UDC 665:664.3

ARTIFICIAL INTELLIGENCE METHODS FOR MODELING AND OPTIMIZING PROCESS PARAMETERS OF A BIOCATALYTIC TECHNOLOGY FOR THE PRODUCTION OF HEALTH-PROMOTING STRUCTURED LIPIDS

https:// doi.org/ 10.15673/fst.v20i1.3358

P. Nekrasov[1], Doctor of Sciences in Engineering, Professor
O. Nekrasov[2], Candidate of Sciences in Engineering, Professor
N. Tkachenko[3], Doctor of Sciences in Engineering, Professor
T. Berezka[1], Candidate of Sciences in Engineering, Associate Professor
O. Gudz[1], Candidate of Sciences in Engineering, Associate Professor
S. Molchenko[1], Candidate of Sciences in Engineering, Associate Professor
[1]Department of technology of fats and fermentation products
[2]Department of physical chemistry
National Technical University "Kharkiv Polytechnic Institute"
2 Kyrpychova Str., Kharkiv, 61002, Ukraine
[3]Department of Technology of Milk, Oil and Fat Products and the Beauty Industry
Odesa National University of Technology
112 Kanatna Str., Odesa, 65039, Ukraine

Correspondence:
P. Nekrasov
E-mail: pavlo.nekrasov@khpi.edu.ua

Cite as Vancouver style citation

Artificial intelligence methods for modeling and optimizing process parameters of a biocatalytic technology for the production of health-promoting structured lipids /Nekrasov P. et al. Food Science and Technology. 2026;20(1):28-39 https:// doi.org/ 10.15673/fst.v20i1.3358

Цитування згідно ДСТУ 8302:2015

Artificial intelligence methods for modeling and optimizing process parameters of a biocatalytic technology for the production of health-promoting structured lipids /Nekrasov P. et al.// Food science and technology. 2026. Vol. 20 Issue 1. P. 28-39. https://doi.org/10.15673/fst.v20i1.3358

Abstract. This study substantiates and reliably demonstrates the application of artificial intelligence for determining the optimal parameters of a biocatalytic technology for producing structured lipids. The process involves the enzymatic acidolysis of sunflower oil with caprylic acid in a continuous packed-bed reactor. The biocatalyst selected was the immobilized lipase preparation Lipozyme RM IM (Novozymes, Denmark). This preparation consists of an sn-1,3-specific microbial lipase from Rhizomucor miehei, immobilized on a macroporous anion exchange resin. The optimization criterion was the maximal incorporation of caprylic acyl groups into the sn-1,3 positions of triacylglycerols. The selected control factors were: molar ratio of acid to oil (3:1 to 8:1), temperature (30–75 °C), and hydrodynamic residence time (15–60 min). The methodological framework of the research combined artificial intelligence models with an evolutionary optimization method, specifically the Particle Swarm Optimization algorithm. Experimental data were generated according to an orthogonal maximin Latin hypercube design and used to compare nine regression models employing nested cross-validation. It was established that support vector regression with a radial basis function kernel provided the best accuracy and the lowest inter-fold variability. The selected model was used as a fitness function within the Particle Swarm Optimization algorithm, enabling the determination of the optimal process conditions for enzymatic acidolysis: a molar ratio of caprylic acid to sunflower oil of 5:1, a temperature of 60 °C, and a residence time of 45 minutes. A verification experiment with five independent replicates confirmed the adequacy of the obtained optimum.

Key words: artificial intelligence; acidolysis; lipase; structured lipid; Particle Swarm Optimization algorithm

Introduction. Formulation of the problem

Modern industry is increasingly embracing the principles of "smart" manufacturing, where data, predictive modeling, and algorithmic process optimization play a pivotal role [1–5]. Many food technology processes are characterized by non-linearity, factor interactions, and high raw material variability, which complicates the application of purely empirical approaches and drives the adoption of intelligent methods for optimum-seeking [6–11].

Among such processes in the oils and fats industry, the production of structured lipids (SLs) holds a distinct place.

In a broad sense, structured lipids are lipids that have been chemically or otherwise modified from their original biosynthetic form. In a narrower, more commonly used sense, the term "structured lipids" refers to triacylglycerols (TAGs) that have been altered from

their natural state by incorporating new fatty acids or rearranging the acyl group positions within the molecule, as well as to artificially synthesized novel triacylglycerols [12–15]. SLs can be regarded as a new generation of fats that function as nutraceuticals – components of food products that enhance their nutritional value or confer health benefits, including the potential for disease prevention and treatment [16–19]. These properties are achieved by the presence of long-chain polyunsaturated fatty acids (PUFAs) and medium-chain fatty acids (MCFAs) within the SL structure [20–21].

Analysis of recent research and publications

In the human body, polyunsaturated fatty acids serve two primary functions: they are structural components of phospholipids in all cell membranes, upon which impulse transmission and receptor function depend, and they act as precursors for the synthesis of lipid mediators that are crucial for regulating numerous physiological processes [22]. Medium-chain fatty acids, when incorporated into SLs, are essential for rapid energy release and absorption, which is particularly important for infant nutrition, athletes, patients, and individuals with malabsorption disorders [23]. These characteristics stem from the smaller molecular size of these fatty acids and their greater water solubility compared to long-chain fatty acids. MCFAs are not incorporated into chylomicrons nor stored in the human body; instead, they are converted directly into energy. They undergo oxidation in the liver very rapidly (even faster than glucose) and represent a highly concentrated energy source [24].

Beyond the fatty acid composition of SLs, the positional distribution of acyl groups within the molecule constitutes a critical factor influencing their functionality. To enhance the absorption of any given fatty acid, it is esterified at the sn-2 position of the glycerol molecule [25]. Positioning PUFAs at the sn-2 position enables the full benefits of omega acids to be conferred to the lipid molecule, as PUFAs are hydrolyzed less efficiently by pancreatic lipase from the sn-1 and sn-3 positions compared to fatty acids with shorter carbon atoms. Concurrently, PUFAs are more effectively absorbed as monoacylglycerols when they occupy the sn-2 position [26]. This factor is of considerable importance in the production of structured lipids for both food and therapeutic purposes. While the consumption of polyunsaturated fatty acids is beneficial, their bioavailability in cases of impaired absorption is significantly enhanced when they are present as SLs with the PUFAs located at the sn-2 position. The chain length and degree of unsaturation of the fatty acids occupying the sn-1 and sn-3 positions define the rate of hydrolysis and fat absorption in the body. MCFAs have a shorter biochemical reaction time and are consequently hydrolyzed more rapidly by 1,3-positionally specific pancreatic lipase than are PUFAs [27].

A promising route for obtaining SLs is enzymatic acidolysis, which, owing to the 1,3-specificity of lipases, enables controlled modification of TAG structure and the formation of the desired triacylglycerol profile with fewer by-products compared to chemical approaches [28, 29]. The economic efficiency of this process is enhanced by modern immobilization technologies and the reuse of lipases, thereby expanding the potential for industrial implementation of biocatalytic fat and oil modification [30].

A key prerequisite for the industrial realization of enzymatic acidolysis is the selection of an appropriate reactor configuration and ensuring the stable operation of the immobilized biocatalyst. Packed-bed reactors have been proposed as a suitable solution [31] for continuous biocatalytic transformations due to their feasibility for repeated enzyme use, relatively straightforward process control, scalability, and the ability to maintain a high volumetric fraction of immobilized biocatalyst within the reaction system. This enhances specific productivity and promotes substrate conversion intensification. Furthermore, immobilized lipases in such reactors exhibit substantially enhanced stability, thereby prolonging their operational lifespan. The fixed spatial positioning of the enzyme minimizes its loss and simplifies product recovery, eliminating the need for additional processing steps and thus optimizing the overall technological workflow [32, 33].

However, the biocatalytic synthesis of SLs remains a non-trivial optimization problem due to its multiparametric nature and the potential existence of local extrema [34]. Under such conditions, it is advisable to integrate experimental design with predictive modeling and global optimization, aligning with the contemporary trend of employing artificial intelligence in food technology applications [7, 10, 11].

