A data-driven prediction of lifetime resilience of dairy cows using commercial sensor data collected during first lactation

Wijbrand Ouweltjes,[1]* ● **Mirjam Spoelstra,[2] Bart Ducro,[2]** ● **Yvette de Haas,[2]** ● **and Claudia Kamphuis[2]** ●
[1]Wageningen University and Research, Animal Health and Welfare, PO Box 338, 6700 AH, Wageningen, the Netherlands
[2]Wageningen University and Research, Animal Breeding and Genomics, PO Box 338, 6700 AH, Wageningen, the Netherlands

ABSTRACT

Reliable prediction of lifetime resilience early in life can contribute to improved management decisions of dairy farmers. Several studies have shown that time series sensor data can be used to predict lifetime resilience rankings. However, such predictions generally require the translation of sensor data into biologically meaningful sensor features, which involve proper feature definitions and a lot of preprocessing. The objective of this study was to investigate the hypothesis that data-driven random forest algorithms can equal or improve the prediction of lifetime resilience scores compared with ordinal logistic regression, and that these algorithms require considerably less effort for data preprocessing. We studied this by developing prediction models that forecast lifetime resilience of a cow early in her productive life using sensor data from the first lactation. We used an existing data set from a Dutch experimental herd, with data of culled cows for which birth dates, insemination dates, calving dates, culling dates, and health treatments were available to calculate lifetime resilience scores. Moreover, 4 types of first-lactation sensor data, converted to daily aggregated values, were available: milk yield, body weight, activity, and rumination. For each sensor, 14 sensor features were calculated, of which part were based on absolute daily values and part on relative to herd average values. First, we predicted lifetime resilience rank with stepwise logistic regression using sensor features as predictors and a *P*-value of <0.2 as the cut-off. Next, we applied a random forest with the 6 features that remained in the final logistic regression model. We then applied a random forest with all sensor features, and finally applied a random forest with daily aggregated values as features. All models were validated with stratified 10-fold cross-validation with 90% of the records in the training set and 10%
in the validation set. Model performances expressed in percentage of correctly classified cows (accuracy) and percentage of cows being critically misclassified (i.e., high as low and vice versa) ± standard deviation were 45.1 ± 8.1% and 10.8% with the ordinal logistic regression model, 45.7 ± 8.4% and 16.0% with the random forest using the same 6 features as the logistic regression model, 48.4 ± 6.7% and 10.0% for the random forest with all sensor features, and 50.5 ± 6.3% and 8.4% for the random forest with daily sensor values. This random forest also revealed that data collected in early and late stages of first lactation seem to be of particular importance in the prediction compared with that in mid lactation. Accuracies of the models were not significantly different, but the percentage of critically misclassified cows was significantly higher for the second model than for the other models. We concluded that a data-driven random forest algorithm with daily aggregated sensor data as input can be used for the prediction of lifetime resilience classification with an overall accuracy of ~50%, and provides at least as good prediction as models with sensor features as input.
Key words: dairy cow, resilience, sensors, machine learning, random forest

INTRODUCTION

Resilience of animals refers to their capacity to cope with short-term perturbations in their environment and return to their undisturbed status (Colditz and Hine, 2016). Lifetime resilient dairy cows are characterized as animals that have a high probability of completing multiple lactations, exhibit good productive and reproductive performance, face few health problems that they overcome easily, and are efficient and consistent in their milk production (Adriaens et al., 2020). Improving resilience in dairy cows yields substantial advantages: it contributes to animal health and welfare (Mulder and Rashidi, 2017), improves the productivity of the farm (Colditz and Hine, 2016), reduces the environmental impacts of the sector (Herzog et al., 2018), and reduces the need for antibiotic usage (König and May, 2019).

Additionally, intensification of the livestock sector, with increasing herd size and limited availability of labor, results in a need for easy-to-manage and healthy herds (Elgersma et al., 2018).

Early predictions of lifetime resilience can be used as input for decision support tools for farmers and to improve resilience through breeding (Berghof et al., 2019). However, an early identification of cows that will smoothly go through multiple lactations remains difficult. The challenges an animal will encounter cannot be predicted and will differ from individual to individual, from farm to farm. On top of that, challenges will vary in seriousness (e.g., udder infections with minor versus major pathogens), and indicators for general resilience have currently not been defined (Berghof et al., 2019). The ability to cope with these challenges is most clearly shown at the end of productive life, and is affected by management decisions (van Pelt et al., 2016).

The development of sensor technologies has enabled frequent measurement of physiological, behavioral, and production indicators on individual cows. These data can be used by farmers to improve cow management while applying systems known as precision livestock farming (Berckmans, 2014). Sensor systems are regularly used on dairy farms nowadays (Steeneveld and Hogeveen, 2015; Gargiulo et al., 2018; Groher et al., 2020; Dela Rue et al., 2020), and this is expected to increase in the near future. So far, the majority of management support tools using sensor data are targeted at detecting specific events [e.g., indicating estrous or health events (Friggens and Thorup, 2015)], but there is also growing interest to exploit sensor data for predicting more complex phenotypes. Poppe et al. (2020) used fluctuations in daily milk production records as indicators for resilience and found that log-transformed variance of deviations from fitted curves was the most promising parameter with a heritability of 0.20 to 0.24 and favorable genetic correlations with functional traits. Recent technological developments and increased usage of other sensors give the opportunity to combine continuous data recordings from different sensors to further improve the prediction of resilience. As an example, Adriaens et al. (2020) reported that adding activity sensor data significantly ($P < 0.01$) improved prediction accuracies for resilience compared with predictions based on daily milk features alone. Both Poppe et al. (2020) and Adriaens et al. (2020) tried to construct resilience indicators from a biological point of view. Data processing and the construction of biologically meaningful features from these sensor data to be included in regression models played a crucial role in their approaches. Currently, machine learning algorithms play a central role in the recognition of patterns in large and complex data sets in other domains (van

der Heide et al., 2019). One of these algorithms, random forest, showed high prediction accuracy and potential to unravel conditional relations between variables and interactions between variables for subsets of samples when applied for pattern recognition, while requiring minimal effort for data processing (Touw et al., 2013). Machine learning algorithms have not yet been adopted for the purpose of predicting resilience in dairy cows, but were used to predict survival to second lactation (van der Heide et al., 2019), which is a similar trait.

