Journal of Dairy Research

cambridge.org/dar

Review and assessment of sensor-based disease detection in dairy cattle: conceptual and practical challenges

Morteza Hosseini Ghaffari[1], Harald Hammon[1] and Ute Müller[2]

[1]Working Group Metabolic Health, Research Institute for Farm Animal Biology (FBN), Dummerstorf, Germany and [2]Institute of Animal Science, University of Bonn, Bonn, Germany

Review Article

Cite this article: Hosseini Ghaffari M, Hammon H and Müller U (2026) Review and assessment of sensor-based disease detection in dairy cattle: conceptual and practical challenges. *Journal of Dairy Research*, 1–8. https://doi.org/10.1017/S0022029926102258

Received: 10 December 2025
Revised: 25 February 2026
Accepted: 26 February 2026

Keywords:
dairy cattle; disease detection; precision livestock farming; sensor systems

Corresponding authors:
Morteza Hosseini Ghaffari; Email: hosseini-ghaffari@fbn-dummerstorf.de and Ute Müller; Email: ute-mueller@uni-bonn.de

Abstract

Recent advances in sensor technologies and machine learning have increased expectations for actionable disease detection in dairy cattle, yet farm-level adoption remains limited. This conceptual review outlines an implementation trajectory from proof of concept to deployment as follows: (i) case definitions and defensible reference standards for diseases such as mastitis, lameness and respiratory disease should be biologically based; (ii) model development should consider features that drive utility and address class imbalance, ensuring that algorithms can reliably detect diseases that occur infrequently within herds; (iii) external validation should occur at the herd level, across seasons and time, and should report prevalence-aware metrics (e.g., recall and precision [positive predictive value]) rather than relying solely on accuracy or receiver operating characteristic area under the curve; and (iv) interoperability and economic viability should be sufficient to support multi-sensor use and sustainable adoption. Economic factors such as hardware and software costs, follow-up time burden, usefulness and return on investment under realistic false-positive rates will primarily determine adoption. Ethical and welfare considerations include maintaining human oversight where possible, avoiding invasive sensing and preventing over-automation. In conclusion, sensor-based systems show promise but remain constrained by inconsistent case definitions, limited cross-herd validation, interoperability gaps and uncertain economics, and many tools still blur the line between detecting animals needing treatment and ancillary predictions. With regulatory and ethical safeguards, progress will depend on reliability-focused standard operating procedures and end-user–oriented design; until then, most systems remain closer to hype than independent clinical technologies.

The rapid development of sensor data systems, combined with advances in computational methods such as machine learning, has renewed strong expectations for earlier and more actionable disease detection in dairy cattle. Early efforts to automate disease detection began in the late 1970s and 1980s (Gebre-Egziabher *et al.*, 1979; Fernando *et al.*, 1982). Building on these foundations, current approaches aim to integrate continuous data streams from milking systems, wearable sensors, imaging technologies and environmental monitoring to predict disease events in dairy cows before clinical signs appear (Eckelkamp, 2019). This predictive capability enables timely intervention, optimized treatment protocols, improved herd productivity and enhanced animal welfare. As a result, the dairy industry places considerable expectations on data-driven and predictive analytics to transform herd health management (Neethirajan, 2024).

Regulatory and ethical context shapes implementation. In the European Union, Council Directive 98/58/EC (1998) mandates daily visual inspection by the responsible stockperson, emphasizing that disease detection is a legal and welfare priority and that automated systems should support, not replace, routine husbandry. Despite significant advances in sensor technologies and increasing commercial investment in applications, tools and products, the uptake of sensors and real-time solutions on farms remains limited. This gap between innovation and practical application for prediction of diseases underscores the challenges of integrating new technologies into livestock production. This review evaluates whether sensor data systems have progressed from research and development to practical deployment. We (i) synthesize insights across data streams categorized as milking-system, wearable, imaging and environmental; (ii) assess generalizability and clinical utility using prevalence-aware performance metrics (recall, precision/positive predictive value [PPV], calibration); and (iii) present a translational 'algorithm-to-barn' framework that includes a deployability checklist with biologically standardized case definitions and reference standards, validation (out-of-sample, cross-herd, temporal), alert-to-action pathways, interoperability and data governance, and cost–benefit

© The Author(s), 2026. Published by Cambridge University Press on behalf of Hannah Dairy Research Foundation.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 04:13:41, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0022029926102258

