Scientific Research Publishing

Advances in Bioscience and Biotechnology, 2026, 17(2), 62-79
https://www.scirp.org/journal/abb
ISSN Online: 2156-8502
ISSN Print: 2156-8456

Ensuring Optimum Conditions for the Enzymatic Transesterification of Anthocyanin Mixture from Roselle (*Hibiscus sabdariffa* L.) Calyx Using RSM and ANN

Marie Liliane Mouto Kalla[1,2*], Guifo Joseph Kayem[1], Emmanuel Jong Nso[1], Bineesh Prakasan Nisha[2]

[1]Department of Process Engineering, National School of Agro-Industrial Sciences (NSAIS), University of Ngaoundéré, Ngaoundéré, Cameroon
[2]Department of Agro-Processing and Natural Product, National Institute for Interdisciplinary Science and Technology (NIIST), Trivandrum, India
Email: *moutomarili@gmail.com

How to cite this paper: Kalla, M.L.M., Kayem, G.J., Nso, E.J. and Nisha, B.P. (2026) Ensuring Optimum Conditions for the Enzymatic Transesterification of Anthocyanin Mixture from Roselle (*Hibiscus sabdariffa* L.) Calyx Using RSM and ANN. *Advances in Bioscience and Biotechnology*, **17**, 62-79.
https://doi.org/10.4236/abb.2026.172005

Received: December 22, 2025
Accepted: January 31, 2026
Published: February 3, 2026

Abstract

The sugar moieties of anthocyanins can be acylated via enzymatic transesterification, which improves their solubility in lipids and enhances their stability to external factors such as light, pH, and high temperatures by inducing structural changes. Process parameters, such as reaction time, temperature, and initial reaction conditions, define the reaction yield. Optimal conditions for the highest conversion yield were investigated during the enzymatic transesterification of a mixture of anthocyanins from Roselle calyx extract, using CAL-B as the catalyst and two acyl donors (methyl palmitate, MP, and vinyl laurate, VL). To model and optimise the reaction conditions, Response Surface Methodology (RSM) and a Variational Autoencoder coupled with an Artificial Neural Network (VAE-ANN) were applied to 15 data points generated using a Circumscribed Centred Composite (CCC) design. ANN was trained on 265 data points (15 experimental and 250 artificially generated using a VAE) and demonstrated better fit and predictive performance than RSM. Experimentally, the maximum conversion yields were 85.69% and 96.73% for VL and MP, respectively, whereas the predicted maximum values were 78.87% and 93.85%, respectively. Optimal conditions were found to be the same experimentally and by prediction for VL, using both RSM and VAE-ANN, whereas for MP, VAE-ANN yielded different conditions. The developed ANN exhibited Root Mean Square Error (RMSE) and Coefficient of Determination (R^2) values of 1.60 and 0.98 for MP, and 3.27 and 0.96 for VL, while RSM gave 6.82, 0.93 and 4.10, 0.96, respectively. The optimal conditions for the transesterifi-

cation of Roselle anthocyanins with VL (RSM, VAE-ANN) and MP (RSM) were 66˚C for 54 h 24 min and a substrate-to-enzyme ratio of 6.01. Both optimisation techniques could therefore be used to ensure optimal conditions for achieving the highest reaction yield, but VAE-ANN showed relative superiority to RSM.

Keywords

Hibiscus Sabdariffa, Anthocyanins, Optimum Transesterification Conditions, RSM, VAE-ANN

1. Introduction

Anthocyanins are a family of glycosylated derivatives of flavylium cations (2-phenylbenzopyrilium), which is the core structure of flavonoids, a subclass of polyphenols. They belong to the family of polyphenols and are responsible for the bright colours of flowers. However, due to their low solubility in non-aqueous solvents in their natural form, they show limited application in the biotechnological and pharmaceutical industries [1]. Moreover, these compounds generally exhibit poor stability under external conditions, including high temperatures, light, and pH [2]. In attempts to solve these problems, studies related to acylated anthocyanins with higher stability and increased solubility in organic solvents have shown that the acylation of anthocyanins can improve their stability and Hydrophilic Lipophilic Balance [1]-[4].

Acylation by enzymatic transesterification, due to mild and eco-friendly conditions, is a suitable method by which water-soluble anthocyanins can be converted to liposoluble compounds. This interfacial reaction is achieved in an emulsified system by linking an acyl group to the sugar moieties of anthocyanidins with lipase as a catalyst. Lipases are hydrolases used in biocatalysis to catalyse reactions *in vitro*, which differs from their natural function, such as transesterification [5]-[8]. Very stable in aqueous, organic, ionic, supercritical, and deep eutectic solvent media, they exhibit a peculiar mechanism of action known as "interfacial activation" within the large hydrophobic pocket surrounding the active centre [6] [9]. Among lipases, *Candida antarctica* lipase B (CAL-B) is a widely used biocatalyst in both its free and immobilised forms due to its high stability, tolerance to solvents, and high enantioselectivity, which accelerates various organic reactions and biotransformations with high yield [10]-[12]. A mixture of anthocyanins may present substrates with varying functional groups, leading to differences in selectivity and reaction rates due to the type of alcohol and steric hindrance [13] [14]. The nature of the acyl chain and the resulting fatty acid ester may also affect the linearity between reaction parameters and yield, underscoring the suitability of response surface methodology (RSM) for optimisation.

Factors such as medium temperature, reaction time, and enzyme-to-substrate

ratio can influence the reaction rate and thus the conversion yield. Response Surface Methodology (RSM) and Artificial Neural Network (ANN) are modelling methods used to develop predictive models. RSM is based on polynomial models that do not take account of nonlinearity [15] [16]. Currently, artificial intelligence, coupled with evolutionary computing, is a well-suited tool for learning and prediction in biotechnology, food science, and environmental science. Inspired by biological systems, Artificial Neural Network (ANN) is a non-linear data analysis tool which has been widely used for predicting enzyme-catalysed reactions [17]-[23].

Many types of ANNs have been developed and used, including Backpropagation Neural Networks (BP-ANNs). This network propagates backwards the difference (error) between the predicted and expected outputs, adjusting the weights of the connections between neurons via a gradient-descent iterative process that terminates when predictions reach a specified accuracy.

Data for RSM and ANN modelling can be obtained by implementing an experimental design. Due to the limited number of experiments, real laboratory experiments in enzymology yield small datasets. While these datasets may be sufficient for implementing RSM, they are insufficient for ANN and may lead to overreliance on the original data and loss of generalisation ability. Data augmentation can be performed to artificially generate data and thus increase the size and diversity of a training dataset by applying various transformations or modifications to the existing data samples. Algorithms such as the Variational Auto-Encoder (VAE) and Generative Adversarial Network (GAN) are widely used [24]. VAEs generate data samples from the original data distribution, whereas GANs generate data by training two networks to compete with one another. Numerous studies have shown that both VAEs and GANs improve model performance on classification problems [25]-[26]. Their application to regression problems is recent and rarer [24] [27] [28].

Ohno [27] addressed the issue of small data sizes in regression problems using multi-task learning (auto-encoding and regression tasks). By conducting experiments on eight datasets, he improved the generalisation performance of the multivariable linear regression model using multi-task learning for VAEs. In predicting product molecular weight in a bio-polymerisation system, Wei *et al.* [24] observed that ANN + VAE augmentation was more stable than ANN + GAN. This observation was made because the VAE was trained on the original distribution, resulting in data with strong regularity. Additionally, it was challenging to match the distribution of the original data using GAN augmentation.