We hypothesized that random forest algorithms, particularly suited for situations with a large number of explanatory variables (Breiman, 2001), can improve the prediction of lifetime resilience scores and that these algorithms require less effort for data preprocessing compared with the traditional method (i.e., stepwise logistic regression). We studied this by developing prediction models that forecast lifetime resilience of a cow early in her productive life using sensor data from the first lactation. First, we developed an ordinal logistic regression model (considered as reference) with predictive features based on the entire lactation period similar to Adriaens et al. (2020). We then compared this with several random forest models, where we successively changed the set of predictors. This set of predictors was either based on the entire lactation period (similar to the features used in the regression model), or on daily sensor measurements.

MATERIALS AND METHODS

Data

Data from 2,127 cows born between March 1993 and May 2014 were collected at the research facility Dairy Campus, Lelystad, the Netherlands (formerly known as Waiboerhoeve). Birth dates, insemination dates, calving dates, culling dates, and health treatments were retrieved from the farm database for all cows for which at least one milk yield record was available. Moreover, 4 types of sensor data were retrieved: milk yield, BW, activity, and rumination. The latter 2 were obtained from SCR tags provided by Lely Industries. Milk yield was recorded either in the milking parlor (twice daily) or in a Lely Astronaut automatic milking system (**AMS**; at each milking). Body weight was recorded at entry of the parlor or in the AMS.

Calculation of Lifetime Resilience Scores and Resilience Classification

From the initial 2,127 cows, lifetime resilience scores (**LRS**) could be calculated for 1,800 cows that fulfilled the following requirements:

- The cow was culled at the time the data were collected for this study's purpose (i.e., before July 2016)
- Complete data regarding inseminations, calving dates, disease, and treatment records were available throughout the cows lifetime.
- Age at first calving had to be between 640 and 1,100 d.
- Similar to the approach of Adriaens et al. (2020), LRS was calculated according to the following equation:

$$\text{LRS} = 500 \times n - \left(A1 - \overline{A1}\right) + \sum_{p=2}^{p=n}\left(\overline{CI_p} - CI_p\right) - \sum_{p=1}^{p=n}\left(TD_p\right) + FP_n,$$

[1]

where LRS refers to the lifetime resilience score as a summation of scores for lactations 1 to n, 500 is the standard credit points for each calving, $A1$ = age at first calving in days, $\overline{A1}$ = herd average for age at first calving in days, CI_P = calving interval in days for all succeeding parities ($P > 1$), $\overline{CI_p}$ = herd average for calving interval in days for all succeeding parities ($P > 1$; so, calvings with below herd average values receive additional credit points and animals with above herd average values get points subtracted), TD_P = number of days the cow was curatively treated for a health problem during parity p, and $FP_n = 25 \times$ number of inseminations in the last lactation. When an animal was culled within 100 d after the last calving the remaining days up to d 100 were regarded as days the cow was treated for a health problem.

The $\overline{A1}$ for the farm was 735 d and $\overline{CI_p}$ was 414 d. The LRS of the 1,800 cows with 5,771 calvings (1–11 per cow) ranged from 31 to 6,031, with an average value of 1,518. These cows were classified evenly into 3 categories: high (33% highest LRS), medium, and low (33% lowest LRS).

Selection and Processing of Sensor Data

From all available sensor data, only data from d 1 to 300 during first lactations were included for processing. Raw sensor data were aggregated to daily values; for activity and rumination this was done by summing all bihourly measures per day; records with incomplete information were not taken into account. For BW the average of all weight measurements per day was taken (usually the cows were weighed at each milking). Fi-

nally, for milk yield, 24-h yields were obtained by summing all yields during a calendar day, dividing this sum by the time interval in minutes from the last milking on the preceding day until the last milking of this calendar day and multiplying by 1,440. In this procedure, daily yields were discarded when intervals exceeded 2,160 min (36 h). Moreover, for all 4 sensors, outliers of these daily values (mean ± 4 SD) were removed and set as missing. For analysis, a data set with records per cow and DIM was constructed with data from cows that had at least 100 full-day observations for each of the 4 sensors. Only 370 of the 1,800 cows (21%) met these criteria and were included in this study, with first calvings between 2009 and 2016. The main reason not to include so many lactations was that activity and rumination data were only available from 2007 onward. It was not required that days with sensor information were identical for all sensors (e.g., cows could have milk yield data for d 1–300, BW for d 1–200, and activity and rumination data for d 150–300). Milk yield information was nearly complete (available for 99.4% of the records), activity, rumination, and BW were available for 78.1%, 78.8%, and 79.4% of the records, respectively. Rumination data were always available when activity data were available and 92% of the records with activity data also had BW data. Of the data set, 71.7% had information for all 4 sensors. In total, the sensor data set had 106,689 daily records (288.3 per lactation on average, ranging from 129 to 300). The final data set included cow ID, date of recording, DIM, recorded sensor value, and herd average for that sensor value (based on all sensor information available for the 1,800 cows for which LRS were constructed for this DIM in first lactation) and relative (to herd) sensor value calculated as recorded value − herd average value for each sensor (activity, rumination, milk yield, and weight). A general overview of the sensor data used for this study and the herd mean values is shown in Table 1.

