—Technology Report—

An attempt at estrus detection in cattle by continuous measurements of ventral tail base surface temperature with supervised machine learning

Shogo HIGAKI[1], Hongyu DARHAN[1], Chie SUZUKI[1], Tomoko SUDA[1], Reina SAKURAI[1] and Koji YOSHIOKA[1]

[1]*National Institute of Animal Health, National Agriculture and Food Research Organization, Ibaraki 305-0856, Japan*

Abstract. We aimed to determine the effectiveness of estrus detection based on continuous measurements of the ventral tail base surface temperature (ST) with supervised machine learning in cattle. ST data were obtained through 51 estrus cycles on 11 female cattle (six Holsteins and five Japanese Blacks) using the tail-attached sensor. Three estrus detection models were constructed with the training data (n = 17) using machine learning techniques (random forest, artificial neural network, and support vector machine) based on 13 features extracted from sensing data (indicative of estrus-associated ST changes). Estrus detection abilities of the three models on test data (n = 34) were not statistically different among models in terms of sensitivity and precision (range 50.0% to 58.8% and 60.6% to 73.1%, respectively). The relatively poor performance of the models might indicate the difficulty of separating estrus-associated ST changes from estrus-independent fluctuations in ST.
Key words: Body surface temperature, Estrus detection, Supervised machine learning, Wearable sensor

In cattle, body core (typically vaginal) temperature decreases transiently a few days before estrus followed by a sharp increase at estrus [1, 2]. The proestrus decrease in temperature is suggested to reflect corpus luteum regression [3] and the estrual rise in temperature is closely associated with the preovulatory luteinizing hormone surge [4]. Several approaches to detect the temperature changes during estrus have been studied using wearable sensors applied to different body parts; such as the rumen [5], vagina [2, 6], and ear canal [7]. Monitoring the body surface temperature (ST) at the ventral tail base was offered as a less invasive and easy-to-apply (the fitting of the sensor device requires only a few minutes without strict hygiene management and restriction of body movement) approach for monitoring body temperature [8, 9], and estrus could be detected as high as sensitivity (true-positive/[true-positive + false-negative]) of 78% and precision (true-positive/[true-positive + false-positive]) of 70% by setting appropriate cut-offs [8]. Supervised machine learning techniques have been proposed as alternatives to simple thresholding methods and seem to be good candidates for handling wearable sensing data for estrus detection [2, 10]. In the use of a vaginal sensor, we previously developed an efficient estrus detection model (both sensitivity and precision were 94%) showing the superiority to combine two parameters (vaginal temperature and conductivity) [2]. However, these algorithms have never been applied to ST data for bovine estrus detection. In the present study, we attempted to develop estrus detection models with a single parameter ST by using appropriate features, hyperparameters, and machine learning algorithms used for modeling.

First, to confirm the feasibility of detecting estrus-associated ST changes by the tail-attached sensor, we measured ST throughout 17 estrous cycles. The standing estrous periods of these cycles were confirmed by testing with herd-mates. To exclude the influence of the circadian rhythm of ST, the actual raw ST values were calculated as residual ST (rST = hourly maximum ST value − mean ST value for the same hour on the previous 3 days) and analyzed. The mean rST started to increase from approximately 12 h before the onset of estrus and the values during −1 to +15 h from the onset of estrus were higher than that of the control period (−192 h to −121 h) (P < 0.05; Fig. 1). Similar to the previous study [8], rST showed the highest value at 6 h after the beginning of estrus. However, the transient decrease a few days before estrus observed in the previous study was absent. This discordant result might be explained by the different sensors used as described below and/or by individual cow difference in ST changes, because it was reported that 10% to 16% of the cycling cows did not show the proestrus decrease in body temperature [6, 11]. Nevertheless, substantial rST changes around estrus were detected successfully by the present tail-attached sensor.

