2: Proximate analysis of the defatted soy flour (DSF) and the ultrafiltered soy protein.

Parameters	DSF	Ultrafiltered soy protein
Protein (%)	56	88 ± 0.3
Moisture content (% w.b.)	5.5 ± 0.1	5.7 ± 0.03
Fat (%)	0.45 ± 0.01	0.4 ± 0.02
Ash (%)	6.77 ± 0.11	4.4 ± 0.03

cess parameters. All the quadratic models were observed to be significant at $p < 0.006$, and lack of fit was nonsignificant.

4.2.1. Prediction of Permeate Flux by Box-Behnken Method.

The model has shown significant overall significance at an extremely low p value ($p < 0.0001$), indicating its reliability. Among the model terms, certain factors like ultrafiltration membrane modules, transmembrane pressure, interaction effect of ultrafiltration membrane modules, and flow rate, as well as square terms of ultrafiltration membrane modules, transmembrane pressure, and flow rate, have emerged as statistically significant. The "lack of fit F value" of 2.17 implies that the lack of fit is not significantly relative to the pure error. This suggests that the model accurately predicts the responses. The "Pred R-squared" of 0.8100, which denotes the predictive ability of the model, aligns reasonably well with the "Adj R-squared" of 0.9093. This indicates that a substantial portion of the variability in the data can be explained by the model. The coefficient of variation (C.V) calculated at 15.53% suggests a slightly lower precision due to a higher C.V value compared to the ideal threshold of 10%. A decrease in the root mean square error (RMSE), sum of squared errors (SSE), percent error, and χ^2 values (Table 4) indicates the enhanced appropriateness of employing this model for the prediction of permeate flux.

The ultrafiltration membrane module exhibits a significant effect on permeate flux ($p < 0.0001$). Initially, there was an increase in permeate flux with the rise in membrane module, followed by a subsequent decline. Optimal permeate flux is achieved within the range of 15.50 to 22.75 kDa (Figure 3). The highest permeate flux was observed with the 10 kDa membrane, surpassing both 1 kDa and 30 kDa membranes. This finding parallels Sagu et al.'s [33] study on banana juice ultrafiltration using hollow fibre membranes. They observed minimal pore blocking in the 10 kDa membrane, resulting in higher flux compared to higher cutoff membranes (27 and 44 kDa membranes). Transmembrane pressure also significantly impacts permeate flux ($p < 0.0031$), showcasing an increase in flux with rising pressure. In contrast, volume concentration ratio and flow rate did not show a significant impact on permeate flux. However, diafiltration exhibited a significant effect ($p < 0.0819$) on permeate flux. The interaction effect between the ultrafiltration membrane module and feed flow rate significantly affects permeate flux at a 5% level ($p < 0.0494$). These findings resonate with prior studies. Blonigen's [34] research on protein mixture ultrafiltration with a 30 kDa membrane revealed a similar curvature in permeate flux concerning transmembrane

TABLE 3: Regression coefficients and significance of each variable on membrane process parameters.

Source	Permeate flux		Protein content of retentate		Protein retention percentage	
	Coefficient	<i>p</i> value	Coefficient	<i>p</i> value	Coefficient	<i>p</i> value
Intercept	11.67	<0.0001	86.71	<0.0001	0.99	0.0060
X_1	2.63	<0.0001	5.50	<0.0001	0.00	<0.0001
X_2	0.95	0.0031	0.51	0.3124	0.00	0.4862
X_3	-0.39	0.1874	-0.50	0.3209	0.00	0.4045
X_4	0.52	0.0819	-0.75	0.1434	0.00	0.4996
X_5	-0.30	0.3078	0.32	0.5282	0.00	0.4361
X_1X_2	-0.71	0.1540	0.26	0.7593	0.00	0.8387
X_1X_3	-0.13	0.7949	0.04	0.9635	0.00	0.5930
X_1X_4	-0.16	0.7376	-0.43	0.6076	0.00	0.2977
X_1X_5	-1.00	0.0494	-0.37	0.6623	0.00	0.8812
X_2X_3	-0.08	0.8761	-0.97	0.2792	0.00	0.4267
X_2X_4	-0.35	0.4982	0.31	0.7281	0.00	0.4009
X_2X_5	0.45	0.3881	-0.25	0.7795	0.00	0.9465
X_3X_4	0.53	0.3078	-1.07	0.2337	0.00	0.4465
X_3X_5	-0.42	0.4202	-0.27	0.7580	0.00	0.5952
X_4X_5	0.27	0.6009	-1.14	0.2053	0.00	0.3912
X_1^2	-8.17	<0.0001	-11.27	<0.0001	0.00	0.0030
X_2^2	-0.79	0.0303	0.16	0.7836	0.00	0.2221
X_3^2	-0.25	0.4690	0.26	0.6651	0.00	0.0362
X_4^2	-0.67	0.0648	-1.93	0.0032	0.00	0.0012
X_5^2	-0.75	0.0387	-1.75	0.0068	0.00	0.2946
R^2	0.95		0.94		0.70	
Adj R^2	0.91		0.88		0.46	
<i>F</i> value	23.56		18.01		2.93	
C.V (%)	15.53		2.21		0.33	
LOF	2.17	NS	3.1	NS	0.2	NS