The methodological foundations of this approach have also been established in the authors' previous work, which addressed the biocatalytic processing of oil raw materials and the optimization of process parameters, kinetic analysis of enzymatic transformations, the application of neural network modeling and evolutionary algorithms for identifying rational operating regimes, and the development of fat systems with targeted property modifications (including oleogels and systems with reduced trans-isomer content) [35–40].

The objective of this study was to determine the optimal parameters for a biocatalytic technology for producing structured lipids, using the maximal incorporation of caprylic acyl groups into the sn-1,3 positions of the resulting triacylglycerols as the optimization criterion. This technology is based on the enzymatic acidolysis of sunflower oil with medium-chain caprylic acid ($C_{8:0}$) in a packed-bed bioreactor.

To achieve this objective, a combined application of artificial intelligence models and an evolutionary

optimization method – namely, the Particle Swarm Optimization (PSO) algorithm – was proposed for processing the experimental data.

This approach represents one of the most effective mathematical frameworks for optimizing complex, multiparameter functional dependencies [41–43], a category to which the biocatalytic acidolysis process belongs.

Nine artificial intelligence models were compared in this study: Multiple Linear Regression (MLR), Partial Least Squares Regression (PLSR), Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Regression with a radial basis function kernel (SVR), Kernel Ridge Regression (KRR), Extreme Gradient Boosting (XGB), Random Forest Regression (RFR), Extra Trees Regression (ETR), and Gradient Boosting Regression (GBR).

MLR, LASSO, and PLSR describe the response–factor relationship through a linear combination of predictors, ensuring high interpretability of parameters and enabling statistical analysis of individual variable contributions [44]. MLR, as a baseline approach, is appropriate as a reference model when the data structure is approximately linear; however, it may suffer from instability in the presence of multicollinearity or a large number of features. To enhance robustness, LASSO applies L1 regularization, which simultaneously reduces coefficient estimate variability and performs feature selection by shrinking some coefficients to zero. PLSR offers an alternative strategy for addressing multicollinearity: the original predictors are projected onto a low-dimensional latent variable space $T = XW$, constructed to maximize covariance with the response, after which regression is performed in the T space. This approach is particularly effective when variables are highly correlated and when the number of features is comparable to, or exceeds, the number of observations.

Support Vector Regression (SVR) is a highly effective kernel-based model. It is especially useful for handling complex datasets, as it efficiently manages cases with a high number of input features [45, 46]. The primary objective of SVR is to construct an optimal hyperplane in the feature space. To build a regression model, SVR utilizes a vector of target variable values $\in \mathbb{R}^n$ and vectors of input features $x_i \in \mathbb{R}^p$ (i = 1,…,n), as shown in Equation (1):

$$F(x) = w^T \varphi(x) + b \qquad (1)$$

where w is the vector of model weights, F is the function defined by the regression model, and b is the invariable.

Kernel Ridge Regression (KRR) combines the kernel method with ridge regression for modeling nonlinear processes [47]. KRR minimizes the sum of squared errors (SSE) with regularization, which helps prevent overfitting and ensures compliance with necessary constraints. The mathematical formulation of KRR is presented in Equation (2):

$$arg\ min\ f\ \frac{1}{m}\sum_{i=1}^{m}(f_i - y_i)^2 + \lambda\|f\|H \qquad (2)$$

where m is the dimension of the kernel matrix, λ is a hyperparameter; and f_i and y_i correspond to the observed and predicted values of the target variable.

Extreme Gradient Boosting (XGB) is a modern method incorporating tree-based models and gradient boosting [48]. Its effectiveness stems from the additive combination of multiple weak learners into a single strong learner. XGB employs autonomous parallel computing to enhance training accuracy and computational efficiency. The prediction function of XGB at a given time step t is presented in Equation (3):

$$f_i^t = \sum_{k=1}^{t} f_k(xi) = f_i^{t-1} + f_t(xi) \qquad (3)$$

where $f_i^{(t)}$ denotes the prediction at step t; $f_i^{(t-1)}$ is the prediction at the previous step; $f_t(x_i)$ is the new learner added at step t; x_i is the input feature vector, and $f_k(x_i)$ belongs to a specific functional space of models.

Random Forest Regression (RFR) employs a bootstrap aggregating (bagging) approach, which combines aggregation with bootstrapping [49]. Each decision tree in the construction of a random forest model receives its own random subsample of the input data (a bootstrap sample) and is trained independently of the other trees. The collective predictive strength obtained from individual decision trees significantly influences the accuracy of the random forest model. The prediction of the random forest model is given by Equation (4):

$$f(x) = \frac{1}{K}\sum_{k=1}^{K} DT_k(x) \qquad (4)$$

where K is the number of decision trees (DT) in the random forest, and f(x) is the function defined by the regression model.

The Extra Trees Regression (ETR) model is similar to random forest but further randomizes the position of split thresholds, which further reduces the variance of the ensemble. Each tree is built using the entire training sample, and splits are chosen from a random set of thresholds [50]. This results in high training speed and good generalization capability. However, the model offers lower interpretability and may exhibit underfitting on very smooth functions.

Gradient Boosting Regression (GBR) implements sequential construction of an additive model by approximating the negative gradient of the loss function. The prediction update at iteration t takes the form:

$$\hat{y}^{(t)}(x) = \hat{y}^{(t-1)}(x) + \eta\, f_t(x), \qquad (5)$$

where $f_t(x)$ is a base regressor (typically a decision tree), and η is the learning rate. At each step, the base model is trained on pseudo-residuals, i.e., the values of the negative gradient of the loss function computed for the current ensemble, which ensures progressive error reduction. GBR is capable of representing complex nonlinear relationships and factor interactions and often demonstrates high predictive accuracy [51]. Limitations include sensitivity to hyperparameter selection (number of trees, tree depth, η) and a propensity for overfitting in the absence of regularization.

PSO belongs to the family of swarm intelligence computational methods and is suitable for solving optimization problems involving continuous multimodal functions, with or without constraints. Its key advantage lies in operating directly in the space of continuous real variables without requiring gradient information of the objective function [52, 53]. The algorithm effectively balances global exploration of the search space with local investigation. Mimicking the collective behavior of a flock of birds or a school of fish, PSO employs a population of "particles" that iteratively search the solution space. Each particle corresponds to an individual candidate solution and progressively approaches the global optimum by dynamically adjusting its velocity and position. PSO is initialized by defining an initial population of particles randomly distributed throughout the search space. Particles are characterized by a position vector, corresponding to the current point under investigation in the hyperparameter space, and a velocity vector, which determines the direction and intensity of their movement during the iterative optimization process. At each iteration (or generation) of the algorithm, each particle adjusts its velocity and position based on its own experience and that of its "neighbors". This adjustment is governed by two key components: the best position known to the particle (personal best) and the best position found by any particle in the swarm (global best).

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot (p_{i,best} - x_i(t)) + c_2 \cdot r_2 \cdot (p_{g,best} - x_i(t)) \quad (6)$$

where $v_i(t)$ is the velocity of particle i at iteration t; w is the inertia weight, determining the influence of the previous velocity; c_1 and c_2 are acceleration coefficients responsible for the cognitive and social components, respectively; r_1 and r_2 are random variables uniformly distributed on the interval [0; 1]; $p_{i,best}$ is the personal best position of particle i; $p_{g,best}$ is the global best position found by any particle; and $x_i(t)$ is the position of particle i at iteration t.

After updating the velocities, the particle positions are recalculated using the following equation:

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (7)$$

The algorithm runs for a specified number of iterations or until a satisfactory solution quality is achieved.

Research materials and methods

Model mixtures consisted of sunflower oil and caprylic acid; the molar ratio of acid to oil was varied from 3:1 to 8:1 (substrate ratio, mol/mol).

