Computers and Electronics in Agriculture 180 (2021) 105904

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Predicting the milk yield curve of dairy cows in the subsequent lactation period using deep learning

Arno Liseune[a,*], Matthieu Salamone[b], Dirk Van den Poel[a], Bonifacius van Ranst[b], Miel Hostens[b]

[a] Faculty of Economics and Business Administration, Ghent University, Tweekerkenstraat 2, B-9000 Ghent, Belgium
[b] Faculty of Veterinary Medicine, Ghent University, Salisburylaan133, B-9820 Merelbeke, Belgium

A R T I C L E I N F O

Keywords:
Milk yield prediction
Animal monitoring
Deep learning

A B S T R A C T

Existing lactation models predict milk yields based on a fixed amount of observed milk production in early lactation. In contrast, this study proposes a model to predict the entire lactation curve of dairy cows by leveraging historical milk yield information observed in the preceding cycle. More specifically, we present a deep learning framework to encode the model inputs, predict the latent representation of the milk yield sequences and generate the corresponding lactation curves. Results show that the proposed framework outperforms the baseline models and that during the first 26 days of lactation, the model's predictions are more accurate than those of a state-of-the-art lactation model which is able to leverage the observed milk yields. As a result, the framework presented in this study allows farmers to increase their forecast horizon with respect to predicting its herd's total production and hence facilitates optimal herd management. Additionally, the model can be used to compare a cow's actual and expected milk yield over the entire course of the lactation cycle. This in turn can help to accelerate disease detection and enhance current animal monitoring systems. Finally, as the model incorporates the impact of health and reproduction events as well as herd management on the cow's productivity, future earnings and costs can be estimated more accurately.

1. Introduction

Forecasting milk yield is an important asset for dairy farmers as it can lead to improved decision making for optimal herd management (Dematawewa et al., 2007; Grzesiak et al., 2006). In particular, lactation models help to forecast the dairy farm's income (Ehrlich et al., 2011; Grzesiak et al., 2003b), determine the required nutrition and energy consumption (Murphy et al., 2014), optimize selection and culling decisions (Njubi et al., 2010; Sharma et al., 2006) and enhance animal monitoring systems (Adriaens et al., 2018; Silvestre et al., 2006).

Early lactation models were determined by mathematical functions describing the general milk yield pattern of homogeneous groups of animals (Brody et al., 1923). Lactation was modeled as a function of time with an increasing phase until a peak yield, followed by a more steady decline (Olori et al., 1999). Several mathematical functions have been widely used for predicting dairy milk yields, including an incomplete gamma, (Wood, 1967), a polynomial (Ali and Schaeffer, 1987), an exponential (Wilmink, 1987) and a Legendre polynomial (Kirkpatrick et al., 1994). Over time, the need to model individual variations from the mean lactation curve increased as more animal records were collected and farm management software improved (Macciotta et al., 2011). This led to several authors developing new models in order to fit more complex shapes and to include more input features (Murphy et al., 2018). For example, Grzesiak et al. (2003a) presented a multivariate regression model (MLR) that in addition to days in milk, also used test-day (TD) records, month of calving and the percentage of Holstein-Friesian (HF) genes as features to predict the 305d milk yield. Vasconcelos et al. (2004) and Macciotta et al. (2002) proposed autoregressive (AR) models in order to predict a milk yield based on a sequence of preceding TD records. Græsbøll et al. (2016) on the other hand presented a robust prediction model for cow level milk yield using lactation curves with reduced number of parameters, which is useful in case of sparse data.

Later, several artificial neural networks (ANN) have been proposed

* Corresponding author at: Tweekerkenstraat 2, B-9000 Ghent, Belgium.
E-mail addresses: Arno.Liseune@ugent.be (A. Liseune), Matthieu.Salamone@ugent.be (M. Salamone), Dirk.VandenPoel@ugent.be (D. Van den Poel), Bonifacius.vanRanst@ugent.be (B. van Ranst), Miel.Hostens@ugent.be (M. Hostens).

https://doi.org/10.1016/j.compag.2020.105904
Received 1 September 2020; Received in revised form 13 November 2020; Accepted 16 November 2020
Available online 10 December 2020

to predict milk yield. Lacroix et al. (1995) trained the first successful multilayer perceptron model (MLP) to predict the 305d yield based on 16 variables. In subsequent studies, this model was improved by applying more sophisticated data preprocessing techniques (Lacroix et al., 1997) and by training multiple networks each assigned to make specific predictions (Salehi et al., 1998). Furthermore, MLPs have been used to predict the 305d milk yield (Grzesiak et al., 2003a; Gorgulu, 2012), the 305d milk yield of the first lactation (Sharma et al., 2006, 2007; Njubi et al., 2010), the daily milk yield (Grzesiak et al., 2006; Torres et al., 2005) and dairy herd's total production (Murphy et al., 2014; Sanzogni and Kerr, 2001). In contrast to MLPs, recent research has shown that convolutional neural networks (CNN), most commonly applied for processing image data, can also be of great value for time series analysis (Zhao et al., 2017; Zheng et al., 2014). This was also shown by Liseune et al. (2020) who used a sequential autoencoder (SAE) to interpolate as well as predict missing milk yields along the entire lactation cycle by leveraging a latent representation of all the information available in the lactation cycle.

With the advent of individual curve fitting models, animal monitoring systems improved significantly. More specifically, by comparing a cow's expected and actual milk yield, diseases such as mastitis and ketosis could be detected more accurately (Grzesiak et al., 2003a; Adriaens et al., 2018). Such early detection systems are very valuable for the farmer since there can be a lot of costs associated with diseases of dairy cattle, e.g. lower production, discarded milk, treatment and culling or death (Wilson et al., 2004; Gröhn et al., 2004). Furthermore, the knowledge of the expected lactation curve made it easier to assess the impact of different treatments (Tekerli et al., 2000). However, in order to infer the expected lactation curve to which the actual milk yield can be compared, an initial number of milk yields recorded in early lactation is generally required. As a result, a reliable reference in early lactation is often unavailable which makes herd management as well as health monitoring of dairy cows particularly difficult in the period immediately after calving.