X_1 : ultrafiltration membrane module; X_2 : transmembrane pressure; X_3 : volume concentration ratio; X_4 : diafiltration; X_5 : flow rate.

TABLE 4: Performance criteria used for BBD modelling.

Response variable	RMSE	SSE	Percent error	Chi-square (χ^2)	Absolute average deviation
Permeate flux	1.15	1.31	7.9	0.96	0.74
Protein in retentate	3.88	0.64	0.75	0.99	0.43
Protein retention	0.0034	0.08	0.24	0.99	0.22

pressure. The initial flux increase was attributed to enhanced cross-flow rate with pressure. Understanding these influences on permeate flux is pivotal for optimizing the ultrafiltration process, especially in selecting membrane modules and adjusting operating parameters to achieve desirable flux rates while minimizing pore blocking and flux decline over time.

4.2.2. Prediction of Protein in Retentate by Box-Behnken Method. In the analysis of the experimental data, the model *F* value of 18.01, indicating significance at $p < 0.0001$, establishes the overall significance of the model. Notably, values of “Prob > *F*” less than 0.0500 underscore the significance of model terms. Specifically, among these, the main effect of ultrafiltration membrane modules, including their square terms, as well as diafiltration and flow rate, emerged as noteworthy contributors within the model. Furthermore, the assessment of the “lack of fit *F* value” at 3.10 indicates its lack of significance concerning pure error, signifying the model’s adequacy in fitting the observed data. The comparability between the “Pred R-squared” (0.7732) and the “Adj R-squared” (0.8832) suggests a reasonable agreement in predicting outcomes and a robust fit of the model. Additionally, the other performance parameters (Table 4) contribute to elucidating the model’s aptness in predicting protein levels in the retentate. Interestingly, concerning membrane

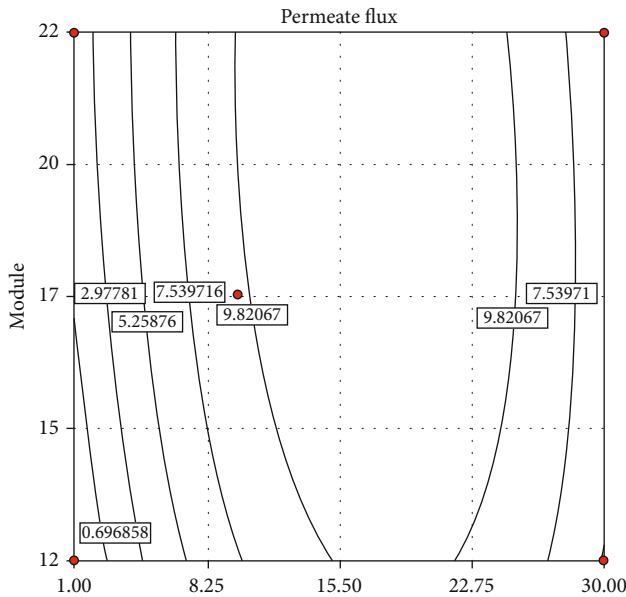


FIGURE 3: Contour plot representing the effect of membrane module and transmembrane pressure on permeate flux.

module size (kDa), a trend was observed wherein an initial increase in the retentate's protein content was followed by a subsequent decrease. However, this change was not solely linear, implying a nuanced relationship between module size and protein content. Moreover, the examination of interaction effects among independent parameters revealed their nonsignificant impact on the retentate's protein content, suggesting that their combined influence did not significantly alter the protein content outcomes.