Because a given numerical value for rST seemed not to be necessarily indicative of estrus without monitoring its sequential changes, we next extracted the features listed in Table 1 to follow up the changes in rST around estrus, and to build the estrus detection models through machine learning techniques. As examples of features, the differences between the current smoothened rST and the minimum and maximum values during the last 12 h of smoothened rST should both be positive values during 12 h before the onset of estrus, since rST showed a continuously increasing trend during this period. Using these features, binary classification models— namely, classification into "in estrus" or "not in estrus" for a particular time point— were developed based on the 17 estrous period-labeled datasets ("training

datasets"). Following development of the models, 34 unlabeled datasets ("test datasets") were applied to the models, and these were used in attempts to detect the day of estrus. We tested the three widely used supervised machine learning algorithms with wearable sensing data: random forest (RF), artificial neural network (ANN), and support vector machine (SVM) [12], because of the lack of *a priori* distinctions between these algorithms [13].

Table 2 shows the performance of the three estrus detection models. The ranges of sensitivity and precision of the models (50.0 to 58.8% and 60.6 to 73.1%, respectively) were not statistically different among models and these rates were similar to or lower than those of the previous study using a simple thresholding method on rST (78 and 70%, respectively) [8]. The lack of the proestrus decrease in rST observed in the previous study [8] was at least partially responsible for the relatively low estrus detection abilities of the present estrus detection models, because this depression, thought to be related to corpus luteum regression [3], should be useful in confirming that a subsequent rST increase is a true positive [8].

Figure 2 shows the representative cases of animals correctly detected and falsely undetected in estrus. Several false positives in the present study occurred a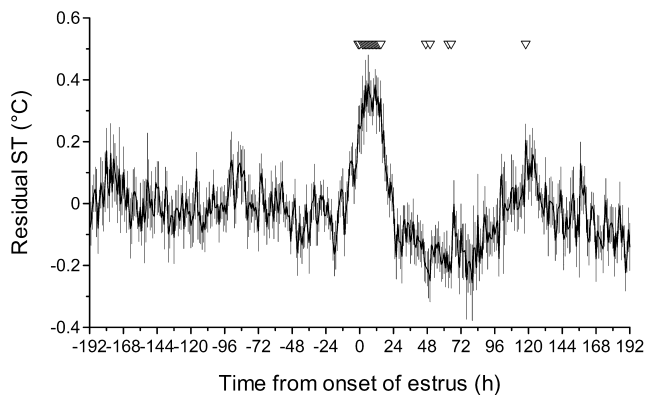ssociated with sudden rST elevations such as caused by fever (Fig. 2B). Supervised machine learning is the search for algorithms that reason from externally supplied instances (training data) to produce general hypotheses (models), which then make predictions about unknown instances (test data) [14]. Therefore, the relatively poor performance of our models might indicate the difficulty of separating estrus-associated rST changes from estrus-independent fluctuations in rST; in other words, estrus detection based only on ST data might be difficult by the present method. Further studies may be needed for improving estrus detection models by minimizing the influence of estrus-independent fluctuations in rST and/or adding much more training data that covers more possible situations in terms of season, cattle breed, rearing condition, etc. Otherwise, by adding other parameters that could be used to detect estrus, such as activity, and recumbent time and bouts [10]. Because, theoretically, having more features should result in more discriminative power, when the features highly correlated with the target class (*i.e.*, estrus) and uncorrelated with each other [15].

Fig. 1. Residual tail surface temperature (rST) changes around estrus. Inverted-triangles indicate the periods with differences between rST at the indicated time point and the mean rST during the control period (from 192 to 121 h before the beginning of estrus). Data were standardized to the onset of standing estrus (0 h). Because of variation in the length of estrous cycles, the number of animals included in each time point varied between 6 and 17. Values are presented as the mean (bold line) ± standard error (vertical bar).

Table 1. Description of the features used for building estrus detection models

Feature
Current smoothened residual-tail surface temperature (rST) [*]
Minimum value during the last 12 h of smoothened rST (12 h min.)
Minimum value during the last 24 h of smoothened rST (24 h min.)
Minimum value during the last 48 h of smoothened rST (48 h min.)
Maximum value during the last 12 h of smoothened rST (12 h max.)
Maximum value during the last 24 h of smoothened rST (24 h max.)
Maximum value during the last 48 h of smoothened rST (48 h max.)
Difference between the current smoothened rST and 12 h min.
Difference between the current smoothened rST and 24 h min.
Difference between the current smoothened rST and 48 h min.
Difference between the current smoothened rST and 12 h max.
Difference between the current smoothened rST and 24 h max.
Difference between the current smoothened rST and 48 h max.

