ACTA IMEKO
ISSN: 2221-870X
March 2024, Volume 13, Number 1, 1 - 7

Short review of current limits and challenges of application of machine learning algorithms in the dairy sector

Lucia Trapanese[1], Miel Hostens[2], Angela Salzano[3], Nicola Pasquino[1]

[1] *Department of Electrical Engineering and Information Technologies, University of Naples Federico II, Italy*
[2] *Department of Population Health Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, Netherlands*
[3] *Department of Veterinary Medicine and Animal Production, University of Naples Federico II, Naples, Italy*

ABSTRACT
In the last years, the livestock sector is moving towards a more sustainable animal-based industry, mitigating the environmental impact of livestock while meeting the demand for high-quality food. To achieve these goals, farms are using a more technological approach, adopting algorithms to manipulate the vast amount of data from sensors and routine operations. The results will be useful for making more objective decisions. In this context, machine learning – a branch of Artificial Intelligence applied to the study of prediction, inference, and clustering algorithms - can be successfully employed. Nowadays, machine learning algorithms are successfully used to solve many issues in the livestock sector, such as early disease detection, and they are expected to be employed in the future for welfare monitoring. This brief review gives an overview of the current state of the art of the most popular applications for dairy science and the most widely used and best-performing algorithms, highlighting the challenges and obstacles for broad acceptance of these techniques in the dairy sector.

Section: RESEARCH PAPER

Keywords: Precision livestock farming; machine learning; dairy sector

Citation: L. Trapanese, M. Hostens, A. Salzano, N. Pasquino, Short review of current limits and challenges of application of machine learning algorithms in the dairy sector, Acta IMEKO, vol. 13 (2024) no. 1, pp. 1-7. DOI: 10.21014/actaimeko.v13i1.1725

Section Editor: Leopoldo Angrisani, Università degli Studi di Napoli Federico II, Naples, Italy

Received December 6, 2023; **In final form** March 4, 2024; **Published** March 2024

Funding: This study was carried out within the Agritech National Research Center and received funding from the European Union Next-Generation EU (Piano Nazionale di Ripresa e Resilienza (PNRR) Missione 4 Componente 2, Investimento 1.4 D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors views and opinions; neither the European Union nor the European Commission can be considered responsible for them.

Corresponding author: Lucia Trapanese, e-mail: lucia.trapanese2@unina.it

1. INTRODUCTION

Nowadays, the livestock sector is facing many challenges, such as making animal-based food production more sustainable and efficient, mitigating the environmental impact of livestock, and satisfying the demand for high-quality food. To achieve these goals, the sector is moving toward a more technological approach by introducing a plethora of different solutions, such as *i*) automated milking systems to control milk quantity and quality, *ii*) electric feeders able to give the precise amount of food required, *iii*) wearable or environmental sensors to monitor animal health and welfare. These systems allow scientists, technicians, and breeders to control most aspects of the herd. This new approach is called precision livestock farming (PLF).

According to Berckmans *et al.* [1], PLF is the "management of livestock by continuous automated real-time monitoring of production/reproduction, health and welfare of livestock, and its environmental impact". The adoption of new technologies in this field, such as embedded and wearable sensors, biosensors, and digital imaging systems, gives the opportunity to collect a large amount of data regarding animal conditions and behaviour. To manage this amount of information, a classical statistical approach would not represent a suitable solution. Indeed, this approach does not work properly with "big data" as it was designed for a few input variables and sample sizes [2]. Moreover, a priori hypothesis and knowledge of the data and observed phenomena are needed to build a useful statistical model [3].

In this context, machine learning (ML) can successfully overcome these limitations. ML is a branch of Artificial Intelligence applied to studying algorithms for forecasting, inference, and clustering. ML techniques can manage big data and are quite adaptable with non-linear, complex, noisy, and imprecise datasets such as dairy production (production, reproduction, and animal welfare) and sensors data.

These techniques could efficiently extract useful information from fully automated data recording systems [4]. Among all sub-categories of ML, unsupervised and supervised learning represent the traditional ML techniques that are very common categories of algorithms in livestock application [5]. The major difference between them is the presence of labelled information in the supervised method. The supervised algorithms aim to classify or predict a certain value depending on whether the variable is categorical or numerical [6]. This approach allows for managing, processing, and detecting patterns and correlations among complex and unrelated data to develop decision support systems useful for PLF. ML has been used in many domestic species, such as pigs [7], poultry [8], beef [9], and dairy cattle [10]. Despite the growing interest in all these sectors, this work focuses on dairy cattle and their food production, an industry that has started to use many innovative technologies.