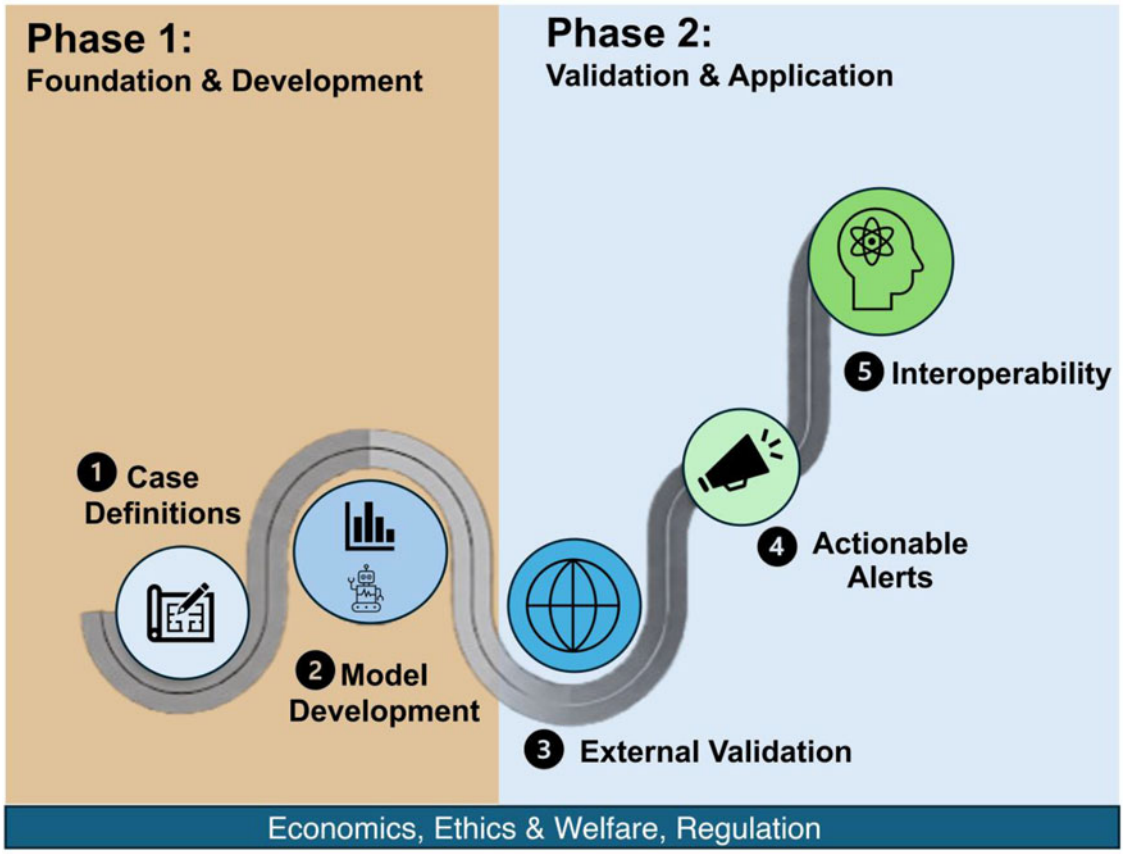


Figure 1. Translational 'algorithm-to-barn' roadmap for sensor-based disease detection in dairy cattle. A graphic summary of the sequential phases required to advance from proof-of-concept algorithms to field-ready detection systems. Phase 1 (Foundation & Development) establishes biologically defensible case definitions and rigorous model development. Phase 2 (Validation & Application) focuses on external, cross-herd validation and converting model outputs into actionable alerts with clear decision thresholds. The roadmap also incorporates considerations of economics, ethics, welfare and regulation in the development of sensor-based detection systems.

transparency. To provide a structured orientation, Figure 1 visually maps the entire trajectory, from case definition and model construction (Phase 1) through validation and actionability to interoperability (Phase 2). This framing is essential because the principal question is not whether algorithms can detect disease-linked signals, but under which real-world conditions such systems generate reliable, interpretable and actionable decision support on commercial dairy farms.

From concept to deployment: a translational framework

The development of a sensor-based health detection system is a sequential, multi-stage process that begins with identifying a biologically meaningful signal or behaviour associated with disease, such as reduced milk intake or activity in calves, or decreased milk yield in cows, followed by the design and implementation of sensors and algorithms capable of reliably capturing and interpreting that signal. Subsequently, a prototype is produced and evaluated through controlled pilot studies to verify the feasibility of the concept-representing the first stage of biological validation. For example, a calf disease detection algorithm could be tested using feeding behaviour data from automated milk feeding systems to detect conditions such as diarrhoea or respiratory disease (Ghaffari *et al.*, 2022). If successful, the next step is to validate the system under real farm conditions. However, robust external validation is crucial, as a systematic review found that only 18 out of 129 commercially available sensors have been externally validated in scientific studies (Stygar *et al.*, 2021). Continuous development is required for successful implementation, with the practical experience of companies considered during the process. Many prototypes with great potential have not been deployed on a larger scale due to issues such as sensor robustness, data incompatibility and lack of rigorous testing (Eastwood and Renwick, 2020; Kaur *et al.*, 2023). Several unproven sensor-based systems have found their way onto the market, leaving farmers with a disappointing experience and losing confidence in sensor technology (Eastwood and Renwick, 2020). The need to remedy this is supported by some strong scientific evidence, pilot studies and large-scale validations that have established a systematic framework to ensure that only effective technologies are used (Post *et al.*, 2020).

Defining diseases and modelling requirements

A central challenge in sensor-based disease detection is establishing a clear and biologically meaningful definition of the 'disease case' that the model is intended to identify, given that conditions such as mastitis and lameness are defined inconsistently across studies. Mastitis can range from subclinical conditions, typically indicated by elevated somatic cell count (SCC), to clinical infections with visible changes in milk and/or udder inflammation (Brand *et al.*, 2021; Carvalho-Sombra *et al.*, 2021). As a result, studies use different SCC thresholds and clinical criteria to define

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 04:13:41, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0022029926102258

mastitis. More inclusive (lower) SCC thresholds increase the apparent prevalence of 'cases' and may improve sensitivity (recall), but they can also increase false positives and reduce specificity and PPV, making comparisons between studies and models trained under different case definitions difficult (Cavero *et al.*, 2008).

Rumination-derived metrics deserve explicit consideration because they are a core input in many commercially available automated health monitoring systems and frequently trigger early 'health deviation' alerts. In a randomized controlled trial comparing automated alerts with visual observation during early lactation, automated monitoring based on rumination, activity and milk yield increased both the number of cows selected for clinical examination and the number of clinical disorders diagnosed, demonstrating that rumination-driven alerting can alter detection intensity and downstream management workload (Rial *et al.*, 2024). Similarly, evaluations of sensor-based monitoring around parturition indicate that rumination-based alerts can achieve moderate sensitivity but often have limited specificity, particularly near calving when physiological changes in rumination and a high density of early-postpartum alerts complicate interpretation (Simoni *et al.*, 2024).