This paper uses RSM and VAE + ANN to model and determine the optimum conditions for the transesterification reaction during the production process of an acylated anthocyanidin from Hibiscus sabdariffa L. calyx water extract and two acyl donor groups (methyl palmitate, MP and vinyl laurate, VL) with CAL-B as catalyst. Before transesterification, the anthocyanin compounds present in the extract were identified using preparative and analytical HPLC coupled to mass spectrometry.

2. Materials and Methods

2.1. Materials

Freshly harvested Roselle (*Hibiscus sabdariffa* L. var. vimto) calyxes in Tchatibali, a locality in Mayo-Danaï, Far North region, Cameroon, were oven-dried for 48 h at 45˚C before use. The acyl donors used in this study were Methyl Palmitate (MP) and Vinyl Laurate (VL), both from Sigma. Tert-amyl alcohol (Sigma) was used as a solvent. The immobilised lipase B from *Candida antarctica* on acrylic resin (≥ 5000 U/g, recombinant, expressed in Aspergillus niger, Sigma) was used as the catalyst.

2.2. Methods

2.2.1. Production Process of Roselle Calyx Acylated Anthocyanins and Data Collection

Purified Anthocyanins were obtained from Roselle calyx water extract and submitted to transesterification following the production process diagram presented in Figure 1.

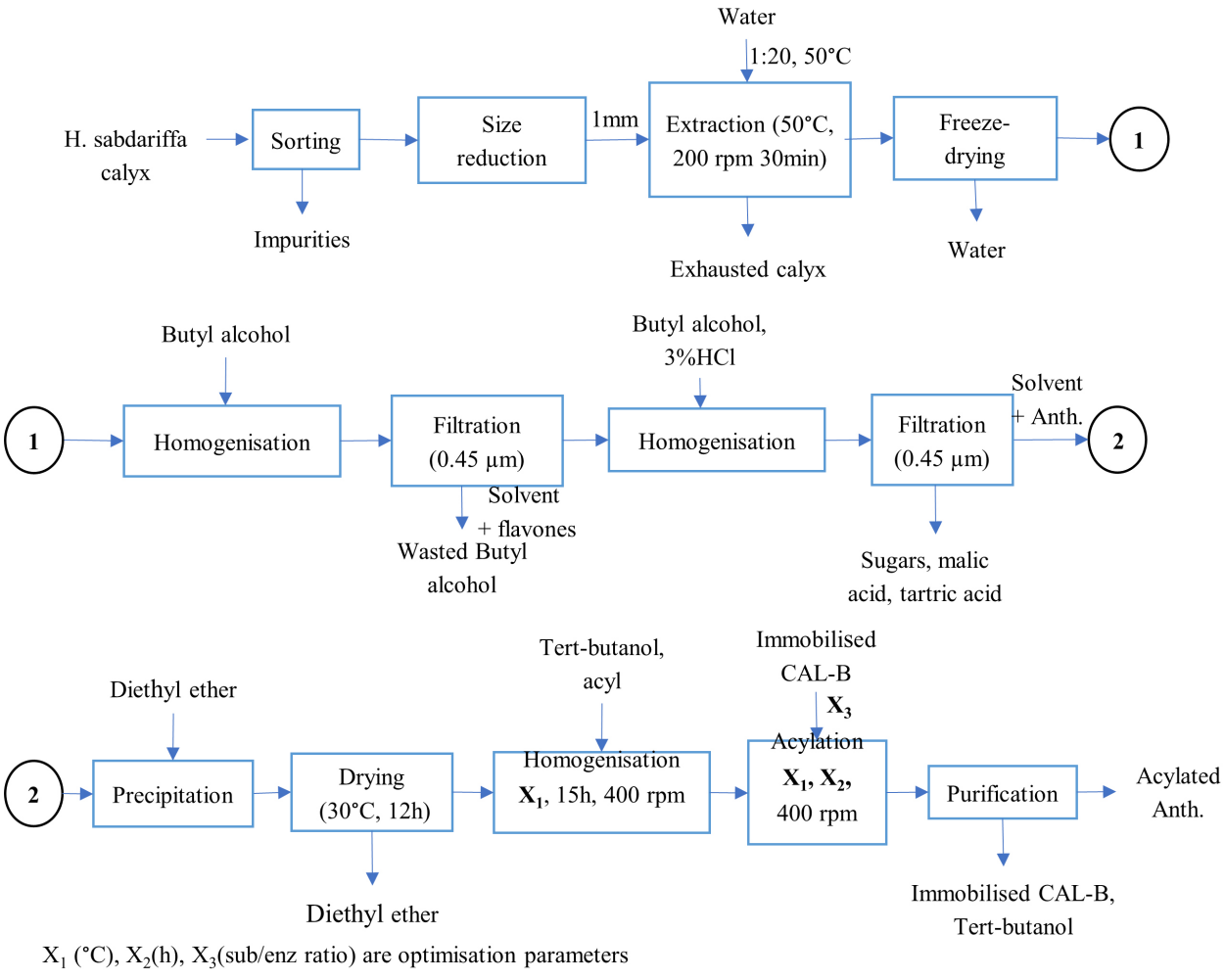

Figure 1. Process diagram of the production of acylated anthocyanins from *H. sabdariffa* L. calyx.

The transesterification reaction was conducted in the dark in a 50 mL stirred reactor. For each run, 15 mL of solvent was added to 50 mg of freeze-dried anthocyanidin extract, followed by 2 mL of acyl donor. The experiments were conducted using a Circumscribed Centred Composite (CCC) experimental design. The reaction was initiated by adding a specified quantity of immobilised lipase with a defined specific activity. Control experiments were performed without enzymes. Calcium chloride was used in a guard tube to prevent the reaction medium from absorbing water.

100 µL of suspension was withdrawn from the reaction medium at the relevant time, dissolved in 0.9 mL of methanol, and filtered with a 0.2 µm syringe filter. A volume of 20 µL of this solution was then injected into the HPLC system to evaluate the conversion yield. At the end of the reaction time, the biocatalysts were removed by centrifugation. Each experiment was conducted three times, and the mean values were used for calculations.

2.2.2. Identification of Anthocyanins and Transesterification Reaction Monitoring

HPLC was performed using a Shimadzu HPLC system (Kyoto, Japan; LC-20AD pump, CTO-20A column oven, and SPO-M20A diode-array detector). A C18 column of 4.6 mm diameter and 250 mm length (Phenomenex Gemini) was used for the separation. Data were analysed using LC Real Time Analysis software. Preparative HPLC was also performed, and an aliquot was collected based on the retention time (RT). MS was performed to identify the compounds. The maximum absorption wavelength (λ_{max}) at which data were to be collected was determined by analysis of the absorption spectrum at pH one and pH 4.5 of a 1/50 to 1/5 diluted solution of 1 mg/mL anthocyanins contained in the pellets.

Elution for quantitative reaction monitoring was performed using a low-pressure gradient of 5% formic acid in Milli-Q water (solvent A) and ACN (solvent B) at a flow rate of 0.8 ml/min and a column temperature of 40˚C. The program was as follows: 2 min (10% B), 5 min (20% B), 10 min (30% B), 15 min (33% B), 20 min (30% B), 25 min (40% B), 30 min (50% B), 32 min (100% B), and 35 min (10% B). The initial quantity of solvent B was 10%. The different compounds were quantified at 330 and 520 nm. Each time, 20 µL of the sample was injected, and the resulting chromatogram was compared to the reference (at t = 0 s).

The conversion yield was calculated with the following formula:

$$C_{yield} = \frac{\left(A_{t=0} - A_t\right)}{A_{t=0}} \quad (1)$$

where $A_{t=0}$ is the peak area at t = 0 s, and A_t is the peak area at a given time.

