CowPain Check: AI-Based Facial Expression Analysis for Dairy Cow Welfare

Running title: AI-Based Facial Grimace Scales Review

Shivam Patel[1] & Suresh Neethirajan[1,2,*]

[1]Faculty of Computer Science, Dalhousie University, 6050 University Avenue, Halifax, NS B3H 4R2, Canada
[2]Faculty of Agriculture, Dalhousie University, Truro, NS B3H 4R2, Canada
*Corresponding author: Email: sneethir@gmail.com

Abstract
Pain management in dairy cattle remains a persistent challenge, hindered by subjective assessments and inherent observer biases that compromise animal welfare and impose significant economic burdens due to conditions such as mastitis and lameness. Emerging artificial intelligence (AI) technologies, integrated with computer vision and mobile platforms, now offer transformative solutions through objective, automated facial expression analysis. Advancements in neurobiological research have elucidated the mechanisms underlying bovine pain expression, enabling the development of robust grimace scales validated by high sensitivity and specificity (e.g., UCAPS, sensitivity/specificity: 0.78–0.85). Recent AI models employing advanced architectures such as YOLOv8-Pose (achieving 96.9% mAP in landmark detection) and transformer-based frameworks (demonstrating 98.36% accuracy in facial recognition tasks) significantly surpass conventional methodologies in accuracy, reliability, and scalability. Moreover, multimodal approaches fusing RGB and thermal imaging have demonstrated remarkable efficacy (81–95% accuracy) in capturing nuanced physiological indicators of pain. Edge-optimized deployment strategies now enable real-time, field-level applications, delivering rapid classifications at up to 24 frames per second with classification accuracies of 94.2%. Yet, substantial challenges persist, particularly in accounting for breed-specific variability and environmental interferences that limit universal applicability. Critical future research avenues include transfer learning for improved crossbreed adaptability, multimodal integration for chronic pain detection, and the advancement of longitudinal monitoring frameworks within precision livestock farming. The practical implications of these technologies are profound, promising significant welfare improvements through timely interventions, reduced economic losses, and the broader ethical advancement of AI-driven veterinary partnerships. The integration of automated facial expression-based pain detection in dairy operations thus holds immense potential to redefine standards in animal welfare and establish a new paradigm for sustainable and ethically aligned global dairy production.

Keywords: Artificial Intelligence in Dairy Farming; Automated Pain Detection; Facial Expression Analysis; Precision Livestock Farming; Grimace Scales; Computer Vision in Agriculture.

1. Introduction

The welfare of dairy cattle represents an urgent priority for producers, veterinarians, regulatory bodies, and consumers worldwide, driven by both ethical responsibilities and significant economic considerations. Pain-related health issues, notably mastitis and lameness, exact profound economic tolls and substantially compromise animal well-being. Traditional methodologies for assessing pain in dairy cattle - primarily behavioural observations, physiological indicators, and clinical scoring systems, are hindered by inherent limitations including observer bias, subjectivity, invasiveness, and insufficient temporal sensitivity. The advancement of artificial intelligence (AI), computer vision, and mobile technologies offers new avenues for precise, objective, and scalable pain monitoring, thereby significantly enhancing animal welfare and economic sustainability in dairy farming through precision livestock farming (PLF) initiatives.

1.1. Significance of Pain Management in Dairy Cattle Welfare

Effective pain management is increasingly recognized as a cornerstone of dairy cattle welfare, significantly influencing both animal well-being and production efficiency. Mastitis, one of the predominant diseases affecting dairy herds globally, imposes extensive economic consequences extending far beyond the direct expenses associated with treatment. Recent studies underscore that mastitis profoundly reduces the net present value (NPV) of dairy operations due to decreased milk yield, impaired reproductive capabilities, and increased culling rates [1]. Moreover, the negative economic ramifications of mastitis extend notably into reproductive outcomes, as cows afflicted with mastitis display substantially reduced conception rates compared to their healthy counterparts: notably lower first-service (41.7% vs. 58.2%), third-service (30.2% vs. 45.3%), and cumulative conception rates across multiple services (36.4% vs. 49.2%)[1].

Lameness is another significant contributor to pain-associated economic losses within dairy operations. Recent evidence positions lameness as the third most economically damaging health issue among dairy cattle, preceded only by mastitis and fertility disorders [2]. The economic impact of lameness manifests clearly through immediate and considerable reductions in milk production shortly after the onset of symptoms [3]. Beyond direct production losses, lameness triggers additional financial burdens from increased treatment costs, prolonged calving intervals, and the necessity of premature culling [2]. Collectively, the multidimensional nature of pain-related economic impacts underscores the necessity of developing effective, scalable, and precise methods for timely detection and intervention.

1.2. Limitations of Traditional Pain Assessment Methods

Current practices for pain assessment in dairy cattle rely predominantly on subjective behavioural observations, physiological measurements, and clinical scoring techniques. These methods exhibit significant methodological constraints that compromise their reliability and applicability. Systematic reviews have documented considerable subjectivity inherent in behavioural assessments, with variations in observer outcomes heavily influenced by factors such as evaluator experience, gender, age, and contextual nuances [4]. Meta-analyses have quantified this subjectivity, demonstrating substantial discrepancies in reported pain scores between studies employing diverse scales and terminologies [4].

Observer bias and inter-rater variability further weaken the reliability of conventional pain assessment tools. Recent evaluations employing the COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) guidelines highlight that only a small subset of behavioural-based instruments achieve consistently high validation scores across all essential measurement criteria [5]. These findings emphasize the critical challenges associated with obtaining dependable and uniform pain assessments, particularly within commercial dairy environments where evaluators face stringent time constraints and limited training opportunities.

Furthermore, traditional assessment approaches frequently fail to capture the dynamic and transient nature of pain expressions. Most conventional evaluations provide only episodic, snapshot observations, which may overlook brief but clinically meaningful expressions of pain. Research indicates that pain-related facial expressions in livestock often last between 0.3 and 0.7 seconds, rendering periodic manual observations insufficiently sensitive to identify early-stage or subclinical pain conditions [6]. This temporal limitation presents significant gaps, allowing undetected pain states to escalate unchecked between observation intervals.

Additionally, invasive traditional methods raise critical animal welfare concerns. Physiological indicators, such as blood sampling or rectal temperature measurement, may induce stress reactions in animals, inadvertently confounding pain assessment outcomes [7]. The necessary handling and restraint involved in invasive assessments can mask or artificially amplify expressions of pain, thus undermining both the accuracy and ethical justification of such procedures [7].

1.2. Emergence of Artificial Intelligence and Mobile Technology in Veterinary Medicine

The integration of artificial intelligence and mobile technologies into veterinary medicine has witnessed rapid acceleration since 2021, significantly advancing capabilities in automated animal welfare monitoring and pain detection systems. Recent innovations in computer vision techniques demonstrate impressive accuracy in cattle biometric identification and behaviour monitoring. For example, Vision Transformer (ViT) architectures applied to the Opencows2020 dataset achieved cattle identification accuracy rates as high as 99.79%, while YOLO-based frameworks combining YOLOv5 with ViT have attained detection precision (mean average precision, mAP) of 97.8% and identification accuracy of 96.3% in practical farm settings [8,9].

Precision livestock farming (PLF) represents a paradigm shift in dairy farm management, incorporating AI-driven sensors, computer vision, and big data analytics to monitor animal health, behavior, and welfare continuously. Contemporary PLF systems leverage diverse sensing modalities such as RFID tags, accelerometers, thermal imaging, and computer vision analytics to deliver comprehensive, real-time insights into cattle welfare. Research indicates that accelerometer-based systems effectively detect movement patterns indicative of lameness or stress-related behaviors, while multimodal sensor integration consistently outperforms single-modality systems in terms of detection accuracy and reliability [10].

The proliferation of mobile technology has facilitated widespread accessibility to advanced monitoring capabilities, empowering farmers to deploy PLF solutions effectively, even without extensive technical expertise or substantial infrastructural investments. Recent deployments of mobile applications for livestock welfare have demonstrated high usability and practical feasibility in commercial production scenarios, enabling farmers to swiftly interpret data and respond proactively to welfare alerts[11]. The seamless integration of mobile technology with advanced AI algorithms has successfully addressed temporal limitations inherent in traditional pain assessment methodologies, enabling timely interventions and enhancing animal welfare outcomes in dairy cattle.

Thus, the convergence of AI, computer vision, and mobile platforms holds remarkable promise to address current limitations in pain assessment, facilitating objective, scalable, and ethically responsible improvements in dairy cattle welfare and economic viability.

1.3 Methodology

This systematic review followed PRISMA 2020 guidelines to identify and synthesize peer-reviewed research on AI-based animal pain detection systems. A comprehensive search was conducted across three major databases (PubMed, Web of Science, Scopus) covering publications from January 2021 to July 2025.

Search Strategy: Boolean combinations of terms including ("artificial intelligence" OR "machine learning" OR "computer vision" OR "deep learning") AND ("pain detection" OR "facial expression" OR "grimace scale") AND ("cattle" OR "livestock" OR "cat" OR "dog" OR "horse" OR animal species terms) were employed with database-specific syntax optimization. A representative PubMed search string.

("artificial intelligence"[Title/Abstract] OR "machine learning"[Title/Abstract] OR "deep learning"[Title/Abstract] OR "computer vision"[Title/Abstract]) AND ("pain detection"[Title/Abstract] OR "facial expression"[Title/Abstract] OR "grimace scale"[Title/Abstract]) AND ("cattle"[Title/Abstract] OR "cow"[Title/Abstract] OR "livestock"[Title/Abstract] OR "sheep"[Title/Abstract] OR "horse"[Title/Abstract] OR "dog"[Title/Abstract] OR "cat"[Title/Abstract])

Eligibility Criteria: Studies were selected according to predefined inclusion and exclusion criteria, as outlined in Table 1.

Table 1. Inclusion and exclusion criteria for selecting publications for a systematic review

Domain	Inclusion Criteria	Exclusion Criteria
Publication type	Peer-reviewed journal articles	Preprints, non-peer-reviewed works, conference abstracts
Language	English	Non-English publications
Study type	Primary research reporting automated AI/ML approaches for animal pain detection	Reviews, opinion papers, or studies not involving automated methods
Indicators	Facial action units (FAUs), facial expressions, or facial-based indicators	Studies using only physiological or behavioral (non-facial) indicators
Performance reporting	Quantitative performance metrics (accuracy, sensitivity, specificity, AUC, etc.)	Studies lacking validation or performance reporting
Ground truth	Veterinary assessment or validated pain scales used as gold standard	Studies without validated ground truth

Study Selection: Search results were exported to Zotero reference manager. Duplicate records were removed. Screened titles and abstracts for relevance, followed by full-text assessment against eligibility criteria.

Data Synthesis: A total of 112 high-quality studies met inclusion criteria, encompassing multiple species (cattle, sheep, horses, cats, dogs, rabbits, rodents) and AI approaches (CNNs, Vision Transformers, YOLO architectures). Performance metrics were systematically extracted and synthesized both quantitatively and narratively to provide comprehensive coverage of current AI capabilities in automated animal pain detection. The selection process is documented in the PRISMA 2020 flow diagram (Figure 1), detailing the number of records identified, screened, excluded (with reasons), and included in the final synthesis.

Figure 1. PRISMA 2020 flow diagram for study selection in the systematic review on AI-based facial action unit analysis for pain detection in dairy cattle.

2. Pain Assessment in Dairy Cattle: Foundations and Limitations

Effective pain assessment in dairy cattle involves a complex interplay of animal welfare science, veterinary practice, agricultural economics, and ethical considerations. Historically viewed primarily as a welfare-focused issue, pain detection and management have evolved into multidimensional challenges that significantly impact the economic sustainability and social acceptability of modern dairy farming operations.

2.1 Importance of Pain Detection in Dairy Cattle Welfare and Economics

Detecting pain in dairy cattle is critical not only for animal welfare but also for its profound economic implications across the dairy industry. Unaddressed pain negatively affects animal behavior and physiological health, triggering stress responses that diminish productivity, growth rates, milk yield, and reproductive efficiency [12]. These physiological impacts such as hormonal stress responses, metabolic disruptions, and immune system suppression directly compromise animal health, thus reducing overall farm profitability [7].

Economic incentives further underscore the importance of effective pain mitigation. Zoltick et al. (2024) highlight that reducing pain through proactive analgesia enhances production efficiency sufficiently to offset the associated analgesic costs, thereby incentivizing farmers toward improved animal welfare management [7]. Additionally, stress responses triggered by pain negatively affect nutrient absorption, reproductive function, and general body condition, collectively translating into measurable economic losses at the herd level [7].

2.2 Impact on Productivity, Longevity, and Economic Losses

Economic losses attributed to pain-related conditions in dairy cattle are substantial, with mastitis, lameness, and ketosis identified as the primary economic burdens globally, costing the dairy sector approximately $65 billion annually [13]. These losses encompass direct treatment expenses and significant indirect impacts, including reduced productivity, reproductive failures, premature culling, and impaired herd longevity.

Mastitis-Associated Economic Impact

Globally, mastitis remains one of the most financially devastating dairy cattle diseases, incurring losses estimated between $20 and $30 billion annually [14]. Economic analyses show that clinical mastitis causes significant individual losses through decreased milk production, impaired fertility, and increased culling rates, with subclinical mastitis alone accounting for roughly 70-80% of the total mastitis-related economic burden [14].

The COVID-19 pandemic intensified these economic pressures. Research indicates dairy farms globally experienced increased mastitis-related losses due to disrupted veterinary care access, constrained market channels, and falling milk prices. These factors amplified the disease's financial burden, emphasizing systemic vulnerabilities within dairy supply chains and reinforcing the economic importance of early, accurate pain detection methods [14].

2.2.1. Lameness and Production Performance

Lameness is another prominent pain-related condition severely impacting dairy farm profitability. Recent longitudinal research demonstrated that lameness significantly reduces milk yield, with lame cows producing approximately 161-183 kg less milk per lactation compared to their healthy counterparts [15]. Lameness also prolongs calving-to-conception intervals, with affected cows experiencing significantly longer delays—approximately 38 additional days if lame before the first service and up to 87 days if lame afterward [15].

Moreover, the timing of lameness occurrences further amplifies its economic implications. Early lactation lameness typically triggers severe inflammatory responses, reducing feed intake, rumination times, and milk production efficiency. Such behavioral changes negatively affect energy balance and ovarian activity, thereby delaying postpartum reproductive cyclicity [15]. These cumulative productivity losses underscore lameness's profound economic consequences.

Lameness also indirectly exacerbates mastitis risks, creating additional economic complexity. Lame cows spend increased time lying on contaminated bedding, heightening bacterial exposure risks and subsequently raising mastitis incidence rates [16]. Thus, lameness indirectly contributes to economic losses through diminished milk quality and increased treatment expenses, reflecting interlinked disease management challenges.

2.3 Traditional Methods of Pain Assessment

Traditional pain assessment approaches predominantly rely on direct behavioral observations, physiological biomarkers, and structured clinical scoring systems. Despite recent methodological improvements, these approaches carry inherent limitations affecting their practical effectiveness.

2.3.1. Behavioral Indicators and Observational Methods

Behavioral observations remain a cornerstone of cattle pain assessment. Typical indicators include abdominal discomfort behaviors, altered locomotion, posture changes, and interaction disruptions. Recent advancements, such as accelerometer-based movement analyses, enhance behavioral assessment objectivity, capturing precise mobility pattern alterations associated with pain [7].

Tools like the Cow Pain Scale, validated in recent literature, systematically identify behavioral indicators—including reduced environmental interaction, altered posture, and decreased responsiveness—that effectively signal pain [17]. Despite validation, these tools heavily depend on observer training and experience, often leading to subjective variability [4].

2.3.2. Physiological Measures and Biomarker Assessment

Physiological biomarkers, notably cortisol, offer quantifiable pain detection metrics. Recent validation demonstrates plasma cortisol's diagnostic reliability, achieving receiver operating characteristic (ROC) curves (AUC >0.7) at specific post-pain stimulus intervals [18]. Additionally, hair cortisol provides robust chronic stress assessments by reflecting prolonged hypothalamic-pituitary-adrenal (HPA) axis activation, offering retrospective pain measures superior to acute assessments [18].

Infrared thermography (IRT) has gained traction as a non-invasive physiological pain indicator, demonstrating reliable diagnostic accuracy at specific post-intervention intervals (e.g., 72 hours, AUC >0.7). However, environmental factors, including ambient temperature and humidity, substantially impact IRT accuracy, requiring stringent calibration [19].

2.3.4. Advanced Physiological Monitoring Technologies

Pressure algometry quantifies mechanical nociceptive thresholds, effectively distinguishing pain states such as digital dermatitis in cattle. Recent studies confirmed its reliability, though practical constraints—including animal restraint requirements and specialized training—limit widespread implementation [20]. Integration of multiple physiological indicators, as recent research suggests, may enhance assessment accuracy, given that single biomarkers rarely offer definitive pain discrimination [18].