These findings collectively underscore that rumination signals are biologically relevant yet inherently non–disease-specific. They are influenced by numerous non-pathological factors, including calving dynamics, dietary transitions, heat stress, oestrus, regrouping and management interventions. Therefore, when rumination is included as a model feature or as part of a composite alert, case definitions and reference standards must be biologically defensible and independent of the sensor output. Validation frameworks should explicitly quantify the impact of non-disease events on false-positive rates and follow-up workload. In practice, rumination metrics often provide the greatest utility when integrated with complementary data streams – such as milk yield or electrical conductivity, body temperature, activity, and feeding or drinking behaviour – to enhance specificity and guide actionable follow-up. Performance reporting should include prevalence-aware metrics (e.g., PPV/precision and alert rate per cow-day), rather than sensitivity alone, because limited specificity directly increases examination burden and contributes to alert fatigue at the herd level (Eckelkamp and Bewley, 2020). Importantly, rumination sensors also illustrate the practical value of relatively non-specific indicators in herd management. Even when alerts lack strong disease specificity, retrospective, cow-level rumination trajectories can provide clinically meaningful context. For cows already flagged for closer attention, due to recent treatment, calving, observable behavioural changes or other management events, reviewing individual rumination patterns can support nuanced, animal-specific decision-making. Thus, beyond prospective alerting, the retrospective interpretation of longitudinal rumination data represents a substantial and often underappreciated management asset.

The recognition of lameness is also associated with similar challenges, although the thresholds for classification are different. Lameness is traditionally assessed by a subjective visual evaluation of gait, which is time consuming and inconsistent (Jones, 2017). As Van Nuffel *et al.* (2015) point out, inconsistent criteria for classifying lame and non-lame cows compromise sensor-based detection and affect the sensitivity and specificity of the models. Furthermore, the lack of standardized definitions makes it difficult to establish the 'ground truth' or gold standard for sensor comparisons. Recent evidence highlights these diagnostic limitations: using the Agriculture and Horticulture Development Board (AHDB) mobility scoring system (threshold ≥ 2), sensitivity for detecting foot lesions ranged from 0.18 to 0.43 depending on the definition of lesion severity, while specificity remained high (0.94–0.96) (Logan *et al.*, 2024). Accordingly, biologically meaningful and transparent case definitions are essential for cross-study comparability and the development of reliable sensor models.

These considerations also imply that labelling strategies must align with the intended application, because automated mobility technologies (e.g., accelerometers, computer vision, micro-Doppler radar) can be optimized for either individual-cow case detection or herd-level monitoring, with fundamentally different threshold and operating-point trade-offs (Jonsson and Linardopoulou, 2023). The AHDB mobility scheme is a 4-point (0–3) visual gait scoring method applied during routine cow movement (typically on exit from the parlour) to categorize mobility impairment and support both action-oriented management (e.g., prioritizing cows for examination or trimming) and herd-level reporting or benchmarking; therefore, score groupings and cut-points should be selected according to the specific objective (surveillance, intervention or model training).

In the UK, standardized mobility scoring has been promoted to reduce observer variability, harmonize actionable case definitions and support timely intervention by identifying emerging mobility impairment, generating prioritized examination or trimming lists, and tracking herd-level trends over time (Archer *et al.*, 2010). Within supply-chain governance, mobility scores also serve as auditable welfare outcome indicators for documenting monitored welfare performance (Archer *et al.*, 2010). At the animal level, moderate to severe mobility impairment is closely linked to pain and thus represents a significant welfare concern, supporting systematic scoring as a trigger for targeted assessment, treatment and follow-up (O'Connor *et al.*, 2023). However, welfare impact depends on how thresholds are applied: audit-oriented cut-points typically prioritize specificity and comparability, while intervention-oriented thresholds aim for higher sensitivity at the cost of increased follow-up workload and more false-positive alerts (Bartlett *et al.*, 2023). This distinction is critical for sensor-model development because algorithms trained on mobility-score labels will inherit these operating characteristics, which may not align with goals such as early detection of painful or treatable pathology.

Once target diseases and detection methods are standardized, robust sensor-based models should be developed within a system-oriented framework that explicitly defines the intended detection level (general welfare deviation, distress or health deviation, or specific disease), guides indicator selection and specifies how behavioural, physiological and environmental signals are combined to generate decisions (Stachowicz and Umstätter, 2021). This approach includes rigorous data preprocessing, extraction of biologically meaningful features and appropriate validation strategies. Sensors commonly used in dairy production generate high-dimensional, noisy time-series data, which require comprehensive data cleaning, normalization and temporal aggregation before downstream modelling and analysis (Eradus and Jansen, 1999; van der Voort *et al.*, 2021). In addition to preprocessing, selected features should be mechanistically linked to disease processes. For example, walking speed, step frequency and recumbency improve predictive value for lameness compared with raw accelerometer signals alone (O'Leary *et al.*, 2020). Model performance also depends on large, diverse and accurately labelled

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 04:13:41, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0022029926102258

datasets that reflect variability across farms, seasons, physiological states and health status. Furthermore, many target conditions for sensor-based detection have a low per-day event probability at the cow level (i.e., low cow-day incidence), creating pronounced class imbalance that must be explicitly addressed during model development and validation to reduce biased learning and overfitting (Kaur *et al.*, 2023). In this context, 'low prevalence' refers to conditions with daily incidence on the order of only a few percent per cow-day, consistent with the 'low daily risk' framework applied to treatment-event prediction in dairy systems (Post *et al.*, 2021; Ghaffari *et al.*, 2022). Critically, such low base rates mathematically limit the attainable PPV under field conditions, meaning that even models with acceptable sensitivity and specificity may yield a substantial proportion of false-positive alerts (Ghaffari *et al.*, 2022). The development of benchmark datasets would benefit from transparent disease definitions, standardized scoring metrics and open-source code to improve reproducibility and accelerate progress across the field (Romero *et al.*, 2022). These elements collectively strengthen generalizability and enable more consistent comparisons across studies.