Calculation of Sensor Features

For each of the sensor time series, 14 sensor features (i.e., curve parameters) were calculated from the daily aggregated measurements, totaling 56 sensor features. First, the absolute daily values were made relative to the herd average, creating relative curves (Figure 1). Subsequently, a linear regression line was fitted through this relative curve (Figure 1). The mean, minimum, maximum, 25th percentile, 50th percentile, 75th percentile, standard deviation, skewness, kurtosis, and autocorrelation (lag1) were computed based on the relative curve of each cow. The slope and intercept were obtained from the regression line through the relative
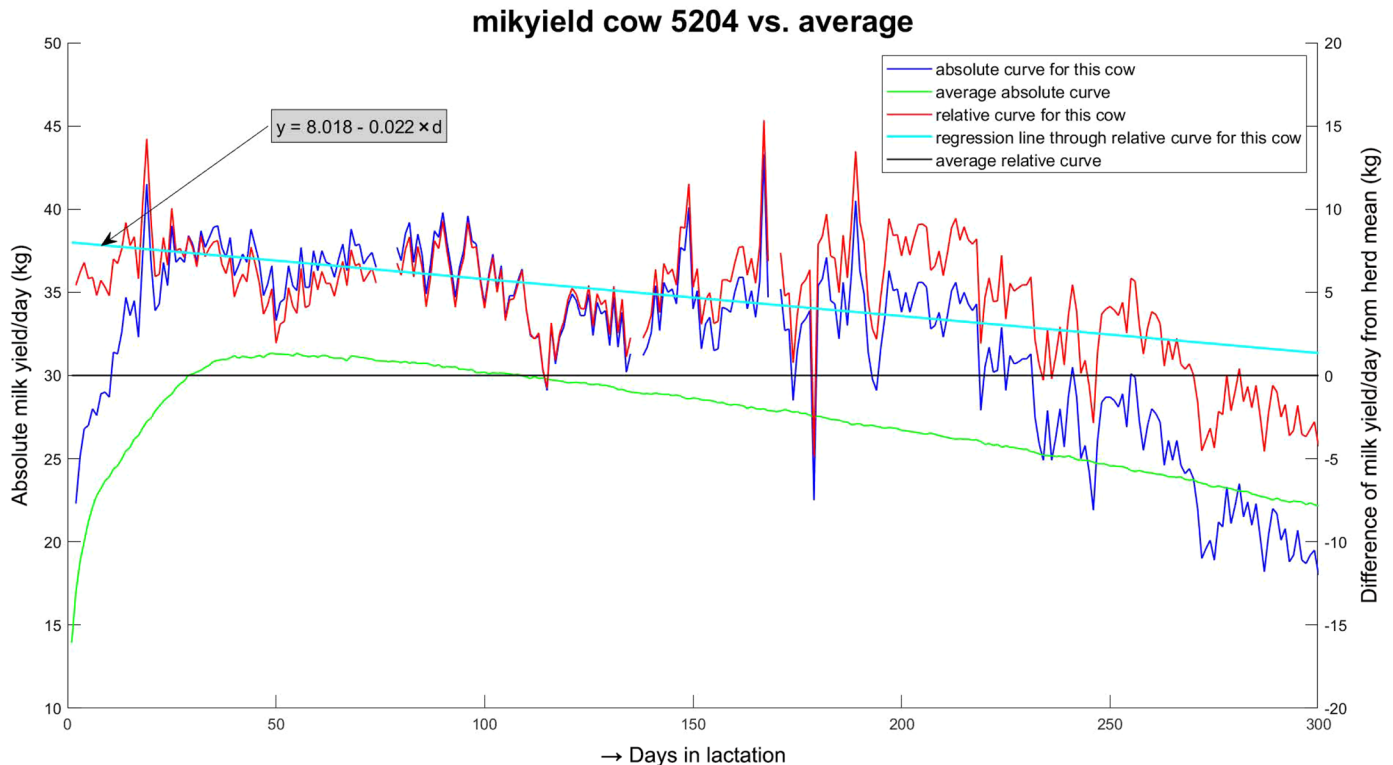
Figure 1. Assessment of curve parameters per sensor based on the relative curve (red line), or its regression line (light blue line) of a cow. The absolute values of a cow are in dark blue, the absolute herd mean values in green, and the relative herd mean values in black.

curve of each cow and the residual standard deviation was calculated from the residuals of the regression. Correlation between the relative curve values and fitted values was calculated as an additional feature. Additionally, absolute daily sensor values and their lactation averages were also used as features, totaling 1,204 sensor features, but these were only included in the random forest analysis.

Statistical Analysis

A logistic regression and 3 random forests (Figure 2) were constructed to predict LRS classes (high, medium, or low) based on the set of 56 sensor features as described in the previous section to explore the predictive power of the sensor data. Random forest (Breiman,

2001) models were parameterized using a random grid search (for each model separately). All models were validated with stratified 10-fold cross-validation with 90% of the records in the training set and 10% in the validation set.