4.2.3. Prediction of Protein Retention by Box-Behnken Method. The model F value of 2.93 demonstrates significance at $p < 0.0060$, affirming the overall significance of the model. Ultrafiltration membrane modules, their square terms, volume concentration ratio, and diafiltration emerged as pivotal model terms at a significance level of $p < 0.0500$, underscoring their substantive impact within the model. The "lack of fit F value" at 0.20 indicates its lack of significance concerning pure error, validating the model's suitability in accurately representing the observed data. Furthermore, the coefficient of variation (C.V) below ten percent reflects the experiments' precision, ensuring reliable and consistent results. The values of RMSE, SSE, absolute average deviation, etc. (Table 4), further underscore the appropriateness of this model in predicting protein retention through ultrafiltration membrane. There observed an inverse relationship between protein retention and ultrafiltration membrane module size (Figure 4). Remarkably, the highest protein retention was attained using a membrane module with a lower molecular weight cutoff. Similar observations align with prior research. For instance, Vijayasanthi et al. [35] examined protein recovery from coconut milk whey using ultrafiltration, noting 83% (w/w) protein retention with a 300 kDa membrane and 86-90% (w/w) retention with membranes at 5 and 50 kDa molecular

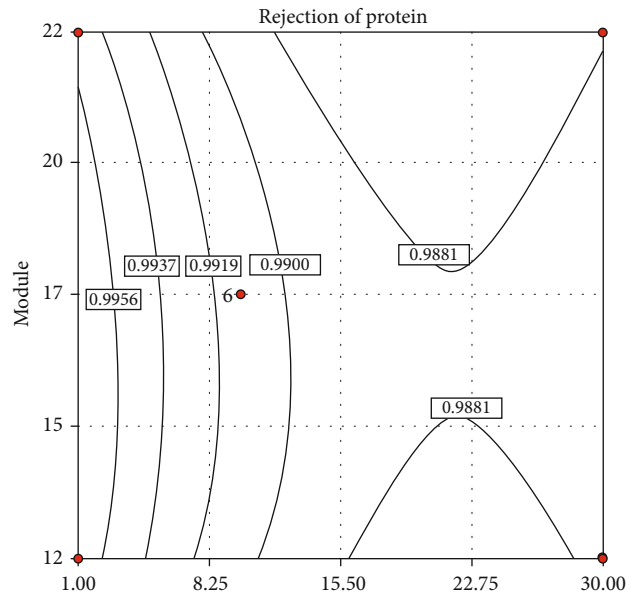


FIGURE 4: Contour plot representing the effect of ultrafiltration membrane module and volume concentration ratio on protein retention.

weight cutoffs. Similarly, Machado et al. [36] explored ultrafiltration (10 and 30 kDa) for protease separation and purification, highlighting superior purification with smaller MWCO membranes. The interaction effects among independent variables did not significantly influence protein retention, suggesting that their combined impact did not markedly alter the observed protein retention outcomes.

4.2.4. Optimization of Process Parameters. The polynomial model derived through response surface methodology underwent validation via three trials conducted at the numerically optimized point, maximizing desirability. The optimal ultrafiltration process parameters, determined using the desirability function approach with a desirability score of 0.767, were as follows: membrane module size: 13.38 kDa; transmembrane pressure: 117 kPa (17 PSI); volume concentration ratio: 3.6; diafiltration: 1; and flow rate: 63.04% of pump capacity. At these optimized conditions, the resulting performance indicators were as follows: permeate flux: 11.13 LPH; protein content in retentate: 85.52%; and protein retention: 98.99%. The identified parameters and their respective values showcase an enhanced performance, maximizing permeate flux while maintaining high protein content in the retentate and achieving substantial protein retention levels.

4.3. ANN Modelling

4.3.1. Prediction of Permeate Flux by ANN. In the experiment, the ANN model was developed to predict the target variable "permeate flux" using the "predictors" UMM (ultrafiltration membrane module), TMP (transmembrane pressure), VCR (volume concentration ratio), diafiltration, and FR (flow rate). Initially, data from the BBD parameters, including the response variable permeate flux, were