2.4. Facial Expressions and Grimace Scales: Bridging Traditional and Automated Methods

Facial expressions constitute one of the most fundamental, evolutionarily conserved communication mechanisms for pain across mammalian species. The development of standardized grimace scales has significantly enhanced objective pain assessment in veterinary medicine, overcoming traditional limitations related to observer subjectivity. This section systematically examines the neurobiological mechanisms underlying facial expressions of pain, the rigorous development and validation processes for grimace scales across domestic, laboratory, and farm animal species, and addresses ongoing challenges in their clinical applicability and reliability for livestock welfare management.

2.4.1 Neurobiological Basis of Pain Expression

Neural Pathways and Facial Action Unit Activation

Facial expressions of pain involve intricate interactions among nociceptive processing, emotional regulation, and motor control pathways. These systems collectively produce observable facial muscle responses indicative of pain states. Recent neuroscientific advancements have identified critical neural circuits translating pain perception into facial action units (FAUs), thus providing foundational scientific justification for grimace scale methodologies.

Current evidence underscores the amygdala’s pivotal role in generating pain-related facial expressions due to its extensive connections with sensory processing and motor control regions [21]. The central nucleus of the amygdala (CeA) serves as an integrative hub, receiving direct inputs from nociceptive regions such as the parabrachial nucleus, and projecting to brainstem motor centres that regulate facial musculature [22]. Optogenetic studies reveal that targeted CeA circuit activation elicits distinct pain-associated facial expressions, whereas inhibition reduces such responses, confirming functional links between pain perception and facial motor output [23].

The trigeminal nerve complex further supports pain-related facial expressions, facilitating both sensory detection and motor responses via the trigeminal motor nucleus, which governs critical muscles involved in grimacing behaviors [24].Thus, the amygdala-trigeminal circuitry is instrumental in generating specific facial pain behaviors.

Recent molecular-level insights highlight the contribution of non-neuronal elements, particularly astrocytes within the CeA, to facial expression regulation during chronic pain states. Elevated glial fibrillary acidic protein (GFAP) levels correspond with sustained facial pain behaviors, and selective inhibition of amygdala astrocytes reduces these expressions, indicating glial involvement in pain signalling and expression modulation [22].

Species-Specific Neural Control Mechanisms

Although foundational neural circuits for facial pain expressions remain evolutionarily conserved, species-specific variations in facial musculature and innervation patterns significantly impact observable expressions. In cattle, anatomical studies reveal unique facial muscle arrangements and nerve supply patterns distinct from human or rodent models, emphasizing the necessity for species-specific grimace scales [6].

2.4.2 Development and Validation of Grimace Scales Across Species

Evolution of Standardized Assessment Approaches

Grimace scale development has evolved significantly, transitioning from initial observational methodologies to rigorously validated, standardized instruments providing quantifiable pain metrics. Key developmental principles—identification of consistent FAUs correlating with pain, standardized scoring criteria for trained observers, and validation against established pain indicators—have maintained consistency across various species [25]. This systematic approach enhances scientific rigor and practical applicability across different animal groups.

Laboratory Animal Applications and Refinements

Grimace scales in laboratory animals, particularly rodents, have benefited from substantial refinement and validation. The Mouse Grimace Scale (MGS) now demonstrates optimized accuracy with fewer facial action units; notably, orbital tightening consistently exhibits strong predictive accuracy across pain models [26].

Advanced quantitative methods employing machine learning have further improved rodent grimace scale accuracy. Automated Rat Grimace Scale (RGS) scoring, leveraging advanced computational techniques, achieves precision and recall rates above 97%, closely matching human expert assessments (ICC of 0.82) [27]. Training protocols significantly enhance inter-rater reliability in rat grimace assessments, indicating sustained improvements over extended periods and emphasizing the durability of standardized training programs [28].

Feline Pain Assessment Advances

Recent advancements in feline pain assessment have demonstrated high reliability and practical applicability of the Feline Grimace Scale (FGS). Validation across diverse user groups—veterinarians, veterinary nurses, students, and caregivers—confirms robust inter-rater reliability, with intraclass correlations consistently between 0.65 and 0.69 [29]. Structured training substantially improves observer consistency, elevating reliability metrics to excellent levels (ICC 0.75–0.80) [30].

Furthermore, automated feline pain recognition using deep learning techniques has achieved promising accuracy (>70%), employing precise landmark-based analysis derived from feline facial action coding systems [31]. Nevertheless, continued validation remains critical to address variability across datasets and individual cat populations.

Equine Grimace Scale Development and Challenges

Equine grimace scales face distinct challenges, particularly related to the brief temporal dynamics of equine pain expressions, with approximately 75% of FAUs lasting only 0.3–0.7 seconds [32]. This underscores the importance of temporal resolution in equine pain assessments, favouring video-based analyses over static photographic methods.

Comparative reliability studies involving multiple equine pain scales—including HGS, EQUUS-FAP, EPS, and CPS—indicate varying inter-rater consistency, with the Composite Orthopedic Pain Scale displaying the highest reliability (ICC up to 0.75) [32]. Breed-specific differences in pain expression among horses—such as Friesians demonstrating reduced pain responsiveness compared to Quarter Horses—highlight the necessity for breed-sensitive grimace scales [33]. Recent investigations also suggest limited effectiveness of equine grimace scales for chronic pain states, such as gastric ulcers, reinforcing the importance of distinguishing scale utility across pain conditions [6].

Bovine Pain Assessment Developments

In bovine pain assessment, the Unesp-Botucatu Cattle Pain Scale (UCAPS) represents a landmark development, achieving robust validation and high reliability across diverse breeds [34,35]. Recent developments have expanded this approach to calves, creating the Calf Grimace Scale (CGS), which reliably identifies pain-associated FAUs following painful procedures like castration [36,37].

Advanced bovine validation methodologies incorporate comprehensive criteria—expression specificity, construct validity, responsiveness—to rigorously evaluate facial FAUs during painful conditions, notably clinical mastitis [38]. Real-time versus video-recorded assessment comparisons using UCAPS demonstrate high consistency (ICC ≥0.81), informing standardized clinical assessment protocols [39]. Fig illustrates the temporal dynamics of FAU activation across a 72-hour postoperative period in dairy cows (n = 45).

Fig 2. Temporal Dynamics of Facial Action Unit Activation with Error Bars and Statistical Significance During 72-Hour Postoperative Period in Dairy Cows (n=45)[34]

2.4.3 Reliability, Validity, and Limitations of Facial Scoring Systems

Inter-rater Reliability Achievements and Challenges

Inter-rater reliability remains critical for clinical grimace scale implementation, yet observer variability persists across species and contexts. Systematic analyses confirm significant improvements following structured training protocols; however, reliability gains vary considerably across species-specific contexts [30]. Table 1 summarizes the comparative validation metrics of contemporary grimace scales across species, including inter-rater and intra-rater reliability, sensitivity, and specificity as reported in recent studies. Feline scales consistently demonstrate high reliability, whereas equine grimace assessments vary notably with pain type and breed specificity [40]. Studies in macaques reinforce that while moderate-to-good reliability is achievable, extensive observer training and standardized protocols remain essential, especially for cognitively complex species [41].

Table 2: Comparative validation metrics of contemporary grimace scales across different species, highlighting inter-rater and intra-rater reliability, sensitivity, and specificity as reported in recent peer-reviewed studies.

Species	Scale	Sample Size	Inter-rater ICC	Intra-rater ICC	Sensitivity	Specificity	Reference
Feline	FGS	1,262 caregivers	0.65-0.69	>0.90	Not reported	Not reported	[29]
Feline	FGS (trained vets)	7 veterinarians	0.75-0.80	Not reported	Not reported	Not reported	[30]
Equine	HGS	8 horses	0.52	Not reported	Variable by condition	Variable by condition	[32]

Equine	HGS (dental disease)	12 horses	0.27	Not reported	Poor for chronic pain	Poor for chronic pain	[40]
Rat	RGS (automated)	Multiple cohorts	0.82 vs humans	Not applicable	81-93% weighted accuracy	81-93% weighted accuracy	[27]
Macaque	CMGS	43 animals	0.67 ± 0.28	0.79 ± 0.14	Not reported	Not reported	[41]
Donkey	DOPS	44 animals	0.56-0.66	0.88-0.96	80-98% at M1	90-97% at M0	[42]

Sensitivity and Specificity Performance

Diagnostic performance varies considerably among species-specific grimace scales, with sensitivity and specificity metrics heavily dependent upon pain type, duration, and assessment timing. Advanced ROC curve analyses confirm high diagnostic accuracy (AUC >0.70) in cattle when optimally timed post-intervention [18]. Notably, donkey scales exhibit particularly robust diagnostic accuracy (AUC = 0.91), providing clear analgesic intervention thresholds for clinical use [42]. Temporal dynamics significantly influence grimace scale sensitivity, particularly as acute pain transitions to chronic pain, requiring temporal optimization in clinical protocols to maintain assessment precision [43,44].

Methodological Limitations and Technological Solutions

Methodological limitations, notably static photographic assessments and subjective observer scoring, constrain grimace scale reliability and clinical utility [45,46]. Automated assessment systems utilizing machine learning and computer vision techniques demonstrate potential to significantly reduce observer variability, enhancing real-time monitoring and accuracy [27].

Multimodal assessment integration—combining facial analysis with physiological and behavioral data further improves detection precision, surpassing single-method approaches [47]. However, breed-specific anatomical and behavioral variations require continued validation and tailored scoring criteria across genetically diverse cattle populations [39].

Grimace scales represent critical advancements toward objective, species-specific pain assessment across diverse animal taxa. Achieving widespread clinical implementation necessitates ongoing refinement, comprehensive observer training, integration of advanced technological methodologies, and continual breed-specific validation efforts. These multidisciplinary approaches will ensure reliable, accurate pain measurement, significantly enhancing animal welfare management practices in veterinary medicine. Table 2 summarizes key factors influencing grimace-scale reliability and validity, detailing variables, their impacts on assessment performance, and proposed strategies for improving accuracy and consistency across species.

Table 3: Summary of key factors influencing the reliability and validity of grimace scales, highlighting specific variables, their impacts on assessment performance, and suggested strategies to enhance accuracy and consistency across species.

Factor Category	Specific Influences	Impact on Performance	Mitigation Strategies	Reference
Training Effects	Structured training programs	Moderate to good improvement in ICC	Standardized protocols, ongoing education	[27]
Species Differences	Anatomical variations, behavioral patterns	Requires species-specific validation	Species-appropriate scale development	[40,41]
Pain Type	Acute vs chronic, visceral vs somatic	Acute pain shows better detection	Condition-specific assessment tools	[32,48]
Temporal Factors	Duration of expression, assessment timing	Optimal windows for detection	Video analysis, temporal optimization	[32]
Observer Experience	Professional vs lay observers	Experience improves consistency	Training programs, standardization	[26]
Breed Variations	Genetic differences in expression	Requires breed-specific consideration	Diverse training datasets	[36]

2.5 Limitations and Challenges of Conventional Approaches

Despite methodological advancements, traditional pain assessment faces practical and conceptual constraints that impede widespread effectiveness.

2.5.1. Subjectivity and Observer Bias

Observer variability significantly undermines traditional pain assessment reliability. Recent systematic reviews and meta-analyses clearly demonstrate that observer training, personal biases, scale usage differences, and terminology variations significantly impact scoring consistency [7]. Even structured training protocols fail to completely eliminate observer bias, limiting assessment reliability.

2.5.2. Species-Specific and Environmental Challenges

Cattle's evolutionary inclination to mask pain, derived from predator-avoidance behaviors, severely complicates clinical assessments, leading to frequent underestimation of pain severity [7]. Environmental factors such as housing conditions, handling practices, and social interactions further obscure accurate pain detection, complicating the differentiation between general stress and specific pain behaviours [7]. Similarly, environmental conditions significantly influence physiological indicators such as thermography accuracy [19].

2.5.3. Physiological Indicator Constraints

Physiological biomarkers frequently demonstrate specificity limitations, failing to achieve consistently high diagnostic accuracy across varied pain states and individual animal variability

(AUC often <0.7) [18]. Chronic pain conditions further complicate biomarker assessments, with adaptive physiological responses reducing biomarker reliability [18].

2.5.4. Practical Implementation Barriers

Operational challenges significantly limit traditional assessment feasibility. Comprehensive assessments require intensive labour, substantial training, and expensive specialized equipment, restricting their scalability across large commercial herds [12]. Invasive assessment methods, such as blood sampling, further introduce ethical and practical dilemmas by inducing additional stress and potentially confounding pain assessments [7]. Table 3 presents a comparative evaluation of traditional pain assessment methods, outlining their primary strengths, methodological limitations, and key references.

Table 4. Comparative evaluation of traditional pain assessment methods used in dairy cattle, highlighting assessment types, primary strengths, methodological limitations, and representative references from recent peer-reviewed literature.

Method	**Type**	**Strengths**	**Limitations**	**Reference Example**
Behavioral Observation	Visual/Manual	Widely accessible; non-invasive; captures species-specific behaviours	Subjective; observer bias; time-consuming; low throughput	[43]
Physiological Biomarkers (Cortisol)	Biochemical	Objective; quantifiable; hair cortisol offers chronic-stress measure	Requires sampling; invasive (blood); temporal variability; lab analysis	[18]
Pressure Algometry	Mechanical Nociceptive Threshold	Quantifies mechanical sensitivity; reliable thresholds	Requires restraint; operator-dependent; localized assessment	[20]
Infrared Thermography	Thermal Imaging	Non-invasive; detects physiological heat changes; real-time	Affected by environment (temperature, humidity); calibration needed	[19]
Facial Expression/Grimace Scales	Visual Scoring	Rapid; non-invasive; sensitive to acute pain	Requires training; semi-subjective; limited to acute responses	[37]

Collectively, these critical limitations emphasize the urgent need for accurate, minimally invasive, objective pain assessment solutions capable of continuous monitoring without extensive human intervention. The integration of AI, computer vision, and mobile technologies offers promising pathways toward overcoming traditional assessment challenges, providing practical, scalable, and ethically responsible alternatives for modern dairy cattle pain management.

3. AI and Computer Vision Foundations for Animal Pain Detection

The integration of artificial intelligence with computer vision represents a paradigm shift from subjective human observation to objective, automated pain assessment in livestock. This section examines the foundational AI architectures that have been successfully applied to animal pain detection, with particular emphasis on recent advances from 2021-2025 that demonstrate measurable improvements in accuracy and practical deployment capabilities.

3.1 Convolutional Neural Networks: Architectural Evolution and Performance

Convolutional Neural Networks remain the cornerstone of automated animal pain detection systems, with recent studies demonstrating substantial improvements through architectural refinements and species-specific optimizations. The foundational strength of CNNs lies in their hierarchical feature extraction capabilities, enabling the identification of subtle facial patterns associated with pain expressions across multiple livestock species [49,50].

ResNet Architectures and Transfer Learning

ResNet-based models have shown remarkable versatility in cross-species applications. A comprehensive study on rabbit pain detection achieved 87% accuracy using ResNet-50 architectures combined with novel temporal processing techniques [50]. The study employed Grayscale Short-Term stacking (GrayST) methodology, which incorporates temporal information by combining consecutive frames into single composite images, effectively capturing the dynamic nature of pain expressions that static analysis often misses [50].

For cattle facial landmark detection, ResNet-101 demonstrated superior performance on RGB imagery, achieving 94.37% average precision (AP) on the CattleFace-RGBT benchmark dataset [51]. However, performance degraded significantly when applied to thermal imagery (64.60% AP), highlighting the modality-specific challenges that plague cross-spectral applications [51]. This performance disparity underscores the need for specialized training approaches when working with multimodal data.

More sophisticated CNN variants have emerged to address livestock-specific challenges. The IWOA-CNN model, incorporating an improved whale optimization algorithm, has shown superior performance compared to traditional CNN approaches by optimizing critical hyperparameters including dropout probability, L2 regularization parameters, and dynamic learning rates [52]. This algorithmic enhancement addresses the fundamental issue of manual hyperparameter tuning, which often results in suboptimal performance for animal-specific applications.

Recent studies in facial recognition for livestock have further demonstrated the viability of CNNs in real-world farm settings. YOLOv5 for cow face detection combined with a Vision Transformer for identification in a 77-cow herd, achieving 97.8% detection AP and 96.3% ID accuracy [53]. Similarly, CFR-YOLO based on YOLOv7, which achieved 96.27% mean average precision and 98.46% precision [54]. These models processed video at real-time speeds (~50 fps), validating their feasibility for continuous on-farm monitoring. Additionally, combined YOLOv4-tiny and MobileNetV2 on edge devices for cow recognition, reached a detection F1 of 0.98 and ID accuracy of 0.97 under practical farm conditions [55].

