Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

# A machine learning framework to predict the next month's daily milk yield, milk composition and milking frequency for cows in a robotic dairy farm

*Boyu Ji* [a], *Thomas Banhazi* [b,*], *Clive J.C. Phillips* [c,d], *Chaoyuan Wang* [e], *Baoming Li* [e]

[a] *Digital Innovation Team, Pig Improvement Company, Genus, Shanghai, China*
[b] *Faculty of Health, Engineering and Science, University of Southern Queensland, Toowoomba, QLD, Australia*
[c] *Curtin University, School of Design and the Built Environment, Perth, WA, Australia*
[d] *Estonian University of Life Sciences, Institute of Veterinary Medicine and Animal Sciences, Kreutzwaldi 1, Tartu 51014, Estonia*
[e] *College of Water Resources and Civil Engineering, China Agricultural University, Beijing, China*

A R T I C L E I N F O

*Article history:*
Received 30 July 2021
Received in revised form
13 February 2022
Accepted 18 February 2022
Published online 7 March 2022

*Keywords:*
Robotic dairy farm
Dairy cow
Machine learning
Thermal comfort index

Robotic milking systems (RMS) are increasingly utilised by modern livestock farmers because they can reduce labour costs, and they have the potential to collect data that will improve animal welfare and animal productivity through better monitoring. Sensors and devices installed in RMS enable farmers to routinely collect data on environment conditions, individual animal's behaviours, health, productivity, and milk quality. This dataset can be used to train artificial intelligence algorithms to predict trends in these variables. This study developed a machine learning framework using 5 years' behaviour, heath and productivity data from 80 cows in a robotic dairy farm. Here we demonstrate the development of a framework to automatically train models with up-to-date farm data and predict daily milk yield, composition (fat and protein content) and frequency of individual cow milking during the subsequent 28 days. A time series cross-validation was applied to simulate the application of this framework under commercial conditions and to evaluate the performance. A high accuracy of prediction ($R^2 > 0.90$ and overall accuracy > 80%) was achieved with the models created by this framework. The practical potential of using such frameworks to enhance the management efficiency and animal welfare in robotic dairy farms is discussed.



* *Corresponding author.*
E-mail address: Thomas.Banhazi@usq.edu.au (T. Banhazi).

https://doi.org/10.1016/j.biosystemseng.2022.02.013
1537-5110/

**Nomenclature**

| | |
|---|---|
| API | Application programming interface |
| BM | Body mass (kg) |
| CMS | Conventional milking systems |
| DIM | Days in milk |
| DMI | Dry matter intake (kg $d^{-1}$) |
| DMY | Daily milk yield (kg $cow^{-1}$ $d^{-1}$) |
| DOY | Day of year |
| FC | Fat content (%) |
| MF | Milking frequency |
| MT | Milk temperature (°C) |
| PC | Protein content (%) |
| PLF | Precision livestock farming |
| RF | Random forest |
| RFID | Radio frequency identification |
| $RH_{max}$ | Maximum relative humidity (%) |
| $RH_{min}$ | Minimum relative humidity (%) |
| RMS | Robotic milking systems |
| RT | Rumination time (min) |
| $T_{max}$ | Maximum temperature (°C) |
| $T_{min}$ | Minimum temperature (°C) |
| $THI_{max}$ | Maximum temperature and humidity index |
| $THI_{min}$ | Minimum temperature and humidity index |
| VP | Vapour pressure (hPa) |
| VPD | Deficit vapour pressure (hPa) |

## 1. Introduction

Robotic milking systems (RMS) are becoming a strong competitor to conventional milking systems (CMS) in recent years as farmers are increasingly adopting Precision Livestock Farming (PLF) technologies on conventional farms (Banhazi et al., 2012; McCullough, 2019). The reported benefits of RMS include decreased labour cost (Rodenburg, 2012), improved animal welfare (Heyden, 2015), and better milk quality (Bear & Holloway, 2019). However, investing in costly RMS can lead to financial risk or business failure (Simões Filho et al., 2020). The cost-benefit of adopting RMS can vary widely as a result of variation in management (Hermans, Ipema, Stefanowska, & Metz, 2003), variation in climatic conditions due to climate change (Speroni, Pirlo, & Lolli, 2006) and other biometric factors (Rodenburg & Wheeler, 2002). Regardless of these financial risks, other practical aspects of RMS (such as the ability of these systems to collect a vast amount of data on various aspects of dairy production) can still make investment in RMS an attractive proposition (Ji, Banhazi, Ghahramani, et al., 2020a; Jorquera-Chavez et al., 2019). The database of the changed health status and production performance of livestock can be used to create predictive models for lameness and oestrus detection (Klaas, Rousing, Fossing, Hindhede, & Sorensen, 2003; Pastell & Madsen, 2008), heat stress assessment (Ji et al., 2019a, 2019b; Ji, Banhazi, Ghahramani, et al., 2020a, 2020b) and milk yield prediction (Fuentes et al., 2020; Yan, Chen, Akcan, Lim, & Yang, 2015).

Predictive models related to cow behaviours, health, productivity and milk quality are important components of the modern dairy industry, as these models can assist with decision making on farms and product marketing. For example, prediction of daily milk yield (DMY) is valuable for identifying cows that have failed to attain their predicted trajectory in milk yield for reasons of disease or other welfare issue. It can also assist in selecting dairy cows earlier for genetic improvements and breeding programmes (Kettunen, Mäntysaari, Strandén, Pösö, & Lidauer, 1998; Schaeffer, 1994). Lactation curves can assist farmers to pinpoint cows under negative energy balance and identify animals that are more resistant to metabolic stress so that they can optimise the feeding operation, including targeting cows that are losing too much weight by feeding them high quality supplements, such as bypass protein (Dekkers, Ten Hag, & Weersink, 1998; Jakobsen et al., 2002; Swalve, 2000). Prediction of milk composition (in particular fat and protein content) provides essential information about the anticipated purchase price of milk (Klopčič, Malovrh, Gorjanc, Kovač, & Osterc, 2003; Mayeres, Stoll, Bormann, Reents, & Gengler, 2004), which can help farmers to make better business decisions. Moreover, the prediction of milking frequency (MF) is another key piece of information that can be helpful in evaluating cow welfare and management efficiency in relation to optimising the traffic system management in RMS (Hart, McBride, Duffield, & DeVries, 2013; Ji, Banhazi, Ghahramani, et al., 2020a; Wildridge et al., 2018). Regression models are the usual way of building predictive models with cow-related factors (Grzesiak, Zaborski, Szatkowska, & Królaczyk, 2021; Ji et al., 2019a, 2019b; Ji, Banhazi, Ghahramani, et al., 2020a, 2020b; McParland et al., 2019). Core algorithms of regression models mainly involve both conventional statistics and machine learning. The former category of algorithms depends on standard linear regression techniques, such as stepwise multiple linear regression, segmented multi-phases regression and generalised linear regression (Rubinfeld, 1998). However, the practical application and prediction power of such linear regressions can be severely limited due to their structural restrictions (in particular the number of possible input variables) and quality of the available data (especially the type of data distribution) (Fang, Hanna, Haque, & Spillman, 2000; Kominakis, Abas, Maltaris, & Rogdakis, 2002), Machine learning algorithms are increasingly preferred by companies seeking to build practical applications, due to their improved flexibility and better accuracy (Sharma, Sharma, & Kasana, 2007). So far, predictive models have been developed for the dairy industry using supported vector regression (Hsieh, Hung, & Kuo, 2011), tree-based (random forest) regression (Fenlon et al., 2016) and neural network regression (Grzesiak et al., 2021). Significant amounts of data generated by RMS in many countries in recent decades are available to model developers. Published predictive models mainly focus on prediction at a herd level (i.e. 305-day milk yield) (Sharma et al., 2007), or if prediction is attempted at an individual level (e.g. next day milk yield) this is usually done over a very short time frame (McParland et al., 2019). Newly developed predictive models are expected to be able to predict factors related to individual cows rather than the whole herd and over longer time periods, to allow immediate and precise problem solving on farms. This includes the identification of individual needs (e.g. heat stress mitigation, nutrition supply or medical treatment), which in turn reduces the chances of long term productivity loss (Ji, Banhazi, Perano, et al., 2020c; Mayeres et al., 2004). However, So far, few research groups have developed models aimed at predicting important production-related parameters on an individual cattle level over a long time horizon. Furthermore, few studies have reported the linking of predictive models with databases supplied by RMS (Fuentes et al., 2020).

