

Impact-based Nondestructive Sensing and Machine Learning for Eggshell Thickness Prediction in Food Processing

Chia-Chun Lai,¹ Ting-En Wu,² Sih-Hao Huang,³ and Chia-Hung Lai^{3*}

¹AI Engineering and Cyber-Physical Learning Office, National Taichung University of Science and Technology, No. 129, Section 3, Sanmin Road, North District, Taichung City 404336, Taiwan (R.O.C.)

²Department of Aeronautics and Astronautics, National Cheng Kung University, No. 1, University Road, Tainan City 70101, Taiwan (R.O.C.)

³Department of Intelligent Automation Engineering, National Chin-Yi University of Technology, No. 57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung 411030, Taiwan (R.O.C.)

(Received July 28, 2025; accepted October 14, 2025)

Keywords: machine learning, duck eggs, eggshell thickness

We present a nondestructive and data-driven method for predicting eggshell thickness using impact-based sensing and machine learning. A custom low-speed impact module was developed to simulate mechanical responses of duck eggshells, and the Hertzian contact theory was employed to interpret deformation behavior. Three machine learning models—random forest (RF), XGBoost, and K-nearest neighbors (KNN)—were implemented and optimized with metaheuristic algorithms, including particle swarm optimization (PSO). Among them, the RF model obtained by PSO demonstrated superior prediction accuracy with an R^2 of 0.65155 and a mean squared error (MSE) of 0.00044. The proposed approach offers a scalable, cost-effective alternative to traditional eggshell assessment techniques and can be readily integrated into industrial egg grading systems to enhance food quality monitoring and reduce product waste.

1. Introduction

With the rapid development of machine learning technology, its application scope has expanded to various scientific and engineering fields, and the food industry has also begun to benefit from the transformations brought about by this technology.⁽¹⁾ Eggshell thickness, as one of the important indicators for measuring egg quality,⁽²⁾ not only affects the durability and processability of eggs but is also closely related to egg safety. However, traditional methods for measuring eggshell thickness often rely on destructive testing or high-cost equipment, posing challenges for large-scale egg production and quality monitoring. For example, terahertz waves⁽³⁾ and ultrasound⁽⁴⁾ nondestructive thickness measurement techniques can achieve noninvasive measurement, but their equipment costs are relatively high, and operation is complex. Therefore, establishing an efficient, accurate, and nondestructive method for predicting eggshell thickness holds significant practical value for the food industry and animal husbandry.

*Corresponding author: e-mail: chlai@ncut.edu.tw
<https://doi.org/10.18494/SAM5864>

The earliest applications of machine learning were in the industrial field.⁽⁵⁾ Convolutional neural network (CNN) was used to identify the nonlinear behavior of a robot arm with an accuracy of 99.5%.^(6,7) CNN was used to identify welding defects with an accuracy of more than 97%. In recent years, machine learning technology has shown broad application prospects in the field of food quality inspection.^(8,9) We aim to apply machine learning algorithms to the prediction and evaluation of egg product quality to improve detection efficiency and accuracy. However, research on predicting eggshell thickness is still relatively limited, especially in terms of integrating optimization algorithms to enhance prediction performance, leaving much room for exploration.

The innovation of this study lies in the introduction of multiple machine learning models, such as XGBoost, random forest (RF), and K-nearest neighbors (KNN), combined with particle swarm optimization (PSO) to optimize the models, aiming to improve the accuracy and generalization ability of eggshell thickness prediction. Traditional eggshell measurement methods are often physical measurements or destructive tests. In this study, nondestructive low-speed impact signal acquisition technology, combined with Hertzian contact theory, is used as input features for the machine learning model. This not only provides a more efficient and accurate measurement method but also enables the machine learning model to more effectively learn the relationship between eggshell thickness and physical properties. Additionally, the optimization method in this study differs from previous approaches that rely solely on a single model for prediction. It employs multiple optimization techniques for comparative analysis to identify the best algorithm combination, thereby improving prediction accuracy and computational efficiency.

