Using test day data and neural network regression to rank cows based on their future lifetime milk yield revenue

L. Fadul[1], R. Lacroix[1], D. Warner[1], M. Ayat[1], R. Cue[2] and D. Lefebvre[1]

[1]Lactanet, Dairy Center of Expertise, Sainte-Anne-de-Bellevue, QC, Canada
[2]McGill University, Animal Science Department, Sainte-Anne-de-Bellevue, QC, Canada

Abstract

Increasing the longevity of dairy cows can lead to a more sustainable dairy industry and enhance farm profitability. Farmers often face important herd management decisions, such as determining which animals to keep. Integrating data with advanced analytics can help streamline this process. This study aimed to develop a model to predict future cow lifetime revenue based on yields using Dairy Herd Improvement (DHI) data. Data was sourced from Lactanet (Canadian Network for Dairy Excellence), covering 2,296 herds over eight years (2017-2024). A multilayer perceptron regressor from the scikit-learn package in Python was used to predict the lifetime cumulative yields of milk, fat, and protein at any given age in the future. These predicted yields were then used to estimate projected income. The model utilized 14 input features, including age on test day, parity, days in milk, fat and protein yields, somatic cell count linear score, current lactation and lifetime yields, and age at test. Up to eight consecutive test day records were randomly selected for each cow for prediction, with a minimum of one test. The inclusion of multiple breeds ensured the model's breed-agnostic nature. The model demonstrated high accuracy in predicting milk, fat, and protein yields, with R^2 values of 0.97 for each yield (MAPE: 5.68%, 5.92%, and 5.51%, respectively). Additionally, the projected income from the predicted yields showed a strong correlation with the observed income (R^2 = 0.97, MAPE = 5.53%). Consistent results were found across breeds (e.g., Holstein R^2=0.97, Jersey R^2=0.94, Ayrshire R^2=0.97). The model effectively predicted cumulative yields at any future age, with an average prediction range of 5.4 years ahead. These findings suggest that neural networks can help farmers make informed decisions by predicting future cow lifetime yields and ranking animals within a herd across all age classes.

Introduction

In most modern dairy production systems, there is a general agreement on the importance of improving longevity, from a consumer acceptability and a sustainability perspective. The longevity of a cow can be defined as the length of its productive life and total lifespan (Schuster *et al.,* 2020). Economic decisions are the primary driver of culling and, therefore, of the life cycle of a cow (Vries, 2020). Lifetime milk production, as the kg of milk produced by the cow during her lifetime, represents a more global perspective (i.e., economic and environmental) on longevity (Pritchard *et al.,* 2013).

The amount of available data on dairy herds continues to increase, providing an opportunity to support informed decision-making. However, the lack of data integration renders it difficult to fully utilize the data's potential (Cockburn, 2020; Ferris *et al.,* 2020). The use of digital technologies can help better explore the added value of the data and develop tools that support the decision-making process. Therefore, the aim was to develop a model to predict future cow lifetime revenue based on yields using Dairy Herd Improvement (DHI) data.

Materials and methods

Data was sourced from Lactanet (Canadian Network for Dairy Excellence), covering 2,296 herds over eight years (2017-2024), with a total of 539,072 cows and 12.5 million test records. For model training, 80% of the data (1,836 herds) was used, while 20% (460 herds) was reserved for testing. A multilayer perceptron regressor from the scikit-learn package in Python was used to predict the lifetime cumulative yields of milk, fat, and protein at any given age in the future. These predicted yields were then used to estimate projected income. The model utilized 14 input features, including age on test day, parity, days in milk, fat and protein yields, somatic cell count linear score, current lactation and lifetime yields, and age at test at any given age in the future. These predicted yields were then used to estimate projected income. For training and testing, only cows with at least 10 test-day records were included. The number of available records per cow ranged from 10 to 74 in the testing data set, with an average of 23 records per cow. Up to eight consecutive test-day records were randomly selected from these records for each cow for prediction, with a minimum of one test-day record per cow. The inclusion of multiple breeds ensured the model's breed-agnostic nature.

Results

The model demonstrated high accuracy in predicting milk, fat, and protein yields, with R^2 values of 0.97 for each yield (MAPE: 5.68%, 5.92%, and 5.51%, respectively; Figure 1). Additionally, the projected income from the predicted yields showed a strong correlation with observed income (R^2 = 0.97, MAPE = 5.53%). Consistent results were found across breeds (Figure 2), which confirms that the model is breed-agnostic. We observed a cluster of cows on the left-upper side of the graph, for which the predictions were not as good as for most cows. Those outliers appear to be due to errors in the lifetime records. They represent only 1.4 % of cows, which does not significantly affect the overall performance of the model. We are currently working on setting up rules to exclude such cases for the implementation of the model.