Indeed, Shine *et al.* [11], in their review, identified the agriculture sector as one in which the techniques were most advanced and, ML approaches most widespread. There are already some papers describing the use of ML in the dairy sector.

For example, Benos *et al.* [12] focused their work on applying the best ML algorithms to improve the management of water resources, crop, and soil. The authors also assessed the diffusion of this data-driven approach in livestock. More specifically, they summarised the ML algorithms employed in the livestock species and their performances and found out that most research papers were focused on ML applications in cattle, with secondary attention given to sheep and goats. The most of tasks were related to livestock management and animal welfare assessment. Among the ML techniques utilized, the most impressive performances were observed with artificial neural networks (ANN), ensemble methods, and Support Vector Machines (SVM).

Garcia *et al.* [13] showed a geographic overview of the ML approach and described the main environmental and animal health monitoring applications. Moreover, the work discussed the main challenges to developing ML in livestock. Although a lot of work describes the state of the art of ML algorithms in this field, to the author's knowledge, there are no papers that clearly express the current limitations of ML techniques and the general obstacles in disseminating this computational approach.

This work offers a different point of view based on the current issues and problems in ML application in dairy cow breeding. This work aims to summarise the ML state of the art and propose alternative solutions to overcome current limitations.

To highlight the main trends and provide a clear overview of this topic, this work is arranged as follows:

- Description of the main applications of unsupervised ML in the dairy sector;
- Description of the main application of supervised ML in the dairy sector;
- Overview of the main limitations and problems for both techniques.

2. TRENDS IN MACHINE LEARNING APPLICATIONS

Since the variability of livestock data, different algorithms have been developed. A preliminary analysis showed that supervised algorithms were more widely employed than unsupervised [14]. Supervised ML can predict several characteristics of animal products, recognize sick animals and timely detect outliers. Unsupervised techniques can be employed for cluster analysis of the herd, find hidden patterns, and carry out exploratory analysis. Finally, ML, with different approaches, aimed to increase the management, reproductive, and productivity performances of dairy (e.g. discriminating sick animals from healthy ones or productive from unproductive ones). In addition, ML is employed in combination with near-infrared spectroscopy (NIR) or electronic noses to recognize and predict the quality of raw material, food fraud, diet components, composition of manure, and diseases [15].

In the following sections, the main data source, algorithms, and applications of unsupervised and supervised techniques are described.

3. UNSUPERVISED LEARNING TECHNIQUES

Cluster analysis in livestock is a valuable tool to detect hidden patterns among animals or herds and describe their performances. Distance-based algorithms are very suitable for this task. K-means, hierarchical, and principal component analysis (PCA) are the most widespread techniques for grouping herds or animals with similar performances. In this work, the main applications of unsupervised ML algorithms were grouped into two classes:

- Cluster analysis of herd data;
- Cluster analysis of sensors data.

3.1. Analysis of herd data

Brotzman *et al.* [16] employed two unsupervised techniques (PCA and hierarchical) to cluster 557 herds based on 22 features. The PCA analysis was used to find a meaningful subset of variables in the routine data. The algorithm identified an optimal subset of 16 variables. Among them, the risk of subclinical intramammary infection at first examination, age at first calving, the rate of cure of intramammary infection in the dry period, and the number of days in milk represented the most important variables to describe the herds. Then, the agglomerative hierarchical algorithm with Ward's linkage method was employed to cluster the herds. To decide the optimal number of clusters, the hierarchical clustering with Ward's linkage and Fourier plots, maximum simple structure index, and minimum error sum of squares optimization, using the methodology of Borcard, were employed. Based on these analyses, the hierarchy was cut to obtain six clusters. The algorithm grouped the herds based on their size, milk frequencies, milk production, and reproduction performance. The results showed that four clusters differed in herd size and milking frequencies. One cluster grouped the herd with high somatic cell count, while herds with the worst production performance were in the last cluster. This analysis returned a broad overview, useful for discovering trends in performance characteristics in large Upper Midwest dairy herds.

Alessio *et al.* [17] employed cluster analysis through a hierarchic approach with the Ward linkage. The aim was to assess the influence of lactose on the somatic cell count and other milk components such as fat, protein, and total bacterial count. The analysis returned three clusters. In the first cluster, an increase in the somatic cells was associated with a reduction in lactose. Cluster three showed decreased lactose content associated with increased somatic cells and total bacterial count. Using this technique, it was possible to identify the cluster with a negative correlation between somatic cells and lactose confirming previous results [18].