Therefore, a study was initiated to test the development of a machine learning framework for forecasting DMY, components and MF of individual cows over a one-month (28 days) period. It is envisaged that, if successful, a predictive model would be available for optimisation for animal production and improvement of cow welfare.

## 2. Methodology

### 2.1. Data collection

This study used production and behaviour data from 80 lactating dairy cows in a robotic dairy farm over a 5-year period, 2013–2018. The dairy farm selected was in Gatton, Queensland, Australia, which contained an RMS with free traffic (Fig. 1a). The traffic between resting, milking and feeding areas was controlled by an electrical gate with an embedded programmer. Water troughs were placed in the milking area and in the walkways leading back to the feeding area. When a cow wanted to drink water, she would enter Walkway 1 and be recognised by her ear tag via radio frequency identification (RFID) techniques. The programme in the electrical gate then retrieved her daily milk production. If she did not meet the expected milk production, she would be returned to the feeding area after being milked via walkway 2. If she met the expectations, the gate would be switched to let her return to feeding area directly through route 3. In the milking area, 3 milking robots, shown in Fig. 1b (LELY Astronaut, LELY Industries, NV, Maassluis, The Netherlands), were installed to serve the cows. These robots measured the DMY, milk quality (protein and fat content, PC and FC) and MF of individual cows and saved these data in the herd management system (LELY T4C, LELY Industries NV, Maassluis, The Netherlands). It also measured the milk temperature (MT) as an indicator of each animal's core body temperature. Moreover, the intake of feed provided when the cows were being milked by the robot as a nutrient supplement was recorded as dry matter intake (DMI), while their basic diet was provided in feeding area with free access. The herd management system also documented age, body mass (BM) and days in milk (DIM) of individual cows. Daily rumination time (RT) of each cow was monitored using a neck band and collar (Lely Qwes-HR, Lely Industries BV, Maassluis, the Netherlands), with values uploaded to the herd management system as well. The data collection of this part was processed by directly downloading from the database of herd management system.

Climate data was downloaded from a government climate website (http://www.bom.gov.au/), which converted data measured by local weather stations into climatic grid points. The grid point selected in this study was approximately 2 km from the farm. For the time period of this study (2013–2017), climate data comprised daily maximum and minimum dry bulb temperature ($T_{max}$ and $T_{min}$), maximum and minimum relative humidity ($RH_{max}$ and $RH_{min}$), vapour pressure (VP), deficit vapour pressure (VPD), solar radiation (SR) and rainfall. Daily maximum and minimum temperature and humidity index ($THI_{max}$ and $THI_{min}$) were calculated using the downloaded temperature and humidity data

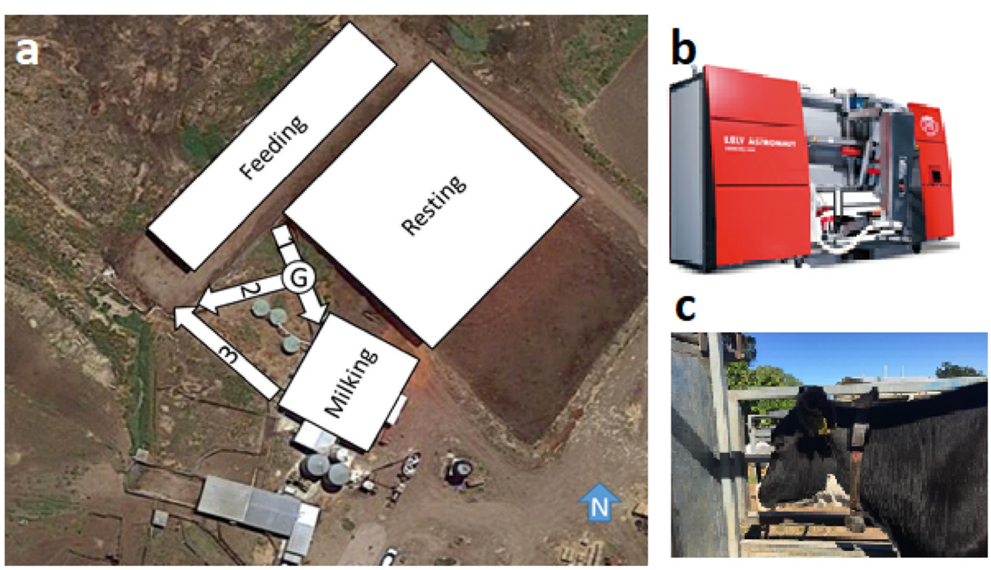


**Fig. 1 – Surveyed robotic dairy farm and animal measurement, [a] – farm layout and traffic control, G = electrical gate with an embedded programmer (as labelled with "G"). Water troughs were placed in the milking area and in the walkways leading back to the feeding area ("2" and "3") to motivate animal movements. When a cow wanted to drink water, she would enter walkway "1" and be recognised by her ear tag via radio frequency identification (RFID) techniques. [b] – LELY Astronaut milking robot and [c] – LELY Qwes-HR tag monitor (adapted from Ji et al., 2019a; Ji, Banhazi, Ghahramani, et al., 2020a).**

([$T_{max}$, $RH_{min}$] and [$T_{min}$, $RH_{max}$]) as inputs into the equation developed by Ji et al. (2019a) and Ji, Banhazi, Ghahramani, et al. (2020b).

### 2.2. XGBOOST algorithm

Tree–based algorithms, such as Random Forest (RF) (Liaw & Wiener, 2002), have been applied in many machine learning projects. In tree-based regression models, a number of regression trees (sub-models) are created (Fig. 2). Each regression tree has its own structure applying different logic to allocate a leaf with a score related to the input features. The final output or prediction is generated by summing the scores for the leaves from different trees. Similar to the tree-based regression model, the tree-based classification model just changes the score of leaves to class tag, and the final output is the class tag with the highest proportion on the allocated leaves. The key process of a tree-based algorithm is to optimise the number and structure of these regression trees providing the best accuracy and stability of prediction, which is called tree boosting. XGBOOST is one of latest tree boosting algorithms using a novel sparsity-aware algorithm and weighted quantile sketch for gradient tree learning (Chen & Guestrin, 2016). In this study, a XGBOOST regression model was built to predict milk yield, protein and FC (Fig. 2), and a classification model was built to predict MF (Fig. 3). Cow features, climate features and time series features were used as the input to these XGBOOST models. Climate features were calculated as daily average and the historical 60 days' average. To consider the seasonality, day of year (DOY) and month were transformed to sin and cos values using trigonometric functions. The importance of these features in building and boosting tree-based models were calculated and compared for model interpretation (Casalicchio, Molnar, & Bischl, 2018; Zien, Krämer, Sonnenburg, & Rätsch, 2009).

### 2.3. Framework design and cross-validation

To implement machine learning algorithms in practical farm management conditions, a framework was designed using Python (Van Rossum, 2007) to automatically collect and process data, train machine learning models and make predictions, as shown in Fig. 4. The framework depended on an Application Programming Interface (API) provided by different data sources (herd management system and government climate website) to automatically retrieve data as data frame objects (McKinney, 2011). Then data was processed to merge different data frame objects into one object using date as the key index, so that the merged data frame object could be used to train machine learning models. The data frame is in the form of a table with columns involving the features and targets described in Fig. 4. Rows containing null values or error values were removed from the merged data frame. After that, the framework trained four sets of XGBOOST machine learning models, including three sets of regression models for predicting DMY, PC and FC, and one set of classification models to predict MF. In each set, 28 XGBOOST models were built to predict 28 days' values, so that model #1 would predict the next day's value (i.e. $DMY_{+1}$), model #2 would predict value for the second day's value (i.e. $DMY_{+2}$), and model #28 would predict the value of the 28th day in the future (i.e. $DMY_{+28}$). Predictions from these models were combined into one data frame as the prediction of 1 month (28 days) into the future (i.e. $DMY_{+1}$, $DMY_{+2}$ … $DMY_{+28}$).

