Eurasia Specialized Veterinary Publication

International Journal of Veterinary Research and Allied Science

ISSN:3062-357X

2026, Volume 6, Issue 1, Page No: 168-180

Available online at: www.esvpub.com/

Wearable and Contactless Technologies for Detecting Pain, Lameness, Stress, Disease, and Behavioral Change in Animals: An Evidence-Mapping Review of Sensor Modalities, Validation Maturity, Clinical Readiness, and Research Gaps, 2017–2026

Sara Ben Youssef[1]*, Amal Trabelsi[1], Karim Boudiaf[2], Nabil Jebali[3]

[1]Department of Veterinary Wearable Technologies and Sensor Validation, Faculty of Veterinary Medicine, University of Tunis, Tunis, Tunisia.
[2]Department of Veterinary Contactless Monitoring and Animal Behavior, Faculty of Veterinary Medicine, University of Sousse, Sousse, Tunisia.
[3]Department of Veterinary Clinical Readiness and Implementation Science, Faculty of Veterinary Medicine, University of Sfax, Sfax, Tunisia.

***E-mail** ✉ sara.benyoussef@inat.ucar.tn

ABSTRACT

Wearable and contactless sensing technologies increasingly quantify animal movement, posture, activity, temperature, facial characteristics, vocal or respiratory events, and other observable signals potentially relevant to pain, lameness, stress, disease, and behavioral change. However, technical recognition performance, biological interpretation, external validation, and clinical actionability are frequently treated as if they represent the same evidential stage. This evidence-mapping review examined sensor modalities, species coverage, target outcomes, validation maturity, and clinical readiness in peer-reviewed veterinary and animal-health literature published from 2017 to 2026. The verified search identified 105 result appearances; 10 duplicate appearances were removed, leaving 95 unique records for screening. Forty-five reports underwent detailed eligibility assessment and 36 evidence units were included. The map comprised 26 primary empirical studies and 10 reviews or broader syntheses, with primary evidence concentrated in cattle. Wearable accelerometry, inertial sensing, computer vision, thermal imaging, bioacoustic monitoring, and multimodal systems addressed behavior, lameness, pain, health alerts, disease-related signs, and related states, but the evidence was uneven across species and endpoints. A proposed validation-maturity classification separates proof-of-concept and internal validation from independent-animal, external, multisite, and prospective implementation evidence; a complementary proposed clinical-readiness classification separates technical measurement from decision-support and demonstrated outcome utility. The evidence indicates that strong within-study classification does not by itself establish transportability, diagnosis, or clinical readiness. Major boundaries concern reference standards, individual-animal leakage, case-mix shift, environmental variability, measurement stability, and reliance on behavioral or physiological proxies. Future progress requires prospective multisite validation tied to clinically meaningful reference standards and explicit evaluation of how sensor-derived alerts alter veterinary decisions and outcomes.

Keywords: Precision livestock farming, Accelerometry, Computer vision, Infrared thermography, Animal pain, Validation

Received: 29 March 2026
Revised: 19 June 2026
Accepted: 25 June 2026

How to Cite This Article: Ben Youssef S, Trabelsi A, Boudiaf K, Jebali N. Wearable and Contactless Technologies for Detecting Pain, Lameness, Stress, Disease, and Behavioral Change in Animals: An Evidence-Mapping Review of Sensor Modalities, Validation Maturity, Clinical Readiness, and Research Gaps, 2017–2026. Int J Vet Res Allied Sci. 2026;6(1):168-80. https://doi.org/10.51847/DeRDwCPyGK

Introduction

Automated animal monitoring now encompasses wearable, imaging, acoustic, and algorithmic systems intended to recognize lameness, health change, and behavior across veterinary and production settings [1–3]. Contactless computer vision has further expanded observation without requiring a device on the animal, while introducing environmental and deployment dependencies [4]. The central problem is therefore no longer whether sensors can generate discriminative signals, but what those signals establish. This review separates measurement from interpretation, internal performance from transportability, and detection from clinical actionability, and maps how technology, species, outcome, validation maturity, and readiness intersect across the contemporary evidence base.

Rationale for evidence mapping

The literature contains fundamentally different evidential objects: validation of a behavioral sensor, classification of an observable state, prediction of a disease-associated event, and evaluation of an alert against clinical assessment. Treating these as one performance literature obscures where evidence is concentrated and where translation remains incomplete. Evidence mapping was therefore used to preserve modality, species, endpoint, setting, reference-standard, and validation differences while allowing cross-domain comparison. The resulting maturity and readiness classifications are proposed analytical tools for this review rather than externally validated grading systems.

Review questions

The review addressed five linked questions. Which wearable and contactless technology classes have been evaluated in animals? Which species, production or clinical settings, and target outcomes dominate the evidence? What reference standards and validation designs were used? How far has each evidence pattern progressed from technical recognition toward external validation and clinically interpretable use? Finally, where do gaps remain in species coverage, transportability, measurement stability, biological interpretation, and implementation? These questions intentionally distinguish sensor-observable signals from the clinical or welfare states that investigators may seek to infer from them.

Materials and Methods

An evidence-mapping review was conducted for literature published from 2017 through 2026. Searches were executed on 22 August 2026 through web-indexed scholarly retrieval, with PubMed and publisher records used to verify article identity, bibliographic metadata, and DOI information. Eligible journals were checked for Q1 standing in a relevant recognized category. Records were screened for peer review, date, sensor relevance, species and outcome compatibility, methodological interpretability, DOI availability, and inferential fit. Study-level charting covered technology, species, outcome, comparator or reference standard, validation setting, external-validation status, proposed maturity class, and proposed clinical-readiness class.

Information sources

The search deliberately included evidence addressing technical performance and the conditions that alter field interpretation. Extensive livestock systems, for example, introduce connectivity, range, power, and sampling constraints that influence automatic monitoring feasibility [5]. Validation scholarship was included because precision, bias, and reproducibility are inconsistently assessed across dairy behavior-sensor studies [6]. Computer-vision literature was also searched beyond strictly diagnostic applications because cattle imaging systems increasingly span behavior, production, health, and management functions while still requiring real-farm validation [7]. This breadth was necessary to distinguish measurement capability from practical readiness.