3.2. Analysis of sensors data

Shahriar *et al.* [19] used k-means to detect estrus in dairy cows. This work proposed an innovative approach based on the clustering of the time series of a wearable accelerometer located on the cow's collar.

The authors were able to group activity in high, middle, and low levels. The results showed that the clustering algorithm properly grouped the observation with an accuracy of 82% - 100% compared to the ground truth (presence of estrus).

In summary, the main application of unsupervised learning was employed for the exploration of hidden patterns and natural clusters of animals or herds. The original approach employed data regarding wearable sensors to classify dairy cows' activity and, in general, behaviour. Nowadays, the main sources of routinely collected data are herd records, milk recording, fertility, and clinical parameters.

4. SUPERVISED LEARNING TECHNIQUES

Supervised learning encompasses several useful techniques. These techniques have been employed for tasks such as monitoring the behaviour and location of animals, early detection of diseases, outlier detection, and finally, the prediction of milk yield and environmental impact of herds and animals [20].

To provide a clear overview and describe the different applications for each topic, this work grouped the supervised learning algorithms into four general fields of applications:

- Behaviour recognition, classification, and analysis;
- Disease detection;
- Prediction of performance;
- Food analysis.

4.1. Behaviour recognition, classification, and analysis

Monitoring cows' behaviour provides information about the animal's health and physiological conditions.

In the last few years, the availability and low costs of embedded sensors and monitoring devices allowed the recording of data on cow activity [21]. Neck collars and pedometers represent the main data source for behaviour recognition. The latter devices are accelerometers that provide the acceleration value on the three axes. Data from this device are used by supervised ML to recognize and classify different types of cow behaviours such as rumination, feeding, lying, and standing [22].

For example, Tamura *et al.* [23] assessed the association between observed dairy cattle behavior and collar data by employing the random forest algorithm to recognize three different behaviors: eating, rumination, and lying. The authors collected the data from four different farms using neck collars. The results showed irregular and continuous accelerations for eating behavior in all three directions. During rumination, homogeneous and regular acceleration amplitudes were measured in all three directions. Finally, for lying, acceleration was near zero. The authors applied decision trees on the data set from one farm and used the data sets from the other farms to validate the model. In this application, a high value of precision (99.2%) was achieved.

In another application, Benaissa *et al.* [24] aimed to classify cows' behavior through three supervised approaches: i) Naïve Bayes, ii) k-nearest neighbors (KNN), iii) SVM by comparing leg and neck-mounted accelerometers. In addition, they studied the influence of the sampling rate and the number of accelerometer axes on the final prediction performance. The results showed that the SVM algorithm performed better than the Naïve Bayes and KNN. Moreover, the three axes generally performed better in classification accuracy, while no significant differences were noticed when only one or two axes were used. The differences among three axes accelerometers and other configurations were emphasized when the sensor was mounted on the neck collar. It was also possible to employ neck collars and pedometers to obtain information on the time budget of dairy [25].

Combining the typical pattern behavior of each cow and ML algorithms, several papers focused on detecting anomalies in the health status and timely recognizing diseases or events such as estrus or calving [26], [27]. Wang *et al.* [28] assessed the combination of location, acceleration, and supervised ML to predict estrus. Four ML algorithms: i) KNN, ii) back propagation neural network (BPNN), iii) linear discriminant analysis (LDA), and iv) classification and regression tree (CART) were tested. In addition, the impact of time-windows length was also assessed. The authors found that the BPNN algorithm with a 0.5 h time window better denoted the estrus in cows in terms of sensitivity, precision, accuracy, specificity, and F1-score measure compared to the other algorithms. F1 is the harmonic mean of precision and sensitivity, and ranges between 0 and 1. This measure is very suitable when different algorithms are compared because it allows to maximize both precision and recall [24], [29], [30].

Regarding the prediction of calving, Borchers *et al.* [31] studied the possibility of characterizing the peripartum behavior to predict the calving time starting with the automated activity, lying, and rumination monitoring. They quantified activity, rumination, and lying behavior before calving using two devices: neck collars and leg-mounted accelerometers. Moreover, the authors compared these behaviors to existing literature and determined the calving prediction efficacy of these technologies, both individually and in combination, using three ML algorithms: i) random forest (RF), ii) ANN, and iii) LDA. Results showed that primiparous and multiparous expressed different activity behaviours beginning seven days before calving. Primiparous cattle became more restless before calving. Regarding the ruminating time, all animals showed a decrease two days before calving. For ML application, the best score of specificity, positive predictive values (PPV) and NPV values were achieved by combining both neck collars and leg-mounted accelerometers and ANN.

