

Machine Learning for Predicting Somatic Cell Count in Holstein Cattle

Mostafa Ghaderi-Zefrehei

mghaderi@yu.ac.ir

Yasouj University

Effat Nasre Esfahani

Payame Noor University

Morteza Jafaroghl

Payame Noor University

Mahdi Talebian-Shahreza

Isfahan University of Technology

Alireza Abdolmohammadi

Razi University

Fatemeh Esmaeili

Vahdat Isfahan Industrial Farmers and Livestock Breeders Cooperative

Saeid Ansari-Mahyari

Isfahan University of Technology

Reza Reisi-Vanani

Isfahan University of Technology

Research Article

Keywords: Somatic cell count, Machine learning, Bayesian Neural Network, Random Forest, Dairy cattle

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11210373/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Machine Learning for Predicting Somatic Cell Count in Holstein Cattle

Mostafa Ghaderi-Zefrehei¹, Effat Nasre Esfahani², Morteza Jafaroghl², Mahdi Talebian-Shahreza⁴, Alireza Abdolmohammadi³, Fatemeh Esmaeili⁵, Saeid Ansari-Mahyari^{4*}, Reza Reisi-Vanani⁴

¹ Department of Animal Sciences, Faculty of Agriculture, Yasouj University, Yasouj, Iran

² Department of Agriculture, Payame Noor University, Tehran, Iran

³ Department of Animal Science, Faculty of Agriculture, Razi University, Kermanshah, Iran

⁴ Department of Animal Science, Faculty of Agriculture, Isfahan University of Technology, Isfahan, Iran

⁵ Vahdat Isfahan Industrial Farmers and Livestock Breeders Cooperative, Isfahan, Iran

Corresponding Authors:

Mostafa Ghaderi-Zefrehei

P. O. Box: 75914, Fax: +98 741 222 4840, Tel: +98 741 222 4840,

Email: Mosmos741@yahoo.com, mghaderi@yu.ac.ir

Saeid Ansari-Mahyari

Email: s.ansari@cc.iut.ac.ir

ORCID:

Effat Nasre Esfahani: 0000-0002-5445-4759

Alireza Abdolmohammadi: 0000-0002-9785-2255

Mahdi Talebian-Shahreza: 0009-0008-3117-5985

Saeid Ansari-Mahyari: 0000-0003-1115-5043

Reza Reisi-Vanani: 0009-0005-5218-4109

Abstract

Somatic cell count (SCC) is a key indicator of udder health and milk quality in dairy cattle, but predicting it from routinely recorded data is difficult because SCC is highly skewed and depends on production, reproductive, and management factors that interact with one another. Seven modeling approaches were evaluated for predicting natural log-transformed SCC (LSCC) in Holstein cows using data from a commercial dairy herd. The approaches included Multiple Regression (MR) and Generalized Linear Model (GLM) as linear benchmark methods, a Support Vector Machine (SVM) using a Radial Basis Function kernel, and four nonlinear methods, including Bayesian Neural Network (BNN), Random Forest (RF), Chi-squared Automatic Interaction Detection (CHAID), and Classification and Regression Tree (C&R Tree). After multivariate outlier removal, 9,915 records were retained and stratified into five lactation groups, and models were developed using on-farm data. Nonlinear models generally showed higher testing correlations and lower relative errors than MR, GLM, and SVM. The best-performing models achieved testing correlations of approximately 0.62 to 0.70 across the first four lactation groups, whereas predictive performance declined in cows with more than four lactations. BNN and RF showed broadly comparable predictive performance across lactation groups. Average protein, milk, and fat yields were among the most influential predictors of LSCC in both models, with average milk yield becoming the most important predictor in cows with more than four lactations. Overall, nonlinear models, particularly BNN and RF, provided better prediction of LSCC from routinely recorded data than the conventional approaches, although prediction accuracy decreased in later lactations.

Keywords: Somatic cell count; Machine learning; Bayesian Neural Network; Random Forest; Dairy cattle

1. Introduction

Somatic cell count (SCC) remains among the most widely used indicators of udder health and milk quality in dairy cattle. Elevated SCC reflects mammary inflammation and intramammary infection, and it is associated with reduced milk yield and economic losses at the herd level [1, 2]. Because SCC is typically measured only periodically through milk recording rather than monitored continuously, there has been growing interest in using routinely collected cow-level information to fill the gaps between sampling occasions. Predicting SCC from production and management variables that are already recorded on farms could complement conventional testing and help flag cows that warrant closer attention. The growing volume of production, reproductive, management, and sensor data generated on dairy farms has opened new possibilities for predictive modeling. Machine-learning (ML) methods have already been applied to several dairy health and production problems, from mastitis detection to reproductive performance and the prediction of economically important traits. Goyachee et al. [3], for instance, used ML techniques together with routinely available herd and cow records to identify cows with persistently elevated SCC and support udder health management decisions. Bobboe et al. [4] compared eight ML methods for predicting subsequent udder health status and found that Neural Network, Random Forest, and linear models performed comparably well when udder health was defined using an SCC threshold of 200,000 cells/mL and previous test-day cow and milk records were used as predictors. Similarly, Ebrahimi et al. [5] found that deep learning and gradient-boosted tree models outperformed several other machine-learning approaches for predicting subclinical mastitis. More recently, Pakrashi et al. [6] showed that ML models incorporating cow-level production, milk composition, parity, and mastitis-history variables could predict subclinical mastitis up to 7 d before its occurrence. Together, these studies demonstrate the potential of ML to extract predictive signal from data dairies already

collect, although most have targeted classification of mastitis-related events rather than prediction of SCC as a continuous trait.

This distinction matters because SCC is not simply a binary marker of disease status. Sorting cows into healthy and diseased categories around a fixed threshold discards information contained in the actual magnitude of SCC. Anglart et al. [7] addressed this directly by comparing generalized additive models, Random Forest, and multilayer perceptron models for predicting cow composite SCC from routinely recorded quarter-level milking data. They found that continuous SCC could be predicted from variables such as milk flow and conductivity, and highlighted the value of predicted SCC between routine sampling dates for udder-health management. These studies differ considerably in how the outcome is defined, which predictors are used, and how models are validated, which makes it difficult to compare approaches directly across the existing literature. Despite this expanding literature, comparative evaluation of different modeling approaches for predicting continuous SCC from routinely recorded production, reproductive, and management traits, while accounting for lactation group, remains limited. Most studies have instead framed the outcome as a classification problem or relied on detailed sensor-derived milking variables, as shown by Bobbo et al. [4] and Anglart et al. [7], respectively. Likewise, Thompson et al. [8] used lifetime milk-recording data from over 100 commercial herds to predict elevated composite SCC status after calving, treating SCC as a binary rather than continuous outcome. Whether routinely recorded production and reproductive information can effectively predict the natural logarithm of SCC (LSCC), and whether different ML approaches provide consistent advantages across lactation groups, therefore remains unclear. This question is also practically relevant because routinely recorded variables could be incorporated into existing herd-recording systems without additional sensor infrastructure or specialized measurements.

Accordingly, this study compared the predictive performance of Support Vector Machine (SVM), Multiple Regression (MR), Generalized Linear Model (GLM), Bayesian Neural Network (BNN), Random Forest (RF), Chi-squared Automatic Interaction Detection (CHAID), and Classification and Regression Tree (C&R Tree) for predicting LSCC in Holstein dairy cows using on-farm data. Models were fitted separately across five lactation groups to account for potential parity-related differences, while BNN architecture and predictor importance were also evaluated across models and lactation groups.

2. Material and Methods

2.1. Dataset

Data were obtained from the FOKA Holstein dairy herd (Isfahan, Iran), previously described by Ghaderi-Zefreheiet al. [9]. It comprised 13 production, reproductive, and management records for Holstein dairy cows, including days in milk (DIM), age at first calving (AFC), calving interval (CI), dry period length (DP), days from calving to first service (CTFS), gestation length (GL), days open (DO), days from first service to conception (FSTC), milk yield, fat yield, protein yield, somatic cell count (SCC), and lactation number. All data preprocessing and statistical analyses were performed using R version 4.1 [10]. Multivariate outliers were identified using the `pcout()` function in the `mvoutlier` package, which implements the PCOut method of Filzmoseret al. [11] for robust outlier identification in high-dimensional data. This method computes semi-robust principal components from sphered data and combines robust estimates of location and scatter to assign an outlier weight to each observation. Three production-derived variables were created to better describe lactation performance, including average milk yield (MY, kg), average fat yield (FY, kg), and average protein yield (PY, kg). Each variable was calculated by averaging cumulative production records and their corresponding test-day measurements to capture both overall production

potential and performance during the current lactation. Following preprocessing, records were stratified into five lactation groups (1, 2, 3, 4, and >4). Separate prediction models were developed for each lactation group to account for parity-related biological differences. Because SCC was positively skewed, it was natural log-transformed prior to analysis:

$$LSCC = \ln(SCC) \quad (1)$$

where LSCC is the log-transformed of SCC. LSCC was used as the response variable in all subsequent analyses. Descriptive statistics for all predictor variables and LSCC are presented in Table 1.