The acidolysis process was conducted at temperatures ranging from 30 to 75 °C in a continuous packed-bed reactor containing the immobilized enzyme preparation Lipozyme RM IM (Novozymes, Denmark). This preparation is an sn-1,3-specific microbial lipase from *Rhizomucor miehei*, immobilized on a macroporous anion exchange resin.

The reactor consisted of a glass column with an internal diameter of 2.0 cm and a height of 20 cm, equipped with a jacket for thermostated water circulation to maintain a constant process temperature. A stainless steel mesh (50 mesh) served as the support for the immobilized lipase bed, which had a mass of 15.0 g. The formed packed enzyme bed exhibited a void fraction of 0.45.

The reaction medium containing sunflower oil and caprylic acid was prepared in a thermostated vessel equipped with a magnetic stirrer to ensure homogeneity. The mixture was fed to the bottom of the column using a peristaltic pump, with an upward flow direction to prevent bed compaction and channeling. The hydrodynamic residence time was varied within the range of 15–60 minutes, corresponding to a volumetric flow rate range of 0.4 to 1.5 cm³/min.

The selection of the enzyme preparation and the parameter variation intervals was based on the results of preliminary studies.

In each experiment, samples for subsequent analysis were collected after conditioning the packed enzyme bed with three void volumes of the substrate mixture.

Triacylglycerols (TAGs) present in the reaction products were initially isolated by thin-layer chromatography (TLC). For this purpose, collected samples were dissolved in chloroform, applied to a TLC plate, and eluted with a solvent system of petroleum ether/diethyl ether/acetic acid (90:10:1, v/v/v). After chromatography, the plate was air-dried, and the separated zones were visualized under ultraviolet light. The band corresponding to the TAG fraction was scraped from the plate, and TAGs were extracted three times with anhydrous diethyl ether. The diethyl ether was removed by nitrogen flushing until complete evaporation of the solvent. The dry residue of the isolated TAG fraction was weighed and used for subsequent analysis.

The fatty acid composition of the isolated TAG fraction was analyzed by gas-liquid chromatography following sample derivatization via methylation. A Clarus 500 Gas Chromatograph (Perkin-Elmer) equipped with a flame ionization detector was employed. The column was a Restek RT 2560 capillary column with the following specifications: length 100 m, internal diameter 0.25 mm, stationary phase thickness 0.2 μm. The stationary phase was bis-cyanopropyl polysiloxane. Chromatographic analysis conditions: temperature program – 140 °C (4 min), 8 °C/min to 179 °C (0 min), 4 °C/min to 240 °C (7 min); injector temperature – 250 °C; detector temperature – 270 °C. Hydrogen was used as the carrier gas at a flow rate of 3 cm³/min. All analyses were performed in duplicate. Peak identification and calibration were carried out using Merck standards.

To determine the fatty acid composition at the sn-2 position of triacylglycerols, hydrolysis with pancreatic

lipase was employed. The analysis was performed as follows. 5 mg of the isolated TAG fraction was mixed with 2 ml of 1 M Tris–HCl buffer (pH 7.6), 0.5 ml of 0.05% sodium cholate solution, and 0.2 ml of 2.2% calcium chloride solution; the mixture was then thoroughly mixed. Following the addition of pancreatic lipase (20 mg), the mixture was incubated at 40 °C for 2 min, vigorously agitated for 30 s, after which 5 ml of 6 M hydrochloric acid was added. Subsequently, the pancreatic lipase hydrolysis products were extracted twice with 15 ml of anhydrous diethyl ether and passed through a column containing anhydrous sodium sulfate. The hydrolysates were then applied to a TLC plate and separated using a hexane/diethyl ether/acetic acid (50:50:1, v/v/v) solvent system. The plate was air-dried, sprayed with 0.2% methanolic solution of 2,7-dichlorofluorescein, and the zones were visualized under ultraviolet light. The band corresponding to sn-2 monoacylglycerols was scraped from the plate, methylated using 14% boron trifluoride in methanol, and analyzed by gas-liquid chromatography as described above.

All analyses were performed in duplicate.

Based on the total caprylic acid content ($C_{8:0}$) in the triacylglycerols of the acidolysis product and its content at the sn-2 position, the incorporation of acyl groups of this fatty acid into the sn-1,3 positions of triacylglycerols (mol %) was calculated using the following equation:

$$Incorporation\ of\ C_{8:0}\ into\ sn-1{,}3-positions = \\ = \frac{3 \times \%\ total\ content\ of\ C_{8:0} - \%\ content\ of\ C_{8:0}\ at\ sn-2\ position}{2} \quad (8)$$

The obtained experimental data (mean values of duplicate analyses) were used as initial data for training and validating the artificial intelligence models.

The training set was constructed using *Orthogonal Maximin Latin Hypercube Sampling* (LHS), which enabled representative coverage of the multidimensional factor space with a limited number of experimental runs [54]. The range of each controlled factor was preliminarily stratified into *N* equiprobable subintervals (where *N* is the number of experimental points), after which one level of the corresponding factor was selected from each subinterval, and the complete design matrix was formed by permuting the levels across factors. This procedure ensured nearly uniform one-dimensional representation of each factor within its specified range and prevented excessive concentration of experimental points in local regions of the design space. To enhance the informativeness of the sample, the LHS configuration was optimized according to the maximin criterion by maximizing the minimum pairwise Euclidean distance between any two experimental points in the factor space. From the set of candidate LHS configurations, the one for which the minimum distance between points was the largest possible was selected, thereby reducing the formation of local clusters and minimizing the likelihood of gaps in the factor space, ensuring a more homogeneous spatial filling of the investigated domain. Additionally, factor orthogonality was ensured by minimizing the off-diagonal elements of the correlation matrix between factors. This reduced correlation among input variables and diminished the risk of multicollinearity in the training data, which, in turn, enhanced the numerical stability and reproducibility of artificial intelligence model training, reduced the sensitivity of quality estimates to random data partitioning during cross-validation, and ensured correct comparison of alternative artificial intelligence algorithms on an identical training set.

To mitigate the influence of systematic errors arising from external conditions, the sequence of experiments was randomized.

Feature standardization was applied for models sensitive to the scale of input variables (SVR, PLSR, LASSO, MLR, KRR) in order to unify factor scales and ensure correct and computationally efficient training in the presence of differences in data ranges. Transformation was performed using z-standardization (μ=0, σ=1):

$$x' = \frac{x-\mu}{\sigma} \quad (9)$$

where x is the original feature value, x' is the standardized value, μ is the feature mean, and σ is its standard deviation. To avoid data leakage during validation, the parameters μ and σ were calculated solely on the training subset of the corresponding split and applied to the test data within the processing pipeline. For tree-based ensembles (XGB, RFR, ETR, GBR), feature standardization was not applied, as these models are based on partitioning the feature space by threshold values, and linear scaling of variables does not affect the structure of such partitions or prediction quality.

To obtain an unbiased estimate of the predictive capacity of the models while simultaneously tuning their hyperparameters, nested cross-validation was employed. In the outer cross-validation loop, a k-fold scheme with k=6 was used: at each iteration, an outer-train set and an outer-test set were formed, with performance metrics evaluated on the latter. Hyperparameter selection was performed in the inner cross-validation loop using a k-fold split (k=5) via an exhaustive grid search procedure (GridSearchCV), conducted exclusively on the outer-train set. After selecting the optimal hyperparameters, the models were retrained on the entire outer-train set and evaluated on the outer-test set.

Model quality was assessed using standard regression metrics: the coefficient of determination R^2, root mean square error RMSE, and mean absolute error MAE:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\widehat{y_i})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (11)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (12)$$

where y_i is the experimentally measured response value for the i-th observation, $\hat{y}_i$ is the model-predicted response value for the i-th observation, $\bar{y}$ is the mean response value in the sample, n is the number of observations; and i=1,…,n is the observation index.

