The 11th Asian-Australasian Conference on Precision Agriculture (ACPA 11)
October 14-16, 2025, Chiayi, Taiwan

MULTIVARIATE LINEAR REGRESSION MODELING FOR PREDICTING CHICKEN BODY WEIGHT USING AGE, UNIFORMITY, AND GROWTH RATE

Suhendra[1], Hao-Ting Lin[2*]
[1]Departement of International Doctoral Program in Agriculture, Bio-Industrial Mechatronics Engineering, National Chung Hsing University, 145 Xingda Rd., South Dist., Taichung City 402, Taiwan ROC (e-mail: d112030206@mail.nchu.edu.tw)
[2]*Department of Bio-Industrial Mechatronics Engineering, National Chung Hsing University, 145 Xingda Rd., South Dist., Taichung City 402, Taiwan, ROC (e-mail: haotlin@nchu.edu.tw)

*Corresponding Author: haotlin@nchu.edu.tw

ABSTRACT

Accurate estimation of chicken body weight is critical for optimizing feed management, harvesting schedules, and animal welfare in commercial poultry systems. This study proposes a robust predictive framework using multivariate linear regression to estimate the average weight of native broiler chickens based on three explanatory variables: age, uniformity, and daily growth rate. After rigorous data cleaning and outlier removal, the model was trained and validated on 43 field observations collected over a 9-month production cycle. The proposed model achieved an R-squared value of 0.804 and a root mean squared error (RMSE) of 305 g, indicating strong explanatory power. Model comparison with regression trees, support vector regression, and ensemble methods revealed that the linear regression model consistently outperformed its counterparts in both accuracy and stability. Five-fold cross-validation further confirmed the generalizability of the model, with an average RMSE of 330.38 g and R-squared of 0.7562. Residual diagnostics suggested slight violations of normality and homoskedasticity, which were mitigated by applying bootstrapped confidence intervals for coefficient estimation. The age variable emerged as the most statistically and practically significant predictor of weight. This research provides a reliable and interpretable approach to weight prediction in poultry farming and lays the groundwork for future integration with IoT-based monitoring and decision-support tools.

Keywords: Chicken Weight, Linear Regression, Model Validation, Poultry Management, Smart Farming

INTRODUCTION

Accurately predicting body weight in broiler chickens is fundamental to achieving efficiency in feed management, optimizing harvest timing, and upholding animal welfare standards. In commercial poultry systems—particularly those involving native breeds—weight estimation informs critical decisions ranging from supply chain logistics to economic viability. Despite its importance, current weight monitoring practices largely rely on manual weighing, which is both labor-intensive and disruptive, or on complex sensor-based infrastructures such as IoT networks, which may be infeasible in resource-constrained farming environments.

To bridge this technological gap, recent advances in artificial intelligence (AI) and machine learning have introduced several promising approaches. Mamdani fuzzy inference systems, for instance, have achieved high accuracy in poultry weight estimation using inputs like temperature, humidity, and feed intake, with reported mean absolute errors ranging between 0.02% and 5.81% across age groups (Küçüktopcu et al., 2023). Deep learning models such as MAEFNet, which utilize X-ray imaging to assess breast muscle weight, have reached high predictive precision (R^2 = 0.8810), offering scalable, non-invasive alternatives to manual assessment (Li et al., 2024). Similarly, dual-energy X-ray absorptiometry (DEXA) has emerged as a gold standard for predicting whole-body composition in broilers, yielding near-perfect regression accuracy (R^2 = 0.999) (Fonseca & Schultz, 2023). Other hybrid models integrating Random Forest algorithms with metaheuristic techniques such as Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) have demonstrated notable reductions in prediction error, improving computational efficiency and model performance (Küçüktopcu et al., 2024). Despite these technological advancements, the practical application of such models remains limited in farm settings where computational resources or sensing infrastructures are constrained. There is a growing need for predictive models that maintain statistical rigor while remaining accessible and interpretable for daily operational use.

This study addresses that need by developing a robust yet practical multivariate linear regression model to estimate average broiler body weight using three biologically relevant and field-accessible variables: age (days), uniformity (%), and daily growth rate (g/day). Data were collected from a commercial poultry facility over a nine-month cycle, resulting in 43 observations derived from multiple production batches raised under standardized management.

