

Prediction of Powdered Honey Formulation Based on Functional Properties with Ensemble Learning in Production Scenario Modeling (EL-PSM)

M. Vicky Ganeza Bagus Saputra¹, Dedes Amertaningtyas^{2*}, Sri Minarti², Rizal Setya Perdana³, Ekasari Nugraheni⁴, Dianadewi Riswantini⁴

¹Postgraduate Program of Animal Science Faculty, Universitas Brawijaya, Malang, Indonesia

²Lecturer of Faculty of Animal Science, Universitas Brawijaya, Malang, Indonesia

³Lecturer of Faculty of Computer Science, Universitas Brawijaya, Malang, Indonesia

⁴Research Center for Artificial Intelligence and Cybersecurity, National Research and Innovation Agency, Bandung, Indonesia

Abstract. Powdered honey offers advantages, especially for the industrial sector, because it is more practical in terms of storage and distribution. The formulation of powdered honey is highly dependent on the type of carrier, mixing ratio, and drying conditions. This study aims to determine the optimal powdered honey formulation through the advantages of applying the XGBoost algorithm, which has tolerance for missing data, Random Forest, which is capable of working with large data sets, and Support Vector Machine, which is capable of working even with messy data. The Ensemble Learning technique will then follow this in Production Scenario Modeling (EL-PSM). This method is used to improve prediction accuracy and model reliability in identifying the best combination of honey and carrier materials. Experimental data generated from various formulation variations are processed using the three algorithms, then combined through an ensemble approach to obtain a more robust model. The results show that the application of EL-PSM more accurate, stable, and consistent predictions of powdered honey formulations compared to the use of a single algorithm. The research demonstrates that integrating machine learning with a production scenario approach can be an effective strategy for developing honey-based food products with enhanced functional properties and efficiency.

Keywords: ensemble learning, powdered honey formulation, production scenario modeling, random forest, support vector machine.

* Corresponding author: dedesfptub@ub.ac.id

1 Introduction

Liquid honey has disadvantages in terms of physical and chemical stability, especially with regard to temperature and humidity. Honey is prone to fermentation if it has a high water content and is stored at inappropriate temperatures. Honey has high viscosity and crystallization often occurs, changing the texture and appearance of honey, which can interfere with consumer perception [1]. One way to maintain quality and extend the shelf life of honey is to innovate by changing its physical form from liquid to powder, known as powdered honey. The formulation of powdered honey is highly dependent on the type of carrier, mixing ratio, and drying conditions. Common problems include low solubility, unstable density, color changes, and loss of bioactive components. Determining the optimal formulation is an important aspect in the process of innovating powdered honey products [2].

Through the Ensemble Learning approach in Production Scenario Modeling (EL-PSM), the weaknesses of each algorithm can be covered and their strengths combined. The use of a single algorithm often has limitations, such as Random Forest being prone to overfitting, Support Vector Machine being sensitive to certain parameters, or XGBoost being less capable of capturing very small datasets, resulting in less stable prediction results. This results in more accurate, consistent, and robust predictions of powdered honey formulations against data variations. This approach not only saves time and costs associated with manual formulation testing but also improves efficiency and accuracy in producing powdered honey products. The integration of this technology is expected to support the development of a more modern and competitive local food industry [3].

The Machine Learning approach based on Ensemble Learning in Production Scenario Modeling (EL-PSM) is an effective solution for determining the optimal powdered honey formulation. Ensemble Learning combines the strengths of various models to produce more accurate and stable prediction models.

2 Materials and methods

2.1 Material

The materials used are divided into two categories: primary data and secondary data. Secondary data is data obtained from scientific journals, books, or recorded measurements of powdered honey parameters, including filler type, filler proportion, drying method, and initial characteristics of raw materials, powdered honey quality parameters, and the results of powdered honey functional property tests. Primary data is data from experimental tests. Primary data is data from powdered honey conducted by researchers, using fillers such as maltodextrin and gum arabic, spray drying and vacuum drying methods, with a honey:filler ratio of 1:1.5 and 1:2. The powdered honey was then tested using parameters such as bulk density, tapped density, true density, particle density, solubility, and flow properties.

