**Application**

Feeding the world’s growing population sustainably is a global challenge. The feed efficiency of dairy cattle is an important breeding goal for environmental and economic improvement of the dairy industry. Measuring feed intake is limited by the cost of specialist equipment to identify individual cattle and weigh the food consumed. Using machine learning to predict feed intake from other readily available data will enable us to produce more feed intake phenotypes and more effectively identify the most efficient dairy cattle.

**Introduction**

Producing meat and milk sustainably for a growing population can be achieved through more efficient use of feed. Feed costs also represent a large proportion of the costs on a dairy farm making feed efficiency an important breeding goal for the dairy industry (Berry *et al.*, 2014; Li *et al.*, 2021). Breeding for feed efficiency has the potential to make a permanent and cumulative, population wide improvement. Feed intake is expensive to measure requiring investment in specialist equipment which for many commercial producers is not viable. Milk sampling of dairy cattle is a non-invasive process, routinely carried out to measure milk fat and protein along with somatic cell count (an indicator for mastitis). Mid infrared (MIR) spectroscopy is a low cost method of analysing the composition of milk samples and the spectra are therefore abundantly available. MIR spectra of milk have been widely used for the prediction of expensive and difficult to measure traits such as feed intake (Denholm *et al.*, 2020).

Previous studies have indicated good potential for predicting dry matter intake (DMI) in dairy cattle from milk MIR but are limited in their dataset and computational requirements. The present study aims to (i) use an efficient ML method (XGBoost) to predict DMI from milk MIR data and (ii) assess the benefits of adding key animal traits (live weight, milk yield) in predicting DMI in dairy cattle. We used 330,336 records from 980 lactating Holstein cows, which, to our knowledge is the largest database of this kind.

**Materials and Methods**

The study was based on data from UK Holstein dairy cattle in the Langhill herd at SRUC Dairy Research Centre that has been running since 1973. Cows were milked twice a day and milk sampling was carried out once a week. All milk samples were analysed using MIR spectroscopy and processed according to standard procedures (Grelet *et al.*, 2015; Friedrichs *et al.*, 2015). Feed was offered ad libitum and refusal was measured using individual feed bins (HOKO-system, Insentec B.V.). Feed intake was measured for housed cattle for a period of 130 days on average for each lactation. The feed was sampled, and dry matter content of feed was analysed regularly to obtain daily DMI per cow.

After the removal of spurious records from the data and aligning the feed intake and MIR spectra, the final dataset consisted of 330,336 records. Training and testing datasets were created by splitting the whole dataset by a ratio of 9:1 from random sampling which was repeated for 10 fold cross validation. The machine learning method of XGBoost was used to train the prediction model to predict DMI from MIR data and predictor traits using the training dataset. The best prediction model from training was then used to predict DMI in the testing dataset, where the prediction accuracy for DMI was measured by RMSE. The initial predictor variables for DMI were the 1060 wavepoints of the MIR spectra and this was considered the base model. Further predictor variables added to MIR data included feed group (high- and low-energy forage), genetic group (control-line and select-line), milk yield, and liveweight. The machine learning was run on an NVIDIA DGX computer running a Linux operating system

**Results**

DMI was predicted from milk MIR spectra with a high pearson correlation coefficient of 0.79 as seen in Table 1. The addition of other predictor variables into the model improved the correlation further to 0.82-0.88.

**Table 1.**  Model performance for predicting dry matter intake (DMI) using different predictor traits (r=pearson correlation coefficient of predicted and actual DMI, SE=standard error, RMSE=root mean squared error)

|  |  |  |  |
| --- | --- | --- | --- |
| **Model variables** | **r** | **SE** | **RMSE** |
| MIR (only) | 0.79 | 0.0036 | 3.60 |
| MIR + milk yield | 0.82 | 0.0037 | 3.32 |
| MIR + live weight | 0.87 | 0.0027 | 2.91 |
| MIR + milk yield +live weight | 0.88 | 0.0027 | 2.77 |

**Conclusions**

We have been able to predict dairy DMI from milk MIR spectra with high prediction accuracy. Including extra predictor traits of dairy cows (feed groups, milk yield, liveweight) improved the accuracy of prediction.

**References**

Berry, D. P., Coffey, M. P., Pryce, J. E., de Haas, Y., Løvendahl, P., Krattenmacher, N., Crowley, J. J., Wang, Z., Spurlock, D., Weigel, K., Macdonald, K. & Veerkamp, R. F. (2014). *Journal of Dairy Science* **97**, 3894–3905.

Denholm, S. J., Brand, W., Mitchell, A. P., Wells, A. T., Krzyzelewski, T., Smith, S. L., Wall, E. & Coffey, M. P. (2020). *Journal of Dairy Science* **103**, 9355–9367.

Friedrichs, P., Bastin, C., Dehareng, F., Wickham, B. & Massart, X. (2015). *Biotechnol. Agron. Soc. Environ.*

Grelet, C., Pierna, J. A. F., Dardenne, P., Baeten, V. & Dehareng, F. (2015). *Journal of Dairy Science* **98**, 2150–2160.

Li, B., Mrode, R., Id-Lahoucine, S., Winters, M. & Coffey, M. (2021). *Interbull Bulletin* 125–130.