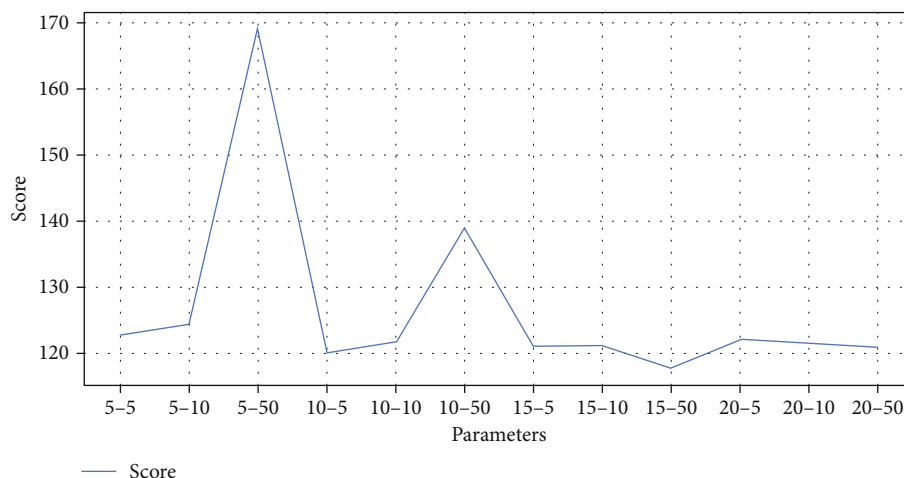


FIGURE 5: Batch size-epoch vs. score graph corresponds to permeate flux.

imported into the model development environment. To evaluate the model, the dataset underwent a split into training and testing sets. Subsequently, multiple configurations with different numbers of hidden layers and neurons were explored and tested to identify the optimal ANN model. The determination of the best combination of epochs and batch size involved a grid search approach, and a graphical representation was created to pinpoint the most suitable parameters (Figure 5).

However, despite these efforts, the performance of the developed ANN regression model in predicting permeates flux remained suboptimal. The accuracy of the model was noted to be only 39 percent, reflecting a significant discrepancy between predicted and actual values. Additionally, other performance metrics utilized to assess the efficiency of the ANN model (Table 5) did not align favourably, further suggesting inadequacy in using this particular ANN model for predicting permeate flux accurately. This outcome suggests that the current ANN architecture or parameter configuration might not sufficiently capture the complex relationships between the predictors and the target variable permeate flux.

4.3.2. Prediction of Protein in Retentate by ANN. An ANN model was employed to predict the “protein in retentate.” Figure 6 displays the batch size-epoch vs. score graph correlating with the prediction of protein in retentate. The model was trained using the combination of epoch and batch size values that corresponded to the highest peak value on the graph.

Upon training the ANN model, its accuracy was assessed and found to be 96.41 percent, indicating a high degree of alignment between predicted and actual values. The maximum absolute percentage error was 9.8 percent, with variability in absolute percentage error (APE) values ranging from 1.7 to 9.8 percent. Additionally, other metrics such as sum of squared errors (SSE) and root mean square error (RMSE) were calculated and obtained as 15.06 and 9.8 percent, respectively (Table 5). These metrics collectively suggest the suitability and effectiveness of the ANN model in

accurately predicting the protein content in the retentate. Furthermore, the optimization process identified the ultrafiltration parameters associated with achieving the maximum protein content in the retentate as follows: membrane module: 10 kDa; transmembrane pressure: 117 kPa (17 PSI); volume concentration ratio: 3.5; diafiltration: 1; and flow rate: 65% of pump capacity. This optimized parameter configuration yielded an absolute percent error value of 2.81. Consequently, with these optimized process parameters, the predicted value of protein content in the retentate was calculated to be 80.49. These findings underscore the capability of the ANN model to accurately predict the protein content in the retentate and highlight the identified optimal parameters for maximizing protein retention during ultrafiltration processes.

4.3.3. Prediction of Protein Retention by ANN. An ANN model was employed to predict “protein retention” using batch size and epoch configurations, identified from the graph (Figure 7). The model was trained using these optimal values, subsequently tested on a separate testing dataset. The accuracy of the model was determined to be 99.61 percent, indicating an exceptional alignment between predicted and actual protein retention values. The maximum absolute percentage error was noted at 0.74 percent, with APE values varying between 0.07 and 0.74 percent across different predictions. Additional performance metrics, as outlined in Table 5, further corroborate the efficacy of the ANN model for predicting protein retention through ultrafiltration. Notably, the root mean square error (RMSE) and sum of squared error (SSE) values approaching zero signify a close fit between the observed and predicted values, emphasizing the model’s accuracy and reliability in predicting protein retention [37].

