Kafkas Universitesi Veteriner Fakultesi Dergisi
Journal Home-Page: http://vetdergikafkas.org
E-ISSN: 1309-2251

Kafkas Univ Vet Fak Derg
32 (3): 395-399, 2026
DOI: 10.9775/kvfd.2026.36392

Research Article

Prediction of Daily Egg-Laying in Japanese Quails Using Machine Learning

Kemal ESKİOĞLU [1], Berkant İsmail YILDIZ [1], Demir ÖZDEMİR [1(*)], Mustafa AKŞİT [2]

[1]Akdeniz University, Faculty of Agriculture, Department of Agricultural Biotechnology, TR-07058 Antalya - TÜRKİYE
[2]Aydın Adnan Menderes University, Faculty of Agriculture, Department of Animal Science, TR-09070 Aydın - TÜRKİYE

(*) **Corresponding author:**
Demir Özdemir
Phone: +90 505 818 12 79
E-mail: dozdemir@akdeniz.edu.tr

How to cite this article?
Eskioğlu K, Yıldız Bİ, Özdemir D, Akşit M: Prediction of Daily Egg-Laying in Japanese Quails Using Machine Learning. *Kafkas Univ Vet Fak Derg*, 32 (3): 395-399, 2026.
DOI: 10.9775/kvfd.2026.36392

Article ID: KVFD-2026-36392
Received: 10.02.2026
Accepted: 23.05.2026
Published Online: 02.06.2026

Abstract

Accurate monitoring of individual productivity in quail farming is essential for maximizing economic returns and shaping effective, sustainable breeding programs. However, daily variation in egg-laying behavior across individuals presents significant challenges to consistent productivity prediction. While conventional methods often struggle to account for these irregular and long-term behavioral trends, machine learning techniques offer a promising alternative through their capacity for individualized, data-driven modeling. In this study, we propose a machine learning-based approach to forecast daily egg-laying outcomes by identifying production patterns from a longitudinal dataset comprising 193 days of continuous observations on 371 quails. The predictive feature set includes rolling averages, cumulative production metrics, and prior-day laying status engineered from each bird's historical production data. Three supervised classification models, Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost) were employed to predict egg-laying on a daily basis. Performance was assessed using accuracy, precision, recall, and F1-score. Among the models, XGBoost outperformed the others, achieving an F1-score of 91.2% and a recall of 97% on the test set. Feature importance analysis identified the 7-day rolling mean as the most influential predictor. These findings underscore the value of machine learning approaches in modeling individual-level laying patterns and demonstrate their potential application in selection decisions, flock-level management, and automated performance monitoring systems in poultry production.

Keywords: Egg-laying prediction, Individual productivity, Machine learning, Precision livestock farming, Quail

Introduction

The Japanese quail *(Coturnix coturnix japonica)* is a poultry species of considerable economic and scientific importance, widely farmed for its rapid growth, early sexual maturity, and high egg production [1,2]. Despite its small body size, the species' prolific laying capacity has made it a valuable model in both commercial production and research settings. In particular, the quail's suitability for confined rearing conditions and efficient feed conversion has positioned it as a promising source of sustainable animal protein, especially in resource-limited contexts [3]. Egg production is a key determinant of profitability in quail farming. However, laying performance is shaped not only by genetic factors but also by a range of extrinsic and intrinsic variables, including environmental conditions, nutrition, health status, and individual behavioral tendencies [4,5]. This complexity underscores the need for continuous and individualized monitoring systems that can support effective flock management and improve the precision of genetic selection strategies.