3.2 Vision Transformers: Global Context and Attention Mechanisms

The introduction of Vision Transformers (ViTs) has fundamentally challenged CNN dominance in animal facial analysis. ViTs excel at capturing long-range dependencies and global contextual information, characteristics particularly valuable for understanding complex facial expression patterns in livestock [56]. The ViT-Sheep model, incorporating LayerScale modules and transfer learning strategies, achieved 97.9% accuracy for sheep face recognition, demonstrating the architecture's potential for livestock applications [56].

CLIP-Based Pain Detection

A groundbreaking study in sheep pain recognition demonstrated that CLIP (Contrastive Language-Image Pre-training) encoders significantly outperformed human expert assessment [57] The AI pipeline achieved an AUC of 0.82 for binary pain classification, significantly exceeding human facial scoring performance (AUC difference = 0.115, $p < 0.001$) when provided with identical visual information (frontal and lateral face images) [57]. The system utilized 768-dimensional CLIP embeddings concatenated from both viewing angles, processed through Naive Bayes classifiers with leave-one-animal-out cross-validation [57].

Swin Transformers for Multimodal Processing

Swin Transformers represent a particularly promising advancement, combining the global attention mechanisms of transformers with CNN-like hierarchical processing. In pig recognition and segmentation tasks, Swin Transformers achieved 93.0% recognition accuracy and 86.9% segmentation accuracy, maintaining excellent performance even under challenging conditions including overlapping, occlusion, and deformation [58]. These results suggest that transformer architectures may be particularly well-suited for handling the complex environmental conditions typical of farm settings.

3.3 YOLO Architectures: Real-Time Detection and Multi-Object Tracking

You Only Look Once (YOLO) frameworks have become indispensable for real-time livestock monitoring applications, offering optimal balance between detection speed and accuracy essential for practical farm deployment [59].

YOLOv8 Advancements

Recent implementations of YOLOv8 have demonstrated exceptional performance in livestock applications. A modified YOLOv8-CBAM system for cattle detection achieved 95.2% precision and 82.6% mAP@0.5:0.95, representing a 2.3% improvement over baseline YOLOv8 across diverse camera configurations [60]. The integration of Convolutional Block Attention Modules (CBAM) enhanced the model's ability to focus on relevant facial features while suppressing background noise [60].

For sheep head recognition, YOLOv8-CBAM achieved 97.7% mean average precision with an F1 score of 0.94, demonstrating consistent improvements over multiple YOLO variants: 0.5% over YOLOv8n, 1.4% over YOLOv5n, and 2.4% over YOLOv10n [61]. The attention mechanism proved particularly effective for recognizing facial color patterns essential for breed identification and individual recognition [61].

CFR-YOLO for Cattle Face Recognition

A specialized cattle face recognition system based on YOLOv7 improvements (CFR-YOLO) achieved remarkable performance metrics of 96.27% mean average precision while maintaining real-time processing capabilities at approximately 50 fps [62]. The system incorporated several key optimizations: replacement of CIoU loss with SIoU loss functions, integration of FReLU activation functions, and inclusion of Receptive Field Block (RFB) modules in the backbone network [62].

3.4 Multimodal Fusion: RGB-Thermal Integration

The combination of RGB and thermal imaging represents a significant advancement in automated pain detection, providing complementary information streams that enhance overall system robustness and accuracy[51].

CattleFace-RGBT Benchmark Dataset

The development of the CattleFace-RGBT dataset, consisting of 2,300 RGB-thermal image pairs with 13 annotated facial landmarks, has established a critical benchmark for multimodal livestock analysis [51]. The dataset covers key facial regions including ears, eyes, muzzle, nostrils, and mouth, enabling comprehensive welfare assessment through both visual and thermal indicators [51].

Performance analysis reveals significant modality-specific differences: while RGB processing achieves superior accuracy (ResNet-101: 94.37% AP), thermal processing remains challenging (ResNet-101: 64.60% AP). However, transformer architectures show better thermal performance, with Swin-B achieving 73.16% AP on thermal imagery [51].

Fusion Strategies and Implementation

Three primary fusion approaches have been evaluated: early fusion (feature-level integration), late fusion (decision-level combination), and mixture of experts (dynamic weighting) [63]. Early fusion enables cross-modal learning during feature extraction but requires careful calibration between modalities. Late fusion processes modalities independently before higher-level integration, providing greater flexibility for handling modality-specific preprocessing requirements[64].

The thermal imaging component provides unique physiological information invisible to RGB cameras, particularly useful for detecting inflammation and temperature variations associated with pain states. However, environmental factors including ambient temperature, humidity, and airflow significantly impact thermal measurement reliability, necessitating sophisticated calibration protocols.

3.5 Technical Implementation Challenges and Solutions

Edge Computing and Deployment Constraints

Real-world deployment faces substantial computational constraints, particularly in rural environments with limited connectivity and power availability. Successful edge implementations using Nvidia Jetson Nano devices have demonstrated feasibility, maintaining high performance (96.1% accuracy) while operating within 10W power envelopes [49]. Model compression techniques, including quantization-aware training and pruning, have achieved up to 86% reduction in model size while preserving accuracy above 95% [65].

Cross-Species Generalization

Recent research has demonstrated both the potential and limitations of cross-species model transfer. A CNN trained for pig pneumonia detection achieved substantial agreement (Cohen's kappa: 0.65-0.71) when applied to lamb lung assessment, with sensitivity (0.87-0.88) and specificity (0.88-0.91) comparable to expert veterinary assessment [66]. However, facial expression models show greater species-specificity, with accuracy drops of 15-20% when applied across species without fine-tuning [67].

Scalability and Farm Integration

Commercial operations involving thousands of animals introduce scalability challenges beyond typical applications. Multi-camera systems, sophisticated tracking algorithms, and data fusion techniques offer potential solutions, though they increase calibration complexity [68]. Effective farm integration necessitates alignment with existing management systems, including user-friendly interfaces, real-time alerts, decision-support tools, and mobile application integration, addressing computational limitations inherent to smartphone hardware [69].

The development of appropriate sensitivity thresholds and human-centered design considerations remains essential to avoid alert fatigue, maintaining user trust, and ensuring widespread adoption of advanced AI-based livestock pain detection systems in real-world agricultural settings.

3.6 Practical comparison of AI architectures for farm implementation

A critical question for adoption is not which architecture attains the highest benchmark score in controlled experiments, but which architecture reliably performs under real farm constraints (variable lighting, occlusion, dirt, overlapping animals), runs on available hardware (edge devices, low-power systems), and generalizes across herds and barns. Below we compare Convolutional Neural Networks (CNNs), YOLO-family detectors, Vision Transformers (ViTs) and multimodal fusion approaches against practical implementation criteria supported by recent peer-reviewed farm or near-farm studies.

3.6.1 Detection & classification performance in farm/field tests

- YOLO-family detectors (e.g., YOLOv5–v8 variants) show high detection performance in real or semi-real farm deployments while maintaining high frame rates suitable for continuous monitoring. Recent farm-targeted studies report mean average precision (mAP) in the mid-90s for cattle detection/landmark tasks and sustained inference speeds (20–50 fps) on embedded hardware after optimization (quantization/TensorRT). These deployments achieved realistic classification accuracies in the 90–95% range for biometrics and health-related labels in independent test sets[70].

- CNN backbones (ResNet, MobileNet, EfficientNet) remain highly effective for landmarking and facial feature extraction in field conditions. Lightweight CNN variants (MobileNet, pruned/quantized ResNets) have been successfully deployed on Jetson-class devices with accuracy often exceeding 90% for face detection/landmark tasks while keeping power consumption <10 W, making them practical for continuous barn operation[71].

- Vision Transformers (ViT / Swin) demonstrate excellent representational power and sometimes outperform CNNs on large, curated datasets, but peer-reviewed farm implementations report limited on-device feasibility due to higher compute and data requirements; where deployed, hybrids (CNN encoder + transformer blocks) have shown improved accuracy while reducing latency compared with pure ViTs. Field-oriented transformer work for livestock remains emerging but promising[72].

- Multimodal fusion (RGB + thermal / sensors) increases robustness to lighting and can improve physiological detection (inflammation/fever), but thermal performance and fusion require careful calibration in farm environments and entail higher system complexity and cost. Cattle RGB-thermal benchmark studies show strong RGB AP but substantially lower thermal AP unless advanced transformer fusion or calibration is used[73].

3.6.2 Robustness to farm conditions (lighting, occlusion, dirt, overlap)

- YOLO and modern CNN detectors tolerate moderate occlusion and variable lighting when trained with augmentations and multi-site data, but performance degrades when animals overlap densely or when reflective surfaces and dust produce spurious detections—practical fixes include optimized camera placement and exposure control. Farm deployment reports recommend per-camera tuning and occasional re-calibration[73].

- Transformer models benefit from global attention and can be more robust to certain contextual variations if trained on very diverse datasets; however, in most peer-reviewed farm trials such large, diverse pretraining corpora are not yet available, limiting ViT robustness in practice[72].

3.6.3 Edge feasibility, latency and power constraints

- Practical farm systems prioritize on-device inference to avoid latency and connectivity dependence. Studies like *Dairy DigiD* demonstrate that lightweight YOLO/CNN stacks, combined with INT8 quantization and TensorRT, can achieve ~24 fps on Jetson NX/Nano devices while preserving high classification accuracy (~94%), making them feasible for continuous on-farm operation. Such optimizations (pruning, quantization) are essential to make modern architectures practical on farms [70].

- Pure ViT pipelines currently require cloud or high-end accelerators for real-time operation; thus, unless offloading or hybrid architectures are used, ViTs are less feasible for always-on edge monitoring at present [72].

3.6.4 Recommendations for practitioners (evidence-based)

1. For continuous, real-time monitoring on typical dairies: deploy optimized YOLOv8 / YOLOv7 or compressed CNN backbones (MobileNet/ pruned ResNet) with INT8 quantization; these achieve the best trade-off of accuracy, fps and edge power envelope in peer-reviewed deployments[70].
2. For research or centralized analytics with ample compute and large datasets: explore Transformer / hybrid models to leverage their superior context modeling for cross-farm generalization—provided extensive pretraining or multi-farm data are available[72].
3. For low-light or physiological signs (inflammation): consider RGB+thermal fusion, but include temperature/humidity calibration protocols and expect higher annotation and hardware costs[73].
4. Always validate with LOAO and farm-fold tests and report per-fold sensitivity/specificity and confidence intervals; real farm readiness requires inter-farm robustness, not just within-dataset accuracy[74].

4. Current AI Applications in Livestock Pain Recognition

The application of artificial intelligence for automated pain detection has expanded significantly across multiple animal species since 2021, with validated systems demonstrating clinical feasibility for both livestock and companion animals. Fig 2 compares mean accuracy of AI-based pain detection systems across laboratory, livestock, and companion species, highlighting key performance differences among these groups .

4.1 Feline Pain Detection Systems

Automated pain recognition in cats has achieved remarkable progress through multiple complementary approaches. The landmark-based methodology achieved 77% accuracy in pain detection using manually annotated geometric landmarks positioned relative to underlying facial musculature, significantly outperforming deep learning approaches that reached only 65% accuracy on the same heterogeneous dataset [75]. This study utilized 84 client-owned cats of different breeds, ages, sexes, and varying medical conditions, representing a substantial advancement over previous homogeneous datasets limited to single breeds.

Video-based automation marked a significant technological leap with the development of end-to-end AI pipelines requiring no manual image selection or landmark annotation [76]. The system achieved over 70% and 66% accuracy respectively on two different cat pain datasets, outperforming previous landmark-based approaches using single frames under similar conditions. The pipeline integrated YOLOv8 for face detection, ensemble landmark detection, and XGBoost classification with moving window analysis.

Smartphone-applicable systems represent the current clinical frontier, utilizing deep neural networks and machine learning models trained on 3,447 cat face images annotated with 37 landmarks [77]. The best CNN model (ShuffleNetV2) achieved 16.76% Normalized Root Mean Squared Error for landmark prediction, while XGBoost models reached 95.5% accuracy and 0.0096 mean squared error for Feline Grimace Scale score prediction. The system demonstrated excellent discriminatory capability between painful and non-painful cats, enabling practical veterinary applications.

4.2 Non-Human Primate Pain Recognition

Macaque facial expression analysis achieved groundbreaking automation through the first prototype for automatic MaqFACS (Macaque Facial Action Coding System) coding [78]. The system achieved high performance in recognition of six dominant action units, demonstrating generalization between conspecific individuals (Macaca mulatta) and even between species (Macaca fascicularis). The method showed concurrent validity with manual MaqFACS coding, supporting automated applications in social and affective neuroscience research.

Japanese macaque pain detection utilizing ResNet50 architectures achieved varying accuracy depending on extraction methodology [79]. Box extraction using RetinaFace resulted in test accuracies between 48-54%, while contour extraction using Mask R-CNN improved performance to 64% through preprocessing and fine-tuning. The study utilized 30-60 minutes of video footage from macaques undergoing laparotomy, recorded before surgery (No Pain) and one day post-surgery before analgesic administration (Pain).

Geometric morphometric approaches complemented automated systems by revealing subtle facial shape variations in female Japanese macaques following experimental laparotomy [80]. The study identified pain-associated changes including orbital tightening, asymmetrical eye aperture, lip tension, and elongated mouth lines, providing anatomical foundation for automated detection algorithms.

4.3 Rodent Pain Assessment Systems

Mouse Grimace Scale Automation

Automated mouse grimace scale assessment achieved impressive performance through Vision Transformer architectures trained on manually scored datasets [81]. The system achieved 97% weighted accuracy for binary pain classification, with attention heatmaps revealing model focus on eye and ear regions as primary pain indicators. Individual action unit classifiers demonstrated weighted accuracies of 81-93% for orbital tightening, nose bulge, cheek bulge, ear position, and whisker changes[81].

4.4 Canine Emotional State Recognition

Dog emotional state recognition achieved significant progress through dual-approach methodologies comparing DogFACS-based and deep learning systems [82]. The DogFACS-based approach utilizing Decision Tree classifiers reached 71% accuracy, while deep learning techniques achieved 89% accuracy for positive/negative emotional state classification. The study analyzed 29 Labrador Retrievers under experimentally induced emotional states of positive anticipation and frustration.

Continuous facial dynamics analysis introduced novel automated methods for measuring dog facial behavior through video-based tracking of 46 facial landmarks [83]. The system revealed distinct patterns between brachycephalic (Boston Terrier) and normocephalic (Jack Russell Terrier) dogs, with brachycephalic dogs exhibiting consistently lower facial dynamics across all tested contexts and facial regions compared to normocephalic dogs.

Table 5. Performance overview of AI-based pain detection systems across animal species (2021-2025).Values are specific to individual studies and not statistically comparable because of heterogeneous datasets, imaging conditions, and validation protocols. Performance patterns reflect methodological differences in dataset design, validation rigor, and species-specific facial expressivity.

Species	Primary Reference(s)	Model / Methodology	Dataset Characteristics	Reported Performance	Validation Strategy / Methodological Notes
Cat	Feighelstein et al. 2023 [75]; Martvel et al. 2024 [76]; Steagall et al. 2023 [77]	Landmark-based CNN; YOLOv8 + XGBoost; ShuffleNetV2 + Feline Grimace Scale	84 client-owned cats (heterogeneous breeds, ages, health); 3 447 annotated face images	65 – 95 % accuracy range depending on architecture	Heterogeneous validation (train/test split or k-fold); some studies lacked independent test sets; lighting and breed variability affect generalization
Dog	Boneh-Shitrit et al. 2022 [82]; Martvel et al. 2025 [83]	DogFACS + Decision Tree; Deep CNN; Video-based landmark tracking	29 Labradors and multi-breed cohorts (brachycephalic vs. normocephalic)	71 – 89 % accuracy	Leave-one-video-out or within-subject cross-validation; performance influenced by breed morphology and reduced facial mobility

					in brachycephalic dogs
Sheep	Feighelstein et al. 2023 (CLIP encoders)	CLIP encoder + Naïve Bayes classifier	Controlled post-surgical dataset, frontal + lateral views	AUC = 0.82 (≈ 82 % accuracy)	Leave-one-animal-out validation minimized identity bias; consistent lighting and scoring; model outperformed human experts
Macaque (Primate)	Morozov et al. 2021 [78]; Gris et al. 2024 [79];	ResNet50; Mask R-CNN; Automatic MaqFACS coding	30 – 60 min per subject (pre- and post-surgery); 6 action units annotated	48 – 64 % accuracy	Cross-session validation; limited sample size; subtle facial muscle differences across species reduce transferability
Rodent (Mouse/Rat)	Arnold et al. 2023 [81];	Vision Transformer; Automated Grimace Scale	Controlled laboratory imagery with manual grimace labels	89 – 97 % weighted accuracy	Randomized cross-validation; standardized grimace scoring ensured high inter-rater consistency; results robust under uniform lighting

4.5 Comprehensive Species Validation

Cross-Species Performance Metrics

Current automated pain detection systems demonstrate species-specific performance variations, with accuracy ranges reflecting both methodological approaches and validation rigor. Sheep pain recognition using CLIP encoders achieved the highest reported accuracy (>82%), significantly outperforming human expert assessment 14. Cat pain detection systems showed moderate performance (65-77%) depending on approach methodology[75,84]. Primate systems achieved variable results (48-64%) reflecting the complexity of facial morphology and expression subtlety.