We used a structured approach in developing the different models. First, an ordinal logistic regression model was constructed with all 56 sensor features. In a stepwise selection procedure the sensor features with the highest P-values were deleted from the model (one at a time), until a final model remained with features with P-values ≤ 0.2. Second, a random forest approach using those sensor features that were kept in the final ordinal logistic regression model was applied. This was done to enable a one-to-one comparison of a random forest and a logistic regression approach in their ability

Table 1. Descriptive statistics of the first-lactation sensor data of dairy cows (n = 370) used for this study

| Sensor	Mean	SD	Minimum	Maximum	Herd average
Activity (count)	417.7	97.2	0.0	838.0	413.4
Rumination (count)	439.7	124.0	15.0	816.0	436.1
Milk yield (kg/d)	27.0	6.0	0.1	67.5	27.5
Weight (kg)	574.1	57.0	351.0	943.0	569.0

to predict LRS classifications using exactly the same information. Third, a random forest approach using all 56 lactation sensor features that were calculated was applied. This was to test whether including features that were considered insignificant in the stepwise selection for the ordinal logistic regression could improve the predictive performance of the random forest model. Finally, a random forest approach that did not include the sensor features but that used all daily sensor values and their lactation averages was applied. This was to investigate whether the efforts for preprocessing could be reduced while maintaining predictive ability. All of the data editing and modeling was performed with the programming language Python. For both the ordinal logistic regression and the random forests an open source machine learning library, scikit-learn (https://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html), was used. Furthermore, the following Python packages and modules were used: Pandas, NumPy, Os, SciPy, Matplotlib, Seaborn, and Statsmodels.

Model performances were primarily evaluated by the accuracy (**ACC**) of predictions, which is the percentage of correctly classified cows. Additionally, we looked
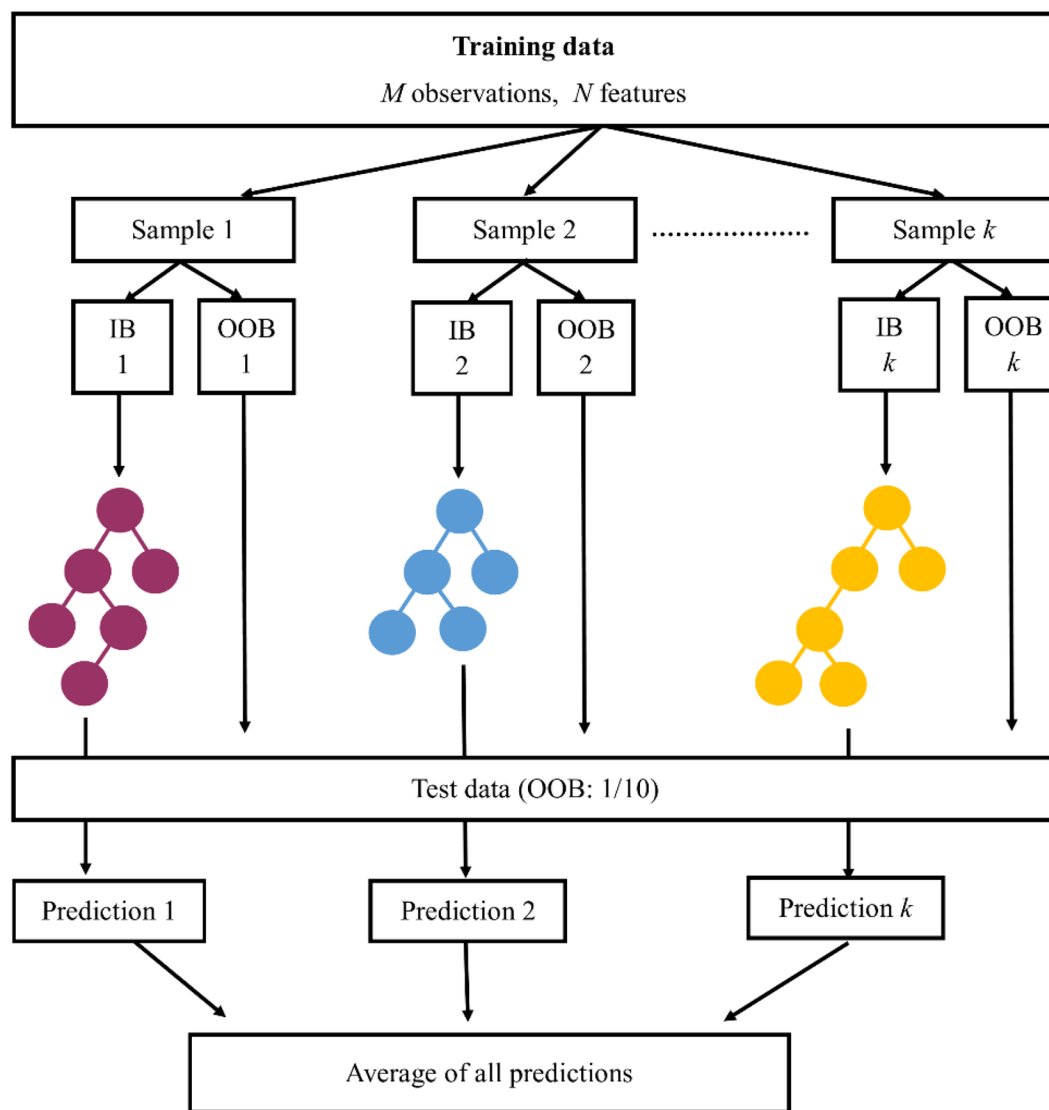

Figure 2. Structure of a random forest algorithm on a data set containing N features, 370 observations, and k trees. Random forest is an ensemble of many individual decision trees; each individual tree randomly samples from the data set with replacement, resulting in different trees. In bag (IB) is used as training data, whereas out of bag (OOB) is used to test the model. The average of all predictions results in the final prediction.

at the proportion of cows that were misclassified in opposing categories (i.e., high in low and vice versa). This will be referred to as critically misclassified (**CritMis**). For the last random forest model, feature importances were calculated with the Python scikit-learn package. The ACC and CritMis of the 4 models were compared with a McNemar test to estimate significance levels of the differences.