To evaluate the performance of the framework under practical farm management condition, cross-validation was

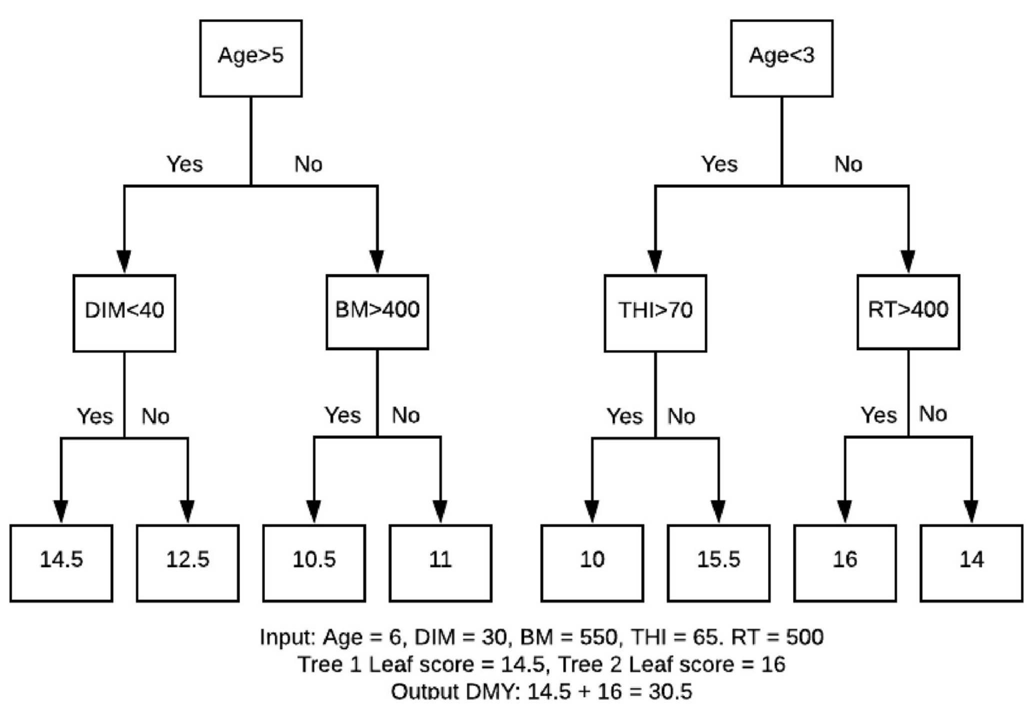


**Fig. 2 – An artificial example of tree-based regression model using 2 trees each with 2 depths to predict daily milk yield (DMY) in kg of individual cow. Values are randomly assigned to only show how a tree-based model is developed.**

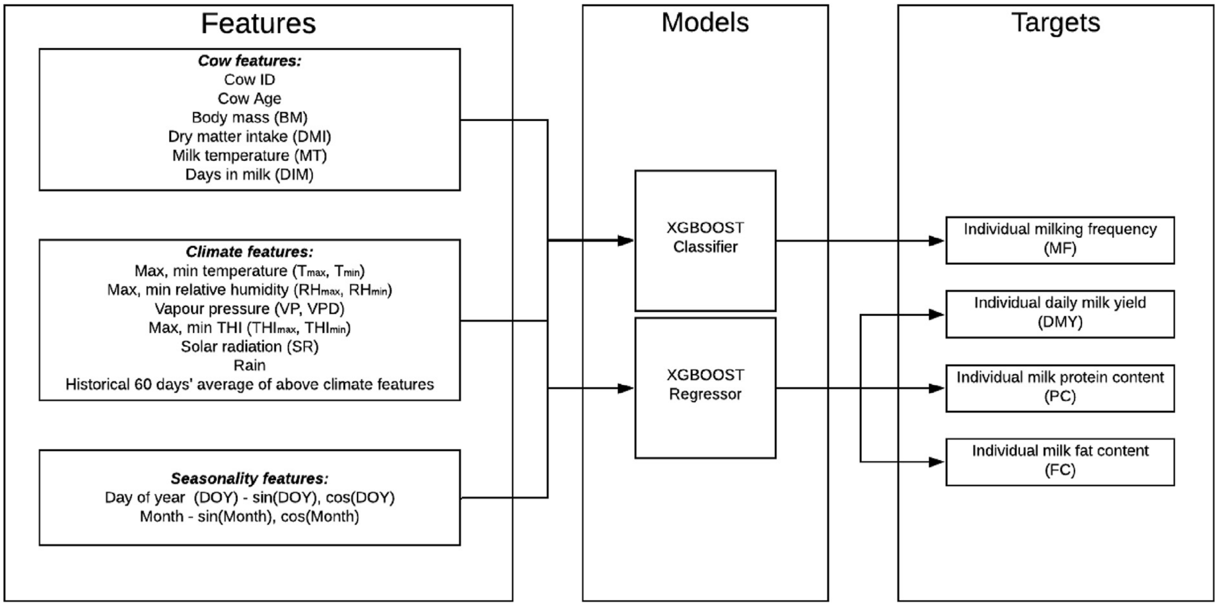


**Fig. 3 – XGBOOST models to predict milking frequency (MF), daily milk yield, protein content (PC) and fat content (FC). Input features include cow, climate and seasonality elements.**

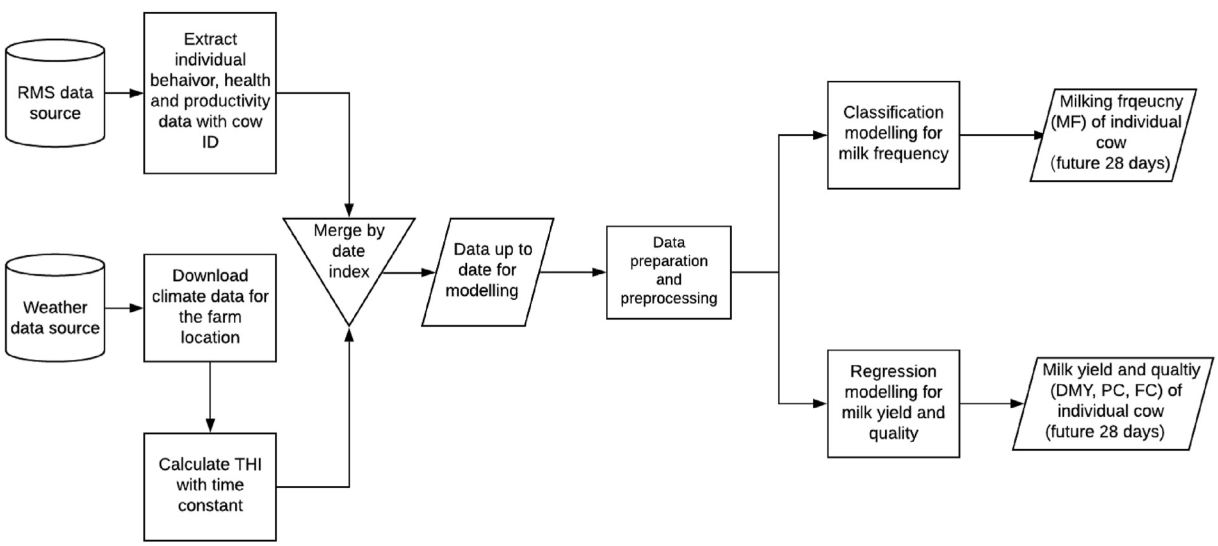


**Fig. 4 – Machine learning framework to enable automatic data collection, preparation, model training and prediction.**

conducted as described by Roberts et al. (2017). Given that each prediction generated by the framework was a series of production values for a one-month period (28 days), a time series rolling window with a size of 28 days was applied to create dynamic train and test dataset, as shown in Fig. 5. The initial training dataset used the data from May 2013 to May 2014 to ensure that the machine learning had sufficient training. The initial test data was used to predict values for the June after May 2014. After training the XGBOOST models, the accuracy of model prediction was calculated as the R squared ($R^2$) between actual and predicted values (Miles, 2014). Then the rolling window moved forward for a one-month period, updated the training and test datasets, refitted machine learning models and recalculated the accuracy values ($R^2$). This loop stopped at the end date of the full dataset (November 2017) and finally generated a set of accuracy values to evaluate the framework performance. Furthermore, the mean residuals between actual and predicted values were calculated, along with the distribution of prediction errors, as an extra evaluation of the framework performance.