Validation and impact of low daily disease incidence

Following development, validation under real-life conditions is essential to assess whether the theoretical advantages of the sensor system are realized in practice. Validation should be performed on independent farms and should span diverse seasons and management systems to ensure generalizability. Even when models perform well in initial test settings (often using balanced datasets), performance can deteriorate in new contexts due to differences in herd management, housing, feeding, sensor placement and animal-level variation, underscoring the need for external validation. Although many commercial companies developing sensor technologies have dedicated teams for sensor development, analysis and troubleshooting, academia remains essential because several requirements for trustworthy deployment are effectively public-good functions that benefit from independent, cross-platform assessment. Independent academic evaluations can strengthen external validity by benchmarking performance across diverse herds, seasons and management systems using transparent and reproducible evaluation strategies (Stygar *et al.*, 2021). Moreover, independent work is central to standardization, including biologically meaningful case definitions and harmonized reporting metrics, which enable credible cross-system comparisons and evidence synthesis (Stachowicz and Umstätter, 2021). Finally, because clinical utility depends on how farmers respond to alarms, independent studies that evaluate not only discrimination but also alert actionability and downstream consequences (e.g., workload generated by false positives under realistic daily incidence) are needed to support sustained adoption and responsible use (Post *et al.*, 2021).

A central determinant of practical utility is the low base rate of many health events at the cow-day level, which mathematically constrains the PPV (precision) even when sensitivity and specificity appear acceptable. Many diseases occur infrequently per cow per day, and thus a substantial proportion of alerts may be false positives unless specificity is very high. For example, a machine learning system for respiratory diseases and diarrhoea in calves only achieved a PPV of 37% despite a sensitivity of 83% and a specificity of 79%. The relatively low PPV was attributable to the low disease prevalence, resulting in frequent false alarms (Ghaffari *et al.*, 2022). The same base-rate constraint applies to mastitis and lameness: Post *et al.* (2021) highlighted the 'low daily risk' and reported that mastitis and lameness treatments accounted for approximately 3–4% and 5–6% of treatment-positive labelled cow-days, respectively (i.e., cow-day label frequency rather than a biological daily hazard) (Post *et al.*, 2021). Restricting application to higher-risk subgroups increased PPV to about 0.20 for mastitis and 0.15 for lameness, yet false positives remained substantial (Post *et al.*, 2021). For systems intended to identify animals requiring follow-up, sensitivity or recall should generally be high (e.g., $\geq$0.80) to limit missed cases, but performance assessment must also emphasize precision and alert burden under field prevalence. Accordingly, composite metrics such as the *F*1 score (harmonic mean of sensitivity and precision) should be interpreted cautiously because they conflate recall (driven largely by false negatives in the biological-validation setting) with precision (strongly determined by farm-level base rates and the operational incidence at the cow-day resolution).

For evidence-based evaluation of disease detection tools, a practical framework is recommended. First, prioritize independent external validation across multiple farms, seasons and management systems and report performance stratified by context (Stygar *et al.*, 2021). Second, predefine and report biologically meaningful case definitions and the reference standard, including the timing of confirmatory examinations, because inconsistent definitions and imperfect standards can bias estimates and limit cross-study comparability (Stachowicz and Umstätter, 2021). Third, report sensitivity and specificity together with PPV and negative predictive value under the observed daily disease incidence, since low incidence largely determines the false-alarm burden (Post *et al.*, 2021; Ghaffari *et al.*, 2022). Fourth, quantify actionability and downstream workload, such as checks per true case, under real farm routines, because these factors strongly influence adoption and responsible use (Post *et al.*, 2021).

A commonly mentioned aspect of validation is the ability to recognize cases at an 'early' stage. When conducting tests, it must be clearly defined in advance during the planning stage at which stage of a disease progression the sensor system can display corresponding indications and whether this is also a stage of the disease progression for corresponding control examinations (to distinguish between true positives and false positives) and subsequent measures (in the case of true positives) for improvement. 'Early' detection occurs when subclinical signs of disease can be detected by the sensor system (e.g., subclinical mastitis in the milk of a udder quarter), if, on the one hand, the subsequent control measure can also be carried out at the subclinical level (e.g., a cytobacteriological analysis of the milk from the udder quarter) and, on the other hand, the subsequent treatment based on these subclinical signs is accepted for diagnosis (by the veterinarian) and the treatment can be carried out even without the need for clinical symptoms. However, if the person responsible for the animals checks for true-positive animals based on examinations for clinical signs (such as locomotion or damage to the hooves or limbs in automatic lameness detection systems), it is also essential in the development of the sensor system to detect animals with corresponding clinical signs during the validation studies, as otherwise the subsequent measures will not be appropriate. In this case, however, it should not be referred to as 'early' detection in the strict sense. Studies have already been conducted that specifically linked the alarm to clinical signs of disease several days in advance (Cantor *et al.*, 2024). The only question is how the user should respond to the alarms received regarding clinical diseases (e.g., lameness or diarrhoea in a calf). These animals would then have to remain on the alarm

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 04:13:41, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0022029926102258

lists for several days until the clinical signs appear, so that the user knows for sure who to treat themselves or for whom to contact the veterinarian.