The model effectively predicted cumulative yields at any future age, with an average prediction range of 5.4 years forward. Furthermore, the number of tests used for predictions did not affect the model's performance (Figure 3).

The prediction accuracy depended on the age of the cow at the time of prediction, the time to prediction, and breed (Figure 4). Therefore, we calculated the prediction error for each cow by estimating the standard deviation (SD) of the prediction inherent to the age class, time to prediction, and breed using a linear mixed-effects regression model. For this purpose, the NLME package in R was used to allow for heterogeneous variances. Briefly, the differences between observed and projected lifetime milk yield, fat yield, and protein yield were each modelled based on the aforementioned factors, including all possible interactions, and considering the herd as a random effect. Weight statements in the NLME models were defined to allow for different variances for each level of the factor. Various models with different model structures were tested, and the model with the lowest Bayesian Information Criterion (BIC) was selected. Finally, the combined SD of the residuals and of each modelled parameter (i.e. for the three factors age class, time to prediction and breed, as well as the interactions) was computed. This procedure was applied to the projected lifetime yields for milk, fat and protein. To estimate the prediction error of the projected lifetime income, an additional computation step was included. As the market prices for the achieved yields

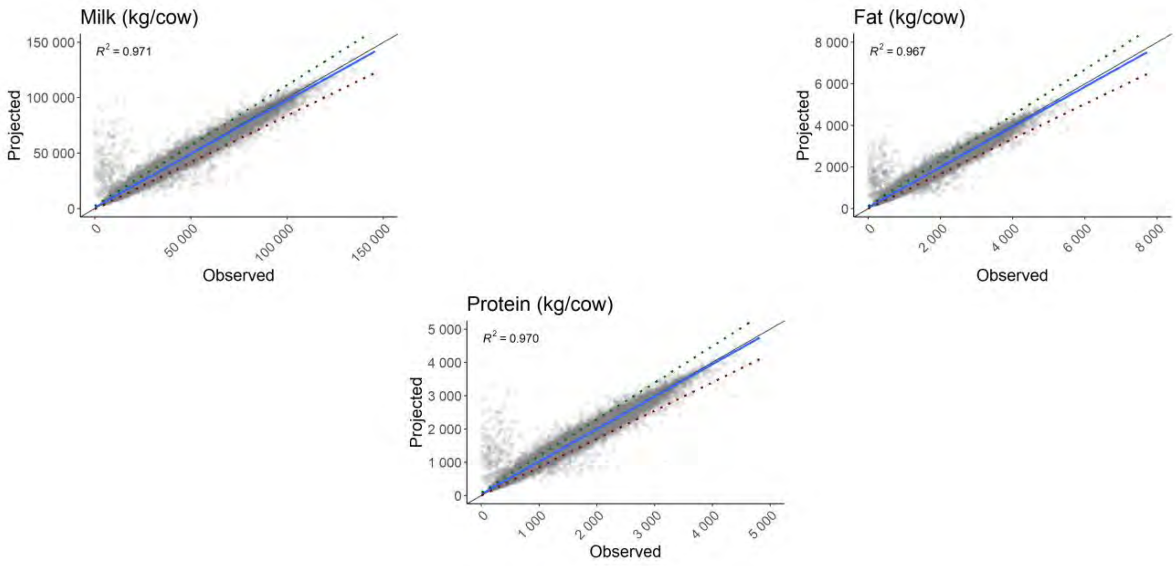


Figure 1 Results of the projected vs observed cumulative milk, fat and protein yields with the confidence interval (dot line), identity line (black) and regression line (blue).