In the experimental design of this study, a complete data analysis process was established by measuring the curvature and eggshell thickness of duck eggs. The collected data were divided into training and testing sets, and multiple performance evaluation metrics such as mean squared error (*MSE*), root mean squared error (*RMSE*), and R-squared (R^2), were introduced to compare the accuracy and robustness of different models and optimization methods. Through the analysis results, the impact of different models and their optimization methods on prediction accuracy was identified, and the most suitable solution for this application scenario was selected. The findings of this study not only demonstrate the potential of machine learning and optimization algorithms in theory but also provide a fast and reliable method to address practical agricultural needs. In the future, this model can be further applied to the quality inspection of other agricultural products, laying the foundation for the development of smart agriculture.

2. Theoretical Framework

2.1 Hertzian contact theory

Hertzian contact theory is the foundation of all mechanical models,⁽¹⁰⁾ but it is limited to frictionless surfaces and perfectly elastic solids.⁽¹¹⁾ Hertz proposed that the contact area is typically elliptical, and to calculate the localized deformation caused by the contact, the two objects can be considered as elastic half-spaces on an elliptical region.⁽¹²⁾ The maximum

Hertzian contact pressure is given by Eq. (1), where q_0 is the maximum pressure intensity and a is the contact radius. The total displacement of the sphere after contact is given by Eq. (2). The total deformation of the sphere after contact is given by Eq. (3), where r is the radius of the sphere.

$$P = \frac{3q_0}{2\pi a^2} \quad (1)$$

$$S = (k_1 + k_2) \iint q ds d\theta \quad (2)$$

Here, $k_1 = \frac{1-\nu_1}{\pi E_1}$ and $k_2 = \frac{1-\nu_2}{\pi E_2}$. E is the elastic modulus and ν is Poisson's ratio.

$$\delta = \frac{a^2}{R} = \left(\frac{3}{2}\right)^{\frac{2}{3}} \left(\frac{q_0}{E^2 r}\right)^{\frac{1}{3}} \quad (3)$$

2.2 Machine learning models

2.2.1 XGBoost

XGBoost is an end-to-end tree boosting system.⁽¹³⁾ XGBoost combines K-class classification and regression trees, with the final prediction being the sum of the outputs from each tree.⁽¹⁴⁾ The loss function is expanded using a second-order Taylor series to improve prediction accuracy and reduce model complexity.⁽¹⁵⁾ The XGBoost model has advantages such as strong generalization ability and fast computation.⁽¹⁶⁾ The objective function is given by Eq. (4), where l is the loss function, $\Omega(f_t)$ is the regularization term, and $Const$ is a constant. The regularization term is given by Eq. (5), where γ and λ are penalty coefficients.

$$O(\varphi) = \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + Const \quad (4)$$

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \omega_j^2 \quad (5)$$

By performing a second-order Taylor expansion of Eq. (5), the objective function becomes

$$O(\varphi) = \sum_{i=1}^n \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) + Const. \quad (6)$$

Here, $g_i = \frac{\partial l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}}$ and $h_i = \frac{\partial^2 l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)^2}}$. Simplifying Eq. (6), we obtain the form

$$O(\varphi) = \sum_{i=1}^n \left[g_i \omega_q(x_i) + \frac{1}{2} h_i \omega_q^2(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2. \quad (7)$$

Taking the derivative of the objective function with respect to ω_j , we obtain ω_j as

$$\omega_j = -\frac{g_i}{h_i + \lambda}. \quad (8)$$

Substituting Eq. (8) into Eq. (7), the optimal solution of the objective function is obtained as

$$O(\varphi) = -\frac{1}{2} \sum_{j=1}^T \frac{g_i^2}{h_i + \lambda} + \lambda T. \quad (9)$$

The conceptual diagram of XGBoost is shown in Fig. 1.