Search strategy

Six targeted retrieval clusters were used: wearable accelerometry and inertial measurement; multimodal precision-livestock sensing; computer vision and video; infrared thermography and contactless physiological sensing; acoustic or audio-visual monitoring; and validation, transportability, clinical-readiness, and implementation evidence. Searches combined veterinary or animal terms with technology terms and the outcomes pain, lameness, stress, disease, health alert, and behavior. Records were consolidated by article identity and DOI before screening. Search terms were intentionally broad enough to capture studies measuring precursor or proxy signals while eligibility assessment preserved the distinction between those signals and confirmed clinical states.

Eligibility criteria

Eligible evidence comprised peer-reviewed journal articles published from 2017 to 2026 in journals meeting the prespecified Q1 criterion, with a verifiable DOI and direct relevance to at least one mapped technology, outcome, validation, or implementation domain. Species, setting, measurement process, and target construct had to be sufficiently described to bound interpretation. Conference abstracts, dissertations, preprints, noneligible journals, duplicate article versions, and topic-adjacent papers without an exact evidential role were excluded. Behavioral or physiological proxies were retained only when their relationship to the claimed health or welfare state could be represented without treating the proxy as equivalent to diagnosis.

Technology taxonomy

Technologies were charted into wearable accelerometry, inertial measurement technologies, location or proximity sensing, RGB or video computer vision, thermal or infrared imaging, bioacoustic sensing, contactless physiological sensing, and multimodal fusion. Classification followed the primary measurement architecture rather than the commercial purpose of the product. Systems integrating two or more independently informative streams could receive a multimodal designation while preserving their component modalities. This taxonomy is proposed for evidence mapping; it does not imply equivalent biological meaning, performance, invasiveness, or clinical utility across technology classes.

Animal and species classification

Species was treated as an analytical boundary rather than a descriptive label. Independent-animal validation of equine inertial sensing showed that performance differed markedly by target behavior within the same system [8]. In grazing dairy cows, collar accelerometry was validated for specific behaviors under a particular production setting, supporting setting-bounded rather than universal interpretation [9]. A sheep accelerometry proof-of-concept further illustrated the danger of extending technically successful behavior recognition to biological stress inference when the stress-validation subset was very small [10]. Species, management context, phenotype, and target behavior were therefore charted jointly.

Table 1 presents species, setting, and target behavior represented in the veterinary sensing evidence map.

Table 1. Species, setting, and target behavior represented in the veterinary sensing evidence map

Species and setting	Sensor target	Data-generating conditions	Biological meaning	Clinical or screening role	Transfer test
Dairy cattle in managed housing	Activity, locomotion or posture change	Herd management, flooring, disease prevalence	Signal may reflect behavior or health change	Treat as screening or monitoring evidence	**Transfer limit:** Farm and case-mix dependence **Revalidation:** Recheck against animal examination or reference assessment
Grazing dairy cattle	Grazing and rumination pattern	Pasture conditions, device placement, management	Validity is behavior- and environment-specific	Use only for validated target behaviors	**Transfer limit:** Intensive-housing transfer not assumed **Revalidation:** Revalidate after major management change
Stabled horses	Recumbency, standing or weight-shifting behavior	Behavior frequency and class imbalance	Performance differs among behaviors	Avoid one accuracy label for all states	**Transfer limit:** Stable context and behavior set **Revalidation:** Reassess poorly recognized behaviors separately
Sheep or lambs	Activity or sickness-associated behavior	Small biological-validation groups, environment	Behavioral classification may exceed stress-state validation	Interpret stress inference cautiously	**Transfer limit:** Cohort and challenge context **Revalidation:** Require independent biological validation
Cats in clinical populations	Facial configuration associated with pain	Individual appearance and clinical heterogeneity	Individual variation may challenge generalization	Use as adjunct rather than autonomous diagnosis	**Transfer limit:** Dataset and patient mix **Revalidation:** Reassess in independent heterogeneous patients
Dogs assessed thermally	Ocular or facial thermal signal	Acquisition conditions and measurement stability	Temperature variation requires technical contextualization	Confirm before clinical attribution	**Transfer limit:** Camera, environment and anatomical site **Revalidation:** Repeat or

					standardize acquisition when unstable
Pigs or poultry under camera/acoustic monitoring	Movement, posture or respiratory-event signal	Housing, occlusion, background noise	Observable event may not identify underlying disease	Escalate to confirmatory assessment when warranted	**Transfer limit:** Facility and endpoint specificity **Revalidation:** Revalidate after deployment-context shift

Note: The interpretive and reassessment categories constitute a proposed synthesis of the mapped evidence and are not validated clinical scoring rules.

Outcome classification

Outcomes were coded according to the level of inference required. Directly observable behavior, posture, movement, temperature, facial configuration, or acoustic events were distinguished from composite states such as lameness, pain, stress, sickness, and diagnosed disease. A further distinction separated detection or classification from prediction and from sensor-generated alerts intended to prompt action. This hierarchy prevents a technically accurate proxy from being presented as a confirmed clinical state. Multi-label coding was allowed when a study genuinely evaluated more than one level, but the strongest inference permitted by its reference standard remained explicit.

Study selection

The verified search produced 105 result appearances. Ten duplicate article or DOI appearances were removed, leaving 95 unique records for title/abstract-equivalent screening. Fifty records were excluded at that stage and 45 reports advanced to detailed eligibility assessment; none was unavailable for assessment. Nine detailed reports were excluded, leaving 36 included evidence units. Detailed exclusions comprised four redundant records, three records with outcome or unit-of-analysis mismatch, one surrogate cough record with unresolved eligibility for the final evidence set, and one record providing insufficient sensing-specific contribution. The numerical transitions reconciled exactly.

Data charting

Each included primary study was charted for technology class, species or population, setting, sample where reported, target state, comparator or reference standard, analytical method, temporal design, validation procedure, external-validation status, and limitations. Reviews were charted for their methodological or synthesis role rather than treated as primary validation units. Data not explicitly reported were retained as not reported rather than inferred. Particular attention was given to whether validation separated animals, herds, sites, or time periods, because apparent performance can change when information from the same individuals or environment influences both model development and evaluation [6].