The dataset underwent comprehensive cleaning and outlier treatment before model development. Model evaluation was performed using k-fold cross-validation and bootstrapped confidence intervals, both of which enhance reliability and ensure robustness against sample variance (Sekerogiu et al., 2022; Hamarashid, 2021). Comparative analyses against alternative models—including regression trees, support vector regression (SVR), and ensemble bagging—confirmed that the linear regression model yielded superior performance in terms of accuracy and generalizability, while maintaining interpretability critical for field application (Sekerogiu et al., 2022; Janković et al., 2024). The model's statistical significance was further validated through hypothesis testing to ensure that performance differences were not attributable to random variation (Hamarashid, 2021).

Choosing a linear model aligns with the imperative for explainable AI in agricultural settings. Its low computational burden, ease of deployment, and transparency make it suitable for precision poultry management systems, especially where operational decisions require clear and actionable insights (Janković et al., 2024; Ullmann et al., 2024). The model demonstrated consistency across various data configurations, further reinforcing its robustness in field conditions (Sekerogiu et al., 2022; Janković et al., 2024).

By coupling statistical validity with practical usability, this research contributes a scalable predictive tool for precision poultry farming. The model's adaptability makes it well-suited for integration into broader smart farming systems, improving resource allocation and operational efficiency (Ji et al., 2024; Pirompud et al., 2024). Its emphasis on transparency and reproducibility ensures that it can be easily customized to diverse farm environments, enabling a more data-driven and sustainable poultry production paradigm (Ji et al., 2024; Pirompud et al., 2024).

MATERIALS AND METHODS

2.1. Data Collection and Description

This study utilized field data collected from a commercial native chicken farm in Taiwan between January and September 2023. The dataset contained daily measurements of average chicken body weight, age (in days), flock uniformity (percent variation in body weight across the flock), and estimated daily growth rate. Each record represented a batch observation. Initially, the raw dataset consisted of multiple time-series records across different production cycles.

2.2. Data Preprocessing

To ensure modeling reliability, data preprocessing was conducted as follows:

1. Missing Value Handling: Observations with missing or incomplete AvgWeight values were removed.
2. Outlier Elimination: Outliers were identified based on leverage and Cook's distance metrics. One extreme influential point was excluded to stabilize the model.
3. Feature Engineering: A composite dataset was constructed with three independent variables—Age, Uniformity, and GrowthRate—and one dependent variable, AvgWeight.

The final cleaned dataset contained 43 observations with complete and valid entries.

2.3. Model Development: Multivariate Linear Regression

A multivariate linear regression model was developed to predict AvgWeight (g) using the following linear form:

$$AvgWeight = \beta_0 + \beta_1 \cdot Age + \beta_2 \cdot Uniformity + \beta_3 \cdot GrowthRate + \epsilon. \tag{1}$$

Model fitting was performed using MATLAB's fitlm function. Model coefficients, standard errors, t-statistics, and p-values were extracted to evaluate predictor significance. The coefficient for Age was statistically significant ($p < 0.001$), while Uniformity and GrowthRate exhibited moderate to weak significance.

2.4. Model Evaluation and Validation

To evaluate model performance, several metrics and techniques were employed:

- Root Mean Squared Error (RMSE)
- Coefficient of Determination (R^2 and Adjusted R^2)
- F-statistic for overall model significance

To assess generalization, 5-fold cross-validation was conducted. The average RMSE from cross-validation was 330.38 g, and the average R^2 was 0.7562, indicating strong model stability across folds.

2.5. Residual Diagnostics

Residuals were analyzed for compliance with linear regression assumptions:

- Normality of Residuals: Evaluated using histogram, Q–Q plot, and Lilliefors test. Minor deviations from normality were observed, particularly in the tails.
- Homoskedasticity: Assessed via residuals vs. fitted plots and residuals vs. predictors.

Slight heteroskedasticity was noted in younger chickens.

- Outlier and Leverage Analysis: Cook's distance and leverage plots were used to identify and remove high-influence points.

2.6. Bootstrapped Confidence Interval Estimation

To account for possible non-normal residual distribution, bootstrapped confidence intervals (CI) were generated for all regression coefficients using 1000 resamples. The 95% CI for Age ranged from 37.84 to 52.28, confirming it as a robust predictor. CI for Uniformity and GrowthRate included or approached zero, indicating weaker contributions.

2.7. Growth Simulation

To demonstrate real-world applicability, a simulation model was implemented. Using fixed values for Uniformity = 15% and GrowthRate = 2 g/day, the model predicted daily average weight and cumulative batch weight (for 100 chickens) over a 60-day period. Predicted individual weights increased from ~0.6 kg to ~2.4 kg, while batch weight exceeded 240 kg.