2.2.3. System Predictive Modelling and Optimisation

Response Surface Methodology (*RSM*)

A 3-factor circumscribe central composite (CCC) experiment design was used. The system response (y) was the conversion yield. The reaction temperature X_1,

the reaction time X_2, and the substrate-to-enzyme ratio X_3 were the explanatory variables. Absolute values were obtained by converting coded values with the formula:

$$U_j = x_j \Delta U_j + U_j^0 \tag{2}$$

where U_j and x_j are respectively a real variable j and its corresponding coded value. ΔU_j which is the variation step, was calculated with:

$$\Delta U_j = \frac{\left(U_j \max - U_j^0\right)}{x_j \max} = \frac{\left(U_j \min - U_j^0\right)}{x_j \min} \tag{3}$$

Max and min were meant to represent the maximum and minimum value of a variable j (Real or coded), while U_j^0 is the natural value at the centre of the range for each factor (Table 1).

Table 1. Process variable employed in the CCC design.

variable	Real values				Coded values	
	U_j max	U_j min	U_j^0	Δu_j	$X_j$min	$X_j$max
X_1 Temp (˚C)	70	50	60	5.95	−1.68	1.68
X_2 Time (h)	72	0	36	21.40	−1.68	1.68
X_3 Sub/Enz*	10	5	7.5	1.49	−1.68	1.68

*Mass ratio.

The centre point was replicated 5 times, for a total of 20 experiments. The relationship between the response and the independent variables was obtained by fitting the acquired data into the following second-order polynomial equation:

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \sum\sum_{i<j}^{k} \beta_{ij} x_i x_j + \varepsilon \tag{4}$$

Analysis of variance (ANOVA) was used to evaluate the variation of the response (y). The data were analysed using Statgraphic Centurion 15.1.

***Artificial neural networks* (*ANNs*)**

CCC Additional runs were obtained through data augmentation. Before augmentation, dimensional effects were removed by normalisation to the range 0 – 1 of both input (original CCC matrix values) and output variable (experimental response, Yield). After generation, the synthetic data were denormalised to the original physical units to enable direct comparison with experimental observations. The following relations were used:

$$X^* = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}} \tag{5}$$

$$x_{ij} = X^* \left(x_{\max} - x_{\min}\right) + x_{\min} \tag{6}$$

Python Code was used to define a Variational Autoencoder (VAE) model in TensorFlow. For each acyl, the VAE model was trained on the CCC matrix and the experimental system response to generate new rows in Google Colab

(https://colab.research.google.com). The mean response of the centre points was used for the VAE; thus, only a 15-line matrix was used. The VAE used an encoder consisting of a single dense hidden layer with 64 neurons and ReLU activation, followed by a two-dimensional latent space parametrised by mean and log-variance vectors. The decoder mirrored this structure with a dense layer of 64 neurons and a sigmoid output layer. The model was trained for 100 epochs using the Adam optimiser (learning rate 0.001), minimising the sum of the binary cross-entropy reconstruction loss and the Kullback-Leibler divergence. Boundary filtering and physical consistency were used to ensure chemical plausibility.

A self-organising feature map network was used to predict the enzymatic conversion yields. The architecture of the feedforward incremental backpropagation network consisted of an input layer with three neurons (X_1, X_2, and X_3), one hidden layer, and an output layer. The incremental mode was chosen to achieve convergence within a few epochs, thereby reducing computational time. Several networks were developed to determine the number of neurons in the hidden layer by varying the number of neurons and observing changes in mean squared error (MSE) and R^2. The training was performed using a Levenberg-Marquardt backpropagation algorithm in MATLAB R2015b. The transfer functions chosen for the response (output layer) and the hidden layers were the linear function and the hyperbolic Sigmoid Tangent function (tansig). Training and testing were performed on the randomised experimental and augmented data, with 70% allocated to training, 15% to validation, and 15% to testing. The trained network was saved as a MATLAB. mat file for subsequent use.

Models' validation and comparison tools

Predicted values obtained from RSM and ANNs were compared with experimental data using statistical metrics such as the coefficient of determination (R^2), the chi-square test statistic (χ^2), the Absolute Average Deviation (AAD), the Root Mean Squared Error (RMSE), and the Standard Error of Prediction (SEP). They were calculated using the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{p-}y_e\right)^2}{\sum_{i=1}^{n}\left(y_{e-}y_{em}\right)^2} \tag{7}$$

$$\chi^2 = \sum_{i=1}^{n}\frac{\left(y_p - y_e\right)^2}{y_p} \tag{8}$$

$$AAD = \frac{\sum_{i=1}^{n}\frac{\left|\left(y_p - y_e\right)\right|}{y_e}}{n} \tag{9}$$

$$RSME = \sqrt{\frac{\left[\sum_{i=1}^{n}\left(y_p - y_e\right)^2\right]}{n}} \tag{10}$$

$$SEP = \frac{100}{y_{pm}}\sqrt{\frac{\sum_{i=1}^{n}\left(y_p - y_e\right)^2}{n}} \tag{11}$$

where n is the number of experiments, y_e and y_p, experimental and predicted system response, and y_{em}, y_{pm} their corresponding mean values.

3. Results and Discussion

3.1. Identification of Anthocyanins Present in Roselle Extract

Anthocyanin pellets extracted from Roselle aqueous freeze-dried extract represented up to 1.33 ± 0.05 g/5 g of dried calyx powder, and the λ_{max} for their analysis was found to be 518 nm. Three major compounds detected and identified in aliquots collected by preparative HPLC were analysed by FTMS-ESI in positive ion mode, yielding interesting results. The obtained results are presented in the following Table 2.

Table 2. Mass details of identified compounds.

Compound name	Dp-3-sambubioside	Cy-3-sambubioside	Cy-3,5-diglucoside
m/z ([M⁺])	597.14	581.80	610.80
[M - Sambubioside]⁺	303.05	287.10	287.10
Molecular weight (g.mol⁻¹)	597.14	581.50	611.52
Percentage (%)	56.2	29.8	14.0

Apart from the identified compounds, there can be other anthocyanins, which are unidentified, in the studied sample (others may not have been detected because of very low concentration). Although other studies have confirmed the presence of these compounds in Roselle calyx samples, our findings align with the most frequently reported compounds, Dp and Cy. After developing a selective extraction method for anthocyanins from dried Roselle calyces from Guerlé, Senegal, Segura-Carretero *et al.* [29] identified various components by exact mass and fragmentation patterns using CE-TOF-MS in positive-ion ESI mode. They identified Delphinidin-3-sambubioside and cyanidin-3-sambubioside as the main components. Cyanidin-3,5-diglucoside was identified as a minor component along with cyanidin-3-O-rutinoside and delphinidin-3-O-glucoside. Using HPLC with diode array detection coupled to electrospray TOF and IT tandem MS, Rodriguez-Medina *et al.* [30] also identified, using positive ionisation mode, Delphinidin-3-sambubioside and cyanidin-3-sambubioside as the main anthocyanins present in Roselle dried calyx. In a study published in 2013, Omotuyi *et al.* identified up to five classes of compounds and acetylated malvidin (m/z 391.33) as major anthocyanidins in H.S hydrolysate using Mass spectrophotometry [31]. The concerned were pelargonidin (m/z 271.23), Cyanidin (m/z 287), Peonidin (m/z 299.28), Delphinidin (m/z 301), and petunidin (m/z 317.31). In the present study, the significant component, Delphinidin, was followed during transesterification.

