



Determination of Optimal Extraction Conditions of Pine Honey Using a Pi-Sigma Artificial Neural Network–Genetic Algorithm Approach

Huseyin SAHIN*¹ , Erol EGRIOGLU² , Eren BAS² 

¹Department of Basic Science, Faculty of Engineering, Giresun University, Giresun, 28200, Türkiye

²Department of Data Science and Analytics, Faculty of Arts and Science, Giresun University, Giresun, 28200, Türkiye

• Received: 13.03.2026

• Accepted: 16.05.2026

Abstract

Pine honey is a high-value natural product distinguished by its unique phenolic composition and diverse bioactive properties. Among these compounds, protocatechuic acid has been identified as a major and characteristic phenolic marker in pine honey samples originating particularly from the Aegean region of Türkiye. Despite its importance, systematic optimization of extraction conditions for this compound from pine honey matrices remains limited. Therefore, the present study aimed to determine the optimal extraction conditions for protocatechuic acid using an artificial intelligence–based optimization framework integrating Pi-Sigma artificial neural networks (PS-ANN) with a genetic algorithm (GA). In this approach, critical extraction parameters, including solvent type, extraction temperature, extraction time, and medium acidity, were evaluated. Experimental data obtained at different parameter levels were used to construct a response surface through the PS-ANN model, enabling the modeling of nonlinear relationships between extraction variables and protocatechuic acid yield. The optimal extraction conditions were subsequently identified using the GA optimization strategy. This hybrid AI-based modeling and optimization approach provides a robust and efficient tool for identifying optimal extraction conditions in complex food matrices. The findings demonstrate that the integration of PS-ANN and GA not only improves prediction accuracy but also significantly enhances the efficiency of phenolic compound extraction processes. Overall, this study highlights the potential of AI-driven optimization strategies for advancing analytical methodologies and improving the recovery of bioactive compounds from natural products such as pine honey.

Keywords: Genetic algorithm, Pine honey, Pi-sigma artificial neural network, Protocatechuic acid, Response surface methodology

1. INTRODUCTION

Türkiye plays a dominant role in global pine honey production (nearly 90%), accounting for a significant proportion of the world's supply (Uçurum et al., 2023). Pine honey is a highly valued natural product due to its unique chemical composition and the wide range of bioactive properties attributed to its phenolic constituents (Kahraman et al., 2024). Previous studies have demonstrated that phenolic compounds in honey contribute to various biological activities such as antioxidant, antibacterial, antiradical, and anti-inflammatory effects (Kahraman et al., 2024; Eraslan et al., 2010; Tsavea et al., 2022). The concentration and composition of these bioactive components can vary depending on several factors, including honey type, geographical origin, harvesting period, floral diversity, and surrounding vegetation (Joshi et al., 2000; Küçük et al., 2007; Silici et al., 2010; Kolaylı et al., 2025).

Honey can generally be classified into two main types based on its origin. The first type is blossom honey, which is produced from the nectar of flowering plants, while the second type is honeydew honey, which originates from the secretions of plant-sucking insects belonging to the order Hemiptera that live on plants (Sahin et al., 2022; Ülgentürk et al., 2013). Pine honey belongs to the honeydew honey group and is mainly produced from secretions of *Marchalina hellenica*, an insect species that feeds on pine trees (Sahin et al., 2022). Due to its distinctive chemical profile and higher mineral and phenolic content compared to many blossom honeys, pine honey has attracted considerable scientific attention in recent years (Tsavea et al., 2022).

*Huseyin SAHIN; huseyin.sahin@giresun.edu.tr

Among the phenolic compounds identified in pine honey, protocatechuic acid has been reported as a major and characteristic phenolic component, particularly in samples obtained from Türkiye (Sahin et al., 2022). The accurate extraction and quantification of this compound are therefore important for understanding the bioactive profile of pine honey and for improving analytical approaches used in honey characterization studies (Sahin et al., 2025). However, the extraction efficiency of phenolic compounds is influenced by several experimental parameters, including solvent type, extraction temperature, extraction time, and medium acidity (Taibi et al., 2025). Determining the optimal combination of these parameters represents a complex optimization problem because of the nonlinear interactions among variables.