RESULTS

The stepwise selection of the ordinal logistic regression model resulted in the inclusion of 6 significant sensor features. This model yielded an ACC (calculated for each replicate) of 45.1 ± 8.1% and a weighted F_1-score (calculated from the aggregated results) of 44.5%. The features that were included in the final model were slope of the regression line through the relative milk

yields, standard deviation of the relative BW, minimum of the relative activities, skewness of the regression line through the relative activities, maximum of the relative rumination values, and skewness of the regression line through the relative rumination values. Prediction techniques for categorical variables such as ordinal logistic regression models calculate probabilities for each of the classes and the class with highest probability is presented as the predicted class. Relationships of the significant features with predicted resilience are given in Figure 3, where the range of the feature for which the probabilities are plotted is scaled from 0 (minimum value as observed in the data set) to 1 (maximum value as observed in the data set). The slope of the milk yield curve ranged from −0.051 to +0.073 in our data set, standard deviation of the relative BW from 9.0 to 59.1 kg, minimum of the activity curve from −441 to 185 min/d, skewness of the activity curve from −2.45 to
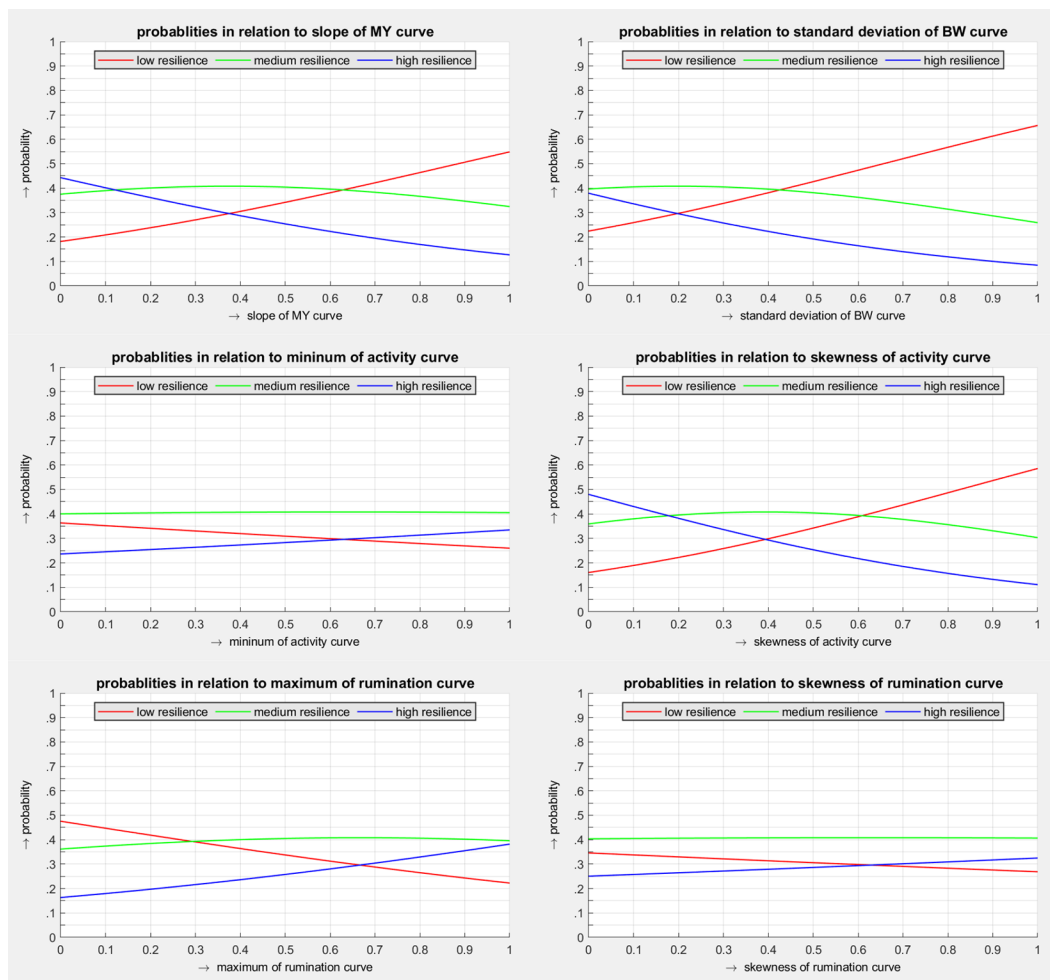
Figure 3. Calculated probabilities of belonging to low, medium, or high resilience categories for observed ranges (0 = minimum, 1 = maximum) of sensor features that are kept in the final ordinal logistic regression prediction model (assuming the other features are at average values). MY = milk yield in kilograms.

Table 2. Normalized confusion matrix, including the percentage of critically misclassified animals (CritMis[1]) for each model[2]

Model	True classification	Predicted classification			CritMis
		Low	Medium	High	
1	Low	0.103	0.159	0.062	
	Medium	0.084	0.216	0.081	
	High	0.046	0.116	0.132	
					0.108
2	Low	0.143	0.089	0.092	
	Medium	0.111	0.159	0.111	
	High	0.068	0.073	0.154	
					0.160
3	Low	0.135	0.127	0.062	
	Medium	0.103	0.178	0.100	
	High	0.038	0.086	0.170	
					0.100
4	Low	0.170	0.103	0.051	
	Medium	0.100	0.165	0.116	
	High	0.032	0.092	0.170	
					0.084

[1]Cows that were misclassified in opposing categories (i.e., high in low and vice versa).

[2]1 = ordinal logistic regression with 6 significant sensor features, 2 = random forest with 6 significant sensor features from ordinal logistic regression, 3 = random forest with all 56 sensor features, 4 = random forest with absolute sensor values.

5.99, maximum of the rumination curve from −192 to 383 min/d, and skewness of the rumination curve from −2.61 to 1.89. To create these plots, all other feature values were fixed at average values observed in the data set.