Traditional approaches to productivity assessment often rely on weekly or periodic averages. While useful at a flock level, these methods tend to obscure the high variability present in individual laying behavior. They are especially limited in capturing the dynamic, day-to-day fluctuations that are characteristic of long-term production cycles. Accurate modeling of such temporal variability requires analytical tools capable of handling sequential data structures and uncovering subtle behavioral patterns [6]. In this context, artificial intelligence -and machine learning (ML) techniques in particular- have emerged as a powerful means of analyzing large-scale, temporally structured data. ML methods have been increasingly applied in animal production systems, including in disease detection, feed intake prediction, and productivity forecasting [7]. For example, Ahmad [6] compared three neural network architectures for predicting weekly egg production in commercial laying hens, identifying the general regression neural network (GRNN) as the most accurate model. More recently, Lemke et al. [8] used multivariate time-series data to forecast egg production in Hy-Line chickens, applying a sliding window strategy with algorithms such as Ridge Regression, Random Forest, XGBoost, and Multi-Layer Perceptron (MLP). Their findings showed that Ridge Regression, when used with a 7-day window, minimized forecasting errors to as low as 3.81% in mean absolute percentage error (MAPE).

Despite the growing body of literature on ML-based productivity forecasting in poultry, existing studies have predominantly focused on chickens and ducks. To date, there is a conspicuous absence of research applying machine learning to long-term, individual-level daily egg-laying data in quails. This represents a significant gap in the development of precision breeding tools for this species. The present study aims to address this gap by applying a pattern-based machine learning framework to forecast daily egg-laying behavior in Japanese quails. Using a longitudinal dataset comprising 193 consecutive days of egg-laying records from 371 individuals, a range of time-dependent features were engineered -such as 7-day rolling averages, previous-day laying status, and cumulative production. These features were then used to train and evaluate three supervised classification algorithms: Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost). Model performance was assessed using standard metrics including accuracy, precision, recall, and F1-score, with the goal of identifying an effective and generalizable framework for individual productivity modeling in quails.

Material and Methods

Ethical Statement

This study did not involve any procedures requiring ethical approval.

Data Collection and Dataset Description

This study utilizes a longitudinal dataset designed to monitor individual productivity trends in Japanese quails *(Coturnix coturnix japonica)*. A total of 371 quails were individually housed under standardized environmental conditions, including controlled lighting, temperature, and feeding regimes. The egg-laying status of each bird was recorded daily over a 193-day period, with egg presence coded as 1 and absence as 0. Missing entries -resulting from mortality or recording errors- were marked as NaN.

The raw dataset was originally structured in a wide format, where rows represented individual birds and columns denoted daily observations. To facilitate temporal analysis, the data were reshaped into a long format, with each row representing a single bird-day observation. This restructuring enabled more flexible feature engineering and temporal modeling.

Feature Engineering

To effectively capture individual productivity dynamics, three temporal features were engineered from each bird's production history:

- RollingMean_7: The 7-day rolling average of egg-laying status, reflecting recent productivity trends
- PrevDay: Binary indicator of egg-laying on the previous day
- CumSum: Cumulative egg count up to the current day

Missing values in predictor variables were imputed using feature-wise median imputation calculated across the dataset, while preserving the overall variance structure of the data. Records containing missing target values or incomplete rolling-window history were excluded from the analysis. In total, approximately 1,200 bird-day observations were excluded during preprocessing. The target for prediction was defined as the egg-laying status (1 or 0) on a given day. After preprocessing, the final dataset comprised over 70,000 valid bird-day observations.

Model Development

Continuous explanatory variables were standardized using z-transformation (mean = 0, standard deviation = 1) prior to model training.

To predict daily egg-laying behavior at the individual level, three supervised machine learning algorithms were applied:

- Logistic Regression (LR)
- Random Forest (RF)
- Extreme Gradient Boosting (XGBoost)

The models were trained to predict the laying status on day *t+1* using features derived from day *t*. The dataset was randomly split into training (80%) and test (20%) subsets, with stratification to preserve class distribution. As the class labels (laying vs. non-laying) were relatively balanced, no resampling or weighting adjustments were necessary.

