J. Dairy Sci. 103:8433–8442
https://doi.org/10.3168/jds.2020-18320

Comparison of methods for predicting cow composite somatic cell counts

Dorota Anglart,[1,2]* Charlotte Hallén-Sandgren,[1] Ulf Emanuelson,[2] and Lars Rönnegård[3,4]
[1]DeLaval International AB, PO Box 39, SE-147 21, Tumba, Sweden
[2]Swedish University of Agricultural Sciences, Department of Clinical Sciences, PO Box 7054, SE-750 07 Uppsala, Sweden
[3]School of Technology and Business Studies, Dalarna University, SE-791 88 Falun, Sweden
[4]Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, PO Box 7023, SE-750 07 Uppsala, Sweden

ABSTRACT

One of the most common and reliable ways of monitoring udder health and milk quality in dairy herds is by monthly cow composite somatic cell counts (CMSCC). However, such sampling can be time consuming, and more automated sampling tools entail extra costs. Machine learning methods for prediction have been widely investigated in mastitis detection research, and CMSCC is normally used as a predictor or gold standard in such models. Predicted CMSCC between samplings could supply important information and be used as an input for udder health decision-support tools. To our knowledge, methods to predict CMSCC are lacking. Our aim was to find a method to predict CMSCC by using regularly recorded quarter milk data such as milk flow or conductivity. The milk data were collected at the quarter level for 8 wk when milking 372 Holstein-Friesian cows, resulting in a data set of 30,734 records with information on 87 variables. The cows were milked in an automatic milking rotary and sampled once weekly to obtain CMSCC values. The machine learning methods chosen for evaluation were the generalized additive model (GAM), random forest, and multilayer perceptron (MLP). For each method, 4 models with different predictor variable setups were evaluated: models based on 7-d lagged or 3-d lagged records before the CMSCC sampling and additionally for each setup but removing cow number as a predictor variable (which captures indirect information regarding cows' overall level of CMSCC based on previous samplings). The methods were evaluated by a 5-fold cross validation and predictions on future data using models with the 4 different variable setups. The results indicated that GAM was the superior model, although MLP was equally good when fewer data were used. Information regarding the cows' level of previous CMSCC was shown to be important for prediction, lowering prediction error in both GAM and MLP. We conclude that the use of GAM or MLP for CMSCC prediction is promising.

Key words: generalized additive model, multilayer perceptron, random forest, udder health

INTRODUCTION

Udder health monitoring is one of the most important tasks on a dairy farm. Udder health-related issues affect many things, such as animal welfare and farm economy (Halasa et al., 2007; Hogeveen et al., 2011), milk quality (Forsbäck et al., 2009), and production volume (Dürr et al., 2008). The primary way to monitor udder health on farms is by counting somatic cells in milk; that is, the presence of white blood cells indicating an inflammatory process (Pyörälä, 2003; International Dairy Federation, 2013). The SCC reflects not only udder health status but also milk quality (Schukken et al., 2003). Accordingly, many countries apply penalties if milk with elevated SCC is delivered to dairies. Sampling individual cows for cow composite somatic cell count (**CMSCC**) is a way to monitor herd status and is also useful in identifying the cows influencing the bulk milk SCC. Sampling is normally done monthly, but there is a risk of udder health misclassification because normal variations in CMSCC can affect the results (Quist et al., 2008). One reason for increased CMSCC variation between days could be an IMI (Chagunda et al., 2006) causing clinical or subclinical mastitis. Such information is missed if the sampling rate is low. Screening more frequently for CMSCC at the herd level could improve the ability to manage udder health (Sørensen et al., 2016). However, frequent CMSCC sampling can be time consuming, and the more automated sampling methods currently available entail additional costs. These factors together often limit the sampling frequency (Pyörälä, 2003).

Machine learning (**ML**) prediction models have been extensively explored in several areas because they are well-suited for handling the "big data" analyses required for accurate predictions and complex relationships be-

Received February 6, 2020.
Accepted April 9, 2020.
*Corresponding author: dorota.anglart@delaval.com

tween variables. Evaluation of ML methods is common in dairy health management (Lokhorst et al., 2019), specifically in mastitis detection research (Rutten et al., 2013). In recent decades, various mastitis prediction methods have been considered. However, CMSCC has mainly been used as input for the models evaluated for mastitis detection (e.g., Kamphuis et al., 2010; Jensen et al., 2016; Sørensen et al., 2016) and as the gold standard (i.e., the "true" reference) for mastitis cases (e.g., Chagunda et al., 2006; Cavero et al., 2008). Applications capable of predicting CMSCC between regular samplings could reduce workload and costs and, more importantly, provide additional information for udder health decision-support tools. Prediction of CMSCC cut-off levels has been investigated to some extent, mainly by comparing ML methods (Mammadova and Keskin, 2015; Sitkowska et al., 2017; Ebrahimi et al., 2019). However, the usefulness of cut-off levels has been questioned (Ruegg, 2003) and any cut-off level will have advantages and disadvantages (Schukken et al., 2003). For example, Bach et al. (2019) demonstrated that blanket cut-off points such as 200,000 cells/mL are inefficient as a diagnostic tool regardless of composite or quarter SCC evaluation. Hence, models predicting actual CMSCC values should be more useful as udder health decision-support tools, because predictions of CMSCC values make it possible to detect increases in CMSCC or recovery from elevated CMSCC.