3.2. Transesterification Reaction and Yield

As described in the methods, the transesterification reaction was carried out on

the Roselle calyx water extract, which contained the identified compounds at specified percentages. During the process, some of the anthocyanins present in the extract were not converted. Meanwhile, as shown in Figure 2, the chromatographic monitoring of Delphinidin-3-sambubioside, the most abundant compound, indicated complete disappearance. The corresponding conversion yield did not reach 100%. This apparent discrepancy can be attributed to competing hydrolysis reactions that produce non-acylated products, to the formation of non-detectable side esters at the monitoring wavelength, and/or to thermal degradation of anthocyanins under prolonged heating. The consumption of the parent substrate via parallel pathways has been shown to occur alongside the formation of target acylated derivatives [2] [32]. These pathways may include anthocyanin degradation and secondary reactions during enzymatic modification.

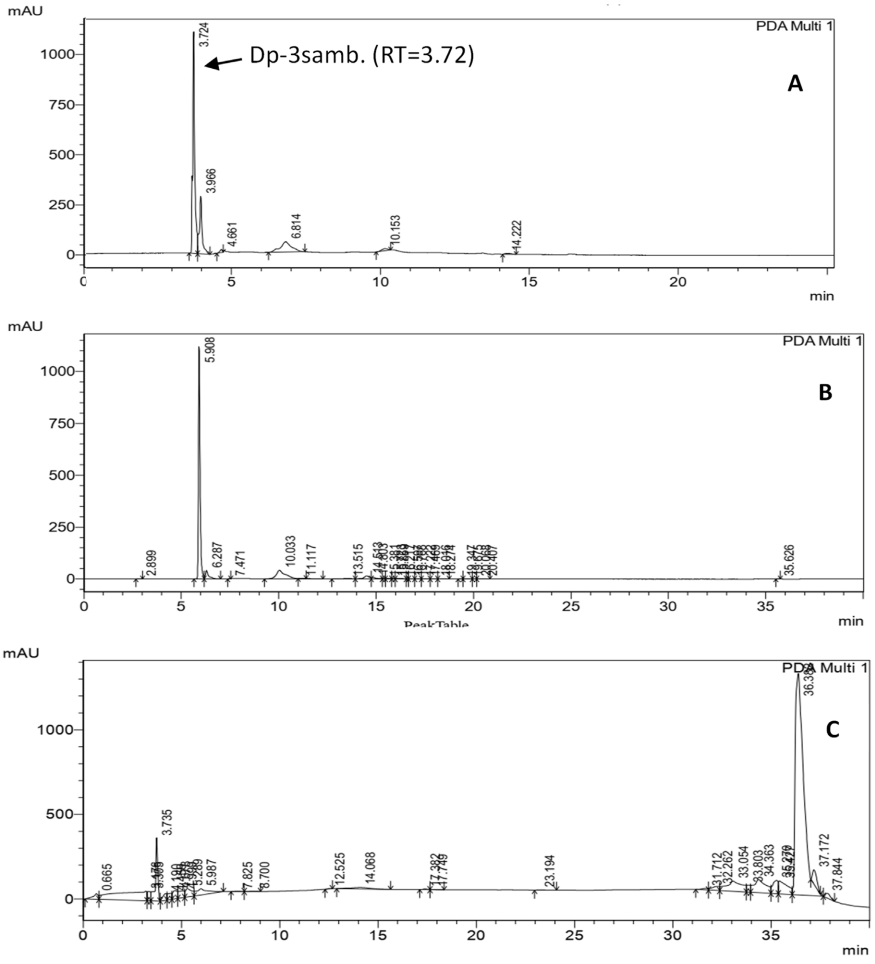

Figure 2. Chromatograms of Hs anthocyanins before reaction (A); acylated Hs anthocyanidin after 72 h of acylation reaction with VL (B); acylated Hs anthocyanidins after 72 h of acylation reaction with MP (C).

3.3. Modelling of the Transesterification of Dp-3-Sambubioside with CAL-B

As shown in Figure 3, the generated RSM and VAE-ANN data and predictions

indicated optimal operating parameters for maximum conversion yield.

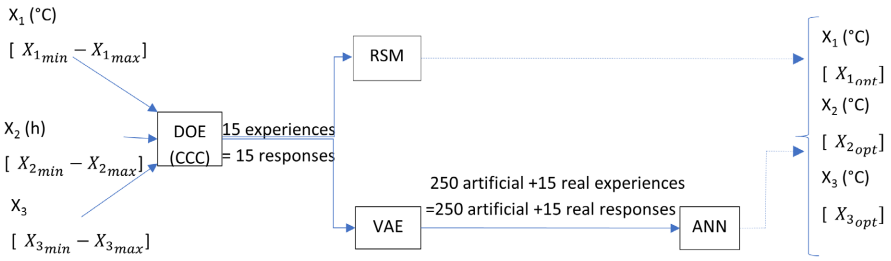

Figure 3. Data generation and computing approach.

The conversion yield obtained for each experiment without augmentation is presented in **Table 3**.

Table 3. Obtained experimental and predicted conversion yields from the CCC matrix without augmentation.

Run	Coded values (RSM)			Normalised values (ANN)			Real matrix values			Experimental Yield (%)		Predicted Yield (%)			
												RSM		ANN	
	X_1	X_2	X_3	X_1	X_2	X_3	X_1 (°C)	X_2 (h)	X_3	VL	MP	VL	MP	VL	MP
1	−1	−1	−1	0.2	0.2	0.2	54	14.60	6.01	8.36	8.67	6.06	2.96	7.73	8.43
2	1	−1	−1	0.8	0.2	0.2	66	14.60	6.01	31.75	49.44	23.71	39.94	29.25	47.97
3	−1	−1	1	0.2	0.2	0.8	54	14.60	8.99	14.45	17.76	14.38	23.13	13.34	17.25
4	1	−1	1	0.8	0.2	0.8	66	14.60	8.99	14.63	45.08	18.78	47.09	13.50	43.75
5	−1	1	−1	0.2	0.8	0.2	54	57.40	6.01	54.49	66.46	49.74	67.19	50.17	64.48
6	1	1	−1	0.8	0.8	0.2	66	57.40	6.01	**85.69**	93.89	**85.16**	**91.26**	78.87	91.09
7	−1	1	1	0.2	0.8	0.8	54	57.40	8.99	33.08	55.85	40.52	68.09	30.48	54.20
8	1	1	1	0.8	0.8	0.8	66	57.40	8.99	61.00	70.70	62.68	79.15	56.16	68.60
9	0	−1.68	0	0.5	0	0.5	60	0	7.5	0	0	3.44	5.98	0.04	0.02
10	0	1.68	0	0.5	1	0.5	60	72	7.5	79.60	**96.73**	77.01	86.86	73.27	93.85
11	0	0	−1.68	0.5	0.5	0	60	36	5	46.47	60.02	55.48	71.52	42.79	58.24
12	0	0	1.68	0.5	0.5	1	60	36	10	51.75	93.68	43.60	78.29	47.65	90.89
13	−1.68	0	0	**0**	**0.5**	**0.5**	50	36	7.5	7.42	24.67	6.94	18.47	6.86	23.95
14	1.68	0	0	1	0.5	0.5	70	36	7.5	39.06	56.51	40.40	58.82	35.97	54.84
15	0	0	0	**0.5**	**0.5**	**0.5**	60	36	7.5	40.72	68.67	40.61	68.79	37.50	66.63
***16**	0	0	0	0.5	0.5	0.5	60	36	7.5	40.77	68.77	40.61	68.79	37.50	66.63
***17**	0	0	0	**0.5**	**0.5**	**0.5**	60	36	7.5	40.91	68.86	40.61	68.79	37.50	66.63
***18**	0	0	0	0.5	0.5	0.5	60	36	7.5	40.31	68.33	40.61	68.79	37.50	66.63
***19**	0	0	0	**0.5**	**0.5**	**0.5**	60	36	7.5	40.93	68.76	40.61	68.80	37.50	66.63
***20**	0	0	0	0.5	0.5	0.5	60	36	7.5	40.17	68.67	40.61	68.79	37.50	66.63

*Replication of the centre point.