Figure 3 shows that with increasing slope of the milk yield curve, increasing standard deviation of the BW curve and increasing skewness of the activity curve the probability decreased that a cow will realize a high LRS. For a cow with the lowest (most negative) slope of the milk yield curve and average values for the other features, the predicted probability that it will realize a high LRS is around 45%, and that it will realize a low LRS is less than 20%. For a cow with the highest (most positive) slope of the milk yield curve and average values for the other features, the predicted probabilities are around 12% and close to 60%, respectively. Increasing the minimum of the activity curve and increasing the maximum and skewness of the rumination curve coincide with increasing probability that a cow will realize a high LRS.

Confusion matrices of the 4 models are presented in Table 2 and significance levels of the differences for ACC and CritMis are in Table 3. Cross-validation accuracies reported for the other models were 45.7 ± 8.4% for the random forest using the 6 significant features from the logistic regression model, 48.4 ± 6.7% for the random forest with all 56 available curve features, and 50.5 ± 6.3% for the random forest with absolute sensor values as input variables. None of the differences in ACC values was significant, but the second model had

significantly higher CritMis than the other models ($P = 0.001$ for comparison of models 1 and 2, $P = 0.003$ for comparison of models 2 and 3, and $P = 0.000$ for comparison of models 2 and 4). On average, model 4, which requires the least amount of preprocessing, had at least equal performance compared with the other models.

The 1,204 "features" (absolute daily sensor values) that were used in the last model were ranked for their importance. There were relatively few features with considerable importance, and many that only contributed marginally to the model. The 20 most important features in the model with absolute daily sensor data (Table 4) contained 12 activity features, 5 rumination features, 3 milk yield features, and no BW features. The majority of these features were from early lactation

Table 3. P-values for pairwise comparison of accuracy and percentage of critically misclassified animals (CritMis[1]) for the different models[2]

Models compared	Accuracy	CritMis
1 and 2	0.908	0.001
1 and 3	0.327	0.775
1 and 4	0.135	0.272
2 and 3	0.407	0.003
2 and 4	0.171	0.000
3 and 4	0.530	0.430

[1]Cows that were misclassified in opposing categories (i.e., high in low and vice versa).

[2]1 = ordinal logistic regression with 6 significant sensor features, 2 = random forest with 6 significant sensor features from ordinal logistic regression, 3 = random forest with all 56 sensor features, 4 = random forest with absolute sensor values.

Table 4. Identification and importance of the 20 most important features in the random forest model with absolute daily sensor values as features

Absolute daily value	Days in lactation	Rank	Importance (× 100)
Activity	33	1	1.311
Rumination	33	2	1.109
Rumination	242	3	0.817
Activity	41	4	0.740
Milk yield	246	5	0.674
Rumination	247	6	0.645
Activity	51	7	0.644
Rumination	41	8	0.586
Milk yield	300	9	0.530
Activity	49	10	0.489
Activity	38	11	0.459
Activity	242	12	0.438
Milk yield	244	13	0.432
Activity	66	14	0.428
Rumination	26	15	0.412
Activity	30	16	0.403
Activity	57	17	0.395
Activity	245	18	0.375
Activity	299	19	0.369
Activity	26	20	0.368

(DIM 26–66) and some were from late lactation (DIM 242–300), and none were from mid lactation. Milk yield variation only seemed to be related to resilience at the end of the lactation, whereas activity and rumination contributed both in early and late lactation. Overall, the summed importance of all 301 activity features was 0.315, the summed importance of all 301 rumination features was 0.258, the summed importance of all 301 milk yield features was 0.232, and the summed importance of all 301 BW features was 0.195.

DISCUSSION

Comparison of Different Models

Classification performances for resilience in the current study varied between 45.1 and 50.5% for the different models that were developed with data from just one (research) farm. This result is in line with a reported accuracy of 55.5% (ranging from 46.7 to 84.0% for 27 farms) for a logistic regression model developed by Adriaens et al. (2020) that also used predictive features based on first lactation milk yield and activity sensor data and used a similar definition of resilience. The ACC of the logistic regression and the random forest model using the same features were similar (45.1% and 45.7%, respectively), but the CritMis was significantly higher for this random forest (16.0%, compared with 10.8% for the logistic regression). This indicates that the random forest model performed worse when comparing these 2 models. However, including features that were deemed insignificant by the ordinal logistic regres-

sion improved the ACC of the random forest to 48.4% and reduced the CritMis to 10.0%, although these values were not significantly different from those of the logistic regression model. To put this in perspective: for an average Dutch herd with 100 cows 11 cows would be completely misclassified with the logistic regression model (5 that are high classified as low resilient and 6 that are low classified as high resilient) versus 10 with this random forest (4 that are high classified as low resilient and 6 that are low classified as high resilient). For comparison, the logistic regression model was also evaluated with all 56 features included. This resulted in an ACC of 45.7 ± 6.0% and 11.6% CritMis. These findings suggest that predicting resilience using a random forest yields similar results compared with a logistic regression. This is consistent with a recent study from van der Heide et al. (2019) that concluded that regression and random forest could predict survival to second lactation for dairy cows with similar performances. Nearly equal prediction performance in our study was achieved with a random forest that used daily average sensor values. This indicates that the added value of preprocessing to determine sensor features with potential biological meaning for the predictions was very limited. Our results show that LRS classifications can be predicted using a random forest and daily sensor values with similar accuracies as more time-consuming methods such as a stepwise ordinal logistic regression and sensor features, although this result should be verified for more than one farm.