Validation-maturity classification

A proposed six-level maturity structure was used to prevent internal technical performance from being conflated with translation. V0 denotes synthesis without new validation; V1, proof-of-concept or internal validation; V2, independent-animal or reference-standard validation; V3, temporal, herd, site, or heterogeneous-population transportability testing; V4, prospective multisite field validation in realistic workflow; and V5, implementation or impact validation. Multi-farm camera evaluation illustrates a higher evidential state than a single internal dataset while remaining short of impact evidence [11]. Thermal measurement stability [12] and internally validated accelerometer classification [13] further show that validation concerns differ by modality.

Clinical-readiness classification

Clinical readiness was coded independently of validation maturity. The proposed classes are C0, exploratory signal or technical feasibility; C1, research-use recognition or measurement; C2, a decision-support candidate requiring confirmatory assessment; C3, a field-deployable adjunct embedded in a defined confirmatory workflow; and C4, a prospectively validated integrated system with demonstrated decision or outcome utility. Review articles were not assigned a primary-study readiness level. A system can therefore be technically mature for a narrow measurement task without being ready to diagnose disease or direct treatment. The classification is interpretive and does not represent a regulatory or professional accreditation scheme.

Evidence-mapping approach

The final map crossed five principal dimensions: technology, species or production context, target outcome, validation maturity, and clinical readiness. A study could occupy more than one technology or outcome category when justified, but each maturity and readiness judgment was anchored to the evidence actually reported. Counts were generated from the study-level coding ledger rather than inferred from narrative prominence. The approach was designed to expose asymmetry—for example, abundant technical recognition evidence but little external validation—without converting heterogeneous designs into a pooled performance estimate or implying that a numerically larger evidence cluster is necessarily stronger.

Results and Discussion

Thirty-six evidence units met the eligibility criteria: 26 primary empirical studies and 10 reviews or broader syntheses. Among primary studies, cattle predominated, with 17 units; sheep or lambs and cats contributed two each, pigs two, and horses, dogs, and goats one each. Poultry was represented principally through review-level computer-vision evidence rather than an included primary validation study. The mapped literature spanned wearable, imaging, thermal, acoustic, and multimodal sensing, but distributions were uneven across species, outcomes, and validation depth. The most consequential result was therefore not a single performance estimate, but the recurrent separation between signal-recognition capability and evidence for external transportability or clinically actionable use.

Table 2 presents what evidence density, behavior recognition, external testing, disease proxies, and impact studies each establish.

Table 2. What evidence density, behavior recognition, external testing, disease proxies, and impact studies each establish

Mapped evidence cluster	Unit of evidence	Generalisability dimension	Closest legitimate use	Missing validation layer
Dense cattle evidence	Multiple sensing approaches studied in cattle	**Population/system:** Dairy and production-system concentration **Interpretation:** Strongest evidence density is species-specific	Prioritize cattle-specific conclusions	**Unresolved:** Transfer to other species uncertain **Needed next:** Replicate in underrepresented species
Successful behavior recognition	Sensors can classify defined behaviors	**Population/system:** Behavior frequency, setting, device placement **Interpretation:** Technical measurement evidence	Use for validated behavior task	**Unresolved:** Clinical meaning may remain indirect **Needed next:** Link to independent biological reference
High internal model performance	Algorithm separates study-defined classes	**Population/system:** Dataset composition and individual overlap **Interpretation:** Internal discrimination, not transportability	Do not infer deployment accuracy	**Unresolved:** Leakage and case-mix shift **Needed next:** External-animal or external-site testing
Multi-farm or heterogeneous testing	Model evaluated across broader populations	**Population/system:** Herd, site and patient heterogeneity **Interpretation:** Stronger evidence of transportability	Consider as adjunct if reference standard is adequate	**Unresolved:** Geographic and workflow limits remain **Needed next:** Prospective field confirmation
Disease-associated proxy or alert	Behavior, temperature or another signal changes near disease	**Population/system:** Timing, specificity, background prevalence **Interpretation:** Screening or escalation signal	Trigger confirmatory assessment	**Unresolved:** Signal may have alternative causes **Needed next:** Clinical adjudication
Evidence short of impact validation	Technical or field validity reported without outcome study	**Population/system:** Workflow and user response **Interpretation:** Readiness remains conditional	Avoid claims of improved health or welfare	**Unresolved:** Action-effect chain untested **Needed next:** Prospective implementation evaluation

Note: Response and reassessment categories are proposed interpretations derived from the evidence map; they are not validated clinical directives.

Study-selection map

The selection flow preserved a distinction between retrieval volume and the final evidence map. The 45 detailed reports contained studies and syntheses spanning poultry computer vision, predictive infrared thermography, and audio-visual cough monitoring, illustrating the breadth of technologies considered before final synthesis [14–16].

Nine were excluded at detailed assessment using the prespecified inferential and eligibility boundaries, leaving 36 mapped evidence units. **Figure 1** is therefore an accounting figure rather than a conceptual model: its scientific role is to make every transition, exclusion stage, and final inclusion count visible without implying that selection volume measures evidential strength.

Figure 1 provides the auditable pathway from search retrieval to the final technology–species evidence map while keeping screening exclusions separate from scientific classification.