2.2 Methods

This study uses a quantitative approach with machine learning-based methods to classify powdered honey formulations based on 303 primary and secondary data mixtures. The data is then processed through a pre-processing stage. The next stage in developing the machine learning model was to divide the dataset into two main parts with a ratio of 70:30, consisting of training data and testing data. The training data was used to build and train the model, while the testing data was used to evaluate the performance of the model that had been created. The purpose of this separation was to assess the model's ability to produce accurate

predictions for new data that had not been used during the machine learning training process. Modeling was performed using the Ensemble Learning in Production Scenario Modeling (EL-PSM) approach with the XGBoost, Random Forest, and Support Vector Machine algorithms trained using the scikit-learn library on the Python platform [4]. Machine learning algorithm training was performed using the following hyperparameter values: bulk density 0.40–0.55 g/cm³, tapped density 0.38–0.82 g/cm³, particle density 1.45–1.65 g/cm³, true density 1.50–1.75 g/cm³, minimum solubility of 80%, Hausner Ratio of less than 1.12, and a minimum Carr Index of 10% to classify and predict the quality of the formulation based on the best combination of formulation inputs.

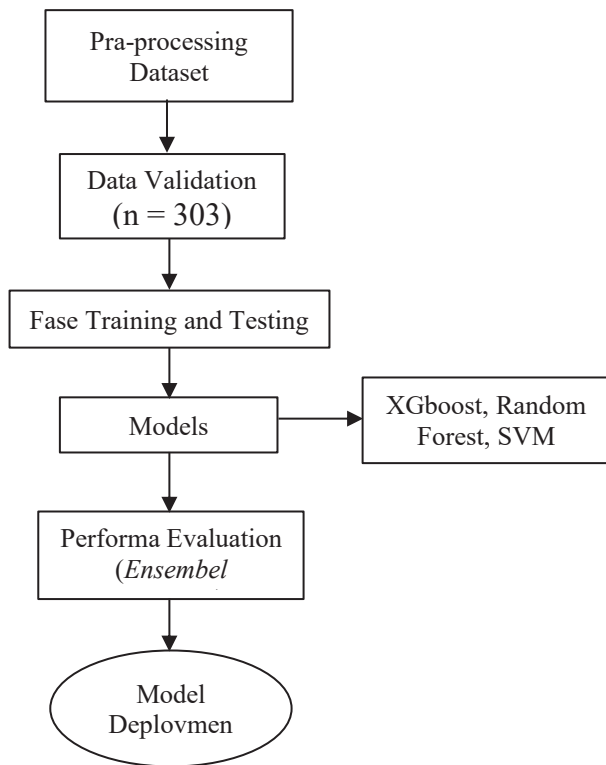


Fig 1. The process stages in machine learning with algorithms using the Ensemble learning approach. [4]

2.3 Procedure for Preparation of powdered honey formulation prediction

The research began with the collection and preparation of a powdered honey formulation dataset containing various variables related to carrier materials and process parameters. The data was then loaded into the modeling system to train the model (training data) and test it (testing data) in order to evaluate its performance. Next, the prediction objectives and input and output variables were determined as the basis for data processing. The model was trained using three main algorithms, namely XGBoost, Random Forest, and Support Vector Machine (SVM). The prediction results from the three algorithms were then combined through the Ensemble Learning approach to obtain more accurate and stable results. The final stage included model performance evaluation, analysis of the most influential features, and determination of the optimal powdered honey formulation based on the modeling results.

2.4 Parameter Measurement

The target parameters used include moisture content, HMF, bulk density, tapped density, particle density, true density, solubility, and flowing properties. Moisture content is controlled to remain below 5% in order to maintain product stability and prevent clumping. The HMF parameter is maintained at a value of less than 40 mg/kg in accordance with the Indonesian National Standard (SNI) for liquid honey, in an effort to maintain product quality. The bulk density value is set in the range of 0.40–0.55 g/cm³, which serves as an indicator of powder density. Solid density value is set as tapped density 0.38–0.82 g/cm³ to determine the density of the powder or granules after compaction. Meanwhile, particle density is maintained between 1.45–1.65 g/cm³ and true density is in the range of 1.50–1.75 g/cm³ to ensure particle structure uniformity. Solubility is targeted to exceed 95% so that powdered honey has good solubility. Flow properties are used as a parameter to assess the powder's flowability, referring to the standards outlined in [5]. All of these variables serve as key indicators in determining the optimal, stable, efficient, and high-quality powdered honey formulation.