Biological Interpretation of Features Used in Ordinal Logistic Regression

A drawback of a best-prediction oriented approach could be that biological interpretation of the relationships between sensor data and resilience is obscured. Although these predictions provide insight into model performance, they do not clearly illustrate how model features are linked with outcomes [e.g., how LRS is expected to change when the slope of the regression line through the relative milk yields changes from 0 (smallest value) to 1 (largest value)]. It can be debated if the relationships between sensor features and LRS as determined from ordinal logistic regression in the current study and illustrated in Figure 3 are biologically interpretable. The study of Adriaens et al. (2020) revealed that models were farm specific. This could be a matter of relative importance of features, but it could also mean that relationships between features and LRS were different for different farms. The large variability in correlation coefficients between predicted lifetime resilience rank at farm level and sensor features suggest the latter. Our ordinal logistic regression results

indicate that animals with a relatively large positive slope of the first-lactation relative milk yield curve (so, lactating heifers with above average persistency) are more likely to have low resilience than animals with below average persistency. This could be related to the antagonistic relationship between milk yield and fertility (which is a strong driver of resilience). In the low resilience group there will be animals included that were culled after first lactation because they were not in calf. Thus, the question remains whether high persistency is causally related to low resilience. Increasing variation of BW was linked with decreasing resilience. This could indicate increased sensitivity for environmental changes, but if this is the underlying cause of the relationship the same could be expected for the variation of data from the other sensors. Therefore, it is remarkable that for BW only, the standard deviation was kept in the logistic regression model after the parameter selection. Increasing skewness of activity (from negative to positive values), decreasing skewness of rumination, increasing minimum of activity, and increasing maximum of rumination were linked with decreasing resilience. This suggests that animals that have relatively many days with below average activity or relatively many days with above average rumination counts have a higher resilience than animals that show the opposite. We have no biological interpretation for these phenomena.

Interpretation of Results of Random Forest Models

Our analyses with random forest models have revealed some interesting findings. The lactation curve features that were used in the final ordinal logistic regression model and the first random forest model were not the features with the highest feature importance in the random forest model with all lactation curve features included (results not shown). This, but also the fact that adding features on average improved the predictions, shows that the stepwise feature selection procedure used with ordinal logistic regression is probably not the optimal way to subtract information from the data. The random forest with absolute daily sensor data, that does not require preprocessing in contrast with the other analyses, showed at least equal performance. Moreover, this model also revealed that particularly for activity and rumination information from both early and late lactation data contributed more to the model predictions than information from mid lactation. Weighing information dependent on stage of lactation was not included in the calculation of features at the lactation level. It should be verified with data from other herds if this phenomenon is consistent. Pre-

liminary logistic regression calculations with features derived for specific stages of lactation (d 1 −7, d 1 −60, d 61–150, and d 151–300) did not show clearly improved predictions compared with features derived for d 1 to 300.

Resilience Predictions from Sensor Data

Although LRS predictions could also be useful in relation to health and welfare monitoring, in a lifetime-approach of resilience the fact that challenges and vulnerability are changeable throughout the lifetime is ignored. Because the objective was to compare predictions early in the productive life of animals, we only used first-lactation sensor data from d 1 to 300 after first calving for predictions. Potentially, sensor data from later lactations could have improved predictions, but this would require the need to account for parity effects. Another approach of resilience is described by van Dixhoorn et al. (2018). They focused on the transition period as an informative period where a cow is considerably challenged. Usually cows are not milked around 6 to 8 wk before a calving, so milk yield will often not be available to monitor cows close to calving. The other sensor data used in this study, however, could be obtained throughout the transition phase (also before first calving), but were not available due to the routines on the experimental farm. Therefore, we could not explore their potential to improve the prediction of LRS, but we expect that sensor data collected during transition periods are very useful, at least to increase biological understanding of the concept of resilience.

Limitations of the Study

The data used for this study were obtained during a rather long time span, but we have ignored possible changes in average age at first calving and calving intervals as well as in first-lactation sensor data [e.g., an increase in 305-d milk yield over the years (2009–2016)]. Moreover, experiments and changeable external circumstances may have affected culling decisions. A factor that potentially affects the milk yield sensor data in particular is that part of the milkings were done with an AMS with variable milking frequency, and part in a parlor with milking 2 times a day. Although we filtered the sensor data for outliers, we encountered substantial irregularities in the patterns for individual cows. Perhaps we should reconsider the way to remove erroneous records from the data. Care should be taken so that true outliers that can be informative for resilience are not removed. It is likely that low resilient cows are overrepresented in the final data set because we only

included cows that were culled at the time this data set was constructed. This may have affected the results, which is not taken into account.

Our definition of LRS is similar to that of Adriaens et al. (2020), but differs in some aspects. We did not include milk yield because reliable 305-d yields were not available at the time of construction of the data set. This could have affected the ranking, but the correlations between LRS calculated with and without account for 305-d yields were highly positive (0.854). Moreover, average first-lactation 305-d yields were similar for the low, medium, and high groups (8,100, 8,275, and 8,082 kg, respectively). Also the average production in comparison with the peers (lactation value) was similar for the 3 groups (102, 105, and 104, respectively). Therefore, we expect that including milk yield in our LRS calculation would not have altered our main findings. The main argument of Adriaens et al. (2020) to include 305-d yield in calculating LRS was that animals with higher yields probably had better health. However, we have directly included treatment days into our calculation, and on average there was hardly any difference between the groups of animals with low, medium, or high resilience in the number of days they were treated for a disease (2.1, 2.1, and 1.9, respectively). This suggests that differences in health during first parity were relatively small. The events we included were diagnosed and treated; therefore, we cannot completely rule out that differences in health during first lactation have contributed to differences in milk yield. Fertility is taken into account through age at first calving and calving intervals, and number of inseminations in the last parity. It is assumed that the majority of the animals that were not inseminated after the last calving were culled for other reasons than poor fertility, whereas for animals that were inseminated after the last calving fertility problems presumably were one of the reasons for culling. In first lactation, the animals with low LRS on average had 3.6 inseminations, the animals with medium LRS on average had 2.3 inseminations, and the animals with high LRS on average had 2.0 inseminations.