2.2. Machine Learning Models

Within each lactation group, records were randomly partitioned into training (75%) and testing (25%) subsets using a fixed random seed to ensure reproducibility. The same partitioning strategy was applied to all models to enable a fair comparison of predictive performance.

LSCC was predicted using Support Vector Machine (SVM), Multiple Regression (MR), Bayesian Neural Network (BNN), Random Forest (RF), Generalized Linear Model (GLM), Chi-squared Automatic Interaction Detection (CHAID), and Classification and Regression Tree (C&R Tree) models. BNN and RF were implemented in R using the `brnn` and `randomForest` packages, respectively. The BNN models used Bayesian regularization to control model complexity [12], and models with two to five hidden neurons were fitted for each lactation group to evaluate the effect of network architecture on predictive performance. SVM models employed the Radial Basis Function (RBF) kernel, whereas GLMs assumed a Gaussian distribution with an identity link function. CHAID and C&R Tree models were developed using IBM SPSS Modeler. Mathematical formulations and key properties of the seven modeling approaches are summarized in Table 2.

2.3. Model Evaluation

Predictive performance was evaluated for both the training and testing subsets using four complementary metrics, including Residual Sum of Squares (RSS), Root Mean Square Error (RMSE), the Pearson correlation coefficient (r), and Relative Error (RE), calculated as:

$$RSS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (4)$$

$$r = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(\hat{Y}_i - \bar{\hat{Y}})}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2}} \quad (5)$$

$$RE = \frac{\frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|}{\frac{1}{n} \sum_{i=1}^n |Y_i|} \quad (6)$$

where Y_i and $\hat{Y}_i$ are the observed and predicted log-transformed somatic cell count (LSCC) for the i th observation, respectively; $\bar{Y}$ and $\bar{\hat{Y}}$ are the means of the observed and predicted values, respectively; and n is the total number of observations. Lower values of RSS, RMSE, and RE indicate better predictive accuracy, whereas higher Pearson correlation coefficients (r) indicate stronger agreement between observed and predicted LSCC values.

3. Results

3.1. Descriptive Statistics

The descriptive statistics of all variables across the five lactation groups are presented in Table 1. After data cleaning and multivariate outlier detection, a total of 9,915 records were retained for analysis and subsequently classified into first, second, third, fourth, and greater than fourth lactation groups. The number of observations decreased progressively with increasing lactation number, from 4,147 records in the first lactation to 338 records in cows with more than four lactations. As expected, the average age at first calving increased steadily across successive lactations, reflecting the aging structure of the herd. Milk production traits also varied among

lactation groups. Average milk yield generally increased from 10,882.32 kg in first-lactation cows to 11,740.47 kg in cows with more than four lactations. Similar increasing trends were observed for average fat yield and average protein yield, indicating greater production capacity in mature cows. Reproductive and management variables showed moderate variation among lactations. Mean calving interval ranged from 373.28 to 393.08 days, while average dry period length remained relatively stable, varying between 53.71 and 57.69 days. Likewise, gestation length exhibited minimal variation across lactation groups, averaging approximately 278 days. Log-transformed somatic cell count (LSCC) generally increased from first to fourth lactation, with mean values of 5.00, 5.63, 5.84, and 5.96, respectively, before slightly decreasing to 5.74 in cows with more than four lactations. This trend suggests that udder health challenges generally increase with advancing parity, although the slight decline observed in the oldest cows may be due to the substantially smaller sample size in this group.

3.2. Comparison of Model Performance across Lactation Groups

Among the evaluated machine learning algorithms, the Bayesian Neural Network (BNN), Chi-squared Automatic Interaction Detection (CHAID), and Classification and Regression (C&R) Tree models generally achieved higher predictive accuracy, as measured by the Pearson correlation coefficient (r) and relative error (RE), than the Support Vector Machine (SVM), Multiple Regression (MR), and Generalized Linear Model (GLM) across all lactation groups (Table 3).

In the first lactation, the BNN produced the highest testing correlation ($r = 0.6994$) together with the lowest testing relative error ($RE = 0.5137$), indicating superior predictive performance. CHAID and C&R Tree also showed competitive performance, with testing correlations of 0.6579 and 0.6676, respectively. In contrast, the SVM, MR, and GLM models achieved lower testing correlations ranging from 0.5936 to 0.6150. A similar trend was observed in the second

and third lactations. The BNN consistently maintained the highest testing correlations ($r = 0.6300$ and 0.6176 , respectively), whereas CHAID and C&R Tree provided comparable performance. SVM, MR, and GLM exhibited relatively lower predictive ability throughout these lactation groups. Model performance declined progressively with increasing lactation number. During the fourth lactation, testing correlations ranged from 0.4663 for SVM to 0.6101 for C&R Tree, while relative errors increased across all models. In cows with more than four lactations, predictive performance further decreased, with testing correlations ranging from 0.4508 to 0.5090 and relative errors exceeding 0.77 for most models. Across all lactation groups, CHAID and C&R Tree generally maintained higher predictive accuracy than SVM, MR, and GLM, particularly in later lactations. The BNN consistently achieved the highest or near-highest prediction accuracy during the first three lactations, whereas C&R Tree showed comparatively better performance in cows with four or more lactations. *Figure 1* compares the predictive performance of the evaluated machine learning models using r in the training dataset across different lactation groups. Overall, CHAID and C&R Tree together with the BNN achieved the highest correlation values, whereas SVM, MR, and GLM showed comparatively lower predictive performance. The highest correlations were generally observed during the first and second lactations, while predictive performance gradually declined in later lactations, particularly after the fourth lactation. This reduction is likely associated with the smaller sample size and greater biological variability among older cows. Although all models exhibited reduced accuracy in advanced lactations, the nonlinear methods maintained relatively higher correlation values, indicating greater robustness when modeling complex relationships within heterogeneous dairy production data.

Figure 2 presents the predictive performance of the evaluated models on the independent testing dataset. Similar to the training results, CHAID, C&R Tree, and the BNN consistently produced the highest correlation coefficients across most lactation groups, demonstrating good

generalization ability. A moderate reduction in correlation from the training to testing datasets was observed for all models, indicating the expected decrease in predictive accuracy when evaluated on unseen data. Predictive performance also declined with increasing lactation number, particularly in cows with more than four lactations, where greater biological variability and fewer observations likely limited model performance.

Figure 3 compares model performance using RE in the training dataset. Lower RE values indicate better predictive accuracy. Across most lactation groups, the BNN, CHAID, and C&R Tree achieved the lowest relative errors, whereas SVM, MR, and GLM generally produced larger prediction errors. RE tended to increase as lactation number increased, reflecting the greater difficulty of accurately predicting LSCC in older cows. Nevertheless, the nonlinear methods consistently achieved lower prediction errors than the other approaches, indicating their greater capacity to capture complex nonlinear relationships among production and management variables.

Figure 4 illustrates RE of the evaluated models on the testing dataset. The overall pattern broadly resembled that observed during training, with BNN, CHAID, and C&R Tree tending to produce lower prediction errors than SVM, MR, and GLM, although which of the three performed best varied by lactation group. CHAID's testing RE was comparatively high in the third lactation, and BNN's was comparatively high in the fourth. The training-to-testing increase in RE was modest for most models but was more pronounced for C&R Tree in cows with more than four lactations (RE = 0.50 in training versus 0.81 in testing), consistent with a greater risk of overfitting for this model in the smallest sample group. Prediction accuracy declined noticeably in cows with more than four lactations, likely because of reduced sample size and increased biological heterogeneity. Collectively, the results from Figures 1–4 demonstrate that BNN, CHAID, and C&R Tree generally provided more accurate prediction of LSCC than SVM, MR, and GLM across the different lactation groups.