Traditionally, statistical hypothesis tests such as the Kruskal–Wallis H test and Mann–Whitney U test have been employed to evaluate differences between experimental conditions (Guo et al., 2013; McIntosh et al., 2010). For instance, Sahin et al. (2025) determined the best extraction conditions for protocatechuic acid from pine honey samples using these nonparametric statistical approaches. However, when a large number of experimental parameters and their combinations are involved, classical statistical methods may become limited in modeling complex nonlinear relationships within experimental datasets.

In recent years, artificial intelligence-based modeling techniques have emerged as powerful tools for solving complex optimization problems in food science and natural product chemistry (Xiong & Xu, 2026). Artificial neural networks (ANNs) are particularly effective in capturing nonlinear relationships between experimental variables and response outputs (Yener et al., 2017). Among these models, Pi-Sigma artificial neural networks (PS-ANN) offer several advantages, including a simpler network structure, faster learning capability, and fewer adjustable parameters compared with conventional multilayer neural networks (Egrioglu & Bas, 2023). In addition, genetic algorithms (GA) are widely used evolutionary optimization techniques capable of efficiently searching large solution spaces and identifying global optimum solutions (Naaman et al., 2025). In this study, a hybrid optimization approach integrating Pi-Sigma artificial neural networks and genetic algorithms was employed to determine the optimal extraction conditions for protocatechuic acid from pine honey samples. The experimental dataset used in this research was obtained from the study of Sahin et al. (2025) and served as the basis for modeling and optimization. In the referenced study, five pine honey samples were initially analyzed, and the sample exhibiting the highest bioactive potential was selected based on total phenolic content (TPC) and total flavonoid content (TFC) results. Subsequently, extraction experiments were conducted by varying four parameters—extraction time, solvent type, temperature, and pH—resulting in 81 different extraction combinations. The extracts obtained under these conditions were analyzed using high performance liquid chromatography (HPLC), and the concentrations of protocatechuic acid, identified as a major phenolic component in pine honey, were determined. The experimental results obtained from these analyses constitute the dataset used for the artificial intelligence–based modeling and optimization performed in the present study (Sahin et al., 2025).

The main objective of this study is therefore to model the relationship between extraction parameters and protocatechuic acid yield using a Pi-Sigma neural network model and to determine the optimal extraction conditions through genetic algorithm-based optimization. By combining artificial intelligence modeling with evolutionary optimization techniques, this approach aims to provide a more efficient and reliable strategy for optimizing phenolic compound extraction from complex food matrices such as pine honey.

2. MATERIAL AND METHOD

In this section, the Pi-Sigma artificial neural network, genetic algorithm, and the dataset used in this study are introduced sequentially.

2.1. Pi-Sigma artificial neural network

The Pi-Sigma Artificial Neural Network (PS-ANN) model was first proposed by Shin and Ghosh (1991). The fundamental concepts of the Pi-Sigma neural network are explained below.

Architecture and output:

The basic operating principle of the Pi-Sigma neural network is that the network output is obtained as the product of several linear combinations of the input variables.

Degree concept:

The number of linear combinations used in the model is referred to as the degree of the Pi-Sigma neural network. Due to this structure, the Pi-Sigma neural network is classified as a higher-order neural network model.

Relationship between complexity and performance:

Increasing the degree of the Pi-Sigma neural network leads to a more complex functional structure with a greater number of parameters describing the relationship between input and output variables. Theoretically, this may improve prediction performance. However, using a large number of parameters may also lead to overfitting problems and increased computational time during the training process.

Let w_{ij} ($i = 1, 2, \dots, p, j = 1, 2, \dots, m$) denote the connection weights and θ_j ($j = 1, 2, \dots, m$) represent the bias terms. The final output of the network is obtained by multiplying the outputs of the hidden layer units and applying a transformation function, as expressed in Equation 1

$$output = \frac{1}{1 + \exp(-\prod_{j=1}^m (\sum_{i=1}^p w_{ij}x_i + \theta_j))} \quad (\text{Equation 1})$$

2.2. Genetic algorithm

Genetic Algorithms (GA) are population-based artificial intelligence optimization techniques developed by John Holland in the 1970s, based on the principle of natural selection derived from Darwin's theory of evolution. This approach simulates evolutionary processes in order to identify optimal solutions within complex search spaces by applying the concept of "survival of the fittest."

Unlike traditional optimization methods that operate on a single candidate solution, GA works with a population of potential solutions, enabling a broader exploration of the solution space. This characteristic reduces the risk of convergence to local minima and improves the algorithm's capability to search for global optimal solutions.