In the outer loop of the nested cross-validation, the values of the R^2, RMSE, and MAE metrics were calculated on each outer fold and summarized as the arithmetic mean ± standard deviation (mean ± std) across folds. Additionally, out-of-fold (OOF) predictions from the outer loop were generated, and aggregated estimates of R^2 (CV, OOF), RMSE (CV, OOF), and MAE (CV, OOF) were computed as single metric values for all observations predicted by models that had not encountered the corresponding observations during training.

Model comparison was performed based on minimization of RMSE and MAE and maximization of R^2, with emphasis on R^2 (CV, OOF) as a more stable aggregated estimate. This is because R^2 calculated on individual outer folds may assume negative values when the model on the corresponding fold predicts worse than a naive baseline predictor with a constant prediction $\hat{y} = \bar{y}_{train}$, where $\bar{y}_{train}$ is the mean response value on the training subset of the respective outer fold; aggregation via OOF predictions provides a more consistent generalized estimate of the metric by reducing the influence of inter-fold variability on the final R^2 value.

The model that yielded the best values of the selected quality metrics according to the nested cross-validation results was identified as the most suitable. Subsequently, final hyperparameter tuning was performed using an exhaustive grid search procedure (GridSearchCV), and the model was trained with the identified parameters on the complete dataset. The resulting model was then employed as a fitness function within the Particle Swarm Optimization (PSO) algorithm for process parameter optimization.

Modeling and optimization were carried out using the Python programming language, with the NumPy and pandas libraries for data processing, scikit-learn for model implementation, nested cross-validation, and metric evaluation, xgboost for XGB implementation, and matplotlib for plotting.

Results of the research and their discussion

Table 1 presents a fragment of the orthogonal maximin Latin Hypercube Sampling design matrix along with the experimental response values, which served as the initial data for training and cross-validation verification of the artificial intelligence models.

Results of the comparative assessment of the predictive capacity of the models are presented in Table 2.

Table 1 – Fragment of the orthogonal maximin Latin hypercube design matrix and experimental response values

Run order	LHS design point ID	Substrate ratio, mol/mol	Temperature, °C	Residence time, min	Incorporation of caprylic acid into sn-1,3-positions, mol %
1	36	7	32	48	39.8
2	27	3	33	38	40.4
3	34	4	70	57	73.3
4	22	4	40	42	63.7
5	17	7	35	30	35.9
6	18	4	47	26	55.5
7	35	4	46	15	21.6
8	40	8	64	35	73.9
9	9	4	39	16	12.6
10	32	7	50	60	65.7
…					
39	29	3	32	46	40.1
40	5	5	46	54	72.6
41	30	3	59	52	81.8
42	20	7	48	24	49.3
43	42	4	73	47	77.8
44	31	6	45	45	74.8
45	26	5	31	21	8.3
46	2	6	55	52	83.1
47	24	3	60	24	57
48	19	5	49	45	81.6

Table 2 – Summary model performance metrics from nested cross-validation: R^2, RMSE, and MAE (mean ± standard deviation across outer folds) and aggregated estimates computed from outer-loop out-of-fold predictions (CV, OOF)

Model	R^2 (CV, mean ± std)	RMSE (CV, mean ± std)	MAE (CV, mean ± std)	R^2 (CV, OOF)	RMSE (CV, OOF)	MAE (CV, OOF)	R^2 (Train)	RMSE (Train)	MAE (Train)
MLR	0.439 ± 0.416	12.42 ± 0.64	10.74 ± 0.97	0.659	12.43	10.74	0.701	11.63	10.00
PLSR	0.436 ± 0.409	12.56 ± 0.81	10.83 ± 1.14	0.650	12.59	10.83	0.701	11.63	10.01
LASSO	0.431 ± 0.411	12.61 ± 0.72	10.89 ± 1.15	0.648	12.62	10.89	0.701	11.63	10.00
SVR	0.997 ± 0.004	0.97 ± 0.44	0.66 ± 0.20	0.998	1.06	0.66	1.000	0.32	0.19
KRR	0.989 ± 0.007	1.99 ± 0.85	1.45 ± 0.51	0.990	2.13	1.45	0.999	0.66	0.52
XGB	0.914 ± 0.051	5.16 ± 1.32	4.03 ± 1.00	0.938	5.29	4.03	1.000	0.03	0.03
RFR	0.864 ± 0.079	6.48 ± 2.11	5.27 ± 1.61	0.899	6.76	5.27	0.988	2.38	1.83
ETR	0.919 ± 0.033	5.19 ± 1.20	3.93 ± 0.81	0.938	5.31	3.93	1.000	0.00	0.00
GBR	0.932 ± 0.045	4.53 ± 2.11	3.67 ± 1.63	0.947	4.92	3.67	1.000	0.45	0.37

Analysis of the data presented in Table 2 indicates that the linear models (MLR, PLSR, LASSO) exhibit insufficient and unstable predictive capability: the mean R^2 values across the outer folds are comparable in magnitude to their standard deviations, indicating substantial inter-fold variability in the estimates. Concurrently, their aggregated R^2 (CV, OOF) values at the level of ~0.65–0.66 are combined with high error metrics (RMSE ≈ 12.43–12.62; MAE ≈ 10.74–10.89), pointing to the inadequacy of the linear hypothesis for describing the investigated nonlinear dependence of the response on the factors.

Nonlinear models provide substantially higher prediction accuracy, with the best and simultaneously most stable result achieved by SVR with an RBF kernel: R^2 (CV, mean ± std) = 0.997 ± 0.004, exhibiting the lowest inter-fold variability among all models, and an aggregated R^2 (CV, OOF) = 0.998 accompanied by minimal errors RMSE (CV, OOF) = 1.06 and MAE (CV, OOF) = 0.66. Practically, this signifies that SVR reproduces the major portion of the experimental response variance and yields the smallest typical absolute prediction error, with results consistently reproduced across different outer splits.

A group of strong competitors is represented by the XGB, KRR, ETR, and GBR models. These demonstrate high aggregated R^2 (CV, OOF) ≈ 0.938–0.990 and moderate errors (RMSE ≈ 2.13–5.31; MAE ≈ 1.45–4.03). The RFR model exhibits a moderate level of generalization (R^2 (CV, OOF) = 0.899) with relatively higher errors (RMSE = 6.76; MAE = 5.27). At the same time, analysis of the "generalization gap" between training and cross-validation estimates indicates varying degrees of susceptibility to overfitting: for XGB, GBR, and tree-based ensembles in general, nearly perfect training metrics are observed, whereas cross-validation quality is noticeably lower, being a typical manifestation of high model flexibility and the need for careful regularization. For KRR and SVR, training performance is also high; however, their gap from cross-validation estimates is more moderate, indicating better management of the bias–variance trade-off and, consequently, more reliable predictive capability.

A visual interpretation of the results presented in Table 2 is provided in Fig. 1 in the form of "actual vs. predicted" plots for the out-of-fold predictions from the outer loop of the nested cross-validation.

The format of the diagrams in Fig. 1 allows for the assessment not only of the integral level of accuracy but also of the nature of the errors: the presence of systematic biases, the dependence of scatter on the response level, and the influence of isolated atypical observations. For the linear models, a noticeable "amplitude compression" effect is observed, meaning that predictions tend to gravitate toward mean values, with poorer reproduction of the extreme zones of the response range, manifesting as wide dispersion around the ideal fit line $y = x$. Nonlinear approaches form a substantially more compact band along $y = x$; however, they differ in the uniformity of agreement: for tree-based ensembles, areas of increased scatter of test points and isolated significant deviations are occasionally observed, indicating local sensitivity of predictions to individual observations and subsamples.

The most consistent appearance among the diagrams is demonstrated by SVR: both training and testing points predominantly adhere to the diagonal without pronounced systematic distortions across the entire response range, and deviations are more random than structural in nature. Combined with the summary metrics in Table 2, this indicates that SVR provides not only high average accuracy but also stable calibration of predictions based on the out-of-fold estimates from the outer loop of nested cross-validation. Consequently, support vector regression with a radial basis function kernel was substantiated as the selected baseline model and subsequently applied as the fitness function within the Particle Swarm Optimization (PSO) algorithm for optimizing the parameters of the enzymatic acidolysis process.