3.3. Bayesian Neural Network Performance

The predictive performance of BNN models was evaluated using architectures containing two to five hidden neurons (Table 4). Model performance was assessed using r and RMSE for both the training and testing datasets. For the first lactation, predictive accuracy improved as the number of hidden neurons increased from two to five. The five-neuron architecture achieved the highest training correlation ($r = 0.7026$) and the lowest training RMSE (1.8006), while the corresponding testing performance remained stable ($r = 0.6547$; RMSE = 1.8933). A similar pattern was observed during the second lactation, where the four-neuron architecture produced the highest testing correlation ($r = 0.6548$) with a testing RMSE of 1.9193. Increasing the number of neurons beyond four did not further improve predictive performance. For the third and fourth lactations, model accuracy also increased with network complexity. In the third lactation, the five-neuron model achieved the highest testing correlation ($r = 0.6248$) and the lowest testing RMSE (1.9166). Likewise, the highest training correlation in the fourth lactation ($r = 0.6769$) was obtained using five hidden neurons, although the three-neuron model produced the highest testing correlation ($r = 0.6774$). In contrast, model performance remained relatively unchanged in cows with more than four lactations. Testing correlations ranged from 0.5904 to 0.5939 across all network architectures, while RMSE values varied only marginally, indicating little influence of additional hidden neurons on predictive accuracy for this lactation group.

3.4. Random Forest Performance

The predictive performance of the Random Forest model across the five lactation groups is presented in Table 5. The highest training correlation was observed during the first lactation ($r = 0.6796$, RMSE = 1.8585), together with a testing correlation of 0.6593 (RMSE = 1.8885). The highest testing correlation, however, was observed during the second lactation ($r = 0.6746$),

despite a lower training correlation ($r = 0.6438$) than in the first lactation. Model performance remained relatively stable during the third and fourth lactations. Testing correlations were 0.6153 and 0.6614, respectively, while testing RMSE values ranged between 1.8558 and 1.9254. The lowest predictive performance was observed in cows with more than four lactations. The training correlation decreased to 0.4532, although the testing correlation remained at 0.5732. RMSE values also increased compared with earlier lactation groups, indicating reduced prediction accuracy in this subset.

3.5. Variable Importance Analysis

Table 6 summarizes the relative importance of predictor variables estimated by the Random Forest model using both the percentage increase in mean squared error (%IncMSE) and increase in node purity (IncNodePurity). Across all lactation groups, PY, MY, and FY were consistently identified as the most influential predictors of LSCC. Among these variables, PY exhibited the highest importance across nearly all lactation groups according to both importance measures. Among the remaining predictors, AFC and DP showed relatively high importance during the first and second lactations, whereas their contribution gradually declined in later lactations. Variables associated with reproductive performance, including CTFS, DO, CI, and FSTC, exhibited moderate importance throughout the study.

Table 7 presents the variable importance obtained from the BNN model. Similar to the Random Forest results, PY, MY, and FY consistently ranked as the three most influential predictors of LSCC across all lactation groups. Among these variables, PY showed the highest relative importance, particularly during the first lactation, indicating that protein production was the strongest contributor to model predictions. The importance of DP increased from the first to the third lactation before declining in later lactations, suggesting that its influence on LSCC may vary with parity. AFC was relatively important during early lactations but became less

influential in older cows. Several predictors exhibited very low or slightly negative %IncMSE values in cows with more than four lactations, which is likely attributable to the limited number of observations and increased variability within this group rather than a true lack of biological relevance.

3.6. Comparative Performance of BNN and RF

Table 8 summarizes the predictive performance of the BNN and RF models across lactation groups. RF showed slightly higher testing correlations in the first two lactations, whereas BNN showed marginally higher testing correlations in the third, fourth, and greater-than-fourth lactation groups. The differences between the two models were generally small, indicating no consistent predictive advantage of either model across all lactation groups. Table 9 presents the variable importance estimated by the BNN and RF across different lactation groups. In both models, PY, MY, and FY consistently ranked among the most influential predictors according to both %IncMSE and IncNodePurity. Production-related variables generally exhibited higher importance than reproductive and management traits across all lactation groups. The importance of several predictors decreased in cows with more than four lactations, whereas moderate variation in the ranking of secondary predictors was observed between the two machine learning algorithms.

Discussion

4.1. Model Performance and BNN

Across the five lactation groups, BNN, CHAID, and C&R Tree, the three most flexible algorithms in the general model comparison, produced higher testing correlations and lower testing relative errors than SVM, MR, and GLM (Table 3; Figures 1 to 4). The gap was clearest in the first lactation, where BNN reached a testing correlation of 0.70, compared with

approximately 0.59 to 0.61 for MR and GLM, and a similar pattern was observed in the second and third lactations. RF, evaluated separately from the general model comparison, showed a similar pattern, with testing correlations exceeding those of SVM, MR, and GLM across the lactation groups and performance comparable to CHAID and C&R Tree in several groups (Table 5). Depending on lactation group, the gap in testing correlation between the best-performing flexible model and the best-performing of the remaining three methods ranged from about 0.04 to 0.15. The relatively better performance of the more flexible models may be related to the complex and heterogeneous nature of the factors associated with SCC. Somatic cell count is influenced by stage of lactation, parity, milk production, udder health, and management conditions, and the relationships among these factors are not necessarily constant across cows or stages of lactation. In particular, the association between SCC and milk production has been shown to vary with parity and stage of lactation, suggesting that a single linear relationship may not fully capture the complex relationship between SCC and production-related variables [13, 14]. This type of structure may favor algorithms that can accommodate nonlinear relationships and interactions among predictors. In contrast, MR and the Gaussian-identity GLM used in the present study impose a more restricted functional form unless nonlinear terms or interactions are explicitly incorporated, whereas BNN and tree-based approaches can learn more complex relationships directly from the predictor space. SVM was implemented here with a Radial Basis Function kernel (Section 2.2) and is therefore not a linear method in a strict methodological sense; its empirical performance nonetheless tracked MR and GLM more closely than the tree- and network-based methods in the present comparison, which may reflect the specific kernel and hyperparameter settings used here rather than a general limitation of kernel-based methods. Previous research in dairy cattle provides support for the potential value of flexible prediction methods for udder-health-related traits. Gianola et al. [15] demonstrated that Bayesian neural networks could capture nonlinear relationships in

dairy cattle quantitative traits and showed that predictive performance and effective model complexity varied across network architectures and according to the predictor information used. Nørstebøet al. [16] showed that variation in online SCC was associated with lactation stage, parity, milk yield, previous-milking SCC, conductivity, season, intramammary infection status, and genetic lineage, although individual factors explained only a small proportion of the total variation. Xu et al. [17] found that ML models could use routinely recorded cow-level production and management variables to predict metabolic status in dairy cows, with the importance of production-related predictors varying across lactation weeks. These findings, together with the present results, suggest that flexible algorithms may be advantageous when predicting SCC-related traits because they can accommodate complex relationships among routinely recorded variables. Nevertheless, the moderate testing correlations observed in the present study indicate that these variables explain only part of the variation in LSCC, and that additional information related to udder health, environment, management, or previous disease history may be required to achieve substantially higher predictive accuracy. The single BNN configuration reported in Table 3 performed less consistently across lactation groups than CHAID, C&R Tree, and RF. Its testing correlation declined steadily with increasing lactation number, from 0.70 in the first lactation to 0.51 in the fourth and 0.48 in cows with more than four lactations. This configuration was the lowest of any nonlinear method in the fourth lactation, although CHAID's testing correlation (0.47) was slightly lower in cows with more than four lactations. Exploring BNN architectures with two to five hidden neurons (Table 4) showed that predictive performance varied across lactation groups and network architectures. The final BNN architecture for each lactation group was selected using the predefined validation procedure, and the selected models were subsequently evaluated on the independent testing dataset (Table 8). With the best-performing architectures, BNN's testing correlation in the fourth lactation reached 0.68 with three hidden neurons ($RMSE = 1.81$), while it reached

0.59 with four hidden neurons in cows with more than four lactations (RMSE = 1.85), improvements of roughly 0.10 to 0.17 correlation points over the single configuration summarized in Table 3. BNN and RF produced almost equivalent testing accuracy in every lactation group (Tables 4, 5, and 8). RF held a modest edge in the first two lactations (testing r of 0.66 and 0.67 versus 0.65 for the best BNN configuration), while the BNN was marginally ahead in the third, fourth, and oldest parity groups. These differences, on the order of 0.01 to 0.02 in testing correlation, are small relative to the differences between these models and the conventional regression approaches evaluated in Table 3. In the present study, the first-lactation group contained substantially more training records than the group with more than four lactations, which may have contributed to the different BNN performance observed across lactation groups. RF, by contrast, uses an ensemble of decision trees and does not require the same hidden-layer architecture selection as BNN, which may have contributed to its more consistent performance across lactation groups (Table 5). These results indicate that RF provided a relatively consistent alternative across lactation groups, whereas BNN performance depended more strongly on the selected network architecture.