The GA process begins with the generation of a random initial population consisting of individuals known as chromosomes, which represent potential solutions to the optimization problem. The performance of each individual is evaluated using a fitness function specifically defined for the problem under consideration.

During the generation of new populations, three fundamental genetic operators are applied: selection, crossover, and mutation. These operators iteratively evolve the population by selecting the fittest individuals, recombining genetic information, and introducing random variations. The evolutionary process continues until a predefined number of iterations is reached or the desired error threshold is achieved.

As a result of this iterative evolutionary process, the individual with the highest fitness value within the final population is considered the optimal solution to the problem.

2.3. Proposed methodology and application

The method developed to solve the problem addressed in this study consists of two main stages. In the first stage, the following optimization problem was formulated and solved. The variables used in the optimization problem are defined in Table 1.

Table 1. Definition of variables used in the optimization model

Variable	Solvent Type* (x1)	Time (hour) (x2)	Temperature (°C) (x3)	pH (x4)	Protocatechuic Acid (µg/100 g sample) (y)
Values	1	3	20	7	-
	2	6	40	1	
	3	12	60	2	
Min Value	1	1	20	1	29.6748
Max Value	3	72	60	7	528.1818
Variable Type	Qualitative	Quantitative	Quantitative	Quantitative	Quantitative

*Solvent type was coded as follows: 1 = methanol; 2 = ethanol; 3 = water.

The objective of the study is to determine the optimal solvent type, extraction time, temperature, and pH values that produce the maximum protocatechuic acid concentration.

This process was carried out in two stages. First, extraction experiments were performed using all possible combinations of the four experimental variables—solvent type (type 1: methanol, type 2: ethanol, type 3: water), time (1h, 3h, 6h), temperature (20°C, 40°C, 60°C), and pH (pH₇, pH₁, pH₂). In total, 81 different combinations were generated. Each extraction experiment was repeated three times, resulting in 243 experimental observations (Sahin et al., 2025).

The experimental results obtained from these combinations were analyzed, and the distribution of the data is presented using Box–Plot graphs in Figure 1.