In this research, we propose a novel methodology to predict the entire lactation curve of a cow. More specifically, this paper contributes to previous research in multiple ways. Firstly, we present a model that predicts a lactation curve by using the sequence of milk yields generated in the preceding cycle. Moreover, instead of using the raw sequence of milk yields, the corresponding latent representation is used in order to disentangle the sequential information and to reduce the feature dimensionality. Secondly, we formulate a framework that models the impact of animal and herd Key Performance Indicators (KPI), lactation number and the sequence of health and reproduction events the cow encountered during the preceding cycle on the milk production. Finally, a new prediction approach is presented that generates the entire lactation curve non-sequentially. In particular, an MLP is used to generate the curve's latent encoding which is subsequently converted back into its corresponding milk yield sequence. The predictions obtained by the proposed model can be used to calculate the milk losses immediately after calving and hence support animal monitoring systems. In addition, the framework enables farmers to increase their forecast horizon with respect to the farm's future profitability.

2. Materials and methods

2.1. Data

The data were obtained from 104 different farms between 2013 and 2019. Every farm was equipped with a Herd Management Systems (HMS) that records milkmeter information and lifetime event records such as calving, heat, pregnancy and death. The HMS data sources were streamed and standardized using a cloud-based dairy analysis application (www.mmmooogle.com). In total, 59,122 lactations of 35,133 cows were collected, with an average of 216 recorded daily milk yields per lactation. Furthermore, 304,742 recordings of 13 unique lifetime events were collected. Based on the milk and event recordings, several animal and herd KPIs were calculated. In addition to KPIs summarizing the milk production, a score was developed to quantify the recording quality of the lactation cycle (Liseune et al., 2020). Averaging this score over all the lactations of a specific herd resulted in the herd's average sequence quality score. Finally, by averaging the milk yields of the entire herd per parity, the average lactation curve per herd and per parity was obtained. The final dataset was constructed by extracting every available pair of consecutive lactation cycles. This resulted in a total of 23745 observations, with each observation's dependent variable comprising the sequence of milk yields generated in the predicted lactation cycle. The features included the sequence of milk yields, animal KPIs and events generated in the preceding cycle together with the lactation number, herd KPIs and the herd's average lactation curve corresponding to the predicted lactation cycle. Milk yields and events recorded after 305 days in lactation were removed and days on which no events occurred were represented by a special `PAD` token. Events that did occur in the validation or test set but not in the training set were labeled as `UNKNOWN` and were considered as a rare but unknown event. The animal and herd KPIs were normalized between 0 and 1 and missing values were imputed with the mean value of the variable. From the entire collection of observations containing no missing milk yields in the predicted period, 2000 randomly sampled observations were assigned to the validation set and 2000 randomly sampled observations were assigned to the test set. The remaining observations with complete information in the predicted period were assigned to the training set together with the collection of observations containing missing milk yields in the predicted lactation cycle. An overview of all the variables used in this study is given in Table 1.

2.2. Feature extraction

2.2.1. Convolutional neural network

In general, a CNN's architecture exists out of a sequence of blocks, with each block typically comprising a convolutional layer, followed by a non-linear activation function for feature extraction. Pooling layers on the other hand are specifically designed for reducing the dimensionality of the hidden representation and making the network invariant to small translations in the input by obtaining the most prominent features. Fig. 1 gives an example of how a CNN with one convolutional block is applied on a sequence with S time steps and N features at each time step. A kernel of length $K < S$ and width N is slided over the entire input sequence and the dot products between the entries of the kernel and the input at any position are calculated. By sliding more than one kernel over the input sequence, multiple one-dimensional sequences are produced, with the elements of the sequences containing the response values of the corresponding kernels at every time step. Subsequently, an element-wise non-linear activation function is applied, followed by a pooling layer which summarizes the feature response in a certain time window. Finally, the output of the pooling layer is flattened and results in a vector that represents all the features extracted by every kernel at every time step and hence can be used for upstream tasks.

In this study, a similar architecture was applied on the sequence comprising the M last reproduction and health events encountered by the cow in the preceding lactation cycle. Before feeding the events directly to the CNN however, an embedding matrix of size 14 x k was used to convert each possible event occuring in the sequence into its corresponding numeric vector. Hence, the sequence of M events was converted to a M x k sequence with each time step of the sequence containing the corresponding event's embedding. In order to find the optimal event representations, the values of the embedding matrix were considered as network parameters and were updated during training. The sequences of embedding vectors were passed to a CNN of which each block consisted out of a linear transformation followed by a batch normalization layer. This layer normalizes the hidden activations and supports faster training as well as regularization (IIoffe and Szegedyoffe

Table 1
Variables used in this study. Lactation Cycle = the cycle from which the data was obtained. Dimension = the number of features belonging to the feature group and the number of time steps at which the features were measured.

Variable Group	Lactation Cycle	Dimension	Variable Name
Milk Yields	Preceding	1 x 305	
			Preceding Milk Yield
Milk Yields	Predicted	1 x 305	
			Predicted Milk Yield
Herd Yields	Predicted	1 x 305	
			Avg Milk Yield Per Herd Per Parity
Herd KPIs	Predicted	10 x 1	
			Avg 21d Milk
			Avg 75d Milk
			Avg 305d Milk
			Avg Milk
			Avg Days Dry
			Avg Days Open
			Avg Days Pregnant
			Avg Days In Milk
			Avg Calving Interval
			Avg Quality Sequence
Animal KPIs	Preceding	13 x 1	
			Age At First Calving
			Age At First Insemination
			Days Pregnant
			Days In Milk
			Minimum Milk Yield
			Maximum Milk Yield
			Total Milk Yield
			305d Milk Yield
			75d Milk Yield
			21d Milk Yield
			Avg Milk
			Std Milk
			Quality Sequence
Events	Preceding	1 x 305	
			Mastitis
			Abort
			Breeding
			Stop Breeding
			Pregnancy Negative
			Pregnancy Positive
			Calving
			Disease
			Died
			Heat
			Cull
			Dryoff
			PAD
			UNKNOWN
Parity	Predicted	1 x 1	
			Lactation Number

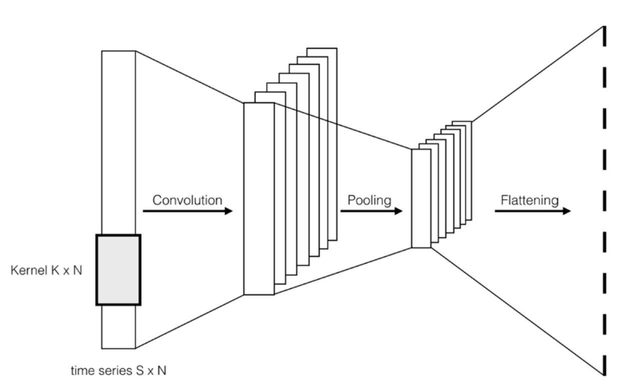


Fig. 1. Convolutional neural netowork for time series (Liseune et al., 2020).