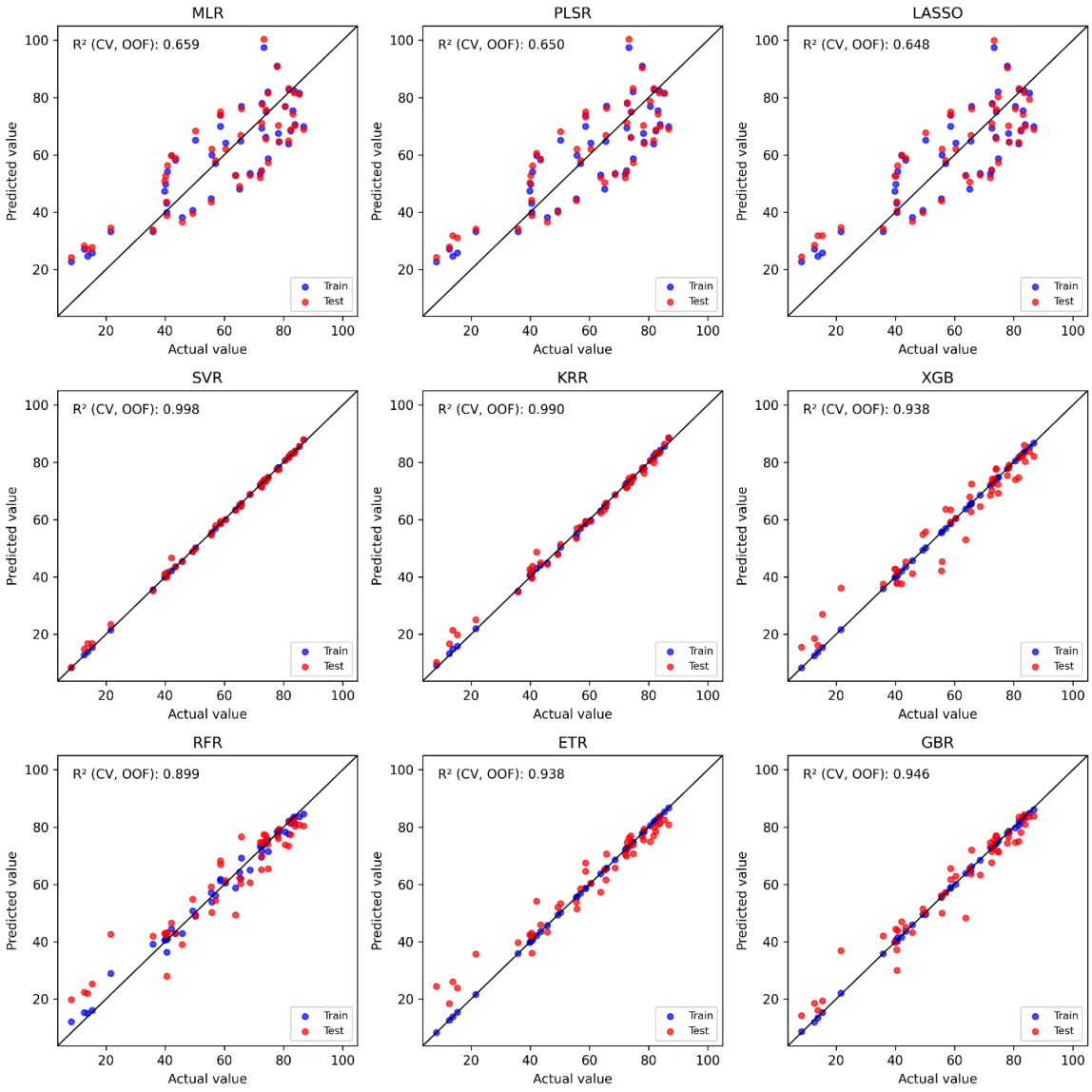

Fig. 1. Comparison of actual vs. predicted response values for the AI models based on nested cross-validation (out-of-fold predictions from the outer loop): blue markers denote training points and red markers denote test points; the solid line indicates the ideal agreement (y = x)

Subsequently, final hyperparameter tuning of SVR was performed on the complete experimental dataset using a pipeline consisting of StandardScaler → SVR and k-fold cross-validation (k=8) as the internal parameter selection procedure. Within the grid search, the following hyperparameter values were identified as optimal: C = 1000, γ = 0.03, ε = 0.01. The adequacy of the obtained SVR model was assessed both by cross-validation metrics and by the quality of reproduction of the available experimental data. The mean R^2 value in k-fold cross-validation was 0.998±0.002, consistent with the estimates obtained at the nested cross-validation stage (Table 2) and confirming the reproducibility of the model across different splits. After retraining the model with the optimal hyperparameters on the complete dataset, high reproduction accuracy (R^2=0.9998) and low prediction errors (RMSE=0.24, MAE=0.14) were achieved, with virtually no systematic bias (the mean residual value was −0.0104, and its standard deviation was 0.2459). Moreover, the maximum absolute error did not exceed 0.892. Additional indicators of prediction reliability revealed that 100% of observations exhibited $|e| \leq 1.0$ (where e is the prediction error: $e = y - \hat{y}$), which is important for subsequent optimization, as PSO concentrates the search in regions of the factor space with the highest predicted response.

The PSO algorithm was initialized with 40 particles and 100 generations; the inertial coefficient was set to $w = 0.7$, and the cognitive and social coefficients were set to $c_1 = c_2 = 1.5$ (random seed 42). This configuration represents a compromise between global exploration (sufficient number of particles and moderate inertia) and the ability to stabilize in the region of the maximum (symmetric c_1 and c_2 promote a balance between "individual experience" and "collective attraction" of the swarm).

The evolution of the best predicted response value and the narrowing of the factor ranges within the population are presented in Fig. 2.

The convergence behavior is typical for problems with a localized region of high response values: the global best value of the objective function, estimated by the SVR model, increases sharply during the first 8–10 generations, after which it reaches a plateau and subsequently remains virtually unchanged. This stabilization of the best prediction indicates the attainment of a stationary search regime and local refinement of the maximum coordinates without significant drift in the solution. The range plots (min–max) of the values for each controlled factor (substrate molar ratio, temperature, and residence time) further confirm the convergence of the swarm with respect to particle coordinates – from nearly complete coverage of the boundaries in the early generations to narrow intervals near the optimal values in the final phase of optimization.

The PSO results yielded the optimal combination of parameters: a molar ratio of caprylic acid to oil of 5 mol/mol, a temperature of 60 °C, and a residence time of 45 min, for which SVR predicts the maximum response – incorporation of caprylic acyl groups into the sn-1,3 positions of the resulting triacylglycerols at $\hat{y}$ = 88.8 mol %. For experimental validation of the modeling and PSO optimization results, an additional verification experiment was performed at the optimum point in five independent replicates with randomized run order. The experimental results yielded $\bar{y}_{exp} = 89.1 \pm 1.2$ mol % (mean ± SD), corresponding to an absolute deviation $|e|=0.3$ mol % and a relative deviation $\delta = 0.34$ %. This discrepancy is substantially smaller than the expected model error based on the out-of-fold estimates from the outer loop of nested cross-validation (Table 2), and the variability of the results across replicates was low (relative standard deviation, RSD, was 1.35%). This confirms the adequacy of SVR in the optimum region and the methodological correctness of employing the model as a fitness function within PSO for process parameter search.

