De Castro, A., Chebel, R., Yu, H. (2026) Utility of multi-modal integration for predicting postpartum diseases in dairy cattle with imbalanced data, *Proceedings 13th World Congress on Genetics Applied to Livestock Production*
23630: https://doi.org/10.31274/wcgalp.23630 ©2026 De Castro, et al. CC BY-NC-ND

Utility of multi-modal integration for predicting postpartum diseases in dairy cattle with imbalanced data

A.L. De Castro[1*], R.C. Chebel[2], and H. Yu[1]

[1]Department of Animal Sciences, University of Florida, Gainesville, USA; [2]Department of Large Animal Clinical Sciences, University of Florida, Gainesville, USA; [*]Corresponding author: decastro.a@ufl.edu

Introduction

Postpartum diseases can negatively affect health and performance in dairy cattle and consequently cause substantial economic losses to the dairy industry (Liang *et al*., 2017; Rasmussen *et al*., 2024). Early detection of these diseases is important for timely intervention, yet current diagnostic approaches rely primarily on subjective visual observation and clinical evaluation, which are labor-intensive and often delay diagnosis. Advances in wearable sensor technology now allow continuous and objective monitoring of individual animals in real time. Time-series rumination and activity data derived from wearable sensors have been widely used for disease detection in dairy cattle (Merenda *et al*., 2021; Prim *et al*., 2024). However, there are limited studies that combine time-series sensor data with single time-point on-farm phenotypic data, which may provide additional information for postpartum disease prediction. This is mainly due to the lack of appropriate modeling approaches capable of integrating time-series sensor data and single time-point phenotypes. In addition, disease cases are typically much fewer than healthy controls, creating imbalanced datasets that limit model prediction accuracy. The objectives of this study were to 1) develop a multi-modal deep learning model that integrates time-series behavioral sensor data with single time-point management phenotypes for postpartum disease prediction, and 2) evaluate the effects of representation learning and class-imbalance handling strategies on disease prediction performance.

Materials & Methods

Data and preprocessing. Data were collected from 9,540 Holstein cows across three commercial dairy farms located in Georgia and Florida during the peripartum period. For each cow, data were recorded from 21 days before calving to 2 days after calving, resulting in a 24-day observation window. The target variable of postpartum diseases was defined as the occurrence of any uterine disease, metabolic disorder, digestive disorder, or other health condition within 30 days after calving. Overall disease incidence was 27.8%, including 2,652 diseased cows and 6,888 healthy cows. The predictor variables included time-series behavioral rumination and activity data derived from wearable sensors, single time-point on-farm management phenotypes, and additional time-series covariates related to environmental conditions and lactation stage. Environmental conditions were represented by the temperature-humidity index (THI), and the lactation stage was represented by days in milk (DIM). The single time-point phenotypes consisted of parity, calving problems, cohort calving number, calving ease score, and calf gender. To prepare the data for model development, preprocessing was performed by removing cows with missing single time-point management phenotypes and excluding time-series records with missing behavioral sensor values. In addition, cows with fewer than 24 days of observations were retained by assigning zero values to missing days to ensure a consistent observation window for model development.

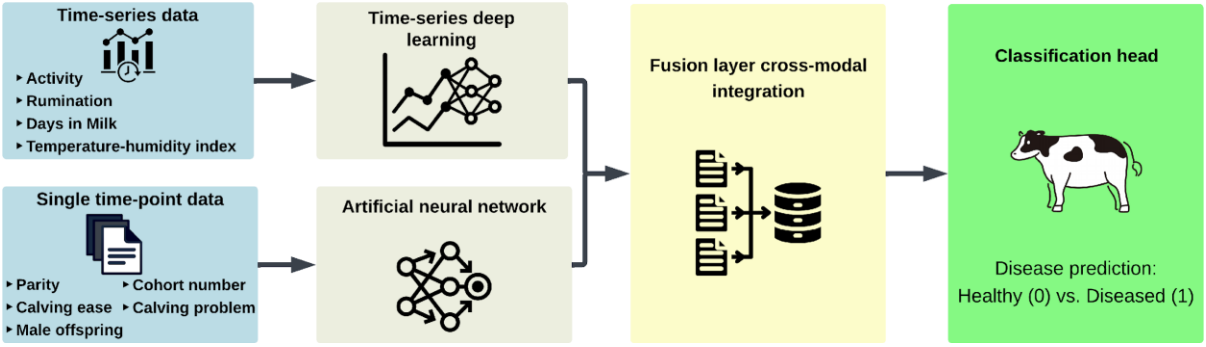

Figure 1. Overview of the multi-modal DL architecture for postpartum disease prediction.