and Szegedy, 2015). More specifically, if $\mathbf{X}$ represents a batch of hidden activations of a certain layer, then the normalization can be obtained as follows:

$$\widehat{\mathbf{X}} = \frac{\mathbf{X} - \mu}{\sigma}$$

with μ and σ being the activation's means and standard deviations respectively. The output of the batch normalization layer can then be calculated as follows:

$$y = \gamma\widehat{\mathbf{X}} + \beta$$

with γ and β being trainable parameters. Subsequently, a Leaky ReLU activation function was applied and is defined as follows:

$$f = max(\alpha\boldsymbol{x}, \boldsymbol{x})$$

with α often being set at a small constant value. As a result, the Leaky ReLU function will produce positive gradients for its entire range and hence facilitates gradient-based optimization. In addition, a dropout layer was applied to the output of the non-linear activation. With this mechanism, each neuron is dropped from the network with a certain probability and hence enforces the neurons to perform well, regardless of which other neurons are present in the network (Hinton et al., 2012). Dimensionality reduction of the hidden representation was obtained by applying max pooling layers which extract the most activated presence of a feature in a specific time interval. The last convolutional block's output was flattened and resulted in the static feature representation of the sequence of events.

2.2.2. *SAE encoder*

The sequential autoencoder (SAE), as proposed by Liseune et al. (2020), is an artificial neural network (ANN) specifically designed to infer missing milk yields along the lactation curve. More specifically, the SAE comprises a CNN which is used to transform the lactation cycle's sequential information into an extensive set of time-dependent features. A neural autoencoder then extracts a latent representation from the lactation curve and uses this encoding to get a reconstruction of the input features. Finally, a deconvolutional neural network (DNN) converts the reconstructed features back into the corresponding sequence of milk yields. As a result, the SAE is able to infer missing milk yields by making use of all the information available in a lactation cycle, irrespective of the length of the time intervals between the different observations. In Fig. 2, a schematic overview of the SAE is given. The SAE's encoder, which comprises the CNN and the autoencoder's compression side, takes an incomplete milk yield curve as input and extracts a latent representation from this curve. Subsequently, the autoencoder's reconstruction side and DNN (i.e., the SAE's decoder) convert the latent representation back into the reconstruction of the milk yield curve. In this study, the SAE was used to impute missing values along the milk yield curves of the preceding lactation cycles. The reconstructed lactation curve however is large in dimensionality and needs to be processed sequentially. Hence, instead of applying the entire SAE's architecture, the output of the SAE's encoder was used. This output comprises the latent representation of the lactation curve and hence entails its most prominent traits in a low-dimensional vector. Likewise, the SAE's encoder was applied to convert the herd's average lactation curve per parity into its low-dimensional representation.

2.3. *Milk yield prediction*

In order to predict the lactation curve of the predicted lactation cycle, a standard MLP was used. In this study, each hidden layer comprised a linear transformation followed by a batch normalization layer, a Leaky ReLU function and a dropout layer. The MLP's inputs comprised the latent representation of the preceding and the herd's

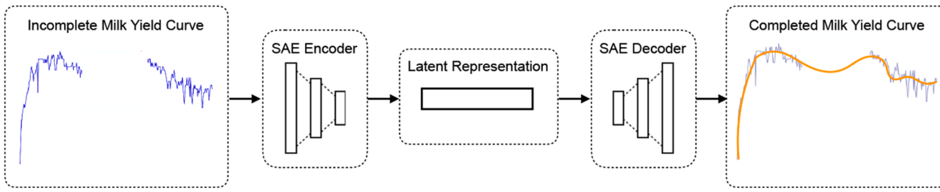


Fig. 2. Schematic overview of the SAE.

average milk yield sequence as well as the processed sequence of health and reproduction events, the animal and herd KPIs and the lactation number. Instead of predicting the entire sequence of milk yields however, the MLP's output was constrained to be of the same dimensionality as that of the latent encoding generated by the SAE's encoder. This vector was then fed to the SAE's decoder, which generated the reconstruction of the corresponding lactation curve. As a result, the MLP was trained to predict the lactation curve's latent encoding, rather than the entire sequence which would be time-consuming and prone to overfitting. The combination of all the different model components to forecast the lactation curve resulted in the Subsequent Lactation Milk Yield Predictor (SLMYP). A schematic overview of the SLMYP is depicted by Fig. 3.

2.4. Training

The entire SLMYP model was trained by making use of the backpropagation algorithm, firstly introduced by Rumelhart et al. (1986). More specifically, the inputs were first propagated through the entire network to produce a lactation curve. A loss between the model's predicted and true lactation curve was then calculated and the gradient of the loss function was propagated backward through the network. The weights were then updated by applying the Adam gradient-based optimization algorithm (Kingma and Ba, 2014). In order to avoid overfitting, an early stopping procedure was applied in which the model was trained until its performance on the validation set started to degrade. Furthermore, as neural networks are typically characterized by a large range of possible configurations, a Bayesian optimization procedure was applied to find the optimal hyperparameter setting. More specifically, once the early stopping procedure was terminated, the next configuration was determined by a trade-off between the exploration of a parameter setting with uncertain results against the exploitation of a point in parameter space with high model performance. In addition to the hyperparameters defining the CNN's and MLP's model architecture, the inclusion of all the model inputs except for the preceding lactation curve were also set as hyperparameter in order to obtain only those features with a significant predictive power. Furthermore, a boolean hyperparameter was included that determined whether the data should be balanced with respect to the lactation number. In the case of data balancing, observations with rare lactation numbers corresponding to the predicted lactation cycles were upsampled during training such that each training batch consisted out of an equal number of observations per lactation number. Finally, each model was trained with one out of three possible loss functions which weighted each daily prediction in the predicted period differently. A uniform loss assigned the same weight to each daily prediction such that the resulting model tried to fit each milk yield of the lactation curve equally well. In order to solely focus on generating predictions during the period for which current lactation models generally lack predictions, a step loss was used that assigned a uniform weight to the first 30 predictions of the predicted period while disregarding later days. This resulted in a model that was trained to approximate milk yields in early lactation as best as possible, while ignoring the predictive accuracy in the period after early lactation. Finally, a logarithmic loss assigned weights to every milk yield in the