4.2. Variation in Model Performance across Lactations

Regardless of the algorithm applied, predictive accuracy was highest in the first three lactations and lowest in cows with more than four lactations (Tables 3, 4, and 5). Because a similar decline was observed across the evaluated algorithms, the pattern appears to be related to differences among lactation groups rather than to the behavior of a single modeling approach. The most direct explanation is sample size. After outlier removal, 4,147 records were available for first-lactation cows but only 338 for cows with more than four lactations, and the training sets used for the machine learning models were smaller by the same proportion, down to roughly 250 records for the oldest parity group. The smaller number of training records in the highest lactation group provided less information for estimating the relationships between the

predictors and LSCC, which may have contributed to the lower predictive performance observed in this group. The variation in model performance across lactation groups may partly reflect biological differences associated with parity and lactation history. Hagnestam-Nielsen et al. [13] showed that multiparous cows had a geometric mean SCC approximately 73% higher than primiparous cows (95,000 vs. 55,000 cells/mL) and that the effect of SCC on milk yield varied with parity, indicating that the relationship between SCC and production changes across lactations. Laevens et al. [18], however, found that parity, stage of lactation, and their interaction were not significant determinants of SCC among bacteriologically negative cows, whereas these effects became significant when infected cows were included. This suggests that some of the variation in SCC across parity and lactation stage may be attributable to differences in infection status and associated inflammatory responses. Similarly, Williams et al. [19] analyzed lactation yield and somatic cell score trajectories in more than 1.1 million dairy cows across successive parities and reported substantial genetic variation in these trajectories, supporting the view that udder-health-related phenotypes can change systematically as cows advance in lactation number. In the present dataset, such biological heterogeneity may have contributed to the weaker relationship between the recorded predictor variables and LSCC in the highest lactation group, in addition to the smaller sample size. These results support the decision to fit separate models for each lactation group rather than a single pooled model. The differences in production and reproductive traits across lactation groups shown in Table 1, together with the observed changes in predictive performance, suggest that the relationships between these predictors and LSCC may differ across lactation groups. A parity-specific modeling strategy is better aligned with the underlying biology, even though it necessarily produces smaller, and therefore less stable, training sets for the higher-parity groups.

4.3. Important Predictors of LSCC

PY, MY, and FY were consistently the three most important predictors of LSCC in both the RF and BNN models (Tables 6, 7, and 9). PY was the top-ranked predictor in RF across the first four lactations and in BNN across the first two; in later parities (the oldest group for RF, and from the third lactation onward for BNN) MY became the leading production trait instead, though all three production variables remained far more informative than the reproductive and management traits throughout. The consistent identification of these three production traits by two fundamentally different modeling approaches suggests that their predictive contribution was robust to the choice of algorithm. The result is consistent with the established association between milk production and udder health. Intramammary infection can alter mammary epithelial function and milk secretion, changing both milk yield and composition, while production level is itself associated with physiological and metabolic conditions that may influence udder health. Hagness-Nielsen et al. [13] showed that the association between SCC and milk yield changes across parity and lactation stage, supporting the relevance of production traits when predicting SCC. In addition, Halasa et al. [20] reviewed evidence showing that both clinical and subclinical mastitis are associated with reductions in milk production, illustrating the close relationship between udder-health status and productive performance. Because SCC reflects intramammary inflammation, this production-udder-health relationship provides a biological basis for the strong predictive contribution of milk production variables observed here. These variable-importance measures, however, reflect predictive contribution within the fitted models rather than evidence of causal effects. In particular, the RF %IncMSE and IncNodePurity measures can be influenced by correlations among predictors, which is relevant here because MY, FY, and PY are biologically related production traits. Because MY, FY, and PY are themselves correlated production traits, their individual importance values should therefore not be interpreted as independent contributions to LSCC, and the present analysis cannot determine how much of their predictive information is shared.

The importance of PY and FY may also reflect the fact that mammary inflammation can modify milk composition in addition to reducing milk volume. Leitner et al. [21] reported that intramammary infection altered milk composition and increased casein proteolysis, demonstrating that udder inflammation can affect the protein-related characteristics of milk. Likewise, Alhussien and Dang [22] reported that intramammary infection and the associated inflammatory response can alter milk composition and quality, supporting the potential value of milk protein- and fat-related traits as indicators of mammary health. These findings provide additional biological context for the importance of MY, FY, and PY in the present prediction models, although the direction and magnitude of their relationships with LSCC cannot be inferred from variable-importance values alone. AFC and DP followed different, secondary patterns. DP's estimated importance rose from the first to the third lactation in BNN (from 23.6 to 44.5 %IncMSE; Table 7), while its importance in RF followed a similar rise-and-decline pattern, peaking earlier before falling in the oldest cows (Table 6). This pattern is biologically plausible, since the dry period is associated with mammary involution, tissue remodeling, and changes in susceptibility to intramammary infection around the dry-off and calving periods [23]. Thus, variation in DP may therefore carry information related to udder health in the subsequent lactation, although the present analysis cannot establish a causal relationship between DP and SCC. AFC, by contrast, was most informative in the first one to two lactation groups and declined thereafter, becoming slightly negative in the BNN analysis for cows with more than four lactations. Together with the other reproductive and management traits, AFC contributed comparatively little predictive information for LSCC relative to the production variables, though this lower importance does not imply that reproductive performance is biologically unrelated to udder health.

4.4. Implications, Limitations, and Future Research

459 The predictor variables used in this study are all routinely recorded in commercial dairy herd
460 management systems, so the RF and BNN models could in principle be integrated into existing
461 recording systems without additional sensor infrastructure. In practice, predicted LSCC values
462 could be used to prioritize cows for additional monitoring or diagnostic testing, complementing
463 existing mastitis control programs based on SCC [1]. Because the dominant predictors
464 identified here, PY, MY, and FY, are production traits rather than SCC itself, these models
465 may be particularly useful for supporting SCC monitoring between routine sampling dates;
466 realizing this potential, however, would require validation under commercial conditions
467 beyond the single herd studied here. Several limitations should be considered when interpreting
468 these results. The data originated from a single commercial Holstein dairy enterprise, so
469 external validation in herds with different management systems, environments, and genetic
470 backgrounds is needed before the models can be generalized more broadly. The relatively small
471 number of records available for cows with more than four lactations ($n = 338$, corresponding
472 to training sets of approximately 250 records) limited the amount of information available for
473 both model training and independent evaluation in this group, and the resulting performance
474 estimates should therefore be interpreted cautiously. Multivariate outlier removal was
475 performed before model development, which reduced the number of observations available for
476 analysis; however, the specific influence of this preprocessing step on model performance,
477 particularly in the smaller higher-lactation groups, was not evaluated. Finally, the predictor set
478 was restricted to routinely recorded production, reproductive, and management variables;
479 future studies could evaluate whether adding direct health and treatment records, genomic
480 information, or sensor-derived behavioral measurements improves prediction, particularly in
481 higher-lactation cows. Future work should prioritize external validation using data from
482 multiple herds and production systems, and should evaluate whether a properly tuned BNN, an
483 ensemble combining BNN and RF predictions, or additional predictor variables can further

narrow the performance gap observed in the highest-parity group. Because the BNN and RF showed broadly similar predictive performance across lactation groups, future studies should also compare their computational requirements and ease of integration into herd-management systems, as practical deployment depends on factors beyond predictive accuracy alone.

Conclusion

This study compared seven modeling approaches for predicting LSCC in Holstein dairy cows from routinely recorded production, reproductive, and management traits, with separate models fitted for five lactation groups: MR and GLM as linear benchmarks, a Radial-Basis-Function SVM, and four nonlinear methods (BNN, RF, CHAID, and C&R Tree). Overall, BNN, RF, CHAID, and C&R Tree showed better predictive performance than SVM, MR, and GLM, and BNN and RF achieved comparable predictive accuracy, both matching or exceeding CHAID and C&R Tree in most lactation groups. Predictive accuracy generally declined with increasing lactation number, particularly in cows with more than four lactations, which may have been influenced by the smaller training sets and greater biological heterogeneity in the higher-lactation groups. Average protein yield, average milk yield, and average fat yield were consistently the most informative predictors of LSCC in both RF and BNN and generally ranked substantially higher than the reproductive and management traits evaluated. Average protein yield ranked first in most lactation groups, although average milk yield became the leading production trait among the oldest cows. Dry period length and age at first calving provided secondary, parity-dependent predictive information, with their relative importance varying across lactation groups. These findings indicate that routinely recorded production, reproductive, and management variables contain useful information for predicting LSCC and may help identify cows at risk between routine SCC tests. Because the models were developed and evaluated using data from a single herd, and because cows with more than four lactations were represented by comparatively few records, external validation across independent herds

and production systems is needed before the findings can be generalized or considered for practical application.