Model performance was evaluated using the following classification metrics:

- Accuracy
- Precision
- Recall
- F1-score
- Area Under the Receiver Operating Characteristic Curve (AUC)

Table 1. Comparison of training and test performance metrics of machine learning models

Model	Train Accuracy	Train Precision	Train Recall	Train F1	Test Accuracy	Test Precision	Test Recall	Test F1
Logistic Regression	0.864	0.869	0.944	0.905	0.863	0.868	0.945	0.905
Random Forest	0.881	0.870	0.973	0.918	0.867	0.859	0.965	0.909
XGBoost	0.878	0.866	0.973	0.916	0.871	0.860	0.970	0.912

In addition to overall metrics, feature importance scores were extracted for the tree-based models (RF and XGBoost) to interpret predictor contributions. Confusion matrices and ROC curves were also plotted to compare classifier performance visually.

Computational Environment and Software

All analyses were performed in Python 3.9 within a Jupyter Notebook environment on a Linux-based computational server. The following open-source libraries were employed throughout the analysis:

- pandas and numpy for data manipulation and numerical operations
- scikit-learn for model development and performance evaluation
- xgboost for gradient boosting implementation
- matplotlib and seaborn for data visualization

The entire analytical workflow was managed using version control to ensure transparency and reproducibility of data preprocessing and modeling procedures.

Results

The predictive performance of three supervised machine learning algorithms -Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost)- was assessed using daily egg-laying records obtained from 371 individually monitored Japanese quails. The feature set comprised the 7-day rolling average of laying status (RollingMean_7), the previous day's laying outcome (PrevDay), and the cumulative egg count (CumSum). The dataset was partitioned into training (80%) and testing (20%) subsets to evaluate generalization performance.

All models demonstrated high classification performance on the training data. Logistic Regression achieved an F1-score of 90.5%, serving as a robust linear baseline. Random Forest and XGBoost yielded slightly higher F1-scores of 91.8% and 91.6%, respectively. Notably, Random Forest attained the highest recall (97.3%), indicating strong sensitivity in detecting laying events.

Model performance remained stable on the test dataset, suggesting minimal overfitting. Logistic Regression preserved an F1-score of 90.5%, whereas Random Forest and XGBoost achieved F1-scores of 90.9% and 91.2%, respectively. XGBoost exhibited the highest recall (97.0%) among all models. Precision values were comparable across methods (86-87%), reflecting balanced prediction behavior with controlled false-positive rates *(Table 1)*.

Confusion matrix analysis further characterized model behavior. All algorithms effectively classified laying (positive) days; however, Logistic Regression showed relatively lower sensitivity in identifying non-laying (negative) days compared to tree-based models. Random Forest demonstrated a more symmetrical classification profile, achieving high true positive and true negative counts. XGBoost minimized false negatives to the greatest extent, albeit with a marginal increase in false positives *(Fig 1)*.

Receiver Operating Characteristic (ROC) analysis produced AUC values of approximately 0.78 for LR