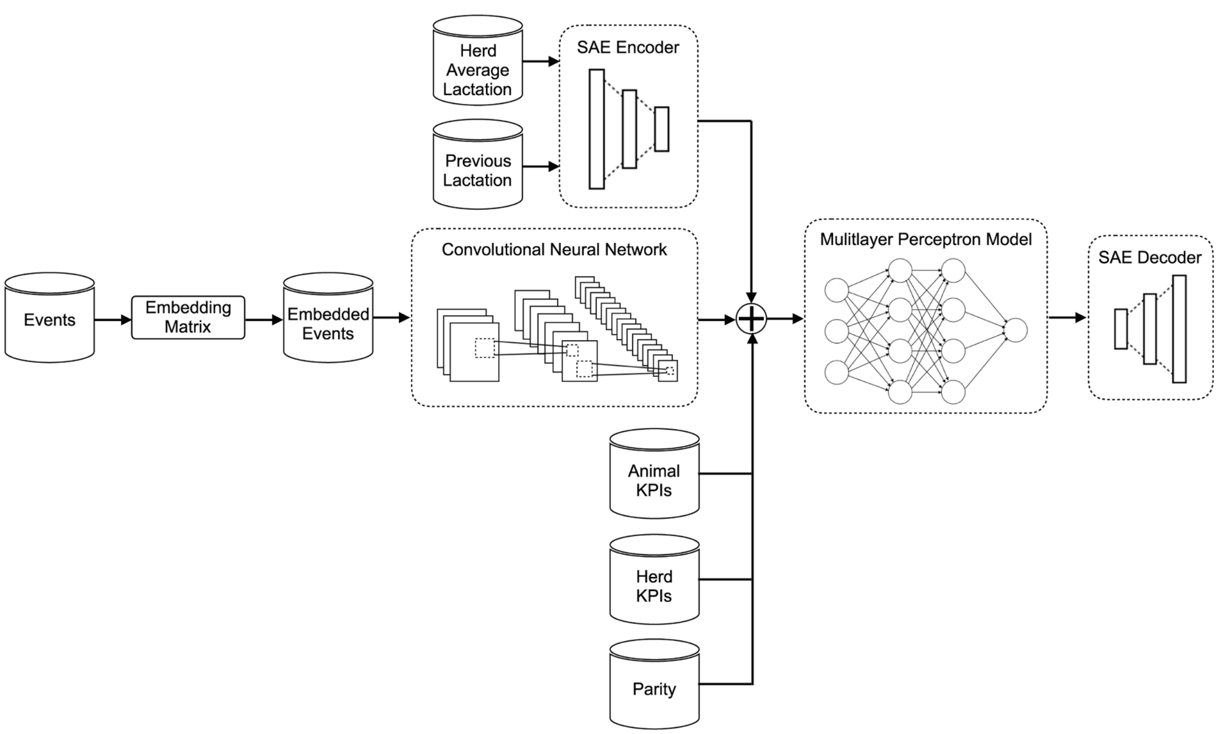


Fig. 3. Schematic overview of the SLMYP.

lactation curve, but with a lower weight for each subsequent milk yield. As a result, the model trained with the logarithmic loss function was particularly focused on generating accurate predictions in early lactation, yet without ignoring the predictions made for the remaining part of the lactation cycle. The weights used by each loss function are depicted by Fig. 4.

2.5. Benchmark models

The SLMYP was compared with three models used to predict the lactation curves of future cycles with no recordings. The first benchmark model uses the milk yields generated in the preceding cycle as forecast. A second model approximates the lactation curve by the herd's average milk yield curve corresponding to the predicted period. Third, a Wood's curve was fitted on the lactation data for each distinct parity. The predictions were then generated by the Wood's curve corresponding to the parity of the predicted lactation cycle. Finally, the SLMYP's predictions were also compared with the predictions generated by the SAE for increasing windows of observed milk yields in the predicted period.

2.6. Model evaluation

The performance of the models were evaluated by four metrics frequently used in similar research: the Pearson correlation coefficient (ρ), the root mean squared error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE) (Liseune et al., 2020; Lacroix et al., 1995; Grzesiak et al., 2003b). In contrast to the MAE and RMSE, which are absolute measures of fit, the MAPE is not scale dependent and indicates how much the predictions deviate from the true values on average. It is defined by the following formula:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \widehat{y}_i|}{y_i} \times 100\%$$

2.7. Variable importance

The variable importance (VI) score measures the relative increase of the model's error for a random perturbation of a specific feature (Breiman, 2001). More specifically, if e is the error of the model using all the features and $\widetilde{e}_{k,i}$ is the error of the model for the ith random perturbation of feature k, then the VI of feature k can be calculated as follows:

$$VI_k = \frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{e}_{k,i}}{e}$$

Likewise, the relative increase of the model's error and hence the VI score for an entire feature group can be calculated by randomizing all the features belonging to that feature group.

3. Results

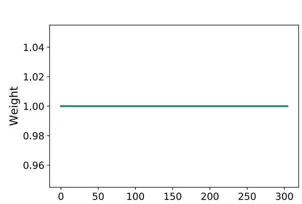
3.1. Model selection

The Bayesian optimization procedure was initialized by evaluating 50 randomly sampled model configurations. Each parameter setting was evaluated on the validation set after every 1000 training iterations with a batch size of 32. Every time the validation RMSE decreased, the model's weights were saved and training was terminated when the performance did not improve for 10 consecutive times. The best performing model was retrained on the training and validation set and evaluated on the test set. This model included the lactation number, the latent representation of the herd's average milk yield curve, the animal and herd KPIs as well as the sequence of the last 300 health and reproductions events encountered by the cow in the preceding lactation cycle. The events were embedded into 5-dimensional vectors and were passed to the CNN which consisted out of 4 blocks with 16 and 32 kernels of size 3 in the first two and last two blocks respectively. The output of each block's non-linear activation was dropped with a probability of 0.5 and a max pooling layer of size 4 was applied in block 1 and 4. The MLP contained 2 layers with 100 and 50 neurons and a dropout probability of 0.2 in each layer. The α of the Leaky ReLU layers in both the CNN and MLP was 0.5 and the initial learning rate of the Adam optimization algorithm was set at 0.001. Furthermore, the best results were obtained when the training data was balanced with respect to the lactation number.