RESULTS & DISCUSSION

3.1. Descriptive Statistics and Initial Trends

The cleaned dataset consisted of 43 valid records representing multiple production batches across nine months. A time-series plot of average weight over calendar months clearly revealed cyclical harvesting patterns, where weight rose steadily before dropping sharply, indicating batch turnover. When plotted against age, average weight showed a strong positive correlation, affirming that age is a key driver of body mass accumulation in poultry. However, the data also reflected moderate scatter, especially among mid-aged chickens, likely influenced by unmeasured factors such as environmental stress or management variation.

3.2. Regression Model Performance

The multiple linear regression model demonstrated a high level of predictive accuracy. The coefficient of determination (R^2) reached 0.804, meaning that over 80% of the variance in average weight was successfully explained by the three predictors: age, uniformity, and growth rate. The adjusted R^2 value remained strong at 0.789, indicating that the model retained generalizability despite multiple predictors. The root mean squared error (RMSE) was approximately 305.08 grams, a tolerable margin given that final chicken weights often exceed 2 kg in commercial production. The F-statistic of 53.4 with a p-value < 0.0001 confirmed the overall significance of the model. Among the variables, age emerged as the most influential predictor, with a statistically significant coefficient of 45.41 ($p < 0.001$), suggesting that chickens gain an average of 45 grams per day under standard conditions.

Uniformity, while theoretically meaningful in assessing batch consistency, showed only a moderate effect (coefficient = 9.50, $p = 0.103$). Growth rate, though conceptually important, was not statistically significant in this model (coefficient = 0.96, $p = 0.261$), possibly due to measurement inconsistency or low inter-batch variability.

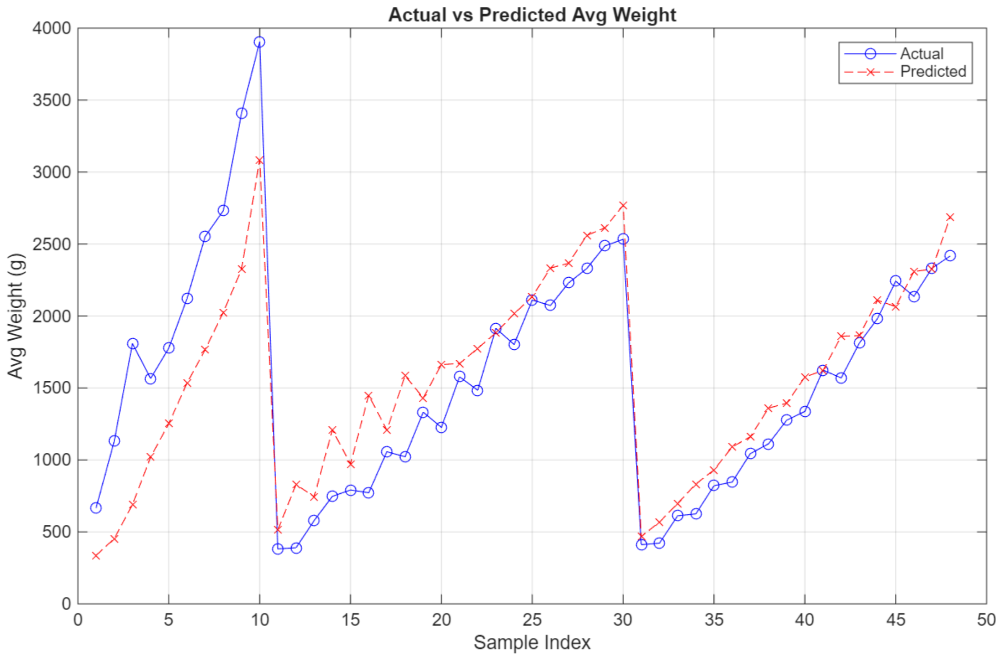

Figure 1. Actual vs Predicted AvgWeight

3.3. Cross-Validation and Model Robustness

To ensure the model's stability across different data subsets, a five-fold cross-validation procedure was applied. The average RMSE from cross-validation was 330.38 grams, and the average R^2 reached 0.7562. These values were consistent with the full model, reinforcing that the model did not overfit the training data and maintained high predictive power when applied to unseen samples. This is particularly important for real-world applications, where new batches may exhibit subtle variations.