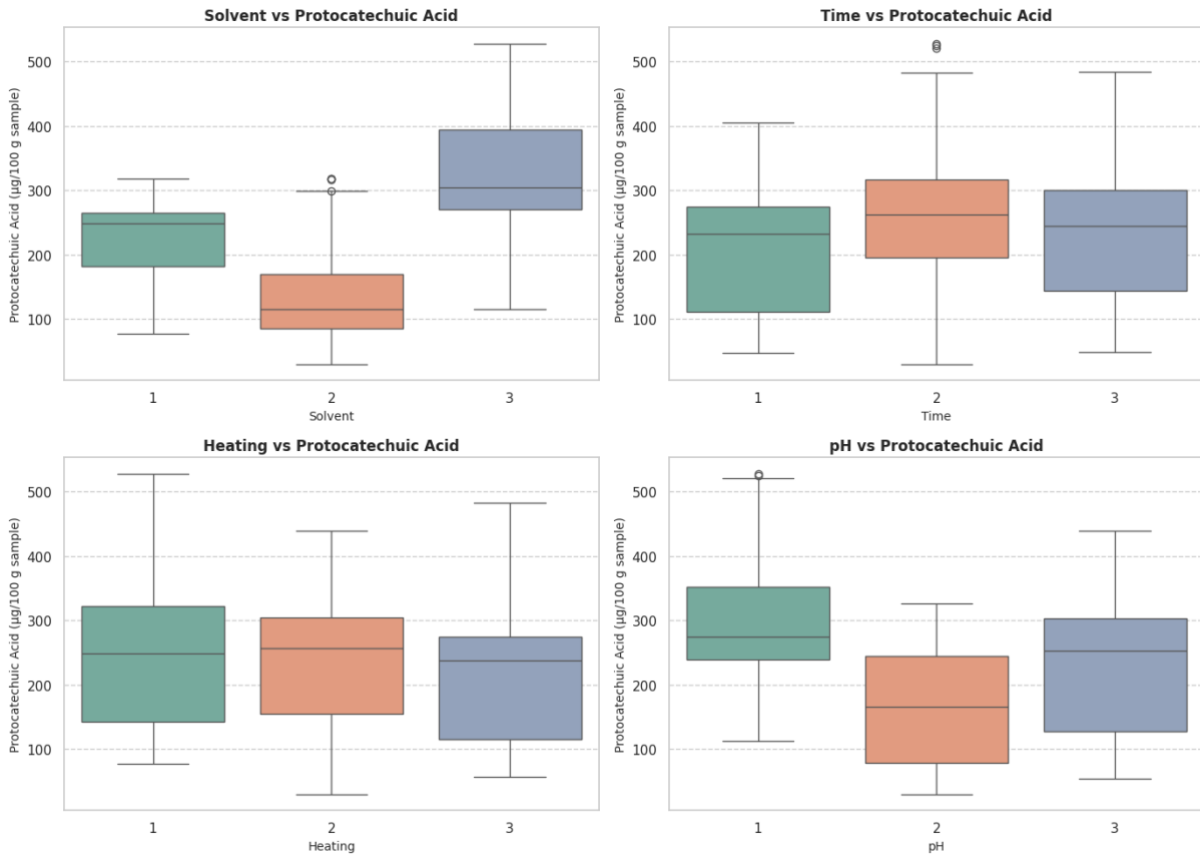


Figure 1. Box-plot representation of experimental results showing the effects of extraction parameters on protocatechuic acid content

A response surface model was constructed using a Pi-Sigma artificial neural network based on the available experimental data.

$$\min_{w_{ij}, \theta_j} \left(y - \frac{1}{1 + \exp(-\prod_{j=1}^m (\sum_{i=1}^p w_{ij} x_i + \theta_j))} \right)^2 \quad (\text{Equation 2})$$

In the first stage, the optimization problem presented in Equation 2 corresponds to the training of the Pi-Sigma artificial neural network. The dataset was divided into training (80%) and test (20%) subsets. The training process was performed using the Particle Swarm Optimization (PSO) algorithm. Each architecture is run 5 times with different initial random weights. The best architecture is selected based on the test performance. The number of hidden layers is 5 in the best architecture. The prediction performance is observed with (root mean square error (RMSE) and mean of absolute percentage error (MAPE) values. In the test set, *RMSE* is 50.5651, and *MAPE* is 0.2153.

1	40	38	4.3216	528.1818
1	70	40	5.2293	528.1818
3	8	20	7	455.5619
1	61	44	2.5990	528.1818
1	71	24	6.8757	528.1818
1	64	21	1.5275	528.1818
1	49	49	6.8817	528.1818
1	72	20	5.0468	527.2924
3	29	20	7	512.3791

*Solvent type was coded as follows: 1 = methanol; 2 = ethanol; 3 = water.

3. RESULTS AND DISCUSSION

In this study, an integrated optimization approach based on Pi-Sigma Artificial Neural Networks (ANN) and Genetic Algorithms (GA) was proposed to determine the optimal extraction conditions (solvent type, extraction time, temperature, and pH) for obtaining protocatechuic acid from pine honey samples. The model trained using the experimental dataset obtained by Sahin et al. (2025) successfully captured the complex and nonlinear relationships among the extraction parameters. Compared with classical statistical methods, the proposed artificial intelligence-based approach provided a more flexible and comprehensive modeling framework for the extraction process.

The Pi-Sigma ANN used in the modeling stage effectively represented the interactions between variables due to its higher-order structure. Since solvent type is a categorical variable, dummy variable encoding (Type 1 = methanol, Type 2 = ethanol, Type 3 = water) was used to enable the neural network to process solvent effects numerically. During the training stage, Particle Swarm Optimization (PSO) was employed to optimize the network weights and bias values, improving the generalization capability of the network while minimizing the risk of overfitting.

The response surface generated by the trained model expressed the relationship between extraction parameters and protocatechuic acid concentration as a continuous function. This function was then used as the fitness function in the genetic algorithm, allowing the search for global optimum solutions within the defined parameter constraints. The GA optimization process produced 20 candidate optimal solutions (Table 2), demonstrating that the proposed method provides not only a single optimal point but also multiple alternative operating conditions that may offer flexibility for industrial applications.

The optimization results revealed that the maximum achievable protocatechuic acid concentration within the investigated parameter ranges was approximately 528.1818 $\mu\text{g}/100$ g sample. This value was obtained in 16 of the 20 optimal scenarios, indicating that the algorithm successfully converged toward the global optimum solution.