To our knowledge, no previous studies have attempted to evaluate methods for predicting CMSCC values. The objective of this study was therefore to find a method for CMSCC prediction using regularly recorded quarter milk data as the model input. This was done by comparing the CMSCC prediction performance of 3 methods: the generalized additive model (**GAM**), random forest (**RF**), and multilayer perceptron (**MLP**).

MATERIALS AND METHODS

Data Collection

The data were collected during an 8-wk trial period from an agricultural college dairy farm in Germany. The 372 Holstein-Friesian cows were milked twice daily in an automatic milking rotary (24-unit platform with 5 robotic arms; DeLaval International AB, Tumba, Sweden). The bulk tank SCC ranged between 160,000 and 200,000 cells/mL during the year, and average milk production was 11,500 kg/cow per year. The cows were kept in a loose housing system and fed a TMR; calving was year-round.

Animal information, such as DIM and parity, was extracted from the herd management system, together with information from each milking during the 8 wk (henceforth, "milking data"). Milking data at the quarter level comprised conductivity (mS/cm), blood in milk (mg/kg), milk yield (g), expected milk yield (g), mean and peak milk flow (g/min), cups kicked off during milking (yes/no), and incompletely (yes/no) and not-milked (yes/no) quarters. Milking data at the cow composite level comprised milking duration (min), milking unit number, mastitis detection index (**MDi**; unitless, an index giving the likelihood of mastitis by incorporating different phases of conductivity during milking together with blood in milk), and udder counters (a counter triggered by MDi >1.4).

Sampling for CMSCC was done once weekly during the afternoon milkings throughout the 8 trial weeks. Each cow's foremilk was visually inspected by strip milking before the milking cups were attached. The cows present at the milking were sampled for CMSCC by a milk sampler (milk meter MM6, DeLaval International AB) attached to each milking unit, collecting a representative sample form each cow separately throughout the milking. Bronopol (Fisher Scientific, Schwerte, Germany) and Kathon (Fisher Scientific) were used as preservatives for the milk samples. Samples were stored at room temperature (15–20°C) during the milking session and transported to the laboratory (Jena, Germany) immediately after milking to be analyzed within 24 h. The samples were analyzed for CMSCC according to ISO/IEC (2005) in the laboratory using a Fossomatic 7, DC 600 system (Foss, Hillerød, Denmark). During the trial period, the distribution of the number of cows over parities 1, 2, and ≥3 was 132, 101, and 139, respectively, and the average DIM on the testing day was 191. The geometric average CMSCC was 68,000 cells/mL and the interquartile range was 26,000 to 125,000 cells/mL.

Data Preparation

Information such as parity and DIM together with data collected from each milking during the 8 wk of the trial was extracted from the database of the herd management system. The raw milking data contained 30,734 records from 372 cows, where each record contained information on 87 variables. The data cleaning steps are summarized in Table 1 and were performed accordingly: data from cows present in the database but not included in the weekly CMSCC sampling were removed from the milking data. The cows were categorized into parity 1, 2, and ≥3, and all milking events for cows during the first week of lactation were removed. Mean and peak milk flow values classified as outliers according to boxplots (i.e., outside 1.5 × interquartile

Table 1. Steps in preparing the data set before model development

Steps	Obs.[1]	Var.[2]	n cows[3]
Raw milk data	30,734	87	372
Cleaned milking data	28,634	87	372
Creating day lag records as variables	2,384	934	372
Cow composite SCC sample data	2,384	1	372
Merging the milking data with the sample data	2,384	935	372
Removing cows lacking complete setup of explanatory variables 14 milking sessions before the cow composite SCC sampling event	1,758	935	319
Variable screening	1,758	905	319
Removing variables $P > 0.001$	1,758	268	319
Adjusting for multiple comparison using Bonferroni correction; removing variables $P > 0.010$	1,758	158	319
Test for multicollinearity; removing variables with variance inflation factor >8	1,758	105	319
Variables with lagged records included as predictor variables in models for method comparison	1,758	84	319
Creating dummy variables	1,758	404	319

[1]Number of available observations.

[2]Number of variables retained.

[3]Number of unique remaining cows.

range above the upper and below the lower quartiles) were removed. Quarter conductivity values <3 and >10 mS/cm were removed with support from plots of conductivity data from each quarter, to balance what is considered to be biologically reasonable with keeping important information regarding variation.

From the cleaned milking data, new variables combining quarter variables were created and named with a suffix as follows: max—the highest value of a variable within cow and milking session; mean—the arithmetic mean of a variable within cow and milking session; diff—the difference in a variable between quarters; var—the variance of a variable within cow between quarters; diff.milkings—the difference in the milk yield variable between milking sessions; and min—the lowest quarter value of a variable within cow and milking session. Milk yield-associated variables such as quarter milk yield and composite milk yield were removed before variable screening, because milk yield was considered an intervening variable on the causal path between the other predictor variables and CMSCC. Variables not used for analysis, such as cows' birth date, transponder ID, and so forth, were also removed from the milking data set. Past-period records (lags) were added as lagged variables to all variables for 7 d (i.e., 14 milking sessions; thus, the milking of the CMSCC sample event corresponded to milking session 0 and so on) for each of the 8 CMSCC sampling events. Thus, the complete and cleaned data set contained 2,384 observations of 372 milking cows with 934 potential explanatory variables available (i.e., 840 variables with day lag records and 94 variables for milking session 0).

The CMSCC values in the sample data set were divided by 1,000 and $\log_{10}$-transformed (henceforth, $\log_{10}$CMSCC) to achieve normally distributed residuals from the statistical analyses. The milking data containing the created variables with lag records were merged with the sample data. Finally, CMSCC observations that did not include a complete set of explanatory variables for 14 milking sessions before the CMSCC sampling event were removed, leaving 319 cows with 1,758 cow observations for variable screening.