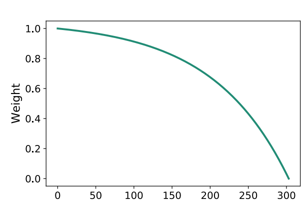
3.2. Model performance

The predictive performance on the daily as well as 305d yield of the SLMYP trained with the uniform loss function and the baseline models on the test set is depicted in Table 2. For the daily yields, the metrics were obtained by taking into account the errors between the non-missing milk yields in the predicted period and the corresponding model predictions. The performance scores for the 305d yields were obtained by comparing the true 305d milk yield with the predicted 305d yield (i.e., the summation of all the predicted yields in a certain cycle). The worst performing baseline model was the Wood's model, with an average RMSE and a MAPE of 9.58 kg and 30% for the daily milk yield and 2338.56 kg and 18% for the 305d yield, respectively. Using the average curves per herd and per parity as predictions on the other hand resulted in a RMSE of 7.97 kg and a MAPE of 25% for the daily predictions. The SLMYP performed best on the daily as well as 305d yield predictions with respect to every metric. On average, the SLMYP's absolute error between the daily milk yield predictions and the true values comprised 5.58 kg. For the 305d milk yield, the prediction error obtained by the SLMYP was 11% and was 2 percentage points lower than the best performing baseline model. Fig. 5 visualizes the predictions made by the baseline models as well as by the SLMYP for two random examples from the test set. Fig. 5a shows how the SLMYP is better able to model the peak of the lactation curve compared to the predictions made by the baseline models. Fig. 5b on the other hand shows that in contrast to the baseline models, the SLMYP is able to predict lower milk yield returns than expected.

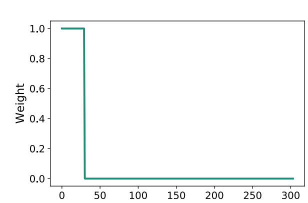
The impact of the different loss functions on the SLMYP's performance for different forecasting windows is displayed in Table 3. As expected, the SLMYP trained with the step loss function achieved the lowest MAE for the first recorded milk yields in the predicted lactation cycle. In particular, by applying the step loss function, the MAE was



(a) Uniform loss



(b) Logarithmic loss

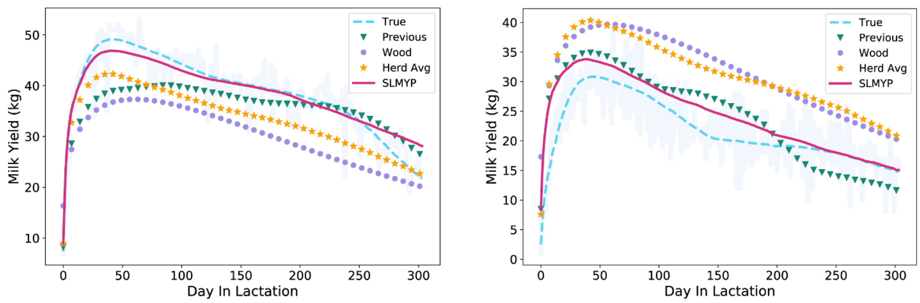


(c) Step loss

Fig. 4. Loss functions.

Table 2
Performance on daily as well as 305d yield of the SLMYP and baseline models. Baseline 1 = lactation curve of preceding cycle, Baseline 2 = average lactation per herd per parity, Baseline 3 = Wood's curve.

	Daily yield				305d yield			
Model	RMSE	MAE	MAPE	ρ	RMSE	MAE	MAPE	ρ
Baseline 1	9.22	7.13	0.26	0.62	2071.33	1081.01	0.16	0.61
Baseline 2	7.97	6.21	0.25	0.70	1705.09	1346.82	0.13	0.61
Baseline 3	9.58	7.67	0.30	0.53	2338.56	1887.68	0.18	0.18
SLMYP	7.38	5.58	0.23	0.75	1448.95	1093.30	0.11	0.73



(a) (b)

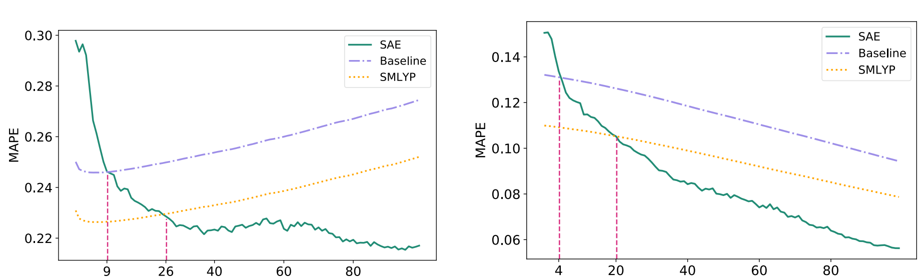
Fig. 5. Visualization of SLMYP and baseline predictions for two random examples of test set.

Table 3
MAE performance of SLMYP trained with different loss functions for multiple forecast horizons.

	Forecast horizon												
Loss	7	14	21	30	60	90	120	150	180	210	240	270	300
Uniform	5.67	5.58	5.63	5.72	5.82	5.78	5.74	5.71	5.68	5.65	5.62	5.59	5.58
Logarithmic	5.58	5.50	5.56	5.67	5.79	5.76	5.73	5.70	5.67	5.65	5.62	5.59	5.59
Step	5.57	5.49	5.56	5.67	6.81	9.75	10.02	10.07	9.99	9.89	9.87	9.91	10.05

5.57 kg for the first week of lactation and 5.67 kg for the first month of lactation. For larger windows however, the performance obtained by the step loss started to decrease rapidly, with the MAE being 6.81 kg for the first 60 days and 10.05 kg for the entire lactation cycle. In contrast, the MAE of the SLMYP trained with the uniform and logarithmic weights never exceeded 6.0 kg for every possible window. Furthermore, the model trained with the logarithmic loss function performed better than the model trained with the uniform loss weights for the first 180 days of lactation. For larger windows, both models performed equally well except for the largest possible forecasting window in which the model with the uniform loss achieved the best results overall.

Furthermore, the performance of the SLMYP trained with the logarithmic loss function as well as the best performing baseline model was compared with the SAE presented in Liseune et al. (2020) during the predicted period. More specifically, Fig. 6 compares the SAE's predictions for increasing windows of observed milk yields in the predicted period with the SLMYP's predictions made in the preceding cycle. For each possible window of observed yields, the performance of both models was calculated for the remaining unknown part of the lactation curve. As expected, the SAE's daily as well as 305d predictions deviated a lot from the true values in the beginning of the lactation cycle but became more accurate as more milk yields were observed. More specifically, the SAE's MAPE for the daily predictions decreased from 30% when no observations were available to 23% when 20 milk yields were recorded. On the contrary, the MAPE of the SLMYP remained below 23% for the first 20 days of lactation. However, as the beginning of the forecasting window was shifted towards the end of the lactation period, the MAPE of the SLMYP for the daily yields increased gradually as the MAE remains more or less constant while the average milk yield becomes smaller towards the end of the lactation cycle. From 26 days



(a) Daily performance (b) 305d performance

Fig. 6. Daily and 305d milk yield performance of the SAE, SLMYP and baseline models for different windows of observed data in the predicted period.

onwards, the SAE outperformed the SLMYP in terms of MAPE by leveraging the observed data. On the other hand, the MAPE of the best performing baseline model that predicted a subsequent lactation curve as the herd's average milk yield per parity was already surpassed by the SAE after 9 days of observed data. For the 305d predictions, the performance of the SAE, SLMYP and baseline model increased for larger windows of recorded milk yields as the predicted 305d yield comprises the cumulative predicted yield as well as the cumulative observed yield. Yet, in this case, the SAE's performance surpassed the baseline model already at the 4th day of lactation while it surpassed SLMYP's performance at the 20th day of lactation, with the MAPE of both models being around 11% at that day.