Regarding the effect of solvent type, most of the scenarios that reached the maximum yield (approximately 75%) involved the use of Solvent Type 1 (methanol) (1, 0, 0). However, Solvent Type 3 (water) also achieved the same maximum yield under specific parameter combinations, particularly at 72 h extraction time, 20°C temperature, and pH 7.

The interaction between temperature and extraction time showed that high temperatures are not always required to achieve maximum extraction efficiency. For example, when Solvent Type 1 (methanol) was used, near-maximum yields (527.2924 $\mu\text{g}/100$ g sample) were obtained even at a relatively low temperature (20°C) by increasing the extraction time (72 h) or adjusting the pH level (pH 5.04). Conversely, at higher temperatures, such as 60°C, shorter extraction times (31 h) were sufficient, but extremely low pH values (pH 1) resulted in slightly reduced yields (509.2606 $\mu\text{g}/100$ g sample).

The results also indicate that pH alone is not the dominant factor determining extraction efficiency. Instead, the optimal pH value varied depending on solvent type and temperature conditions. In high-yield scenarios, pH values ranged between 1.53 and 7.00, suggesting strong interaction effects between pH and other extraction parameters.

In the previous study conducted by Sahin et al. (2025), optimal extraction conditions were determined using nonparametric statistical tests such as the Kruskal–Wallis and Mann–Whitney U tests. However, these statistical approaches are limited to comparisons among discrete experimental levels. In contrast, the ANN–GA approach proposed in this study explores the entire parameter space and can identify optimal solutions at intermediate parameter values that were not explicitly tested in the experiments.

The optimization results further demonstrate that achieving the maximum protocatechuic acid yield does not rely on a single “golden rule.” Instead, multiple parameter combinations—such as low-temperature/long-duration or high-temperature/short-duration extraction strategies—can produce comparable results. This flexibility is particularly advantageous for industrial applications where equipment capacity, energy consumption, and production time may impose practical constraints.

5. CONCLUSIONS

In conclusions, the combination of Pi-Sigma Artificial Neural Networks and Genetic Algorithms proved to be an effective optimization tool for modeling and optimizing the extraction of bioactive compounds from pine honey. The proposed approach provides a reliable, cost-effective, and flexible framework that can support process optimization in natural product extraction and functional food research.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Turkish Central Beekeepers Association (TAB) and the Muğla Provincial Beekeepers Association (MAYBİR) for providing the pine honey samples used in this study. This work was supported and funded by Giresun University, Scientific Research Projects (BAP) with the FEN-BAP-A-250221-34 project number.

AUTHOR CONTRIBUTION

Conceptualization, E.E.; Methodology and investigation, H.S., E.E., and E.B.; Formal analysis, H.S. and E.E.; Data curation, H.S.; Writing—review and editing, H.S., E.E., and E.B.; Supervision, H.S. and E.E. All authors have reviewed the manuscript and approved its final version for publication.

DECLARATIONS

Ethics Approval and Consent to Participate

This study did not involve any human participants or animals. Therefore, ethics committee approval was not required.

Conflict of Interest

The authors declare that there is no conflict of interest.

REFERENCES

- Egrioglu, E., & Bas, E. (2023). Modified Pi Sigma artificial neural networks for forecasting. *Granular Computing*, 8, 131–135 (2023). <https://doi.org/10.1007/s41066-022-00320-7>
- Eraslan, G., Kanbur, M., Silici, S., & Karabacak, M. (2010). Beneficial effect of pine honey on trichlorfon induced some biochemical alterations in mice. *Ecotoxicology and environmental safety*, 73(5), 1084–1091. <https://doi.org/10.1016/j.ecoenv.2010.02.017>