Data Analysis

Variable screening was performed using a GAM (Hastie and Tibshirani, 1990). Hence, $\log_{10}$CMSCC was set as the response variable, and the potential confounders parity (factor), DIM (continuous covariate), and cow (random factor) were added to the screening model. Subsequently, all potential explanatory variables were added one by one. In the next step, Bonferroni correction was performed and variables with $P > 0.01$ were removed. The remaining variables were tested for multicollinearity using the variance inflation factor (**VIF**) according to Fox and Monette (1992). Variables with VIF >8 were removed, leaving 1,758 observations of 105 variables from 7 d before the CMSCC sample for analysis.

The modeling methods considered in evaluating the abilities of different ML methods to predict CMSCC were GAM, RF, and MLP. The response variable in each model was weekly $\log_{10}$CMSCC. Variables included in all models as predictors, for evaluating methods, were cow, DIM, parity, diverse conductivity variables, peakflow.min, MDi, and diff.milkings (Table 2). Thus, a data set containing 1,758 observations of 84 variables was used in the model development. The numerical variables, including the response variable $\log_{10}$CMSCC, were scaled; that is, normalized with a mean value of

Table 2. Predictor variables used in model development together with cow number, DIM, and parity

Variable	Definition	Milking session lag[1]
Conductivity	Quarter conductivity (mS/cm)	{1 4}
Conductivity.max	Highest conductivity (mS/cm) value (quarter) within cow and milking session	{0 1 3 4 6 7 8 9 10 11 13 14}
Conductivity.mean	All-quarter conductivity (mS/cm); arithmetic mean within cow and milking session	{0 4 9 11 14}
Conductivity.diff	Highest conductivity (mS/cm) value; lowest conductivity (mS/cm) value within cow and milking session	{0 1 2 3 4 5 6 7 8 9 10 11 12 13 14}
Conductivity.var	Conductivity (mS/cm) variance between quarters within cow and milking session	{0 1 2 3 4 5 6 7 8 9 10 11 12 13 14}
Diff.milkings	Deviation in quarter yield (kg) from previous corresponding milking session	{0}
Peak flow.min	Lowest (quarter) value (g/min) of the maximum milk flow within cow and milking session	{0 1 2 11 12 13}
MDi	Mastitis detection index based on conductivity and blood in milk	{0 1 2 3 4 5 6 7 8 9 10 11 12 13 14}

[1]Number of past-period lags in milking sessions before the composite SCC (CMSCC). Zero indicates the milking session of the CMSCC sampling event, whereas 14 indicates the 7 d before the CMSCC sampling event.

zero and standard deviation of 1; however, results (figures) are presented on the original scale. Four predictor variable setups were evaluated for each of the 3 modeling methods: data with 7-d lags (**7D**), data with 3-d lags (**3D**), and removing cow as predictor variable from both day lag variations. The 7D data were also used in tuning the hyperparameters (i.e., values set before the learning process) for RF and MLP. Furthermore, cow number was converted to dummy variables for these 2 methods.

ML Methods for Evaluation of Prediction Performance

The choice of methods for the evaluation was based on the properties of each main method; that is, regression (GAM), decision tree (RF), and artificial neural network (MLP). The methods were aligned with previously proposed approaches to predicting CMSCC cutoff levels (e.g., Panchal et al., 2016; Ebrahimi et al., 2019).

Generalized Additive Model

Generalized additive models are flexible additive regression models in which smooth functions allow the relationship between the response and predictor to be nonlinear (Hastie and Tibshirani, 1990). The GAM was fitted according to

$$y_i = \beta_0 + parity\beta_{LN} + DIM\beta_{DIM} + \alpha_{Cow} + \sum_{j=1}^{p} f_j \left[\!\left[(X)\right]\!\right]_{ij} + \varepsilon_i, \quad [1]$$

$$\alpha_{\text{Cow}} \sim N\left(0, \sigma^2_{\text{Cow}}\right), \ \varepsilon_i \sim N\left(0, \sigma^2_{\varepsilon}\right),$$

where y is the response variable, *parity* (factor), *DIM* (linear variable) and α_{Cow} (random factor) are potential confounders, $f_j(X)$ is the nonparametric spline functions of the potential nonlinear predictors, β symbolizes the regression coefficients, i is the ith observation, j is the jth variable, ij is the jth variable of the ith observation, p is the number of variables, ε is the error term, and σ is the variance.

To make predictions with GAM for cows not sampled on all sampling occasions (due to normal circumstances such as dry off and sickness), the estimated random effect of cow (i.e., α_{Cow}) in GAM was set to zero in cases in which cow was missing. Because the expectation of a random effect without any information is zero, we could make predictions for all cows regardless of whether a particular cow was sampled. The smoothing parameter estimation method used in all model variations was REML. The "mgcv" package in R was used for GAM model development (R Development Core Team, 2018).

Random Forest

Random forest is a tree-based ensemble learning method that can be used for both classification and regression problems. The method combines bagging with multiple random decision trees; that is, each decision tree is trained on a different data sample, with sampling done through replacement, which prevents overfitting (Breiman, 2001).