In conclusion, behaviour recognition is an important task for livestock management and supervised ML is one of the most suitable techniques for these analyses.

4.2. Disease detection

Diseases lead to important economic loss, so their early detection is an important topic in dairy management. Data about calving events, previous disorders, genetics information, milk production, farm characteristics, and other types of information are recorded and employed by farmers or technicians [14]. It was possible to also use data obtained from neck collars and pedometers to assess the presence of disease linked to a change in natural behaviour such as lameness [32].

For example, Warner *et al.* [33] compared different ML approaches such as CART, gradient boosting machine (GBM), extreme gradient boosting (XGB), RF models, and linear regression to predict lameness based on twenty routinely pre-collected variables such as calving interval, somatic cell count and age at first calving. The results showed that the ML approach achieved better results than a linear model, and CART and RF had better performance than the XGB regarding specificity.

Hyde *et al.* [34] provided another example that assessed RF performances for classifying the source (environmental or contagious) of mastitis infections using a large dataset with 1000 British herds and 229 features regarding routinely collected data, herds characteristics and veterinary diagnosis of mastitis infection. In addition, the authors also explored the performance of the ensemble method to classify the moment (lactation or dry period) in which the mastitis occurred. The results showed that the RF achieved a 98% accuracy for the prediction of environmental or contagious mastitis and 78% accuracy for the moment of infection.

Finally, also data from other technologies such as NIR are successfully utilised. Mastitis can be detected from NIR data processed through ML as Ramirez-Morales *et al.* [35] described. They assessed the feasibility of developing a low-cost, real-time, field-applicable tool to detect the presence and the severity of bovine mastitis. Data collections are performed with a low cost and portable NIR and KNN algorithms are employed to perform two different models. The first predicted the presence or absence of mastitis while the second classified the sample according to the severity of the disease. For both applications, high values of accuracy are achieved (91 % and 95 %, respectively).

Zhou *et al.* [36] aimed to explore eight supervised prediction models for predicting naturally occurring health disorders in cows. The features employed for training the supervised models were based on the data obtained from the accelerometer devices located on the cow's neck and on automated milking systems data. The first features provide information on feeding behaviour, while the second ones provide information about the milk yield and quality. Their result showed that, in general, the best performance was achieved by the Rpart algorithm with 93 % precision and 80 % specificity. However, also SVM, RF, and XGB models obtained good performances in terms of specificity (> 80 %) and accuracy (80 %).

In conclusion, for disease detection, ML algorithms achieved good performance for different diseases and could be utilized to improve the management and wellness of animals through the integration of artificial intelligence and dairy cattle welfare, within a computerized decision support tool.

4.3. Prediction of performance

The ML algorithms are also suitable to predict dairy cows' performance in terms of milk yield, quantity of greenhouse gases, amount of pollutant emission or productive longevity of animals. The main data sources are very similar to the previous one and strictly depend on the aim of the task.

Salamone *et al.* [37] aimed to predict the milk yield on the first test day, starting with the routine data of previous lactations. They employed the random forest regression (RFR) algorithm on three different datasets to denote the best configuration for achieving higher results in terms of root mean squared error (RMSE), mean absolute error (MAE), and R^2. The first dataset concerned production features, the second production and herd features, and the third production, herd, and reproduction features. Starting with these datasets, a RFR was trained, and performances were compared. All three models achieved similar performance regarding RMSE, MAE, and R^2. In addition, the variable importance showed the same results for each RFR model: days in milk, cumulative milk yield after 305 days, and milk yield at the fifth and fourth test days are the features that achieved the highest importance score. The dataset employed for the prediction of the environmental impact of dairy cows regards milk production, diet information, and values from different gas sensors.

Regarding the environmental impact, several ML algorithms aimed to predict methane production or manure excretion. In the work of Chen *et al.* [3], the authors compared the performance of multiple linear regression and three regression algorithms: i) ANN, ii) RFR, and iii) support vector regression (SVR) to predict nitrogen excretion from manure in dairy cows. The dataset concerned information about the total diet digestibility, such as nitrogen intake and forage proportion in the diet, and routine data, such as milk yield. The results showed that ANN performed better than other ML algorithms and multiple linear regression (MLR) in terms of RMSE. Moreover, regarding the concordance correlation coefficient (CCC), the results of ANN were better than those from other algorithms. The performance of all models increased after applying the algorithm of feature selection. The results denoted a general higher performance of ANN to explore animal and diet factors influencing manure excretion in lactating dairy cows.