Sensor alerts and thresholds for decision-making

Even the best algorithm is of little use if it cannot be implemented in such a way that it fits the farm management. The ability to act after receiving sensor alerts is an important prerequisite. This means that an alarm should indicate a problem that the farmer can do something about immediately. For example, a mastitis alert is actionable if it prompts a milk sample for culturing or a check that leads to earlier treatment, preventing a severe case. The ability to act also depends on how the information is communicated, e.g., a simple list of 'cows to check' each morning could be more useful than a constant stream of phone notifications. Farmers have indicated that warnings need to be clear and credible to warrant a response.

Warning thresholds are crucial for effective detection. Developers need to find the right balance: a low threshold (sensitive model) detects more true cases but also increases the number of false positives, while a high threshold reduces the number of false positives but may miss some sick animals. The optimal balance depends on the specific needs and circumstances of each farm, with the recall value required to be at least 0.8. Lower values result in more undetected animals, which is relevant to animal welfare. Thus, customizable alarm thresholds or priority settings are highly desirable (O'Leary *et al.*, 2020). With many sensor systems, one farm might opt for a 'strict' setting that only flags cows with very strong signs of disease (lower sensitivity, higher specificity), while another farm that has more manpower available might prefer to catch every possible case, even if it means that some controls turn out negative. This adaptability is one way of tailoring the sensor systems to the risk tolerance and management style of the farm in question. For animal protection reasons, however, this should only be possible to a limited degree.

In practice, dealing with false alarms is not only a technical issue but also a psychological one. As already mentioned, a flood of false alarms quickly leads to farmer fatigue and a loss of confidence in the system. This is a common situation: farmers will start to ignore alarms if most of them turn out to be false alarms, which defeats the whole purpose of the technology. When using sensor systems, it should be recognized that an alert must be followed by a clinical check to distinguish true-positive from false-positive alerts, and that the proportion of true positives may be low because of the low disease prevalence. For example, some mastitis systems require an alarm to persist over two milkings before recommending action (to avoid one-off spikes). Similarly, integrating multiple sensor inputs can act as a filter by, for example, only triggering an alarm when a cow has both an elevated temperature and reduced feed intake, rather than either alone. The idea is to improve the PPV of the composite alarm. Farmers themselves often develop heuristics: they may only respond to an alert when certain risk factors are present (e.g., a high-risk period such as early lactation or the cow is not looking well). Thus, the farmer's management knowledge becomes an unofficial layer of the detection system, but a very important expert layer. This synergy is the reason why many authors consider sensor alerts as decision aids rather than automated decision-makers (Post *et al.*, 2021). The best implementations facilitate the farmer's decision by providing relevant information (e.g., 'cow 1234 is marked; her last recorded rumination has decreased by 30% and the temperature has increased by 0.5°C') but leave the decision to the human.

Operational reliability and after-sales support

Implementation support, system uptime and after-sales service are decisive determinants of realized on-farm value and must be evaluated alongside algorithmic performance. In practice, sensor deployments are vulnerable to multiple non-algorithmic failure modes, including device loss or physical damage, battery depletion, intermittent connectivity, calibration drift, mounting or placement errors, and software or firmware incompatibilities. These factors can compromise data integrity, distort model inputs and inflate false-positive alert rates, thereby eroding user trust and perceived utility.

Robust implementation therefore requires a structured service architecture. This includes standardized onboarding and user training, covering alarm interpretation and the practical use of retrospective, cow-level data for targeted follow-up, along with preventive maintenance and calibration protocols. Rapid troubleshooting capacity is also essential and should be supported by remote diagnostics, defined escalation pathways and explicit performance benchmarks such as service-level agreements specifying response and resolution times. Without such operational safeguards, even technically sound algorithms may fail to deliver sustainable value under commercial farm conditions. Ongoing support should include routine software updates, model performance monitoring and recalibration when warranted, and transparent documentation of system changes to preserve interpretability and comparability over time. These operational and support requirements directly influence alert burden, user trust and adoption persistence and should be explicitly incorporated into cost–benefit assessments along with hardware, subscription and follow-up labour costs.

Data management and interoperability challenges

Another important aspect is data management and interoperability. Modern dairies often use multiple sensor systems, e.g., for rumination, activity, milk metrics and cameras, which usually come from different manufacturers and store data in separate software. The lack of common standards for data interoperability in precision livestock farming (PLF) makes it difficult to integrate data from different systems (Niloofar *et al.*, 2021). This is a significant obstacle. For example, a lameness detection algorithm could be improved if it also knows the changes in milk yield (to recognize that a lame cow possibly milks less), but if the activity monitor and milking parlour software do not communicate with each other, such integration is difficult. The industry is aware of this problem: there are efforts (by standardization bodies and research consortia) to develop open data standards and application programming interfaces for sensors in agriculture. Until these are fully developed, farmers should commit to a brand to get an integrated data picture or should laboriously merge the data manually.