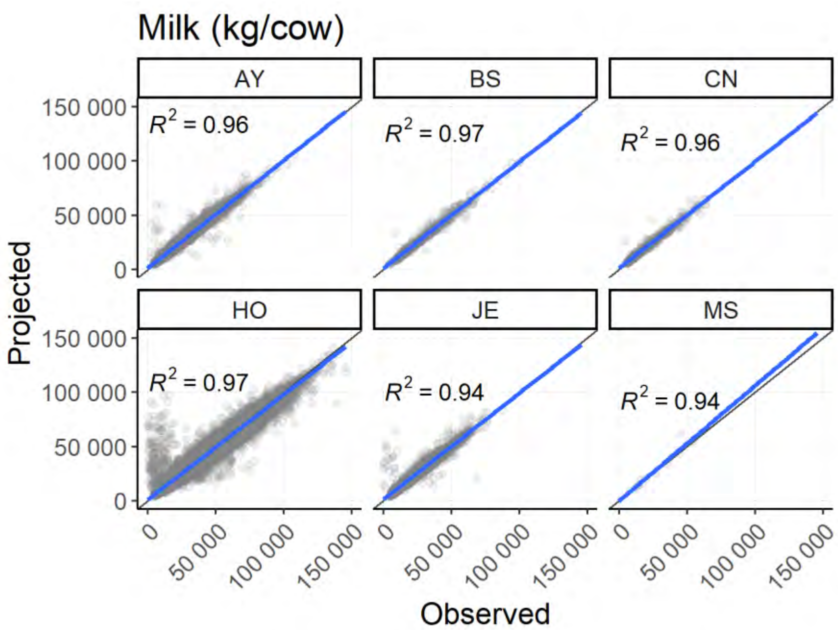


Figure 2 Results of the projected vs observed cumulative milk, fat and protein yields for different breeds (AY= Ayrshire, BS = Brown Swiss, CN= Canadienne, HO = Holstein, JE= Jersey, MS= Milk Shorthorn) with the confidence interval (dot line), identity line (black) and regression line (blue).

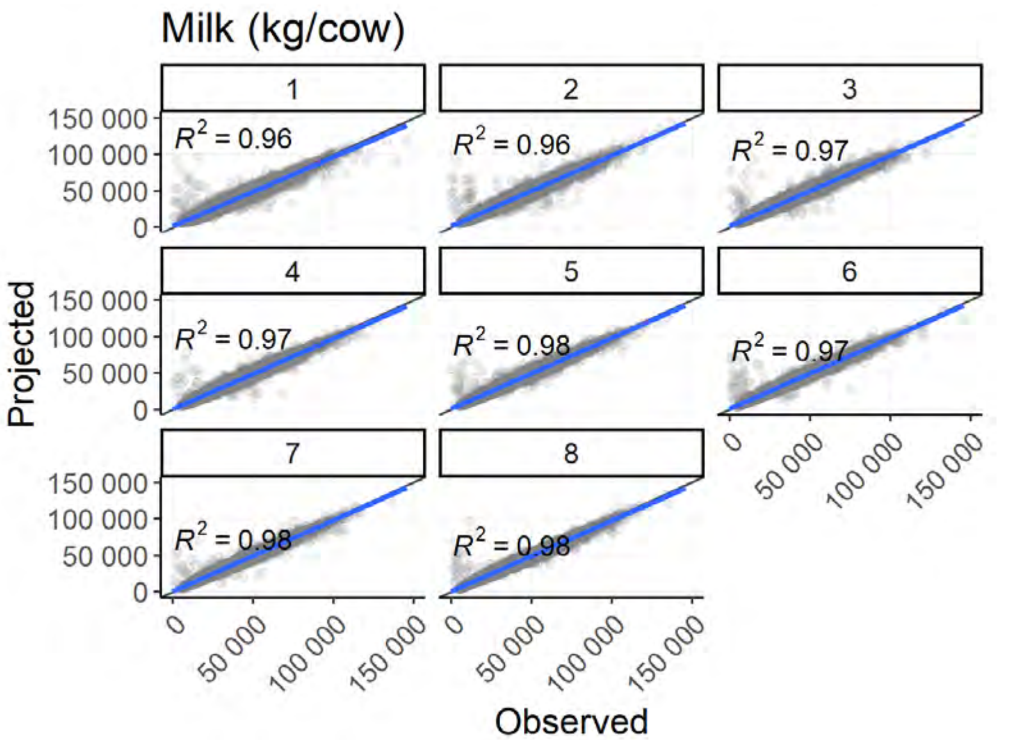


Figure 3 Results of the projected vs observed cumulative milk, fat and protein yields for different number of tests (1 to 8) with the confidence interval (dot line), identity line (black) and regression line (blue).