Moreover, the optimized ultrafiltration parameters associated with maximizing protein retention were identified as follows: membrane module: 10 kDa; transmembrane pressure: 117 kPa (17 PSI); volume concentration ratio: 3.5; diafiltration: 1; and flow rate: 65% of pump capacity. These optimized process parameters exhibited an absolute percent

TABLE 5: Performance criteria used for ANN modelling.

Response variable	RMSE	SSE	Percent error	Chi-square (χ^2)	Absolute average deviation
Permeate flux	15.45	238.81	37.71	1.34×10^{-7}	12.20
Protein in retentate	3.88	15.06	4.18	0.9	2.88
Protein retention	0.0034	1.15×10^{-5}	0.34	1	0.005

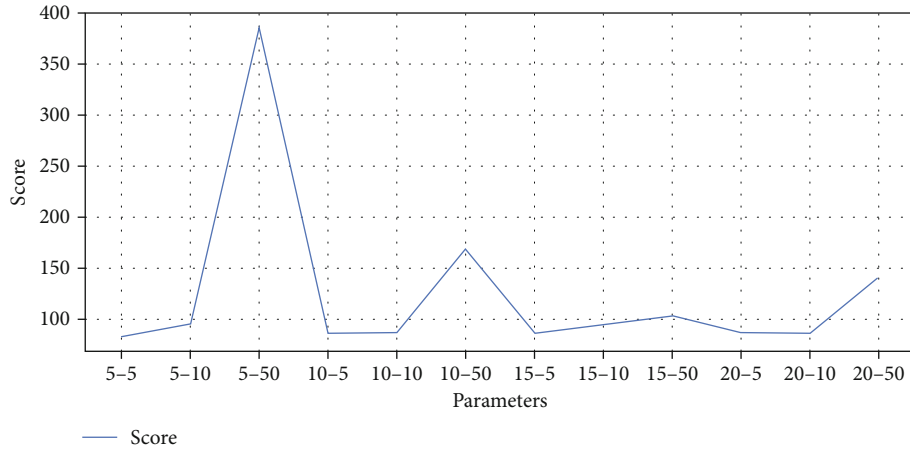


FIGURE 6: Batch size-epoch vs. score graph corresponds to protein in retentate.

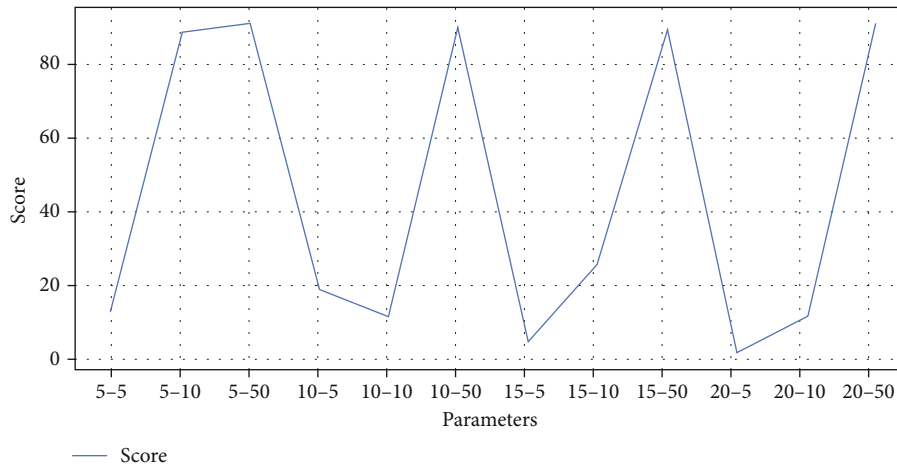


FIGURE 7: Batch size-epoch vs. score graph corresponds to protein retention by membrane.

TABLE 6: Comparison of the predicted optimum process parameters of ultrafiltration.

Ultrafiltration process parameters	RSM (BBD)	ANN
Membrane module	13.38 kDa	10 kDa
Transmembrane pressure	117 kPa	117 kPa
Volume concentration ratio	3.6	3.5
Diafiltration	1	1
Flow rate	63.04% of pump capacity	65% of pump capacity
Predicted values of response variable and accuracy		
Permeate flux	11.13 LPH (desirability 0.767)	39% accuracy
Protein in retentate	85.52% (desirability 0.767)	80.49 (96.41% accuracy)
Protein retention	98.99% (desirability 0.767)	99% (99.61% accuracy)

error value of 0.07. Consequently, using these parameters, the predicted value of protein retention was estimated to be 99 percent. These findings underscore the robustness of the ANN model in accurately predicting protein retention and highlight the identified optimal parameters for maximizing protein retention in ultrafiltration processes.