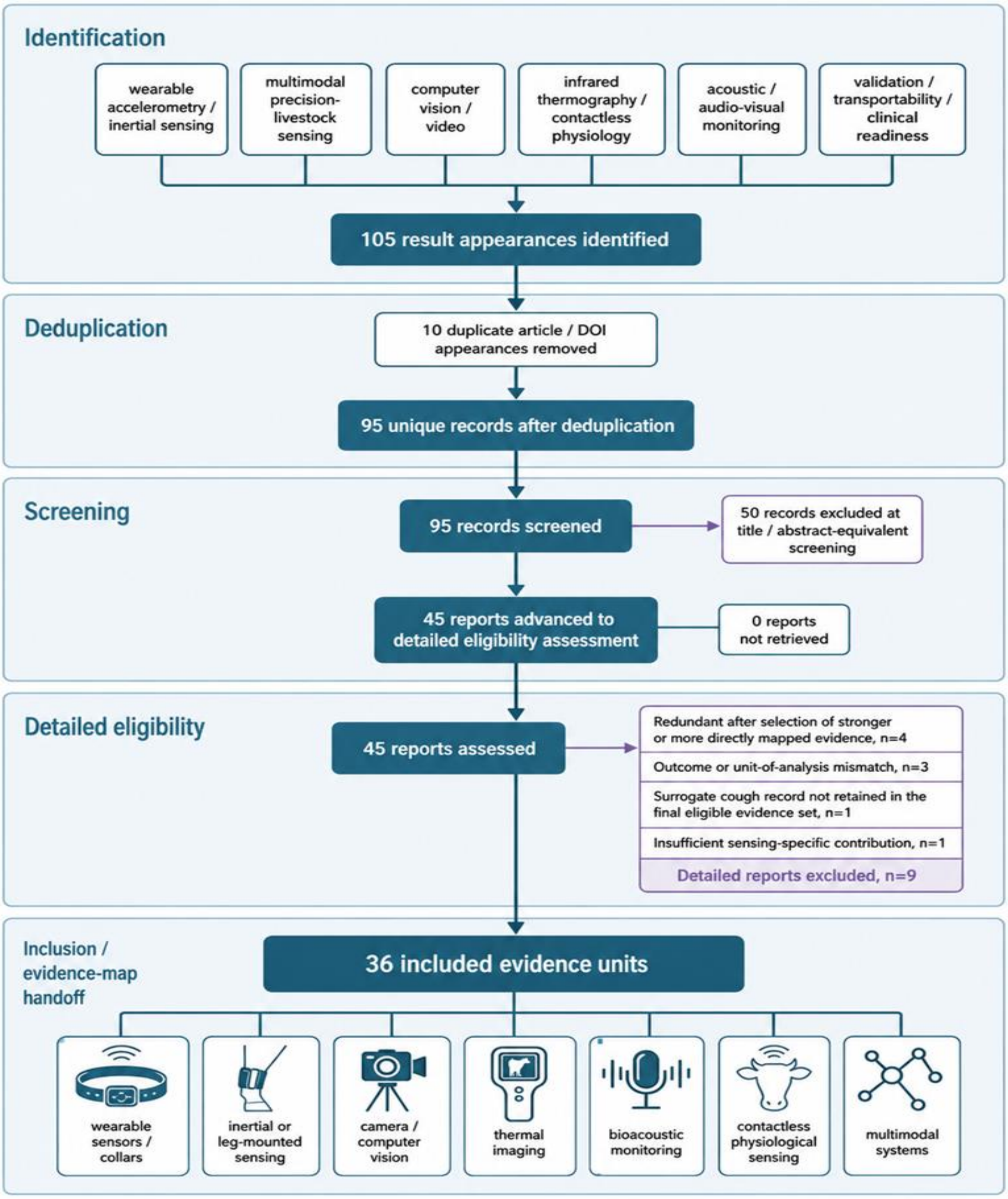

Figure 1. Arithmetic path from 105 retrieved result appearances to 36 mapped wearable and contactless animal-sensing evidence units

Species distribution

Primary empirical evidence was strongly concentrated in cattle: 17 of 26 primary studies involved cattle, whereas sheep or lambs and cats contributed two studies each, and pigs, horses, dogs, and goats were represented by smaller numbers. This distribution is scientifically important because species differences in anatomy, behavior, husbandry, clinical presentation, and sensor placement can alter both measurement and interpretation. Review-level poultry evidence broadened the technological landscape but did not compensate for the scarcity of primary non-cattle validation. Cross-species transfer should therefore be treated as an empirical question rather than assumed from shared sensor architecture.

Wearable accelerometry

Accelerometry provides continuous, relatively low-burden measurement of movement and behavior, but its strongest evidence concerns recognition of defined activity states rather than direct diagnosis. Grazing-collar validation demonstrates that device performance can be credible for specified behaviors within a defined production context [9]. Likewise, machine-learning classification from body-mounted accelerometers supports automated posture and behavior recognition while remaining primarily an internal measurement-validation problem [13]. The evidence map therefore treats accelerometer-derived activity change as an observable layer that may contribute to lameness, sickness, stress, or welfare assessment only when an appropriate biological or clinical reference establishes the additional inferential step.

Inertial measurement technologies

Inertial measurement units extend accelerometry by combining movement signals that can characterize posture, orientation, and transitions. Independent-animal testing in stabled horses demonstrated that recognition quality differed substantially among behaviors, with some states captured more reliably than others [8]. This variation is analytically important: additional sensor channels do not remove dependence on behavior frequency, placement, reference annotation, or species-specific movement patterns. IMU evidence should therefore be interpreted at the level of the validated behavior and population rather than as a general measure of animal state.

Location and proximity sensors

Location and proximity technologies occupy a less densely validated region of the mapped primary evidence than accelerometry or imaging. Wearable-sensor reviews nevertheless identify spatial sensing as part of the expanding animal-monitoring ecosystem [3]. Extensive production systems create particular constraints because communication range, connectivity, power availability, and sampling architecture may determine whether spatial data can be collected consistently [5]. Location therefore provides context about movement and interaction rather than direct proof of pain, disease, or welfare state; its clinical value depends on how spatial change is linked to independent biological evidence.

Computer vision

Computer vision spans visible lesion recognition, locomotor analysis, and automated behavioral measurement. Comparative algorithms have been evaluated for real-time digital-dermatitis detection, demonstrating that model choice matters within a defined disease-imaging task [17]. A single side-view camera has also supported automated cattle lameness classification under study conditions [18]. Top-view keypoint extraction can quantify movement features against visual mobility scores [19]. Conversely, strong internal classification of lameness from back-shape characteristics does not establish external farm performance or clinical utility [20]. The mapped evidence therefore separates image-derived measurement, classifier validation, transportability, and actionability.