2.2.2 RF

RF⁽¹⁷⁾ is composed of predefined binary trees, where each tree in the forest is grown using training data. Assuming that each feature vector has N features, a subset of f features is randomly selected during the growth of trees.⁽¹⁸⁾ As training and testing progress, weaker decision trees are combined to construct a model with higher predictive performance.⁽¹⁹⁾ In RF, the computation formula for the internal nodes of each decision tree is shown as Eq. (10), where c represents the number of unique classes and x denotes the prior probability for each given class. The conceptual diagram of RF is shown in Fig. 2.

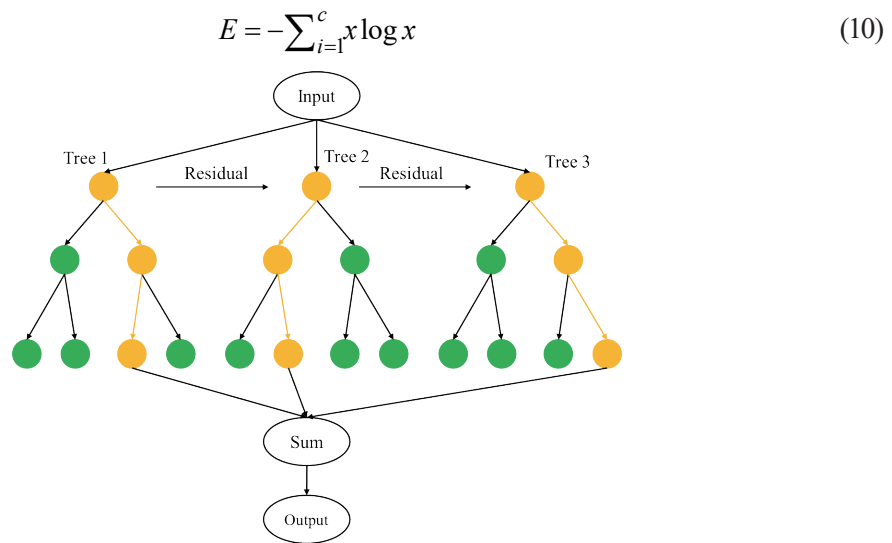


Fig. 1. (Color online) Conceptual diagram of XGBoost.

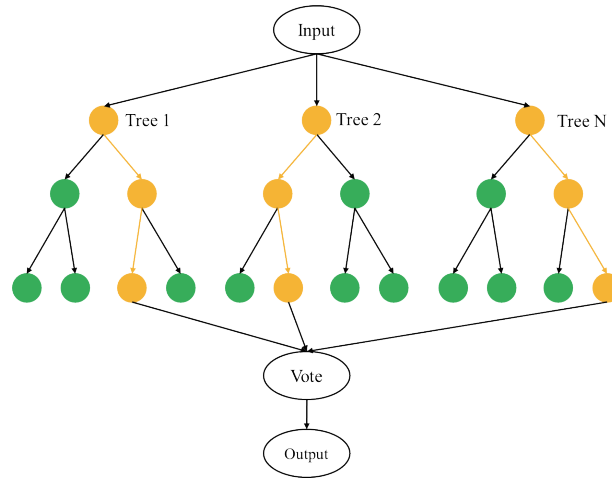


Fig. 2. (Color online) Conceptual diagram of RF.

