Lesson learned in big data for dairy cattle: advanced analytics for heat stress detection

Stefano Benni*, Marco Bovo, Miki Agrusti, Mattia Ceccarelli,
Alberto Barbaresi, Daniele Torreggiani, Patrizia Tassinari

University of Bologna, DISTAL – Viale G. Fanin, 48 – 40127 Bologna, Italy

* stefano.benni@unibo.it

Introduction

The potential of data analysis in livestock farming

New challenges in Precision Livestock Farming (PLF) have been arising as a consequence of the increasing pressure on the livestock sector to meet the growing demand from an increasing population with rising incomes. This is highlighting the awareness of the need to reduce the pressure over natural resources and the environmental impact, while safeguarding animal welfare and the sustainability in animal-derived products. Simultaneously, climate change poses a significant challenge to the dairy industry, and it will become an increasingly critical issue to solve in order to ensure its economic, environmental, and social viability.

In order to gain a comprehensive understanding of these phenomena, and to monitor and control production processes in relation to climate change, sophisticated and high-throughput data acquisition systems are required, with highly heterogeneous and multi-channel datasets ranging from time-resolved sensors for animal behaviour monitoring to detailed measurements related to environmental data, as well as health parameters and animal products, such as milk (John et al., 2016).

Available data recording processes usually include the use of on-farm sensors (cow and herd level), fine-milk composition based on mid-infrared spectrometry, data recording other performances such as fertility, weather, and environmental data (Halachmi et al., 2019). Moreover, the information from different sources is almost never crossed among different systems and different processing procedures (Bonora, Benni, et al., 2018).

The major difficulty is to be able to efficiently gather, store, harmonize, organize, integrate, and analyse these data, in order to answer critical questions regarding production challenges and the discovery of optimal solutions, helpful to a variety of end-users. Furthermore, in the context of animal production, heat stress is one of the most serious threats to animal welfare and productivity in dairy cattle farms (Bonora, Pastell, et al., 2018), with significant implications for milk quantity and quality (Piwczyński et al., 2020), as well as the efficiency with which natural resources are used and the amount of energy that is required for milk production.

A contribution of PLF towards sustainability

As it is well known, PLF is also in charge of providing a significant contribution to environmental sustainability and to the development of strategies for adaptation to climate change. The global trend of rising average and peak temperatures is lowering cow efficiency and, as a result, indirectly worsening the environmental effect of dairy cattle production, necessitating new methods to long-term sustainability. The availability of collected databases has aided the development of novel and effective computational approaches aimed at precisely characterizing the driving factors and quantifying the consequences of heat stress in cattle, in terms of productive, and behavioural characteristics (Heinicke et al., 2021).

The increasing importance of digital solutions to support farmers in the EU is reflected in the Digital Single Market strategy and the Declaration of cooperation signed by 24 EU countries (2019) stating that "European agriculture and rural areas have to benefit fully from the ongoing digital transformation of our economies and societies." There is increasing pressure on farmers to meet certain standards: nitrogen emissions, methane emissions, animal health and welfare, etc. A new wave of flexible and cheaper smart sensors is bringing new opportunities to monitor the livestock and the environment real-time. Smart data management followed by informed decisions is expected to track the supply chain, up to the consumer, showing if and to what extent climate and sustainability goals are met and how the dairy sector is resilient and sustainable. Digital transformation in agriculture demands very specific solutions. Customizing appropriate digital technologies such as the Internet of things, Big Data, Artificial Intelligence, robotics, drones and satellites, are increasing automation, enhancing resource use efficiency, decreasing adverse environmental impacts, reducing food production costs and optimizing production and quality, through more precise system management.

Many devices have been introduced in livestock farms to monitor and control environmental conditions, animal behavior, and production parameters (Lovarelli et al., 2020). Similarly, milk quality reports at dairies and labs are often logged for long-term analysis. This enables the development of data-driven platforms and solutions to improve the sustainability of the agri-food system. In fact, livestock production has a significant impact on local and global environmental balances. Furthermore, consumers are becoming more environmentally conscious, and they are demanding higher-quality food products. The attention of the cattle farming sector has recently been drawn to two key issues: food safety, in terms of reduction of antibiotics and toxins, and environmental sustainability. Today, thanks to modern technologies adopted by farmers, including the use of software and sensors, users at all scales can address these issues by ensuring improved product quality through efficient farm supervision and smart resource management (Cockburn, 2020).