CONCLUSIONS

Performance of ordinal logistic regression and random forest to predict lifetime resilience classification from sensor features was similar. Instead of features derived from sensor data, absolute daily sensor data can be used in combination with a random forest algorithm for the prediction of lifetime resilience classification with similar predictive performance. This has the advantage that laborious preprocessing is no longer required. In addition, random forest prediction using daily sensor

data revealed that early and late stages of first lactation are of particular importance in the prediction. Further research on the effect of lactation stages when data are collected on the model performance is required to confirm this finding, but it is due to the application of machine learning that we obtained this additional insight. Filtering of outliers from raw sensor data due to recording errors while keeping resilience- or challenge-related deviations deserves more attention.

REFERENCES

Adriaens, I., N. C. Friggens, W. Ouweltjes, H. Scott, B. Aernouts, and J. Statham. 2020. Productive life span and resilience rank can be predicted from on-farm first-parity sensor time series but not using a common equation across farms. J. Dairy Sci. 103:7155–7171. https://doi.org/10.3168/jds.2019-17826.

Berckmans, D. 2014. Precision livestock farming technologies for welfare management in intensive livestock systems. Rev. Sci. Tech. 33:189–196. https://doi.org/10.20506/rst.33.1.2273.

Berghof, T. V., M. Poppe, and H. A. Mulder. 2019. Opportunities to improve resilience in animal breeding programs. Front. Genet. 9:692. https://doi.org/10.3389/fgene.2018.00692.

Breiman, L. 2001. Random forests. Mach. Learn. 45:5–32. https://doi.org/10.1023/A:1010933404324.

Colditz, I. G., and B. C. Hine. 2016. Resilience in farm animals: Biology, management, breeding and implications for animal welfare. Anim. Prod. Sci. 56:1961–1983. https://doi.org/10.1071/AN15297.

Dela Rue, B. D., C. Eastwood, J. Edwards, and S. Cuthbert. 2020. New Zealand dairy farmers preference investments in automation technology over decision-support technology. Anim. Prod. Sci. 60:133–137. https://doi.org/10.1071/AN18566.

Elgersma, G. G., G. de Jong, R. van der Linde, and H. A. Mulder. 2018. Fluctuations in milk yield are heritable and can be used as a resilience indicator to breed healthy cows. J. Dairy Sci. 101:1240–1250. https://doi.org/10.3168/jds.2017-13270.

Friggens, N., and V. Thorup. 2015. From monitoring to precision phenotyping: Towards a systemic use of precision livestock measures in dairy herds. Proc. N.Z. Soc. Anim. Prod. 75:145–148.

Gargiulo, J. I., C. Eastwood, S. Garcia, and N. Lyons. 2018. Dairy farmers with larger herd sizes adopt more precision dairy technologies. J. Dairy Sci. 101:5466–5473. https://doi.org/10.3168/jds.2017-13324.

Groher, T., K. Heitkämper, and C. Umstätter. 2020. Digital technology adoption in livestock production with a special focus on ruminant farming. Animal 14:2404–2413. https://doi.org/10.1017/S1751731120001391.

Herzog, A., C. Winckler, and W. Zollitsch. 2018. In pursuit of sustainability in dairy farming: A review of interdependent effects of animal welfare improvement and environmental impact mitigation. Agric. Ecosyst. Environ. 267:174–187. https://doi.org/10.1016/j.agee.2018.07.029.

König, S., and K. May. 2019. Invited review: Phenotyping strategies and quantitative-genetic background of resistance, tolerance and resilience associated traits in dairy cattle. Animal 13:897–908. https://doi.org/10.1017/S1751731118003208.

Mulder, H. A., and H. Rashidi. 2017. Selection on resilience improves disease resistance and tolerance to infections. J. Anim. Sci. 95:3346–3358. https://doi.org/10.2527/jas.2017.1479.

Poppe, M., R. Veerkamp, M. van Pelt, and H. Mulder. 2020. Exploration of variance, autocorrelation, and skewness of deviations from lactation curves as resilience indicators for breeding. J. Dairy Sci. 103:1667–1684. https://doi.org/10.3168/jds.2019-17290.

Steeneveld, W., and H. Hogeveen. 2015. Characterization of Dutch dairy farms using sensor systems for cow management. J. Dairy Sci. 98:709–717. https://doi.org/10.3168/jds.2014-8595.

Touw, W. G., J. R. Bayjanov, L. Overmars, L. Backus, J. Boekhorst, M. Wels, and S. A. van Hijum. 2013. Data mining in the life sciences with random forest: A walk in the park or lost in the jungle? Brief. Bioinform. 14:315–326. https://doi.org/10.1093/bib/bbs034.

van der Heide, E. M., R. Veerkamp, M. van Pelt, C. Kamphuis, I. Athanasiadis, and B. Ducro. 2019. Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle. J. Dairy Sci. 102:9409–9421. https://doi.org/10.3168/jds.2019-16295.

van Dixhoorn, I. D., R. de Mol, J. van der Werf, S. van Mourik, and C. van Reenen. 2018. Indicators of resilience during the transition period in dairy cows: A case study. J. Dairy Sci. 101:10271–10282. https://doi.org/10.3168/jds.2018-14779.

van Pelt, M., G. De Jong, and R. Veerkamp. 2016. Changes in the genetic level and the effects of age at first calving and milk production on survival during the first lactation over the last 25 years. Animal 10:2043–2050. https://doi.org/10.1017/S1751731116001282.

ORCIDS

Wijbrand Ouweltjes ● https://orcid.org/0000-0001-5455-0110
Bart Ducro ● https://orcid.org/0000-0002-0866-0199
Yvette de Haas ● https://orcid.org/0000-0002-4331-4101
Claudia Kamphuis ● https://orcid.org/0000-0001-5552-036X