2.2.3 KNN

KNN⁽²⁰⁾ is a classification method that identifies the K closest objects in the dataset. This method is known as the majority rule. Compared with machine learning models with multiple training stages, KNN omits the learning process and directly performs classification.⁽⁸⁾ The KNN decision process is shown by Eq. (11), where x represents an unknown class sample. The KNN classification algorithm uses Euclidean distance, as shown in Eq. (12), where x and y represent two samples. The conceptual diagram of KNN is shown in Fig. 3.

$$g_i(x) = \min g_j(x) \quad j = 1, 2, 3, \dots, C \quad (11)$$

$$D(x, y) = \sqrt{\sum_{j=1}^D (x_j - y_j)^2} \quad (12)$$

2.3 PSO

PSO⁽²¹⁾ is a continuous nonlinear function optimization method. By simulating the collective synchronization behavior of a flock of birds, it aims to maintain an optimal distance, forming the basis of its conceptual development. In PSO, the position of each particle and the swarm's overall position are continuously stored in memory.⁽²²⁾ All particles in the swarm iteratively update their positions, with the calculation formulas shown by Eqs. (13) and (14). Here, x represents the position vector, v denotes the velocity vector, ω is the inertia weight, z_1 and z_2 are optimization parameters, s_1 and s_2 are randomly generated values within a defined range, and P represents the best position.

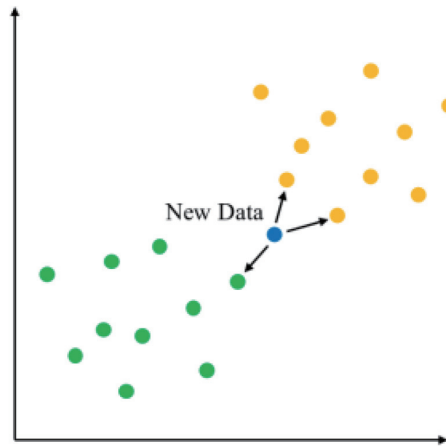


Fig. 3. (Color online) Conceptual diagram of KNN.

$$x_{j+1}^i = x_j^i + v_{j+1}^i \quad (13)$$

$$v_{j+1}^i = \omega v_j^i + z_1 s_1 (p_j^i - x_j^i) + z_2 s_2 (p_j^q - x_j^i) \quad (14)$$

3. Experimental Setup

We designed a novel eggshell thickness prediction model. First, 60 brown Muscovy duck eggs were selected, and their curvature and thickness were measured. The radii at three locations—the egg’s apex, equator, and blunt end—were recorded. A steel ball wrapped in a soft material was used to impact the eggshell at a low speed. An accelerometer was attached to the steel ball to capture the signals generated upon contact with the egg.

Next, the captured signals were used for machine learning with an 8:2 train–test split ratio. In this experiment, three machine learning models—XGBoost, RF, and KNN—were trained and used for prediction. To enhance model performance, PSO was applied to optimize the machine learning models. For model performance evaluation, MSE , $RMSE$, and R^2 were used as evaluation metrics. MSE and $RMSE$ effectively reflect the difference between the model’s predicted and actual values. R^2 is a measure of the model’s ability to recognize data variance. Finally, we compare the prediction accuracy and performance of different machine learning models and their optimization methods to determine the most suitable combination for this dataset. The schematic diagram of the experimental setup is shown in Fig. 4.

4. Results and Discussion

The experiments are conducted on brown Tsaiya duck eggs, whose curvature and eggshell thickness are measured at three locations: the tip, middle, and base. The collected data are used

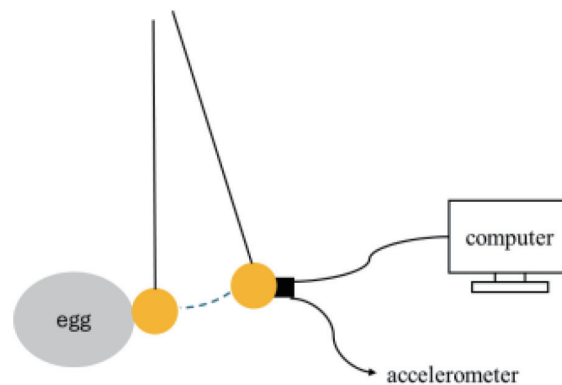


Fig. 4. (Color online) Schematic diagram of the experimental setup.

to train and test machine learning models. Then, PSO methods are applied to enhance the models. The prediction performance of the models is evaluated using MSE , $RMSE$, and R^2 .

