

Novel approach to develop robust artificial neural network for FT-MIR prediction of enteric methane emissions in dairy cows

Soyeurt Hélène^[1], Dehareng Frédéric^[2], Gengler Nicolas^[1], McParland Sinead^[3], Kreuzer Michael^[4], Lund Peter^[5], Martin Cécile^[6], Kuhla Björn^[7], Vanlierde Amélie^[2]

[1] ULiège - GxABT, Gembloux, Belgium, [2] CRA-W, Gembloux, Belgium, [3] Teagasc, Moorepark, Ireland, [4] ETH Zurich, Zurich, Switzerland, [5] Aarhus University, Aarhus, Denmark, [6] Université Clermont Auvergne - INRAE, Saint-Genès-Champanelle, France, [7] Research Institute for Farm Animal Biology, Dummerstorf, Germany

Previous studies have demonstrated the potential of milk Fourier-transform mid-infrared (FT-MIR) spectroscopy to predict indicators related to enteric CH₄ emissions in dairy cows. Early models mainly relied on PLS regression, a method well suited to smaller datasets with highly collinear variables. However, sufficient accuracy of CH₄ reference measurements to reliably capture the underlying population trend requires many observations. The acquisition of such data is limited by financial, practical, and confidentiality constraints. Transfer learning strategies, particularly federated learning (FL), offer a promising solution. FL enables a neural network trained by one research group to be transferred to another without sharing raw data. The second group can further train the model on its own data and share the updated network, preserving data privacy while improving predictive performance. The development of a robust artificial neural network is therefore essential.

The objective of this study was to compare PLS regression with multilayer perceptrons (MLP) and convolutional artificial neural networks (cANN). A dataset of 1,089 records was used, with a subset reserved for external validation. A total of 289 spectral variables were selected and preprocessed using a first derivative, followed by regression on second-order Legendre polynomials according to days in milk. Three MLP architectures (one to three hidden layers) and three cANN models were evaluated, the latter using spectral data reshaped into 3 × 289 images.

Cross-validation performances were consistent with expectations (R^2 between 0.60 and 0.80; mean error around 51 g CH₄/day). The MLP with two hidden layers achieved the best results. Neural networks slightly outperformed PLS and, more importantly, open new perspectives for implementing FL frameworks to improve CH₄ prediction models without sharing sensitive data.