ULS Timisoara
Multidisciplinary Conference on Sustainable Development
15-16 May 2025

Biochemical Predictors of Udder Health: Linking Milk Composition to Somatic Cell Distribution

Bratu Daniel George [1,2], Blaga Șerban[1,2] , Vizitiu Dorin Alexandru[1,2] , Lungu Bianca[1,2] , Ilie Elena Daniela[3] , Mizeranschi Eugeniu Alexandru[3] , Ludovic Toma Cziszter[1,3] , Spătaru Irina[1] , Torda Iuliu[1] , Mircu Călin[1] and Huțu Ioan [1,2]

1. University of Life Science "The King Michael I" – Faculty of Veterinary Medicine, Bioengineering Faculty of Animal Resources, 119th Calea Aradului, Timisoara, 300645, RO.
2. Horia Cernescu Research Unit – USV Timisoara, RO.
3 . Research and Development Station for Bovine Arad, 310059 Arad, Romania
**Correspondence: bianca.lungu@fmvt.ro*

Abstract: *Somatic cell count (SCC) and differential somatic cell count (DSCC) are widely recognized as reliable indicators of udder health and inflammation in dairy cows. Recent advances in statistical modeling and biochemical analysis have enabled more refined pre-dictions of these indicators using milk composition traits. The objective of this study was to investigate the associations of key milk biochemical parameters (fat, protein, casein, lactose, solids non-fat, total dry matter, pH, urea, acetone, and β-hydroxybutyrate-BHB) with udder health, and to identify the most robust predictive model for SCC. A total of 272 milk samples were collected from cows under the Official Performance Recording Milk Production program in Arad County, Romania, and an-alyzed using the CombiFoss™ FT+ system. After outlier removal, 251 samples were retained for analysis. Data preprocessing in-cluded standardization and log-transformation of SCC to improve model assumptions. Statistical modeling involved stepwise re-gression with interaction terms, as well as Ridge and Lasso regularization techniques. The best results were obtained using the Lasso model for log(SCC), which achieved the highest predictive accuracy (R^2 = 0.655), selecting biologically relevant predictors such as protein, BHB, casein, lactose, and DSCC. The model's predictions aligned closely with measured values and confirmed known correlations, such as the negative association between lactose and SCC and the positive association of BHB and protein with SCC. Despite the Lasso model for DSCC showed lower predictive power (R^2 = 0.175), it selected key predictors, protein, ace-tone, and BHB, similarly to other models, underlining the complex physiological basis of DSCC. Ridge regression confirmed these trends, supporting the robustness of the selected variables. These findings emphasize the utility of regularized regression, partic-ularly Lasso, in developing practical tools for early mastitis screening based on routinely collected milk composition data, with potential applications in herd health monitoring and decision-making on commercial dairy farms.*

• Introduction

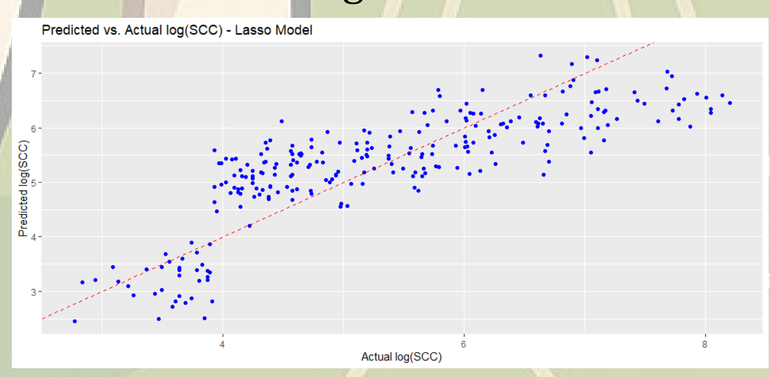
Udder health affects milk quality, animal welfare, and farm profitability. Somatic cell count (SCC) and differential SCC (DSCC) are key indicators of mastitis, a common inflammatory condition that may occur subclinical. Changes in milk composition, such as fat, protein, casein, lactose, urea, and ketone bodies can signal early inflammation.
This study had two main goals: to assess how milk components relate to SCC and DSCC, and to identify the best predictive model for udder health, comparing stepwise regression with interaction terms, Ridge, and Lasso techniques.

• Material and method

Milk samples were collected from dairy cows enrolled in the official performance recording program (COP) in Arad County, Romania. Analyses were performed at the Milk Quality Control Foundation (Cluj-Napoca) using the CombiFoss™ FT+ system, measuring fat, protein, casein, lactose, urea, BHB, acetone, SCC, and DSCC.
Data preprocessing included standardization, outlier removal (±3 SD), and log-transformation of SCC. From 272 initial samples, 251 were retained for analysis.
All processing was performed in R (v4.4.0) with tidyverse and modeling packages. Statistical modeling involved stepwise linear regression with interaction terms, Ridge, and Lasso regression. Models were cross-validated (10-fold), and performance was evaluated using R^2 and residual analysis. AI tools (Scite.ai and GPT-4) were used to support literature synthesis and manuscript clarity, with all outputs reviewed for accuracy.

• Results and discussions

From 272 milk samples, 251 were retained after quality control. Due to skewed distribution, SCC values were log-transformed for modeling. Milk components such as protein, fat, casein, and lactose showed strong intercorrelations, while ketone bodies (BHB, acetone) were more independent. These patterns are illustrated in the correlation heatmap , where casein–protein (r = 0.76) and solids–nonfat (r = 0.91) stood out. Stepwise regression explained 46% of the variance in log(SCC), highlighting interactions like fat × lactose and casein × pH.
Lasso regression achieved the best performance (R^2 = 0.655), selecting protein, BHB, DSCC, and lactose as key predictors. The scatter plot of predicted vs. actual log(SCC) shows high accuracy and model fit. These results confirm that milk composition data can help detect mastitis risk early. Lasso regression provides a powerful, interpretable tool for integrating routine milk testing into udder health monitoring.



• Conclusions

Lasso regression showed strong predictive power for somatic cell levels (log(SCC)) based on milk composition, identifying protein, BHB, lactose, casein, and DSCC as key variables. DSCC was harder to predict, likely due to its biological complexity, but still added value as an udder health indicator. Ridge models confirmed similar trends, reinforcing the robustness of results. These findings support the use of milk biochemical data and regularized regression for early mastitis risk detection in dairy cows.

Acknowledgement: The author acknowledge for database and technical support given by the SCDCB ARAD.