MSE is the average of the squared differences between the predicted and actual values. A smaller MSE indicates more accurate predictions by the model. $RMSE$ is the square root of MSE . A smaller $RMSE$ indicates higher model performance. R^2 evaluates the proportion of variation explained by the model relative to the total variation. When R^2 is closer to 1, it indicates a better fit of the model.

For this experiment, XGBoost, RF, and KNN were selected as machine learning models, with optimization performed by PSO. After training and optimizing the XGBoost model, we compared the actual and predicted precision values, as shown in Fig. 5.

From Fig. 5(a), it can be seen that some data points deviate significantly from the dashed line, indicating that XGBoost's prediction capability is limited without optimization and may not fully meet practical requirements. Figure 5(b) shows an even closer scatter distribution to the dashed line after PSO, indicating an improvement in prediction accuracy. The XGBoost model performance indicators are shown in Table 1.

From Table 1, it can be observed that in terms of MSE , PSO results in values lower than those of the unoptimized XGBoost. In $RMSE$, PSO still outperforms the unoptimized XGBoost. In R^2 , PSO still surpasses the 36.5% variance explained by the unoptimized XGBoost. For the RF model, the actual and predicted accuracies after training and optimization are shown in Fig. 6

From Fig. 6(a), it can be seen that the unoptimized RF model has data points distributed near the dashed line, but there are still significant prediction errors. In Fig. 6(b), after PSO, the data points are more densely clustered near the dashed line, indicating a higher prediction accuracy. The RF model metrics are shown in Table 2.

From Table 2, it can be observed that in terms of MSE , the model obtained by PSO has the smallest prediction error. In terms of $RMSE$, PSO results in the smallest value, indicating that the predicted values are closest to the actual values. In terms of R^2 , the model obtained by PSO achieves the highest R^2 , indicating that the model can explain 65.2% of the variance. The KNN model's training and optimization results, including actual and predicted values, are shown in Fig. 7.

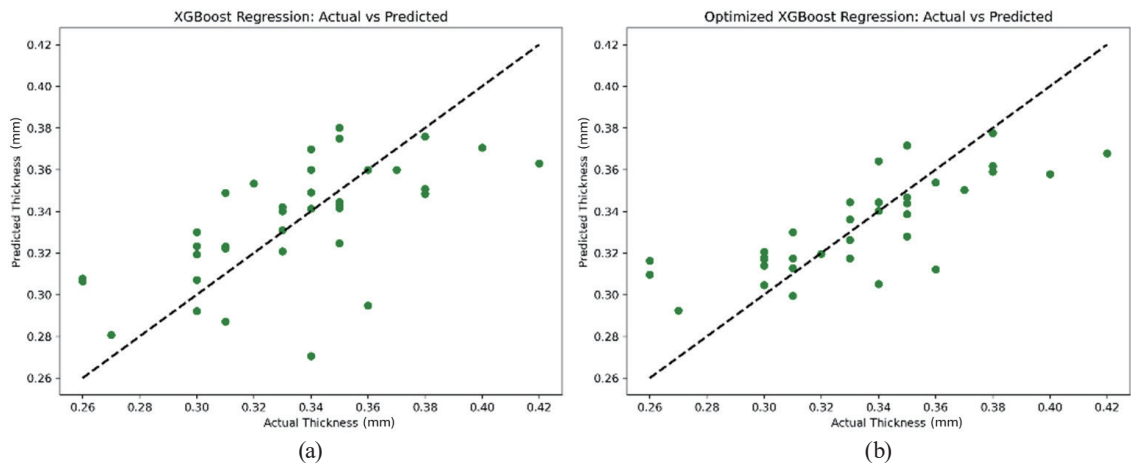


Fig. 5. (Color online) (a) XGBoost prediction results. (b) PSO_XGBoost prediction results.