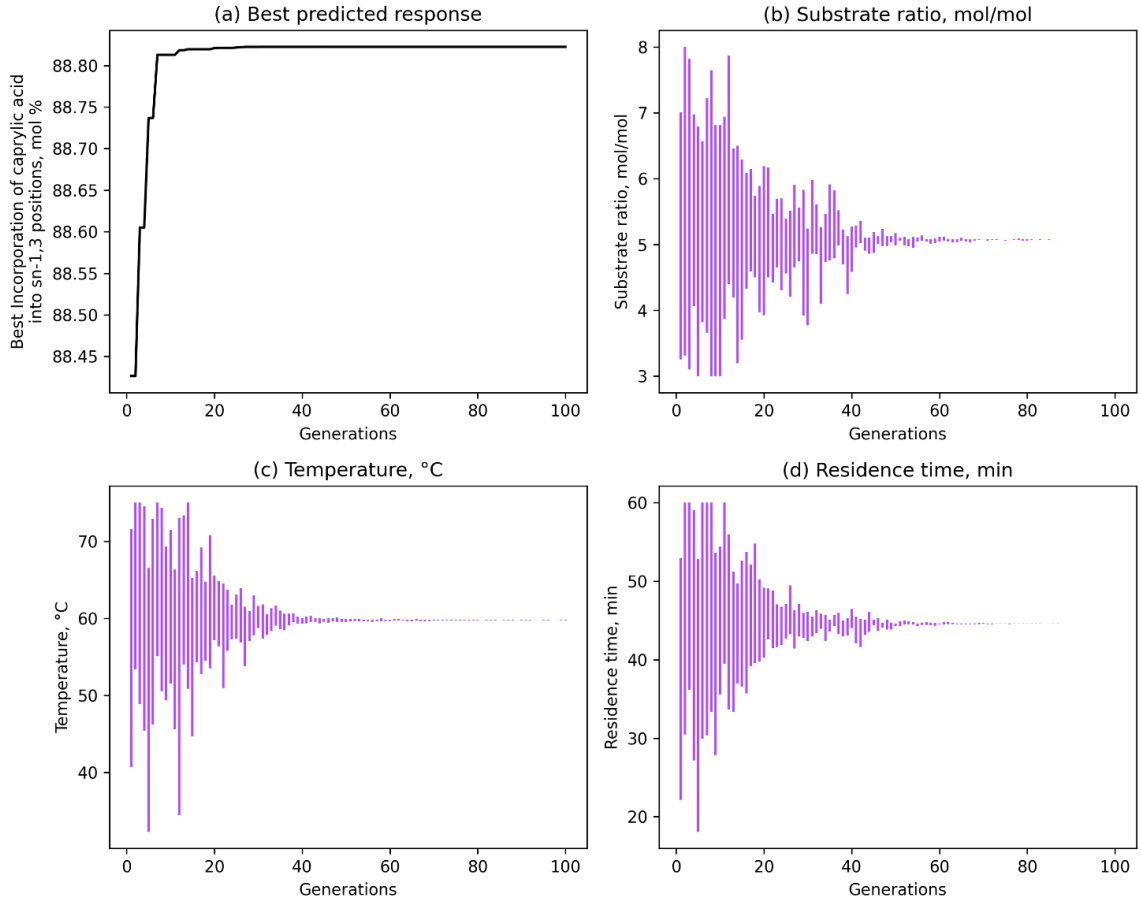

Fig. 2. Optimization of the biocatalytic technology for structured lipids production

Conclusion

The feasibility and effectiveness of employing artificial intelligence in the development of optimal operating regimes for food technologies have been demonstrated. This is confirmed through the case study of optimizing the enzymatic acidolysis of sunflower oil with caprylic acid for the production of structured lipids in a continuous packed-bed bioreactor.

Based on the results of nested cross-validation, SVR with an RBF kernel exhibited the highest and most stable predictive performance (R^2(CV, OOF)= 0.998; RMSE=1.06; MAE=0.66), substantiating its selection as the baseline model for the process.

The integration of SVR as a fitness function with global PSO optimization enabled the determination of the optimal conditions for the enzymatic interesterification process: a substrate molar ratio of 5 mol/mol, a temperature of 60 °C, and a residence time of 45 minutes.

A verification experiment conducted in five independent replicates confirmed the adequacy of the model in the optimum region: the incorporation of caprylic acyl groups into the sn-1,3 positions of the structured lipids was 89.1 ± 1.2 mol %, with a relative deviation δ = 0.34 %.

References:

1. Dantas TET, de-Souza ED, Destro IR, Hammes G, Rodriguez CMT, Soares SR. How the combination of circular economy and Industry 4.0 can contribute towards achieving the Sustainable Development Goals. Sustain Prod Consum. 2021;26:213-27. doi:10.1016/j.spc.2020.10.005.
2. Raven DB, Chikkula Y, Patel KM, Al Ghazal AH, Salloum HS, Bakhurji AS, et al. Machine learning and conventional approaches to process control and optimization: industrial applications and perspectives. Comput Chem Eng. 2024;189:108789. doi:10.1016/j.compchemeng.2024.108789.
3. Castro-Amoedo R, Granacher J, Maréchal F. A combined genetic algorithm and active learning approach to build and test surrogate models in process systems engineering. Comput Chem Eng. 2024;181:108517. doi:10.1016/j.compchemeng.2023.108517.
4. Hassoun A, Prieto MA, Carpena M, Bouzembrak Y, Marvin HJP, Pallarés N, et al. Exploring the role of green and Industry 4.0 technologies in achieving sustainable development goals in food sectors. Food Res Int. 2022;162(Pt B):112068. doi:10.1016/j.foodres.2022.112068.
5. Ghobakhloo M, Iranmanesh M, Tseng ML, Grybauskas A, Stefanini A, Amran A. Behind the definition of Industry 5.0: a systematic review of technologies, principles, components, and values. J Ind Prod Eng. 2023;40(6):432-47. doi:10.1080/21681015.2023.2216701.
6. Abass T, Itua EO, Bature T, Eruaga MA. Innovative approaches to food quality control: AI and machine learning for predictive analysis. World Journal of Advanced Research and Reviews. 2024;21(3):823-8. doi:10.30574/wjarr.2024.21.3.0719.
7. Esmaeily R, Razavi MA, Razavi SH. A step forward in food science, technology and industry using artificial intelligence. Trends Food Sci Technol. 2024;143:104286. doi:10.1016/j.tifs.2023.104286.
8. Harris N, Gonzalez Viejo C, Barnes C, Fuentes S. Non-invasive digital technologies to assess wine quality traits and provenance through the bottle. Fermentation. 2023;9(1):10. doi:10.3390/fermentation9010010.
9. Zhou Q, Zhang H, Wang S. Artificial intelligence, big data, and blockchain in food safety. Int J Food Eng. 2022;18(1):1-4. doi:10.1515/ijfe-2021-0299.
10. Barthwal R, Kathuria D, Joshi S, Kaler RSS, Singh N. New trends in the development and application of artificial intelligence in food processing. Innov Food Sci Emerg Technol. 2024;92:103600. doi:10.1016/j.ifset.2024.103600.
11. Bidyalakshmi T, Jyoti B, Mansuri SM, Srivastava A, Mohapatra D, Kalnar YB, et al. Application of artificial intelligence in food processing: current status and future prospects. Food Eng Rev. 2025;17:27-54. doi:10.1007/s12393-024-09386-2.
12. Alves V, de Figueiredo Furtado GDF, Roman MAS, Delboni LDMP, Alves Macedo JA, Vianna CLDC, et al. Palm oil-free structured lipids: a novel structuring fat for sandwich cookie fillings. Foods. 2026;15(1):178. doi:10.3390/foods15010178.
13. Witasari LD, Jason B, Salim JS, Pratama Y, Nusantoro BP. Enzymatic synthesis of zero-trans structured lipids from red palm oil (Elaeis guineensis) and sacha inchi oil (Plukenetia volubilis) blends. Food Biosci. 2025;69:106981. doi:10.1016/j.fbio.2025.106981.
14. Prasanna Rani KN, Kaki SS, Kezia Rani K, Kranthi Kumar B, Anjaneyulu E, Thirupathi A, et al. Production of healthier palm-based edible oil blends and enzymatic interesterified structured lipids for cooking and trans-free fat formulation applications. J Am Oil Chem Soc. 2024;101(3):269-82. doi:10.1002/aocs.12752.
15. Yuan T, Guo L, Gao X, Song G, Wang D, Li L, et al. Enzymatic synthesis and characterization of structured lipids: medium- and long-chain triacylglycerols enriched with lauric acid and diverse long-chain fatty acids. Food Biosci. 2025;64:105956. doi:10.1016/j.fbio.2025.105956.
16. Ijaz H, Sun S. A review on preparation and application of low-calorie structured lipids in food system. Food Sci Biotechnol. 2025;34(1):49-64. doi:10.1007/s10068-024-01689-8.
17. Guo Y, Cai Z, Xie Y, Ma A, Zhang H, Rao P, et al. Synthesis, physicochemical properties, and health aspects of structured lipids: a review. Compr Rev Food Sci Food Saf. 2020;19(2):759-800. doi:10.1111/1541-4337.12537.
18. Hong CR, Jeon BJ, Park KM, Lee EH, Hong SC, Choi SJ. An overview of structured lipid in food science: synthesis methods, applications, and future prospects. J Chem. 2023;2023:1222373. doi:10.1155/2023/1222373.
19. Mota DA, Rajan D, Heinzl GC, Osório NM, Gominho J, Krause LC, et al. Production of low-calorie structured lipids from spent coffee grounds or olive pomace crude oils catalyzed by immobilized lipase in magnetic nanoparticles. Bioresour Technol. 2020;307:123223. doi:10.1016/j.biortech.2020.123223.
20. Zou X, Jiang X, Wen Y, Wu S, Nadege K, Ninette I, et al. Enzymatic synthesis of structured lipids enriched with conjugated linoleic acid and butyric acid: strategy consideration and parameter optimization. Bioprocess Biosyst Eng. 2020;43(2):273-82. doi:10.1007/s00449-019-02223-5.
21. Ji S, Wu J, Xu F, Wu Y, Xu X, Gao H, et al. Synthesis, purification, and characterization of a structured lipid based on soybean oil and coconut oil and its applications in curcumin-loaded nanoemulsions. Eur J Lipid Sci Technol. 2020;122(10):2000086. doi:10.1002/ejlt.202000086.
22. Ji S, Xu F, Zhang N, Wu Y, Ju X, Wang L. Dietary a novel structured lipid synthesized by soybean oil and coconut oil alter fatty acid metabolism in C57BL/6J mice. Food Biosci. 2021;44:101396. doi:10.1016/j.fbio.2021.101396.

