**Development and validation of a fully automated 2D imaging system generating body condition scores for dairy cows using machine learning**

Nektarios Siachos1, Mark Lennox2, Alkiviadis Anagnostopoulos1, Bethany E. Griffiths1, Joseph M. Neary1, Robert F. Smith1, Georgios Oikonomou1.

1Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool

2CattleEye Ltd, The Innovation Centre, Queens Road, Belfast

**Application**

Autonomous livestock monitoring applications offer opportunities for consistent and objective assessment of various health and welfare indicators, labour-demanding and time-consuming tasks when dependant on humans.

**Introduction**

Monitoring body condition of dairy cows at specific stages of their lactation is a valuable aid to improve feed efficiency, fertility, and milk production and to minimize metabolic disorders. Despite that being emphasized over the last decades, most farms rarely conduct regular body condition scoring. CattleEye Ltd (Belfast, UK) has recently developed a fully automated system for lameness detection using a 2D camera mounted over an exit race (Anagnostopoulos et al., 2023). Our aim was to develop and evaluate the performance of a machine learning algorithm that would generate real-time body condition scores (BCS) using the same equipment.

**Materials and methods**

The study was conducted in 7 dairy farms located in west England and Wales which were equipped with a camera. Herd size ranged from 600 to 2,300 lactating Holstein cows. Manual BCS scoring (MAN\_BCS) data for training of the algorithm was collected by 5 human assessors (HAs) who performed a total of 34 whole-milking-herd BCS sessions. A total of 34,150 MAN\_BCS with correct cow identification were recorded using the 1-5-point scale method with 0.25 increments (Ferguson et al., 1994). The algorithm was developed using the training dataset comprising various farms, which allowed us to build robust image datasets for predicting a cow's BCS. During one milking at the same days the HAs visited the farms, top-down videos of the cows exiting the milking parlour were gathered and processed by a pre-trained cow detection and pose estimation pipeline to track and identify the key points on each animal. An ordinal regression model was used when developing the algorithm. A state-of-the-art deep learning architecture, EfficientNetV2, was used to form the algorithm’s backbone. Subsequently, data for testing was collected in 4 of the participating farms at least 2 months after the last training session to allow conditions to change. The MAN\_BCS collected by one HA in the milking parlour, who performed two whole-herd BCS sessions on each farm 30 days apart, were used as the ground truth. At the end of the study, MAN\_BCS were merged with the automated BCS (CE\_BCS) using the cow ID, resulting in 9,657 single BCS. A total of 3,817 cows were scored twice 30 days apart and the change in their BCS (ΔBCS) was calculated. To assess the intra-observer reliability, the HA visited the same farm 24 h after a whole-herd BCS scoring session to score a subset of the same cows. Automated BCS (CE\_BCS) recordings from the same days were also stored to assess the precision of the automated system at the same farm. The intra- and inter-rater agreement between MAN\_BCS and CE\_BCS for the single BCS and the ΔBCS was estimated by calculating the weighted kappa (κw) for the exact score agreement and the percentage agreement (PA) at the 0.00, ±0.25 and ±0.50 -unit of BCS error range. Moreover, a Bland-Altman plot was created to examine for any systematic or proportional bias of system. The MAN\_BCS of 3 HAs were also used to evaluate the inter-rater agreement between humans. Finally, we measured the backfat thickness (BFT) of 111 randomly selected cows using ultrasonography. Passing-Bablock regressions were used to assess the relationship of MAN\_BCS and CE\_BCS with BFT.

**Results**

The system had an almost perfect repeatability with a κw = 0.99 with 96.2% PA at the exact score agreement, while the HA’s intra-observer agreement produced a κw = 0.94 with 68.1% PA at the exact score agreement. The κw for the exact agreement between single MAN\_BCS and CE\_BCS was 0.69, indicative of substantial agreement, while PA at the 0.00, ±0.25 and ±0.50 -unit error was 44.4, 84.6 and 94.8 %, respectively. A minimal systematic bias was observed (-0.09-unit) with a proportional bias at the extreme scores. Regarding ΔBCS, the κw for the exact agreement between MAN\_BCS and CE\_BCS was 0.20, while PA at the 0.00 and ±0.25 error was 45.7 and 88.2%, respectively. The κw between HAs was 0.77-0.82 and the PA at the 0.00, ±0.25 and ±0.50 -unit error was 29.7-53.4, 75.3-95.8 and 91.7-99.4%, respectively. A strong linear relationship of both MAN\_BCS and CE\_BCS with the BFT measurements was observed, producing Spearman’s rank correlation coefficients of *ρ* = 0.91 and 0.75 (P< 0.001), respectively.

**Conclusions**

Based on our results, automatically generated BCS by the system can predict single BCS and changes in BCS with sufficient accuracy, similar to that observed between trained scorers. Collecting more training data at the extremes could improve the system’s accuracy at the low scores, where it was found to be lower.

**Acknowledgements**

This study was funded by Innovate UK (Farming Innovation Programme – small R&D partnership projects. Project number: 10027372).

**References**

Anagnostopoulos, A., Griffiths, B.E., Siachos, N., Neary, J., Smith, R.F., & Oikonomou, G. (2023). Frontiers in Veterinary Science, 1111057.

Ferguson, J.D., Galligan, D.T., & Thomsen, N. (1994). Journal of Dairy Science, 77, 2695–2703.