In RF model development, the "randomForest" package in R was used (Wiener and Liaw, 2002). Initially, several RF regression models were fitted using the 7D data with predetermined hyperparameters; that is, parameters assigned before training the model. The number of decision trees considered in these models was 250, 500, 750, 1,000, 1,250, 1,500, and 2,000. Analyses

of mean squared error (**MSe**) plots indicated that lowest MSe was obtained in the model using 1,000 trees. The optimal number of variables selected in each tree (mtry) was determined using the "tune RF" function provided in the randomForest package (Wiener and Liaw, 2002). The MSe was very similar between the models (results not shown); hence, the default mtry was used together with 1,000 trees for the 4 predictor variable setups in the final analysis

Multilayer Perceptron

The MLP is a network of linear classifiers containing several perceptrons organized into layers. The perceptron was introduced by Rosenblatt (1958) for use when multiple value inputs are estimated with a single output as the result. Each layer can be described as follows:

$$y = \phi\left(\sum_{i=1}^{n} w_i x_i + b\right) = \phi\left(\mathbf{w^T x} + b\right), \quad [2]$$

where y is the intermediate basis function (response variable in last layer), $\mathbf{w}$ denotes the estimated vector of weights (w), $\mathbf{w^T}$ is the transposed vector $\mathbf{w}$, $\mathbf{x}$ is the vector of input or outputs from previous layer, b is the bias, ϕ is the chosen activation function, i is the ith observation, and n is the number of nodes in the previous layer (Bishop, 1996; Haykin, 2009).

The MLP was constructed with Keras for R (Chollet, 2017), using the Keras model sequentially. Two layers were applied in the model. The number of units in each layer was determined by running several models on the 7D data, divided into 80% training and 20% validation data by random sampling. In searching for the lowest validation MSe, 50 to 500 units were evaluated. The optimal choice was found to be 200 units in the first layer and 100 in the second. The output layer was constructed using one unit, because our model is a regression problem with a single response variable. Furthermore, the rectified linear activation function (relu) was used, being the default activation function for regression problems in Keras; relu was applied in first and second layers but not in the output layer. The option of having dropout between the layers was not used because the difference in the results was negligible.

To configure the learning process, model compilation was done according to the following steps. As optimizer, we chose ADAM (Kingma and Ba, 2014), a stochastic optimization method that works well even with little tuning of the hyperparameters. The loss function, being the object that the model will try to minimize (i.e., showing a difference between the observed and predicted values), was set to MSe. For model training, the default number of times for full forward and backward propagation was used (i.e., epoch = 10). A batch size of 64 was chosen because it gave slightly lower error rates than did the default batch size of 32.

Model Evaluation

Cross Validation. Model performance was evaluated using 5-fold cross validation (**5-CV**), estimating the test error associated with each of the 3 methods. Data were divided by random sampling, where 80% of the data were used for model training and remaining 20% was used for model test; that is, predicting new values. The predicted and observed values of $\log_{10}$C-MSCC were compared and MSe was analyzed for each method. In Keras, model weights are initiated randomly for each subset in the 5-CV, so the MSe for the MLP was calculated as the mean MSe over ten 5-CV runs to obtain more consistent results.

Predictions on Future Data. To make predictions of future CMSCC (predictions of future data, **PFD**), the data set was divided according to the CMSCC sampling events. All data associated with milk sampling events 1 to 6 were used for training, and data associated with milk sampling events 7 and 8 were used for testing in each of the 3 methods evaluated. Models for each method were thus trained on approximately 75% of the data and predictions were made on the remaining 25%. The MSe was calculated for the prediction of each model within methods. The MSe for MLP was calculated as the mean MSe over 10 prediction runs. All statistical procedures were carried out using R (R Development Core Team, 2018).

RESULTS

In the cleaned data set used for model development, 319 cows with CMSCC observations, including a complete set of explanatory variables for 7 d before the CMSCC sampling event, were distributed over parities 1, 2, and ≥3 (119, 90, and 110 cows, respectively). The average DIM on the testing day was 176. The geometric average CMSCC was 61,000 cells/mL and the interquartile range was 26,000 to 112,000 cells/mL.

Cross Validation

The lowest MSe (i.e., 0.09) from the 5-CV of the 3 methods was found for GAM with the 7D predictor variable set as model input (Table 3). Results of the models with the 3D predictor variable also displayed low MSe (0.10) for GAM and MLP. Models including

Table 3. Results of 5-fold cross validation of the generalized additive model (GAM), random forest (RF), and multilayer perceptron (MLP)[1]

	Predictor variable setup[2]			
Method	7D	3D	7D without cow	3D without cow
GAM	0.09	0.10	0.20	0.19
RF	0.16	0.15	0.16	0.16
MLP	0.12	0.10	0.17	0.17

[1]The mean squared error is shown for each method with the 4 predictor variable setups used in the method evaluation.

[2]Where 7D = data set with 7-d lags and 3D = data set with 3-d lags, with and without cow as a predictor variable.

the predictor variable cow (i.e., the cow number connected to the CMSCC value) were the most favorable for GAM and MLP, whereas RF was almost unaffected. Thus, when cow was excluded as the predictor variable from the first 2 methods, MSe increased. The GAM accounted for the largest increase, as MSe increased to 0.20, the highest MSe of all comparisons tested, just by removing cow as the predictor in the 7D predictor variable setup.

The MSe for RF was only weakly affected by changes in the 4 predictor variable setups. However, the lowest MSe (0.16) among the models without cow as the predictor variable was found for RF. Overall, differences in MSe between predictor variable setups within methods were greater than the differences between methods, when compared across the same predictor variable setup for each method.