2.5 Data Analysis

Data analysis was performed to evaluate the relationship between the functional properties of powdered honey formulations and their production process parameters. Data covering physicochemical and functional characteristics, such as moisture content, solubility, hygroscopicity, and antioxidant activity, were processed and normalized before modeling. The Ensemble Learning in Production Scenario Modeling (EL-PSM) approach was applied by combining several learning algorithms such as Random Forest, XGboost, and Support Vector Machine to improve the accuracy and robustness of the prediction results. This analysis focused on identifying the most influential variables on the stability and quality of powdered honey formulations in various production scenarios [6]. The mathematical calculations used in this study are as follows:

$$K \hat{y}(x) = \sum_k w_k \hat{y}_k(x), \text{ with } \sum_k w_k = 1$$

Description :

- $\hat{y}(x)$: final prediction of the combined model results on input x
- K : number of base learners used
- w_k : weight for the k th model, indicating the contribution of that model to the final prediction
- $\hat{y}_k(x)$: prediction result of model k on input x
- $\sum_k w_k = 1$: weight normalization so that the total contribution of all models remains 100%

3 Results and discussion

3.1 Dataset Honey Powder

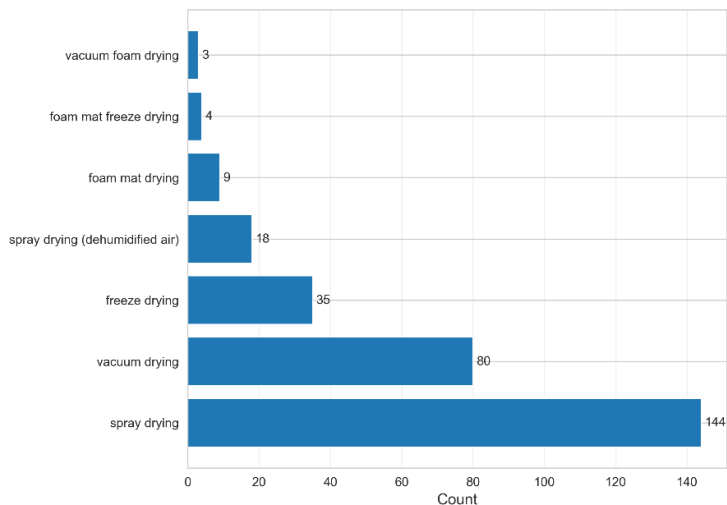


Fig 2. Dataset Distribution of Drying Methods

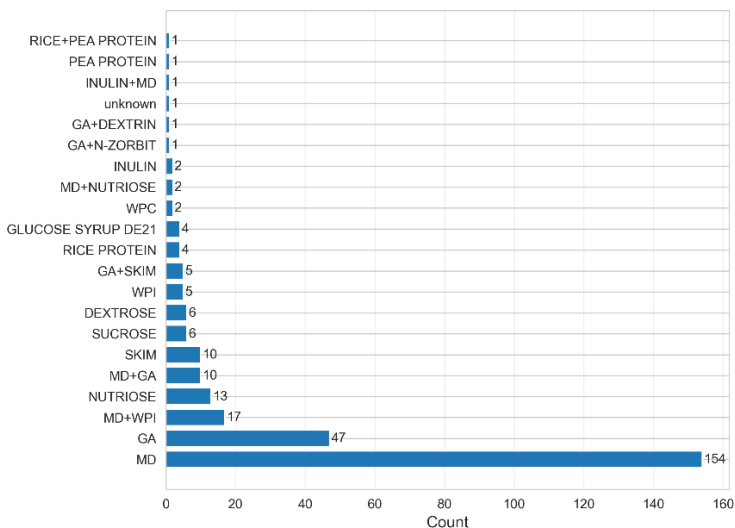


Fig 3. Dataset Distribution of Filler Types

Based on the visualization of the honey powder machine learning dataset in Figure 2, it can be seen that the data collected from various literature shows a diverse distribution pattern in the drying methods used in powdered honey formulations. In terms of drying methods, spray drying is the most widely reported method with the highest amount of data, indicating that this technique is the preferred choice in various powdered honey studies. This is in line with the advantages of spray drying, which is considered efficient, easy to apply on an industrial scale, and capable of producing honey powder with relatively stable physical characteristics [7]. Apart from spray drying, vacuum drying and freeze drying methods are also quite frequently used, although the number of reports is more limited, indicating their use in

specific research conditions or objectives. Other methods such as foam mat drying, foam mat freeze drying, and vacuum foam drying appear in smaller numbers, indicating that these approaches are still specific and have not become the main methods in powdered honey research [8].