Table 1
Performance metrics of XGBoost models.

Model	<i>MSE</i>	<i>RMSE</i>	<i>R</i> ²
XGBoost	0.00080	0.02845	0.36468
PSO_XGBoost	0.00056	0.02362	0.56194

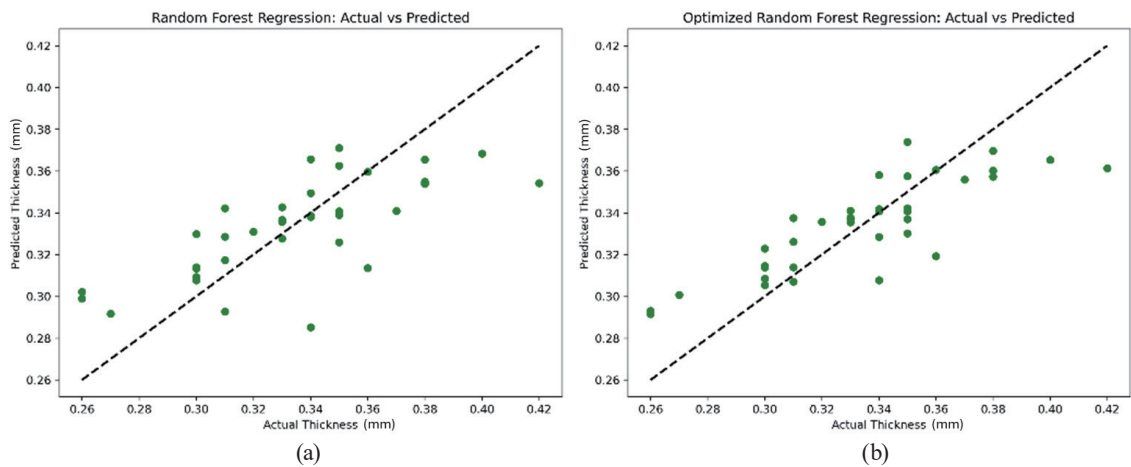


Fig. 6. (Color online) (a) RF prediction results. (b) PSO_RF prediction results.

Table 2
Performance metrics of RF models.

Model	<i>MSE</i>	<i>RMSE</i>	<i>R</i> ²
RF	0.00062	0.02494	0.51188
PSO_RF	0.00044	0.02107	0.65155

From Fig. 7(a), it can be seen that the unoptimized KNN model has a more scattered prediction result, indicating a lower accuracy. Figure 7(b) shows the results obtained after PSO; the data points are more concentrated and closer to the diagonal line, indicating a higher accuracy. However, the performance is still higher than that of the unoptimized KNN model.