Author contributions

Mostafa Ghaderi-Zefrehei: Conceptualization, Methodology, Data curation, Supervision, Writing – original draft, and Writing – review and editing; Effat Nasre Esfahani: Data curation and Writing – review and editing; Morteza Jafaroghl: Methodology, and Writing – review and editing; Mahdi Talebian-Shahreza: Writing – review and editing; Alireza Abdolmohammadi: Writing – review and editing; Fatemeh Esmaeili: Data curation and Writing – review and editing; Saeid Ansari-Mahyari: Supervision, Conceptualization, and Writing – review and editing; Reza Reisi-Vanani: Conceptualization, Validation, Writing – original draft, and Writing – review and editing. All authors read and approved the final manuscript.

Funding

This research received no external funding.

Data availability

The data supporting the findings of this study were obtained from historical dairy cattle production and reproductive records. Due to data ownership restrictions, the dataset is not publicly available but may be obtained from the corresponding author upon reasonable request.

Ethics approval

Not applicable. This study was based exclusively on historical production and reproductive records from dairy cattle. No experimental procedures or interventions involving live animals were performed.

Competing interests

The authors declare that they have no competing interests.

References

1. Ruegg PL. A 100-Year Review: Mastitis detection, management, and prevention. *J Dairy Sci.* 2017;100(12):10381-97.
2. Hadrich J, Wolf C, Lombard J, Dolak T. Estimating milk yield and value losses from increased somatic cell count on US dairy farms. *J Dairy Sci.* 2018;101(4):3588-96.
3. Goyache F, Díez J, López S, Pajares G, Santos B, Fernández I, Prieto M. Machine Learning as an aid to management decisions on high somatic cell counts in dairy farms. *Arch Anim Breed.* 2005;48(2):138-48.
4. Bobbo T, Biffani S, Taccioli C, Penasa M, Cassandro M. Comparison of machine learning methods to predict udder health status based on somatic cell counts in dairy cows. *Sci Rep.* 2021;11(1):13642.
5. Ebrahimi M, Mohammadi-Dehcheshmeh M, Ebrahimie E, Petrovski KR. Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient-Boosted Trees outperform other models. *Comput Biol Med.* 2019;114:103456.
6. Pakrashi A, Ryan C, Guéret C, Berry D, Corcoran M, Keane MT, Mac Namee B. Early detection of subclinical mastitis in lactating dairy cows using cow-level features. *J Dairy Sci.* 2023;106(7):4978-90.
7. Anglart D, Hallén-Sandgren C, Emanuelson U, Rönnegård L. Comparison of methods for predicting cow composite somatic cell counts. *J Dairy Sci.* 2020;103(9):8433-42.
8. Thompson JS, Green MJ, Hyde R, Bradley AJ, O'Grady L. The use of machine learning to predict somatic cell count status in dairy cows post-calving. *Front Vet Sci.* 2023;10:1297750.
9. Ghaderi-Zefrehei M, Rabbanikhah E, Baneh H, Peters SO, Imumorin IG. Analysis of culling records and estimation of genetic parameters for longevity and some production traits in Holstein dairy cattle. *J Appl Anim Res.* 2017;45(1):524-8.

556 10. R Core Team. R: A language and environment for statistical computing. R foundation for
557 statistical computing, Vienna, Austria. 2021.

558 11. Filzmoser P, Maronna R, Werner M. Outlier identification in high dimensions. *Comput*
559 *Stat Data Anal.* 2008;52(3):1694-711.

560 12. Pérez-Rodríguez P, Gianola D, Weigel K, Rosa G, Crossa J. An R package for fitting
561 Bayesian regularized neural networks with applications in animal breeding. *J Anim Sci.*
562 2013;91(8):3522-31.

563 13. Hagnestam-Nielsen C, Emanuelson U, Berglund B, Strandberg E. Relationship between
564 somatic cell count and milk yield in different stages of lactation. *J Dairy Sci.* 2009;92(7):3124-
565 33.

566 14. Gonçalves JL, Cue RI, Botaro BG, Horst JA, Valloto AA, Santos MV. Milk losses
567 associated with somatic cell counts by parity and stage of lactation. *J Dairy Sci.*
568 2018;101(5):4357-66.

569 15. Gianola D, Okut H, Weigel KA, Rosa GJ. Predicting complex quantitative traits with
570 Bayesian neural networks: a case study with Jersey cows and wheat. *BMC Genet.*
571 2011;12(1):87.

572 16. Nørstebø H, Dalen G, Rachah A, Heringstad B, Whist AC, Nødtvedt A, Reksen O. Factors
573 associated with milking-to-milking variability in somatic cell counts from healthy cows in an
574 automatic milking system. *Prev Vet Med.* 2019;172:104786.

575 17. Xu W, van Kneegsel AT, Vervoort JJ, Bruckmaier RM, van Hoeij RJ, Kemp B, Saccenti E.
576 Prediction of metabolic status of dairy cows in early lactation with on-farm cow data and
577 machine learning algorithms. *J Dairy Sci.* 2019;102(11):10186-201.

578 18. Laevens H, Deluyker H, Schukken Y, De Meulemeester L, Vandermeersch R, De
579 Muelenaere E, De Kruif A. Influence of parity and stage of lactation on the somatic cell count
580 in bacteriologically negative dairy cows. *J Dairy Sci.* 1997;80(12):3219-26.

581 19. Williams M, Sleator R, Murphy C, McCarthy J, Berry D. Exploiting genetic variability in
582 the trajectory of lactation yield and somatic cell score with each progressing parity. *J Dairy*
583 *Sci.* 2022;105(4):3341-54.

584 20. Halasa T, Huijps K, Østerås O, Hogeveen H. Economic effects of bovine mastitis and
585 mastitis management: A review. *Vet Q.* 2007;29(1):18-31.

586 21. Leitner G, Krifucks O, Merin U, Lavi Y, Silanikove N. Interactions between bacteria type,
587 proteolysis of casein and physico-chemical properties of bovine milk. *Int Dairy J.*
588 2006;16(6):648-54.

589 22. Alhussien MN, Dang AK. Pathogen-dependent modulation of milk neutrophils
590 competence, plasma inflammatory cytokines and milk quality during intramammary infection
591 of Sahiwal (*Bos indicus*) cows. *Microb Pathog.* 2018;121:131-8.

592 23. Zhao X, Ponchon B, Lanctôt S, Lacasse P. Invited review: Accelerating mammary gland
593 involution after drying-off in dairy cattle. *J Dairy Sci.* 2019;102(8):6701-17.

594 24. Lee Y, Nelder JA. Hierarchical generalized linear models. *J R Stat Soc Series B Stat*
595 *Methodol.* 1996;58(4):619-56.

596 25. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5-32.

597 26. Liaw A, Wiener M. Classification and regression by randomForest. *R News.* 2002;2(3):18-
598 22.

599 27. Boser BE, Guyon IM, Vapnik VN, editors. A training algorithm for optimal margin
600 classifiers. *Proceedings of the fifth annual workshop on Computational learning theory*; 1992.

601 28. Draper NR, Smith H. *Applied regression analysis*: John Wiley & Sons; 1998.

602 29. Kass GV. An exploratory technique for investigating large quantities of categorical data. *J*
603 *R Stat Soc Ser C Appl Stat.* 1980;29(2):119-27.

604 30. Breiman L, Friedman J, Stone CJ, Olshen RA. *Classification and Regression Trees*: Taylor
605 & Francis; 1984.

Table 1. Descriptive statistics of study variables across lactation groups.