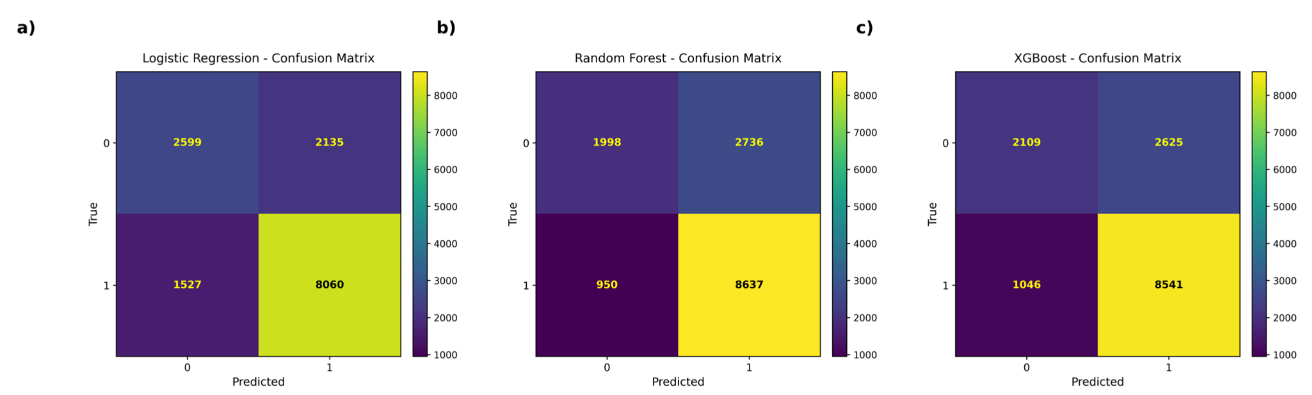

Fig 1. Confusion matrices illustrating classification performance across models: **(a)** Logistic Regression, **(b)** Random Forest, **(c)** XGBoost

and RF, and 0.79 for XGBoost. These results indicate moderate class separability despite strong classification metrics *(Fig 2)*.

Feature importance analysis derived from the Random Forest model identified RollingMean_7 as the dominant predictor, followed by CumSum, whereas PrevDay contributed minimally to overall predictive performance *(Fig 3)*.

Discussion

The consistently high predictive performance across all three algorithms underscores the effectiveness of temporal productivity indicators in forecasting daily egg-laying behavior at the individual level. The strong recall values, particularly those achieved by XGBoost, demonstrate the models' capacity to reliably detect productive laying days, thereby reducing the likelihood of overlooking actively laying individuals.

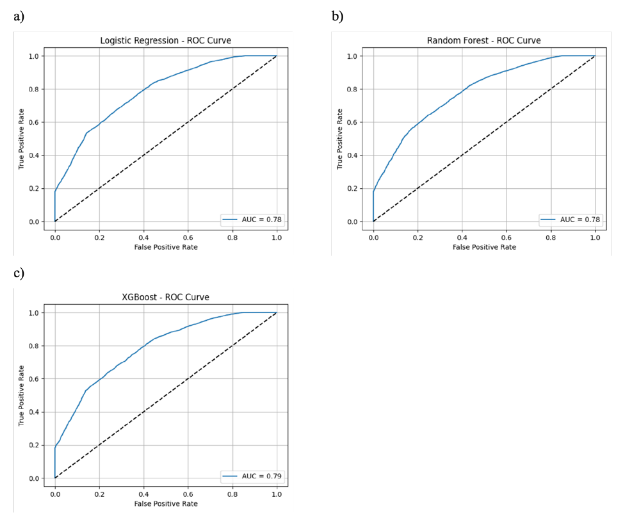

Fig 2. ROC curves of the three machine learning models: **(a)** Logistic Regression, **(b)** Random Forest, **(c)** XGBoost

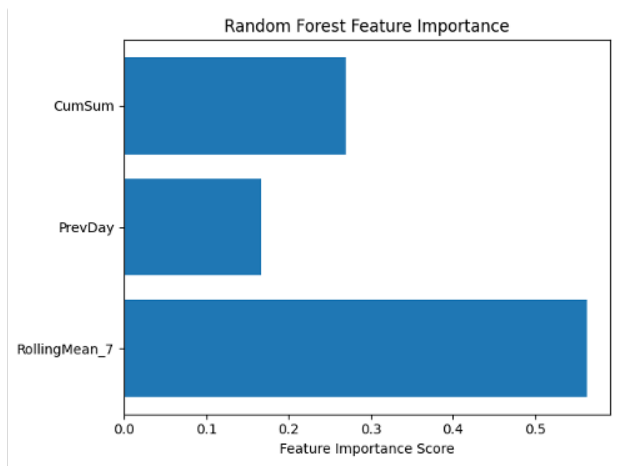

Fig 3. Relative importance of features in predicting daily egg laying according to the Random Forest model

The marginal superiority of tree-based models over Logistic Regression suggests the presence of non-linear interactions within laying behavior patterns that cannot be fully captured by linear decision boundaries. Random Forest provided balanced classification performance, while XGBoost exhibited enhanced sensitivity toward positive events, reflecting its strength in modeling complex feature interactions through gradient boosting mechanisms.