Rodent systems demonstrated strong performance, with mouse grimace scale automation reaching 89-97% accuracy[81].Dog emotional recognition achieved 71-89% accuracy depending on methodological approach [82].

4.6 Dairy Cows:

Recent research has applied computer vision and machine learning to detect pain in dairy cows under various conditions (e.g. lameness, mastitis). These studies use facial and gait indicators (e.g. **orbital tightening**, **ear position**, **back curvature**) extracted from images or video, often combined with sensor data, to train AI models. The Table 4 summarizes post-2021 peer-reviewed studies, detailing pain condition, facial action units (FAUs) or behavioral indicators, sensing methods, AI models, validation design, sample size, and key performance metrics (separating object-detection from pain-classification). All metrics are cited from the primary sources.

Table 6. Recent AI-Based Approaches for Pain Detection and Classification in Dairy Cows (Post-2021 Studies)

Study (Year)	Pain Type / Condition	FAUs or Indicators	Imaging/Sensing	AI Models	Validation & Sample	Performance (Detection vs Classification)
Zhang *et al.* (2025) [85]	Mixed health issues (lameness, metritis, mastitis, pre-birth labor)	Facial regions: *eyes*, *ears*, *muzzle* (key landmarks)	Video (RGB farm footage); frames processed at 1/5 s intervals	YOLOv8-Pose (face+30 facial landmarks), MobileNetV2 (ROI feature extractor), LSTM (temporal classifier)	10 videos (6 pain, 4 no-pain) with 80:20 train/val split; tested on 14 held-out videos.	**Detection:** YOLOv8-Pose achieves bounding box AP@0.5=0.969 (mAP), AP@0.5–0.95=0.899; keypoint AP@0.5=0.838, AP@0.5–0.95=0.590. **Classification:** Validation accuracy ≈99.65% (precision/recall ≈0.9968); unseen-video (video-level) accuracy 64.3%, pain-class precision 0.83, recall 0.56, F1=0.67.

Neupane *et al.* (2024) [86]	Lameness (hoof/leg disorders)	Locomotion features (lying time, steps, changes)	Leg-mounted accelerometer data	Time-series ML models (Random Forest, Naïve Bayes, Logistic Regression, ROCKET)	310 multiparous cows monitored 4 months (daily accelerometer); labeled by claw-trimmer as: healthy, corrective trimming, or lame (therapeutic trimming).	**Classification:** ROCKET classifier (best) for distinguishing healthy vs severely lame cows achieved accuracy >90%, ROC-AUC >0.74, F1 >0.61. For classifying severe vs moderate lameness, ROCKET gave accuracy >85%, ROC-AUC >0.68, F1 >0.44. (No vision-based detection metrics.)
Jia *et al.* (2025)[87]	Lameness (all grades 0–3)	Postural/gait: *arched back*, head bobbing, leg swing, asymmetric gait	Video (milking parlor, 25 fps); head and back keypoints annotated	DeepLabCut pose estimation (DLC pretrained on cow features); spatiotemporal keypoint scoring model	143 videos (dairy cows walking, various lameness levels) split into train/test (20 for testing); also 16 videos from other farms.	**Keypoint Detection:** Mean error ≈4.68 px (90.21% of keypoints correctly tracked). **Classification:** Overall lameness classification accuracy ≈90.2%; by class: 89.0% (normal), 85.3% (mild), 92.6% (moderate),

						100% (severe).
Russello *et al.* (2024) [88]	Lameness (visual gait scoring: healthy vs lame)	Locomotion traits: back posture curvature, head bobbing, tracking distance, stride length, stance/swing durations	Video (side-view walking lane, outdoor)	T-LEAP pose estimator (9 keypoints) + ML classifier on extracted gait features	Cows walking video, scored by 4 observers (5-point scale merged to binary healthy/lame); keypoint model evaluated on diverse lighting.	**Keypoint Detection:** 99.6% of cow keypoints correctly detected. **Classification:** Combining the top 6 locomotion traits yielded 80.1% accuracy (versus 76.6–79.9% using fewer traits) for healthy vs lame detection (binary classification accuracy; no separate AUC reported).

Critical Analysis of Performance Gaps and Generalization Challenges

The performance metrics presented in Table 5 reveal substantial discrepancies between validation accuracies and real-world performance that warrant critical examination. These disparities highlight fundamental challenges in the current state of AI-based cattle pain detection systems and underscore the necessity for more rigorous validation methodologies.
One of the example of these challenges appears in Zhang et al. (2025), where the reported validation accuracy of 99.65% contrasts sharply with the 64.3% accuracy achieved on unseen videos. This 35.35 percentage point performance degradation exemplifies severe overfitting, indicating that the model memorized training-specific patterns rather than learning generalizable pain-related features. The limited training dataset of only 10 videos (6 pain, 4 no-pain) with an 80:20 train/validation split exacerbated this problem by providing insufficient variability for robust feature learning. Such dramatic performance disparities fundamentally undermine the clinical utility of these systems, as the impressive validation metrics provide misleading indications of real-world effectiveness.

This pattern of generalization failure extends beyond Zhang et al., revealing systematic challenges across multiple studies in the literature. Jia et al. (2025) demonstrated similar limitations when their model, achieving 90.2% overall accuracy, experienced performance degradation when tested on videos from different farms, suggesting environment-specific overfitting. The authors' use of only 16 videos from other farms for external validation further highlights the inadequacy of cross-farm validation protocols. Similarly, Neupane et al. (2024) achieved accuracies exceeding 90% using the ROCKET classifier, but these results were obtained exclusively within single-farm validation scenarios using 310 cows from a homogeneous population, raising substantial concerns about cross-farm generalizability and breed-specific applicability.

The methodological approach employed by Russello et al. (2024) illustrates additional concerning patterns in the field. Despite achieving 99.6% keypoint detection accuracy, the subsequent classification performance dropped to 80.1%, indicating substantial information loss during the transition from detection to classification. This 19.5 percentage point gap suggests that high-quality landmark detection does not necessarily translate to effective pain classification, highlighting the complexity of extracting clinically meaningful pain-related features from detected anatomical landmarks.

These performance disparities stem from fundamental methodological limitations prevalent throughout the literature. Sample sizes remain inadequate for robust statistical validation, with most studies employing fewer than 200 animals across all validation phases. Training datasets typically originate from homogeneous environments, lacking the environmental diversity, breed variation, and temporal coverage necessary for meaningful generalization. Cross-validation methodologies frequently employ inappropriate random splits rather than more rigorous approaches such as Leave-One-Animal-Out (LOAO) validation or farm-fold cross-validation that would better assess model generalizability. Additionally, temporal dependencies within animal behavior data are systematically ignored, leading to optimistically biased performance estimates that fail to reflect real-world deployment scenarios.

4.7 Discussion of Factors Influencing Model Performance

Across the reviewed studies, key factors consistently drove differences in reported performance. First, model architecture and design strongly affected outcomes. Convolutional networks often required careful tuning of hyperparameters to avoid overfitting on small datasets. For example, Mao and Liu’s dog-expression study trained a CNN on only 315 images and found that tuning via an improved Whale Optimization algorithm boosted accuracy modestly by ~3 percentage points[52]. This suggests that generic CNN architectures alone may plateau on limited animal-expression datasets. By contrast, transformer-based models and large pre-trained encoders tended to generalize better when data were scarce. Like ViT-based sheep face model (ViT-Sheep) achieved 97.9% accuracy by incorporating architectural enhancements (LayerScale) and transfer learning on 160 sheep images [56]. Similarly, Feighelstein *et al.* used a CLIP encoder (a large-scale vision transformer) to detect pain in sheep, and the AI pipeline significantly *outperformed* expert scoring (AUC 0.82 vs. AUC 0.70 for humans) on the same 48-animal dataset [57]. The benefit of pretraining is clear: models with broad prior knowledge (ViT, CLIP) captured subtle facial cues that smaller CNNs missed.

Second, data quantity and quality were fundamental. Larger, well-annotated datasets yielded higher accuracy. For instance,cow-ID system had high sample diversity (77 cows, numerous face images) and achieved ~98% AP for detection and 96% identity accuracy, likely reflecting the ample data and robust YOLOv5+ViT pipeline used[53]. In contrast, studies with very small animal datasets (e.g. 10–30 individuals) often reported only modest performance (<70–80%). cat-pain pipeline attained only 70% and 66% accuracy on two feline datasets[84], even though they used a video-based approach, because these datasets remained small and heterogeneous. Subjective annotation also added noise: studies relying on human-rated pain scores (grimace scales) were inherently limited by rater inconsistency. The sheep-CLIP study mitigated this by using human scores as “gold standard” for comparison, but AI still outperformed the inconsistent human labels[57].

Third, image modality and preprocessing played a major role. Models trained on RGB imagery nearly always outperformed those on thermal images. In the new CattleFace-RGBT benchmark, ResNet-101 achieved 94.4% AP on RGB face detection but only 64.6% AP on thermal images[51]. Thermal data lack color/texture and suffer from low contrast, making keypoints harder to localize[51]. Transformer architectures fared slightly better on thermal: e.g. Swin-B scored 73.2% AP on thermal (vs. 75.3% on RGB)[51], suggesting that global attention can partly compensate for poor thermal detail. Some authors therefore use *cross-modal transfer*: Coffman *et al.* trained models on RGB and refined them on thermal, using semi-automated annotation to build the thermal landmark set. In video-based systems, temporal preprocessing also helped. Feighelstein *et al.* introduced “Grayscale Short-Term Stacking” (GrayST) to inject motion cues into static CNNs, boosting rabbit-pain recognition from ~67% (ResNet alone) to ~77–81% (with GrayST)[50]. Further filtering of video frames (keeping only high-confidence images) lifted rabbit-pain accuracy above 87%[50]. These examples show that explicit temporal encoding can overcome the lack of color or texture in single frames, at the cost of some complexity.

Fourth, species-specific traits and experimental conditions influenced outcomes. Some species exhibit very subtle facial changes, or wide breed variation, which makes generalization difficult. For example, dog facial morphology varies enormously by breed, so Mao *et al.* note that even with IWOA-CNN their hardest classes (sad, fear) remained under 90% accuracy. Similarly[52], the cat-pain studies point out that facial landmarks in cats are subtle and vary by individual, so performance capped around 70% despite advanced pipelines[84]. By contrast, simpler tasks with distinctive cues yielded higher scores: automated detection of specific behaviors (e.g. cow hoof issues via accelerometers) or identity recognition (hundreds of cow faces) tended to exceed 90% accuracy, showing that modality and task simplicity matter. In experiments with induced pain (e.g. sheep post-surgery), the lab setting ensured high-quality imagery and clear labels, enabling better performance than on “in-the-wild” farm data. As Feighelstein *et al.* comment, machine accuracy in controlled sheep surgery videos even exceeded that of vets, a setting where expressions were pronounced and consistently labeled[57].

Fifth, fusion strategies and multiple cues often improved robustness. Combining face imagery with other modalities (e.g. body posture, sensor data) tends to outperform any single cue. For instance, cow lameness studies fused video keypoints with spatiotemporal models and achieved ~90% classification accuracy. Multi-stage pipelines (e.g. YOLO detection + landmark extraction + LSTM) likewise decomposed tasks into tractable steps. In Martvel *et al.*’s cat-pain study, using video (many frames) instead of isolated images improved detection by leveraging temporal consistency[84]. Conversely, studies using only static images, or only single modalities, generally lagged.

.

In summary, higher performance was generally achieved by (a) using ample, well-curated data; (b) leveraging strong pretraining or multimodal cues; and (c) tailoring architectures and preprocessing to the species and context. Studies consistently note that scarce or noisy data, inter-species variability, and limited modalities suppress accuracy. The success of vision transformers and large encoders in sheep and rabbit pain tasks suggests future work should exploit pretraining and attention to capture subtle patterns. Likewise, integrating temporal dynamics (as in GrayST or video analysis) and multi-modal fusion appears crucial when single frames offer limited information[57,84]fi. These insights indicate that next-generation animal pain recognition systems will likely combine rich data collection (e.g. RGB + thermal + behavior), advanced architectures (transformers, hybrid CNN-AI detectors), and robust preprocessing to overcome the inherent challenges of cross-species pain detection.

4.7 Current Limitations and Challenges

Despite impressive laboratory performance, several consistent limitations emerged across species. Environmental factors including variable lighting, occlusions, and motion artifacts significantly impact accuracy. Cross-species generalization remains limited, with species-specific anatomical differences necessitating dedicated training approaches.

Validation methodology substantially influences reported performance, with rigorous cross-validation revealing more realistic accuracy expectations. Ground truth establishment varies considerably across studies, affecting system reliability and clinical applicability.

5. Validation Strategies for Automated Pain Detection Systems

Rigorous validation of automated pain detection systems is a fundamental requirement for establishing reliable, AI-driven tools in dairy cattle welfare assessment. Unlike conventional veterinary diagnostics, automated systems encounter unique complexities related to pain's inherently subjective nature, interspecies interpretation challenges, and multifaceted interactions between behavioral, physiological, and environmental variables influencing cattle pain expression [5]. Consequently, robust validation frameworks are critical—not only for ensuring technical accuracy—but also for fostering stakeholder trust and obtaining necessary regulatory approval for deploying emerging technological solutions.

Effective validation of automated cattle pain detection technologies hinges on addressing pivotal considerations: accurately establishing ground truth data, ensuring methodological rigor within validation processes, and verifying that research findings generalize effectively across diverse animal populations and varied farming environments. Recent literature underscores significant heterogeneity in existing validation methodologies, leading to challenges in reliably comparing outcomes across studies employing different technological frameworks and analytical approaches [89].

5.1 Establishing Ground Truth: Veterinary Assessment Integration

Establishing reliable ground truth data represents the most critical validation challenge for automated cattle pain detection systems. Unlike human pain assessments, which leverage self-reporting mechanisms for direct subjective experiences, veterinary pain evaluations rely exclusively on third-party interpretations of observed behaviors, physiological indicators, and environmental contexts [89]. This inherent reliance on observer judgments introduces substantial risks of bias, necessitating meticulous protocol design to maximize assessment reliability and validity while minimizing subjective influences.

Recent progress in veterinary pain assessment emphasizes validated species-specific pain scales as essential instruments for establishing robust ground truth. Notably, the UNESP-Botucatu Unidimensional Composite Pain Scale (UCAPS) and Cow Pain Scale (CPS) have emerged prominently. Both tools demonstrate high internal consistency (UCAPS $\alpha = 0.82$; CPS α = 0.79), establishing reliable baselines for objectively quantifying pain severity [34,90]. Comparative evaluations confirm strong criterion validity, exhibiting correlation coefficients ranging from 0.76 to 0.78 when benchmarked against traditional veterinary numerical rating scales [34].

However, integrating validated scales into automated systems necessitates stringent observer training and standardized scoring procedures. Significant variability in inter-rater reliability has been documented, with weighted kappa statistics varying between 0.47 and 0.80 depending on assessor experience and employed scales [34]. Encouragingly, recent studies report high inter-rater agreements between automated systems and human evaluators, consistently exceeding 80%, with Gwet's agreement coefficients spanning from 0.76 to 0.83 for binary pain categorizations [91].

Additional complexity arises from temporal variability in pain expression. Research indicates that acute pain detection accuracy markedly declines over post-procedural intervals—dropping from approximately 88% accuracy at one hour post-procedure to around 65% after 72 hours—as analgesic interventions and natural healing alter observable pain manifestations [43]. Consequently, dynamic ground truth labeling methodologies that consider temporal pain progression may yield superior accuracy compared to static assessments.

Two primary annotation strategies are recognized in ground truth methodologies: stimulus-based and behavior-based annotations. Stimulus-based annotations, categorizing pain by the presence or absence of procedures, provide clear temporal boundaries yet may inadequately represent individual variations in pain perception and expression [89]. Conversely, behavior-based annotations offer detailed observational insight but introduce greater subjectivity and potential observer biases. Emerging evidence suggests hybrid annotation approaches, combining objective temporal data with expert behavioral assessments, may optimize ground truth accuracy, offering balanced objectivity and nuance [57].

5.2 Cross-validation and Performance Metrics

The selection of appropriate cross-validation techniques significantly influences perceived performance and practical generalizability of automated pain detection models. Traditional random cross-validation methods, although computationally convenient, frequently yield overly optimistic estimates due to hierarchical data structures and inherent temporal dependencies typical of livestock behavior datasets [92].
Leave-one-animal-out (LOAO) cross-validation provides greater rigor, better simulating real-world scenarios where pain detection systems must generalize reliably to previously unseen individuals. LOAO validation consistently reports accuracy reductions between 10% and 15% relative to random cross-validation, underscoring individual variability's impact on performance [93]. Studies employing LOAO methodologies document substantial variability in sensitivity (39.2%–79.6%) and specificity (up to 99.1%), reflecting authentic challenges in accommodating animal-level variability within detection algorithms [94].