[*] Residual tail surface temperature (rST) was calculated as actual ST − mean ST for the same hour on the previous 3 days.

Table 2. Performance of the estrus detection models developed by three machine learning algorithms (random forest, RF; artificial neural network, ANN; and support vector machine, SVM) on 34 estrous cycles

Machine learning algorithm	True positive	False positive	False negative	Sensitivity (%)	Precision (%)
RF	20	13	14	58.8	60.6
ANN	19	7	15	55.9	73.1
SVM	17	9	17	50.0	65.4

Sensitivity and precision were calculated as true-positive/(true-positive + false-negative) and true-positive/(true-positive + false-positive), respectively. Sensitivities and precisions of the three estrus detection models were not statistically different (Fisher's exact test and generalized score statistic, respectively).
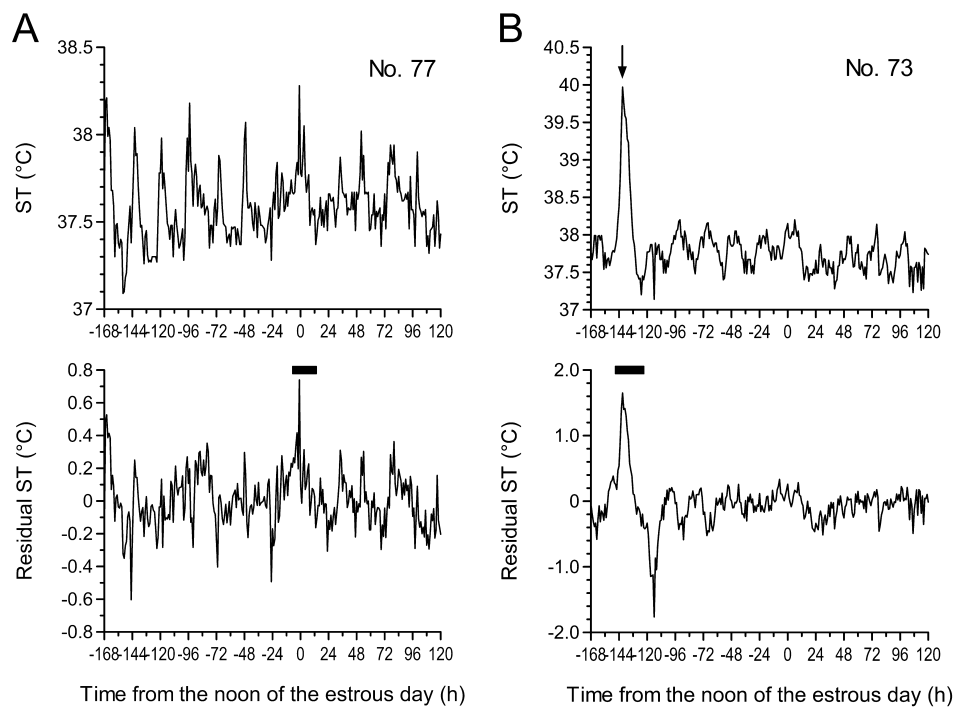
Fig. 2. Representative changes in tail surface temperature (ST) and residual ST (rST) around estrus. A: A representative case of an animal correctly detected in estrus without false positive. B: A representative case of an animal falsely undetected in estrus with one false positive. Arrow indicates a transient ST increase during fever. Thick horizontal bars indicate the periods of estrus alert produced by the machine learning estrus detection model which was developed using the random forest algorithm. Data were standardized to the noon of the estrous day (0 h).

Methods

Animals

This study was conducted at the National Institute of Animal Health, National Agriculture and Food Research Organization in Japan from March 2017 to November 2018. A total of 51 estrous cycles of 11 mature non-pregnant and non-lactating female cattle (six Holsteins and five Japanese Blacks; 15–109 months old, 370–630 kg body weight) were used. The cattle were housed in tie-stalls or individual pens with access to an outdoor paddock during daytime, and were fed twice daily with hay and concentrate to meet the Japanese Feeding Standard recommendations, with *ad libitum* access to water. All procedures were approved by the Institutional Care and Use Committee for Laboratory Animals (Protocol #16-067 and 17-038).