3.3.1. Response Surface Methodology

Significant effects of process variables were determined by analysis of variance. The P-values less than 0.05 showed that four effects, Temperature (X_1), Time (X_2), the quadratic effect of Temperature (X_1^2), and the quadratic effect of Time (X_2^2), were significantly different from zero at the 95.0% confidence level on the conversion yield. It was observed that both temperature and time increased the conversion yield, with the effect of time being more pronounced. However, their quadratic effect reduced the system response due to enzyme behaviour at temperatures above their optimal values. The obtained models, expressed as coded factors, were transformed via variable changes and plotted. The substrate-to-enzyme ratio had the least effect on the conversion yield among the three factors and was therefore fixed at the midpoint of its range. The obtained models with absolute values were:

$$C_{\text{yield (MP)}} = -1257.81 + 39.12X_1 + 3.89X_2 - 0.02X_1X_2 - 0.30X_1^2 - 0.02X_2^2 \quad (12)$$

And,

$$C_{\text{yield (VL)}} = -631.94 + 20.76X_1 - 1.05X_2 + 0.03X_1X_2 - 0.17X_1^2 - 0.00X_2^2 \quad (13)$$

Corresponding response surface areas are presented in Figure 4.

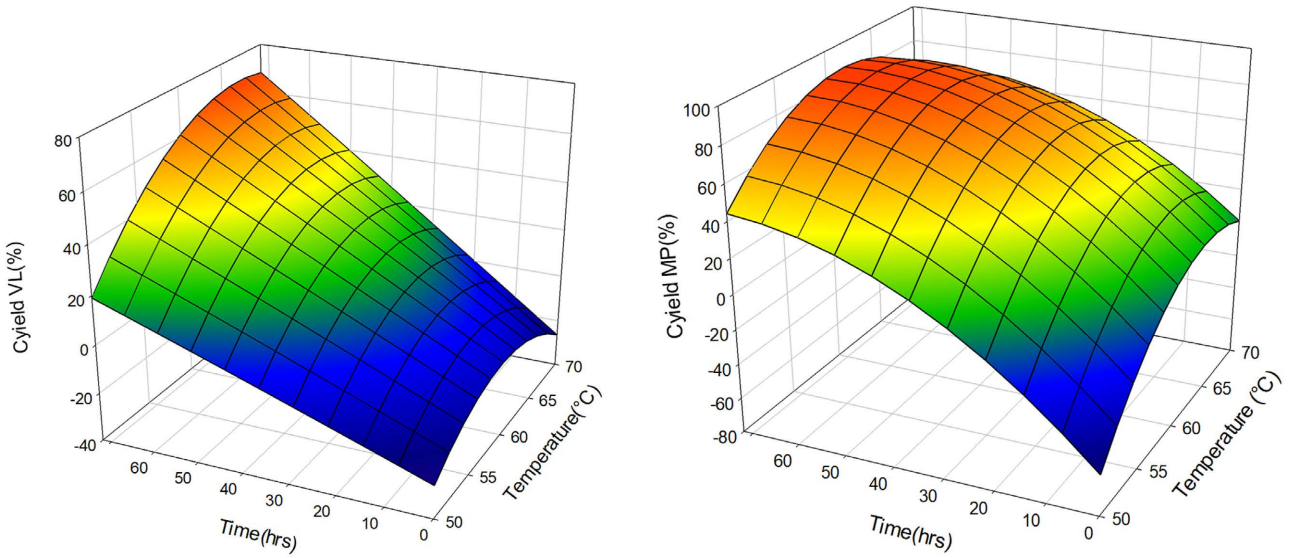

Figure 4. Response surface of conversion yield for (A) VL and (B) MP under the effects of time and temperature.

The temperature was found to affect the reaction yields of both the VL and MP systems. This may be due to changes in the reaction medium's viscosity, substrate solubility, activation, stability, and thermal denaturation of the enzyme [33]. It was observed that, in addition to temperature, time also influences the two systems; the obtained response surfaces show near-linear (VL) and curvature (MP) dependence of yield. The difference in the linearity of response observed between MP and VL systems can be explained by the physicochemical influence of the acyl chain on enzyme-substrate interaction. CAL-B is known to behave as a pocket-limited enzyme, in which acyl donors compete for a confined active site cavity [9] [13]. Longer or bulkier chains impose steric hindrance at the entrance to the hy-

drophobic pocket of CAL-B, thereby altering activation energy and temperature sensitivity. In addition, differences in hydrophobicity modify mass-transfer resistance and substrate solubility in tert-amyl alcohol. Similar steric-governed rate behaviours have been reported for acylated flavonoids and sugar ester synthesis [3] [32] [33] [34]. During the synthesis of phloridzin cinnamate via trans-esterification with Novozym 435® (CAL-B), Enaud *et al.* (2004) [35] showed that the reaction time depended on temperature: 15 h at 60˚C, 4 h at 80˚C, and 2 h at 100˚C. Turner and Vulfson [36] showed that this enzyme remains active without loss up to 130˚C, a temperature not considered in the present work due to the stability of the substrates. Other works [33] [37]-[39] have, however, noted that despite its effectiveness, the effect of temperature lacks clarity and depends on the environment of the enzyme in the case of the synthesis of a sugar ester (as is the case of sambubioside linked to Delphinidin).

3.3.2. ANN Modelling and Optimisation

The AAD, SEP, and RMSE values for the initial non-augmented data and the dataset composed of experimental and augmented data showed that increasing the number of experimental points in the training dataset via VAE improved the network's performance and, thus, its predictive ability.

The number of ANN hidden neurons was determined by trial observation and comparison of mean square error sums (MSE). Across hidden neuron counts from 1 to 20, the optimal VAE-ANN models were obtained with 15 neurons and 19 and 40 epochs for VL and MP, respectively. So, the topology selected for our network was 3-15-1. **Figure 5** shows the validation performances of the networks. For both systems, the best validation performance was achieved at epoch 13, with MP outperforming VL.

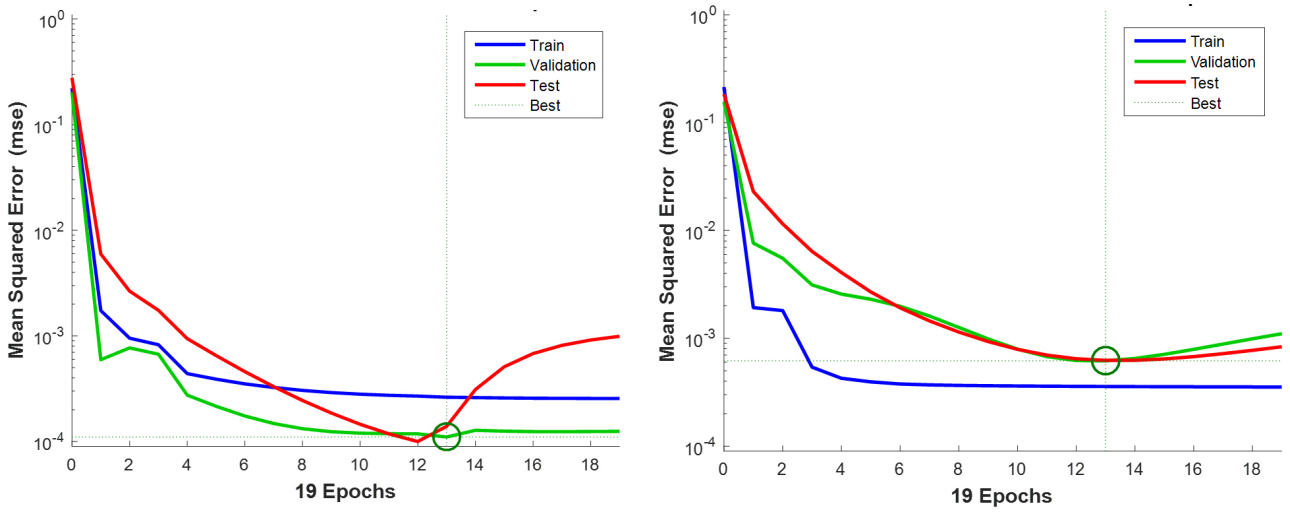

Figure 5. Validation performance of VL (A) and MP (B) built neural networks.