## 3. Results

### 3.1. Descriptive statistics

The descriptive statistics associated with variables included in this study are shown in Table 1. The cow herd had an average

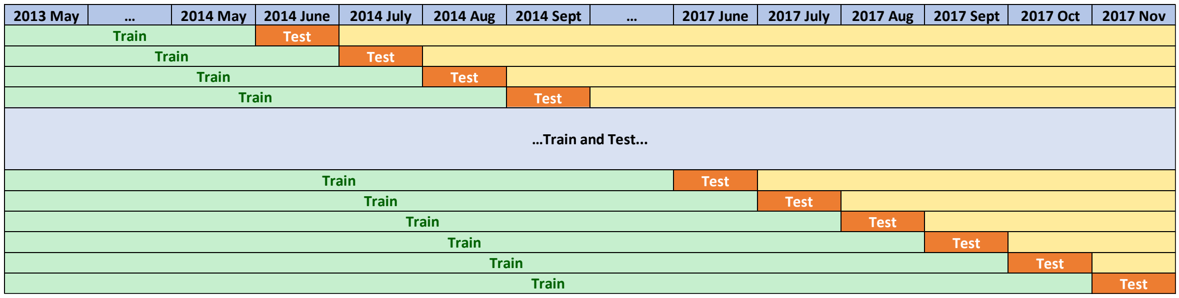


**Fig. 5 – Time series cross-validation to simulate the execution of machine learning framework and evaluate the accuracy of predictions under practical farm management condition.**

**Table 1 – Description of daily aggregated target, cow and climate variables from 80 cows over a five year observation period, presenting mean values and standard deviation. Variables included milking frequency (MF), daily milk yield (DMY), protein content (PC), fat content (FC), dry matter intake (DMI), milk temperature (MT), body mass (BM), days in milk (DIM), age, rumination time (RT), rain, maximum temperature ($T_{max}$), minimum temperature ($T_{min}$), vapour pressure (VP), deficit vapour pressure (VPD), solar radiation (SR), maximum relative humidity ($RH_{max}$), minimum relative humidity ($RH_{min}$), maximum temperature humidity index ($THI_{max}$) and minimum temperature humidity index ($THI_{min}$).**

| Category | Variable | Unit | Mean | Standard deviation |
|---|---|---|---|---|
| Targets | MF | times $d^{-1}$ | 2.8 | 0.9 |
| | DMY | kg $cow^{-1}$ $d^{-1}$ | 30.5 | 8.6 |
| | PC | % | 3.1 | 0.4 |
| | FC | % | 3.8 | 0.9 |
| Cow features | DMI | kg $d^{-1}$ | 4.6 | 1.4 |
| | MT | °C | 39.2 | 0.7 |
| | BM | kg | 682.2 | 76.8 |
| | DIM | d | 136.8 | 89.5 |
| | Age | y | 3.3 | 1.4 |
| | RT | min $d^{-1}$ | 413.3 | 117.4 |
| Climate features | Rain | mm | 45.5 | 9.9 |
| | $T_{max}$ | °C | 27.7 | 5 |
| | $T_{min}$ | °C | 13.2 | 5.7 |
| | VP | hPa | 16.2 | 5.1 |
| | VPD | hPa | 14.9 | 6.1 |
| | SR | MJ $m^{-2}$ | 18.2 | 6.1 |
| | $RH_{max}$ | % | 94.7 | 9.6 |
| | $RH_{min}$ | % | 42.9 | 11.5 |
| | $THI_{max}$ | N/A | 70.2 | 5.5 |
| | $THI_{min}$ | N/A | 54.8 | 6.9 |

BM of 682.2 kg, with a production performance (i.e. DMY) of 30.5 kg $cow^{-1}$ $d^{-1}$, and average age of 3.3 years. Their average lactation duration (i.e. DIM) was 136.8 days, i.e. the mid-lactation stage. The average MF of RMS was 2.8 times $d^{-1}$. During the period of data collection, the average temperature range was between 13.5 and 27.7 °C, while the extreme value was as high as 45.1 °C. The average humidity ranged from 42.9 to 94.7%. The range of THI was between 54.8 and 70.2. The variation between seasons and years has been reported extensively in previous studies (Ji et al., 2019a, 2019b; Ji, Banhazi, Ghahramani, et al., 2020a, 2020b).

### 3.2. *Evaluation of framework performance*

As described in section 2.3 and Fig. 5, the time series cross-validation was applied to evaluate the accuracy of framework predictions under practical conditions. Overall, 42 time series folds were generated from May 2014 to November 2017. Each fold had a combination of training and test datasets. For each target MF, DMY, PC and FC, the framework trained 28 models corresponding to the prediction of +1st to +28th day in each fold. Finally, for each target, 1176 (28 × 42) models were trained and tested.

The mean accuracy of 42 time series folds is shown in Fig. 6. The accuracy of MF prediction was less than 80% before January 2016, with some folds' accuracy values below 60% (Fig. 6[a]). However, the accuracy increased to 90% after January 2017 and remained high for the rest of the cross-validation. This indicated the model might be trained sufficiently to catch most of the individual variation in MF under different production stages. For the prediction of DMY (Fig. 6[b]), the accuracy fluctuated around 10%, but the mean accuracy was higher than 90%. The prediction of PC had greater accuracy than the other three targets, at around 98% (Fig. 6[c]). The second highest accuracy was achieved by the prediction of FC (95%), even though the accuracy before January 2015 was less than 90% (Fig. 6[d]). Figure 7 displays the accuracy drift as prediction moves forward from +1st to +28th day. The decline in accuracy was −4.17% for DMY prediction (Fig. 7[b]), whilst for other predictions, it was less than 1% (Fig. 7[a], [c], and [d]).

Figure 8 shows the accuracy of MF prediction using XGBOOST classification. The prediction was 100% accurate in identifying the day without any milking (i.e. MF equals 0 times $d^{-1}$), and an accuracy of around 70% to predict the day with MF equal to 3. For other MF levels, the lower accuracy might still be caused by the limited amount of data for MF = 1, 2, 4 and 5 times $d^{-1}$ to allow the model to predict such low levels of MF reliably. Regressions between the predicted and actual values of DMY, PC and FC were plotted in Fig. 9[a], [c] and [e], for which $R^2$ were all greater than 0.90. However, as the data size of this study was more than 1 million, the scatter points still did not fully converge on the 1:1 line (the red dashed line). Therefore, the distribution of residuals was applied as an extra test of the accuracy, as shown in Fig. 9[b], [d] and [f]. From these distributions, it was found that the prediction of DMY, PC and FC using XGBOOST regressions will usually overestimate the extremely low values and underestimate the extremely high values. However, within 95%