Thermal imaging

Infrared thermography offers non-contact measurement of surface temperature patterns, but interpretation depends on acquisition stability and biological specificity. Repeated canine measurements show that ocular thermal readings can vary with technical and contextual factors [12]. In dairy cattle, thermal imaging has also been used with artificial intelligence to detect or anticipate digital-dermatitis-associated changes within a defined study setting [15]. These studies occupy different evidential positions: one addresses measurement reliability and the other disease-related prediction. Neither justifies a universal thermal threshold for pain, stress, inflammation, or disease across species or environments.

Bioacoustic monitoring

Bioacoustic sensing can convert vocal or respiratory events into continuously detectable signals without attaching hardware to the animal. Audio-visual multimodal monitoring in pigs demonstrated automated cough-event recognition, supporting the technical feasibility of combining acoustic and visual information [16]. The clinically important boundary is etiological: cough is an observable event with multiple possible causes rather than a diagnosis. Acoustic systems are therefore most defensible as surveillance or escalation tools unless disease attribution is independently adjudicated.

Table 3 presents acoustic respiratory-event evidence from isolated cough detection to synchronized audio–visual confirmation.

Table 3. Acoustic respiratory-event evidence from isolated cough detection to synchronized audio–visual confirmation

Respiratory-event question	Isolated detected cough	Recurrent cough pattern	Audio–visual concordance	Deployment-context change
Signal actually observed	Classified acoustic event	Repeated detected events	Cough plus compatible visual event	Shift in acoustic environment
Veterinary meaning permitted	Respiratory signal only	Possible health concern	Stronger event verification	Model reliability uncertain
Acoustic or housing dependency	Noise, overlap, microphone position	Group density and background acoustics	Camera visibility and synchronization	Housing, equipment, stocking
Escalation use	Continue surveillance	Escalate observation	Prioritize assessment	Withhold strong inference
Confirmation boundary	Recheck persistence and context	Confirm at animal level	Disease cause remains unconfirmed	Revalidate after context shift

Note: The escalation and reassessment logic is a proposed synthesis; it is not a validated clinical decision rule.

Contactless physiological sensing

Contactless physiological sensing is attractive where attachment is impractical or itself alters behavior, yet absence of physical contact does not eliminate measurement error. Thermal-eye measurements in dogs illustrate how acquisition stability, camera conditions, and anatomical targeting can influence recorded physiological signals [12]. Such measurements should therefore remain conceptually separate from the state they are intended to represent. A change in temperature, movement, respiration, or another remote signal becomes clinically interpretable only after the measurement process is sufficiently stable and alternative physiological or environmental explanations have been addressed.

Pain detection

Pain recognition demonstrates particularly clearly why validation design matters. Reviews and video-based studies show technical promise while identifying ground-truth, dataset, and individual-animal generalization constraints; in goats, performance was notably more demanding under subject-wise validation [21–23]. Explainable modeling in a heterogeneous population of client-owned cats provides a more clinically realistic test of facial-pain recognition than narrowly homogeneous datasets, but it still does not establish prospective clinical benefit [24]. The evidence therefore supports pain-recognition assistance as an investigational measurement problem rather than an autonomous diagnostic replacement for veterinary assessment.

Lameness detection

Lameness has one of the broadest technology portfolios in the mapped literature, including camera, locomotor-feature, and wearable approaches. Reviews emphasize continuing heterogeneity in reference standards [1]. Multi-farm camera evaluation moves beyond purely internal model development but remains distinct from intervention-effect evidence [11]. Single-camera classification [18], keypoint-derived mobility measurement [19], and back-shape analysis [20] illustrate complementary visual strategies. Crucially, locomotion scoring, automated classification, and identification of underlying painful foot pathology are different endpoints. Clinical interpretation should retain that distinction rather than equating a mobility signal with lesion diagnosis.

Stress detection

Stress was less directly validated than observable behavior. Sheep accelerometry showed that behaviors could be classified while the biological stress component rested on a very small validation subset [10]. Contactless thermal measurement likewise introduces acquisition and physiological variability that must be resolved before temperature changes are attributed specifically to stress [12]. Accordingly, the evidence map treats stress as a higher-inference state than movement, posture, or temperature itself. Robust stress detection would require independent biological or behavioral reference standards capable of separating stress responses from disease, exertion, handling, environment, and normal individual variation.

Disease detection

Disease-related sensing ranged from lesion imaging to symptom-event recognition. Infrared thermography has been investigated for early digital-dermatitis-associated changes [15], while computer-vision algorithms have been compared for direct visual detection of digital dermatitis [17]. Audio-visual systems can detect cough events [16], but such events remain nonspecific unless linked to an appropriate diagnostic reference. These examples support a layered interpretation: sensors may identify lesions, physiological deviations, or disease-associated behaviors, yet the inferential distance between measured signal and confirmed disease differs materially across tasks and should remain explicit.

Behavioral monitoring

Wearable behavioral monitoring constitutes a comparatively mature technical application, although biological interpretation remains contextual. Ear-tag accelerometry demonstrates that cattle activity patterns vary with management and environmental conditions [25]. Grazing lamb studies show that sickness can generate detectable behavioral changes without making those changes etiologically unique [26]. Deep-learning models can classify defined dairy-cow behaviors from accelerometer data [27]. Sensor-generated rumination alerts around parturition have also been evaluated against standardized clinical monitoring [28]. Together, these studies support continuous observation while reinforcing the distinction between recognized behavior, health alert, and confirmed clinical condition.

Validation maturity

Across modalities, maturity was constrained less by the existence of algorithms than by the independence and realism of validation. Reviews identify inconsistent attention to precision, bias, and reproducibility [6]. Subject-wise pain validation shows how performance can weaken when the test animal is genuinely independent of development data [23]. Multi-farm camera evaluation provides stronger ecological evidence than internal validation while remaining short of prospective impact assessment [11]. The proposed V0–V5 structure therefore treats independent animals, temporal separation, external sites, realistic workflow, and outcome evaluation as progressively different evidential questions rather than a single accuracy continuum.

Table 4 presents validation layers from internal discrimination to sensor-guided outcome impact.