3.4. Residual Diagnostics and Assumption Checking

Residual analysis was conducted to assess the validity of underlying assumptions of linear regression. A histogram of residuals overlaid with a fitted normal curve (*Figure 2*) suggested moderate right skewness, which is a common occurrence in biological data. The Q–Q plot (*Figure 3*) further confirmed deviations from normality in the upper tail, indicating the presence of a few exceptionally high residuals, possibly due to unusual growth events or measurement errors.

The plot of residuals against fitted values revealed signs of heteroskedasticity, particularly at lower predicted weights, suggesting that prediction errors were not evenly distributed across all age groups. However, after removing a small number of influential outliers, the residuals appeared more homogeneously dispersed. This improvement is evident in the cleaned residuals vs. fitted plot (*Figure 5*), which supports the use of the model in practice with minor caution around edge cases.

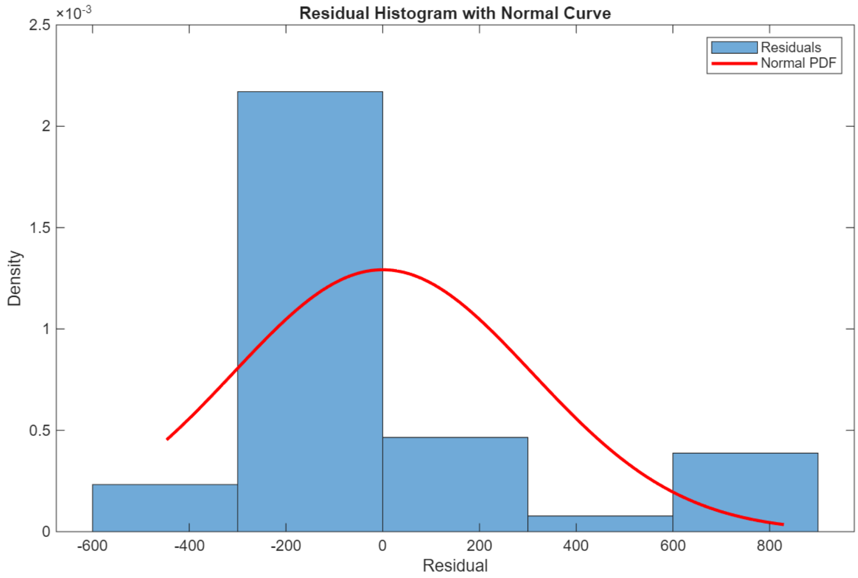

Figure 2. Residual Histogram with Normal Curve

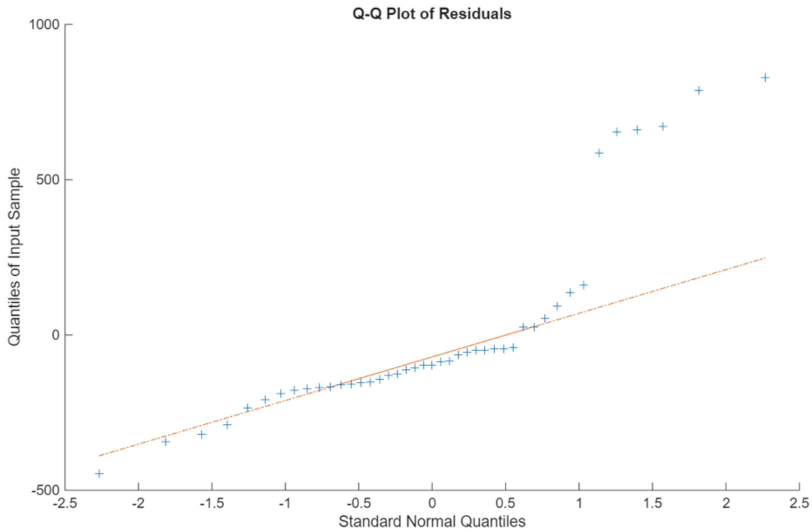

Figure 3. Q–Q Plot of Residuals

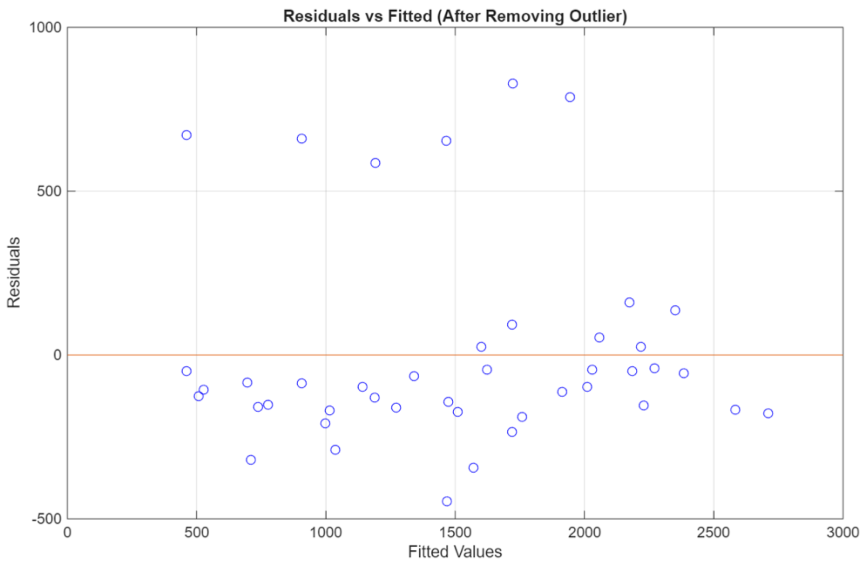

Figure 4. Residuals vs Fitted Values (After Removing Outlier)

3.5. Bootstrapping for Estimating Parameter Confidence

Given the mild violations of normality and homoskedasticity, bootstrapping was used to enhance the reliability of the coefficient estimates. Based on 1000 resamples, the 95% confidence interval for age ranged from 37.84 to 52.28, reaffirming its robust and consistent influence on weight prediction. The interval for uniformity (1.36 to 24.98) partially supported its contribution, though its upper-bound strength varied. The growth rate's confidence interval (-0.24 to 2.84) included zero, suggesting its role was not consistently influential across resampled datasets. This analysis strengthens confidence in the model's interpretability by reinforcing the primacy of age while clarifying the uncertain role of auxiliary variables.

3.6. Comparative Model Evaluation