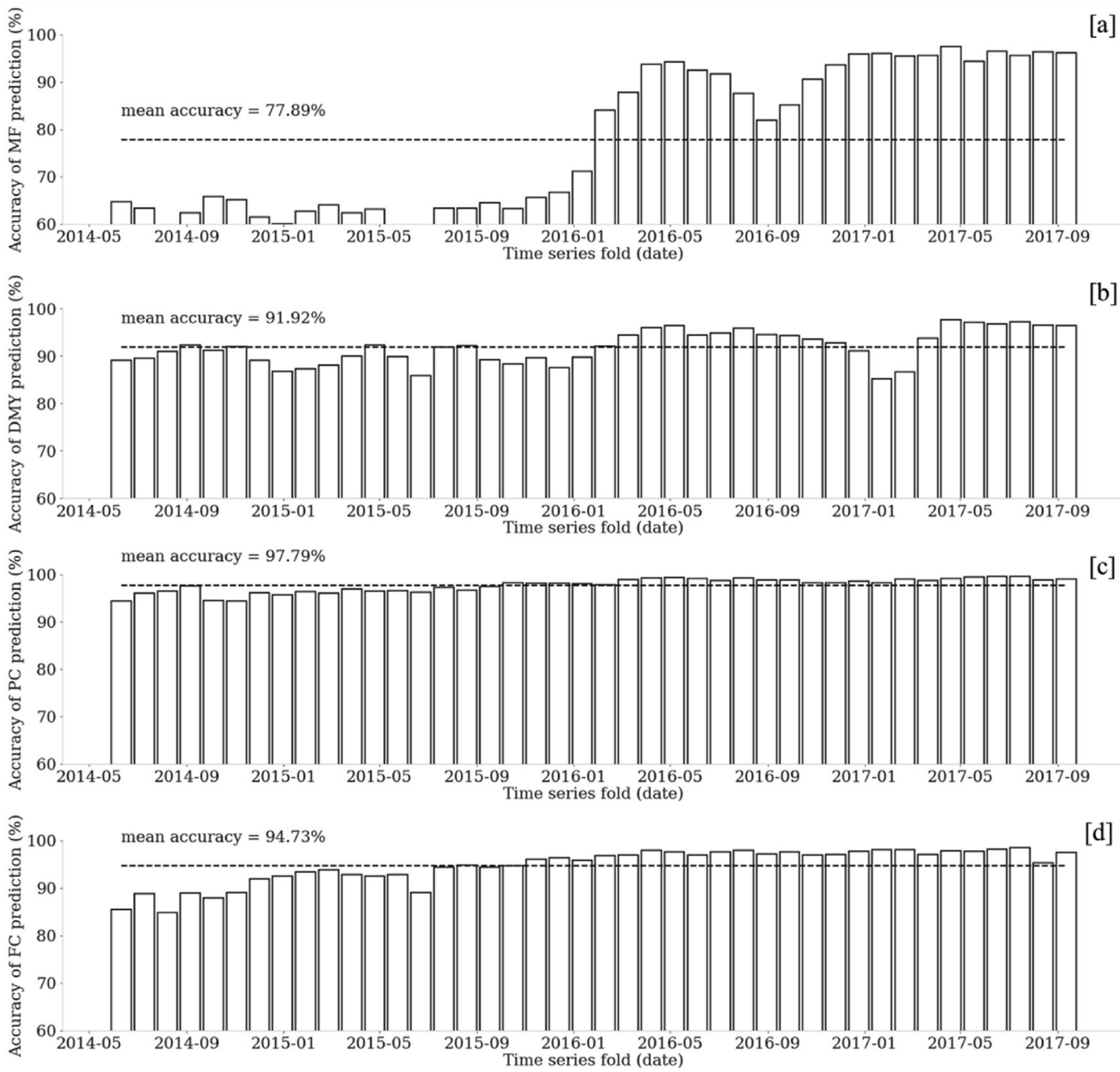


**Fig. 6 – Framework accuracy in cross-validation from June 2014 to September 2017. [a] MF prediction, mean accuracy 77.89%; [b] DMY prediction, mean accuracy 91.92%; [c] PC prediction, mean accuracy 97.79%; and [d] FC prediction, mean accuracy 94.73%.**

proportion of the predictions, as highlighted by the dashed rectangle, mean residuals could be constrained to an acceptable level, which we deemed to be less than 5 kg $cow^{-1}$ $d^{-1}$ for DMY prediction and less than 0.5% for the prediction of PC and FC.

### 3.3. *Feature importance*

The average importance of 31 features in XGBOOST models (i.e. 1176 models for each type of prediction among MF, DMY PC and FC) were calculated and ranked (from #1 to #31) as shown in Fig. 10. The features with higher values and lower ranks had greater power to explain the variation of the target in models. In most of these models, cow ID was identified with the highest value of importance, except the model for DMY prediction, where cow Age achieved slightly higher importance than cow ID. This indicated that the variation of production and behaviour between individual cows is the key driver in training models, such as this one, which have a relatively small herd size.

For MF prediction, DMI, Age and BM had similar importance and were ranked as 2nd, 3rd, and 4th, followed by the 5th most important feature – RT. For DMY prediction, the seasonality (sin[Month]) and lactation status (DIM) had close importance values, which were slightly higher than for BM. For PC prediction, all features other than cow ID had similarly low importance, while for FC prediction the feed intake (DMI) was the 2nd most important feature, with values higher than the other features. The other three features among the top five in FC models were cos(Month), MT and BM.

Despite the $VP_{-60}$ ranking in the 5th position in the PC model (but with comparatively low value), none of the climate features were selected in the top 5 features in the XBOOST models. Interestingly, the last 60 days' mean historical climate features (features with subscript "-60", such as $THI_{min,-60}$) all achieved a higher rank than the corresponding daily climate

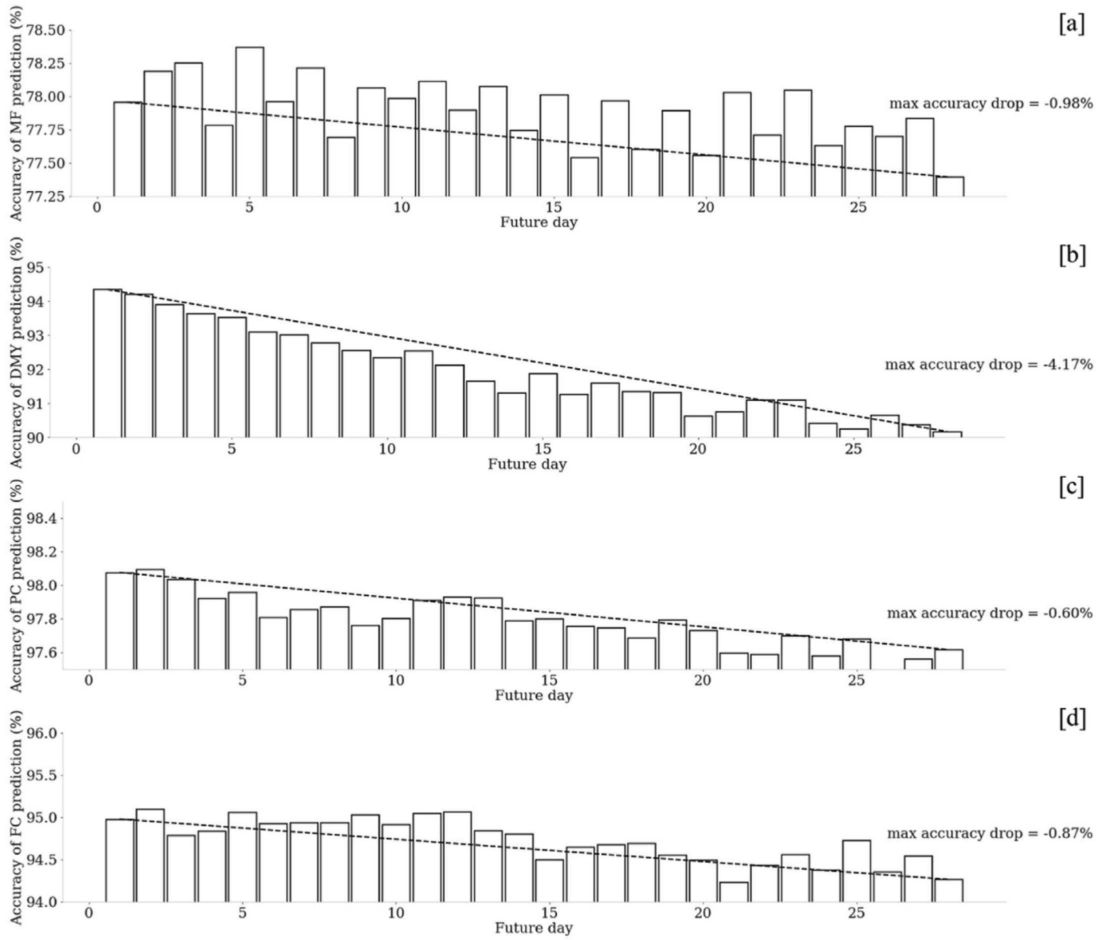


**Fig. 7 – Framework accuracy drift when moving forward from +1 to +28 days in the future. [a] milking frequency (MF) prediction, the maximum (max) accuracy drop was −0.98%; [b] daily milk yield (DMY) prediction, the maximum accuracy drop was −4.17%; [c] PC prediction, the maximum accuracy drop was −0.60%; and [d] FC prediction, the maximum accuracy drop was 0.87%.**

features (such as $THI_{min}$). Furthermore, THI features ($THI_{max}$, $THI_{min}$, $THI_{max,-60}$ and $THI_{min,-60}$) achieved higher rank than other climate features in the prediction of MF, DMY and FC. For PC prediction, historical humidity related features ($RH_{max,-60}$, $RH_{min,-60}$, $VP_{-60}$, and $VPD_{-60}$) were more important than other climate features. In summary, the general order of feature importance grouped by their category was cow features > seasonality features > historical climate features > daily climate features.

## 4. Discussion

### 4.1. *Framework performance*

This study was conducted to demonstrate how to build a framework using machine learning algorithms to predict cow related variables. This is different to analytical modelling used in other studies, where the main aim was to find the relationship between different variables (Fuentes et al., 2020; Ji et al., 2019a, 2019b; Ji, Banhazi, Ghahramani, et al., 2020a, 2020b). This AI based predictive modelling is focused on reliably and accurately predicting variables under practical conditions specific to a particular farm. This contrasts with previous attempts in the last century to model lactation in the cow, which focused on generic models that could be applied across locations (Boujenane & Hilal, 2012; Santos & Silvestre, 2008; Torshizi, 2016). Such individual farm models are likely to be highly accurate due to their ability to take into account the many localised factors, e.g. the weather, affecting the cow's milk production characteristics. These individual models represent the next generation models that are made possible because of the collection of large amounts of data on individual farms.