Table 4. Validation layers from internal discrimination to sensor-guided outcome impact

Validation layer	Population separation achieved	Strongest sensor claim	Remaining transport or workflow threat	Deployment role
V1 internal	Within-dataset discrimination	Technical feasibility	**Threat:** Leakage, imbalance **Next boundary:** Independent-animal test	Research use
V2 independent-animal	Performance on separated animals/reference data	Individual-level validity	**Threat:** Reference quality **Next boundary:** External site or time	Candidate decision support
V3 transportability	Herd, site, temporal, or heterogeneous testing	Broader generalizability evidence	**Threat:** Case mix, management **Next boundary:** Prospective workflow test	Conditional field use
V4 prospective field	Realistic multisite workflow validation	Field validity	**Threat:** User response, prevalence **Next boundary:** Outcome evaluation	Defined adjunctive use
V5 impact	Sensor-guided action linked to meaningful outcomes	Implementation utility	**Threat:** Workflow and intervention fidelity **Next boundary:** Ongoing surveillance	Strongest readiness claim

Note: V1–V5 constitute the proposed review classification and are not regulatory categories.

Clinical-readiness distribution

The evidence map places most primary systems below the level required to demonstrate clinical outcome utility. Some technologies have progressed from technical recognition toward external or field validation, but many remain best understood as research measurements or decision-support candidates. Health-alert evaluation against standardized clinical monitoring demonstrates the importance of preserving a confirmatory step between sensor output and clinical interpretation [28]. Clinical readiness therefore depends not only on discrimination performance but also on reference-standard quality, transportability, alert consequences, workflow integration, and evidence that users can act on outputs without introducing unacceptable error.

Evidence gaps

Four gaps dominate the map. First, primary evidence is heavily concentrated in cattle, limiting cross-species inference. Second, lameness reviews continue to identify reference-standard and pathology-linkage weaknesses [1]. Third, poultry computer-vision synthesis highlights continuing challenges in real-farm generalization despite growing technical capability [14]. Fourth, pain-recognition literature remains constrained by ground truth and heterogeneous individual expression [21]. Across these domains, the largest missing step is often not another internally optimized classifier but independent, multisite validation connecting measurable signals to clinically meaningful and explicitly adjudicated animal states.

Figure 2 condenses these gaps into a layered evidence architecture showing where a sensor signal can legitimately support interpretation and where additional validation or reassessment is required.

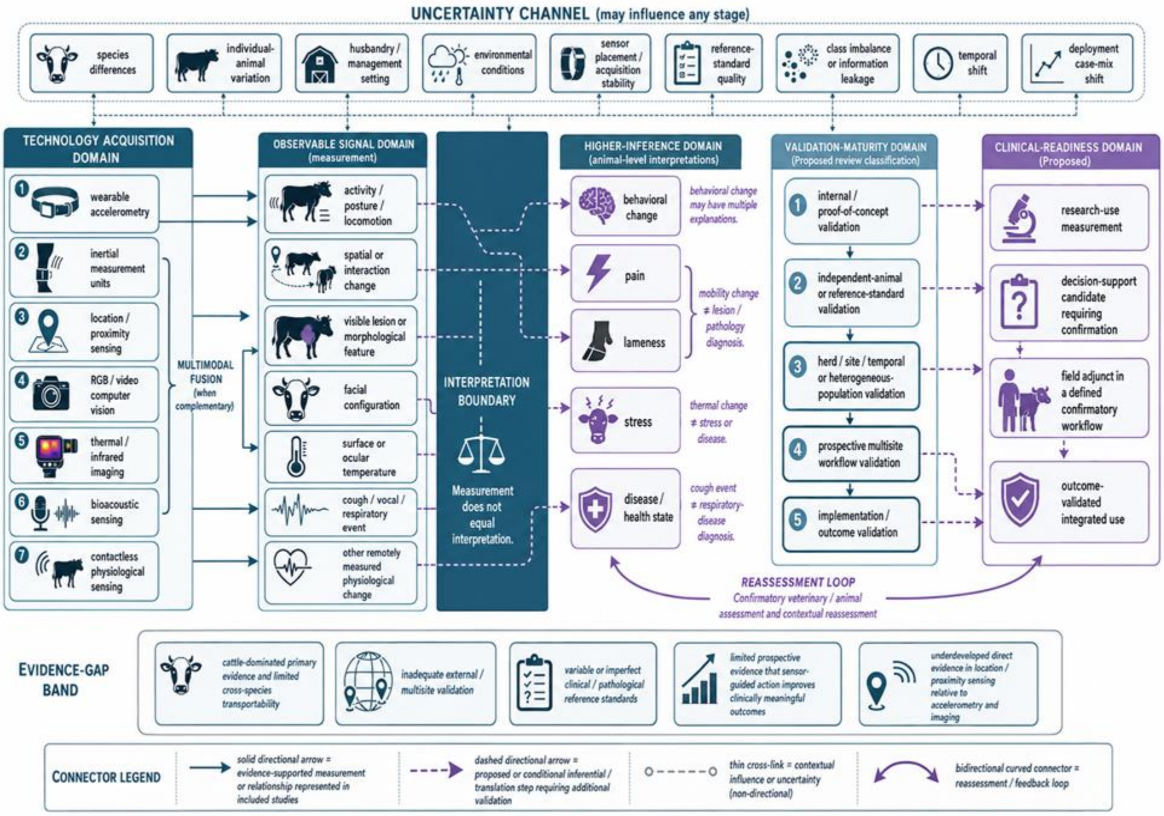

Figure 2. Evidence discontinuities that separate animal-sensor signals from defensible veterinary action

The evidence map indicates that generalization is a distinct scientific problem rather than an automatic consequence of good internal performance: lameness-model precision declined when evaluation expanded to a broader deployment population [29]. Multimodal thermal and behavioral information can provide complementary disease-related signals, as illustrated in bovine respiratory-disease assessment [30]. Controlled-environment pig vision studies demonstrate technical capability while leaving environmental transfer uncertain [31]. Pain-recognition research similarly remains dependent on species-specific expression, labeling, and reference standards [32]. Together, these findings support validation architectures that test where models fail, not only where they discriminate successfully.