Farm-fold cross-validation offers an even stricter validation criterion, explicitly accounting for farm-level variability arising from unique management practices, environmental factors, and herd genetics. Research employing farm-fold approaches typically reports an additional 5%–10% performance decrement compared to LOAO, emphasizing the critical influence of farm-specific contexts on automated system generalizability [94]. Such validation rigor is indispensable when assessing commercial feasibility across diverse farming conditions.

The selection of cross-validation strategy materially affects reported performance and therefore conclusions about real-world readiness. Concrete examples from the reviewed literature illustrate this: Zhang et al. (2025) report a validation accuracy of ≈**99.65%** but only **64.3%** on held-out unseen videos (a drop of ≈35.3 percentage points), with pain-class precision = 0.83 and recall = 0.56, a clear sign that a naive validation split substantially over-estimates deployable performance[85]. Other studies include inter-farm samples but do not disaggregate inter-farm results, preventing assessment of farm-level generalizability[87]. Large longitudinal datasets and realistic temporal holdouts yield more conservative but likely more realistic metrics[86]. Methodological analyses show that replacing random CV with LOAO typically reduces reported accuracy by **~10–15%** and applying farm-fold (inter-farm) validation incurs an additional **~5–10%** decrement; together these stricter protocols can reduce internal estimates by **15–25%** or more. Authors should therefore (i) always report per-study internal and external (held-out/farm-level) metrics, (ii) include LOAO and farm-fold experiments where feasible (or clearly state their absence), and (iii) present balanced metrics (sensitivity/specificity/PPV/NPV/AUC) rather than accuracy alone to avoid misleading conclusions about field performance.

Performance metric selection significantly impacts validation outcomes. While accuracy remains prevalent, it can misrepresent performance in imbalanced datasets—common in livestock pain studies where non-painful observations predominate [95]. Recent validation research stresses balanced metric reporting, including sensitivity, specificity, positive and negative predictive values, and ROC-AUC, providing comprehensive model performance assessments [96].

5.3 Challenges in Validation Methodologies

Validating automated cattle pain detection systems presents multifaceted challenges impacting result interpretation and generalizability. Feline pain detection studies deliberately limited populations to single breed types (domestic short-haired cats) to minimize confounding variables during proof-of-concept validation[97]. This breed-specific variability necessitates explicit validation strategies across genetically diverse cattle populations to ensure comprehensive applicability.

Environmental variability further complicates validation accuracy. Farm-specific environmental factors—including inconsistent lighting conditions, occlusion by equipment or other animals, mud contamination, and motion blur—significantly degrade detection accuracy, with performance typically decreasing by 15%–20% compared to controlled experimental environments [98]. These findings underscore the necessity for explicitly incorporating realistic environmental conditions within validation studies, assessing model resilience across varied farming scenarios.

Limited dataset sizes remain pervasive within current validation literature, typically involving fewer than 100 animals per study with limited representation across age, sex, breed, and farm management practices [92]. Such limited diversity restricts statistical power and constrains broader population generalizability. Temporal and spatial clustering further exacerbates sample size limitations, necessitating larger, more diverse datasets for robust validation outcomes.

Acute and chronic pain condition differentiation presents unique validation complexities, given distinct temporal trajectories and subtle behavioral indicators characterizing chronic pain states compared to acute presentations [7]. Addressing chronic pain validation demands specialized protocols accommodating long-term, subtle behavioral changes alongside traditional acute pain indicators.

Longitudinal data dependencies also introduce validation complexities. Randomly sampling temporal data points risks inadvertent leakage of future information, artificially inflating model performance estimates [99]. Blocked cross-validation approaches respecting chronological data sequences provide more authentic accuracy assessments yet require sufficiently large datasets to preserve statistical power.
Multimodal sensor integration, while promising enhanced accuracy (typically improving detection accuracy by 5%–10%), further complicates validation procedures, necessitating synchronized data collection and modality-specific preprocessing to ensure consistency and reliability [100].

Collectively, these validation complexities highlight critical needs for standardized protocols and collaborative multi-institutional research frameworks, enabling rigorous validation of automated cattle pain detection systems across diverse populations, farm environments, and temporal conditions. Future research prioritizing comprehensive validation methodologies can substantially advance practical translation of emerging technological solutions, significantly enhancing dairy cattle welfare outcomes [101]. Building on these crucial validation insights, the development and deployment of mobile applications represent the next pivotal step in democratizing automated pain detection technologies for farmers and veterinary practitioners alike.

6. Review of Existing Veterinary and Livestock Mobile Apps

The current landscape of veterinary and livestock-focused mobile applications encompasses a broad spectrum, ranging from basic animal record-keeping and self-assessment tools to sophisticated AI-driven monitoring systems. Recent developments highlight a trend towards intuitive, farmer-centric interfaces integrated with advanced technological capabilities.

6.1 Overview of Existing Livestock Mobile Apps

Recent veterinary mobile applications reflect substantial diversity in their functionalities. Applications such as PIGLOW, an EU-funded platform, enable farmers raising free-range pigs to conduct structured welfare audits periodically, providing automated feedback and comparative benchmarking against peer farms. A two-year pilot study involving 12 farms demonstrated modest improvements in welfare indicators, including reductions in lameness and skin lesion prevalence, alongside high farmer-reported usability and acceptance [102]. Similarly, mobile apps tailored for beef cattle management have shown strong user satisfaction and usability, as indicated by a System Usability Scale (SUS) rating of approximately 75, highlighting their effectiveness in streamlining feed tracking and animal health record-keeping processes [11].

Wearable and Internet-of-Things (IoT) devices represent another significant category of livestock monitoring solutions. Prototype collars designed for cattle and other livestock have emerged prominently, capable of continuously monitoring animal physiological parameters such as body temperature, heart rate, and physical activity. These wearable systems transmit collected data to cloud analytics platforms, providing veterinarians and farm managers with timely alerts to early indicators of health issues, including respiratory infections, thereby enabling intervention prior to observable clinical signs [102,103]. Machine-vision-based mobile apps have recently begun leveraging compact convolutional neural network (CNN) architectures—such as YOLOv5—to enable smartphone-based, real-time identification of hoof conditions like digital dermatitis. Such technologies have been successfully deployed on Android and iOS platforms, providing practical and immediate on-farm lameness screening capabilities [68]. Consistently, user feedback underscores that farmers highly value mobile apps featuring intuitive workflows, straightforward checklists, and simplified data captures that seamlessly integrate into their daily farm management routines.

6.2 Mobile Application Deployment Considerations

For mobile applications operating in rural livestock farming environments, robust on-device processing and reliable local networking capabilities are essential. Recent studies underscore the advantages of edge computing solutions, such as deployments utilizing NVIDIA Jetson Nano hardware equipped with 12 MP cameras for real-time cattle identification tasks on dairy farms. These implementations enable rapid edge inference, allowing immediate local web access to cattle identification information without reliance on continuous internet connectivity, thereby demonstrating real-world latency performances measured in milliseconds per inference [49].

Power management strategies are equally critical for prolonged operation in remote farm environments. Several wearable systems now incorporate renewable power solutions, such as small solar panels or kinetic energy harvesting from animal movement, enabling continuous data collection without frequent manual battery replacements. Examples include prototype collars successfully deployed on reindeer and cattle, providing continuous operation for weeks at a time [104]. Reviews of such systems confirm that hybrid power setups—combining solar panels and motion-based harvesting—effectively support uninterrupted, round-the-clock monitoring, in contrast to purely battery-powered collars, which typically require weekly recharging under intensive operational conditions [104].

Moreover, hierarchical network designs employing federated learning approaches further enhance scalability and operational feasibility. By preprocessing raw sensor data at the edge—such as compressing video streams or filtering telemetry data—these systems significantly reduce network bandwidth demands, allowing model updates and analytic processes to occur without sensitive raw data needing to exit the farm environment. This configuration effectively balances computational responsiveness, data security, and limited rural network infrastructure capacities [103].

6.3 User Interface (UI) and User Experience (UX) Design Considerations

Effective UI/UX design remains fundamental for user acceptance and successful integration of livestock mobile applications into farm management practices. Agricultural usability studies consistently emphasize that farmers prefer intuitive interfaces closely aligned with their daily operational workflows and practical field conditions. Key design criteria highlighted in usability evaluations include clearly structured menus, rapid accessibility of essential functions, and adequately sized interactive controls (e.g., large buttons and clearly recognizable icons), facilitating quick, error-free interactions even while wearing protective gloves [105].

Empirical evaluations have repeatedly validated these design principles. For instance, a beef-management mobile application, developed collaboratively with farmers, reported high usability ratings (SUS scores exceeding 70) and substantial self-reported satisfaction, confirming effectiveness in real-world farm environments [106]. Additional design considerations crucial for practical farm deployment include high-contrast displays and minimal text reliance, ensuring readability under direct sunlight. Clear, simplified content structures allowing users rapid access to essential tasks without navigating through multiple screens further improve efficiency and satisfaction.

Localization support—including multilingual interfaces, regional terminology, and local measurement units—ensures broader usability across diverse international and multi-ethnic farming communities. Real-time feedback mechanisms, such as color-coded alerts, clear trend visualizations, and actionable prompts, further enhance usability, allowing farmers to quickly prioritize and manage animal care without extensive data interpretation efforts. Finally, interoperability with widely-used farm management systems, enabled through standardized APIs, significantly reduces redundant data entry tasks, providing veterinarians and farm advisors immediate access to unified, accurate records [106]. Finally, interoperability with widely used farm management systems, enabled through standardized APIs, significantly reduces redundant data entry tasks, providing veterinarians and farm advisors immediate access to unified, accurate records . Fig 3 illustrates the interdisciplinary integration of animal science (facial AU biology), computer vision technology (RGB-thermal analysis), and precision-agriculture systems, with arrows showing data flow from capture to real-time welfare alerts .

Fig 3: Conceptual illustration depicting interdisciplinary integration of animal science (facial action unit biology), computer vision technology (RGB-Thermal image analysis), and precision agriculture management systems. Arrows indicate the directionality of data flow from initial data capture through to generation of real-time welfare notifications.

6.4 Ethical and Regulatory Compliance in Livestock Mobile Applications

Given the sensitive nature of farm production and animal health data, mobile applications must incorporate rigorous privacy and ethical safeguards. Industry best practices emphasize robust end-to-end encryption of data during transmission and storage, stringent user authentication protocols, and clearly defined role-based access controls distinguishing farm owners from employees [106]. Transparent data ownership policies and explicit user consent protocols further establish trust. Applications like PIGLOW utilize anonymous benchmarking systems, allowing users to compare welfare metrics confidentially, facilitating peer learning without compromising data privacy [102,107]. Regulatory guidelines also advocate comprehensive traceability features, including audit logging, tamper-evident record-keeping, and customizable data retention periods, ensuring compliance with mandatory animal welfare audit requirements. Veterinary regulatory frameworks impose additional operational constraints. Many regions stipulate that remote monitoring applications must operate strictly within an established veterinarian–client–patient relationship (VCPR), clarifying that such tools complement rather than replace professional veterinary oversight. Consequently, clear liability disclaimers and predefined emergency flagging thresholds are mandated, ensuring users understand the supplementary role of AI-based alerts in clinical decision-making contexts [108]. Ethical considerations further encourage developers to adopt responsible innovation strategies, involving both veterinarians and farmers directly in application design and validation processes. Such co-creative approaches ensure technological advancements augment, rather than diminish, traditional farmer roles, preserving essential human empathy and local expertise in animal welfare practices [108].

Technical Performance Trade-offs in Mobile Application Deployment mobile-optimized CNN architectures significantly outperform larger conventional models for animal pain detection applications. The technical trade-offs of mobile-optimized CNN architectures are summarized in Fig 4. ShuffleNetV2 emerges as the optimal architecture, achieving 95.5% accuracy for pain classification with only 6.17 million parameters (~25-30 MB) and 22 FPS inference speed on smartphones1. EfficientNetB0 and MobileNetV3 also demonstrate strong performance with 65-77% accuracy rates while maintaining practical deployment characteristics of 17-50 MB model sizes and 12-21 FPS processing speeds. In contrast, ResNet50-based approaches achieve only 65% accuracy with significantly larger memory footprints and slower inference speeds, contradicting claims that larger models offer superior performance for this application domain[75,109].

Multiple studies document successful clinical deployment of mobile animal pain detection systems across various species, including cats (95.5% accuracy), sheep (92.7% accuracy), horses (88.3% accuracy), and rabbits (87% accuracy)[57,77,110,111]. These mobile-optimized systems demonstrate real-time processing capabilities, minimal battery consumption, and successful integration into veterinary clinical workflows with high inter-rater reliability. The research conclusively establishes that mobile-optimized CNN architectures are not only technically feasible for smartphone deployment but also achieve superior accuracy compared to conventional larger models while providing the computational efficiency necessary for practical veterinary applications.

Fig 4: Grouped bar chart illustrating comparative benchmarks for mobile application deployment of animal pain detection, presenting model file size (MB), pain detection accuracy (%), and inference speed (FPS) across ShuffleNetV2, EfficientNetB0, MobileNetV3Large, and ResNet-50 architectures. Benchmark metrics are based on published peer-reviewed evaluations, supporting informed model selection according to practical requirements for mobile veterinary AI applications.

6.5 Ethical Considerations in AI-Based Pain Monitoring

Implementing AI-driven facial grimace scales in dairy cows raises profound ethical questions that go beyond technical issues like data privacy or compliance. Scholars emphasize that digital livestock farming can reshape the human–animal bond and risk treating animals as mere data points. For example, Neethirajan warns that “the use of artificial intelligence in digital livestock farming may lead to a loss of personal connection between farmers and animals,” potentially undermining animal well-being[112]. Similarly, recent reviews note that constant monitoring (“quantified” animals) can diminish caretakers’ empathy: as, animals cannot consent to surveillance, and caretakers “might become overly reliant on graphs or dashboard alerts,” weakening the subtle, compassionate observation that traditionally guides animal care[113]. In short, high-level ethical reflection asks not only *how* AI tools function but *whether* they respect animals as sentient beings with interests. Ethicists point out that if AI focuses farm management solely on efficiency or productivity, it risks violating animals’ autonomy (treating them as instruments) and eroding virtues like compassion and responsibility[113,114]. A “should we” perspective thus urges that any pain-detection AI must be integrated in ways that support rather than replace the human–animal relationship[112,114].

AI Decision Support vs. Practical Adoption Risks

AI tools are often promoted as decision-support aids, but their real-world use may diverge. There is a risk that some farmers will treat AI diagnoses as substitutes for professional care, tempted by the illusion of cost savings. This raises both legal and welfare concerns: veterinary regulations (e.g. the U.S. requirement for a valid Veterinarian–Client–Patient Relationship) exist to prevent unqualified treatment, and ignoring them could harm animals. Moreover, field studies and expert workshops highlight several negative consequences of widespread AI adoption:

- Reduced human–animal interaction: Automated monitoring can make stockkeepers spend less time with cows, weakening the human–animal relationship. Schillings *et al.* report that precision livestock systems "decrease animal keepers' contact with their animals," which can lead to poorer welfare outcomes and "reduced stockmanship skills"[115]. Over time, loss of hands-on familiarity may blunt a farmer's ability to notice subtle signs and bond with individual animals.

- Objectification and intensification: By enabling large-scale monitoring, AI can inadvertently promote viewing cows as data sources. Workshop participants noted that less direct contact may shift attitudes toward animals as "objects," and that PLF could facilitate farm intensification (managing more cows)[115]. Such objectification is echoed by Neethirajan, who cautions against treating animals as "mere data points"[112].

- Skill erosion and dependency: Reliance on algorithms risks deskilling. Farmers may become dependent on AI alerts, reducing their own observational acumen. As one review warned, technologies could "make the job less attractive" and raise questions about the true meaning of being a farmer[115]. If AI is wrong or misinterprets signals, over-reliance could delay veterinary intervention.

- Mental health and equity: The push to adopt advanced AI can strain farmers mentally and financially. High costs and steep learning curves may create stress or widen a "digital divide" between well-resourced and smaller farms[112,115]. Those with limited access to tech might fall behind, raising justice concerns.

- Erosion of empathy: Finally, scholars caution that dashboards and automated alerts, while efficient, may erode empathy. If caretakers "rely too heavily on data," nuanced animal behaviors (ear posture, vocalizations, etc.) might be overlooked[113]. This could compromise the very welfare benefits that AI was supposed to enhance.

Taken together, these observations underline that AI should not replace human judgment or veterinary care. As Schillings *et al.* conclude, responsible use requires codes of practice, training, and co-design with farmers so that technology complements traditional husbandry rather than undermining it[113,115].