This study applied time series cross-validation to evaluate the models generated by the framework. This type of cross-validation ensured the model received no future information in its training, and the prediction of future targets was only dependent on the historical information. Other cross-validations such as random or kFold split might have included leakage of future information between training and test dataset, which could over fit the model or overestimation of the performance (Roberts et al., 2017). Overfitting is a well known potential weakness of all predictive models as

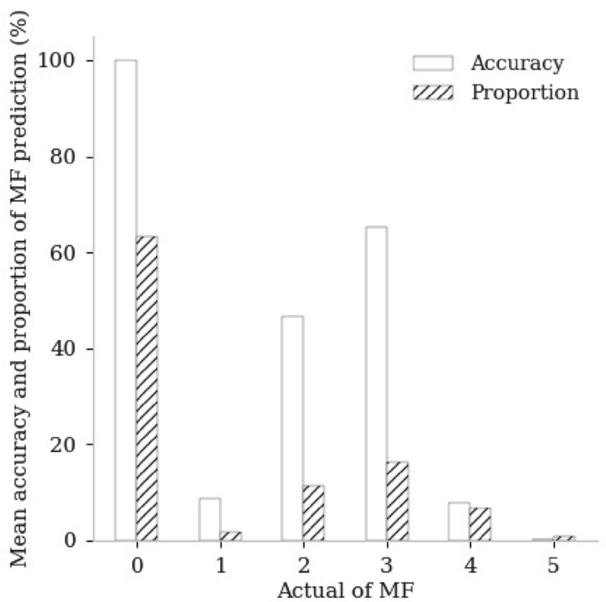


**Fig. 8 – Proportion of milking frequencies in 80 cows and the accuracy of prediction of each milking frequency (MF), using XGBOOST classifications.**

described by Roberts et al. (2017). By using time series cross-validation, the trends and fluctuations of accuracy were determined (Fig. 7). For the prediction of MF, DMY and FC, the accuracy of prediction kept increasing as time advanced, which demonstrates that the framework performs better if there was more data to train the models. This trend of accuracy increase over time was most obvious in the cross-validation of MF prediction. Moreover, the prediction of MF and DMY had about a 10% drop of accuracy between September 2016 and January 2017. By checking with the farm management records, the drop of accuracy might be related to introducing new cows to the herd in these months, thus the framework had to re-train models to catch the production or behaviour patterns of these new cows. A clear rebound of accuracy occurred after a few months, indicating the properties of any new cows, or other factors influencing accuracy, were internalised or absorbed by the machine learning models. The overall accuracy of this framework was demonstrated as reliable in 95% of the predictions (Figs. 8 and 9) from the time series cross-validation. However, for extreme target values or target values with low data size, it was difficult to provide an accurate prediction.

### 4.2. *Model interpretation*

Feature importance was calculated and ranked to understand the tree-based models and enable interpretations. Some interpretations had relatively high agreement with the analytical modelling in previous studies. In XGBOOST models, cow features (i.e. BM, Age, DIM, MT, and DMI) had higher importance than climate features. This was consistent with the

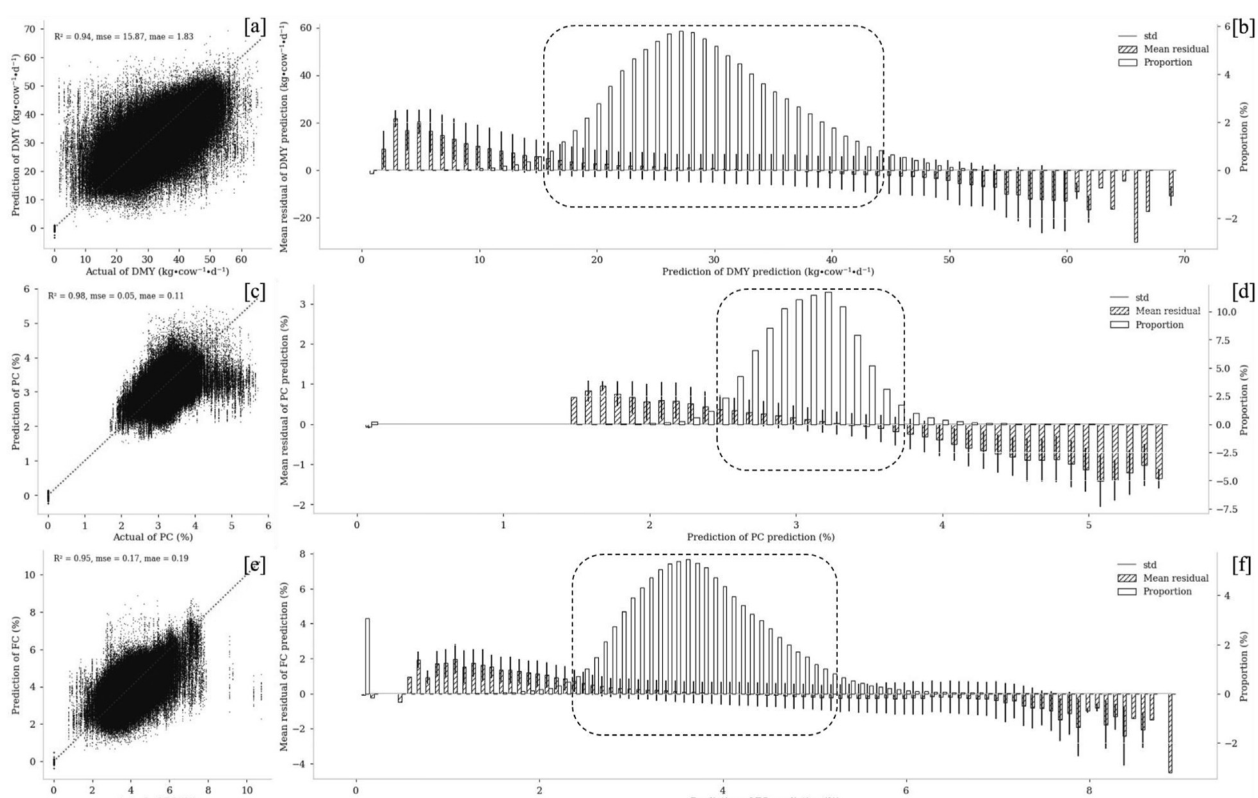


**Fig. 9 – Overall framework accuracy and residuals distribution within XGBOOST regressions. [a] regression between actual and predicted DMY, $R^2 = 0.94$; [b] distribution of residuals in daily milk yield (DMY) prediction; [c] regression between actual and predicted protein content (PC), $R^2 = 0.98$; [d] distribution of residuals in PC prediction; [e] regression between actual and predicted fat content (FC), $R^2 = 0.95$; and [f] distribution of residuals in FC prediction.**

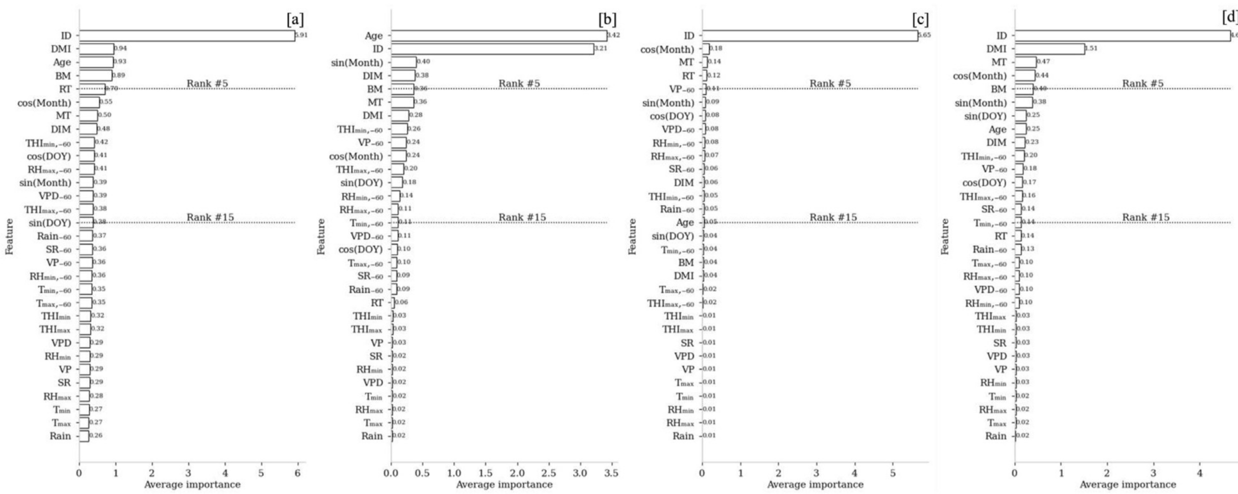


**Fig. 10 – Average feature importance values in the XGBOOST regression models. [a] MF prediction; [b] DMY prediction; [c] PC prediction; and [d] FC prediction.**