Finally, the longevity of dairy cows is an important topic for breeding purposes. Ouweltjes *et al.* [38] compared the performance of the RF algorithm and ordinal logistic regression to predict the resilience and longevity of dairy cows. The authors used feed and activity behaviour data, lactation data, and information regarding the health treatments, culling, and calving dates. The results showed that the two algorithms had the same accuracy. However, the RF algorithms, compared to the ordinal linear regression needed fewer preprocessing phases, so the suitable most method is the suitable most method.

4.4. Food analysis

Predicting and recognizing dairy food quality and safety are among the main tasks of ML application in livestock. For these aims, ML algorithms usually employ data belonging to NIR or electronic nose.

Muniz *et al.* [39] developed a portable tool that allows farmers to assess in real time the quality of raw milk in terms of protein, fat, lactose, and solids-non-fat (SNF) for each cow. A shallow neural network with two hidden layers was built as a regression model. PCA was performed to reduce the dimensionality. Finally, the predictions of NN were compared with the predictions of linear regression. The results showed that NN performs better than linear regression in terms of bias on protein, fat, and SNF prediction. On the lactose, no difference in bias was found. In addition, NN also provided the smallest measure of dispersion.

Regarding food adulteration, a combination of NIR and ML techniques was employed by Ehsani *et al.* [40]. The aim of the work was to assess the performance of two main classes of ensemble learning performed on three different handheld spectrometers to recognize H_2O_2 and NaClO. They employed Random subspace ensemble k-nearest neighbour (RSE-KNN), KNN, random subspace discriminant ensemble (RSDE), random under sampling-boosted ensemble (RUS-BE), ensemble bagged tree (EBT). Each of the algorithms was performed on Linksquare, Tellspec and Neospectra. The performance was assessed by the percentage of accuracy, specificity, sensitivity and Youden index. The results showed that RSE-KNN had the best performance in terms of accuracy and Youden index for detecting H_2O_2 and NaClO on the spectra (Linksquare and Tellspec 95% - 93% and 96% - 93%, respectively). However, when a mixture of both adulterants occurred, ensemble bagging approaches (EBT, RSDE and RSE-KNN) coupled with are handheld spectrometer had the highest performance, in terms of accuracy and Youden index.

Another way to detect the adulteration in milk is the combination of electronic nose and ML. Indeed, Mu *et al.* [41] proposed a combined solution of e-nose and ML to identify the milk source and estimate the milk quality in terms of fat and protein content. For this study, an e-nose with seven sensors was employed. To process data from the array of sensors, SVM, RF, and logistic regression (LR) were used for rapid identification of milk source in combination with LDA to reduce the dimensionality of the dataset. Gradient boosting decision tree (GBDT), XGBoost and RF were applied tasks. For the first task the best performance in terms of accuracy was achieved by the SVM when features belonging to the e-nose and laboratory analysis were employed (95 %). The RF achieved the highest performance in terms of R^2 both for the fat (0.94) and protein (0.93).

As a general conclusion, the main algorithms applied in the livestock context are the ensemble methods and ANN [3]. Regarding the first method, the RF is widely used to classify behaviour and predict events such as estrus or calving [23], [24]. In addition, this algorithm has many advantages, such as a low computational cost and lower data cleaning required. ANN's are a powerful and flexible class of algorithms for classification and regression problems that can find hidden patterns and nonlinear dependence between input and output [42]. For livestock purposes, ANN's are widely employed for regression problems such as the prediction of pollutant emission or asses the quality of milk [3], [39].

Regarding the source of data, accelerometer data are employed for behaviour recognition, diseases such as mastitis, and event detection. For other tasks, daily milk routine data, automated milking systems data, diet information, and herds characteristics are successfully used. Classification algorithms perform well in accuracy, specificity, precision, and F1-measure. For the regression task, only the ANN achieved important results in terms of RMSE and MAE that may justify their use.

5. LIMITATION AND OBSTACLES IN MACHINE LEARNING

Despite their promising results and advantages, there are some limitations in applying ML algorithms. One of the most critical issues is the poor quality and consistency of livestock data. Indeed, data quality affects ML analysis. Data must have these characteristics: validity, consistency, uniformity, accuracy, and completeness. The veterinary operations and, in general, all the activity carried out in the farms are not fully automated, and hence data may not be properly annotated and may be lost and/or inaccurate.

Several works reported the main limitations of their analysis. For example, Salamone *et al.* [37] noted that the main limitation of their work is the absence of a high-quality disease registration dataset. The latter condition has negatively influenced the quality of their results. In this context, as for disease detection, it is challenging to create a labelled dataset, as the label is a diagnosis or the results of complex and invasive exams.