Predictions on Future Data

Comparison of PFD between methods showed that the MSe was the lowest (i.e., 0.09) for GAM models, as long as cow was included as the predictor variable (Table 4). The difference in MSe between GAM and MLP was small, whereas RF had the highest MSe in models including cow as the predictor variable. Removing cow from the models increased the MSe for all methods, for GAM more than for MLP and RF. The differences in MSe between methods were very small for the models from which cow was removed as the predictor variable. All results of the future data predictions for the 4 predictor variable setups can be found in Table 4.

Figure 1 shows the PFD results for the observed versus predicted values of $\log_{10}$CMSCC for each method, estimated by models fitted with the 3D data. Compared with the other methods, GAM (Figure 1a) displayed the most linear pattern, in agreement with the slightly lower MSe for GAM obtained for the future predictions. In addition, both GAM and MLP displayed a more balanced pattern across the predicted axis than does RF. In Figure 1b, PFD by RF displayed predicted $\log_{10}$CMSCC values clustered around 1.5 on the x-axis. This shows that the method overestimates low $\log_{10}$CMSCC values, since there are no predictions for $\log_{10}$CMSCC $\leq$1.4. All 3 methods underestimated values >2.5.

DISCUSSION

The objective of this study was to find the most accurate of 3 ML methods comparing CMSCC prediction. The evaluation was carried out by comparing GAM, RF, and MLP based on 5-CV and PFD. Our results indicated that the best prediction performance was achieved by GAM, in terms of both 5-CV and PFD. This was illustrated by the lowest value of MSe (i.e., 0.09) using the 7D predictor variable setup, and can be seen by comparing the observed and predicted CMSCC values (Figure 1).

The results indicated more disparity in MSe among the different predictor variable setups within methods than between methods. Thus, an equally low MSe (0.10) was obtained for MLP and GAM using the 3D predictor variable setup in the 5-CV. Hence, the average squared error of the predictions was small (0.1) compared with the variation in the observed values (variance of 1). In the study by Ankinakatte et al. (2013), GAM performed slightly better compared with an artificial neural network in detecting clinical mastitis, depending on the input variables used for each method. Despite the different target used by Ankinakatte et al. (2013), this is in agreement with our results, showing that predictor input variables did affect prediction performance. The variable scanning performed by GAM might have benefited the method somewhat, although most of the variables selected for model development were suggested in

Table 4. Results of the predictions on future data for the generalized additive model (GAM), random forest (RF), and multilayer perceptron (MLP)[1]

	Predictor variable setups[2]			
Method	7D	3D	7D without cow	3D without cow
GAM	0.09	0.09	0.18	0.17
RF	0.16	0.16	0.17	0.17
MLP	0.12	0.11	0.18	0.17

[1]The mean squared error is shown for each method with the 4 predictor variable setups used in the method evaluation.

[2]Where 7D = data set with 7-d lags and 3D = data set with 3-d lags, with and without cow as a predictor variable.

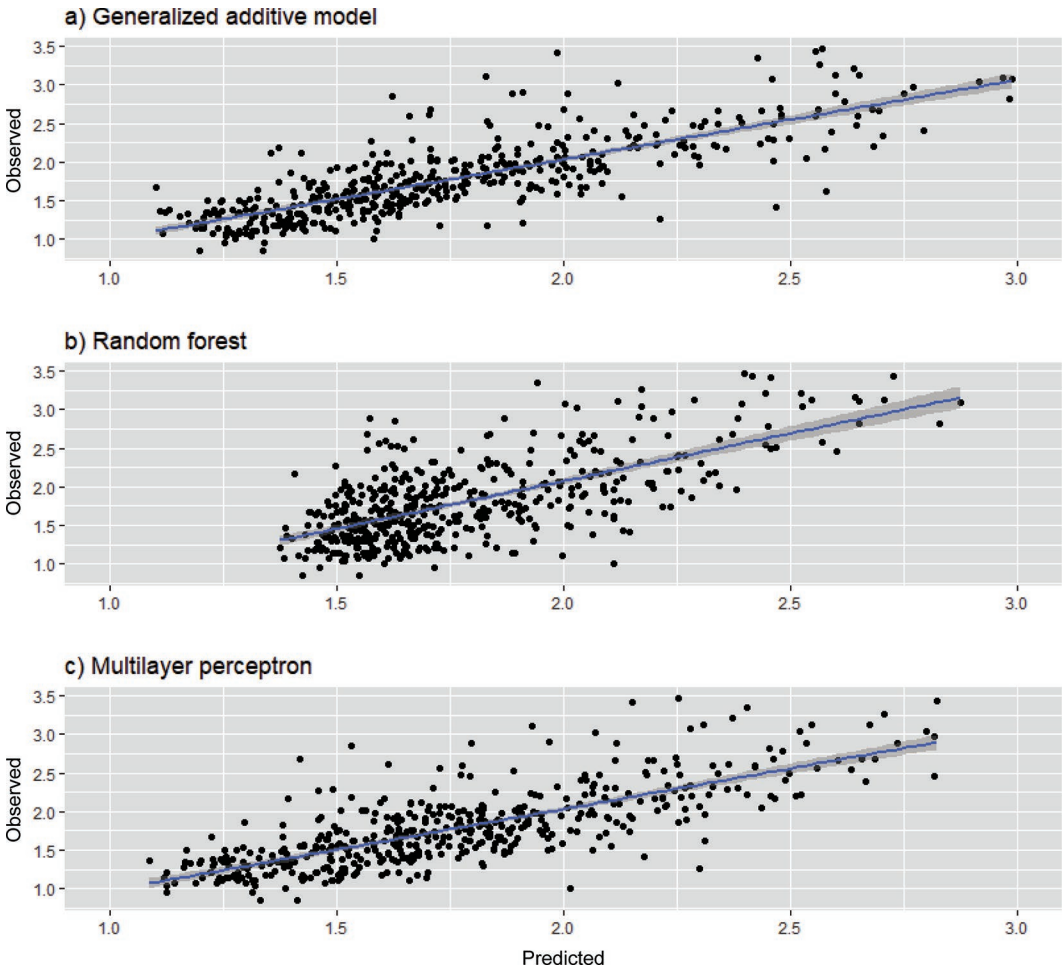

Figure 1. Observed versus predicted values of cow composite SCC divided by 1,000 on a $\log_{10}$ scale estimated by the generalized additive model (a), random forest (b), and multilayer perceptron (c) on future data, based on the combination of predictor variables for 3-d lags.