ST sensing

ST was measured using a tail-attached sensor as described previously [8] with modifications in the position of the thermistor and the housing size (Fig. 3). We used sensors with a thermistor positioned inside the housing (20.0 × 26.0 × 10.0 mm, weighing 5.5 g), rather than those with a thermistor positioned outside the housing (25.0 × 25.0 × 9.6 mm, weighing 7.7 g) described by Miura *et al*. [8]. The temperature resolution of the present sensor was 0.05°C. The sensor was attached to the cattle on around Day 10 of the estrous cycle (Day 0 = the day of ovulation) until around Day 10 of the subsequent estrous cycle and ST was measured at 30-sec or 10-min interval.

The sensor wirelessly transmitted the ST data with a sensor ID to the receiver connected to a personal computer where it was stored.

Confirmation of ovulation and estrus

In all 51 estrous cycles, the day of ovulation was confirmed by daily trans-rectal ultrasonography. Of 17 estrous cycles, the duration of estrus was confirmed based on standing-to-be-mounted behavior; namely, tested cattle were led out of the cattleshed to a paddock and tried to be mounted by hard mates at 3-h interval. The onset and end of estrus were defined as the time when a standing response was first observed and ceased, respectively. Of the remaining 34 estrous cycles, the day of estrus was determined by at least once daily observations of estrus-associated external signs (e.g., standing estrus, clear mucus discharge, or redness and edema of genitalia) with the results of the daily trans-rectal ultrasonography (one or two large follicles with a regressed corpus luteum).

Data preprocessing

In the data preprocessing phase, the maximum hourly ST was extracted to minimize the impacts of rapid momentary decreases of ST in the raw data [8]. After cleaning the data for missing values by forward-filling (*i.e.*, taking the last known value and using this to fill), ST changes were expressed as residual ST (rST = actual hourly ST − mean ST for the same hour on the previous 3 days) to exclude any influence of the circadian rhythm [8]. The exponentially weighted moving average (EWMA) [16] was then applied to the
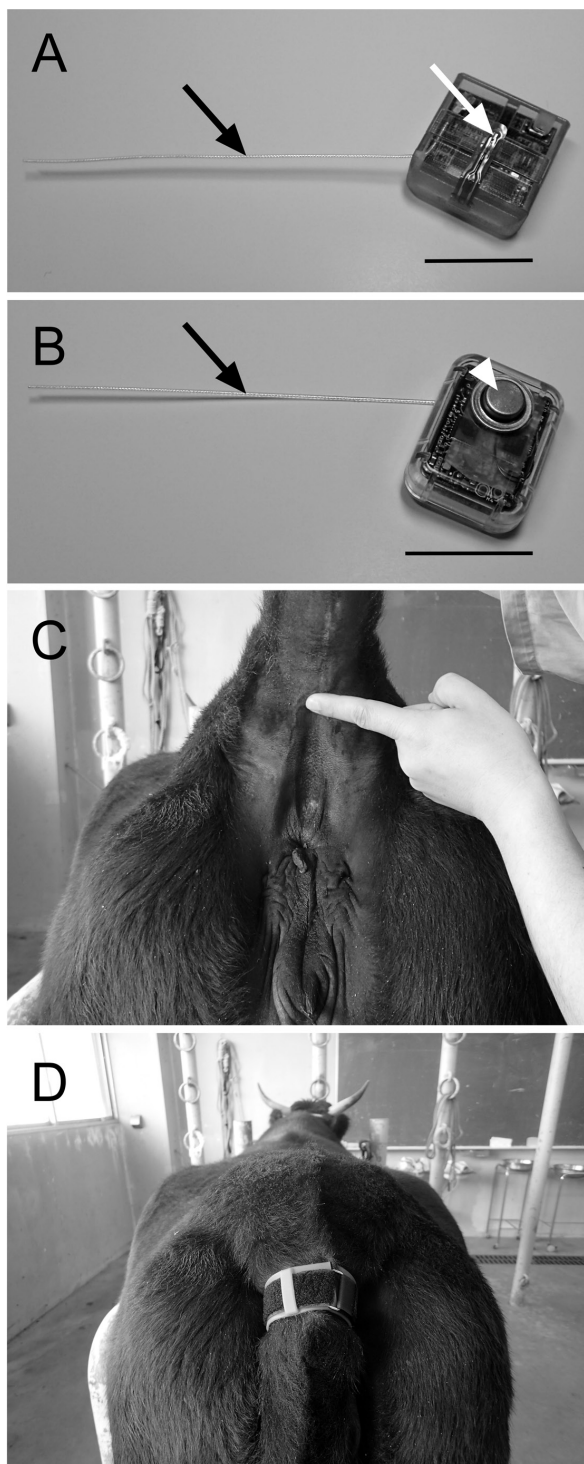
Fig. 3. Structure and usage of the wearable wireless tail base surface temperature (ST) sensor. A: Old type ST sensor (25.0 × 25.0 × 9.6 mm, weighing 7.7 g) used in the previous study [8]. A thermistor was positioned outside the housing (white arrow). B: New type ST sensor (20.0 × 26.0 × 10.0 mm, weighing 5.5 g) used in this study. A thermistor was positioned inside the housing just beneath the stainless steel cup (white arrow head). C: Position of the ST sensor attached the lower surface of the ventral tail base. D: ST senor attached on the surface of the ventral tail base. Black arrows in A and B indicate antenna. Bars in A and B are 2 cm.