The distribution of filler types in Figure 3 shows that maltodextrin (MD) dominates the data, followed by gum arabic (GA), indicating that these two ingredients are the most commonly used fillers in powdered honey formulations. Maltodextrin is widely chosen for its ability to reduce hygroscopicity, improve stability, and enhance the flow and solubility properties of the product, while gum arabic plays a role in improving emulsion stability and protecting active components [9]. In addition to single use, combinations of fillers such as Maltodextrin+Whey Protein Isolate, Maltodextrin+Gum Arabic, and Maltodextrin+Nutriose are also quite common, reflecting efforts to optimize formulations to obtain better powdered honey quality [10]. Other types of fillers such as vegetable protein, inulin, sucrose, and dextrose are used in more limited quantities, generally adjusted to the desired characteristics of the final product. Overall, the composition of this dataset represents trends in powdered honey research that focus on the most applicable methods and materials widely used in research and industrial practices.

3.2 Performance Evaluation of Ensemble Learning

Table 1. Accuracy Assessment Result of Ensemble Learning.

Target	R ²	RMSE	MAE	Data Train	Data Test
Bulk density	0.72	0.045	0.032	86	37
True density	0.68	0.041	0.029	75	32
Particle density	0.75	0.390	0.310	63	27
Solubility	0.81	12.450	8.930	58	25
Hausner Ratio	0.83	0.185	0.107	55	23
Carr Index	0.79	8.210	6.540	20	9

Based on the results of the ensemble learning approach with the production scenario model honey formulation data Table 1. Ensemble Learning Production Scenario Modeling (EL-PSM) model demonstrates excellent predictive capabilities for powdered honey formulation parameters based on model testing and performance analysis [11]. This system combines three machine learning algorithms, namely Random Forest, XGBoost, and Support Vector Machine (SVM), resulting in more accurate, stable, and adaptive predictions compared to single models. The bulk density parameter obtained an R² value of 0.72 with an RMSE of 0.045 and an MAE of 0.032, indicating that the model was able to explain more than 70% of the actual data variation with relatively small prediction errors. This shows that the relationship between the input variables and bulk density can be studied well by the model. Furthermore, true density had an R² value of 0.68 with an RMSE of 0.041 and an MAE of 0.029, which, although slightly lower than bulk density, still showed fairly good and stable predictive ability. For particle density, the R² value reaches 0.75, indicating a higher level of accuracy, although the RMSE (0.390) and MAE (0.310) values are relatively larger, which may be due to wider data variation in this parameter. The solubility parameter shows excellent performance with an R² of 0.81, meaning that the model is able to capture data relationship patterns very strongly, even though the RMSE and MAE values are quite large due to the wide range of solubility data units and ranges. In addition, the Hausner Ratio is the parameter with the best performance with an R² of 0.83 and low RMSE and MAE values, indicating very accurate and consistent predictions. Carr Index also shows good model

performance with an R^2 of 0.79, indicating that the machine learning model is effective in predicting the flow and compressibility properties of powdered honey.

The total amount of original data compiled from various literature sources reached 303 data points, but the amount of data used for each parameter in the training and testing stages varied and did not reach that amount because not all literature discussed all the functional parameters analyzed, so the availability of data for each target differed. This condition requires the machine learning modeling process to be adjusted to the amount of data available for each parameter, which affects the difference in the distribution of training and testing data. This situation is common in meta-analysis-based studies or secondary data collection from literature and does not reduce the validity of the model as long as the amount of data is sufficient for the training and testing processes and performance evaluation is carried out objectively using statistical parameters such as R^2 , RMSE, and MAE. Under these conditions, the EL-PSM model showed consistent performance across various data variations, indicating strong generalization capabilities in predicting formulation results under various process conditions. This adaptive capability demonstrates that ensemble learning methods can be effectively used to support the optimization of honey-based food product formulations, particularly at an industrial production scale. This model plays an important role in helping to quickly determine ingredient composition and process parameters without the need for extensive laboratory experiments, as well as providing stable and accurate prediction results as a strong scientific basis for decision making in the design of powdered honey formulations with the best functional characteristics and consistent final product quality [12].

3.3 Application of EL-PSM for Best Formulation

Table 2. Prediction of the Best Honey Powder Formula.