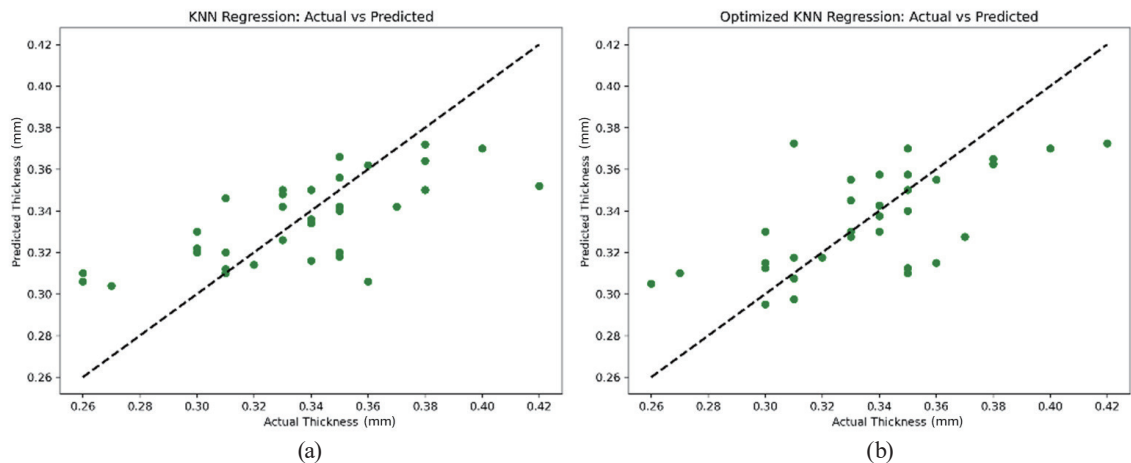


Fig. 7. (Color online) (a) KNN prediction results. (b) PSO_KNN prediction results.

Table 3
Performance metrics of KNN models.

Model	<i>MSE</i>	<i>RMSE</i>	<i>R</i> ²
KNN	0.00068	0.02604	0.46755
PSO_KNN	0.00066	0.02579	0.47798

From Table 3, it can be seen that with regard to *MSE*, the model obtained by PSO has the smallest prediction error. For *RMSE*, PSO gives the smallest value, suggesting that the predicted values are closest to the actual values. In terms of *R*², PSO achieves the highest *R*², indicating that the model can explain 47.8% of the variance.

5. Conclusions

In this study, we predicted the eggshell thickness of brown duck eggs on the basis of curvature and contact time. Three machine learning models, XGBoost, RF, and KNN, were used for prediction, as well as PSO methods. From the results of the above analysis, the following conclusions are drawn:

- (1) In the RF model, the model obtained by PSO exhibited the highest performance, with the *MSE* reaching the lowest value of 0.00044, *RMSE* decreasing to 0.02107, and *R*² improving to 0.65155, significantly higher than that of the unoptimized RF model. This demonstrates that the PSO of the RF model parameters can effectively reduce prediction errors and enhance the model’s robustness and accuracy.
- (2) Predicting eggshell thickness holds significant value for subsequent egg product processing. By accurately determining eggshell thickness, the processing and quality control workflows for egg products can be optimized, improving industry efficiency and product stability. While thinner eggshells may offer advantages in terms of easier processing, they also carry a higher risk of breakage and a shorter shelf life.
- (3) A fast and reliable method for predicting eggshell thickness and optimizing machine learning models without the need for time-consuming and costly trial-and-error experiments was

provided. Future research could combine other machine learning models with IoT-based smart monitoring systems and compare the cost and performance of ultrasonic or terahertz technology.

