

The use of supervised machine learning techniques to identify factors influencing vitamin D bio-enrichment of pork

E.J. Rosbotham¹, D. Rankin², C.I.R. Gill¹, E.J. McDonald³, W.C. McRoberts⁴, H.R. Neill¹, R. Boland¹ and L.K. Pourshahidi¹

¹Nutrition Innovation Centre for Food and Health (NICHE), School of Biomedical Sciences, Ulster University, Coleraine, UK,

²School of Computing, Engineering and Intelligent Systems, Ulster University, Derry-Londonderry, UK,

³Devenish Nutrition Ltd, Belfast, UK and

⁴Agri-Food and Biosciences Institute, Belfast, UK

Vitamin D bio-enriched foods can complement traditional fortification measures to improve sub optimal vitamin D intakes across the population⁽¹⁾. Recent attention has been given to bio-enrichment practices such as fortified livestock feed and/or animal UVB exposure to naturally enhance vitamin D content in meat⁽²⁾. This study aimed to use supervised machine learning techniques to identify features influencing vitamin D response in bio-enriched pork.

In this study, swine (n = 20) were exposed daily to narrowband UVB radiation (irradiance 440 μ W/cm²) (n = 10) or standard lighting (n = 10) for a period of 11 weeks prior to slaughter. Vitamin D₃ and 25-hydroxyvitamin D₃ (25(OH)D₃) concentrations were analysed in cooked pork loin by liquid chromatography tandem mass spectrometry. A dataset capturing the features and outcomes of the on-farm study was collected. Rapidminer Studio (version 9.8) was used to identify the most important features influencing a higher vitamin D response in bio-enriched pork. Feature selection utilising information gain indicated eight features of significance for predicting vitamin D response to the bio-enrichment process: body weight gain (%), body weight at slaughter (kg), behaviour, sex, serum 25(OH)D₃ (μ g/l) at baseline, midpoint and endpoint, and cooking loss (%). These features were used to train five classification models: random forest, Naïve Bayes, decision trees (C4.5 and CART) and KNN. A train/test split of 60/40 was applied with 5-fold cross-validation.

The target variable was vitamin D activity, calculated as [vitamin D₃ + (25(OH)D₃ x 5)]⁽³⁾ in bio-enriched pork, categorised as “high” (>19 μ g/kg) or “low” (\leq 19 μ g/kg). Models obtained F1-scores ranging 57.1–77.1%. KNN and C4.5 performed best (77.1% and 75%). Serum 25(OH)D₃ levels at the midpoint and endpoint of the study, body weight gain and body weight at slaughter were commonly identified by Naïve Bayes, decision trees (C4.5 and CART) and KNN as the four most important factors influencing vitamin D activity in bio-enriched pork. Feature importance results determined by random forest differed, identifying serum 25(OH)D₃ taken at the baseline and midpoint of the study, body weight gain and body weight at slaughter as the four most important features. Results from all five classification models determined sex to be the least important feature influencing the bio-enrichment process.

Supervised machine learning techniques were able to identify the most important features influencing vitamin D response in bio-enriched pork. Preliminary results from the current analysis will be built upon through future studies to enable more robust model development. This will in turn inform bio-enrichment process optimisation for industry to produce novel bio-enriched meat products for consumers.

Acknowledgements

This study was funded by Devenish Nutrition Limited and conducted in collaboration with Agri-Food and Bioscience Institute (AFBI). The first author is in receipt of a Department of Agricultural and Rural Affairs (DAERA) PhD studentship.

References

1. Cashman, K (2020) *Calcif Tissue Int* **106**, 1, 14–29
2. Duffy K, *et al.* (2018) *CAB Rev* **13**, 045, 1–11
3. Cashman, K (2012) *Food Nutr Res*, **56**, 1–8