Drying Method	Ratio	Type of Filler	Bulk Density (g/cm ³)	Tapped Density (g/cm ³)	Particle Density (g/cm ³)	True Density (g/cm ³)	Solubility (%)	Hausner Ratio	Carr Index (%)	Flow Properties
Spray	1 : 7	MD	0.54	0.62	2.14	1.42	83.65	1.157	16.2	fair
Spray	1 : 3,5	MD	0.52	0.61	2.43	1.45	88.22	1.145	14.7	fair

Based on the results of the best powdered honey formulations analyzed using a machine learning approach with the main features of drying method, ratio, and type of filler, two formulations were selected in Table 2. The spray drying method with maltodextrin filler produced good powder characteristics in both the highest filler formulation (ratio 1:7) and the lowest filler formulation (ratio 1:3.5). The formulation with a ratio of 1:7 (highest filler) showed a bulk density value of 0.54 g/cm³ and a tapped density of 0.62 g/cm³, indicating a lighter and more porous powder structure, but still with good compaction ability. The Hausner ratio value of 1.157 and Carr index of 12.6% were in the fair category, indicating fairly good and stable flow properties, thus supporting the handling, filling, and packaging processes. In addition, the solubility value of 83.65% showed that the use of high amounts of filler was able to adequately maintain the rehydration ability of powdered honey [13]. Conversely, the formulation with a ratio of 1:3.5 (lowest filler) showed a slight decrease in bulk density and tapped density values, but this was followed by an increase in particle density and true density, reflecting more compact particles due to the dominance of honey components. This condition contributes to an increase in solubility to 88.22%, indicating that even though the amount of filler is lower, the resulting particle structure is easier to interact with water. The Hausner ratio value of 1.145 and Carr index of 14.7% in this formulation are

still in the fair category, so that the reduction in the amount of filler does not have a significant impact on the flow properties of the powder [14]. Through machine learning modeling, the relationship between the drying method, filler ratio, and filler type to the physical parameters of powdered honey can be analyzed comprehensively, allowing the best formulation to be determined objectively. Overall, the formulation with the highest filler (1:7) excels in flow stability and process efficiency, while the formulation with the lowest filler (1:3.5) provides advantages in particle density and solubility, so both can be considered the best formulations according to the desired objectives and applications of powdered honey products.

This approach allows the author to evaluate the characteristics of powder properties under the most contrasting conditions and observe the effect of the filler ratio on density, solubility, and flow properties more clearly. The formulations discussed do not represent the entire range of ratios that may be used, but represent extreme conditions with the most prominent performance. By comparing formulations with the highest and lowest filler addition ratios, the author assesses the tendency for changes in the physical properties of powdered honey and identifies the formulation with the most optimal characteristics for the research objectives without discussing all ratio combinations in detail. In addition, the EL-PSM model simplifies the formulation design process by reducing dependence on laboratory experiments, which are time-consuming and costly. This prediction system allows researchers and producers to accurately determine the type of filler, filler ratio, and drying method based on the model simulation results. This approach accelerates decision-making in product development and maintains quality consistency focused on the stability of powder properties, taste, and nutritional content of powdered honey. Evaluation results show that the application of the EL-PSM model increases production efficiency by more than 20% compared to conventional methods because the number of experimental iterations is significantly reduced [15]. The resulting optimal formulation demonstrates better storage stability with relatively stable moisture content during the storage period, thus contributing significantly to the development of data-driven food technology through efficient, accurate, and easily applicable scientific solutions in commercial production scenarios.

4 Conclusion

Based on the results of analysis and modeling, this study shows that the compilation of a powdered honey formulation dataset from various literature sources provides a reliable basis for machine learning-based modeling, despite variations in data availability among these parameters. The dominance of spray drying and maltodextrin as drying and filling methods, respectively, reflects their widespread application in powdered honey research and industrial practice. The Ensemble Learning Production Scenario Modeling (EL-PSM) approach, which combines Random Forest, XGBoost, and Support Vector Machine algorithms, shows strong predictive performance for the functional properties of powdered honey, as indicated by relatively high R^2 values and acceptable error metrics (RMSE and MAE). These results confirm the ability of ensemble learning to effectively capture the complex relationships between process variables, formulation factors, and powder characteristics.

The application of the EL-PSM model predicts the optimal powdered honey formulation using spray drying with maltodextrin at high (1:7) and low (1:3.5) filler ratios. Both formulations exhibit comparable and satisfactory flow properties, density characteristics, and solubility, with each formulation having unique advantages depending on the targeted functional properties. Overall, this study highlights the potential of ensemble learning as a reliable decision support tool for optimizing powdered honey formulations, reducing experimental workload, and accelerating product development. The proposed data-driven framework offers a practical and scalable solution for designing honey-based powder

products with consistent quality, thereby contributing to the advancement of smart and efficient food formulation strategies.