4.4. Comparison of Optimized Values from RSM and ANN. The RMSE, SSE, percent error, χ^2 , and absolute average deviation values for both the response surface methodology (RSM) using Box-Behnken design (BBD) and artificial neural network (ANN) models highlight their suitability in predicting ultrafiltration process parameters for soy protein preparation. In Cheok et al.'s study in 2012[38], they utilized both RSM and ANN to optimize phenolic compound extraction from *Garcinia* hull. Their findings favoured ANN as a superior modelling technique for nonlinear data, especially based on performance parameters such as average absolute deviation. Table 6 compares the predicted optimal process parameter values for ultrafiltration, aiming to attain desired response values. This table showcases a comparative analysis between the predictions made by RSM and ANN methodologies.

In Lin et al.'s [39] study, they conducted a comparison between response surface methodology (RSM) employing Box-Behnken design (BBD) and artificial neural network (ANN) models to optimize the ultrafiltration process aimed at removing nickel ions from aqueous solutions. Their findings affirmed the suitability of both tools in predicting ultrafiltration process parameters. Interestingly, the ANN model exhibited higher prediction accuracy compared to RSM. This aligns with observations made by Chakraborty et al. in 2014 [40], where they similarly noted that neural models effectively handled the nonlinear behaviour inherent in the ultrafiltration process.

5. Conclusion

Optimization of the ultrafiltration process for soy protein preparation was carried out using both response surface methodology (Box-Behnken design) and artificial neural network (ANN) techniques. For RSM, a quadratic model was utilized to optimize the ultrafiltration process parameters. The generated polynomial model from the response surface methodology underwent validation at the numerically optimized point, achieving maximum desirability. The resulting optimum process parameters for ultrafiltration were obtained through the desirability function approach, yielding a desirability score of 0.767. However, the ANN regression model faced challenges in accurately identifying the best combinations of predictors to optimize the permeate flux. Despite this limitation, the ANN model demonstrated efficacy in predicting the optimal process conditions for ultrafiltration to maximize both the protein content in the retentate and protein retention. Notably, the ANN model achieved an accuracy of 96.41 percent for predicting the optimal conditions to maximize protein content in the retentate and 99.61 percent accuracy for predicting the conditions maximizing protein retention.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Additional Points

Code Availability. The source code will be made available on request.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

- [1] H. J. Tanudjaja, A. Anantharaman, A. Q. Q. Ng et al., "A review of membrane fouling by proteins in ultrafiltration and microfiltration," *Journal of Water Process Engineering*, vol. 50, article 103294, 2022.
- [2] E. Ratnaningsih, R. Reynard, K. Khoiruddin, I. G. Wenten, and R. Boopathy, "Recent advancements of UF-based separation for selective enrichment of proteins and bioactive peptides—a review," *Applied Sciences*, vol. 11, no. 3, p. 1078, 2021.
- [3] W. Zhang, N. Grimi, M. Y. Jaffrin, L. Ding, and B. Tang, "A short review on the research progress in alfalfa leaf protein separation technology," *Journal of Chemical Technology & Biotechnology*, vol. 92, no. 12, pp. 2894–2900, 2017.
- [4] R. Aguero, E. Bringas, M. San Román, I. Ortiz, and R. Ibanez, "Membrane processes for whey proteins separation and purification. A review," *A review Current Organic Chemistry*, vol. 21, no. 17, pp. 1740–1752, 2017.
- [5] F. Ali, D. Ippersiel, F. Lamarche, and M. Mondor, "Characterization of low-phytate soy protein isolates produced by membrane technologies," *Innovative Food Science & Emerging Technologies*, vol. 11, no. 1, pp. 162–168, 2010.
- [6] S. R. Cabral, B. M. B. Azevedo, M. P. Silva, A. S. Figueiredo, A. P. L. Martins, and M. N. Pinho, "Optimization of cheese whey ultrafiltration/diafiltration for the production of beverage liquid protein concentrates with lactose partially removed," *Journal of Membrane Science and Research*, vol. 5, no. 2, pp. 172–177, 2019.
- [7] S. C. Ferreira, R. E. Bruns, H. S. Ferreira et al., "Box-Behnken design: an alternative for the optimization of analytical methods," *Analytica Chimica Acta*, vol. 597, no. 2, pp. 179–186, 2007.
- [8] Y. Zhao, R. Tian, Z. Xu, L. Jiang, and X. Sui, "Recent advances in soy protein extraction technology," *Journal of the American Oil Chemists' Society*, vol. 100, no. 3, pp. 187–195, 2023.
- [9] H. John, P. Chandra, S. K. Giri, and L. K. Sinha, "Effect of processing methods on 11S/7S protein and nitrogen solubility index of soy protein isolate," *Journal of The Institution of Engineers (India): Series A*, vol. 102, pp. 989–994, 2021.
- [10] H. John and L. K. Sinha, "Quality characteristics and functional properties of soy protein isolate prepared using ultrafiltration," *Journal of Agricultural Engineering*, vol. 56, no. 1, pp. 16–28, 2019.
- [11] M. Mondor, F. Ali, D. Ippersiel, and F. Lamarche, "Impact of ultrafiltration/diafiltration sequence on the production of soy protein isolate by membrane technologies," *Innovative Food Science & Emerging Technologies*, vol. 11, no. 3, pp. 491–497, 2010.