Variable	Unit	Lactation number	n	Mean	SD	Minimum	Maximum
DIM	d	1	4,147	331.91	64.27	239	582
		2	2,836	339.41	59.33	206	627
		3	1,706	328.46	53.06	228	561
		4	888	326.73	51.69	214	515
		>4	338	317.57	43.24	240	518
AFC	d	1	4,147	1131.45	88.78	799	1803
		2	2,836	1545.28	135.13	1116	2221
		3	1,706	1935.33	163.07	1442	2846
		4	888	2313.02	174.33	2010	2933
		>4	338	2682.65	180.17	2377	3340
CI	d	1	4,147	389.75	65.38	294	639
		2	2,836	393.08	63.98	298	632
		3	1,706	383.27	58.59	301	627
		4	888	383.22	56.39	305	596
		>4	338	373.28	49.20	309	595
DP	d	1	4,147	57.69	7.94	1	129
		2	2,836	53.71	17.50	12	143
		3	1,706	54.62	17.15	18	140
		4	888	56.31	18.71	16	136
		>4	338	55.43	18.13	21	133
CTFS	d	1	4,147	61.57	20.65	28	143
		2	2,836	56.14	17.58	26	137
		3	1,706	55.14	16.61	24	136
		4	888	54.00	16.57	26	129
		>4	338	53.37	14.45	31	117
GL	d	1	4,147	277.82	4.92	256	290
		2	2,836	278.37	5.14	256	290
		3	1,706	278.59	5.06	257	290
		4	888	278.75	4.77	258	290
		>4	338	278.70	4.62	259	290
DO	d	1	4,147	111.94	65.12	28	362
		2	2,836	114.71	63.68	29	351
		3	1,706	104.67	58.31	29	341
		4	888	104.47	56.14	32	318
		>4	338	94.58	48.86	35	314
FSTC	d	1	4,147	50.37	62.21	0	299
		2	2,836	58.57	62.53	0	298
		3	1,706	49.53	56.58	0	288
		4	888	50.47	54.85	0	266
		>4	338	41.21	47.00	0	253
MY	kg	1	4,147	10882.32	2019.98	5150.39	18565.48
		2	2,836	11611.63	2198.28	4955.46	19117.10
		3	1,706	11555.83	2175.47	5959.48	18195.54
		4	888	11678.31	2119.19	5592.76	17514.15

		>4	338	11740.47	1835.57	7788.92	17227.60
FY	kg	1	4,147	301.85	66.94	104.77	525.44
		2	2,836	321.66	74.50	103.02	532.94
		3	1,706	322.91	72.04	120.99	564.66
		4	888	330.88	70.61	136.83	527.19
		>4	338	338.07	63.63	185.09	505.79
PY	kg	1	4,147	285.27	75.81	44.24	557.22
		2	2,836	310.41	81.09	63.52	555.54
		3	1,706	310.01	80.00	60.12	577.19
		4	888	315.29	79.68	94.16	564.81
		>4	338	323.13	70.57	86.26	551.67
LSCC	–	1	4,147	5.00	2.52	2.30	13.17
		2	2,836	5.63	2.53	2.59	13.77
		3	1,706	5.84	2.51	2.53	13.82
		4	888	5.96	2.47	2.56	13.19
		>4	338	5.74	2.22	3.05	13.44

Abbreviations: DIM, days in milk; AFC, age at first calving; CI, calving interval; DP, dry period; CTFS, days from calving to first service; GL, gestation length; DO, days open; FSTC, days from first service to conception; MY, average milk yield; FY, average fat yield; PY, average protein yield; LSCC = log-transformed somatic cell count.

612 **Table 2.** Mathematical formulation, key properties, and references for the seven modeling approaches used in this study.

Algorithm	Key properties	Formulation (simplified)	citation(s)
BNN	Single-hidden-layer feed-forward network; weights assigned Gaussian priors and estimated by Bayesian regularization, which limits overfitting without a separate validation set.	$\hat{y} = g(\sum_{j=1}^H w_{2j} h(\sum_{i=1}^d w_{1ji} x_i + b_{1j}) + b_2), w \sim \mathcal{N}(0, \sigma^2)$	Lee and Nelder [24]; brnn package [12]
RF	Ensemble of decision trees, each grown on a bootstrap sample using a random subset of predictors at each split; predictions are averaged across trees.	$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(\mathbf{x})$	Breiman [25]; randomForest package [26]
SVM	Maps predictors into a higher-dimensional feature space via a kernel function and fits a regression hyperplane within an ϵ -insensitive margin; used here with a Radial Basis Function (RBF) kernel.	$\hat{y} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b, K(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma \ \mathbf{x}_i - \mathbf{x}\ ^2)$	Boseret al. [27]
MR	Classical ordinary least-squares regression assuming a linear, additive relationship between predictors and the outcome.	$\hat{y} = \beta_0 + \sum_{j=1}^d \beta_j x_j + \varepsilon$	Classical method; see standard texts [e.g., 28]
GLM	Extends linear regression to allow a specified error distribution and link function relating the linear predictor to the expected outcome; used here with a Gaussian distribution and identity link.	$g(E[Y]) = \beta_0 + \sum_{j=1}^d \beta_j x_j$, with $g(\mu) = \mu$	Lee and Nelder [24]
CHAID	Decision-tree method that recursively partitions the data using chi-square (categorical) or F-test (continuous) significance tests, allowing multi-way splits.	Recursive partitioning by statistical significance testing (no single closed-form equation)	Kass [29]; IBM SPSS Modeler
C&R Tree	Binary decision-tree method (CART) that recursively splits the data to minimize within-node variance.	$\min_{\text{splits}} [\sum_{i \in \mathcal{N}_L} (y_i - \bar{y}_L)^2 + \sum_{i \in \mathcal{N}_R} (y_i - \bar{y}_R)^2]$	Breiman et al. [30]; IBM SPSS Modeler

613 Note: formulations are simplified for summary purposes; see the cited sources for complete derivations. BNN, Bayesian
614 Neural Network; RF, Random Forest; SVM, Support Vector Machine; MR, Multiple Regression; GLM, Generalized Linear
615 Model; C&R Tree, Classification and Regression Tree.

617
618

Table 3. Comparison of machine learning models for predicting log-transformed somatic cell count (LSCC) across lactation groups based on the Pearson correlation coefficient (r) and relative error (RE) in the training and testing datasets.

Lactation Group	Model	Pearson Correlation		Relative Error	
		Training	Testing	Training	Testing
1	SVM	0.571508	0.593571	0.80692	0.806309
	Multiple Regression Model	0.573166	0.611412	0.671481	0.631987
	Bayesian Neural Network	0.673712	0.699396	0.547959	0.513694
	Generalized Linear Model	0.578558	0.61495	0.665271	0.626892
	CHAID	0.69767	0.657918	0.513257	0.568086
	C&R Tree	0.680236	0.66757	0.53728	0.554478
2	SVM	0.599678	0.546472	0.784324	0.792862
	Multiple Regression Model	0.601318	0.511355	0.638417	0.783592
	Bayesian Neural Network	0.666962	0.630025	0.555531	0.612548
	Generalized Linear Model	0.606707	0.512756	0.631907	0.783032
	CHAID	0.686308	0.595572	0.528981	0.654443
	C&R Tree	0.666952	0.574919	0.555175	0.675328
3	SVM	0.58277	0.569092	0.789084	0.794605
	Multiple Regression Model	0.574152	0.570024	0.67035	0.675308
	Bayesian Neural Network	0.633505	0.61764	0.600126	0.622833
	Generalized Linear Model	0.578781	0.573686	0.665013	0.671263
	CHAID	0.658598	0.525753	0.566249	0.741636
	C&R Tree	0.649638	0.575545	0.577971	0.677684
4	SVM	0.535624	0.466254	0.887387	0.967003
	Multiple Regression Model	0.543049	0.500816	0.705098	0.759794
	Bayesian Neural Network	0.569328	0.511366	0.680399	0.772694
	Generalized Linear Model	0.548262	0.511806	0.699409	0.747128
	CHAID	0.68456	0.608306	0.531378	0.648137
	C&R Tree	0.71787	0.61008	0.484663	0.647477
>4	SVM	0.566071	0.450795	0.799422	0.920564
	Multiple Regression Model	0.569731	0.475476	0.675406	0.783211
	Bayesian Neural Network	0.585223	0.483904	0.660072	0.776078
	Generalized Linear Model	0.570517	0.4766	0.674511	0.78248
	CHAID	0.674878	0.472955	0.54454	0.824294
	C&R Tree	0.706395	0.509013	0.501006	0.805208

619

620
621

Table 4. Predictive performance of Bayesian Neural Network (BNN) models for predicting log-transformed somatic cell count (LSCC) across lactation groups using different numbers of hidden neurons.