Translational barriers

Translation is constrained simultaneously by infrastructure, environment, workflow, and population shift. Extensive systems may limit communication range, power, and data continuity [5]. Computer-vision deployment introduces lighting, occlusion, data-integration, and real-time operational requirements [7]. Even after technical deployment, broader farm populations can alter model performance [29]. These barriers operate at different levels and should not be collapsed into a single "implementation" problem. A system may be technically reliable yet operationally fragile, externally valid yet clinically ambiguous, or clinically informative while still lacking evidence that its use improves an outcome.

Minimum validation requirements

The review therefore proposes a minimum sequence before strong readiness claims: clearly define the target construct; use an appropriate independent reference standard; separate animals during validation when individual identity can leak information; test temporal or site transportability; and prospectively evaluate performance within intended workflow. Existing validation reviews support stronger attention to precision, bias, and reproducibility [6]. Subject-wise pain testing demonstrates why independent-animal evaluation matters [23]. Broader-population lameness testing shows why external deployment can reveal performance loss [29]. These requirements are proposed evidential safeguards, not regulatory standards.

Research priorities

Research should move from repeated demonstrations of internal classification toward questions that determine veterinary usefulness. Priority areas include pathology-grounded lameness reference standards, prospective non-cattle validation, heterogeneous clinical pain datasets, and real-farm computer-vision replication. Lameness synthesis identifies persistent reference-standard limitations [1]. Poultry computer-vision evidence emphasizes the need for generalization beyond controlled or study-specific environments [14]. Pain-recognition literature likewise requires stronger ground truth and external testing [21]. Future studies should also document false-alert consequences, clinician responses, animal-level escalation pathways, and whether sensor-guided decisions improve outcomes compared with existing assessment.

Limitations

The map itself inherits limitations from its evidence base. Six-farm longitudinal accelerometer research improves ecological relevance but remains geographically and production-system bounded [33]. Metritis prediction demonstrates that preprocessing choices and model selection can materially change results [34]. Sensor-based welfare work shows weaker performance when models encounter unseen herds and cautions against stand-alone interpretation [35]. Feeding-behavior classification illustrates a broader scope limitation: accurate recognition of one behavior does not validate disease, pain, stress, or welfare inference [36]. The search was targeted rather than a native export from every subscription bibliographic database, so completeness beyond the executed retrieval scope cannot be assumed.

Conclusion

Wearable and contactless technologies can measure increasingly diverse animal signals, but signal recognition, biological interpretation, external validation, and clinical actionability remain distinct evidential stages. The proposed maturity and readiness classifications make those boundaries explicit. Current evidence supports substantial measurement and decision-support potential, especially in cattle, while multisite validation, stronger reference standards, cross-species testing, and prospective evaluation of sensor-guided decisions remain necessary before stronger claims of clinical or welfare benefit.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Siachos N, Neary JM, Smith RF, Oikonomou G. Automated dairy cattle lameness detection utilizing the power of artificial intelligence; current status quo and future research opportunities. Vet J. 2024;304:106091. doi:10.1016/j.tvjl.2024.106091
2. Ferreira C, Todorovic M, Sugrue P, Teixeira S, Galvin P. Review: Emerging sensors and instrumentation systems for bovine health monitoring. Animal. 2025;19(6):101527. doi:10.1016/j.animal.2025.101527
3. Ding L, Zhang C, Yue Y, Yao C, Li Z, Hu Y, et al. Wearable sensors-based intelligent sensing and application of animal behaviors: A comprehensive review. Sensors (Basel). 2025;25(14):4515. doi:10.3390/s25144515
4. Antognoli V, Presutti L, Bovo M, Torreggiani D, Tassinari P. Computer vision in dairy farm management: A literature review of current applications and future perspectives. Animals (Basel). 2025;15(17):2508. doi:10.3390/ani15172508
5. Mancuso D, Castagnolo G, Porto SMC. Cow behavioural activities in extensive farms: Challenges of adopting automatic monitoring systems. Sensors (Basel). 2023;23(8):3828. doi:10.3390/s23083828
6. Lee A, Brause M, Foy D, Cantor MC. Review: Establishing precision, bias, and reproducibility standards for dairy cattle behavior sensors. Animal. 2025;19 Suppl 2:101613. doi:10.1016/j.animal.2025.101613
7. Guarnido-Lopez P, Pi Y, Tao J, Mendes EDM, Tedeschi LO. Computer vision algorithms to help decision-making in cattle production. Anim Front. 2024;14(6):11-22. doi:10.1093/af/vfae028
8. Anderson K, Morrice-West AV, Walmsley EA, Fisher AD, Whitton RC, Hitchens PL. Validation of inertial measurement units to detect and predict horse behaviour while stabled. Equine Vet J. 2023;55(6):1128-38. doi:10.1111/evj.13909
9. Iqbal MW, Draganova I, Morel PCH, Morris ST. Validation of an accelerometer sensor-based collar for monitoring grazing and rumination behaviours in grazing dairy cows. Animals (Basel). 2021;11(9):2724. doi:10.3390/ani11092724
10. Schneidewind SJ, Al Merestani MR, Schmidt S, Waser W, Schmidt T, Wiegard M, et al. Quantifying sheep behaviour using a 3D accelerometer: A proof-of-concept for objective stress assessment. Sensors (Basel). 2026;26(4):1169. doi:10.3390/s26041169
11. Swartz D, Shepley E, Parker Gaddis K, Burchard J, Cramer G. Descriptive evaluation of a camera-based dairy cattle lameness detection technology. J Dairy Sci. 2024;107(11):9847-61. doi:10.3168/jds.2024-24851
12. Rostami S, Woodruff KA, McBride A, Shi J, Smith DR. Evaluation of thermal camera measurement stability and factors associated with infrared eye temperature in dogs. Front Vet Sci. 2026;13:1848874. doi:10.3389/fvets.2026.1848874
13. Balasso P, Marchesini G, Ughelini N, Serva L, Andrighetto I. Machine learning to detect posture and behavior in dairy cows: Information from an accelerometer on the animal's left flank. Animals (Basel). 2021;11(10):2972. doi:10.3390/ani11102972
14. Paneru B, Dhungana A, Dahal S, Ritz CW, Kim W, Liu T, et al. Computer vision models for precision poultry farming: A narrative review of behavioral and welfare monitoring studies. Poult Sci. 2026;105(7):106887. doi:10.1016/j.psj.2026.106887
15. Feighelstein M, Mishael A, Malka T, Magana J, Gavojdian D, Zamansky A, et al. AI-based prediction and detection of early-onset of digital dermatitis in dairy cows using infrared thermography. Sci Rep. 2024;14(1):29849. doi:10.1038/s41598-024-80902-4
16. Chae H, Lee J, Kim J, Lee S, Lee J, Chung Y, et al. Novel method for detecting coughing pigs with audio-visual multimodality for smart agriculture monitoring. Sensors (Basel). 2024;24(22):7232. doi:10.3390/s24227232
17. Aravamuthan S, Cernek P, Anklam K, Döpfer D. Comparative analysis of computer vision algorithms for the real-time detection of digital dermatitis in dairy cows. Prev Vet Med. 2024;229:106235. doi:10.1016/j.prevetmed.2024.106235
18. Myint BB, Onizuka T, Tin P, Aikawa M, Kobayashi I, Zin TT. Development of a real-time cattle lameness detection system using a single side-view camera. Sci Rep. 2024;14(1):13734. doi:10.1038/s41598-024-64664-7
19. Higaki S, Menezes GL, Ferreira REP, Negreiro A, Cabrera VE, Dórea JRR. Objective dairy cow mobility analysis and scoring system using computer vision-based keypoint detection technique from top-view 2-dimensional videos. J Dairy Sci. 2025;108(4):3942-55. doi:10.3168/jds.2024-25545