3.3.3. Comparison between ANN and RSM

RSM and ANN have shown that both can be excellent predictors of conversion

efficiency in D-3-sambubioside transesterification in the ranges of considered factors. Figure 6 presents the predicted-experimental plots for VL and MP using RSM and ANN.

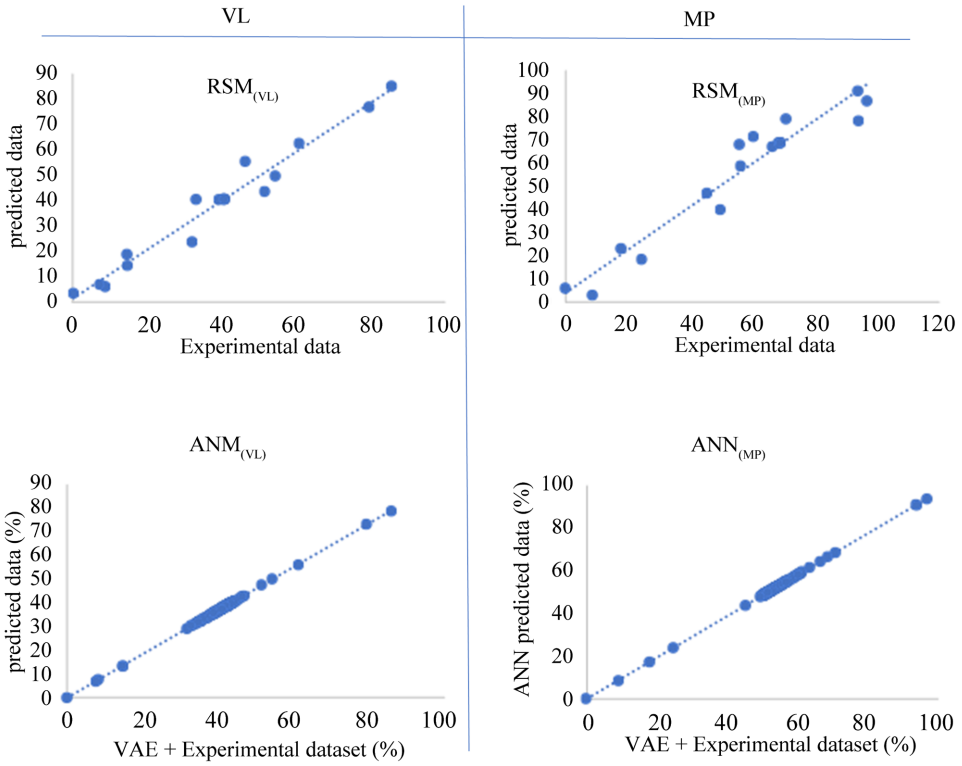

Figure 6. Scatter plots of RSM and ANN Experimental CCC matrix response values versus Predicted values for both VL and MP systems.

The values obtained for the statistical parameters used for comparison are presented in Table 4. Both methods have demonstrated good data-fitting and predictive performance.

Table 4. RSM and ANN Experimental and Predicted Optimum with corresponding statistical parameters

Method		Predicted Optimum				$R^{2*} \geq 0.90$	AAD absolue 0.2* ≤ |0.2|	RMSE	SEP	χ^2
		Yield	Parameters							
			Temp. (˚C)	Time (h)	Ratio sub/enz					
RSM	VL	85.16	66	57.40	6.014	0.96	0.08	4.10	10.64	12.99
	MP	91.26				0.93	0.12	6.82	11.85	31.93
ANN	VL	84.74	66	57.40	6.014	0.94	0.09	5.96	15.20	63.56
	MP	84.75	60	72	7.5	0.96	0.12	6.75	11.79	27.66
VAE-ANN	VL	78.87	66	57.40	6.014	0.96	0.07	3.27	8.70	73.56
	MP	93.85	60	72	7.5	0.98	0.03	1.60	3.09	12.80

*Validation criterion.

The optimal predicted operating parameters for the transesterification of Roselle anthocyanins with VL were identical across both the RSM and ANN methods. With MP, the predicted optimum conditions differed; AAD, RMSE, and SEP values indicated that ANN was more accurate and thus more effective. Although ANN predicted a slightly higher optimum yield than RSM for MP, operating at higher temperatures and longer times may increase energy consumption. Conversely, RSM identified conditions involving lower temperature and reduced residence time for VL. Overall, for both systems and methods, the R^2 and AAD values fell within the acceptable ranges, indicating that both models were valid for prediction. ANNs (ANN and VAE-ANN) demonstrated superior performance for MP. In contrast, RSM exhibited competitive performance for VL, particularly with respect to χ^2, with minor discrepancies between experimental and predicted values, indicating that the two were very close. These observations suggest that different reaction mechanisms induced by acyl donor group structures may have influenced the performance of the fitting algorithms. In other words, the efficiency of fitting by minimising the sum of the squared differences between the experimental and predicted responses (RSM) and Backpropagation combined with the Marquart-Levenberg algorithm, which is also a combination of the gradient descent method and the Gauss-Newton method (ANN), may depend on the behaviour of the system and involved mechanisms. Although ANN is often described as the best-fitting method for nonlinear regression, many authors have found that RSM outperforms ANN, noting lower prediction deviation [40] [41]. When using a Central Composite Rotatable Experimental Design (CCRD) to model lipase production from used engine oil as the substrate in *Burkholderia cenocepacia* ST8, Lau *et al.* [42] found that the RSM model had higher accuracy in predicting lipase activity yield.

4. Conclusion

The optimal reaction conditions for the transesterification of anthocyanins from Roselle by CAL-B with Methyl Palmitate (MP) and Vinyl Laurate (VL) as acyl donors were investigated using an Artificial Neural Network (ANN) combined with a Variational Autoencoder (VAE) and Response Surface Methodology (RSM). By augmenting the dataset with over 250 additional data points, ANN models were trained on a more robust dataset, yielding superior VAE-ANN predictions than simple ANN. For both acyls, the optimal ANN model featured three neurons in the input layer, 15 in the hidden layer, and 1 in the output layer. The optimal conditions for achieving the highest transesterification yields varied between the two acyl donors, with the best results observed for Methyl Palmitate with VAE-ANN. However, RSM outperformed ANNs in predicting conversion yield for the transesterification reaction with VL. This research highlights the effectiveness of advanced modelling techniques in optimising reaction conditions for enzyme-catalysed reactions.

Acknowledgement

Marie Liliane MOUTO KALLA acknowledges the financial assistance provided by the Third World Academy of Science (TWAS) and the Department of Bio-Technology (DBT) India through the scholarship PG2013 TWAS-DBT Fellowship. Marie Liliane MOUTO KALLA acknowledges CSIR-NIIST for hosting and infrastructure facilities. The authors acknowledge Dr Ravi Shankar and Mr Sasikumar P. for their contributions to the identification of the compound.