References

- 1 I. Kumar, J. Rawat, N. Mohd, and S. Husain: J. Food Qual. **2021** (2021) 4535567. <https://doi.org/10.1155/2021/4535567>
- 2 M. Ketta and E. Tümová: Ital. J. Anim. Sci. **17** (2018) 234. <https://doi.org/10.1080/1828051X.2017.1344935>
- 3 A. Khaliduzzaman, K. Konagaya, T. Suzuki, A. Kashimori, N. Kondo, and Y. Ogawa: Sci. Rep. **10** (2020) 1052. <https://doi.org/10.1038/s41598-020-57774-5>
- 4 L. Kibala, I. Rozempolska-Rucinska, K. Kasperek, G. Zieba, and M. Lukaszewicz: Poult. Sci. **94** (2015) 2360. <https://doi.org/10.1038/s41598-020-57774-5>
- 5 C.-C. Wang and Y.-Q. Zhu: Symmetry **13** (2021) 1445. <https://doi.org/10.3390/sym13081445>
- 6 C. H. Lai: J. Chin. Soc. Mech. Eng. **45** (2024) 453.
- 7 S. H. Huang, T. E. Wu, and C. H. Lai: J. Chin. Soc. Mech. Eng. **46** (2025) 205.
- 8 W. Xing and Y. Bei: IEEE Access **8** (2019) 28808. <https://doi.org/10.1109/ACCESS.2019.2955754>
- 9 X. Deng, S. Cao, and A. L. Horn: Annu. Rev. Food Sci. Technol. **12** (2021) 513. <https://doi.org/10.1146/annurev-food-071720-024112>
- 10 M. Machado, P. Moreira, P. Flores, and H. M. Lankarani: Mech. Mach. Theory **53** (2012) 99. <https://doi.org/10.1016/j.mechmachtheory.2012.02.010>
- 11 E. Ciulli, A. Betti, and P. Forte: Lubricants **10** (2022) 233. <https://doi.org/10.3390/lubricants10100233>
- 12 N. D. Londhe, N. K. Arakere, and G. Subhash: J. Tribol. **140** (2018) 021401. <https://doi.org/10.1115/1.4037359>
- 13 T. Chen and C. Guestrin: Proc. 22nd Acm SIGKDD Int. Conf. Knowledge Discovery and Data Mining (2016) 785. <https://doi.org/10.1145/2939672.2939785>
- 14 S. Thongsuwan, S. Jaiyen, A. Padcharoen, and P. Agarwal: Nucl. Eng. Technol. **53** (2021) 522. <https://doi.org/10.1016/j.net.2020.04.008>
- 15 D. Wang, H. Guo, Y. Sun, H. Liang, A. Li, and Y. Guo: Processes **12** (2024) 1660. <https://doi.org/10.3390/pr12081660>
- 16 S. Li and X. Zhang: Neural Comput. Appl. **32** (2020) 1971. <https://doi.org/10.1007/s00521-019-04378-4>
- 17 L. Breiman: Mach. Learn. **45** (2001) 5. <https://doi.org/10.1023/A:1010933404324>
- 18 A. Paul, D. P. Mukherjee, P. Das, A. Gangopadhyay, A. R. Chintha, and S. Kundu: IEEE Trans. Image Process. **27** (2018) 4012. <https://doi.org/10.1109/TIP.2018.2834830>
- 19 C.-C. Wang, P.-H. Kuo, and G.-Y. Chen: Appl. Sci. **12** (2022) 7739. <https://doi.org/10.3390/app12157739>
- 20 S. Zhang: IEEE Trans. Knowl. Data Eng. **34** (2021) 4663. <https://doi.org/10.1109/TKDE.2021.3049250>
- 21 J. Kennedy and R. Eberhart: Proc. ICNN'95-Int. Conf. Neural Networks **4** (1995) 1942. <https://doi.org/10.1109/ICNN.1995.488968>
- 22 F. A. Şenel, F. Gökçe, A. S. Yüksel, and T. Yiğit: Eng. Comput. **35** (2019) 1359. <https://doi.org/10.1007/s00366-018-0668-5>

About the Authors



Chia-Chun Lai is currently a postdoctoral fellow in the Office of Intelligent Engineering Virtual-Real Learning at National Taichung University of Science and Technology, Taiwan. He received doctorate degree from National Chung Hsing University in 2022. His research interests include nondestructive testing, finite element method, mechatronics, and engineering statistics. (s39911050@gmail.com)



Ting-En Wu is currently studying for a M.S degree at the Department of Aeronautics and Astronautics, National Cheng Kung University. His research direction lies in gear design, image recognition, and vibration analysis. (p46144382@gs.ncku.edu.tw)



Sih-Hao Huang is currently studying for a B.S degree at the Department of Intelligent Automation Engineering at National Chin-Yi University of Technology. His research direction lies in vibration analysis, Sensors and image recognition. (3b261070@gm.student.ncut.edu.tw)



Chia-Hung Lai currently serves as an assistant professor in the Department of Intelligent Automation Engineering at National Chin-Yi University of Technology. He received his B.S. and M.S. degrees from National Changhua University of Education, Taiwan, in 2009 and 2011 and his Ph.D. degree from National Cheng Kung University, Taiwan, in 2020. His research interests are in gear design and monitoring, cyber-physics, and sensors. (chlai@ncut.edu.tw)