- [12] E. Ovando, L. Rodríguez-Sifuentes, L. M. Martínez, and C. Chuck-Hernández, "Optimization of soybean protein extraction using by-products from NaCl electrolysis as an application of the industrial symbiosis concept," *Applied Sciences*, vol. 12, no. 6, p. 3113, 2022.
- [13] X. B. Wang and Y. J. Chi, "Preparation of microwave-phosphorylated soy protein isolates through a Box-Behnken model optimization," *CyTA-Journal of Food*, vol. 10, no. 3, pp. 210–215, 2012.
- [14] J. Jawad, A. H. Hawari, and S. Zaidi, "Modeling of forward osmosis process using artificial neural networks (ANN) to predict the permeate flux," *Desalination*, vol. 484, article 114427, 2020.
- [15] A. Malekian and N. Chitsaz, *Concepts, procedures, and applications of artificial neural network models in streamflow forecasting in advances in streamflow forecasting*, Elsevier, 2021.
- [16] M. Asghari, A. Dashti, M. Rezakazemi, E. Joka, and H. Halakoei, "Application of neural networks in membrane separation," *Reviews in Chemical Engineering*, vol. 36, no. 2, pp. 265–310, 2020.
- [17] M. S. Noghabi, S. M. A. Razavi, and S. M. Mousavi, "Prediction of permeate flux and ionic compounds rejection of sugar beet press water nanofiltration using artificial neural networks," *Desalination and Water Treatment*, vol. 44, no. 1-3, pp. 83–91, 2012.
- [18] M. P. Phalak, D. U. Maheshwari, M. S. Singh, and M. S. Kumbhar, "Study of membrane transport for protein filtration using artificial neural networks," *International Journal of Engineering Research and Development*, vol. 13, no. 6, pp. 40–49, 2017.
- [19] Z. Sekulić, D. Antanasijević, S. Stevanović, and K. Trivunac, "Application of artificial neural networks for estimating Cd, Zn, Pb removal efficiency from wastewater using complexation-microfiltration process," *International Journal of Environmental Science and Technology*, vol. 14, no. 7, pp. 1383–1396, 2017.
- [20] M. Mousavi and A. Avami, "Modeling and simulation of water softening by nanofiltration using artificial neural network," *Iranian Journal of Chemistry and Chemical Engineering (IJCCE)*, vol. 25, no. 4, pp. 37–45, 2006.
- [21] M. Razavi, A. Mortazavi, and M. Mousavi, "Application of neural networks for crossflow milk ultrafiltration simulation," *International Dairy Journal*, vol. 14, no. 1, pp. 69–80, 2004.
- [22] J. Sargolzaei, A. N. Seghat, S. M. Mosavi, and M. Khoshnoudi, "Comparative study of artificial neural networks (ANN) and statistical methods for predicting the performance of ultrafiltration process in the milk industry," *Iranian Journal of Chemistry and Chemical Engineering*, vol. 25, no. 2, pp. 67–76, 2006.
- [23] N. Park, Y. L. H. Kim, H. Kim, Y. Lee, Y. Choi, and S. Lee, "Prediction of ultrafiltration membrane fouling using statistical models in pilot and full-scale operations," *Desalin. Water Treat.*, vol. 227, pp. 86–92, 2021.
- [24] M. T. Gaudio, S. Curcio, S. Chakraborty, and V. Calabrò, "Application of artificial neural networks for modelling and control of flux decline in cross-flow whey ultrafiltration," *PRO*, vol. 11, no. 4, p. 1287, 2023.
- [25] R. L. Thomas, L. I. Ndife, H. Shallo, and L. P. Nelles, *U.S. Patent No. 6,313,273*, Washington, DC, U.S. Patent and Trademark Office, 2001.
- [26] S. Abdullah, S. Karmakar, S. Mishra, and R. C. Pradhan, "Ultrafiltration of cashew apple juice using hollow fibers for shelf life extension: process optimization, flux modelling and storage study," *Journal of Food Measurement and Characterization*, vol. 17, no. 3, pp. 2182–2192, 2023.