The lack of interoperability also hinders research, as valuable data are often locked up in proprietary formats. In addition to integration, the volume and complexity of data are also a challenge. Farms generate thousands of data points per cow every day. Managing and analysing these data in real time require robust on-farm software or cloud services. It is not trivial for a farmer to maintain a local server running machine learning models. Consequently, many systems send data to the cloud for processing. This leads to considerations about internet connectivity and data security. In addition, the continuous stream of data requires intelligent filtering: farmers do not want to sift through charts for every

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 04:13:41, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0022029926102258

cow unless there is an anomaly. A good implementation translates the raw data into concise, meaningful alerts or indicators (often via intuitive dashboards or mobile apps). It is also important that the system logs and visualizes the data so that the farmer or veterinarian can review trends (e.g., to confirm that rumination of an alerted cow has actually decreased over days). Standardized interfaces (e.g., dashboards that aggregate multiple sensor outputs) can greatly improve ease of use, but again, this is easiest when systems are interoperable. According to one report, there is currently a lack of data interoperability in PLF and better integration and data exchange frameworks are needed to achieve the desired efficiency gains (Niloofar *et al.*, 2021). Overall, the details of implementation, the way the data get from the sensors to the decisions and the way people interact with the system are as important as the accuracy of the algorithms. Many promising tools have failed not because they were unable to detect diseases, but because they were difficult to use, or configure, or because farmers did not trust the tools on the farm, and this, in turn, for the reasons explained above.

Evaluating sensor performance

To properly evaluate performance, conventional metrics such as accuracy are not sufficient. Due to class imbalance (rare disease events), precision (PPV) and recall (sensitivity), together with a specific economic weighting of false negatives versus false positives, are more meaningful (O'Leary *et al.*, 2020). A system with an 'accuracy' of 95% may simply reflect that 95% of cows are healthy on a given day. What really matters is how many sick cows it detects (sensitivity) and what proportion of its alerts are correct (PPV). Precision-recall (PR) curves have been recommended for evaluating models under such unbalanced conditions, because they focus on the trade-off between PPV and sensitivity and are not dominated by true negatives like receiver operating characteristic (ROC) curves (O'Leary *et al.*, 2020). Indeed, a high area under the ROC curve (AUC) can be misleadingly optimistic in this context. For example, the convolutional neural network (CNN) model for calf disease had an AUC of 0.81, which sounds excellent, but its precision was only 0.37 due to the low frequency of occurrence in this farm (Ghaffari *et al.*, 2022). For the farmer, an AUC of 0.81 is abstract, while 'only 1 in 3 alarms is a really sick calf' is tangible and probably unacceptable. Therefore, stating the PR-AUC or simply the precision at a given recall is more meaningful when discussing performance with the individual end user's situation. Calibration of predictions is another valuable metric: if a system specifies a risk value (probability of disease), these probabilities should reflect reality (e.g., out of 100 cows with a 1% risk, about 1 should actually fall ill within the specified prediction period (e.g., one day)). Well-calibrated models allow farmers to make better decisions. Some advanced systems now favour the presentation of a risk score or probability rather than a binary warning to allow for nuanced decision-making and threshold setting (van der Voort *et al.*, 2021).

Economic considerations

For a sensor-based detection system to move from a novelty to a necessity, it should provide an economic benefit or at least a compelling cost justification. Dairy farming is a low margin business, so any investment in technology is scrutinized for its return on investment. The economic justification for disease detection comes from several angles. Firstly, reducing the incidence or severity of a disease can lead to savings in treatment costs, milk loss, reduced culling and longer animal productive lifespan. For example, the average economic burden of clinical mastitis is $444 per case, with $128 reported as direct costs and $316 as indirect costs (Rollin *et al.*, 2015). Indirect costs, which consist of loss of future milk production (28%) and loss due to premature culling and replacement animals (41%), account for 71% of total costs and reflect the current economic climate in the dairy industry (Rollin *et al.*, 2015). Detection that prevents a case from becoming severe (or, in the case of contagious diseases, prevents the spread) can save some of these costs. Lameness, as mentioned above, is costly: one analysis estimates that a single sole ulcer costs on average about $960 in losses, while even mild foot rot costs about $136 (Dolecheck and Bewley, 2018). If an automated system identifies lame cows immediately upon the first symptoms, which can also be detected during the check-up, and allows for timely treatment, the cow could recover faster and lose less milk, improving the overall economic outcome (O'Leary *et al.*, 2020). However, the flip side of the coin is that these sensor systems themselves come at a cost, in terms of hardware (sensors, infrastructure), software subscriptions and maintenance, and additional labour, including time spent responding to alerts and distinguishing true-positive from false-positive alerts. In addition, the adoption of sensor technology in the dairy and related industries is influenced by the expected future developments related to the technology and uncertainty about the performance of the sensor technology. Research has shown that early investment was beneficial for some sensors, such as automatic heat detection, whereas uncertainty played a greater role in investment decisions for other technologies, such as automatic body-condition scoring (Rutten *et al.*, 2018). This shows that the economic equation is not only about technical efficiency but also about farmers' confidence in the maturity and future performance of the technology (Rutten *et al.*, 2018).

Ethical and animal welfare implications

The use of sensor-based health detection in dairy herds brings with it several ethical and welfare considerations that need to be considered if the technology is to be used responsibly.

While sensor systems improve efficiency and reduce monitoring bias, they are intended to supplement, not replace, routine barnyard inspections of livestock (Schillings *et al.*, 2021). Some critics fear that increasing automation could further reduce human-animal contact, which, in turn, would compromise animal welfare. In addition, humans often discover problems during routine barn inspections that an automated sensor might miss, and gentle handling can have a positive effect on the animals. These discussions are an important aspect of 'smart farming' in relation to animal welfare. The impact on animal welfare also depends on how the sensor is used: non-invasive or invasive. All non-invasive devices placed on an animal should be assessed for potential pain or discomfort, and most sensors used as accessories (e.g., collars, pedometers) generally do not cause significant discomfort once the animal has become accustomed to them. For animal welfare reasons, the assessment should focus on invasive sensors such as rumen boluses, of which a dairy cow may receive more than one. The insertion of an ear tag can also be classified as invasive and carries a risk of inflammation and other consequences if not performed professionally (Hayer *et al.*, 2022). Technology must not be introduced just for the sake of efficiency, as animal welfare must come first. Technology must be used to improve preventative

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 04:13:41, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0022029926102258

management, to support daily animal monitoring and identify problems early enough to limit animal suffering and reduce culling decisions. Identifying risks of abuse can help to mitigate some of the risks.