Credit Authorship Contribution Statement

Marie Liliane Mouto Kalla: Methodology, formal analysis, computation, and production of the original draft. **Joseph Guiffo Kayem** and **Emmanuel Nso Jong**: Supervision and review. **P. Nisha**: Supervision, Resources and review.

Conflicts of Interest

The authors report no declaration of interest.

References

[1] Dulęba, J., Czirson, K., Siódmiak, T. and Marszałł, M.P. (2019) Lipase B from Candida Antarctica—The Wide Applicable Biocatalyst in Obtaining Pharmaceutical Compounds. *Medical Research Journal*, **4**, 174-177. https://doi.org/10.5603/mrj.a2019.0030

[2] Luo, X., Wang, R., Wang, J., Li, Y., Luo, H., Chen, S., *et al.* (2022) Acylation of Anthocyanins and Their Applications in the Food Industry: Mechanisms and Recent Research Advances. *Foods*, **11**, Article 2166. https://doi.org/10.3390/foods11142166

[3] Stevenson, D.E., Wibisono, R., Jensen, D.J., Stanley, R.A. and Cooney, J.M. (2006) Direct Acylation of Flavonoid Glycosides with Phenolic Acids Catalysed by Candida Antarctica Lipase B (Novozym 435). *Enzyme and Microbial Technology*, **39**, 1236-1241. https://doi.org/10.1016/j.enzmictec.2006.03.006

[4] Zeng, S., Lin, S., Jiang, R., Wei, J. and Wang, Y. (2025) Biotechnology Advances in Natural Food Colorant Acylated Anthocyanin Production. *Food Frontiers*, **6**, 698-715. https://doi.org/10.1002/fft2.527

[5] Mokhtar, N.F., Abd. Rahman, R.N.Z.R., Muhd Noor, N.D., Mohd Shariff, F. and Mohamad Ali, M.S. (2020) The Immobilization of Lipases on Porous Support by Adsorption and Hydrophobic Interaction Method. *Catalysts*, **10**, Article 744. https://doi.org/10.3390/catal10070744

[6] Ortiz, C., Ferreira, M.L., Barbosa, O., dos Santos, J.C.S., Rodrigues, R.C., Berenguer-Murcia, Á., *et al.* (2019) Novozym 435: The "Perfect" Lipase Immobilized Biocatalyst? *Catalysis Science & Technology*, **9**, 2380-2420. https://doi.org/10.1039/c9cy00415g

[7] Saik, A.Y.H., Lim, Y.Y., Stanslas, J. and Choo, W.S. (2017) Enzymatic Synthesis of Quercetin Oleate Esters Using Candida Antarctica Lipase B. *Biotechnology Letters*, **39**, 297-304. https://doi.org/10.1007/s10529-016-2246-5

[8] Thangaraj, B., Solomon, P.R., Muniyandi, B., Ranganathan, S. and Lin, L. (2019) Catalysis in Biodiesel Production—A Review. *Clean Energy*, **3**, 2-23. https://doi.org/10.1093/ce/zky020

[9] Uppenberg, J., Hansen, M.T., Patkar, S., and Jones, T.A. (1994) The Sequence, Crystal Structure Determination and Refinement of Two Crystal from *Candida antarctica*.

Structure, **2**, 293-308. https://doi.org/10.1016/S0969-2126(00)00031-9

[10] Chavan, A.S., Kharat, A.S., Bhosle, M.R., Dhumal, S.T. and Mane, R.A. (2022) Novel CAL-B Catalyzed Synthetic Protocols for Pyridodipyrimidines and Mercapto Oxadiazoles. *Journal of Chemical Sciences*, **134**, Article No. 120. https://doi.org/10.1007/s12039-022-02116-3

[11] Elliot, S.G., Andersen, C., Tolborg, S., Meier, S., Sádaba, I., Daugaard, A.E., *et al.* (2017) Synthesis of a Novel Polyester Building Block from Pentoses by Tin-Containing Silicates. *RSC Advances*, **7**, 985-996. https://doi.org/10.1039/c6ra26708d

[12] Passicos, E., Santarelli, X. and Coulon, D. (2004) Regioselective Acylation of Flavonoids Catalyzed by Immobilized Candida Antarctica Lipase under Reduced Pressure. *Biotechnology Letters*, **26**, 1073-1076. https://doi.org/10.1023/b:bile.0000032967.23282.15

[13] Hedfors, C., Hult, K. and Martinelle, M. (2010) Lipase Chemoselectivity towards Alcohol and Thiol Acyl Acceptors in a Transacylation Reaction. *Journal of Molecular Catalysis B*: *Enzymatic*, **66**, 120-123. https://doi.org/10.1016/j.molcatb.2010.04.005

[14] Tindal, R.A., Jeffery, D.W. and Muhlack, R.A. (2024) Nonlinearity and Anthocyanin Colour Expression: A Mathematical Analysis of Anthocyanin Association Kinetics and Equilibria. *Food Research International*, **183**, Article 114195. https://doi.org/10.1016/j.foodres.2024.114195

[15] Ware, K., Kashyap, P., Gorde, P.M., Yadav, R. and Sharma, V. (2025) Comparative Analysis of RSM and ANN-GA Based Modeling for Protein Extraction from Cotton Seed Meal: Effect of Extraction Parameters on Amino Acid Profile and Nutritional Characteristics. *Food and Bioproducts Processing*, **150**, 63-77. https://doi.org/10.1016/j.fbp.2024.12.016

[16] Yang, T., Lai, H., Cao, Z., Niu, Y., Xiang, J., Zhang, C., *et al.* (2022) Comparison of an Artificial Neural Network and a Response Surface Model during the Extraction of Selenium-Containing Protein from Selenium-Enriched *Brassica napus* L. *Foods*, **11**, Article 3823. https://doi.org/10.3390/foods11233823

[17] Baş, D. and Boyacı, İ.H. (2007) Modeling and Optimization II: Comparison of Estimation Capabilities of Response Surface Methodology with Artificial Neural Networks in a Biochemical Reaction. *Journal of Food Engineering*, **78**, 846-854. https://doi.org/10.1016/j.jfoodeng.2005.11.025

[18] Blanco, M., Coello, J., Iturriaga, H., Maspoch, S. and Porcel, M. (1999) Simultaneous Enzymatic Spectrophotometric Determination of Ethanol and Methanol by Use of Artificial Neural Networks for Calibration. *Analytica Chimica Acta*, **398**, 83-92. https://doi.org/10.1016/s0003-2670(99)00373-6

[19] Bryjak, J., Murlikiewicz, K., Zbiciński, I. and Stawczyk, J. (2000) Application of Artificial Neural Networks to Modelling of Starch Hydrolysis by Glucoamylase. *Bioprocess Engineering*, **23**, 351-357. https://doi.org/10.1007/s004499900170

[20] Geeraerd, A.H., Herremans, C.H., Ludikhuyze, L.R., Hendrickx, M.E. and Van Impe, J.F. (1998) Modeling the Kinetics of Isobaric-Isothermal Inactivation of Bacillus Subtilis α-Amylase with Artificial Neural Networks. *Journal of Food Engineering*, **36**, 263-279. https://doi.org/10.1016/s0260-8774(98)00064-8

[21] Manohar, B. and Divakar, S. (2005) An Artificial Neural Network Analysis of Porcine Pancreas Lipase Catalysed Esterification of Anthranilic Acid with Methanol. *Process Biochemistry*, **40**, 3372-3376. https://doi.org/10.1016/j.procbio.2005.03.045