rST data for smoothing by removing short-term (within a few hours) inconsequential fluctuations. EWMA was calculated as:

EWMA(0) = X(0)

EWMA(t) = αX(t) + (1−α) EWMA(t−1), t > 0,

where EWMA(t) is the EWMA at time "t" (t = 0, 1, 2, 3, •••), X(t) is the measured value at time "t," and EWMA(t−1) is the EWMA at time "t−1." The constant parameter α was set to 0.1 based on a previous study [2].

Development of hourly estrus detection models

To build estrus detection models, 13 features were extracted from the sensing data (Table 1). Only features that could be calculated from the current and past data were used to make estrus detection models applicable in actual situations. The 17 datasets corresponding with the standing estrous periods were used as training datasets. The training data during the standing estrous period were labeled as positive instances, and the remaining data were labeled as negative instances. Using the labeled training data, three estrus detection models were developed using RF (hyperparameters mtry = 11 and ntree = 500) [17], ANN (hyperparameters decay = 0.03, size = 17, and maxit = 1000) [18], and SVM (hyperparameters c = 5 and sigma = 0.6) [19]. All models were developed under the "caret" package [20] in R (version 3.4.0 for Windows; https://www.r-project.org) with 25 iterations of bootstrap resampling to minimize the effect of any single anomalous run on the results.

Validation of estrus detection models

Following the development of the three detection models, 34 unlabeled datasets (test datasets) corresponding with the day of estrus were applied to the models and classified into "estrus" or "not in estrus" every hour. At least 10 consecutive hours of "estrus" predictions were regarded as an estrus alert to minimize excessive alerts. When the estrus alert was given within 24 h (0000 – 2300 h) of the day of estrus, the alert was regarded as a true-positive predictor. Non-alerted estrus and alerted non-estrus were regarded as false-negative and false-positive predictors, respectively. Sensitivity and precision were calculated as (true-positive/[true-positive + false-negative]) and (true-positive/[true-positive + false-positive]), respectively.

Statistical analysis

For analyzing the changes in rST around estrus, hourly rST values during −120 h to +192 h from the onset of the estrus were compared with the mean value during the control non-estrous period (−192 h to −121 h) by the non-parametric Steel test. Sensitivities and precisions of the three estrus detection models were compared with Fisher's exact test and generalized score statistic, respectively. Statistical analyses were performed using R, and differences were considered significant at P < 0.05.