- Guo, S., Zhong, S., & Zhang, A. (2013). Privacy-preserving Kruskal–Wallis test, *Computer Methods and Programs in Biomedicine*, 112(1), 135-145. <https://doi.org/10.1016/j.cmpb.2013.05.023>
- Holland, J. H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press.
- Joshi, S. R., Pechhacker, H., Willam, W., & Von Der Ohe, W. (2000). Physico-chemical characteristics of *Apis dorsata*, *A. cerana* and *A. mellifera* honey from Chitwan district, central Nepal. *Apidologie*, 31(3), 367–375. <https://doi.org/10.1051/apido:2000128>
- Kahraman, K., Göcenler, O., & Dağ, Ç. (2024). Characterization of Turkish Pine honey and differentiation from floral honeys by NMR spectroscopy and chemometric analysis. *Journal of Food Composition and Analysis*, 127, 105983. <https://doi.org/10.1016/j.jfca.2024.105983>
- Kolaylı, S., Tunca, E., & Özkök, A. (2025). The Botanic, Physicochemical, Phenolic and antioxidant properties of Muş basin honeys. *Journal of Apitherapy and Nature/Apiterapi ve Doğa Dergisi*, 8(2), 175-192. <https://doi.org/10.35206/jan.1689764>
- Küçük, M., Kolayli, S., Karaoglu, Ş., Ulusoy, E., Baltacı, C., & Candan, F. (2007). Biological activities and chemical composition of three honeys of different types from Anatolia. *Food Chemistry*, 100(2), 526–534. <https://doi.org/10.1016/j.foodchem.2005.10.010>
- McIntosh, A. M., Sharpe, M., & Lawrie, S. M. (2010). Research methods, statistics and evidence-based practice. In E. C. Johnstone, D. C. Owens, S. M. Lawrie, A. M. McIntosh, M. Sharpe, & C. Livingstone (Eds.), *Companion to psychiatric studies - Eighth edition*, (pp. 157-198). <https://doi.org/10.1016/B978-0-7020-3137-3.00009-7>
- Naaman, D. W., Ahmed, B.T., & Ibrahim I. M. (2025) Optimization by nature: A review of genetic algorithm techniques. *The Indonesian Journal of Computer Science*, 14(1), 268-284. <https://doi.org/10.33022/ijcs.v14i1.4596>
- Sahin, H., Karacelik, A. A., Kaltalioglu, K., Bas, E., & Egrioglu, E. (2025). Assessment of extraction conditions on phenolic compounds in Turkish pine honey. *Emirates Journal of Food and Agriculture*, 37, 1-12. <https://doi.org/10.3897/ejfa.2025.141957>
- Sahin, H., Ozkok, A., Tanugur Samanci, A.E., Onder, E.Y., & Kolayli, S. (2022). Identification of the main phenolic markers in Turkish pine honeys and their biological functions. *Chemistry & Biodiversity*, 19, e202200835. <https://doi.org/10.1002/cbdv.202200835>
- Shin, Y., & Ghosh, J. (1991). The pi-sigma network: An efficient higher-order neural network for pattern classification and function approximation. *Proceedings of the International Joint Conference on Neural Networks*, 1, 13-18. <https://doi.org/10.1109/IJCNN.1991.155142>
- Silici, S., Sagdic, O., & Ekici, L. (2010). Total phenolic content, antiradical, antioxidant and antimicrobial activities of rhododendron honeys. *Food Chemistry*, 121(1), 238–243. <https://doi.org/10.1016/j.foodchem.2009.11.078>
- Taibi, A., Mokrani, A., Kadi, A., Bouherour, R., Guermi, N.E.Y., Teffane, M., & Richard, T. (2025), Optimization of extraction conditions of phenolic compounds and antioxidant activity from myrtle (*Myrtus communis* L.) fruit. *Chemistry & Biodiversity*, 22, e202301675. <https://doi.org/10.1002/cbdv.202301675>
- Tsavea, E., Vardaka, F.-P., Savvidaki, E., Kellil, A., Kanelis, D., Bucekova, M., & Mossialos, D. (2022). Physicochemical characterization and biological properties of pine honey produced across Greece. *Foods*, 11(7), 943. <https://doi.org/10.3390/foods11070943>

- Uçurum, H.Ö., Tepe, Ş., Yeşil, E. *et al.* (2023). Characterization of Turkish pine honey according to their geographical origin based on physicochemical parameters and chemometrics. *European Food Research and Technology*, 249, 1317–1327. <https://doi.org/10.1007/s00217-023-04215-y>
- Xiong, H.-H. & Xu, Y.-J. (2026). The application and development of artificial intelligence in functional food research, *Trends in Food Science & Technology*, 170, 105612. <https://doi.org/10.1016/j.tifs.2026.105612>
- Yener, D., Ozgener, O., & Ozgener, L. (2017) Prediction of soil temperatures for shallow geothermal applications in Turkey. *Renewable and Sustainable Energy Reviews*, 70, 71-77. <https://doi.org/10.1016/j.rser.2016.11.065>