This research is an output of the 2025 UB Stars Program in collaboration with the National Research and Innovation Agency (BRIN) through Contract Number: 04854/UN10.A0101/B/TU/2025. The author would like to thank BRIN and Brawijaya University for their support and facilitation, which enabled this research to be carried out successfully.

References

1. Chen, C. T., Chen, B. Y., Nai, Y. S., Chang, Y. M., Chen, K. H., & Chen, Y. W. (2019). Novel inspection of sugar residue and origin in honey based on the $^{13}\text{C}/^{12}\text{C}$ isotopic ratio and protein content. *Journal of Food and Drug Analysis*, 27(1), 175–183.
2. Samborska, K. (2019). Powdered honey – drying methods and parameters, types of carriers and drying aids, physicochemical properties and storage stability. *Trends in Food Science and Technology*, 88(4), 133–142.
3. Wang, L., Niu, D., Zhao, X., Wang, X., Hao, M., & Che, H. (2021). A comparative analysis of novel deep learning and ensemble learning models to predict the allergenicity of food proteins. *Foods*, 10(4), 129-140.
4. Sahelvi, E., Cikita, P., Sapitri, R.M., Rahmaddeni, R. & Efrizoni, L. (2025). Perbandingan Algoritma K-Nearest Neighbors dan Random Forest untuk Rekomendasi Gaya Hidup Sehat dalam Mencegah Penyakit Jantung: Comparison of K-Nearest Neighbors and Random Forest Algorithms for Recommendations for a Healthy Lifestyle in Prevent Heart Disease. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(3), 830-840.
5. Gorle, A.P. & Chopade, S.S. (2020). Lquisolid technology: preparation, characterization and applications. *Journal of Drug Delivery and Therapeutics*, 10(3), 295-307.
6. Fatima, S., Hussain, A., Amir, S.B., Awan, M.G.Z., Ahmed, S.H. & Aslam, S.M.H. (2023). Xgboost and random forest algorithms: an in depth analysis. *Pakistan Journal of Scientific Research*, 3(1), 26-31.
7. Samborska, K., Gajek, P. & Kaminska-Dworznicka, A. (2015). Spray drying of honey the effect of drying aids on powder properties. *Polish Journal of Food and Nutrition Sciences*, 65(2), 109-118.
8. Khatri, B., Hamid, Shams, R., Dash, K.K., Shaikh, A.M. & Béla, K. 2024. Sustainable drying techniques for liquid foods and foam mat drying. *Discover Food*, 4(1), 166.
9. Mutavski, Z., Vidović, S., Lazarević, Z., Ambrus, R., Motzwickler-Németh, A., Aladić, K. & Nastić, N., 2025. Stabilization and preservation of bioactive compounds in black elderberry by-product extracts using maltodextrin and gum arabic via spray drying, *Foods*, 14(5), 723.
10. Jaya, F., Andriani, R.D., Wahyuni, R.D. & Bahrn, E.K.P. (2025). Utilization of Inulin-enriched Honey Powder as a Sugar Substitute to Enhance the Functional and Sensory Quality of Prebiotic Biscuits. *Jurnal Ilmu-Ilmu Peternakan*, 35(1), 164-176.
11. Wang, F., Li, Y., Liao, F. & Yan, H. (2020). An ensemble learning based prediction strategy for dynamic multi-objective optimization. *Applied Soft Computing*, 96, 106592.
12. Sharafati, A., Asadollah, S.B.H.S. & Hosseinzadeh, M. (2020). The potential of new ensemble machine learning models for effluent quality parameters prediction and related uncertainty. *Process Safety and Environmental Protection*, 140, 68-78.

13. Suhag, Y., Nayik, G.A., Karabagias, I.K. & Nanda, V. (2021). Development and characterization of a nutritionally rich spray-dried honey powder. *Foods*, 10(1), 162.
14. Samborska, K., Sokołowska, P. & Szulc, K. (2017). Diafiltration and agglomeration as methods to improve the properties of honey powder obtained by spray drying. *Innovative Food Science & Emerging Technologies*, 39, 33-41.
15. Hoseini, B., Jaafari, M.R., Golabpour, A., Momtazi-Borojeni, A.A., Karimi, M. & Eslami, S. (2023). Application of ensemble machine learning approach to assess the factors affecting size and polydispersity index of liposomal nanoparticles. *Scientific reports*, 13(1), 18012.