- [27] S. Abdullah, R. C. Pradhan, and S. Mishra, "Membrane processing in the food industry," *Preservation, and Decontamination of Foods*, vol. 143, 2023.
- [28] A. Patra, S. Abdullah, and R. C. Pradhan, "Optimization of ultrasound-assisted extraction of ascorbic acid, protein and total antioxidants from cashew apple bagasse using artificial neural network-genetic algorithm and response surface methodology," *Journal of Food Processing and Preservation*, vol. 46, no. 3, Article ID e16317, 2022.
- [29] Association of Official Analytical Chemists AOAC, *Official Methods of Analysis Association of Official Analytical Chemists*, AOAC, Arlington, USA, 14th edition, 1984.
- [30] S. Ranganna, *Handbook of Analysis of Quality Control for Fruit and Vegetable Products*, Tata McGraw Hill Publishing Co, Ltd., New Delhi, 1st edition, 1995.
- [31] A. Haber and R. Runyon, "Manufacturing of soy protein isolate and concentrate by membrane technology," *Yongjae Lee International Journal of Agriculture Innovations and Research*, vol. 3, no. 4, pp. 1150–1158, 2015.
- [32] S. K. Chakraborty, B. K. Kumbhar, S. Chakraborty, and P. Yadav, "Influence of processing parameters on textural characteristics and overall acceptability of millet enriched biscuits using response surface methodology," *Journal of Food Science and Technology*, vol. 48, no. 2, pp. 167–174, 2011.
- [33] S. T. Sagu, S. Karmakar, E. J. Nso, C. Kapseu, and S. De, "Ultrafiltration of banana (*Musa acuminata*) juice using hollow fibers for enhanced shelf life," *Food and Bioprocess Technology*, vol. 7, no. 9, pp. 2711–2722, 2014.
- [34] S. J. Blonigen, *Ultrafiltration behavior of polyelectrolyte and protein mixtures*, [Ph.D. thesis], Iowa State University, 1994.
- [35] J. Vijayasanthi, S. R. Adsare, A. G. Lamdande, A. Naik, K. S. M. S. Raghavarao, and G. Prabhakar, "Recovery of proteins from coconut milk whey employing ultrafiltration and spray drying," *Journal of Food Science and Technology*, vol. 57, no. 1, pp. 22–31, 2020.
- [36] J. R. Machado, E. E. Severo, J. M. de Oliveira, J. D. C. Ores, A. Brandelli, and S. J. Kalil, "One-step ultrafiltration process for separation and purification of a keratinolytic protease produced with feather meal," *International Journal of Chemical Engineering*, vol. 2018, Article ID 6729490, 7 pages, 2018.
- [37] M. Nazerian, M. Akbarzade, P. Ghorbanzade et al., "Optimal modified starch content in UF resin for glulam based on bonding strength using artificial neural network," *Journal of Composites Science*, vol. 6, no. 10, p. 279, 2022.
- [38] C. Y. Cheok, N. L. Chin, Y. A. Yusof, R. A. Talib, and C. L. Law, "Optimization of total phenolic content extracted from *Garcinia mangostana* Linn. hull using response surface methodology versus artificial neural network," *Industrial Crops and Products*, vol. 40, pp. 247–253, 2012.
- [39] W. Lin, L. Jing, and B. Zhang, "Micellar-enhanced ultrafiltration to remove nickel ions: a response surface method and artificial neural network optimization," *Water*, vol. 12, no. 5, p. 1269, 2020.
- [40] D. J. Chakraborty, U. Farooq, J. Sikder, E. Drioli, and S. Curcio, "Experimental analysis, modeling and optimization of chromium (VI) removal from aqueous solutions by polymer-enhanced ultrafiltration," *Journal of Membrane Science*, vol. 456, pp. 139–154, 2014.