Although the F1-scores and recall values indicate high predictive accuracy, the moderate AUC values (0.78-0.79) suggest that class probability discrimination remains limited. This discrepancy implies that while threshold-based classification performs well, the sharpness of probabilistic separation between classes is comparatively modest. The constrained feature space -restricted to short-term rolling averages, cumulative output, and single-day lag variables- likely contributed to this limitation.

The dominance of RollingMean_7 in feature importance analysis highlights the predictive value of short- and medium-term production dynamics. In contrast, the relatively weak contribution of PrevDay indicates that isolated day-to-day continuity provides limited incremental information beyond aggregated temporal trends. This finding emphasizes that laying behavior is better characterized by dynamic temporal patterns rather than single-point historical indicators.

These observations align with prior research highlighting the significance of temporally structured features in poultry productivity modeling [9,10]. Moreover, the superior recall performance of XGBoost corroborates previous reports advocating gradient-boosted frameworks for minimizing false negatives in livestock and behavioral prediction contexts [11,12]. Comparable advantages of XGBoost have been documented in diverse livestock prediction tasks, including slaughter age estimation in pigs, live weight prediction in sheep, and carcass trait modeling in cattle [13-15], reinforcing its robustness in modeling biologically complex and non-linear systems.

Despite the promising findings, the moderate AUC values suggest that predictive refinement is feasible. Incorporating additional physiological attributes (e.g., age, body weight, health status) and environmental variables (e.g., temperature, photoperiod, housing conditions) may enhance discriminative power. Furthermore, advanced sequence modeling approaches such as long short-term memory (LSTM) networks or attention-based architectures could improve temporal dependency capture and provide deeper insights into individual laying dynamics.

Collectively, these results demonstrate that gradient-boosted machine learning models -particularly XGBoost- offer a reliable and scalable framework for individual-level egg-laying prediction. By integrating temporal production metrics into predictive modeling pipelines, this approach has the potential to support precision management and genetic selection strategies in quail production systems.

This study demonstrates that individual-level daily egg-laying behavior in quails can be reliably predicted using machine learning models trained on sequential production data. Among the evaluated algorithms, tree-based ensemble methods outperformed the linear baseline, with XGBoost achieving the highest test performance (accuracy: 87.0%, F1-score: 91.2%) and superior recall (97.0%). These results highlight the effectiveness of gradient-boosted frameworks in capturing non-linear temporal production patterns. Feature importance analysis identified the 7-day rolling average (RollingMean_7) as the most influential predictor, emphasizing the dominant role of short-term temporal dynamics over single-day indicators. The findings underscore the value of integrating time-dependent behavioral features into predictive modeling pipelines for precision livestock management. Such approaches enable early detection of productivity shifts and support individual-based monitoring strategies beyond traditional group-level analyses. Future improvements may be achieved through the inclusion of physiological and environmental variables and the application of advanced sequence modeling architectures such as LSTM or attention-based networks. Although conducted in quails, the proposed framework is transferable to other livestock systems with individual-level behavioral data.

Declaration

Availability of Data and Materials: The datasets used and/or analyzed during the current study are available from the corresponding author (D.Ö.) on reasonable request.

Acknowledgements: The authors declare that no acknowledgements are applicable.

Funding Support: The authors received no funding for this study.

Competing Interest: The authors declare that there are no competing interest.

Declaration of Generative Artificial Intelligence (AI): The authors declare that the article, tables and figures were not written/created by AI and AI-assisted Technologies.

Author Contributions: D.Ö. and M.A. were responsible for data collection and onsite monitoring. K.E. and B.İ.Y. conceived the study, designed the methodology, performed the data analysis, and drafted the manuscript. All authors have read and approved the final version of the manuscript for submission.