3.3. Variable importance

The variable importance scores of each group of features are visualized in Fig. 7. The latent representation of the average milk yield curve per herd and per parity had the highest VI of 1.47, meaning that the SLMYP's total RMSE of 7.38 kg increased to 10.84 kg when the latent features were randomly permutated. The animal KPIs and latent representation of the previous lactation curve were the second and third most discriminative feature groups with VI scores of 1.08. The least discriminative feature group were the herd KPIs with a VI score of 1.02.

The impact of several features on the SLMYP's predictions for a random test observation are depicted by Fig. 8. When mastitis was manually injected at the 170th day of the previous lactation cycle, the SLMYP predicted a slightly lower milk production than when the cow would be healthy. Likewise, the SLMYP adjusted the milk yield curve downwards when the disease event was injected at the 174th day of the previous lactation cycle. For each consecutive parity, the SLMYP predicted a slightly lower milk yield curve for a fixed lactation curve in the preceding cycle. When one of the latent variables of the low-dimensional representation of the previous lactation curve that was related to the cow's persistency was increased to its maximum value of 1, a more gradual decline of the subsequent curve was predicted as well. Finally, when the animal KPI corresponding to the 305d milk yield was decreased to its minimum value of 0, the SLMYP predicted lower returns for the entire lactation cycle. The milk yield curve was shifted upwards when the herd KPI related to the average days open was set to its minimum value of 0.

4. Discussion

Overall, the SLMYP performed better than all the baseline models with an average correlation of 0.75 between the predictions and the true values for the daily milk yields. The SAE presented by Liseune et al. (2020) on the other hand obtained a correlation of 0.77 when 30 days of data were observed. Furthermore, the SLMYP made the most accurate 305d milk yield forecasts with an average prediction error of 11%. This error is slightly higher compared to the results found by Grzesiak et al. (2003a) who reported a prediction error of 9% by making use of a spline model with one test-day (TD) record observed during the first 28 days of lactation. By applying autoregressive models on 2 TD records observed in the first 2 months of lactation, Vasconcelos et al. (2004), Macciotta et al. (2002) reported correlations between the predicted and true 305d milk yields of 0.85 and 0.88 respectively. For a comparable window of available data, the SAE proposed by Liseune et al. (2020) obtained an even higher correlation of 0.90. On the contrary, the SLMYP achieved a correlation of 0.73 between the predicted and true 305d values. Yet, while the previously mentioned studies leveraged data observed during early lactation, the SLMYP generates its milk yield predictions before the start of lactation. As shown by Fig. 6, the SLMYP was still able to produce better predictions for the daily milk yields than the SAE until the 26th day of the lactation cycle. In terms of 305d milk yields, the SLMYP generated better predictions than the SAE until the 20th day of the lactation cycle. Hence, during the first 26 days of lactation, the SLMYP enables farmers to obtain more accurate estimates of milk losses in early lactation and hence facilitate animal monitoring systems. In addition, the model presented in this study allows farmers to increase their forecast horizon with respect to their herd's total productivity with 20 days on average. After that period, predictions of lactation models such as the SAE become more accurate as these model are able to leverage the observed milk yields.

While no studies currently exist that predict the entire lactation curve by using data obtained in the previous cycle, curve fitting models such as those proposed by Wood (1967), Ali and Schaeffer (1987), Wilmink (1987) are able to make forecasts for future cycles by fitting curves on lactation data of homogeneous groups of animals. The resulting parameters of the fitted curves thus describe the group's average production and hence do not incorporate historical information from individual animals. The SLMYP however generates its predictions by taking into account both group statistics as well as individual information regarding historical milk production and reproduction as well as health events. In addition, Silvestre et al. (2006) showed that the accuracy of curve fitting models heavily depends on the sampling properties of the recorded milk yields. This is less a problem for SLMYP as it uses the latent representation of the historical as well as the herd's average milk yield extracted by the SAE. More specifically, as was shown by Liseune et al. (2020), the SAE's MAPE for reconstructing the entire lactation curve decreased by a maximum of 2 percentage points when 60% of the input milk yields were randomly dropped. Using the latent representation instead of the raw milk yield sequences thus makes the SLMYP particularly robust for missing data in the features corresponding to the cow's and herd's lactation curves.

Finally, as shown by Fig. 7, the latent representation of the herd's average milk yield curve corresponding to the predicted lactation cycle contributed the most to the predictions. This could be expected since cows from the same herd usually share the same breed, feeding systems, herd management and climate, which have been shown to significantly affect milk production (Rekik and Gara, 2004). In addition to variables summarizing the average herd production, other features related to herd management also had an impact on the milk production. In Fig. 8f for example, the SLMYP increased its predictions for a lower value of the average number of days open. This is not surprising as a delay in