20. Narli SS, Schmidt H, Firouzabadi A, Schönnagel L, Reich MS, Reitmaier S. Automated detection of lameness in dairy cattle through computer vision analysis of back shape characteristics. Comput Biol Med. 2025;197(Pt A):111038. doi:10.1016/j.compbiomed.2025.111038
21. Chiavaccini L, Gupta A, Chiavaccini G. From facial expressions to algorithms: A narrative review of animal pain recognition technologies. Front Vet Sci. 2024;11:1436795. doi:10.3389/fvets.2024.1436795
22. Martvel G, Lazebnik T, Feighelstein M, Henze L, Meller S, Shimshoni I, et al. Automated video-based pain recognition in cats using facial landmarks. Sci Rep. 2024;14(1):28006. doi:10.1038/s41598-024-78406-2
23. Chiavaccini L, Gupta A, Anclade N, Chiavaccini G, De Gennaro C, Johnson AN, et al. Automated acute pain prediction in domestic goats using deep learning-based models on video-recordings. Sci Rep. 2024;14(1):27104. doi:10.1038/s41598-024-78494-0
24. Feighelstein M, Henze L, Meller S, Shimshoni I, Hermoni B, Berko M, et al. Explainable automated pain recognition in cats. Sci Rep. 2023;13(1):8973. doi:10.1038/s41598-023-35846-6
25. Hu S, Reverter A, Arablouei R, Bishop-Hurley G, McNally J, Alvarenga F, et al. Analyzing cattle activity patterns with ear tag accelerometer data. Animals (Basel). 2024;14(2):301. doi:10.3390/ani14020301
26. Fan B, Bryant RH, Greer AW. Automatically identifying sickness behavior in grazing lambs with an acceleration sensor. Animals (Basel). 2023;13(13):2086. doi:10.3390/ani13132086
27. Balasso P, Taccioli C, Serva L, Magrin L, Andrighetto I, Marchesini G. Uncovering patterns in dairy cow behaviour: A deep learning approach with tri-axial accelerometer data. Animals (Basel). 2023;13(11):1886. doi:10.3390/ani13111886
28. Simoni A, König F, Weimar K, Hancock A, Wunderlich C, Klawitter M, et al. Evaluation of sensor-based health monitoring in dairy cows: Exploiting rumination times for health alerts around parturition. J Dairy Sci. 2024;107(8):6052-64. doi:10.3168/jds.2023-24313
29. Dervić E, Matzhold C, Egger-Danner C, Steininger F, Klimek P. Improving lameness detection in cows: A machine learning algorithm application. J Dairy Sci. 2024;107(12):11550-62. doi:10.3168/jds.2024-24730
30. Perez Marquez HJ, Schaefer AL, Bench CJ. Use of fidget and drinking behaviour in combination with facial infrared thermography for diagnosis of bovine respiratory disease in a spontaneous model. Animal. 2024;18(3):101096. doi:10.1016/j.animal.2024.101096
31. Alves MFA, Pandorfi H, Soares RGF, Medeiros VWC, Santana TC, Grobner VK, et al. Computer vision applied to the analysis of pig behavior patterns in an air-conditioned environment. Animals (Basel). 2026;16(9):1353. doi:10.3390/ani16091353
32. Andersen PH, Broomé S, Rashid M, Lundblad J, Ask K, Li Z, et al. Towards machine recognition of facial expressions of pain in horses. Animals (Basel). 2021;11(6):1643. doi:10.3390/ani11061643
33. Lavrova AI, Choucair A, Palmini A, Stock KF, Kammer M, Querengässer F, et al. Leveraging accelerometer data for lameness detection in dairy cows: A longitudinal study of six farms in Germany. Animals (Basel). 2023;13(23):3681. doi:10.3390/ani13233681
34. Vidal G, Sharpnack J, Pinedo P, Tsai IC, Lee AR, Martínez-López B. Impact of sensor data pre-processing strategies and selection of machine learning algorithm on the prediction of metritis events in dairy cattle. Prev Vet Med. 2023;215:105903. doi:10.1016/j.prevetmed.2023.105903
35. Stygar AH, Frondelius L, Berteselli GV, Gómez Y, Canali E, Niemi JK, et al. Measuring dairy cow welfare with real-time sensor-based data and farm records: A concept study. Animal. 2023;17(12):101023. doi:10.1016/j.animal.2023.101023
36. Bloch V, Frondelius L, Arcidiacono C, Mancino M, Pastell M. Development and analysis of a CNN- and transfer-learning-based classification model for automated dairy cow feeding behavior recognition from accelerometer data. Sensors (Basel). 2023;23(5):2611. doi:10.3390/s23052611