23. Tsuji K, Tsuchiya Y, Yokoi K, Yanagimoto K, Ueda H, Ochi E. Eicosapentaenoic acid and medium-chain triacylglycerol structured lipids improve endurance performance. Nutrients. 2023;15(17):3692. doi:10.3390/nu15173692.
24. Martínez-Galán JP, Ontibón-Echeverri CM, Campos Costa M, Batista-Duharte A, Guerso Batista V, Mesa Salgado VM, et al. Enzymatic synthesis of capric acid-rich structured lipids and their effects on mice with high-fat diet-induced obesity. Food Res Int. 2021;148:110602. doi:10.1016/j.foodres.2021.110602.
25. Bohorquez-Peña MJ, Vargas-Suaza B, Garzón-Alonso KE, Rendón-Londoño JC, Franco-Tobón YN, Gómez-Herrera CL, et al. Production of MLM-type structured lipids from avocado oil catalyzed by immobilized lipase on corn cob powder: assessment of antiproliferative potential on Hep-G2 cell line. Food Biosci. 2025;69:106824. doi:10.1016/j.fbio.2025.106824.
26. Yue C, Tang Y, Chang M, Wang Y, Peng H, Wang X, et al. Dietary supplementation with short- and long-chain structured lipids alleviates obesity via regulating hepatic lipid metabolism, inflammation and gut microbiota in high-fat-diet-induced obese mice. J Sci Food Agric. 2024;104(9):5089-5103. doi:10.1002/jsfa.13344.
27. Zhou Y, Xie Y, Wang Z, Wang C, Wang Q. Effects of a novel medium–long–medium-type structured lipid synthesized using a two-step enzymatic method on lipid metabolism and obesity protection in experimental mice. Food Sci Nutr. 2023;11(8):4516-29. doi:10.1002/fsn3.3410.
28. Yoshioka T, Kuroiwa T. Production of structured lipids via batch and continuous acidolysis of triglycerides using the hydration–aggregation-pretreated lipase loaded onto surface-modified glass bead supports. Process Biochem. 2024;141:82-89. doi:10.1016/j.procbio.2024.03.005.
29. Wan Y, Zhao L, Liu JN, Zhao Z, Tian J, Wu H, et al. Lipase-catalyzed acidolysis of walnut oil as a green strategy to synthesize structured lipids rich in α-linolenic acid: preparation and characterization. Ind Crops Prod. 2026;240:122652. doi:10.1016/j.indcrop.2026.122652.
30. Remonatto D, Miotti RH, Monti R, Bassan JC, de Paula AV. Applications of immobilized lipases in enzymatic reactors: a review. Process Biochem. 2022;114:1-20. doi:10.1016/j.procbio.2022.01.004.
31. Freitas AN, Remonatto D, Miotti Junior RH, do Nascimento JFC, da Silva Moura AC, Ebinuma V de C S. Adsorption of extracellular lipase in a packed-bed reactor: an alternative immobilization approach. Bioprocess Biosyst Eng. 2024;47:1735-49. doi:10.1007/s00449-024-03066-5.
32. Cozentino IDSC, Rodrigues MDF, Mazziero VT, Cerri MO, Cavallini DCU, de Paula AV. Enzymatic synthesis of structured lipids from grape seed (Vitis vinifera L.) oil in associated packed bed reactors. Biotechnol Appl Biochem. 2022;69(1):101-9. doi:10.1002/bab.2085.
33. Pacheco BJS, Andrade GSS, de Paula AV. Lipase from Rhizopus oryzae immobilized on corn cob powder: a new approach for dietetic triglyceride synthesis in a fixed bed reactor. Quim Nova. 2024;47(7):e20240025. doi:10.21577/0100-4042.20240025.
34. Delgado Naranjo JM, Jiménez Callejón MJ, Peñuela Vásquez M, Rios LA, Robles Medina A. Optimization of the enzymatic synthesis of structured triacylglycerols rich in docosahexaenoic acid at sn-2 position by acidolysis of Aurantiochytrium limacinum SR21 oil and caprylic acid using response surface methodology. J Appl Phycol. 2021;33(4):2031-45. doi:10.1007/s10811-021-02464-6.
35. Nekrasov PO, Berezka TO, Nekrasov OP, Gudz OM, Molchenko SM. Improving the efficiency of vegetable oil extraction by biocatalytic treatment of oilseed raw materials. Journal of Chemistry and Technologies. 2025;33(1):146-52. doi:10.15421/jchemtech.v33i1.322422.
36. Nekrasov PO, Berezka TO, Nekrasov OP, Gudz OM, Molchenko SM. Investigation of the basic laws of the kinetics of biocatalytic hydrolysis of vegetable oil. Journal of Chemistry and Technologies. 2024;32(1):191-7. doi:10.15421/jchemtech.v32i1.298927.
37. Nekrasov PO, Berezka TO, Nekrasov OP, Gudz OM, Molchenko SM, Rudneva SI. Optimization of the parameters of biocatalytic hydrolysis of vegetable oil using the methods of neural networks and genetic algorithms. Journal of Chemistry and Technologies. 2023;31(1):140-6. doi:10.15421/jchemtech.v31i1.274704.
38. Nekrasov PO, Berezka TO, Nekrasov OP, Gudz OM, Rudneva SI, Molchenko SM. Study of biocatalytic synthesis of phytosterol esters as formulation components of nutritional systems for health purposes. Journal of Chemistry and Technologies. 2022;30(3):404-9. doi:10.15421/jchemtech.v30i3.265174.
39. Nekrasov PO, Tkachenko NA, Nekrasov OP, Gudz OM, Berezka TO, Molchenko SM. Oxidization resistance and sorption of oleogels as new-generation fat systems. Voprosy khimii i khimicheskoi technologii. 2021;(4):89-95. doi:10.32434/0321-4095-2021-137-4-89-95.
40. Nekrasov PO, Gudz OM, Nekrasov OP, Kishchenko VA, Holubets OV. [Fatty systems with reduced content of trans-fatty acids]. Voprosy khimii i khimicheskoi technologii. 2019;(3):132-138. Ukrainian. doi:10.32434/0321-4095-2019-124-3-132-138.
41. Senthil H, Janve M. AI technologies shaping the future of the cocoa industry from farm to fork: a comprehensive review. Food Sci Biotechnol. 2025;34(14):3127-51. doi:10.1007/s10068-025-01848-5.
42. Guru Prasad MS, Naveen Kumar HN, Jain AK, Syed J, Baig RU. Convergence of improved particle swarm optimization based ensemble model and explainable AI for the accurate detection of food adulteration in red chilli powder. J Food Compos Anal. 2025;143:107577. doi:10.1016/j.jfca.2025.107577.
43. Yakoubi S. Sustainable revolution: AI-driven enhancements for composite polymer processing and optimization in intelligent food packaging. Food Bioprocess Technol. 2025;18(1):82-107. doi:10.1007/s11947-024-03449-2.
44. Kurtanjek Ž. Causal artificial intelligence models of food quality data. Food Technol Biotechnol. 2024;62(1):102-9. doi:10.17113/ftb.62.01.24.8301.
45. Li W, Yu J, Ren N, Huang L, Dang Y, Wu Y, et al. Exploration of the prediction and generation patterns of heterocyclic aromatic amines in roast beef based on genetic algorithm combined with support vector regression. Food Chem. 2025;463(Pt 1):141059. doi:10.1016/j.foodchem.2024.141059.
46. Wang Q, Han F, Wu Z, Lan T, Wu W. Estimation of free fatty acids in stored paddy rice using multiple-kernel support vector regression. Appl Sci (Basel). 2020;10(18):6555. doi:10.3390/app10186555.
47. Long T, Tang X, Liang C, Wu B, Huang B, Lan Y, et al. Detecting bioactive compound contents in Dancong tea using VNIR-SWIR hyperspectral imaging and KRR model with a refined feature wavelength method. Food Chem. 2024;460:140579. doi:10.1016/j.foodchem.2024.140579.
48. Wang L, Niu D, Zhao X, Wang X, Hao M, Che H. A comparative analysis of novel deep learning and ensemble learning models to predict the allergenicity of food proteins. Foods. 2021;10(4):809. doi:10.3390/foods10040809.
49. Moneravilla VB, Rodrigo D, Dharmaratne G, Purasingha S, Pushpakumara A, Abeyasinghe N, et al. Random forest regression-assisted Raman spectroscopy for authenticating the purity of virgin coconut oil. J Food Compos Anal. 2026;149:108784. doi:10.1016/j.jfca.2025.108784.