other studies predicting CMSCC cut-offs (Panchal et al., 2016; Sitkowska et al., 2017; Ebrahimi et al., 2019). This indicates that either GAM or MLP could be suitable for CMSCC prediction models, depending on how many data (i.e., information from previous milkings) are available. Additionally, the convergence speed of MLP was much higher than that of GAM (data not shown), which is a major advantage if the method is to be used for applications operating on real-time data.

Including cow number as a predictor variable, which captures information regarding the cows' overall level of CMSCC based on previous samplings, was found to be important for prediction performance by lowering the MSe, especially for GAM but also for MLP. The variable "cow" was also chosen as the variable with highest importance by the RF (results not shown). However, the MSe changed very little among the RF models when using different predictor variable setups as input. Variable importance is a measure of how much the variable affects the residual sum of squares within the training (Wiener and Liaw, 2002; Louppe et al., 2013) and an explanation for the discrepancy between variable importance and prediction MSe could be that high importance scores were caused by overfitting since several cows had a limited number of observations each. The RF for PFD also failed to predict $\log_{10}$CMSCC $\leq$1.4 (Figure 1b). When the RF grows trees, it uses the best split point from a randomly selected set of input variables. Using random input in regression problems could affect their performance (Breiman, 2001), possibly explaining the lower prediction performance of RF.

To our knowledge, only a few studies have evaluated methods for predicting CMSCC values as output. The SCC prediction results presented by Jędruś et al. (2012) are somewhat difficult to compare with our results due to the different scales used; that is, $\log_{10}$CMSCC and

SCC. However, there are studies investigating methods for predicting subclinical mastitis, which is often approached as a classification problem; that is, predicting CMSCC cut-off levels (Mammadova and Keskin, 2015; Sitkowska et al., 2017; Ebrahimi et al., 2019). Despite the differences in approach to the problem (i.e., regression versus classification), there are some similarities to our findings. For instance, different types of artificial neural network methods were suggested to be superior classifiers compared with decision tree methods (Mammadova and Keskin, 2015; Ebrahimi et al., 2019). This was confirmed by our results, which indicate that RF is the least suitable method due to its higher prediction MSe. Decision tree methods generally have lower prediction accuracy than other methods (Gareth et al., 2013), as confirmed by Sitkowska et al. (2017), who found that the decision tree model had a higher misclassification rate than did logistic regression.

Predictions of CMSCC between ordinary test samplings could improve udder health decision support in 2 ways, either by detecting deviations from the animals' baseline CMSCC or by detecting the recovery of individual cows with elevated CMSCC (Sørensen et al., 2016). Normal variability contributes greatly to the overall variability of SCC in a healthy cow's udder (Quist et al., 2008; Forsbäck et al., 2010; Nørstebø et al., 2019). Accordingly, to distinguish sick from healthy cows, CMSCC prediction methods need to be sensitive enough to accurately predict changes in CSMCC independent of the expected magnitude of the outcome. Our results indicate a low MSe for prediction and, additionally, the capacity to also predict low values of CMSCC (around 12,500 cells/mL) for 2 of the 3 models. This demonstrates the increased applicability of methods predicting CMSCC values compared with methods that predict CMSCC cut-off levels.

For practical reasons, test samplings of CMSCC occur monthly or even less frequently, although additional samplings using simpler methods (e.g., the California Mastitis Test) may be carried out if an IMI is suspected. Important information can be missed due to a low sampling frequency or deficient sampling accuracy. Generally, a single cell count measurement will not yield the same understanding as information from several measurements. Dalen et al. (2019) found that single measurements from an inline cell counter were outperformed in detecting subclinical mastitis by a 7-d rolling average. Predictions of CMSCC could be utilized as input for such models, contributing to the more accurate interpretation of the udder health of individual cows. Furthermore, predicted CMSCC values could be valuable as input for clinical mastitis prediction models, because the performance of clinical mastitis detection models is improved by adding inline CMSCC values (Kamphuis et al., 2008). Predictions of CMSCC between routine monthly samplings could save time and money by reducing the number of samples needed. For example, our results from the 5-CV indicate that by using 3 d of data from 80% of the cows in the herd as GAM input, the CMSCC of the remaining cows could be predicted with an MSe of 0.10.

Another possible application of the prediction results could be to reduce the number of required samples; for example, predicting every fifth CMSCC sample instead of actually sampling the cows. Furthermore, PFD indicates that by using historical information regarding the cows' milkings as model input, the next CMSCC samples can be predicted with a low MSe. The difference in our MSe results between 5-CV and PFD was small, which is likely a good indicator of how the methods would perform under real conditions, with 5-CV being an accepted method for model evaluation and PFD reflecting a "real life" situation with a limited amount of data on which to base future predictions.