Deep learning model architecture. The proposed multi-modal deep learning (DL) model is shown in Fig. 1 and consists of a time-series feature encoder, a single time-point feature encoder, a fusion module, and a classification head. The time-series feature encoder processes 24-day behavioral, THI, and DIM data to produce a 256-dimensional representation, and incomplete sequences with missing records are handled using mask-aware pooling. Three time-series encoder architectures were compared, including long short-term memory networks (LSTM), temporal convolutional networks (TCN), and transformer encoders, using standard baseline configurations. The LSTM consisted of two stacked layers with a dropout rate of 0.2, the TCN consisted of five convolutional blocks with dilation rates of 1, 2, 4, 8, and 16, along with residual connections and a dropout rate of 0.2, and the transformer encoder consisted of two layers with multi-head self-attention, positional encoding, layer normalization, and a dropout rate of 0.2. The single time-point feature encoder used a multi-layer perceptron to process management phenotypes and produce a 256-dimensional representation. The fusion module combined the time-series and single time-point representations using learned attention weights to assign relative importance to each modality for each cow. The classification head used a fully connected layer with a sigmoid activation to output a disease probability.

Representation learning and class-imbalance handling. Using the proposed multi-modal DL model architecture described above as the baseline model, we investigated the impact of representation learning and class imbalance handling strategies on disease prediction performance. Specifically, contrastive multi-modal (CMM) representation learning and three class imbalance handling strategies, including focal loss, positive weighting, and weighted sampling, were evaluated. CMM is a self-supervised pretraining approach that aligns each cow's time-series and single time-point representations without using disease labels by aligning representations from the same cow and separating those from different cows. CMM pretraining was performed for up to 40 epochs with a batch size of 128, with early stopping based on validation loss. The pretrained encoder weights were then used to initialize supervised model training. Three class imbalance handling strategies were evaluated to reduce prediction bias toward the healthy majority class. Focal loss increased the emphasis on misclassified cases using a focusing parameter of 2. Positive weighting applied a class ratio weight such that each diseased cow contributed 2.6 times more than a healthy cow to the loss function. Weighted sampling adjusted mini-batch composition so that diseased cows appeared more frequently based on their class frequency. In total, five training configurations were compared, including the baseline model, the baseline model with CMM, and the baseline model with CMM combined with each of the three class imbalance handling strategies.

Model evaluation. Model prediction performance was measured using two metrics. The area under the precision–recall curve (AUPRC) measures the trade-off between precision and recall and focuses on the minority disease class. The area under the receiver operating characteristic

curve (AUROC) measures how well the model distinguishes diseased from healthy cows. AUPRC was used as the primary evaluation metric because AUROC could overestimate model performance in imbalanced datasets, where the false positive rate is dominated by true negatives from the majority healthy class. Two cross-validation (CV) strategies were used to evaluate model performance. In the hold-out CV, cows from all three farms were combined and randomly split, with 70% assigned to training and validation and 30% assigned to testing. In addition, a cross-farm CV strategy was implemented, in which models were trained on data from the two smaller farms collected during earlier time periods and evaluated on data from the larger farm.

Results & Discussion

The results of five training configurations exploring the effects of representation learning and class imbalance handling are shown in Table 1 based on the transformer encoder. Under hold-out CV, the baseline model achieved an AUPRC of 0.69 and an AUROC of 0.80, while incorporating CMM increased performance to an AUPRC of 0.72 and an AUROC of 0.81. Among the three class imbalance handling strategies, focal loss further improved performance and achieved the highest AUPRC of 0.73 and AUROC of 0.83, whereas positive weighting and weighted sampling did not consistently improve performance beyond the baseline model with CMM. The improvement associated with CMM suggests that aligning time-series behavioral data with single time-point management phenotypes enables the model to learn more informative cow-level representations for postpartum disease prediction under class imbalance by capturing complementary patterns during pretraining. The effectiveness of class imbalance handling depended on the specific strategy, with focal loss showing clear advantages when combined with CMM, likely because it emphasizes misclassified disease cases while reducing the influence of well-classified healthy cows. In contrast, positive weighting and weighted sampling adjust class contributions at a more global level, which may be less effective for capturing subtle disease-related signals in heterogeneous, multi-modal farm data. Prediction performance under cross-farm CV was consistently lower than that under hold-out CV, reflecting a more generalized and challenging evaluation scenario in which models trained on smaller farms and earlier time periods were applied to a different farm with distinct management and environmental conditions (Wang *et al*., 2025).

Table 1. Prediction performance of five training configurations using the transformer encoders under two CV strategies.