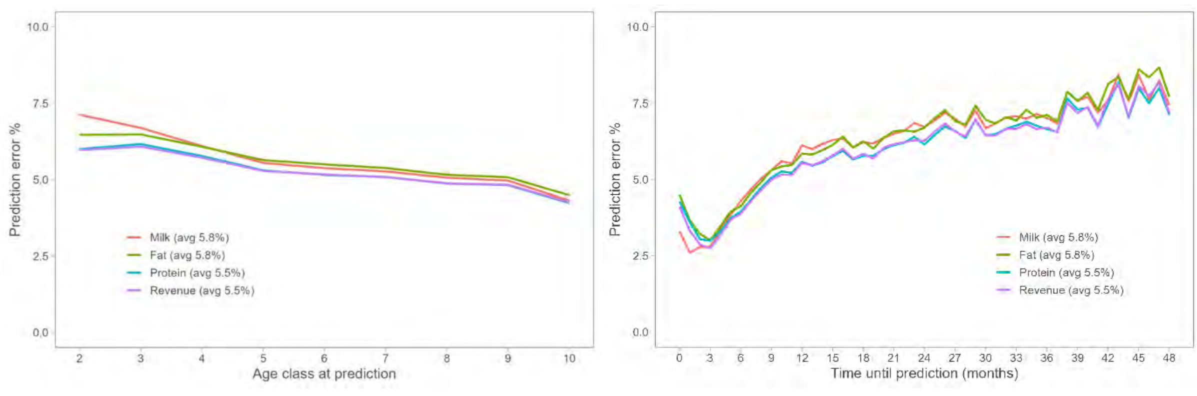
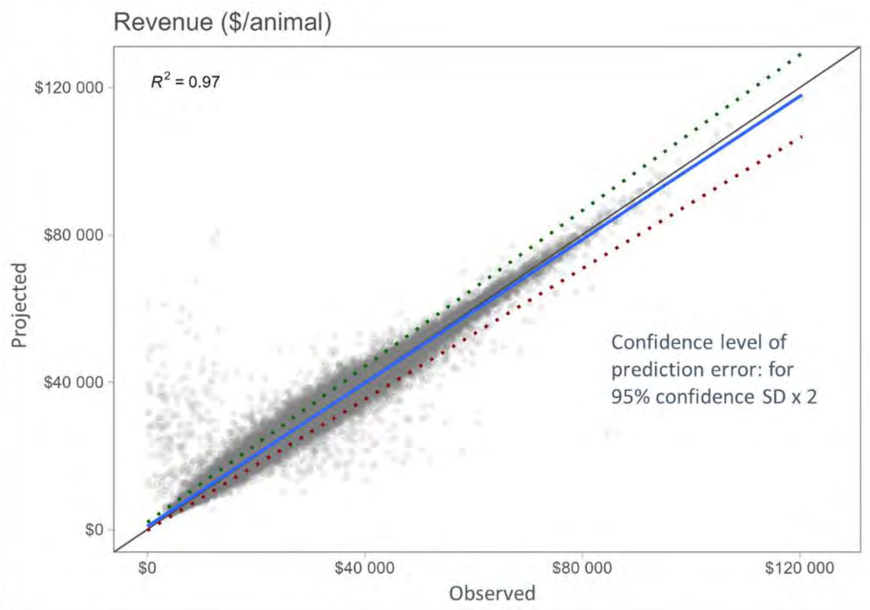


Figure 4 Prediction error based on the age at prediction (left) and time until prediction (right).



e

Figure 5. Prediction error for the projected lifetime revenue with the respective confidence level.

in fat, protein and lactose and other solids had to be considered, the respective market prices were squared and multiplied by the aforementioned (squared) SD for milk, fat and protein. The resulting variances for milk, fat and protein were summed, and finally the combined SD for income was extracted (squared root of the summed variance). To accommodate for the milk price of lactose and other solids in the income equation, the SD of milk was multiplied by 5.74, assuming that, on average, 5.74% of the milk is composed of lactose and other solids.

This procedure allowed for the estimation of the prediction error for each cow. The estimated SD can be used to provide the expected confidence level for the projected lifetime yields for each prediction by considering the time interval until prediction and the age and breed of the animal (Figure 5).

Takeaways

The model effectively predicted cumulative yields at any future age. These findings suggest that neural networks can be utilized to develop tools that assist farmers in making informed decisions by predicting future cow lifetime yields and ranking animals within a herd across all age classes.

Acknowledgements

The authors would like to acknowledge the funding support provided by the "Ministère de l'Agriculture des Pêcheries et de l'Alimentation du Québec" as part of the "Programme de développement territorial et sectoriel 2023-2026".

List of references

Cockburn, M. 2020. Review: Application and Prospective Discussion of Machine Learning for the Management of Dairy Farms. Animals 10. doi:10.3390/ani10091690.

Ferris, M.C., A. Christensen, and S.R. Wangen. 2020. Symposium review: Dairy Brain—Informing decisions on dairy farms using data analytics. Journal of Dairy Science 103:3874–3881. doi:https://doi.org/10.3168/jds.2019-17199.

Pritchard, T., M. Coffey, R. Mrode, and E. Wall. 2013. Understanding the genetics of survival in dairy cows. Journal of Dairy Science 96:3296–3309. doi:10.3168/jds.2012-6219.

Schuster, J.C., H.W. Barkema, A.D. Vries, D.F. Kelton, and K. Orsel. 2020. Invited review: Academic and applied approach to evaluating longevity in dairy cows. Journal of Dairy Science 103:11008–11024. doi:https://doi.org/10.3168/jds.2020-19043.

Vries, A.D. 2020. Symposium review: Why revisit dairy cattle productive lifespan?. Journal of Dairy Science 103:3838–3845. doi:https://doi.org/10.3168/jds.2019-17361.