Regarding the supervised approach, the label requirement is one of the most important obstacles. For example, for behaviour recognition, an experienced operator must recognise the animal behaviour for each time window and record it correctly in a database. This work is particularly demanding, time-consuming, and requires the presence of competent workers. Obtaining a labelled dataset for behaviour recognition is time-consuming and expensive.

Deep learning, such as convolutional neural networks (CNN) or deep neural networks (DNN), have been employed for livestock purposes. Despite the high performances achieved, these techniques have a relevant computational burden. Indeed, since the farms' computational panel is limited, complex ML algorithms have little chance of being applied in the field, even if some attempts prone promising.

6. CONCLUSIONS

In this work, a collection of the most important applications of ML algorithms is shown to provide an overview of the most widespread topic of application, algorithms, and their challenges. From this review, it was found that ensemble methods and NN are the most useful algorithms that achieve the best performance. Sensors and herd data are most suitable for many applications. Finally, the main limitations are the availability of data and their consistency. In the end, ML will become essential for the dairy sector to overcome modern challenges such as the sustainability of the product, the wellbeing of animals, and promising performance.

REFERENCES

[1] D. Berckmans, General introduction to precision livestock farming, Animal Frontiers 7(1) (2017), pp. 6–11. DOI: 10.2527/af.2017.0102

[2] D. Bzdok, N. Altman, M. Krzywinski, Statistics versus machine learning, Nat Methods 15(4) (2018), pp. 233–234. DOI: 10.1038/nmeth.4642

[3] X. Chen, H. Zheng, H. Wang, T. Yan, Can machine learning algorithms perform better than multiple linear regression in predicting nitrogen excretion from lactating dairy cows, Scientific Reports 2022 12(1) (2022), pp. 1–13. DOI: 10.1038/s41598-022-16490-y

[4] G. Morota, R. V. Ventura, F. F. Silva, M. Koyama, S. C. Fernando, Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture, J Anim Sci 96(4) (2018), pp. 1540–1550. DOI: 10.1093/jas/sky014

[5] M. E. Hossain, M. A. Kabir, L. Zheng, D. L. Swain, S. McGrath, J. Medway, A systematic review of machine learning techniques for cattle identification: Datasets, methods and future directions, Artificial Intelligence in Agriculture 6 (2022), pp. 138–155. DOI: 10.1016/j.aiia.2022.09.002

[6] R. Pugliese, S. Regondi, R. Marini, Machine learning-based approach: global trends, research directions, and regulatory standpoints, Data Science and Management 4 (2021), pp. 19–29. DOI: 10.1016/j.dsm.2021.12.002

[7] Y. He, F. Tiezzi, J. Howard, C. Maltecca, Predicting body weight in growing pigs from feeding behavior data using machine learning algorithms, Comput Electron Agric 184 (2021), art. no. 106085. DOI: 10.1016/j.compag.2021.106085

[8] S. Adebayo, H. O. Aworinde, A. O. Akinwunmi, O. M. Alabi, A. Ayandiji, A. B. Sakpere, A. Adeyemo, A. K. Oyebamiji, (+ 2 more authors) Enhancing poultry health management through machine learning-based analysis of vocalization signals dataset, Data Brief 50 (2023), art. no. 109528. DOI: 10.1016/j.dib.2023.109528

[9] W. Grzesiak, D. Zaborski, R. Pilarczyk, J. Wójcik, K. Adamczyk, Classification of daily body weight gains in beef calves using decision trees, Artificial Neural Networks, and Logistic Regression, Animals 13(12) (2023), art. no. 1956. DOI: 10.3390/ani13121956