	CV strategy			
	Hold-out CV (70-30)		Cross-farm CV	
Training configuration	AUPRC	AUROC	AUPRC	AUROC
Baseline	0.69	0.80	0.52	0.70
Baseline + CMM	0.72	0.81	0.59	0.75
Baseline + CMM + Focal loss	0.73	0.83	0.65	0.78
Baseline + CMM + Positive weighting	0.70	0.82	0.59	0.75
Baseline + CMM + Weighted sampling	0.71	0.81	0.56	0.75

Consistently, the best prediction performance for both the LSTM and TCN encoders was also achieved when using the baseline model combined with CMM and focal loss. A comparison of the three time-series encoders under this best-performing configuration is shown in Fig. 2. Under hold-out CV, the transformer encoder achieved the highest AUPRC and AUROC of 0.73 and 0.83, respectively, compared with 0.72 and 0.82 for LSTM and 0.71 and 0.81 for TCN. Under cross-farm CV, a similar pattern was observed, with the transformer encoder again outperforming LSTM and TCN, while the latter two showed comparable performance. Overall,

the transformer encoder consistently achieved higher prediction performance than LSTM and TCN across both CV strategies. This is likely due to the self-attention mechanism, which can capture long-range dependencies across the 24-day behavioral sequences more effectively than sequential processing in LSTM or the fixed local receptive fields used in TCN.

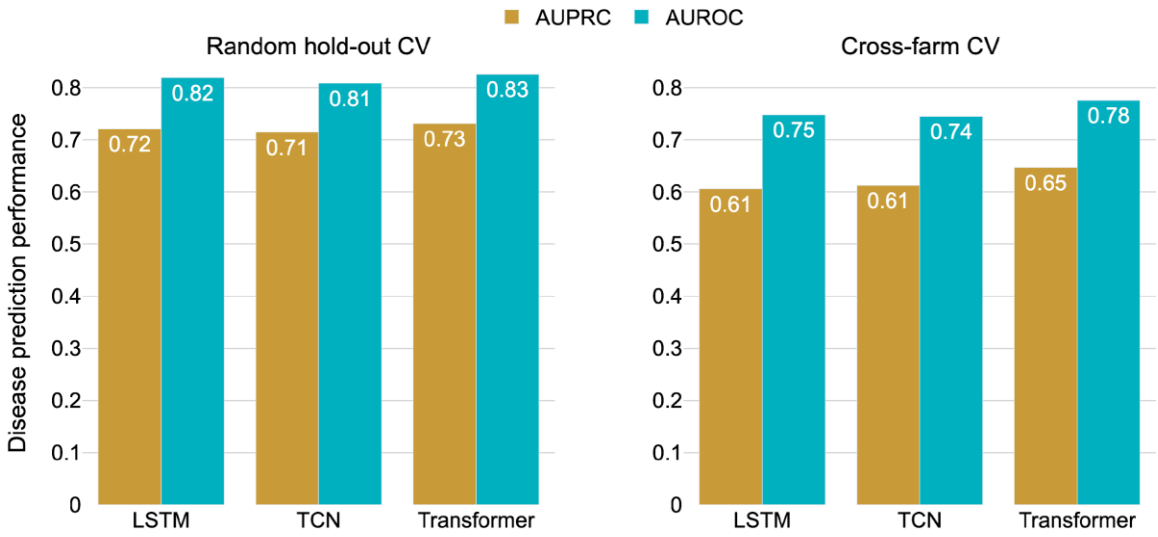

Figure 2. Prediction performance comparison of three time-series encoder models under two CV strategies.

Conclusions

In this study, we developed a multi-modal deep learning model that integrates time-series behavioral sensor data with single time-point management phenotypes for postpartum disease prediction. We showed that combining representation learning with class-imbalance handling strategies can improve prediction performance for minority disease cases. Using both random hold-out CV and cross-farm CV, we evaluated the generalization of the proposed model. Collectively, these results suggest that the proposed approach can support earlier detection of postpartum diseases and has the potential for practical implementation in on-farm dairy systems to support timely health management.

References

Liang D., Arnold L.M., Stowe C.J., *et al*. (2017) J. Dairy Sci. 100(2):1472-1486. doi.org/10.3168/jds.2016-11565

Merenda V.R., Ruiz-Munoz J., Zare A., *et al*. (2021) Prev. Vet. Med. 194:105431. doi.org/10.1016/j.prevetmed.2021.105431

Prim J.G., Casaro S., Mirzaei A., *et al*. (2024) J. Dairy Sci. 107(7):4881-4894. doi.org/10.3168/jds.2023-23611

Rasmussen P., Barkema H.W., Osei P.P., *et al*. (2024) J. Dairy Sci. 107(9):6945-6970. doi.org/10.3168/jds.2023-24626

Wang J., Yu Z., Chebel R.C., and Yu H. (2025) Smart Agr. Tech. 12:101483. doi.org/10.1016/j.atech.2025.101483