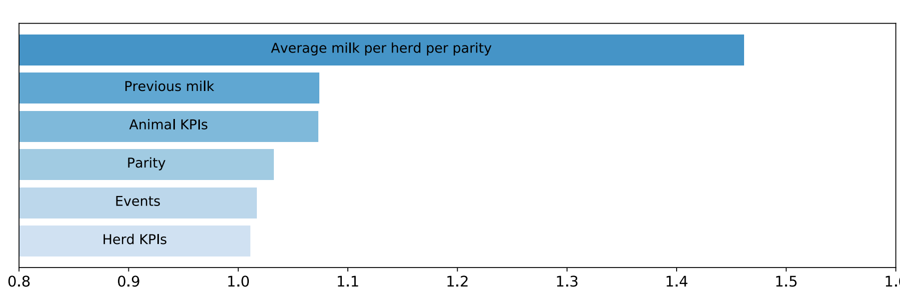
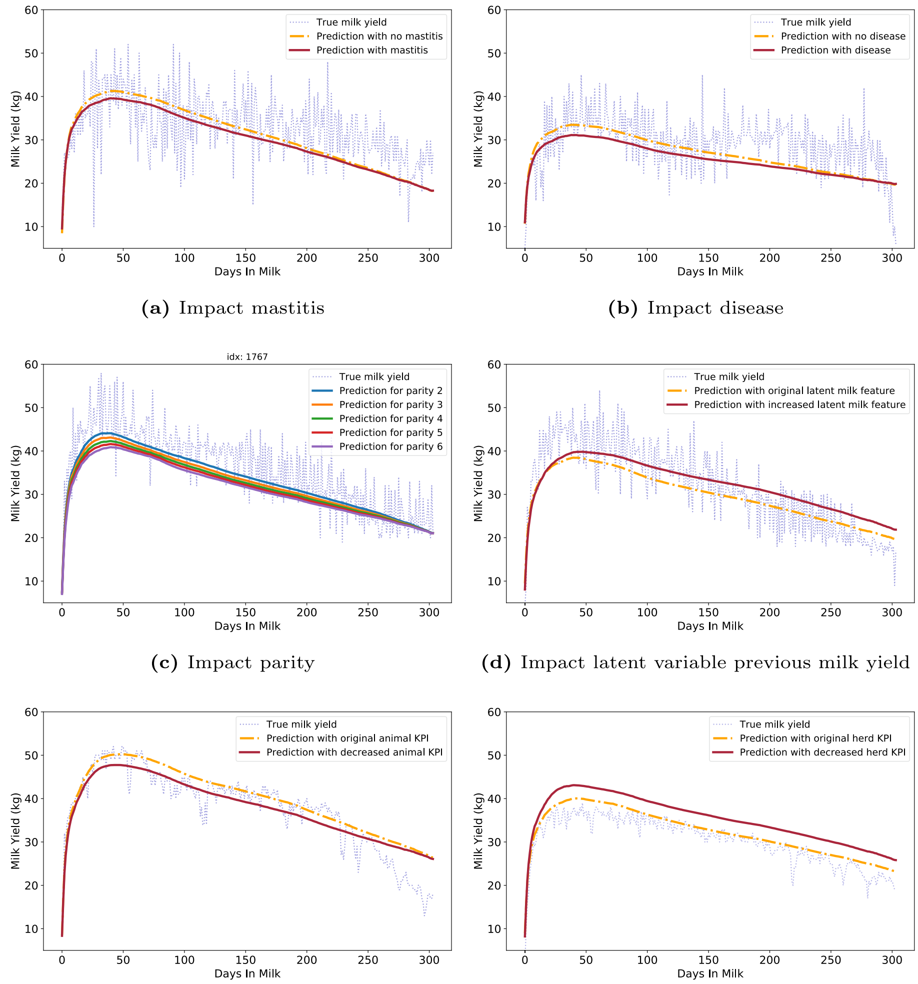


Fig. 7. Variable importance scores of feature groups.



(a) Impact mastitis **(b)** Impact disease

(c) Impact parity **(d)** Impact latent variable previous milk yield

(e) Impact animal KPI 305d milk yield **(f)** Impact herd KPI avg days open

Fig. 8. Variable importance.

pregnancy will inversely influence milk production (Cattaneo et al., 2015). In contrast to previous studies which found that cows generally produced more in each subsequent lactation cycle (Ehrlich et al., 2011; Macciotta et al., 2011), the SLMYP predicted lower milk yield returns for higher parities. Yet, this could be explained by the fact that the curves corresponding to the previous lactation cycle as well as to the herd's average lactation remained fixed and hence became abnormally low for higher parities. Furthermore, the results presented in this research are in line with previous studies in that a cow's historical milk production is positively related to its future production (Ali and Schaeffer, 1987). As shown by Fig. 8e and d, the SLMYP lifted the predicted curve upwards as the total milk production of the preceding cycle was increased or when the persistency of the preceding cycle was positively adjusted. On the contrary, a loss in milk yield can be expected in case of mastitis or disease (Adriaens et al., 2018). This was shown by Fig. 8a and b in which the SLMYP slightly adjusted its predictions downwards when the cow was sick during the preceding cycle. Hence, the SLMYP is able to generate more realistic milk yield predictions by taking into account the sequence of reproduction and health events. As a result, differences between expected and produced milk yields can be calculated more accurately which improves disease detection.

5. Conclusion

Current lactation models rely on a fixed number of milk yields recorded in early lactation to forecast individual milk yield curves. As a result, animal monitoring becomes particularly difficult in early lactation as expected milk yields are often missing in the period immediately after calving. In addition, forecasts of a cow's total productivity can only be obtained from the moment the model's last required milk yield input is observed. Curve fitting models on the other hand are able to generate entire lactation curves. These curves however represent group averages and hence remain constant irrespective of the animal's individual variation. In this study, we present a framework that, in addition to herd statistics, uses the cow's historical sequence of milk yields as well as reproduction and health events in the preceding cycle to predict the cow's entire lactation curve in the subsequent cycle. Results show that by leveraging individual data, the model is able to generate more accurate predictions than by solely using group averages. As a result, the framework presented in this research can be used to assess the impact of herd management, health and reproduction events as well as a cow's historical milk yield on the cow's future productivity. In addition, the model allows to increase the farmer's forecast horizon with respect to the herd's future productivity as well as to improve animal monitoring systems in early lactation.

CRediT authorship contribution statement

Arno Liseune: Conceptualization, Methodology, Software, Investigation, Visualization, Writing - original draft, Writing - review & editing, Formal analysis. **Matthieu Salamone:** Data curation, Writing - review & editing, Software. **Dirk Van den Poel:** Supervision, Resources.

Bonifacius Ranst: Resources. **Miel Hostens:** Supervision, Software, Resources, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adriaens, I., Huybrechts, T., Aernouts, B., Geerinckx, K., Piepers, S., Ketelaere, B.D., Saeys, W., 2018. Method for short-term prediction of milk yield at the quarter level to improve udder health monitoring. J. Dairy Sci. 101, 10327–10336.

Ali, T.E., Schaeffer, L.R., 1987. Accounting for covariances among test day milk yields in dairy cows. Can. J. Anim. Sci. 67, 637–644.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brody, S., Ragsdale, A.C., Turner, C.W., 1923. The rate of decline of milk secretion with the advance of the period of lactation. J. Gen. Physiol. 5, 441–444.