We are aware that our study has limitations that should be considered when interpreting the results. Data for model development were obtained from one farm only, which contributed to an unbalanced distribution of CMSCC values for model training, and there may be farm-specific SCC patterns, as shown by Sørensen et al. (2016) for online cell counts. Thus, our prediction models would need to train on data from the farms for which predictions are to be used to make them valid. Our interpretation of the plots of observed versus predicted CMSCC is that predictions of lower $\log_{10}$CMSCC values (i.e., $\log_{10}$CMSCC ≤ 2, corresponding to 100,000 cells/mL) are more or less in accordance with observed values. This might reflect the small number of milk samples with high cell counts in the data set; that is, the cows were too healthy and the data set contained fewer observations of $\log_{10}$CMSCC ≥ 2. Although the present results can be indicative, predictions of CMSCC exceeding 100,000 cells/mL should be interpreted with caution.

CONCLUSIONS

The lowest prediction error was found for the GAM using data from the 7 preceding days of milkings, although there was no general disadvantage to using data from only the 3 preceding days. We suggest that information regarding cows' previous CMSCC should be used for model training to lower the prediction error. The use of GAM or MLP for CMSCC prediction

appears promising, although the results cannot be generalized broadly due to the limited data used in this study.

ACKNOWLEDGMENTS

Financial support for this study was provided by the Swedish Foundation for Strategic Research (SSF, Stockholm, Sweden). We thank the Kjell & Märta Beijer Foundation (Stockholm, Sweden) for funding Lars Rönnegård. The authors have not stated any conflicts of interest.

REFERENCES

Ankinakatte, S., E. Norberg, P. Løvendahl, D. Edwards, and S. Højsgaard. 2013. Predicting mastitis in dairy cows using neural networks and generalized additive models: A comparison. Comput. Electron. Agric. 99:1–6. https://doi.org/10.1016/j.compag.2013.08.024.

Bach, K. D., A. Sipka, and J. A. A. McArt. 2019. Case study: Evaluating quarter and composite milk sampling for detection of subclinical intramammary infections in dairy cattle. Prev. Vet. Med. 163:51–57. https://doi.org/10.1016/j.prevetmed.2018.12.013.

Bishop, C. M. 1996. Neural networks: A pattern recognition perspective. https://publications.aston.ac.uk/id/eprint/639/1/NCRG_96_001.pdf.

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

Cavero, D., K. H. Tölle, C. Henze, C. Buxadé, and J. Krieter. 2008. Mastitis detection in dairy cows by application of neural networks. Livest. Sci. 114:280–286. https://doi.org/10.1016/j.livsci.2007.05.012.

Chagunda, M. G. G., N. C. Friggens, M. D. Rasmussen, and T. Larsen. 2006. A model for detection of individual cow mastitis based on an indicator measured in milk. J. Dairy Sci. 89:2980–2998. https://doi.org/10.3168/jds.S0022-0302(06)72571-1.

Chollet, F. 2017. Keras (2015). Accessed May 10, 2019. https://keras.io.

Dalen, G., A. Rachah, H. Nørstebø, Y. H. Schukken, and O. Reksen. 2019. The detection of intramammary infections using online somatic cell counts. J. Dairy Sci. 102:5419–5429. https://doi.org/10.3168/jds.2018-15295.

Dürr, J. W., R. I. Cue, H. G. Monardes, J. Moro-Méndez, and K. M. Wade. 2008. Milk losses associated with somatic cell counts per breed, parity and stage of lactation in Canadian dairy cattle. Livest. Sci. 117:225–232. https://doi.org/10.1016/j.livsci.2007.12.004.

Ebrahimi, M., M. Mohammadi-Dehcheshmeh, E. Ebrahimie, and K. R. Petrovski. 2019. Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient-Boosted Trees outperform other models. Comput. Biol. Med. 114:103456. https://doi.org/10.1016/j.compbiomed.2019.103456.

Forsbäck, L., H. Lindmark-Månsson, A. Andrén, M. Åkerstedt, L. Andrée, and K. Svennersten-Sjaunja. 2010. Day-to-day variation in milk yield and milk composition at the udder-quarter level. J. Dairy Sci. 93:3569–3577. https://doi.org/10.3168/jds.2009-3015.

Forsbäck, L., H. Lindmark-Månsson, A. Åndrén, M. Akerstedt, and K. Svennersten-Sjaunja. 2009. Udder quarter milk composition at different levels of somatic cell count in cow composite milk. Animal 3:710–717. https://doi.org/10.1017/S1751731109004042.

Fox, J., and G. Monette. 1992. Generalized collinearity diagnostics. J. Am. Stat. Assoc. 87:178–183. https://doi.org/10.1080/01621459.1992.10475190.

Gareth, J., D. Witten, T. Hastie, and R. Tibshirani. 2013. An Introduction to Statistical Learning with Applications in R. 6th ed. Springer Science & Business Media, New York, NY.

Halasa, T., K. Huijps, O. Østerås, and H. Hogeveen. 2007. Economic effects of bovine mastitis and mastitis management: A review. Vet. Q. 29:18–31. https://doi.org/10.1080/01652176.2007.9695224.

Hastie, T., and R. Tibshirani. 1990. Generalized Additive Models. Chapman and Hall, New York, NY.

Haykin, S. 2009. Neural Networks and Learning Machines. 3rd ed. Pearson Education, Upper Saddle River, NJ.

Hogeveen, H., K. Huijps, and T. J. G. M. Lam. 2011. Economic aspects of mastitis: New developments. N. Z. Vet. J. 59:16–23. https://doi.org/10.1080/00480169.2011.547165.