Conclusion

In summary, while sensor-based systems have the potential to detect disease in dairy cows, inconsistent disease definitions, interoperability limitations and insufficient validation across farming environments must be addressed and clearly communicated to end users, including the fact that disease occurrence is a rare daily event. When using sensor systems, a special distinction must be made between the detection of animals in distress and the prediction of other information about the animal. Appropriate regulatory compliance, ethical issues such as the frequency of human-animal contact and over-reliance on automation, which should be combined with direct human control, should also be considered. For these systems to move from technological promise to routine operational use, reliability, integration and end-user orientation need to be formalized. Currently, they are more hype than reality, but their systematic development will make them a viable part of dairy farming once the technological, practical and ethical barriers are no longer ignored and effectively addressed.

Competing interests. The authors declare no conflicts of interest.

Declaration of generative AI in the writing process. During the preparation of this manuscript, the first author used ChatGPT Plus (GPT-5.5 Thinking; OpenAI) to improve readability and language clarity. The first author used NotebookLM Pro (Google AI-powered research assistant) to interact with curated libraries of uploaded research articles and to support thematic comparison and synthesis during manuscript preparation. Following the use of this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the originality, accuracy and integrity of the final publication.

References

Archer S, Bell N and Huxley J (2010) Lameness in UK dairy cows: A review of the current status. *Practice* **32**, 492–504. https://doi.org/10.1136/inp.c6672

Bartlett H, Balmford A, Holmes MA and Wood JLN (2023) Advancing the quantitative characterization of farm animal welfare. *Proceedings of the Royal Society B: Biological Sciences* **290**, 20230120. https://doi.org/10.1098/rspb.2023.0120

Brand KS, Filor V and Bäumer W (2021) Early inflammatory events of mastitis—a pilot study with the isolated perfused bovine udder. *BMC Veterinary Research* **17**, 356. https://doi.org/10.1186/s12917-021-03029-y

Cantor MC, Welk A, Creutzinger KC, Woodrum MM, Setser JH, Costa HC and Renaud DL (2024) The development and validation of a milk feeding behavior alert from automated feeder data to classify calves at risk for a diarrhea bout: A diagnostic accuracy study. *Journal of Dairy Science* **107**, 3140–3156. https://doi.org/10.3168/jds.2023-23635

Carvalho-Sombra TCF, Fernandes DD, Bezerra BMO and Nunes-Pinheiro DCS (2021) Systemic inflammatory biomarkers and somatic cell count in dairy cows with subclinical mastitis. *Veterinary and Animal Science* **11**, 100165. https://doi.org/10.1016/j.vas.2021.100165

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

Council Directive 98/58/EC (1998) Council directive 98/58/EC of 20 July 1998 concerning the protection of animals kept for farming purposes. *Official Journal of the European Communities* L **221**, 23–27.

Dolecheck K and Bewley J (2018) Animal board invited review: Dairy cow lameness expenditures, losses and total cost. *Animal* **12**, 1462–1474. https://doi.org/10.1017/S1751731118000575

Eastwood CR and Renwick A (2020) Innovation uncertainty impacts the adoption of smarter farming approaches. *Frontiers in Sustainable Food Systems* **4**, 4–24. https://doi.org/10.3389/fsufs.2020.00024

Eradus WJ and Jansen MB (1999) Animal identification and monitoring. *Computers and Electronics in Agriculture* **24**, 91–98. https://doi.org/10.1016/S0168-1699(99)00039-3

Eckelkamp EA (2019) Invited review: Current state of wearable precision dairy technologies in disease detection. *Applied Animal Science* **35**(2), 209–220. https://doi.org/10.15232/aas.2018-01801

Eckelkamp EA and Bewley JM (2020) On-farm use of disease alerts generated by precision dairy technology. *Journal of Dairy Science* **103**, 1566–1582. https://doi.org/10.3168/jds.2019-16888

Fernando RS, Rindsig RB and Spahr SL (1982) Electrical conductivity of milk for detection of mastitis. *Journal of Dairy Science* **65**, 659–664. https://doi.org/10.3168/jds.S0022-0302(82)82245-5

Gebre-Egziabher A, Wood HC, Robar JD and Blankenagel G (1979) Evaluation of automatic mastitis detection equipment. *Journal of Dairy Science* **62**, 1108–1114. https://doi.org/10.3168/jds.S0022-0302(79)83383-4

Ghaffari MH, Monneret A, Hammon HM, Post C, Müller U, Frieten D, Gerbert C, Dusel G and Koch C (2022) Deep convolutional neural networks for the detection of diarrhea and respiratory disease in preweaning dairy calves using data from automated milk feeders. *Journal of Dairy Science* **105**, 9882–9895. https://doi.org/10.3168/jds.2021-21547

Hayer JJ, Nysar D, Schmitz A, Leubner CD, Heinemann C and Steinhoff-Wagner J (2022) Wound lesions caused by ear tagging in unweaned calves: Assessing the prevalence of wound lesions and identifying risk factors. *Animal* **16**, 100454. https://doi.org/10.1016/j.animal.2022.100454

Jones BW (2017) Behavioral Gait Change Characterization and Detection Using Precision Dairy Monitoring Technologies. PhD Diss. Univ. Kentucky Lexington. https://doi.org/10.13023/ETD.2017.264

Jonsson NN and Linardopoulou K (2023) *Digital Systems for Lameness Detection*. Conference presentation, Leipzig, Germany.