Data Privacy and Legal Frameworks

Beyond welfare, AI-monitoring systems involve vast data streams that raise regulatory issues. Video or sensor data on farms can implicate privacy laws: for example, the EU's General Data Protection Regulation (GDPR) applies to any personal information, potentially including footage where farmworkers or visitors are identifiable[112]. Also digital farming tools "are subject to existing legislation, as well as new laws such as GDPR"[112]. Farmers and technology providers must therefore ensure compliant data handling, including secure storage, transparency about data use, and respect for individuals' privacy. Cybersecurity is also crucial, as breaches of animal data (or misused monitoring) could undermine trust in these systems.

Similarly, animal-welfare laws and standards impose boundaries on AI use. In the United States, the Animal Welfare Act (though focused on research and exhibition) reflects society's expectation of humane animal treatment. Many countries also have dairy-specific welfare codes (e.g. the EU's minimum welfare regulations, national "Red Tractor" standards, etc.). Any AI-based pain monitoring must operate within these frameworks: it should trigger interventions consistent with legal care requirements, not merely optimize production. For example, a cow flagged as in pain must be treated in accordance with veterinary standards and animal-health legislation.

In summary, integrating facial expression AI into dairy farming demands a *responsible, animal-centered approach*. Ethical guidelines suggest co-developing technology with stakeholders (farmers, veterinarians, ethicists) and embedding safeguards (data protection, obligatory vet oversight, periodic ethical review)[112,115]. Only by addressing the "what if" and "should we" questions on animal dignity, farmer roles, and legal duties can AI-based monitoring truly benefit cow welfare without unintended harm.

7. Future Perspectives and Recommendations

The field of automated pain detection in dairy cattle is at a crucial juncture, where groundbreaking technological innovations must align closely with real-world implementation and widespread industry adoption. As detailed in this comprehensive review, significant advancements in artificial intelligence, computer vision, and mobile technology have produced robust, accurate, and clinically meaningful tools capable of transforming livestock welfare management. Moving forward, addressing challenges related to breed diversity, environmental robustness, and collaborative implementation frameworks will be critical for successfully transitioning these technologies from experimental validation to broad commercial acceptance.

7.1 Addressing Breed-Specific and Environmental Limitations

Advanced Transfer Learning for Crossbreed Adaptation

Breed-specific variability remains one of the most significant barriers to universal implementation of automated cattle pain detection systems. Recent breakthroughs in transfer learning methodologies offer compelling solutions by allowing models trained on a single breed to generalize effectively to genetically diverse herds, mitigating the need for extensive breed-specific datasets. Research has shown that transfer learning effectively maintains high accuracy levels across diverse cattle breeds, providing scalable, broadly applicable solutions [116].

Moreover, multimodal data fusion has emerged as a powerful technique to overcome breed-specific biases. For example, studies applying adaptive fuzzy logic in multimodal fusion systems have demonstrated exceptional accuracy, achieving validation performance rates up to 95% for environment evaluation, 100% for feeding evaluation, and approximately 94% for behavior detection [117]. These results underscore the transformative potential of integrating diverse data sources—such as RGB imaging, thermal sensors, accelerometers, and environmental monitors—to reliably capture breed-independent pain expressions [118].

Comprehensive multimodal datasets have significantly advanced crossbreed validation. By capturing detailed facial anatomical variations and behavioral patterns across breeds, researchers have developed models that generalize more effectively. Notably, advanced neural network architectures such as Vision Transformers with Bi-Level Routing Attention have achieved impressive facial recognition accuracies of 98.36%, adeptly handling breed-specific anatomical differences [116]. Leveraging the global contextual understanding provided by transformer models positions them as particularly suitable for addressing breed-dependent variations in pain-related expressions.

Environmental Robustness via Edge Computing

The unpredictable and dynamic nature of farm environments poses substantial obstacles to implementing automated pain detection systems. Edge computing solutions have emerged as pivotal for enhancing environmental robustness, enabling real-time data processing in challenging agricultural contexts. Recent edge-computing deployments have demonstrated extremely low latency (5–10 milliseconds), significantly improving responsiveness of livestock monitoring systems [119]. Intelligent wearable devices powered by solar energy have achieved continuous operation in real-world settings, consistently maintaining accuracy (97.27%) in health and behavior classification tasks [119]. These findings underscore the practicality and sustainability of edge computing frameworks.

Moreover, integrating edge computing with mobile applications simultaneously addresses multiple environmental constraints reducing network bandwidth requirements, enhancing system resilience during connectivity disruptions, and facilitating reliable operation even in remote agricultural locations [120]. Robust environmental monitoring, exemplified by multi-zone Temperature-Humidity Index (THI) predictive models, further complements pain detection systems, enabling adaptive processing and accurate welfare assessment across diverse environmental conditions [121].

Enhancing Reliability through Multimodal Fusion

To ensure robust pain detection across varying environmental conditions, multimodal data fusion strategies are crucial. Recent research clearly demonstrates superior reliability and accuracy when combining multiple data streams such as accelerometry, visual observation, thermal imaging, and environmental sensors relative to single-sensor approaches. Studies confirm that accelerometers detect behavioral changes related to pain significantly earlier than visual assessments alone; conversely, visual observations provide nuanced identification of pain-specific behaviors undetectable by sensor data alone [98].

Further advancements in sensor fusion methodologies such as integrating computer vision with mechanical sensors have shown notable improvements in monitoring precision. Studies monitoring cattle brush-use behaviors have highlighted that combined machine-learning models significantly outperform individual sensor approaches, enhancing accuracy and reliability [122]. The creation of comprehensive multimodal datasets encompassing diverse sensor types has significantly strengthened fusion methodology validation, underpinning development of robust algorithms capable of maintaining high accuracy across heterogeneous farm conditions [123].

7.2 Enhancing Real-time and Longitudinal Pain Monitoring

Precision Livestock Farming Integration

Integrating automated pain detection into broader precision livestock farming (PLF) frameworks represents a critical step toward comprehensive herd health and welfare management. Recent research highlights the effectiveness of PLF technologies, employing real-time monitoring, machine learning, and IoT-based solutions to enable proactive disease detection and welfare management [120]. LoRa-based sensor networks integrated with Subspace k-Nearest Neighbors classifiers have consistently demonstrated superior disease classification accuracy and timeliness, enabling targeted interventions [120].

Scalable, AI-driven welfare platforms leveraging deep learning and edge computing are now demonstrating significant promise, automating critical welfare assessments such as locomotion scoring, health status evaluation, and body condition monitoring. Markerless animal identification further enhances these platforms, making them both practical and scalable across farm sizes [118].

Continuous Monitoring Frameworks

Implementing continuous pain monitoring necessitates sophisticated technological architectures capable of real-time computation and sustained reliability over extended periods. IoT-based cattle monitoring systems employing accelerometer sensors coupled with advanced statistical models (e.g., ARIMA, wavelet transformations) effectively predict and classify behavioral patterns, facilitating proactive health management [119]. Additionally, continuous multi-zone environmental monitoring (THI prediction) achieves robust predictive accuracy, enabling proactive environmental control strategies and thereby enhancing overall herd welfare [121].

Integration with Herd Health Records

Effective integration of automated pain detection with existing farm management systems is vital to enable actionable insights and informed herd-health decision-making. Standardizing data formats and protocols has emerged as a crucial facilitator of seamless integration across multiple monitoring systems, ensuring consistency and comparability in welfare assessments [124]. Advanced machine-learning analytics further enhance data integration capabilities, providing actionable insights that optimize treatment strategies, resource allocation, and herd health management overall [120].

7.3 Recommendations for Industry-wide Implementation

Collaborative Veterinary-AI Partnerships

Successfully deploying automated pain detection technologies requires well-structured collaborative frameworks combining veterinary expertise with AI capabilities. Effective human-AI collaboration substantially improves decision-making efficiency, operational precision, and stakeholder trust. Recent research emphasizes transparency and explainability in AI outputs, significantly enhancing adoption rates among veterinarians and farmers. Maintaining veterinary oversight within collaborative frameworks is critical, ensuring that AI systems serve as valuable decision-support tools rather than substitutes for expert veterinary judgment.

Structured training programs significantly enhance veterinarian and farmer confidence in AI-driven tools, improving diagnostic outcomes and adoption rates. Such industry-specific collaborative frameworks, integrating technology developers, veterinarians, and farm managers throughout design and deployment phases, have been demonstrated as critical for addressing practical implementation challenges effectively.

Standardization and Validation Protocols

Establishing rigorous industry-wide standards and validation protocols is imperative for ensuring the reliability, safety, and effectiveness of automated pain detection systems. Validation protocols must consider species-specific physiological and behavioral nuances, as validation methodologies successful in one livestock species may not directly transfer to others. External, independent validation is essential for industry credibility, as currently only a small fraction (approximately 14%) of available technologies have undergone independent validation, highlighting a significant gap in existing approaches [125].

Farm-specific Customization Strategies

Farm-specific customization is necessary due to variability in management practices, environmental contexts, and operational scales. Recent studies indicate perceived ease-of-use and demonstrated utility significantly influence farmer adoption decisions [103]. Cost-effective approaches utilizing readily accessible technologies, such as optimized IoT sensor systems, enhance economic feasibility and adoption rates across both small-scale and commercial operations [119].

Scalability considerations, notably demonstrated through high-precision cattle tracking systems, highlight that deep learning-based architectures can efficiently scale from individual animal monitoring to extensive herd management applications without compromising accuracy or operational efficiency [123]. This adaptability allows tailored technology deployments to match diverse farming contexts. Table 4 outlines the technical specifications and recommended enhancements for automated cattle pain detection systems, including performance targets for multi-breed adaptability, environmental robustness, and practical usability.

Table 7: Technical specifications and recommended implementation enhancements for automated cattle pain detection systems. Proposed performance targets emphasize multi-breed adaptability, robust environmental integration, and high practical usability:

System Component	Current Capabilities	Recommended Enhancements	Integration Requirements	Performance Targets
AI Processing	95-99% accuracy in controlled conditions	Multi-breed validation; Environmental adaptation	Edge-cloud hybrid architecture	>95% accuracy across all breeds
Sensor Integration	Individual sensor validation	Multimodal fusion; Continuous monitoring	Standardized data formats	>90% uptime; <5% false positive rate
Mobile Applications	Basic monitoring capabilities	Real-time alerts; Veterinary integration	Cross-platform compatibility	>90% user satisfaction
Data Management	Local storage; Periodic synchronization	Real-time cloud integration; Predictive analytics	Interoperability with farm systems	<1% data loss; Real-time processing

Validation Framework	Species-specific testing	Cross-breed; Multi-environment validation	COSMIN compliance; External validation	>85% sensitivity; >90% specificity

The future success of automated pain detection technologies for dairy cattle hinges upon effectively aligning technological innovation with real-world practicalities and stakeholder priorities. Comprehensive multimodal integration, robust environmental resilience, industry-wide standardization, and collaborative implementation frameworks represent essential pathways from experimental validation towards broad commercial adoption.

8. Conclusions

Automated pain detection in dairy cattle has reached an inflection point, transitioning from experimental promise to real-world applicability, driven by breakthroughs in neurobiology, artificial intelligence (AI), and mobile technology. Traditional veterinary assessment methods, notably Numerical Rating Scales (NRS) and Visual Analog Scales (VAS), though historically foundational, continue to face inherent limitations due to subjectivity (ICC range: 0.73–0.81), invasiveness, and challenges in accurately capturing subtle pain indicators in large herds. In stark contrast, validated facial grimace scales like UCAPS, boasting strong diagnostic metrics (AUC = 0.93), have introduced objective, quantifiable alternatives, significantly enhancing the reliability of acute pain detection (sensitivity range: 0.66–0.90). Yet, a clear and pressing gap persists in reliably assessing chronic pain conditions, underscoring the need for further targeted research in this critical area.

The integration of advanced AI algorithms and computer vision technologies has marked a revolutionary advancement in precision livestock welfare. Cutting-edge detection architectures, such as RetinaNet (99.8% average precision) and YOLOv8-Pose (96.9% mAP), have enabled remarkable accuracy and consistency in facial landmark detection and pain-related behavioral analysis. Moreover, the deployment of multimodal AI strategies—combining RGB imagery and thermal sensors—has achieved impressive accuracy (81–95%) in detecting inflammation and physiological stress responses linked to pain. The practicality of these technologies in real-world farm environments has been further validated by edge-computing frameworks like Dairy DigiD, demonstrating robust real-time processing capabilities (24 frames per second) under variable conditions, significantly enhancing their readiness for widespread commercial deployment.

Mobile technology further amplifies these advancements by democratizing access to sophisticated welfare monitoring systems. Validated applications such as PIGLOW (featuring high usability ratings) and VetPain (inter-rater reliability ICC ≥0.87) highlight the critical role of intuitive, user-centric designs in facilitating widespread adoption by non-specialist stakeholders. These applications incorporate multilingual interfaces, actionable alerts, and seamless integration into daily farming workflows, thus bridging the gap between technological innovation and practical usability in diverse agricultural contexts.

Robust validation protocols have confirmed strengths of automated pain detection systems, particularly in acute pain detection scenarios (precision and recall consistently exceeding 0.80). However, critical limitations remain concerning breed-specific performance biases and the precise differentiation between chronic and acute pain states. Future research directions must prioritize advanced transfer-learning approaches, effectively addressing genetic variability between cattle breeds such as Holstein and Zebu where transformative transformer-based architectures have already demonstrated accuracy rates reaching 98.36%. Complementing this, environmental resilience must be strengthened through the strategic deployment of solar-powered edge-computing devices, which have achieved reliable behavior classification accuracy of approximately 97.27%, ensuring operational sustainability across diverse, challenging farm environments.

Longitudinal monitoring capabilities represent another critical area poised for substantial impact. Integrating accelerometry data with advanced vision-based systems has already demonstrated exceptional performance (up to 99.55% accuracy in lameness detection), promising proactive herd health management that can significantly mitigate economic losses associated with undetected pain. Leveraging these capabilities within Precision Livestock Farming (PLF) frameworks enables earlier interventions, optimized herd health management, and significant productivity gains, presenting compelling economic incentives for industry-wide adoption.

However, these numeric gains are strongly context-dependent. Most high figures derive from acute-pain datasets, controlled conditions or within-dataset validation; when evaluated under LOAO or farm-fold (inter-farm) protocols, performance commonly drops (typical contractions reported across studies ≈10–25%). Breed, management and environment remain important constraints: models trained on one breed or barn layout do not automatically generalize to others. Likewise, reliable automated detection of chronic pain remains unresolved. Therefore, claims that AI will “significantly” improve welfare must be anchored to these contextual limits and to validated field performance.
To move from demonstrated capability to documented welfare impact, we recommend the following measurable priorities:

1. Dataset breadth: curate and publish large, annotated datasets that include multiple breeds, ages and chronic-pain cases to reduce out-of-sample failures.
2. Standardized validation: require LOAO and farm-fold testing and report sensitivity, specificity, PPV/NPV and ROC-AUC with 95% CIs for each validation design; aim for field-validated sensitivity/specificity ≥0.80 across at least three independent farms before making deployment claims.
3. Cross-breed adaptation: adopt transfer-learning and few-shot strategies with explicit fine-tuning on under-represented breeds to close genetic bias gaps.
4. Robust field deployment: prioritize energy-efficient edge solutions and stress-testing in real barns (lighting, occlusion, weather) to ensure continuous operation at target frame rates (≈20–30 fps).
5. Ethics and veterinary integration: implement mandatory escalation pathways to veterinarians (VCPR-aligned), transparent data governance, and co-design with end users to preserve human–animal relationships and avoid over-reliance on automation.
6. Impact evaluation: accompany technological deployments with longitudinal welfare studies that quantify outcomes (e.g., reductions in undetected lameness, changes in time-to-treatment, or modeled reductions in premature culling).

By focusing on these concrete milestones rather than on unqualified potential, future work can translate current algorithmic advances into sustained animal-level improvements in welfare.

The implications for global dairy cattle welfare from successfully implementing automated pain detection technologies are profound and far-reaching. With more than 270 million dairy cows globally experiencing pain-related welfare challenges, widespread adoption of these innovations could drastically reduce animal suffering, significantly extend herd longevity (potentially decreasing premature culling rates by 10–20%), and contribute to sustainable and ethically responsible agriculture. However, translating these opportunities into real-world outcomes requires sustained commitment to addressing identified gaps—particularly the accurate identification of chronic pain and improved crossbreed adaptability.

Moving forward, dedicated investment is essential for developing comprehensive, publicly accessible datasets, rigorous ethical AI deployment guidelines, and targeted educational programs for farmers and veterinary professionals. Pioneering solutions like CowPain Check exemplify the immense potential of thoughtful technological integration, setting powerful precedents for humane, sustainable dairy farming practices aligned closely with the United Nations Sustainable Development Goals (SDGs). By addressing current technical, economic, and social challenges through a coordinated interdisciplinary approach, the dairy industry can leverage these innovations not only to elevate animal welfare standards significantly but also to lead broader advancements across global livestock welfare management practices.