International Dairy Federation. 2013. Guidelines for the use and interpretation of bovine milk somatic cell count in the dairy industry. Vol. 466. Bulletin: International Dairy Federation, Brussels, Belgium.

ISO. (International Organization for Standardization)/IEC (International Electrotechnical Commission). 2005. General requirements for the competence of testing and calibration laboratories. 2nd ed. Committee on Conformity Assessment, Geneva, Switzerland. 17025:2005.

Jędruś, A., P. Boniecki, J. Dach, and K. Pilarski. 2012. Neural predictive model in the estimation process of somatic cell counts in milk 3 Problem Solution. Pages 58–61 in Proc. 1st Int. Conf. Biologically Inspired Computation (BICA'12). World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, WI.

Jensen, D. B., H. Hogeveen, and A. De Vries. 2016. Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis. J. Dairy Sci. 99:7344–7361. https://doi.org/10.3168/jds.2015-10060.

Kamphuis, C., H. Mollenhorst, A. Feelders, D. Pietersma, and H. Hogeveen. 2010. Decision-tree induction to detect clinical mastitis with automatic milking. Comput. Electron. Agric. 70:60–68. https://doi.org/10.1016/j.compag.2009.08.012.

Kamphuis, C., R. Sherlock, J. Jago, G. Mein, and H. Hogeveen. 2008. Automatic detection of clinical mastitis is improved by in-line monitoring of somatic cell count. J. Dairy Sci. 91:4560–4570. https://doi.org/10.3168/jds.2008-1160.

Kingma, D. P., and J. Ba. 2014. Adam: A Method for Stochastic Optimization. Accessed Oct. 22, 2019. https://arxiv.org/abs/1412.6980.

Lokhorst, C., R. M. de Mol, and C. Kamphuis. 2019. Invited review: Big data in precision dairy farming. Animal 13:1519–1528. https://doi.org/10.1017/S1751731118003439.

Louppe, G., L. Wehenkel, A. Sutera, and P. Geurts. 2013. Understanding variable importances in forests of randomized trees. Pages 431–439 in Proceedings of the 26th International Conference on Neural Information Processing Systems, Volume 1 (NIPS'3). C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, ed. Curran Associates Inc., Red Hook, NY.

Mammadova, N. M., and I. Keskin. 2015. Application of neural network and adaptive neuro-fuzzy inference system to predict subclinical mastitis in dairy cattle. Indian J. Anim. Res. 49(OF):671–679. https://doi.org/10.18805/ijar.5581.

Nørstebø, H., G. Dalen, A. Rachah, B. Heringstad, A. C. Whist, A. Nødtvedt, and O. Reksen. 2019. Factors associated with milking-to-milking variability in somatic cell counts from healthy cows in an automatic milking system. Prev. Vet. Med. 172:104786. https://doi.org/10.1016/j.prevetmed.2019.104786.

Panchal, I., I. K. Sawhney, A. K. Sharma, and A. K. Dang. 2016. Classification of healthy and mastitis Murrah buffaloes by application of neural network models using yield and milk quality parameters. Comput. Electron. Agric. 127:242–248. https://doi.org/10.1016/j.compag.2016.06.015.

Pyörälä, S. 2003. Indicators of inflammation in the diagnosis of mastitis. Vet. Res. 34:565–578. https://doi.org/10.1051/vetres:2003026.

Quist, M. A., S. J. LeBlanc, K. J. Hand, D. Lazenby, F. Miglior, and D. F. Kelton. 2008. Milking-to-milking variability for milk yield, fat and protein percentage, and somatic cell count. J. Dairy Sci. 91:3412–3423. https://doi.org/10.3168/jds.2007-0184.

R Development Core Team. 2018. R: A Language and Environment for Statistical Computing. Accessed March 5, 2019. http://www.r-project.org.

Rosenblatt, F. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. Psychol. Rev. 65:386–408. https://doi.org/10.1037/h0042519.

Ruegg, P. L. 2003. Investigation of mastitis problems on farms. Vet. Clin. North Am. Food Anim. Pract. 19:47–73. https://doi.org/10.1016/S0749-0720(02)00078-6.

Rutten, C. J., A. G. J. Velthuis, W. Steeneveld, and H. Hogeveen. 2013. Invited review: Sensors to support health management on dairy farms. J. Dairy Sci. 96:1928–1952. https://doi.org/10.3168/jds.2012-6107.

Schukken, Y. H., D. J. Wilson, F. Welcome, L. Garrison-Tikofsky, and R. N. Gonzalez. 2003. Monitoring udder health and milk quality using somatic cell counts. Vet. Res. 34:579–596. https://doi.org/10.1051/vetres:2003028.

Sitkowska, B., D. Piwczyński, J. Aerts, M. Kolenda, and S. Özkaya. 2017. Detection of high levels of somatic cells in milk on farms equipped with an automatic milking system by decision trees technique. Turk. J. Vet. Anim. Sci. 41:532–540. https://doi.org/10.3906/vet-1607-78.

Sørensen, L. P., M. Bjerring, and P. Løvendahl. 2016. Monitoring individual cow udder health in automated milking systems using online somatic cell counts. J. Dairy Sci. 99:608–620. https://doi.org/10.3168/jds.2014-8823.

Wiener, A., and M. Liaw. 2002. Classification and regression by random forest. R News 2:18–22.

ORCIDS

Dorota Anglart https://orcid.org/0000-0003-1412-0057
Ulf Emanuelson https://orcid.org/0000-0001-7889-417X
Lars Rönnegård https://orcid.org/0000-0002-1057-5401