Besides, digitalization is accelerating in all activities, including agriculture, where an abundance of data is already collected on farms, particularly in dairy production (Tassinari et al., 2021). Although this data contains a wealth of information, its full potential has yet to be realized. Dairy farmers are continuously presented with unprecedented, difficult issues as a result of these new data flows.

Aims of the study

This report provides an overview of the strategies for data management and data analysis developed within the EU project EIT Food DairySust "Big data and advanced analytics for sustainable management of the dairy cattle sector". The main ambition of this project is to improve sustainability and animal welfare, besides productivity, in dairy farming, through advanced data analytics for every level of stakeholders. Good data management, in terms of acquisition, processing, harmonization and imputation, is required for good modelling for early diagnosis and for the identification of optimal prevention strategies, particularly in fields where monitoring can collect very heterogeneous data, and for which agreed protocols have not yet been standardized. The project investigated the "ecosystem" of data and application strategies for sharing computer resources and information in a secure and organic manner. This research first developed an optimal computational ecosystem based on the integration and harmonization of heterogeneous data types. Classical and advanced modelling strategies were used and compared. The results are suitable to provide the stakeholders with improved decision-making process about animal welfare and sustainability of the production.

This report focuses on the implementation of a numerical model for the assessment of the impact of heat stress on milk production and provides a feedback on it.

Scientific approach and data considered

Data collected and generated in the DairySust project include large datasets on farm equipment, farm structures, management, indoor and outdoor environmental conditions of farm, animal production and animal productivity, animal welfare and health and milk quality. Farming and environmental data have the structure of a multivariate time series.

Two levels of organization for data collection and management were distinguished: local level (farm) and global level on shared infrastructures. At the local level, all farms have been responsible of the data collection (e.g., from local farm equipment or labs), processing (e.g., data wrangling, cleaning, validation) and access (e.g. on secured servers or equipment according to local regulations). Uniformed protocols with Standard Operation Procedures have been defined to optimize data collection and harmonization between the farmers providing data and to allow optimal cross-region analysis with the largest available dataset.

The data used for this research have been collected in the context of previous and ongoing research projects. These datasets contain data of commercial and experimental dairy farms. In particular, data were retrieved from reports files extracted from the farm management systems of twenty Italian dairy farms with automatic milking system or electronic milking parlour. This database contains information of approximately 25.000 lactations linked to approximately 10.000 cows.

The study has been developed with standard Python software languages. At the local level, data have been collected and then transferred to organized folder organigrams that allow for traceability and reproducibility of any data cleaning or processing. An important principle of such data management approach is to keep a version of the raw data to allow for replaying the cleaning, processing and validation pipelines.

By means of the available contacts, the consortium members of the DairySust project put in contact farmers, vets, breeders, dairy industry, livestock building and equipment designers in order to explain the project goals and collect possible suggestions for the improvement of the running project. In particular, useful input about system requirements and functionality of the software have been collected. The end users' preferences have been adopted for the definition of the features of the algorithms developed in the study.

Numerical model for the assessment of the impact of heat stress

The algorithm of the statistical model, used to assess the daily milk yield of a single cow considering the environmental conditions (Bovo et al., 2021), makes use of:

- day in milk of the cow;
- average temperature-humidity index of the current day;
- average temperature-humidity indices of the 5 days before the day of interest.

The Random Forest algorithm allows to assess the heat stress effects at the level of each single cow, by means of seven different features (i.e., predictors) used as input data.

The regression analysis of the available data is performed by using the Random Forest algorithm, an ensemble learning method that makes predictions by averaging over the predictions provided by several independent random models. The algorithm is conceived as a method of combining several classification and regression trees using bagging, and as the name suggests, it is a tree-based ensemble with each tree depending on a collection of random variables. In the present case, the algorithm was adopted for regression purposes by using the Scikit-Learn Python library in order to establish the random forest model (RFM) best fitting the data values of each cow.