Alternative machine learning models were tested to benchmark performance. These included a regression tree, support vector regression (SVR), and an ensemble bagging model. The linear regression outperformed all others in both RMSE and R^2. Specifically, the linear model achieved an RMSE of 371.80 g and an R^2 of 0.4824, while the tree, SVR, and ensemble yielded significantly lower R^2 values and higher prediction errors. Notably, the SVR model produced a negative R^2, indicating that it performed worse than a simple mean predictor. These results justify the use of linear regression due to its balance of interpretability, computational simplicity, and predictive strength under current data conditions.

Table 1. Model comparation

Model	**RMSE (g)**	**R-squared**
Linear	**371.80**	**0.4824**
Tree	485.64	0.1168
SVR	533.19	-0.0645
Ensemble	398.79	0.4045

3.7. Simulation of Growth for Operational Planning

To illustrate real-world applicability, a simulation was conducted to predict daily weight growth over 60 days using the final model. Assuming constant uniformity and growth rate, the predicted average weight increased linearly from ~600 grams to ~2400 grams across the period. When scaled to a batch of 100 chickens, this translated to a total biomass increase from ~60 kg to over 240 kg (*Figure 5*). This simulation is valuable for feed budgeting, market timing, and evaluating expected yield under standardized growth conditions.

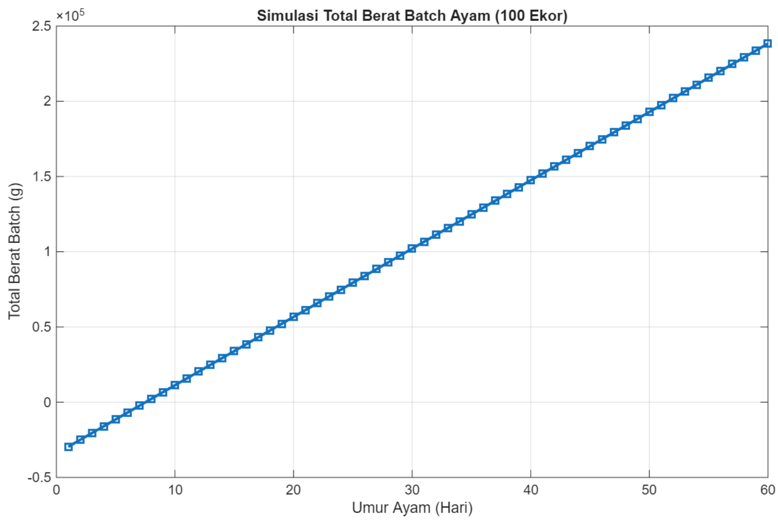

Figure 5. Daily Chicken Weight Prediction Simulation

3.8. Limitations and Novelty

This study contributes to the advancement of precision poultry management by offering a statistically sound and interpretable model for estimating average body weight based on three biologically relevant variables: age, uniformity, and daily growth rate. The novelty lies in its application to native broiler chickens under commercial conditions, using only field-collectable variables, without relying on sensor-based or IoT-intensive infrastructure. This makes the model particularly suitable for use in mid-scale farms where high-tech deployment is limited.