References

1. **Fisher AD, Morton R, Dempsey JM, Henshall JM, Hill JR.** Evaluation of a new approach for the estimation of the time of the LH surge in dairy cows using vaginal temperature and electrodeless conductivity measurements. *Theriogenology* 2008; **70**: 1065–1074. [Medline] [CrossRef]

2. **Higaki S, Miura R, Suda T, Andersson LM, Okada H, Zhang Y, Itoh T, Miwakeichi F, Yoshioka K.** Estrous detection by continuous measurements of vaginal temperature and conductivity with supervised machine learning in cattle. *Theriogenology* 2019; **123**: 90–99. [Medline] [CrossRef]

3. **Redden KD, Kennedy AD, Ingalls JR, Gilson TL.** Detection of estrus by radiotelemetric monitoring of vaginal and ear skin temperature and pedometer measurements of activity. *J Dairy Sci* 1993; **76**: 713–721. [Medline] [CrossRef]

4. **Clapper J, Ottobre J, Ottobre A, Zartman D.** Estrual rise in body temperature in the bovine I. Temporal relationships with serum patterns of reproductive hormones. *Anim Reprod Sci* 1990; **23**: 89–98. [CrossRef]

5. **Cooper-Prado MJ, Long NM, Wright EC, Goad CL, Wettemann RP.** Relationship of ruminal temperature with parturition and estrus of beef cows. *J Anim Sci* 2011; **89**: 1020–1027. [Medline] [CrossRef]

6. **Kyle BL, Kennedy AD, Small JA.** Measurement of vaginal temperature by radiotelemetry for the prediction of estrus in beef cows. *Theriogenology* 1998; **49**: 1437–1449. [Medline] [CrossRef]

7. **Randi F, McDonald M, Duffy P, Kelly AK, Lonergan P.** The relationship between external auditory canal temperature and onset of estrus and ovulation in beef heifers. *Theriogenology* 2018; **110**: 175–181. [Medline] [CrossRef]

8. **Miura R, Yoshioka K, Miyamoto T, Nogami H, Okada H, Itoh T.** Estrous detection by monitoring ventral tail base surface temperature using a wearable wireless sensor in cattle. *Anim Reprod Sci* 2017; **180**: 50–57. [Medline] [CrossRef]

9. **Miwa M, Matsuyama S, Nakamura S, Noda K, Sakatani M.** Prepartum change in ventral tail base surface temperature in beef cattle: comparison with vaginal temperature and behavior indices, and effect of ambient temperature. *J Reprod Dev* 2019; **65**: 515–525. [Medline] [CrossRef]

10. **Borchers MR, Chang YM, Proudfoot KL, Wadsworth BA, Stone AE, Bewley JM.** Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. *J Dairy Sci* 2017; **100**: 5664–5674. [Medline] [CrossRef]

11. **Wrenn T, Bitman J, Sykes JF.** Body temperature variations in dairy cattle during the estrous cycle and pregnancy. *J Dairy Sci* 1958; **41**: 1071–1076. [CrossRef]

12. **Banaee H, Ahmed MU, Loutfi A.** Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. *Sensors (Basel)* 2013; **13**: 17472–17500. [Medline] [CrossRef]

13. **Wolpert DH.** The lack of a priori distinctions between learning algorithms. *Neural Comput* 1996; **8**: 1341–1390. [CrossRef]

14. **Kotsiantis SB.** Supervised machine learning: A review of classification techniques. *In*: Maglogiannis IG, Karpouzis K, Wallace M, Soldatos J (eds.), Emerging Artificial Intelligence Applications in Computer Engineering. Amsterdam: IOS Press; 2007: 3–24.

15. **Hall MA.** Correlation-based feature selection of discrete and numeric class machine learning. *In*: Langley P (ed.), Proceedings of the Seventeenth International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers; 2000: 359–366.

16. **Roberts S.** Control chart tests based on geometric moving averages. *Technometrics* 1959; **1**: 239–250. [CrossRef]

17. **Liaw A, Wiener M.** randomForest: Breiman and Cutler's random forests for classification and regression. R package version 4.6-14. Available online at https://cran.r-project.org/web/packages/randomForest/index.html, 2019.

18. **Ripley B, Venables W.** nnet: Feed-forward neural networks and multinomial log-linear models. R package version 7.3-13. Available online at https://cran.r-project.org/web/packages/nnet/index.html, 2019.

19. **Karatzoglou A, Smola A, Hornik K, Maniscalco M, Teo C.** kernlab: Kernel-based machine learning lab. R package version 0.9-29. Available online at https://cran.r-project.org/web/packages/kernlab/index.html, 2019.

20. **Kuhn M.** caret: Classification and regression training. R package version 6.0-86. Available online at https://cran.r-project.org/web/packages/caret/index.html, 2019.