Lactation Group	Neurons	Training			Testing		
		N	RMSE	R	N	RMSE	R
1	2	3110	1.895513	0.6625239	1037	1.927305	0.6394595
	3		1.824793	0.6928305		1.874573	0.6631869
	4		1.809842	0.6989305		1.895868	0.6537723
	5		1.800633	0.7026284		1.893329	0.6547079
2	2	2127	1.94187	0.6409309	709	1.983995	0.624046
	3		1.896784	0.6616906		1.927345	0.6511419
	4		1.880454	0.6689355		1.919255	0.6548107
	5		1.890172	0.6646455		1.930389	0.6496981
3	2	1279	2.064398	0.5811196	427	2.004532	0.5719624
	3		1.953725	0.6378206		1.951846	0.6040221
	4		1.933169	0.6474803		1.954756	0.6052395
	5		1.877445	0.6724869		1.916566	0.6248245
4	2	666	1.902887	0.6386026	222	1.858909	0.6563965
	3		1.866938	0.6557289		1.812037	0.6773859
	4		1.840013	0.6681687		1.873284	0.6535701
	5		1.820385	0.6769324		1.847855	0.6648775
>4	2	253	1.811449	0.5671613	85	1.853501	0.5909644
	3		1.81205	0.566787		1.854519	0.5904474
	4		1.812916	0.5661692		1.847879	0.5938804
	5		1.812781	0.5662755		1.849862	0.5929381

622

623 **Table 5.** Predictive performance of the Random Forest (RF) model for predicting log-transformed somatic cell count
624 (LSCC) across lactation groups.

Lactation Group	Training			Testing		
	N	RMSE	r	N	RMSE	r
1	3110	1.858526	0.6795979	1037	1.888527	0.6593104
2	2127	1.936976	0.643759	709	1.878207	0.6746033
3	1279	2.002806	0.6142743	427	1.925376	0.6153107
4	666	1.971676	0.6036149	222	1.85584	0.6613924
>4	253	1.963204	0.4532289	85	1.881761	0.5732001

625

626

Table 6. Variable importance of predictors for log-transformed somatic cell count (LSCC) estimated by the Random Forest (RF) model across lactation groups.

Variables	%IncMSE					IncNodePurity				
	Lactation Groups					Lactation Groups				
	1	2	3	4	>4	1	2	3	4	>4
DIM	22.1	18.9	15.2	9.2	7.1	892.1	590.3	362.3	167.3	68.8
AFC	27.3	18.1	15.3	3.8	2.0	1635.1	1027.3	537.1	261.0	71.2
CI	19.6	17.3	15.5	12.7	7.6	853.8	615.1	396.2	183.2	63.2
DP	22.1	27.3	16.9	12.8	7.0	1059.1	1542.8	767.7	394.8	107.9
CTFS	17.5	7.2	9.2	7.7	3.0	834.3	632.5	408.0	210.5	81.0
GL	18.7	7.9	9.4	4.4	-2.2	739.0	517.1	402.4	187.3	58.7
DO	20.7	20.3	14.6	11.8	6.8	844.1	577.7	366.7	184.2	53.4
FSTC	16.6	18.5	11.3	6.9	5.2	575.6	473.1	261.2	111.8	42.4
MY	52.4	35.3	27.5	28.2	14.1	3157.4	2020.0	1348.6	828.7	245.9
FY	36.5	24.3	24.1	19.2	9.1	2138.2	1413.1	819.3	390.8	124.3
PY	71.8	53.0	43.0	29.2	6.3	5327.4	2925.3	1717.7	738.5	142.6

Abbreviations: DIM, days in milk; AFC, age at first calving; CI, calving interval; DP, dry period; CTFS, days from calving to first service; GL, gestation length; DO, days open; FSTC, days from first service to conception; MY, average milk yield; FY, average fat yield; PY, average protein yield.

Table 7. Variable importance of predictors for log-transformed somatic cell count (LSCC) estimated by the Bayesian Neural Network (BNN) model across lactation groups.

Variables	%IncMSE					IncNodePurity				
	Lactation Groups					Lactation Groups				
	1	2	3	4	>4	1	2	3	4	>4
DIM	16.98	21.36	23.13	15.23	13.70	290.58	192.36	99.72	56.87	10.27
AFC	43.89	47.52	18.54	19.70	-0.38	549.17	248.83	113.74	63.42	6.46
CI	16.44	21.89	22.66	17.62	13.91	267.43	171.62	119.41	47.19	14.92
DP	23.64	36.03	44.48	31.77	14.00	355.09	586.67	482.32	203.93	23.89
CTFS	13.80	8.78	7.05	8.43	9.44	121.59	76.84	63.78	30.59	12.12
GL	10.88	7.84	9.72	7.65	-0.70	119.34	64.99	64.27	25.82	4.62
DO	16.84	15.85	20.52	17.22	13.93	216.61	129.00	100.85	55.56	13.80
FSTC	14.53	16.47	14.78	13.85	6.78	103.23	81.81	64.10	31.05	6.20
MY	54.18	41.98	48.10	39.71	35.62	1995.37	1225.61	825.40	537.16	113.90
FY	49.17	42.01	40.94	14.64	20.94	1193.60	789.89	389.01	137.51	49.03
PY	80.04	51.18	47.49	35.87	24.16	4043.95	2254.82	1215.09	519.13	82.46

Abbreviations: DIM, days in milk; AFC, age at first calving; CI, calving interval; DP, dry period; CTFS, days from calving to first service; GL, gestation length; DO, days open; FSTC, days from first service to conception; MY, average milk yield; FY, average fat yield; PY, average protein yield.

639
640

Table 8. Comparison of the predictive performance of Bayesian Neural Network (BNN) and Random Forest (RF) models for predicting log-transformed somatic cell count (LSCC) across lactation groups.

Method	Lactation Neurons		Training N	Training RMSE	Training r	Testing N	Testing RMSE	Testing r
Bayesian Neural Network	1	3	3110	1.824793	0.692831	1037	1.874573	0.663187
Random Forest	1	–	3110	1.858526	0.679598	1037	1.888527	0.659310
Bayesian Neural Network	2	4	2127	1.880454	0.668936	709	1.919255	0.654811
Random Forest	2	–	2127	1.936976	0.643759	709	1.878207	0.674603
Bayesian Neural Network	3	5	1279	1.877445	0.672487	427	1.916566	0.624825
Random Forest	3	–	1279	2.002806	0.614274	427	1.925376	0.615311
Bayesian Neural Network	4	3	666	1.866938	0.655729	222	1.812037	0.677386
Random Forest	4	–	666	1.971676	0.603615	222	1.855840	0.661392
Bayesian Neural Network	>4	4	253	1.812916	0.566169	85	1.847879	0.593880
Random Forest	>4	–	253	1.963204	0.453229	85	1.881761	0.573200

641

642

643
644

Table 9. Comparison of predictor variable importance estimated by Bayesian Neural Network (BNN) and Random Forest (RF) models across lactation groups.

Method	Variables	%IncMSE					IncNodePurity				
		Lactation Groups					Lactation Groups				
		1	2	3	4	>4	1	2	3	4	>4
Bayesian Neural Network	DIM	16.98	21.36	23.13	15.23	13.70	290.58	192.36	99.72	56.87	10.27
	AFC	43.89	47.52	18.54	19.70	-0.38	549.17	248.83	113.74	63.42	6.46
	CI	16.44	21.89	22.66	17.62	13.91	267.43	171.62	119.41	47.19	14.92
	DP	23.64	36.03	44.48	31.77	14.00	355.09	586.67	482.32	203.93	23.89
	CTFS	13.80	8.78	7.05	8.43	9.44	121.59	76.84	63.78	30.59	12.12
	GL	10.88	7.84	9.72	7.65	-0.70	119.34	64.99	64.27	25.82	4.62
	DO	16.84	15.85	20.52	17.22	13.93	216.61	129.00	100.85	55.56	13.80
	FSTC	14.53	16.47	14.78	13.85	6.78	103.23	81.81	64.10	31.05	6.20
	MY	54.18	41.98	48.10	39.71	35.62	1995.37	1225.61	825.40	537.16	113.90
	FY	49.17	42.01	40.94	14.64	20.94	1193.60	789.89	389.01	137.51	49.03
	PY	80.04	51.18	47.49	35.87	24.16	4043.95	2254.82	1215.09	519.13	82.46
Random Forest	DIM	22.1	18.9	15.2	9.2	7.1	892.1	590.3	362.3	167.3	68.8
	AFC	27.3	18.1	15.3	3.8	2.0	1635.1	1027.3	537.1	261.0	71.2
	CI	19.6	17.3	15.5	12.7	7.6	853.8	615.1	396.2	183.2	63.2
	DP	22.1	27.3	16.9	12.8	7.0	1059.1	1542.8	767.7	394.8	107.9
	CTFS	17.5	7.2	9.2	7.7	3.0	834.3	632.5	408.0	210.5	81.0
	GL	18.7	7.9	9.4	4.4	-2.2	739.0	517.1	402.4	187.3	58.7
	DO	20.7	20.3	14.6	11.8	6.8	844.1	577.7	366.7	184.2	53.4
	FSTC	16.6	18.5	11.3	6.9	5.2	575.6	473.1	261.2	111.8	42.4
	MY	52.4	35.3	27.5	28.2	14.1	3157.4	2020.0	1348.6	828.7	245.9
	FY	36.5	24.3	24.1	19.2	9.1	2138.2	1413.1	819.3	390.8	124.3
	PY	71.8	53.0	43.0	29.2	6.3	5327.4	2925.3	1717.7	738.5	142.6

645
646
647

Abbreviations: DIM, days in milk; AFC, age at first calving; CI, calving interval; DP, dry period; CTFS, days from calving to first service; GL, gestation length; DO, days open; FSTC, days from first service to conception; MY, average milk yield; FY, average fat yield; PY, average protein yield.