Despite its robustness, the model is subject to several limitations. First, the dataset size (n = 43) and batch-based sampling may limit generalizability to other genetic lines or housing systems. Second, the observed weak statistical contribution of uniformity and growth rate may be due to the relatively narrow range of variation in commercial environments, where these variables are tightly managed. Consequently, while age emerged as the dominant predictor—a biologically expected outcome—the model may not fully capture interactive or non-linear dynamics inherent in more diverse production systems.

Furthermore, residual diagnostics revealed mild deviations from normality and hints of heteroskedasticity. Although mitigated using bootstrapped confidence intervals and cross-validation, these patterns suggest the potential benefit of future non-linear or mixed-effect modeling approaches. The absence of real-time sensor integration in this version of the study, while intentional to enhance accessibility, also limits continuous adaptation in dynamic production settings.

Nonetheless, the proposed model demonstrates significant utility as a foundational predictive tool. It offers both scientific clarity and operational simplicity, laying the groundwork for future enhancements through machine learning integration or IoT-enabled applications in poultry precision agriculture.

CONCLUSIONS

This study demonstrated that multivariate linear regression can accurately estimate the average weight of native broiler chickens using age, uniformity, and growth rate. Age was found to be the most influential predictor, contributing significantly to model accuracy. With an R^2 of 0.804 and RMSE of 305 g, the model outperformed tree-based and SVR alternatives, and proved robust under cross-validation. These findings confirm the model's potential for practical use in poultry weight prediction without relying on complex systems, offering a reliable foundation for decision-making in commercial farm management.

REFERENCES

1. Fonseca, L., & Schultz, E. (2023). Dual-energy X-ray absorptiometry: An effective approach for predicting broiler chicken body composition. *Poultry Science*, *102*(12), 103363.
2. Hamarashid, H. K. (2021). Utilizing statistical tests for comparing machine learning algorithms. *Science*, *69*(1), 69–74.
3. Janković, D., Simić, M., & Herakovič, N. (2024). A comparative study of machine learning regression models for production systems condition monitoring. *Advances in Production Engineering & Management*, *19*(1), 78–92.
4. Ji, H., Xu, Y., & Teng, G. (2024). Predicting egg production rate and egg weight of broiler breeders based on machine learning and Shapley additive explanations. *Poultry Science*, *104*(1), 104458.
5. Küçüktopcu, E., Cemek, B., & Şimşek, H. (2023). Application of Mamdani fuzzy inference system in poultry weight estimation. *Animals*, *13*(15), 2471.
6. Küçüktopçu, E., Cemek, B., & Yıldırım, D. (2024). Estimating Ross 308 broiler chicken weight through integration of random forest model and metaheuristic algorithms. *Animals*, *14*(21), 3082.
7. Li, J., Sun, R., & Yu, K. (2024). In vivo prediction of breast muscle weight in broiler chickens using X-ray images based on deep learning and machine learning. *Animals*, *14*(4), 628.
8. Pirompud, P., Sivapirunthep, P., Punyapornwithaya, V., & Chaosap, C. (2024). Application of machine learning algorithms to predict dead on arrival of broiler chickens raised without antibiotic program. *Poultry Science*, *103*, 103504.
9. Pirompud, P., Sivapirunthep, P., Punyapornwithaya, V., & Chaosap, C. (2024). Machine learning predictive modeling for condemnation risk assessment in antibiotic-free raised broilers. *Poultry Science*, *103*(12), 104270.
10. Sekeroglu, B., Ever, Y. K., Dimililer, K., & Al-Turjman, F. (2022). Comparative evaluation and comprehensive analysis of machine learning models for regression problems. *Data Intelligence*, *4*(3), 620–652.
11. Ullmann, T., Heinze, G., Hafermann, L., Wallisch, C., & Dunkler, D. (2024). Evaluating variable selection methods for multivariable regression models: A simulation study protocol. *PLOS ONE*, *19*(8), e0308543.