Cattaneo, L., Baudracco, J., Lazzarini, B., Ortega, H., 2015. Methodology to estimate the cost of delayed pregnancy for dairy cows. An example for Argentina. Revista Brasileira de Zootecnia 44, 226–229.

Dematawewa, C., Pearson, R., VanRaden, P., 2007. Modeling extended lactations of Holsteins. J. Dairy Sci. 90, 3924–3936.

Ehrlich, J., et al., 2011. Quantifying shape of lactation curves, and benchmark curves for common dairy breeds and parities. Bovine Practitioner 45, 88.

Gorgulu, O., 2012. Prediction of 305-day milk yield in Brown Swiss cattle using artificial neural networks. South Afr. J. Anim. Sci. 42, 280–287.

Gröhn, Y., Wilson, D., González, R., Hertl, J., Schulte, H., Bennett, G., Schukken, Y., 2004. Effect of pathogen-specific clinical mastitis on milk yield in dairy cows. J. Dairy Sci. 87, 3358–3374.

Græsbøll, K., Kirkeby, C., Nielsen, S.S., Halasa, T., Toft, N., Christiansen, L.E., 2016. Models to estimate lactation curves of milk yield and somatic cell count in dairy cows at the herd level for the use in simulations and predictive models. Front. Vet. Sci. 3, 115.

Grzesiak, W., Blaszczyk, P., Lacroix, R., 2006. Methods of predicting milk yield in dairy cows—predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). Comput. Electron. Agric. 54, 69–83.

Grzesiak, W., Lacroix, R., Wójcik, J., Blaszczyk, P., 2003a. A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. Can. J. Anim. Sci. 83, 307–310.

Grzesiak, W., Wojcik, J., Binerowska, B., 2003b. Prediction of 305-day first lactation milk yield in cows with selected regression models. Arch. Anim. Breed. 46, 213–224.

Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. ArXiv e-prints.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. ArXiv e-prints.

Kingma, D., Ba, J., 2014. Adam: A method for stochastic optimization. ArXiv e-prints.

Kirkpatrick, M., Hill, W.G., Thompson, R., 1994. Estimating the covariance structure of traits during growth and ageing, illustrated with lactation in dairy cattle. Genet. Res. 64, 57–69.

Lacroix, R., Salehi, F., Yang, X., Wade, K., 1997. Effects of data preprocessing on the performance of artificial neural networks for dairy yield prediction and cow culling classification. Trans. ASAE 40, 839–846.

Lacroix, R., Wade, K., Kok, R., Hayes, J., 1995. Prediction of cow performance with a connectionist model. Trans. ASAE 38, 1573–1579.

Liseune, A., Salamone, M., Van den Poel, D., Van Ranst, B., Hostens, M., 2020. Leveraging latent representations for milk yield prediction and interpolation using deep learning. Comput. Electron. Agric. 175, 105600.

Macciotta, N., Vicario, D., Pulina, G., Cappio-Borlino, A., 2002. Test day and lactation yield predictions in Italian Simmental cows by ARMA methods. J. Dairy Sci. 85, 3107–3114.

Macciotta, N.P., Dimauro, C., Rassu, S.P., Steri, R., Pulina, G., 2011. The mathematical description of lactation curves in dairy cattle. Italian J. Anim. Sci. 10, e51.

Murphy, M., O'Mahony, M., Shalloo, L., French, P., Upton, J., 2014. Comparison of modelling techniques for milk-production forecasting. J. Dairy Sci. 97, 3352–3363.

Murphy, M., Zhang, F., Upton, J., Shine, P., Shalloo, L., 2018. A review of milk production forecasting models. pp. 14–61.

Njubi, D., Wakhungu, J., Badamana, M., 2010. Use of test-day records to predict first lactation 305-day milk yield using artificial neural network in Kenyan Holstein-Friesian dairy cows. Tropical Anim. Health Prod. 42, 639–644.

Olori, V., Brotherstone, S., Hill, W., McGuirk, B., 1999. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. Livestock Prod. Sci. 58, 55–63.

Rekik, B., Gara, A., 2004. Factors affecting the occurrence of atypical lactations for Holstein-Friesian cows. Livestock Prod. Sci. 87, 245–250.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536.

Salehi, F., Lacroix, R., Wade, K., 1998. Improving dairy yield predictions through combined record classifiers and specialized artificial neural networks. Comput. Electron. Agric. 20, 199–213.

Sanzogni, L., Kerr, D., 2001. Milk production estimates using feed forward artificial neural networks. Comput. Electron. Agric. 32, 21–30.

Sharma, A.K., Sharma, R., Kasana, H., 2006. Empirical comparisons of feed-forward connectionist and conventional regression models for prediction of first lactation 305-day milk yield in Karan Fries dairy cows. Neural Comput. Appl. 15, 359–365.

Sharma, A.K., Sharma, R., Kasana, H., 2007. Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. Appl. Soft Comput. 7, 1112–1120.

Silvestre, A., Petim-Batista, F., Colaço, J., 2006. The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. J. Dairy Sci. 89, 1813–1821.

Tekerli, M., Akinci, Z., Dogan, I., Akcan, A., 2000. Factors affecting the shape of lactation curves of Holstein cows from the Balikesir province of Turkey. J. Dairy Sci. 83, 1381–1386.

Torres, M., Hervás, C., Amador, F., 2005. Approximating the sheep milk production curve through the use of artificial neural networks and genetic algorithms. Comput. Oper. Res. 32, 2653–2670.

Vasconcelos, J., Martins, A., Petim-Batista, M., Colaço, J., Blake, R., Carvalheira, J., 2004. Prediction of daily and lactation yields of milk, fat, and protein using an autoregressive repeatability test day model. J. Dairy Sci. 87, 2591–2598.

Wilmink, J., 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. Livestock Prod. Sci. 16, 335–348.

Wilson, D., González, R., Hertl, J., Schulte, H., Bennett, G., Schukken, Y., Gröhn, Y., 2004. Effect of clinical mastitis on the lactation curve: a mixed model estimation using daily milk weights. J. Dairy Sci. 87, 2073–2084.

Wood, P., 1967. Algebraic model of the lactation curve in cattle. Nature 216, 164–165.

Zhao, B., Lu, H., Chen, S., Liu, J., Wu, D., 2017. Convolutional neural networks for time series classification. J. Syst. Eng. Electron. 28, 162–169.

Zheng, Y., Liu, Q., Chen, E., Ge, Y., Zhao, J.L., 2014. Time series classification using multi-channels deep convolutional neural networks. In: Web-Age Information Management, pp. 298–310.