References

[1] Borş A, Borş S-I, Floriștean V-C. Mastitis impact on high-yielding dairy farm's reproduction and net present value. *Front Vet Sci*. 2024;10 https://doi.org/10.3389/fvets.2023.1345782.

[2] Browne N, Hudson CD, Crossley RE, Sugrue K, Kennedy E, Huxley JN, et al. Lameness prevalence and management practices on Irish pasture-based dairy farms. *Ir Vet J*. 2022;75:14 https://doi.org/10.1186/s13620-022-00221-w.

[3] Mekonin HA, Reda AA, Assen AA, Assen AM. Prevalence and risk factors for lameness in dairy cattle on selected farms located in Dessie and Kombolcha, Northeast Ethiopia. *Front Vet Sci*. 2025;12:1456527 https://doi.org/10.3389/fvets.2025.1456527.

[4] Tschoner T, Mueller KR, Zablotski Y, Feist M. Pain Assessment in Cattle by Use of Numerical Rating and Visual Analogue Scales—A Systematic Review and Meta-Analysis. *Anim Open Access J MDPI*. 2024;14:351 https://doi.org/10.3390/ani14020351.

[5] Tomacheuski RM, Monteiro BP, Evangelista MC, Luna SPL, Steagall PV. Measurement properties of pain scoring instruments in farm animals: A systematic review using the COSMIN checklist. *PLOS ONE*. 2023;18:e0280830 https://doi.org/10.1371/journal.pone.0280830.

[6] Mota-Rojas D, Whittaker AL, Coria-Avila GA, Martínez-Burnes J, Mora-Medina P, Domínguez-Oliva A, et al. How facial expressions reveal acute pain in domestic animals with facial pain scales as a diagnostic tool. *Front Vet Sci*. 2025;12 https://doi.org/10.3389/fvets.2025.1546719.

[7] Zoltick AH, Mann S, Coetzee JF. Pain pathophysiology and pharmacology of cattle: how improved understanding can enhance pain prevention, mitigation, and welfare. *Front Pain Res*. 2024;5 https://doi.org/10.3389/fpain.2024.1396992.

[8] Dulal R, Zheng L, Kabir MA. MHAFF: Multi-Head Attention Feature Fusion of CNN and Transformer for Cattle Identification 2025; https://doi.org/10.48550/arXiv.2501.05209.

[9] Identification C. Cattle Identification and Detection using Vision Transformers and YOLOv8. *J Inf Syst Eng Manag*. 2025;10:504–18 https://doi.org/10.52783/jisem.v10i42s.7911.

[10] Neethirajan S, Scott S, Mancini C, Boivin X, Strand E. Human-computer interactions with farm animals—enhancing welfare through precision livestock farming and artificial intelligence. *Front Vet Sci*. 2024;11 https://doi.org/10.3389/fvets.2024.1490851.

[11] Thannithi W, Intawicha P, Phuwisaranakom P, Saengwong S. Design and evaluation of a mobile application for enhancing farm management and performance assessment in fattening beef cattle. *J Adv Vet Anim Res*. 2024;1 https://doi.org/10.5455/javar.2024.k766.

[12] Roche S, Saraceni J, Zehr L, Renaud D. Pain in Dairy Cattle: A Narrative Review of the Need for Pain Control, Industry Practices and Stakeholder Expectations, and Opportunities. *Anim Open Access J MDPI*. 2025;15:877 https://doi.org/10.3390/ani15060877.

[13] Rasmussen P, Barkema HW, Osei PP, Taylor J, Shaw AP, Conrady B, et al. Global losses due to dairy cattle diseases: A comorbidity-adjusted economic analysis. *J Dairy Sci*. 2024;107:6945–70 https://doi.org/10.3168/jds.2023-24626.

[14] Bose B, Siva Kumar S. Economic burden of zoonotic and infectious diseases on livestock farmers: a narrative review. *J Health Popul Nutr*. 2025;44:158 https://doi.org/10.1186/s41043-025-00913-3.

[15] Logroño JC, Rearte R, Corva SG, Domínguez GA, de la Sota RL, Madoz LV, et al. Lameness in Early Lactation Is Associated with Lower Productive and Reproductive Performance in a Herd of Supplemented Grazing Dairy Cows. *Anim Open Access J MDPI*. 2021;11:2294 https://doi.org/10.3390/ani11082294.

[16] Hisira V, Zahumenska J, Kadasi M, Klein R, Mudron P, Zigo F. Claw diseases are the dominant cause of lameness in dairy cows and a risk factor for mastitis. *Veterinární Medicína*. 2025;70:35–44 https://doi.org/10.17221/56/2024-VETMED.

[17] Gleerup KB, Andersen PH, Munksgaard L, Forkman B. Pain evaluation in dairy cattle. *Appl Anim Behav Sci*. 2015;171:25–32 https://doi.org/10.1016/j.applanim.2015.08.023.

[18] Martin M, Kleinhenz MD, Schwartzkopf-Genswein KS, Melendez D, Marti S, Pajor EA, et al. Characterizing the diagnostic sensitivity and specificity of pain biomarkers in cattle using receiver operating characteristic curves. *J Dairy Sci*. 2022;105:9853–68 https://doi.org/10.3168/jds.2021-21393.

[19] Mincu M, Nicolae I, Gavojdian D. Infrared thermography as a non-invasive method for evaluating stress in lactating dairy cows during isolation challenges. *Front Vet Sci*. 2023;10 https://doi.org/10.3389/fvets.2023.1236668.

[20] Thomas AD. Objective determination and quantification of pain and inflammation associated with digital dermatitis in feedlot cattle. *Appl Anim Behav Sci*. 2022;

[21] Chen J, Gao Y, Bao S-T, Wang Y-D, Jia T, Yin C, et al. Insula→Amygdala and Insula→Thalamus Pathways Are Involved in Comorbid Chronic Pain and Depression-Like Behavior in Mice. *J Neurosci*. 2024;44 https://doi.org/10.1523/JNEUROSCI.2062-23.2024.

[22] Mazzitelli M, Ponomareva O, Presto P, John J, Neugebauer V. Impaired amygdala astrocytic signaling worsens neuropathic pain-associated neuronal functions and behaviors. *Front Pharmacol*. 2024;15 https://doi.org/10.3389/fphar.2024.1368634.

[23] Singh S, Wilson TD, Valdivia S, Benowitz B, Chaudhry S, Ma J, et al. An inhibitory circuit from central amygdala to zona incerta drives pain-related behaviors in mice. *eLife*. n.d.;11:e68760 https://doi.org/10.7554/eLife.68760.

[24] Shen C-L, Santos JM, Elmassry MM, Bhakta V, Driver Z, Ji G, et al. Ginger Polyphenols Reverse Molecular Signature of Amygdala Neuroimmune Signaling and Modulate Microbiome in Male Rats with Neuropathic Pain: Evidence for Microbiota–Gut–Brain Axis. *Antioxidants*. 2024;13:502 https://doi.org/10.3390/antiox13050502.

[25] Söderlind M, Ask K, Leclercq A, Åkesson T, Valarcher J-F, Hernlund E, et al. Evaluation of a composite pain scale including facial expressions for detecting orthopedic pain in lame dairy cows. *J Dairy Sci*. 2025;108:5153–69 https://doi.org/10.3168/jds.2024-25692.

[26] Aulehner K, Leenaars C, Buchecker V, Stirling H, Schönhoff K, King H, et al. Grimace scale, burrowing, and nest building for the assessment of post-surgical pain in mice and rats—A systematic review. *Front Vet Sci*. 2022;9 https://doi.org/10.3389/fvets.2022.930005.

[27] Arnold B, Ramakrishnan R, Wright A, Wilson K, VandeVord PJ. An automated rat grimace scale for the assessment of pain. *Sci Rep*. 2023;13:18859 https://doi.org/10.1038/s41598-023-46123-x.

[28] Zhang EQ, Leung VS, Pang DS. Influence of Rater Training on Inter- and Intrarater Reliability When Using the Rat Grimace Scale 2019; https://doi.org/10.30802/AALAS-JAALAS-18-000044.

[29] Monteiro BP, Lee NHY, Steagall PV. Can cat caregivers reliably assess acute pain in cats using the Feline Grimace Scale? A large bilingual global survey. *J Feline Med Surg*. 2023;25 https://doi.org/10.1177/1098612X221145499.

[30] Robinson AR, Steagall PV. Effects of training on Feline Grimace Scale scoring for acute pain assessment in cats. *J Feline Med Surg*. 2024;26 https://doi.org/10.1177/1098612X241275284.

[31] 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 https://doi.org/10.1038/s41598-024-78406-2.

[32] Ask K, Andersen PH, Tamminen L-M, Rhodin M, Hernlund E. Performance of four equine pain scales and their association to movement asymmetry in horses with induced orthopedic pain. *Front Vet Sci*. 2022;9 https://doi.org/10.3389/fvets.2022.938022.

[33] Flores JEM, Terrazas A, Lara Sagahon AV, Aleman M. Parasympathetic tone activity, heart rate, and grimace scale in conscious horses of 3 breeds before, during, and after nociceptive mechanical stimulation. *J Vet Intern Med*. 2024;38:2739–47 https://doi.org/10.1111/jvim.17174.

[34] Tomacheuski RM, Oliveira AR, Trindade PHE, Oliveira FA, Candido CP, Teixeira Neto FJ, et al. Reliability and Validity of UNESP-Botucatu Cattle Pain Scale and Cow Pain Scale in Bos taurus and Bos indicus Bulls to Assess Postoperative Pain of Surgical Orchiectomy. *Anim Open Access J MDPI*. 2023;13:364 https://doi.org/10.3390/ani13030364.

[35] Tomacheuski RM, Klostermann C, Frank D, Taffarel MO, Pinho RH, Monteiro BP, et al. Bovine pain scale: A novel tool for pain assessment in cattle undergoing surgery in the hospital setting. *PLOS One*. 2025;20:e0323710 https://doi.org/10.1371/journal.pone.0323710.

[36] Farghal M, Pajor E, Luna SPL, Pang D, Windeyer MC, Ceballos MC. Development of the calf grimace scale for pain and stress assessment in castrated Angus beef calves. *Sci Rep*. 2024;14:25620 https://doi.org/10.1038/s41598-024-77147-6.

[37] Farghal M. Pain Assessment in Cattle during Castration Using Facial Expressions as a Promising Tool. *Anim Behav Welf Cases*. 2024;2024:abwcases20240008 https://doi.org/10.1079/abwcases.2024.0008.

[38] Ginger L, Aubé L, Ledoux D, Borot M, David C, Bouchon M, et al. A six-step process to explore facial expressions performances to detect pain in dairy cows with lipopolysaccharide-induced clinical mastitis. *Appl Anim Behav Sci*. 2023;264 https://doi.org/10.1016/j.applanim.2023.105951.

[39] Fuentes S, Gonzalez Viejo C, Tongson E, Lipovetzky N, Dunshea FR. Biometric Physiological Responses from Dairy Cows Measured by Visible Remote Sensing Are Good Predictors of Milk Productivity and Quality through Artificial Intelligence. *Sensors*. 2021;21:6844 https://doi.org/10.3390/s21206844.

[40] Sidwell AE, Duz M, Hill B, Freeman S, Hole SL. Application of the horse grimace scale in horses with dental disease: Preliminary findings. *Vet Rec*. 2025;196 https://doi.org/10.1002/vetr.4800.

[41] Paterson EA, O'Malley CI, Moody C, Vogel S, Authier S, Turner PV. Development and validation of a cynomolgus macaque grimace scale for acute pain assessment. *Sci Rep*. 2023;13:3209 https://doi.org/10.1038/s41598-023-30380-x.

[42] Oliveira MGC de, Paula VV de, Mouta AN, Lima I de O, Macêdo LB de, Nunes TL, et al. Validation of the Donkey Pain Scale (DOPS) for Assessing Postoperative Pain in Donkeys. *Front Vet Sci*. 2021;8 https://doi.org/10.3389/fvets.2021.671330.

[43] Tomacheuski RM, de Oliveira AR, Trindade PHE, Lopez-Soriano M, Merenda VR, Luna SPL, et al. Real-time and video-recorded pain assessment in beef cattle: clinical application and reliability in young, adult bulls undergoing surgical castration. *Sci Rep*. 2024;14 https://doi.org/10.1038/s41598-024-65890-9.

[44] Dalla Costa E, Minero M, Lebelt D, Stucke D, Canali E, Leach MC. Development of the Horse Grimace Scale (HGS) as a Pain Assessment Tool in Horses Undergoing Routine Castration. *PLoS ONE*. 2014;9:e92281 https://doi.org/10.1371/journal.pone.0092281.

[45] Evangelista MC, Monteiro BP, Steagall PV. Measurement properties of grimace scales for pain assessment in nonhuman mammals: a systematic review. *Pain*. 2022;163:e697–714 https://doi.org/10.1097/j.pain.0000000000002474.

[46] Cheng AJ, Malo A, Garbin M, Monteiro BP, Steagall PV. Construct validity, responsiveness and reliability of the Feline Grimace Scale in kittens. *J Feline Med Surg*. 2023;25 https://doi.org/10.1177/1098612X231211765.

[47] Dhaliwal Y, Bi H, Neethirajan S. Bimodal data analysis for early detection of lameness in dairy cows using artificial intelligence. *J Agric Food Res*. 2025;21 https://doi.org/10.1016/j.jafr.2025.101837.

[48] Mota-Rojas D, Whittaker AL, Bienboire-Frosini C, Buenhombre J, Mora-Medina P, Domínguez-Oliva A, et al. The neurobiological basis of emotions and their connection to facial expressions in non-human mammals: insights in nonverbal communication. *Front Vet Sci*. 2025;12 https://doi.org/10.3389/fvets.2025.1541615.

[49] Dac HH, Gonzalez Viejo C, Lipovetzky N, Tongson E, Dunshea FR, Fuentes S. Livestock Identification Using Deep Learning for Traceability. *Sensors*. 2022;22:8256 https://doi.org/10.3390/s22218256.

[50] Feighelstein M, Ehrlich Y, Naftaly L, Alpin M, Nadir S, Shimshoni I, et al. Deep learning for video-based automated pain recognition in rabbits. *Sci Rep*. 2023;13:14679 https://doi.org/10.1038/s41598-023-41774-2.

[51] Coffman E, Clark R, Bui N-T, Pham TT, Kegley B, Powell JG, et al. CattleFace-RGBT: RGB-T Cattle Facial Landmark Benchmark 2024; https://doi.org/10.48550/arXiv.2406.03431.

[52] Mao Y, Liu Y. Pet dog facial expression recognition based on convolutional neural network and improved whale optimization algorithm. *Sci Rep*. 2023;13 https://doi.org/10.1038/s41598-023-30442-0.

[53] Bergman N, Yitzhaky Y, Halachmi I. Biometric identification of dairy cows via real-time facial recognition. *Anim Int J Anim Biosci*. 2024;18:101079 https://doi.org/10.1016/j.animal.2024.101079.

[54] Meng H, Zhang L, Yang F, Hai L, Wei Y, Zhu L, et al. Livestock Biometrics Identification Using Computer Vision Approaches: A Review. *Agriculture*. 2025;15:102 https://doi.org/10.3390/agriculture15010102.

[55] Chen Y-S, Rustia DJA, Huang S-Z, Hsu J-T, Lin T-T. IoT-based system for individual dairy cow feeding behavior monitoring using cow face recognition and edge computing. *Internet Things*. 2025;33:101674 https://doi.org/10.1016/j.iot.2025.101674.

[56] Zhang X, Xuan C, Ma Y, Su H. A high-precision facial recognition method for small-tailed Han sheep based on an optimised Vision Transformer. *Animal*. 2023;17 https://doi.org/10.1016/j.animal.2023.100886.

[57] Feighelstein M, Luna SP, Silva NO, Trindade PE, Shimshoni I, van der Linden D, et al. Comparison between AI and human expert performance in acute pain assessment in sheep. *Sci Rep*. 2025;15:626 https://doi.org/10.1038/s41598-024-83950-y.

[58] Lu J, Wang W, Zhao K, Wang H. Recognition and segmentation of individual pigs based on Swin Transformer. *Anim Genet*. 2022;53:794–802 https://doi.org/10.1111/age.13259.

[59] Gao G, Ma Y, Wang J, Li Z, Wang Y, Bai H. CFR-YOLO: A Novel Cow Face Detection Network Based on YOLOv7 Improvement. *Sensors*. 2025;25:1084 https://doi.org/10.3390/s25041084.

[60] Araújo VM, Rili I, Gisiger T, Gambs S, Vasseur E, Cellier M, et al. AI-powered cow detection in complex farm environments. *Smart Agric Technol*. 2025;10:100770 https://doi.org/10.1016/j.atech.2025.100770.