[10] K. Ren, G. Bernes, M. Hetta, J. Karlsson, Tracking and analysing social interactions in dairy cattle with real-time locating system and machine learning, Journal of Systems Architecture, 116 (2021), pp. 102139.
DOI: 10.1016/j.sysarc.2021.102139
[11] P. Shine, M. D. Murphy, Over 20 Years of Machine Learning Applications on Dairy Farms: A Comprehensive Mapping Study, Sensors 22(1) (2021), pp. 52.
DOI: 10.3390/S22010052
[12] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, D. Bochtis, Machine Learning in Agriculture: A Comprehensive Updated Review, Sensors 21(11) (2021), pp. 3758
DOI: 10.3390/S21113758
[13] R. García, J. Aguilar, M. Toro, A. Pinto, P. Rodríguez, A systematic literature review on the use of machine learning in precision livestock farming, Comput Electron Agric 179 (2020) pp. 105826
DOI: 10.1016/j.compag.2020.105826
[14] N. Slob, C. Catal, A. Kassahun, Application of machine learning to improve dairy farm management: A systematic literature review, Prev Vet Med 187 (2021), pp. 105237.
DOI: 10.1016/j.prevetmed.2020.105237
[15] C. Evangelista, L. Basiricò, U. Bernabucci, An overview on the use of near infrared spectroscopy (NIRS) on farms for the management of dairy cows, Agriculture 11(4) (2021), art. no. 296.
DOI: 10.3390/agriculture11040296
[16] R. L. Brotzman, N. B. Cook, K. Nordlund, T. B. Bennett, A. Gomez Rivas, D. Döpfer, Cluster analysis of dairy Herd Improvement data to discover trends in performance characteristics in large upper midwest dairy herds, J Dairy Sci 98(5) (2015), pp. 3059–3070.
DOI: 10.3168/jds.2014-8369
[17] D. R. M. Alessio, J. P. Velho, C. M. McManus, D. A. Knob, F. R. Vancin, G. V. Antunes, M. Busanello, F. De Carli, A. Th. Neto, Lactose and its relationship with other milk constituents, somatic cell count, and total bacterial count, Livest Sci 252 (2021), p. 104678.
DOI: 10.1016/j.livsci.2021.104678
[18] I. Berglund, G. Pettersson, K. Östensson, K. Svennersten-Sjaunja, Quarter milking for improved detection of increased SCC, Reproduction in Domestic Animals 42(4) (2007), pp. 427–432.
DOI: 10.1111/j.1439-0531.2006.00803.x
[19] Md. S. Shahriar, D. Smith, A. Rahman, D. Henry, G. Bishop-Hurley, R. Rawnsley, M. Freeman, J. Hills, Heat event detection in dairy cows with collar sensors: An unsupervised machine learning approach, IEEE Sensors, Busan, Korea (South), 1-4 November 2015, pp. 1-4.
DOI: 10.1109/ICSENS.2015.7370528
[20] M. Cockburn, Review: Application and prospective discussion of machine learning for the management of dairy farms, Animals 10(9) (2020), pp. 1690.
DOI: 10.3390/ani10091690
[21] V. Bloch, L. Frondelius, C. Arcidiacono, M. Mancino, M. Pastell, Development and analysis of a CNN- and Transfer-Learning-Based classification model for automated dairy cow feeding behavior recognition from accelerometer data, Sensors 23(5) (2023), pp. 2611.
DOI: 10.3390/s23052611
[22] R. Dutta, D. Smith, R. Rawnsley, G. Bishop-Hurley, J. Hills, G. Timms, D. Henry, Dynamic cattle behavioural classification using supervised ensemble classifiers, Comput Electron Agric 111 (2015), pp. 18–28.
DOI: 10.1016/j.compag.2014.12.002
[23] T. Tamura, Y. Okubo, Y. Deguchi, S. Koshikawa, M. Takahashi, Y. Chida, K. Okada, Dairy cattle behavior classifications based on decision tree learning using 3-axis neck-mounted accelerometers, Animal Science Journal 90(4) (2019), pp. 589–596.
DOI: 10.1111/asj.13184
[24] S. Benaissa, F. A. M. Tuyttens, D. Pletsa , T. de Pessemier, J. Trogha, E. Tanghe, L. Martens, L. Vandaele, (+ 3 more authors) On the use of on-cow accelerometers for the classification of behaviours in dairy barns, Res Vet Sci 125 (2019), pp. 425–433.
DOI: 10.1016/j.rvsc.2017.10.005
[25] P. R. Hut, S. E. M. Kuiper, M. Nielen, J. H. J. L. Hulsen, E. N. Stassen, M. M. Hostens, Sensor based time budgets in commercial Dutch dairy herds vary over lactation cycles and within 24 hours, PLoS One 17(2) (2022)
DOI: 10.1371/journal.pone.0264392