References

1. **Arunrao KV, Kannan D, Amutha R, Thiruvenkadan AK, Yakubu A:** Production performance of four lines of Japanese quail reared under tropical climatic conditions of Tamil Nadu, India. *Front Genet,* 14:1128944, 2023. DOI: 10.3389/fgene.2023.1128944
2. **Haqani MI, Nakano M, Nagano AJ, Nakamura Y, Tsudzuki M:** Association analysis of production traits of Japanese quail *(Coturnix japonica)* using restriction-site associated DNA sequencing. *Sci Rep,* 13 (1):21307, 2023. DOI: 10.1038/s41598-023-48293-0
3. **Goto T, Konno S, Konno M:** Establishment of wild-derived strains of Japanese quail *(Coturnix japonica)* in field and laboratory experiments. *Biology,* 12 (8):1080, 2023. DOI: 10.3390/biology12081080
4. **Pinto R, Ferreira AS, Albino LFT, Gomes PC, Vargas Júnior JGD:** Níveis de proteína e energia para codornas japonesas em postura. *Rev Bras Zootec,* 31 (4): 1761-1770, 2002. DOI: 10.1590/S1516-35982002000700019
5. **Ashour EA, Kamal M, Altaie HA, Swelum AA, Suliman GM, Tellez-Isaias G, Abd El-Hack ME:** Effect of different energy, protein levels and their interaction on productive performance, egg quality, digestibility coefficient of laying Japanese quails. *Poult Sci,* 103 (1):103170, 2024. DOI: 10.1016/j.psj.2023.103170
6. **Ahmad HA:** Egg production forecasting: Determining efficient modeling approaches. *J Appl Poult Res,* 20 (4): 463-473, 2011. DOI: 10.3382/japr.2010-00266
7. **Yildiz BI, Eskioglu K, Özdemir D, Akşit M:** Predicting quail egg quality using machine learning algorithms. *Braz J Poult Sci,* 27:eRBCA-2024, 2025. DOI: 10.1590/1806-9061-2024-2037
8. **Lemke V, Andrade JO, Komati KS:** Machine learning techniques for egg production prediction. **In,** *Proceedings of the 45th Ibero-Latin-American Congress on Computational Methods in Engineering*. 11-14 November, Maceió, Brazil, 2024.
9. **Rosário MFD, Silva MAN, Coelho AAD, Savino VJM:** Estimating and predicting feed conversion in broiler chickens by modeling covariance structure. *Int J Poult Sci,* 6, 508-514, 2007. DOI: 10.3923/ijps.2007.508.514
10. **Bumanis N, Kviesis A, Paura L, Arhipova I, Adjutovs M:** Hen egg production forecasting: Capabilities of machine learning models in scenarios with limited data sets. *Appl Sci,* 13 (13):7607, 2023. DOI: 10.3390/app13137607
11. **Hartono H:** Application of the XGBoost algorithm for predicting the target effective temperature in closed broiler chicken cage. **In,** *Proceedings of the 2nd International Conference on Mathematics, its Applications and Mathematics Education*. 21 September, Yogyakarta, Indonesia, 2024.
12. **Ji H, Xu Y, Teng G:** Predicting egg production rate and egg weight of broiler breeders based on machine learning and Shapley additive explanations. *Poult Sci,* 104 (1):104458, 2025. DOI: 10.1016/j.psj.2024.104458
13. **Alsahaf A, Azzopardi G, Ducro B, Veerkamp RF, Petkov N:** Predicting slaughter weight in pigs with regression tree ensembles. **In,** *Proceedings of the 1st International Applications of Intelligent Systems Conference.* 8-12 January, Las Palmas de Gran Canaria, Spain, 2018.
14. **Coşkun G, Şahin Ö, Altay Y, Aytekin İ:** Final fattening live weight prediction in Anatolian Merinos lambs from some body characteristics at the initial of fattening by using some data mining algorithms. *Black Sea J Agric,* 6 (1): 47-53, 2023. DOI: 10.47115/bsagriculture.1181444
15. **Srivastava S, Lopez BI, Kumar H, Jang M, Chai HH, Park W, Park JE, Lim D:** Prediction of Hanwoo cattle phenotypes from genotypes using machine learning methods. *Animals,* 11 (7):2066, 2021. DOI: 10.3390/ani11072066