Kaur U, Malacco VMR, Bai H, Price TP, Datta A, Xin L, Sen S, Nawrocki RA, Chiu G, Sundaram S, Min BC, Daniels KM, White RR, Donkin SS, Brito LF and Voyles RM (2023) Invited review: Integration of technologies and systems for precision animal agriculture—a case study on precision dairy farming. *Journal of Animal Science* **101**, skad206. https://doi.org/10.1093/jas/skad206

Logan F, McAloon CG, Ryan EG, O'Grady L, Duane M, Deane B and McAloon CI (2024) Sensitivity and specificity of mobility scoring for the detection of foot lesions in pasture-based Irish dairy cows. *Journal of Dairy Science* **107**, 3197–3206. https://doi.org/10.3168/jds.2023-23928

Neethirajan S (2024) Artificial intelligence and sensor innovations: Enhancing livestock welfare with a human-centric approach. *Human-Centered Intelligent Systems* **4**, 77–92. https://doi.org/10.1007/s44230-023-00050-2

Niloofar P, Francis DP, Lazarova-Molnar S, Vulpe MC, Vochin G, Suciu M, Balanescu V, Anestis T and Bartzanas T (2021) Data-driven decision support in livestock farming for improved animal health, welfare, and greenhouse gas emissions: Overview and challenges. *Computers and Electronics in Agriculture* **190**, 106406. https://doi.org/10.1016/j.compag.2021.106406

O'Connor AH, Shalloo L, Bokkers EAM, de Boer IJM, Hogeveen H, Sayers R, Byrne N and Ruelle E (2023) Modeling the economic impacts of mobility scores in dairy cows under Irish spring pasture-based management. *Journal of Dairy Science* **106**, 1218–1232. https://doi.org/10.3168/jds.2021-21531

O'Leary NW, Byrne DT, O'Connor AH and Shalloo L (2020) Invited review: Cattle lameness detection with accelerometers. *Journal of Dairy Science* **103**, 3895–3911. https://doi.org/10.3168/jds.2019-17123

Post C, Rietz C, Büscher W and Müller U (2020) Using sensor data to detect lameness and mastitis treatment events in dairy cows: A comparison of classification models. *Sensors* **20**, 3863. https://doi.org/10.3390/s20143863

Post C, Rietz C, Büscher W and Müller U (2021) The importance of low daily risk for the prediction of treatment events of individual dairy cows with sensor systems. *Sensors* **21**, 1389. https://doi.org/10.3390/s21041389

Rial C, Stangaferro ML, Thomas MJ and Giordano JO (2024) Effect of automated health monitoring based on rumination, activity, and milk yield alerts

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 04:13:41, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0022029926102258

versus visual observation on herd health monitoring and performance outcomes. *Journal of Dairy Science* **107**, 11576–11596. https://doi.org/10.3168/jds.2024-25256

Rollin E, Dhuyvetter KC and Overton MW (2015) The cost of clinical mastitis in the first 30 days of lactation: An economic modeling tool. *Preventive Veterinary Medicine* **122**, 257–264. https://doi.org/10.1016/j.prevetmed.2015.11.006

Romero AR, Deypalan Y, Mehrotra MN, Jungao JT, Sheils NE, Manduchi E and Moore JH (2022) Benchmarking AutoML frameworks for disease prediction using medical claims. *BioData Mining* **15**, 15. https://doi.org/10.1186/s13040-022-00300-2

Rutten CJ, Steeneveld W, Oude Lansink AGJM and Hogeveen H (2018) Delaying investments in sensor technology: The rationality of dairy farmers' investment decisions illustrated within the framework of real options theory. *Journal of Dairy Science* **101**, 7650–7660. https://doi.org/10.3168/jds.2017-13358

Schillings J, Bennett R and Rose DC (2021) Exploring the potential of precision livestock farming technologies to help address farm animal welfare. *Frontiers in Animal Science* **2**, 639678. https://doi.org/10.3389/fanim.2021.639678

Simoni A, König F, Weimar K, Hancock A, Wunderlich C, Klawitter M, Breuer T, Drillich M and Iwersen M (2024) Evaluation of sensor-based health monitoring in dairy cows: Exploiting rumination times for health alerts around parturition. *Journal of Dairy Science* **107**, 6052–6064. https://doi.org/10.3168/jds.2023-24313

Stachowicz J and Umstätter C (2021) Do we automatically detect health- or general welfare-related issues? A framework. *Proceedings of the Royal Society B: Biological Sciences* **288**, 20210190. https://doi.org/10.1098/rspb.2021.0190

Stygar AH, Gómez Y, Berteselli GV, Dalla Costa E, Canali E, Niemi JK, Llonch P and Pastell M (2021) A systematic review on commercially available and validated sensor technologies for welfare assessment of dairy cattle. *Frontiers in Veterinary Science* **8**, 634338. https://doi.org/10.3389/fvets.2021.634338

van der Voort M, Jensen D, Kamphuis C, Athanasiadis I, De Vries A and Hogeveen H (2021) Invited review: Toward a common language in data-driven mastitis detection research. *Journal of Dairy Science* **104**, 10449–10461. https://doi.org/10.3168/jds.2021-20311

Van Nuffel A, Zwertvaegher I, Pluym L, Van Weyenberg S, Thorup VM, Pastell M, Sonck B and Saeys W (2015) Lameness detection in dairy cows: Part 1. *Animals* **5**, 838–860. https://doi.org/10.3390/ani5030387

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 04:13:41, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0022029926102258