50. Lanjewar MG, Morajkar PP, Parab JS. Hybrid method for accurate starch estimation in adulterated turmeric using Vis-NIR spectroscopy. Food Addit Contam A. 2023;40(9):1131-46. doi:10.1080/19440049.2023.2241557.
51. Zheng R, Jia Y, Ullagaddi C, Allen C, Rausch K, Singh V, et al. Optimizing feature selection with gradient boosting machines in PLS regression for predicting moisture and protein in multi-country corn kernels via NIR spectroscopy. Food Chem. 2024;456:140062. doi:10.1016/j.foodchem.2024.140062.
52. Zhang Y, Yu H, Zhang H, Tang X. Bread staling prediction with a multiobjective particle swarm optimization-based bread constitutive modeling method. J Texture Stud. 2023;54(4):498-509. doi:10.1111/jtxs.12775.
53. Sarkar T, Salauddin M, Mukherjee A, Shariati MA, Rebezov M, Tretyak L, et al. Application of bio-inspired optimization algorithms in food processing. Curr Res Food Sci. 2022;5:432-50. doi:10.1016/j.crfs.2022.02.006.
54. James G, Witten D, Hastie T, Tibshirani R, Taylor J. An introduction to statistical learning: with applications in Python. 1st ed. Cham: Springer; 2023. 607 p. doi:10.1007/978-3-031-38747-0. ISBN: 978-3-031-38746-3.

МЕТОДИ ШТУЧНОГО ІНТЕЛЕКТУ В МОДЕЛЮВАННІ І ОПТИМІЗАЦІЇ ПАРАМЕТРІВ БІОКАТАЛІТИЧНОЇ ТЕХНОЛОГІЇ ВИРОБНИЦТВА СТРУКТУРОВАНИХ ЛІПІДІВ ОЗДОРОВЧОГО ПРИЗНАЧЕННЯ

П.О. Некрасов[1], доктор технічних наук, професор , *Email:* Pavlo.Nekrasov@khpi.edu.ua
О.П. Некрасов[2], кандидат технічних наук, професор , *Email:* Oleksandr.Nekrasov@khpi.edu.ua
Н.А. Ткаченко[3], доктор технічних наук, професор, *E-mail:* nataliya.n2013@gmail.com
Т.О. Березка[1], кандидат технічних наук, доцент, *Email:* Tetiana.Berezka@khpi.edu.ua
О.М. Гудзь[1], кандидат технічних наук, доцент, *Email:* Olha.Hudz@khpi.edu.ua
С.М. Мольченко[1], кандидат технічних наук, доцент
Email: Svitlana.Molchenko@khpi.edu.ua
[1]Кафедра технології жирів та продуктів бродіння
[2]Кафедра фізичної хімії
Національний технічний університет «Харківський політехнічний інститут»
вул. Кирпичова, 2, м. Харків, Україна, 61002
[3]Кафедра Технології молока, олійно-жирових продуктів та індустрії краси
Одеський національний технологічний університет
вул. Канатна, 112, м. Одеса, Україна, 65039

Анотація. У статті обґрунтовано та апробовано використання штучного інтелекту при встановленні оптимальних параметрів біокаталітичної технології виробництва структурованих ліпідів шляхом ферментативного ацидолізу соняшникової олії каприловою кислотою у безперервному реакторі з нерухомим шаром біокаталізатора, в якості якого було обрано ферментний препарат Lipozyme RM IM («Novozymes», Данія). Цей препарат є sn-1,3-специфічною мікробною ліпазою Rhizomucor miehei, іммобілізованою на макропористій аніонообмінній смолі. Критерієм оптимізації слугувало максимальне включення ацилів каприлової кислоти у sn-1,3-положення триацилгліцеринів. Як керовані фактори було обрано: мольне співвідношення кислоти до олії (від 3:1 до 8:1), температура (30–75 °C), та гідродинамічний час перебування (15–60 хв.). Методологічну основу дослідження становило поєднання моделей штучного інтелекту та еволюційного методу оптимізації, а саме алгоритму рою частинок. Експериментальні дані сформовано за ортогонально-максимінним планом латинського гіперкуба та використано для порівняння дев'яти регресійних моделей із застосуванням вкладеної крос-валідації. Встановлено, що найкращу точність і найменшу міжфолдову варіабельність забезпечує регресія на основі опорних векторів з радіально-базисним ядром. Обрану модель використано як функцію пристосованості в алгоритмі рою частинок, що дозволило визначити оптимальні умови процесу ферментативного ацидолізу: мольне співвідношення каприлової кислоти до соняшникової олії 5:1, температура 60 °C і час перебування 45 хвилин. Перевірочний експеримент у п'яти незалежних повторностях підтвердив адекватність отриманого оптимуму.

Ключові слова: штучний інтелект; ацидоліз; ліпаза; структурований ліпід; алгоритм рою частинок

Received 13.02.2026 **Reviewed 23.02.2026**
Revised 16.02.2026 **Approved 03.03.2026**