The statistical model used for the interpretation of the collected data is composed of seven predictors: days in milk of the cow, daily average THI of the day of the assessment and those of past five days. In a study performed over a sample of Italian farms, the average relative error provided by the model in the predictions, is about 18% with a single daily yield, whereas it becomes just 2% if the total milk production in the test days is considered.

Conclusions

The results show that the model can detect the drop in the cow's milk yield due to extreme hot conditions inducing heat stress effects. The algorithm allows to assess the expected daily milk yield in presence of heat stress effects, for each single cow. Moreover, it allows to quantify the effects of critical levels of THI on milk production and lastly to assess the consequences of climate change scenarios on production based on projections of expected temperature-humidity index calculated on the basis of climatic data.

The results confirm the obtained RFM can represent a reliable and viable tool for the evaluation of future productive scenarios of dairy cows in the presence of heat stress effects. This could help to develop and improve decision support for farmers to increase both milk yield and animal welfare and, on the other hand, to reduce the resources needed, so to increase the sustainability of the dairy sector.

References

Bonora, F., Benni, S., Barbaresi, A., Tassinari, P., & Torreggiani, D. (2018). A cluster-graph model for herd characterisation in dairy farms equipped with an automatic milking system. *Biosystems Engineering*, *167*, 1–7. https://doi.org/10.1016/j.biosystemseng.2017.12.007

Bonora, F., Pastell, M., Benni, S., Tassinari, P., & Torreggiani, D. (2018). ICT monitoring and mathematical modelling of dairy cows performances in hot climate conditions: a study case in Po valley (Italy). *Agricultural Engineering International: CIGR Journal*, *20*(Special issue: Animal Housing in Hot Climate), 1–12.

Bovo, M., Agrusti, M., Benni, S., Torreggiani, D., Tassinari, P., Guarino, M., & Lovarelli, D. (2021). *Random Forest Modelling of Milk Yield of Dairy Cows under Heat Stress Conditions*. https://doi.org/10.3390/ani11051305

Cockburn, M. (2020). Review: Application and prospective discussion of machine learning for the management of dairy farms. In *Animals*. https://doi.org/10.3390/ani10091690

Halachmi, I., Guarino, M., Bewley, J., & Pastell, M. (2019). Smart Animal Agriculture: Application of Real-Time Sensors to Improve Animal Well-Being and Production. In *Annual Review of Animal Biosciences* (Vol. 7, pp. 403–425). https://doi.org/10.1146/annurev-animal-020518-114851

Heinicke, J., Ott, A., Ammon, C., & Amon, T. (2021). Heat Load-Induced Changes in Lying Behavior and Lying Cubicle Occupancy of Lactating Dairy Cows in a Naturally Ventilated Barn. *Annals of Animal Science*, *21*(4), 1543–1553. https://doi.org/10.2478/aoas-2020-0113

John, A. J., Clark, C. E. F., Freeman, M. J., Kerrisk, K. L., Garcia, S. C., & Halachmi, I. (2016). Review: Milking robot utilization, a successful precision livestock farming evolution. *Animal*. https://doi.org/10.1017/S1751731116000495

Lovarelli, D., Finzi, A., Mattachini, G., & Riva, E. (2020). A survey of dairy cattle behavior in different barns in northern italy. *Animals*, *10*(4). https://doi.org/10.3390/ani10040713

Piwczyński, D., Sitkowska, B., Kolenda, M., Brzozowski, M., Aerts, J., & Schork, P. M. (2020). Forecasting the milk yield of cows on farms equipped with automatic milking system with the use of decision trees. *Animal Science Journal*. https://doi.org/10.1111/asj.13414

Tassinari, P., Bovo, M., Benni, S., Franzoni, S., Poggi, M., Mammi, L. M. E., Mattoccia, S., Di Stefano, L., Bonora, F., Barbaresi, A., Santolini, E., & Torreggiani, D. (2021). A computer vision approach based on deep learning for the detection of dairy cows in free stall barn. *Computers and Electronics in Agriculture*, *182*. https://doi.org/10.1016/j.compag.2021.106030