results of multiple linear regression models reported by Ji, Banhazi, Ghahramani, et al. (2020b), where the $R^2$ of DMY modelling with cow features was 0.77 and only increased by 0.03 when including climate features. Moreover, it has been reported in a number of previous studies that DMY was primarily based on cow lactation status (DIM), (i.e. the lactation curve), while other impacts might only adjust the fluctuation of that curve instead of influencing the primary trend (Amin, 2003; Grzesiak, Błaszczyk, & Lacroix, 2006; Sharma et al., 2007). In our study, cow age and month of the year both had greater importance than lactation stage. Cow age may have assumed more importance in modern intensive production systems, because cows only last a few years in the herd, with pronounced changes in production between years. Month may have been important because of the nutrition changes with time in this herd and also heat stress effects, noting that temperatures exceeded the upper critical temperature for the cattle on most days in summer. Previous studies have demonstrated that milk production could be highly correlated with the joint influence of month of the year and age, as well as the age at first lactation (Miller, Lentz, & Henderson, 1970; Pirlo, Miglior, & Speroni, 2000).

Historical climate features were found to be more important than daily climate features, which suggests a lag or cumulative heat stress (Ji, Banhazi, Ghahramani, et al., 2020b) or the potential for heat stress starting in the dry off period (Fabris et al., 2019) to have greater impact on productivity. However, the greater importance of historic climate features could also stem from high variation between days, which is smoothed by taking data from more days. To simplify the process of this study, historical climate features were not generated using the equations developed by Ji, Banhazi, Ghahramani, et al. (2020b), i.e. this study did not assign weight to different levels of heat stress determined by a THI calculation, which could be involved in future development. The high importance of DMI in FC prediction demonstrates the inextricable link between fibre consumption and FC (Oba & Allen, 1999). Unfortunately, due to the limited information about feed intake and nutritional quality of the diet collected for this study, it was impossible to do further analysis by replacing DMI with more specific nutrients.

### 4.3. *Potential applications*

In this study, the framework provided an acceptable drift of accuracy when predicting MT, DMY, PC and FC from +1st to +28th day, as well as an overall reliable accuracy within the cross-validation. The prediction generated by this framework is different from that produced by traditional lactation curves (Adediran, Ratkowsky, Donaghy, & Malau-Aduli, 2012; Grzesiak et al., 2006; Jones, 1997; Kokate, Dongre, Khandait, & Kale, 2019; Olori, Brotherstone, Hill, & McGuirk, 1999) or test-day analytical models (Ji et al., 2019a, 2019b; Ji, Banhazi, Ghahramani, et al., 2020a, 2020b). It is more dynamic, individual-cow-oriented and relatively long-term, which will facilitate many innovative applications. It can be used by farmers to identify a cow experiencing heat stress or health issues when her production or behaviour (milking time) is different from the prediction. Therefore, precise treatment can be given to individuals, such as nutrient adjustment (Wheelock, Rhoads, Vanbaale, Sanders, & Baumgard, 2010) or misting in RMS for cooling (Fuentes et al., 2020). Moreover, it can be used to generate a simulation of production and behaviour change at herd level after implementing any farm management strategy, such as installing a new ventilation system. It can be used to estimate the effect of changes in DMI under local conditions, without resort to feeding standards, and investigations of optimum milking times. Therefore, the farm manager can estimate the effectiveness of their decisions in advance. Furthermore, by embedding such a model into a bigger scale framework, as introduced by Ji, Banhazi, Perano, et al. (2020), a potential application may include forecasting the milk price or helping genetic selection. Incorporating milk conductivity and cow stepping data into such models will enable better prediction of mastitis and oestrus, respectively.

## 5. Conclusions

This study developed a framework using XGBOOST models to predict MF, DMY, PC and FC of individual cows in advance, within a 28-day period. The framework was evaluated using time series cross-validation. It produced models that were able to provide reliable predictions with high accuracy and stable low residuals. The interpretation of feature importance of XGBOOST models also mostly agreed with the results of previously developed analytical models (Ji et al., 2019a, 2019b; Ji, Banhazi, Ghahramani, et al., 2020a, 2020b). Although the framework still had limitations and could be further improved, its potential usage can provide many benefits for RMS or routine dairy farm management. Further validation of the models could include comparisons with models created on other farms and a sensitivity analysis.

## 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.

## Acknowledgement

This research was supported by a research scholarship granted by the University of Southern Queensland. The authors would also like to thank the technical assistance provided by the farmer and his staff on the study farm. The measurements and procedures used in this project (ID: 17REA007) were approved by the University of Southern Queensland's Animal Ethics Committee.

## REFERENCES

Adediran, S. A., Ratkowsky, D. A., Donaghy, D. J., & Malau-Aduli, A. E. O. (2012). Comparative evaluation of a new lactation curve model for pasture-based Holstein-Friesian dairy cows. *Journal of Dairy Science, 95*(9), 5344–5356. https://doi.org/10.3168/jds.2011-4663

Amin, A. A. (2003). Test-day model of daily milk yield prediction across stages of lactation in Egyptian buffaloes. *Archives Animal Breeding, 46*(1), 35–45.

Banhazi, T. M., Lehr, H., Black, J., Crabtree, H., Schofield, P., Tscharke, M., et al. (2012). Precision Livestock Farming: An international review of scientific and commercial aspects. *International Journal of Agricultural and Biological Engineering, 5*(3), 1–9.

Bear, C., & Holloway, L. (2019). Beyond resistance: Geographies of divergent more-than-human conduct in robotic milking. *Geoforum, 104*, 212–221.

Boujenane, I., & Hilal, B. (2012). Genetic and non genetic effects for lactation curve traits in Holstein-Friesian cows. *Archives Animal Breeding, 55*(5), 450–457.

Casalicchio, G., Molnar, C., & Bischl, B. (2018). Visualizing the feature importance for black box models. In *Paper presented at the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Paper presented at the Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*.

Dekkers, J., Ten Hag, J., & Weersink, A. (1998). Economic aspects of persistency of lactation in dairy cattle. *Livestock Production Science, 53*(3), 237–252.

Fabris, T. F., Laporta, J., Skibiel, A. L., Corra, F. N., Senn, B. D., Wohlgemuth, S. E., et al. (2019). Effect of heat stress during early, late, and entire dry period on dairy cattle. *Journal of Dairy Science, 102*(6), 5647–5656. https://doi.org/10.3168/jds.2018-15721

Fang, Q., Hanna, M., Haque, E., & Spillman, C. (2000). Neural network modeling of energy requirements for size reduction of wheat. *Transactions of the ASAE, 43*(4), 947.

Fenlon, C., O'Grady, L., Dunnion, J., Shalloo, L., Butler, S., & Doherty, M. (2016). *A comparison of machine learning techniques for predicting insemination outcome in Irish dairy cows*.

Fuentes, S., Gonzalez Viejo, C., Cullen, B., Tongson, E., Chauhan, S. S., & Dunshea, F. R. (2020). Artificial intelligence applied to a robotic dairy farm to model milk productivity and quality based on cow data and daily environmental parameters. *Sensors, 20*(10), 2975.

Grzesiak, W., Błaszczyk, P., & Lacroix, R. (2006). Methods of predicting milk yield in dairy cows—Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Computers and Electronics in Agriculture, 54*(2), 69–83.

Grzesiak, W., Zaborski, D., Szatkowska, I., & Królaczyk, K. (2021). Lactation milk yield prediction in primiparous cows on a farm using the seasonal auto-regressive integrated moving average model, nonlinear autoregressive exogenous artificial neural networks and Wood's model. *Animal Bioscience, 34*(4), 770.

Hart, K., McBride, B., Duffield, T., & DeVries, T. (2013). Effect of milking frequency on the behavior and productivity of lactating dairy cows. *Journal of Dairy Science, 96*(11), 6973–6985.

Hermans, G., Ipema, A., Stefanowska, J., & Metz, J. (2003). The effect of two traffic situations on the behavior and performance of cows in an automatic milking system. *Journal of Dairy Science, 86*(6), 1997–2004.

Heyden, T. (2015). *The cows that queue up to milk themselves*. BBC News Magazine.

Hsieh, C.-L., Hung, C.-Y., & Kuo, C.-Y. (2011). Quantization of adulteration ratio of raw cow milk by least squares support vector machines (LS-SVM) and visible/near infrared spectroscopy. In *Engineering applications of neural networks* (pp. 130–139). Springer.