648

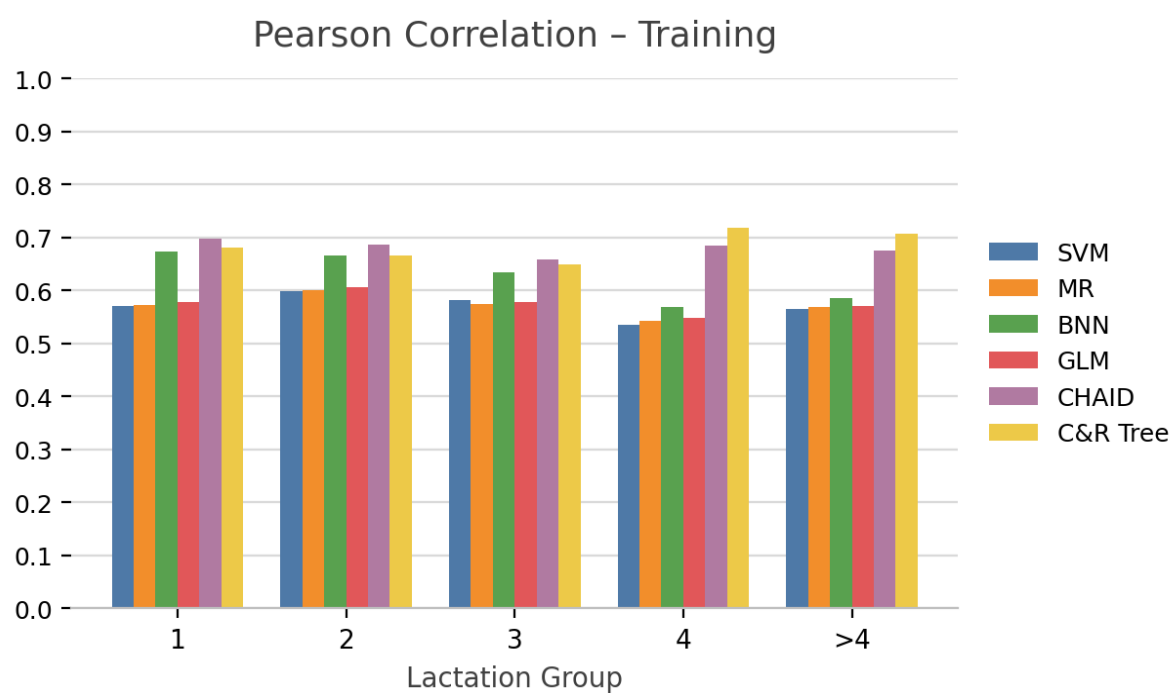


Figure 1. Comparison of machine learning model performance based on Pearson correlation coefficient in the training dataset across lactation groups.

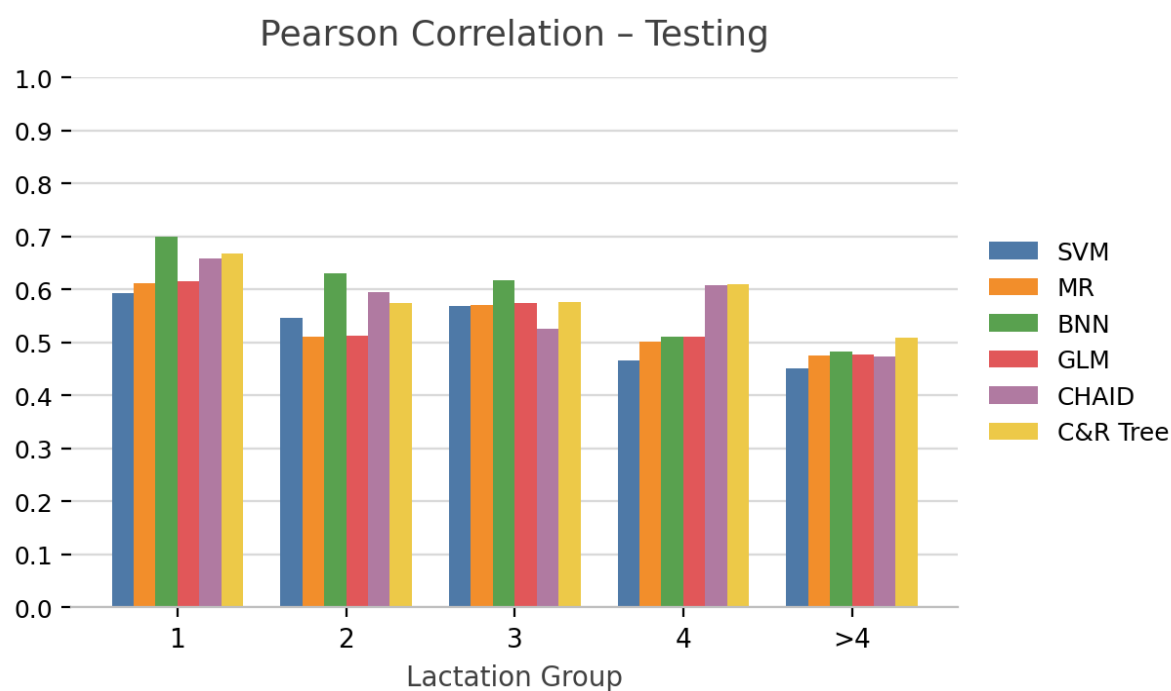


Figure 2. Comparison of machine learning model performance based on Pearson correlation coefficient in the testing dataset across lactation groups.



Figure 3. Comparison of machine learning model performance based on relative error in the training dataset across lactation groups.

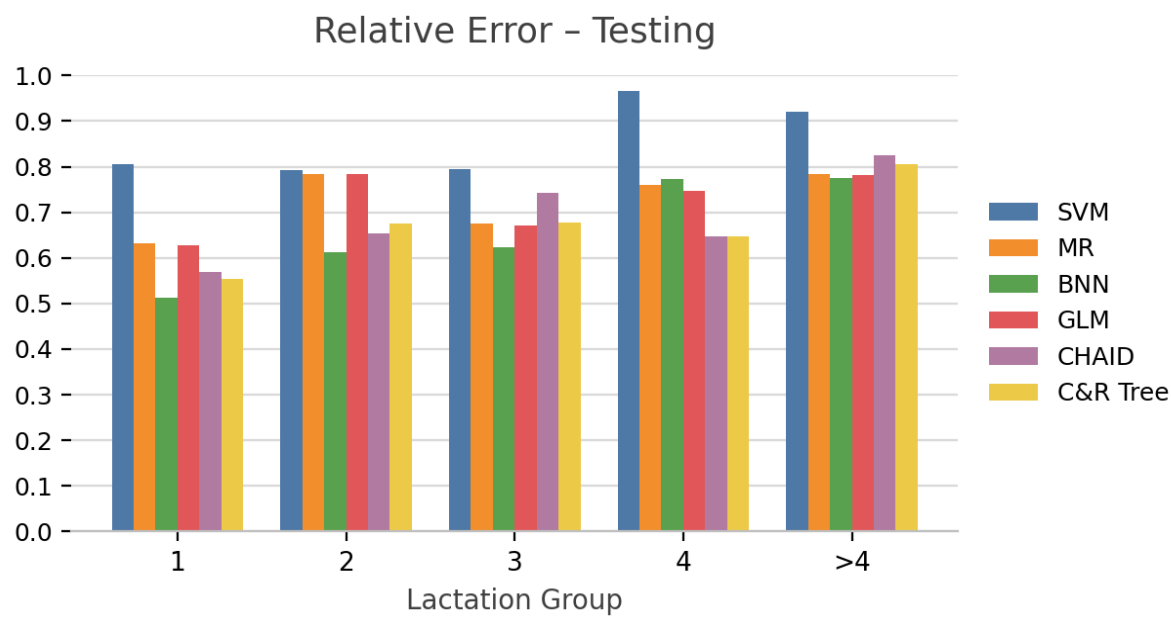


Figure 4. Comparison of machine learning model performance based on relative error in the testing dataset across lactation groups.