[22] Talebian-Kiakalaieh, A., Amin, N.A.S., Zarei, A. and Noshadi, I. (2013) Transesterification of Waste Cooking Oil by Heteropoly Acid (HPA) Catalyst: Optimization and Kinetic Model. *Applied Energy*, **102**, 283-292.

https://doi.org/10.1016/j.apenergy.2012.07.018

[23] Wang, Z., Duan, H. and Hu, C. (2009) Modelling the Respiration Rate of Guava (*Psidium guajava* L.) Fruit Using Enzyme Kinetics, Chemical Kinetics and Artificial Neural Network. *European Food Research and Technology*, **229**, 495-503. https://doi.org/10.1007/s00217-009-1079-z

[24] Wei, S., Chen, Z., Arumugasamy, S.K. and Chew, I.M.L. (2022) Data Augmentation and Machine Learning Techniques for Control Strategy Development in Bio-Polymerization Process. *Environmental Science and Ecotechnology*, **11**, Article 100172. https://doi.org/10.1016/j.ese.2022.100172

[25] Moreno-Barea, F.J., Jerez, J.M. and Franco, L. (2020) Improving Classification Accuracy Using Data Augmentation on Small Data Sets. *Expert Systems with Applications*, **161**, Article 113696. https://doi.org/10.1016/j.eswa.2020.113696

[26] Shorten, C. and Khoshgoftaar, T.M. (2019) A Survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, **6**, Article No. 60. https://doi.org/10.1186/s40537-019-0197-0

[27] Ohno, H. (2020) Auto-Encoder-Based Generative Models for Data Augmentation on Regression Problems. *Soft Computing*, **24**, 7999-8009. https://doi.org/10.1007/s00500-019-04094-0

[28] Rezagholiradeh, M. and Haidar, M.A. (2018) Reg-Gan: Semi-Supervised Learning Based on Generative Adversarial Networks for Regression. 2018 *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), Calgary, 15-20 April 2018, 2806-2810. https://doi.org/10.1109/icassp.2018.8462534

[29] Segura-Carretero, A., Puertas-Mejía, M.A., Cortacero-Ramírez, S., Beltrán, R., Alonso-Villaverde, C., Joven, J., *et al.* (2008) Selective Extraction, Separation, and Identification of Anthocyanins from *Hibiscus sabdariffa* L. Using Solid Phase Extraction-Capillary Electrophoresis-Mass Spectrometry (Time-of-Flight/Ion Trap). *Electrophoresis*, **29**, 2852-2861. https://doi.org/10.1002/elps.200700819

[30] Rodríguez-Medina, I.C., Beltrán-Debón, R., Molina, V.M., Alonso-Villaverde, C., Joven, J., Menéndez, J.A., *et al.* (2009) Direct Characterization of Aqueous Extract of *Hibiscus sabdariffa* Using HPLC with Diode Array Detection Coupled to ESI and Ion Trap MS. *Journal of Separation Science*, **32**, 3441-3448. https://doi.org/10.1002/jssc.200900298

[31] Omotuyi, I.O., Elekofehinti, O.O., Ejelonu, O.C. and Obi, F.O. (2013) Mass Spectra Analysis of *H. sabdariffa L.* Anthocyanidins and Their In-Silico Corticosteroid-Binding Globulin Interactions. *Pharmacology Online. Archives*, **1**, 206-217.

[32] Yang, W., Kortesniemi, M., Ma, X., Zheng, J. and Yang, B. (2019) Enzymatic Acylation of Blackcurrant (*Ribes nigrum*) Anthocyanins and Evaluation of Lipophilic Properties and Antioxidant Capacity of Derivatives. *Food Chemistry*, **281**, 189-196. https://doi.org/10.1016/j.foodchem.2018.12.111

[33] Chebil, L., Humeau, C., Falcimaigne, A., Engasser, J. and Ghoul, M. (2006) Enzymatic Acylation of Flavonoids. *Process Biochemistry*, **41**, 2237-2251. https://doi.org/10.1016/j.procbio.2006.05.027

[34] Martinelle, M., Holmquist, M. and Hult, K. (1995) On the Interfacial Activation of Candida Antarctica Lipase A and B as Compared with Humicola Lanuginosa Lipase. *Biochimica et Biophysica Acta—Lipids and Lipid Metabolism*, **1258**, 272-276. https://doi.org/10.1016/0005-2760(95)00131-u

[35] Enaud, E., Humeau, C., Piffaut, B. and Girardin, M. (2004) Enzymatic Synthesis of New Aromatic Esters of Phloridzin. *Journal of Molecular Catalysis B: Enzymatic*, **27**, 1-6. https://doi.org/10.1016/j.molcatb.2003.08.002

[36] Turner, N.A. and Vulfson, E.N. (2000) At What Temperature Can Enzymes Maintain Their Catalytic Activity? *Enzyme and Microbial Technology*, **27**, 108-113. https://doi.org/10.1016/s0141-0229(00)00184-8

[37] Coulon, D., Girardin, M., Engasser, J.M. and Ghoul, M. (1997) Investigation of Keys Parameters of Fructose Oleate Enzymatic Synthesis Catalyzed by an Immobilized Lipase. *Industrial Crops and Products*, **6**, 375-381. https://doi.org/10.1016/s0926-6690(97)00028-9

[38] Cao, L., Bornscheuer, U.T. and Schmid, R.D. (1999) Lipase-Catalyzed Solid-Phase Synthesis of Sugar Esters. Influence of Immobilization on Productivity and Stability of the Enzyme. *Journal of Molecular Catalysis B*: *Enzymatic*, **6**, 279-285. https://doi.org/10.1016/s1381-1177(98)00083-6

[39] Chamouleau, F., Coulon, D., Girardin, M. and Ghoul, M. (2001) Influence of Water Activity and Water Content on Sugar Esters Lipase-Catalyzed Synthesis in Organic Media. *Journal of Molecular Catalysis B*: *Enzymatic*, **11**, 949-954. https://doi.org/10.1016/s1381-1177(00)00166-1

[40] Awolusi, T.F., Oke, O.L., Akinkurolere, O.O. and Atoyebi, O.D. (2019) Comparison of Response Surface Methodology and Hybrid-Training Approach of Artificial Neural Network in Modelling the Properties of Concrete Containing Steel Fibre Extracted from Waste Tyres. *Cogent Engineering*, **6**, 1-18. https://doi.org/10.1080/23311916.2019.1649852

[41] Nazerian, M., Kamyabb, M., Shamsianb, M., Dahmardehb, M. and Kooshaa, M. (2018) Comparison of Response Surface Methodology (RSM) and Artificial Neural Networks (ANN) towards Efficient Optimization of Flexural Properties of Gypsum-Bonded Fiberboards. *CERNE*, **24**, 35-47. https://doi.org/10.1590/01047760201824012484

[42] Optimization of Fermentation Medium Components by Response Surface Methodology (RSM) and Artificial Neural Network Hybrid with Genetic Algorithm (ANN-GA) for Lipase Production by Burkholderia Cenocepacia ST8 Using Used Automotive Engine Oil as Substrate. *Biocatalysis and Agricultural Biotechnology*, **50**, Article 102696. https://doi.org/10.1016/j.bcab.2023.102696

Highlights

- More than 80% conversion yield is achievable with both acyl donors, with the same optimum conditions for VL (from RSM and ANN) and MP from RSM.
- The nature of the substrate may influence the linearity of CAL-B activity, which in turn affects the suitability of different fitting algorithms for predicting conversion yields.
- Data augmentation, by increasing data volume, yields better fitting of neural networks.
- Accordance between experimental and predicted optimum values for MP (methyl palmitate) and VL (vinyl laurate) systems with RSM and ANN.