Jakobsen, J. H., Madsen, P., Jensen, J., Pedersen, J., Christensen, L., & Sorensen, D. (2002). Genetic parameters for milk production and persistency for Danish Holsteins estimated in random regression models using REML. *Journal of Dairy Science, 85*(6), 1607–1616.

Ji, B., Banhazi, T., Ghahramani, A., Bowtell, L., Wang, C., & Li, B. (2019a). Modelling of heat stress in a robotic dairy farm. Part 1: Thermal comfort indices as the indicators of production loss. *Biosystems Engineering, 199*(1), 27–42.

Ji, B., Banhazi, T., Ghahramani, A., Bowtell, L., Wang, C., & Li, B. (2019b). Modelling of heat stress in a robotic dairy farm. Part 2: Identifying the specific thresholds with production factors. *Biosystems Engineering, 199*(1), 43–57. https://doi.org/10.1016/j.biosystemseng.2019.11.005

Ji, B., Banhazi, T., Ghahramani, A., Bowtell, L., Wang, C., & Li, B. (2020a). Modelling of heat stress in a robotic dairy farm. Part 3: Rumination and milking performance. *Biosystems Engineering, 199*(1), 58–72. https://doi.org/10.1016/j.biosystemseng.2020.02.006

Ji, B., Banhazi, T., Ghahramani, A., Bowtell, L., Wang, C., & Li, B. (2020b). Modelling of heat stress in a robotic dairy farm. Part 4:

Time constant and cumulative effects of heat stress. *Biosystems Engineering, 199*(1), 73–82. https://doi.org/10.1016/j.biosystemseng.2020.07.014
Ji, B., Banhazi, T., Perano, K., Ghahramani, A., Bowtell, L., Wang, C., et al. (2020c). A review of measuring, assessing and mitigating heat stress in dairy cattle. *Biosystems Engineering, 199*(1), 4–26. https://doi.org/10.1016/j.biosystemseng.2020.07.009
Jones, T. (1997). Empirical Bayes prediction of 305-day milk production. *Journal of Dairy Science, 80*(6), 1060–1075. https://doi.org/10.3168/jds.S0022-0302(97)76031-4
Jorquera-Chavez, M., Fuentes, S., Dunshea, F. R., Warner, R. D., Poblete, T., & Jongman, E. C. (2019). Modelling and validation of computer vision techniques to assess heart rate, eye temperature, ear-base temperature and respiration rate in cattle. *Animals, 9*(12), 1089.
Kettunen, A., Mäntysaari, E., Strandén, I., Pösö, J., & Lidauer, M. (1998). Estimation of genetic parameters for first lactation test day milk production using random regression models. In *Paper presented at the World Congress of Genetics Applied to Livestock Production*.
Klaas, I. C., Rousing, T., Fossing, C., Hindhede, J., & Sorensen, J. (2003). Is lameness a welfare problem in dairy farms with automatic milking systems? *Animal Welfare-Potters Bar then Wheathampstead-, 12*(4), 599–604.
Klopčič, M., Malovrh, Š., Gorjanc, G., Kovač, M., & Osterc, J. (2003). Prediction of daily milk fat and protein content using alternating (AT) recording scheme. *Czech Journal of Animal Science, 48*(11), 449–458.
Kokate, L., Dongre, V., Khandait, V., & Kale, S. (2019). Modeling of lactation curves for prediction of standard lactation milk yield in Marathwadi buffaloes. *Indian Journal of Animal Sciences, 89*(8), 909–911.
Kominakis, A., Abas, Z., Maltaris, I., & Rogdakis, E. (2002). A preliminary study of the application of artificial neural networks to prediction of milk yield in dairy sheep. *Computers and Electronics in Agriculture, 35*(1), 35–48.
Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18–22.
Mayeres, P., Stoll, J., Bormann, J., Reents, R., & Gengler, N. (2004). Prediction of daily milk, fat, and protein production by a random regression test-day model. *Journal of Dairy Science, 87*(6), 1925–1933.
McCullough, C. (2019). Global dairy outlook by 2030: Industry. *The Dairy Mail, 26*(11), 12–15.
McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing, 14*(9).
McParland, S., Coughlan, B., Enright, B., O'Keeffe, M., O'Connor, R., Feeney, L., et al. (2019). Prediction of 24-hour milk yield and composition in dairy cows from a single part-day yield and sample. *Irish Journal of Agricultural and Food Research, 58*(1), 66–70.
Miles, J. (2014). R squared, adjusted R squared. In *Wiley StatsRef: Statistics Reference Online*.
Miller, P., Lentz, W., & Henderson, C. (1970). Joint influence of month and age of calving on milk yield of Holstein cows in the northeastern United States. *Journal of Dairy Science, 53*(3), 351–357.
Oba, M., & Allen, M. (1999). Evaluation of the importance of the digestibility of neutral detergent fiber from forage: Effects on dry matter intake and milk yield of dairy cows. *Journal of Dairy Science, 82*(3), 589–596.
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 Production Science, 58*(1), 55–63.
Pastell, M., & Madsen, H. (2008). Application of CUSUM charts to detect lameness in a milking robot. *Expert Systems with Applications, 35*(4), 2032–2040.
Pirlo, G., Miglior, F., & Speroni, M. (2000). Effect of age at first calving on production traits and on difference between milk yield returns and rearing costs in Italian Holsteins. *Journal of Dairy Science, 83*(3), 603–608.
Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography, 40*(8), 913–929.
Rodenburg, J. (2012). The impact of robotic milking on milk quality, cow comfort and labor issues. In *Natl. Mastitis Counc. Annu. Meet. Proc. St. Pete Beach, FL* (pp. 126–137). Madison, WI: Natl. Mastitis Counc.
Rodenburg, J., & Wheeler, B. (2002). Strategies for incorporating robotic milking into North American herd management. In *Paper presented at the Proceedings of the first North American Conference on Robotic Milking, Toronto, Canada* (pp. III-18–III-32).
Rubinfeld, D. L. (1998). *Econometric models and economic forecasts*. Irwin Professional Pub.
Santos, A., & Silvestre, A. (2008). A study of Lusitano mare lactation curve with Wood's model. *Journal of Dairy Science, 91*(2), 760–766.
Schaeffer, L. (1994). Random regressions in animal models for test-day production in dairy cattle. In *Paper presented at the World Congress of Genetics Applied Livestock Production, 1994*.
Sharma, A. K., Sharma, R., & Kasana, H. (2007). Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. *Applied Soft Computing, 7*(3), 1112–1120.
Simões Filho, L. M., Lopes, M. A., Brito, S. C., Rossi, G., Conti, L., & Barbari, M. (2020). Robotic milking of dairy cows: A review. *Semina: Ciências Agrárias, 41*(6), 2833–2850.
Speroni, M., Pirlo, G., & Lolli, S. (2006). Effect of automatic milking systems on milk yield in a hot environment. *Journal of Dairy Science, 89*(12), 4687–4693.
Swalve, H. (2000). Theoretical basis and computational methods for different test-day genetic evaluation methods. *Journal of Dairy Science, 83*(5), 1115–1124.
Torshizi, M. E. (2016). Effects of season and age at first calving on genetic and phenotypic characteristics of lactation curve parameters in Holstein cows. *Journal of Animal Science and Technology, 58*(1), 1–14.
Van Rossum, G. (2007). Python programming language. In *Paper presented at the USENIX annual technical conference*.
Wheelock, J. B., Rhoads, R. P., Vanbaale, M. J., Sanders, S. R., & Baumgard, L. H. (2010). Effects of heat stress on energetic metabolism in lactating Holstein cows. *Journal of Dairy Science, 93*(2), 644–655. https://doi.org/10.3168/jds.2009-2295
Wildridge, A. M., Thomson, P. C., Garcia, S. C., John, A. J., Jongman, E. C., Clark, C. E., et al. (2018). The effect of temperature-humidity index on milk yield and milking frequency of dairy cows in pasture-based automatic milking systems. *Journal of Dairy Science, 101*(5), 4479–4482.
Yan, W., Chen, X., Akcan, O., Lim, J., & Yang, D. (2015). Big data analytics for empowering milk yield prediction in dairy supply chains. In *Paper presented at the 2015 IEEE International Conference on Big Data (Big Data)*.
Zien, A., Krämer, N., Sonnenburg, S., & Rätsch, G. (2009). The feature importance ranking measure. In *Paper presented at the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*.