[26] G. A. Miller, M. Mitchell, Z. E. Barker, K. Giebel, E. A. Codling, J. R. Amory, C. Michie, C. Davison, (+ 3 more authors), Using animal-mounted sensor technology and machine learning to predict time-to-calving in beef and dairy cows, Animal 14(6) (2020) pp. 1304–1312.
DOI: 10.1017/S1751731119003380
[27] H. M. Zebari, S. M. Rutter, E. C. L. Bleach, Characterizing changes in activity and feeding behaviour of lactating dairy cows during behavioural and silent oestrus, Appl Anim Behav Sci 206 (2018), pp. 12–17.
DOI: 10.1016/j.applanim.2018.06.002
[28] J. Wang, M. Bell, X. Liu, G. Liu, Machine-learning techniques can enhance dairy cow estrus detection using location and acceleration data, Animals 10(7) (2020), pp. 1160.
DOI: 10.3390/ani10071160
[29] S. Shahinfar, M. Khansefid, M. Haile-Mariam, J. E. Pryce, Machine learning approaches for the prediction of lameness in dairy cows, Animal 15(11) (2021), pp. 100391.
DOI: 10.1016/j.animal.2021.100391
[30] J. Wang, Y. Zhang, J. Wang, K. Zhao, X. Li, B. Liu, Using machine-learning technique for estrus onset detection in dairy cows from acceleration and location data acquired by a neck-tag, Biosyst Eng 214 (2022), pp. 193–206.
DOI: 10.1016/j.biosystemseng.2021.12.025
[31] M. R. Borchers, Y. M. Chang, K. L. Proudfoot, B. A. Wadsworth, A. E. Stone, J. M. Bewley, Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle, J Dairy Sci 100(7), pp. 5664–5674.
DOI: 10.3168/jds.2016-11526
[32] K. Ito, M. A. G. von Keyserlingk, S. J. LeBlanc, D. M. Weary, Lying behavior as an indicator of lameness in dairy cows, J Dairy Sci 93(8) (2010), pp. 3553–3560.
DOI: 10.3168/jds.2009-2951
[33] D. Warner, E. Vasseur, D. M. Lefebvre, R. Lacroix, A machine learning based decision aid for lameness in dairy herds using farm-based records, Comput Electron Agric 169 (2020), art. no. 105193.
DOI: 10.1016/j.compag.2019.105193
[34] Hyde, R. M., P. M. Down, A. J. Bradley, J. E. Breen, C. Hudson, K. A. Leach, M. J. Green, Automated prediction of mastitis infection patterns in dairy herds using machine learning, Scientific Reports 10(1) (2020) pp. 1–8.
DOI: 10.1038/s41598-020-61126-8
[35] I. Ramirez-Morales, L. Aguilar, E. Fernandez-Blanco, D. Rivero, J. Perez, A. Pazos, Detection of bovine mastitis in raw milk, using a low-cost NIR spectrometer and k-NN algorithm, Applied Sciences 11(22) (2021), art. no. 10751.
DOI: 10.3390/app112210751
[36] X. Zhou, C. Xu, H. Wang, W. Xu, Z. Zhao, M. Chen, B. Jia, B. Huang, The early prediction of common disorders in dairy cows monitored by automatic systems with machine learning algorithms, Animals 12(10) (2002), art. no. 1251.
DOI: 10.3390/ani12101251
[37] M. Salamone, I. Adriaens, A. Vervaet, G. Opsomer, H. Atashi, V. Fievez, B. Aernouts, M. Hostens, Prediction of first test day milk yield using historical records in dairy cows, Animal 16(11) (2022), art. no. 100658.
DOI: 10.1016/j.animal.2022.100658
[38] W. Ouweltjes, M. Spoelstra, B. Ducro, Y. de Haas, C. Kamphuis, A data-driven prediction of lifetime resilience of dairy cows using commercial sensor data collected during first lactation, J Dairy Sci 104(11) (2021), pp. 11759–11769.
DOI: 10.3168/jds.2021-20413

[39] R. Muñiz, M. Cuevas-Valdés, B. de la Roza-Delgado, Milk quality control requirement evaluation using a handheld near infrared reflectance spectrophotometer and a bespoke mobile application, Journal of Food Composition and Analysis 86 (2020), art. no. 103388.
DOI: 10.1016/j.jfca.2019.103388
[40] S. Ehsani, H. Yazdanpanah, H. Parastar, Development of a non-targeted approach using three handheld spectrometers combined with ensemble classifiers for authentication of bovine milk, Chemometrics and Intelligent Laboratory Systems 237 (2023), art. no. 104816.
DOI: 10.1016/j.chemolab.2023.104816
[41] F. Mu, Y. Gu, J. Zhang, L. Zhang, Milk source identification and milk quality estimation using an electronic nose and machine learning techniques, Sensors 20(15), art. no. 4238.
DOI: 10.3390/s20154238
[42] J. V. Tu, Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes, J Clin Epidemiol 49(11) (1996), pp. 1225–1231.
DOI: 10.1016/s0895-4356(96)00002-9