An accurate machine learning approach to predict fat, protein and milk yield before and during lactation from test-day records

Hagens Marit[1], Sol Marc[1], Herrera William Garcia[1], Hidalgo Andre Marubayashi[1], Adriaens Ines*[1]

[1] CRV BV

Lactation models that accurately predict the milk, fat, and protein yield of individual cows over time are important inputs for many farm and dairy processing use cases. For example, they serve as the basis for cow-comparison key figures such as net returns and the lactation index, for herd planning and insemination decisions, and for production planning of dairy plants.

We developed a machine learning model that accurately predicts milk, fat, and protein yields of individual dairy cows up to 500 days in lactation, both at the start of the lactation and after newly incoming milk recording measurements. It uses features based on genetic information, herd and animal levels as compared to population-wide standard curves, pregnancy status, and cow auxiliary data such as parity, calving season, and age. The model is trained and tested on herd-milk recording data from 1.9 million lactations from 8,981 Dutch farms.

For daily predictions of the first 305d of lactation before the lactation has started, the model has a median absolute error (MEDAE) of 3.64, 0.13, and 0.16 kg for milk, protein, and fat yield respectively, and an average percentage error of respectively 15.3, 14.6, and 15.3%. The predictions for later lactation days are slightly poorer (MEDAE of 3.55, 0.14, 0.17 respectively), mainly because the pregnancy status is unknown, and cannot be included in the model at the moment of prediction when done before or at the start of lactation. Once the lactation has started and new milk recordings are included, the median absolute error decreases to respectively 2.45, 0.10, and 0.13 kg, corresponding with a percentage error of 11.3, 11.7, and 12.5% for milk, protein, and fat yield. The predictions at calving provide a reference to compare true performance with the cow's expected performance. The updated predictions, when new measurements come in or when pregnancy is confirmed, can be used as an input to estimate e.g. the dry-off milk yield and the cumulative (305d) milk, fat and protein yield of an animal, herd or parity group. Since the model predicts ongoing as well as future lactations, it can be used for predicting total expected yield of an entire herd or group of herds for horizons up to >1 year.