[61] Qin Q, Zhou X, Gao J, Wang Z, Naer A, Hai L, et al. YOLOv8-CBAM: a study of sheep head identification in Ujumqin sheep. *Front Vet Sci*. 2025;12:1514212 https://doi.org/10.3389/fvets.2025.1514212.

[62] Gao G, Ma Y, Wang J, Li Z, Wang Y, Bai H. CFR-YOLO: A Novel Cow Face Detection Network Based on YOLOv7 Improvement. *Sensors*. 2025;25:1084 https://doi.org/10.3390/s25041084.

[63] Dang J, Tang X, Li S. HA-FPN: Hierarchical Attention Feature Pyramid Network for Object Detection. *Sensors*. 2023;23:4508 https://doi.org/10.3390/s23094508.

[64] Yang X, Guo R, Li H. Comparison of multimodal RGB-thermal fusion techniques for exterior wall multi-defect detection. *J Infrastruct Intell Resil*. 2023;2:100029 https://doi.org/10.1016/j.iintel.2023.100029.

[65] Benato L, Murrell J, Knowles TG, Rooney NJ. Development of the Bristol Rabbit Pain Scale (BRPS): A multidimensional composite pain scale specific to rabbits (Oryctolagus cuniculus). *PLOS ONE*. 2021;16:e0252417 https://doi.org/10.1371/journal.pone.0252417.

[66] Romano A, De Camillis A, Sciota D, Baghini S, Di Provvido A, Rosamilia A, et al. Cross-species AI: shifting a convolutional neural network from pigs to lambs to detect pneumonia at slaughter. *Front Vet Sci*. 2025;12:1591032 https://doi.org/10.3389/fvets.2025.1591032.

[67] 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 https://doi.org/10.3389/fvets.2024.1436795.

[68] Dwivedi A, Henige M, Anklam K, Döpfer D. Real-time digital dermatitis detection in dairy cows on Android and iOS apps using computer vision techniques. *Transl Anim Sci*. 2025;9:txae168 https://doi.org/10.1093/tas/txae168.

[69] Mahato S, Neethirajan S. Integrating Artificial Intelligence in dairy farm management − biometric facial recognition for cows. *Inf Process Agric*. 2024; https://doi.org/10.1016/j.inpa.2024.10.001.

[70] Mahato S, Neethirajan S. Dairy DigiD: An Edge-Cloud Framework for Real-Time Cattle Biometrics and Health Classification. *AI*. 2025;6:196 https://doi.org/10.3390/ai6090196.

[71] Liu F, Zhang Y, Liu Y, Li J, Li M, Yao J. A Novel Lightweight Dairy Cattle Body Condition Scoring Model for Edge Devices Based on Tail Features and Attention Mechanisms. *Vet Sci*. 2025;12:906 https://doi.org/10.3390/vetsci12090906.

[72] Feng X, Zhang J, Qi Y, Liu L, Li Y. CATR-Net: Cattle–Attentive transformer with adaptive and enhanced segmentation and recognition. *Comput Electron Agric*. 2025;239:111038 https://doi.org/10.1016/j.compag.2025.111038.

[73] Yu W, Yang X, Liu Y, Xuan C, Xie R, Wang C. SMEA-YOLOv8n: A Sheep Facial Expression Recognition Method Based on an Improved YOLOv8n Model. *Animals*. 2024;14 https://doi.org/10.3390/ani14233415.

[74] 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*. 2025;15:2508 https://doi.org/10.3390/ani15172508.

[75] Feighelstein M, Henze L, Meller S, Shimshoni I, Hermoni B, Berko M, et al. Explainable automated pain recognition in cats. *Sci Rep*. 2023;13:8973 https://doi.org/10.1038/s41598-023-35846-6.

[76] 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;14:28006 https://doi.org/10.1038/s41598-024-78406-2.

[77] Steagall PV, Monteiro BP, Marangoni S, Moussa M, Sautié M. Fully automated deep learning models with smartphone applicability for prediction of pain using the Feline Grimace Scale. *Sci Rep*. 2023;13:21584 https://doi.org/10.1038/s41598-023-49031-2.

[78] Morozov A, Parr LA, Gothard K, Paz R, Pryluk R. Automatic Recognition of Macaque Facial Expressions for Detection of Affective States. *eNeuro*. 2021;8:ENEURO.0117-21.2021 https://doi.org/10.1523/ENEURO.0117-21.2021.

[79] Gris VN, Crespo TR, Kaneko A, Okamoto M, Suzuki J, Teramae J, et al. Deep Learning for Face Detection and Pain Assessment in Japanese Macaques (Macaca fuscata). *J Am Assoc Lab Anim Sci JAALAS*. 2024;63:403 https://doi.org/10.30802/AALAS-JAALAS-23-000056.

[80] Gris VN, Broche N, Kaneko A, Okamoto M, Suzuki J, Mills DS, et al. Investigating subtle changes in facial expression to assess acute pain in Japanese macaques. *Sci Rep*. 2022;12:19675 https://doi.org/10.1038/s41598-022-23595-x.

[81] Arnold B, Ramakrishnan R, Wright A, Wilson K, VandeVord PJ. An automated rat grimace scale for the assessment of pain. *Sci Rep*. 2023;13:18859 https://doi.org/10.1038/s41598-023-46123-x.

[82] Boneh-Shitrit T, Feighelstein M, Bremhorst A, Amir S, Distelfeld T, Dassa Y, et al. Explainable automated recognition of emotional states from canine facial expressions: the case of positive anticipation and frustration. *Sci Rep*. 2022;12:22611 https://doi.org/10.1038/s41598-022-27079-w.

[83] Martvel G, Eretová P, Přibylová L, Chaloupková H, Pongrácz P, Shimshoni I, et al. Continuous automated analysis of facial dynamics of brachycephalic and normocephalic dogs in different contexts. *BMC Vet Res*. 2025;21:372 https://doi.org/10.1186/s12917-025-04839-0.

[84] 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;14:28006 https://doi.org/10.1038/s41598-024-78406-2.

[85] Zhang S, Sailunaz K, Neethirajan S. Micro-Expression-Based Facial Analysis for Automated Pain Recognition in Dairy Cattle: An Early-Stage Evaluation. *AI*. 2025;6:199 https://doi.org/10.3390/ai6090199.

[86] Neupane R, Aryal A, Haeussermann A, Hartung E, Pinedo P, Paudyal S. Evaluating machine learning algorithms to predict lameness in dairy cattle. *PLOS ONE*. 2024;19:e0301167 https://doi.org/10.1371/journal.pone.0301167.

[87] Jia Z, Zhao Y, Mu X, Liu D, Wang Z, Yao J, et al. Intelligent Deep Learning and Keypoint Tracking-Based Detection of Lameness in Dairy Cows. *Vet Sci*. 2025;12:218 https://doi.org/10.3390/vetsci12030218.

[88] Russello H, van der Tol R, Holzhauer M, van Henten EJ, Kootstra G. Video-based automatic lameness detection of dairy cows using pose estimation and multiple locomotion traits. *Comput Electron Agric*. 2024;223:109040 https://doi.org/10.1016/j.compag.2024.109040.

[89] Chiavaccini L, Gupta A, Chiavaccini G. From facial expressions to algorithms: a narrative review of animal pain recognition technologies. *Front Vet Sci*. 2024;11 https://doi.org/10.3389/fvets.2024.1436795.

[90] de Oliveira FA, Luna SPL, do Amaral JB, Rodrigues KA, Sant'Anna AC, Daolio M, et al. Validation of the UNESP-Botucatu unidimensional composite pain scale for assessing postoperative pain in cattle. *BMC Vet Res*. 2014;10:200 https://doi.org/10.1186/s12917-014-0200-0.

[91] Anagnostopoulos A, Griffiths BE, Siachos N, Neary J, Smith RF, Oikonomou G. Initial validation of an intelligent video surveillance system for automatic detection of dairy cattle lameness. *Front Vet Sci*. 2023;10 https://doi.org/10.3389/fvets.2023.1111057.

[92] Ferdinandy B, Gerencsér L, Corrieri L, Perez P, Újváry D, Csizmadia G, et al. Challenges of machine learning model validation using correlated behaviour data: Evaluation of cross-validation strategies and accuracy measures. *PLoS ONE*. 2020;15:e0236092 https://doi.org/10.1371/journal.pone.0236092.

[93] Riaz MU, O'Grady L, McAloon CG, Logan F, Gormley IC. Comparison of machine learning and validation methods for high-dimensional accelerometer data to detect foot lesions in dairy cattle. *PLOS ONE*. 2025;20:e0325927 https://doi.org/10.1371/journal.pone.0325927.

[94] Coelho Ribeiro LA, Bresolin T, Rosa GJ de M, Rume Casagrande D, Danes M de AC, Dórea JRR. Disentangling data dependency using cross-validation strategies to evaluate prediction quality of cattle grazing activities using machine learning algorithms and wearable sensor data. *J Anim Sci*. 2021;99:skab206 https://doi.org/10.1093/jas/skab206.

[95] Gutierrez R, Garcia-Ortiz J, Villegas-Ch W. Multimodal AI techniques for pain detection: integrating facial gesture and paralanguage analysis. *Front Comput Sci*. 2024;6 https://doi.org/10.3389/fcomp.2024.1424935.

[96] Alshehri H, Al-Nafjan A, Aldayel M. Decoding Pain: A Comprehensive Review of Computational Intelligence Methods in Electroencephalography-Based Brain–Computer Interfaces. *Diagnostics*. 2025;15:300 https://doi.org/10.3390/diagnostics15030300.

[97] Feighelstein M, Shimshoni I, Finka LR, Luna SPL, Mills DS, Zamansky A. Automated recognition of pain in cats. *Sci Rep*. 2022;12 https://doi.org/10.1038/s41598-022-13348-1.

[98] Ledoux D, Veissier I, Meunier B, Gelin V, Richard C, Kiefer H, et al. Combining accelerometers and direct visual observations to detect sickness and pain in cows of different ages submitted to systemic inflammation. *Sci Rep*. 2023;13:1977 https://doi.org/10.1038/s41598-023-27884-x.

[99] Wang J, Yu Z, Chebel RC, Yu H. Impact of cross-validation designs on cattle behavior prediction using machine learning and deep learning models with tri-axial accelerometer data 2025;2025.01.22.634181 https://doi.org/10.1101/2025.01.22.634181.

[100] Arablouei R, Wang Z, Bishop-Hurley GJ, Liu J. Multimodal sensor data fusion for in-situ classification of animal behavior using accelerometry and GNSS data 2022; https://doi.org/10.48550/arXiv.2206.12078.

[101] Pound P, Ritskes-Hoitinga M. Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail. *J Transl Med*. 2018;16:304 https://doi.org/10.1186/s12967-018-1678-1.

[102] Graat EAM, Vanden Hole C, Rodenburg TB, Giersberg MF, Tuyttens FAM. Can a Digital Application for Animal Welfare Self-Assessments by Farmers Help Improve the Welfare of Free-Range and Organic Pigs? *Anim Open Access J MDPI*. 2024;14:3374 https://doi.org/10.3390/ani14233374.

[103] Ganesan LP, Krishnan S. Blockchain-Enabled, Nature-Inspired Federated Learning for Cattle Health Monitoring. *Informatics*. 2025;12:57 https://doi.org/10.3390/informatics12030057.

[104] Lamanna M, Bovo M, Cavallini D. Wearable Collar Technologies for Dairy Cows: A Systematized Review of the Current Applications and Future Innovations in Precision Livestock Farming. *Animals*. 2025;15:458 https://doi.org/10.3390/ani15030458.

[105] Sebald C, Treiber M, Eryilmaz E, Bernhardt H. Usability Testing of Novel IoT-Infused Digital Services on Farm Equipment Reveals Farmer's Requirements towards Future Human–Machine Interface Design Guidelines. *AgriEngineering*. 2024;6:1660–73 https://doi.org/10.3390/agriengineering6020095.

[106] Hazrati M, Dara R, Kaur J. On-Farm Data Security: Practical Recommendations for Securing Farm Data. *Front Sustain Food Syst*. 2022;6 https://doi.org/10.3389/fsufs.2022.884187.

[107] Graat EAM, Vanden Hole C, Rodenburg TB, Giersberg MF, Tuyttens FAM. Can a Digital Application for Animal Welfare Self-Assessments by Farmers Help Improve the Welfare of Free-Range and Organic Pigs? *Anim Open Access J MDPI*. 2024;14:3374 https://doi.org/10.3390/ani14233374.

[108] Doidge C, Frössling J, Dórea FC, Ordell A, Vidal G, Kaler J. Social and ethical implications of data and technology use on farms: a qualitative study of Swedish dairy and pig farmers. *Front Vet Sci*. 2023;10 https://doi.org/10.3389/fvets.2023.1171107.

[109] Feighelstein M, Shimshoni I, Finka LR, Luna SPL, Mills DS, Zamansky A. Automated recognition of pain in cats. *Sci Rep*. 2022;12:9575 https://doi.org/10.1038/s41598-022-13348-1.

[110] Lencioni GC, de Sousa RV, de Souza Sardinha EJ, Corrêa RR, Zanella AJ. Pain assessment in horses using automatic facial expression recognition through deep learning-based modeling. *PLoS ONE*. 2021;16 https://doi.org/10.1371/journal.pone.0258672.

[111] Feighelstein M, Ehrlich Y, Naftaly L, Alpin M, Nadir S, Shimshoni I, et al. Deep learning for video-based automated pain recognition in rabbits. *Sci Rep*. 2023;13:14679 https://doi.org/10.1038/s41598-023-41774-2.

[112] Neethirajan S. The Significance and Ethics of Digital Livestock Farming. *AgriEngineering*. 2023;5:488–505 https://doi.org/10.3390/agriengineering5010032.

[113] Neethirajan S. Agency in Livestock Farming—A Perspective on Human–Animal–Computer Interactions. *Hum-Centric Intell Syst*. 2025;5:291–307 https://doi.org/10.1007/s44230-025-00108-3.

[114] Neethirajan S, Scott S, Mancini C, Boivin X, Strand E. Human-computer interactions with farm animals—enhancing welfare through precision livestock farming and artificial intelligence. *Front Vet Sci*. 2024;11 https://doi.org/10.3389/fvets.2024.1490851.

[115] Schillings J, Bennett R, Rose DC. Animal welfare and other ethical implications of Precision Livestock Farming technology. *CABI Agric Biosci*. 2021;2:17 https://doi.org/10.1186/s43170-021-00037-8.

[116] Tun SC, Onizuka T, Tin P, Aikawa M, Kobayashi I, Zin TT. Revolutionizing Cow Welfare Monitoring: A Novel Top-View Perspective with Depth Camera-Based Lameness Classification. *J Imaging*. 2024;10:67 https://doi.org/10.3390/jimaging10030067.

[117] Ajibola G, Kilders, V, Erasmus MA. A peep into the future: artificial intelligence for on-farm poultry welfare monitoring. *Anim Front Rev Mag Anim Agric*. 2025;14:72–5 https://doi.org/10.1093/af/vfae031.

[118] Michielon A, Litta P, Bonelli F, Don G, Farisè S, Giannuzzi D, et al. Mind the Step: An Artificial Intelligence-Based Monitoring Platform for Animal Welfare. *Sensors*. 2024;24:8042 https://doi.org/10.3390/s24248042.

[119] Zurnawita, Prabowo C, Kurnia R, Elfitri I. Application of Internet of Things Technology in Cattle Farm Monitoring Systems. *2024 2nd Int Symp Inf Technol Digit Innov ISITDI*. 2024;247–52 https://doi.org/10.1109/ISITDI62380.2024.10796296.

[120] Umeh II, Umeh KC. Redefining the future of data processing with edge computing. *World J Adv Res Rev*. 2024;24:2865–77 https://doi.org/10.30574/wjarr.2024.24.2.3463.

[121] Perez Garcia CA, Bovo M, Torreggiani D, Tassinari P, Benni S. Indoor Temperature Forecasting in Livestock Buildings: A Data-Driven Approach. *Agriculture*. 2024;14:316 https://doi.org/10.3390/agriculture14020316.

[122] Sadrzadeh N, Foris B, Krahn J, Keyserlingk MAG von, Weary DM. Automated monitoring of brush use in dairy cattle. *PLOS ONE*. 2024;19:e0305671 https://doi.org/10.1371/journal.pone.0305671.

[123] Luo W, Zhang G, Yuan Q, Zhao Y, Chen H, Zhou J, et al. High-precision tracking and positioning for monitoring Holstein cattle. *PLOS ONE*. 2024;19:e0302277 https://doi.org/10.1371/journal.pone.0302277.

[124] de Jong IC, Ouweltjes W, Llonch P, Martin Valls GE, Ko H-L, Spoolder H, et al. A review of existing scientific literature on welfare assessment of farmed species applied in commercial practice: identification of strengths, weaknesses, and areas for further development. *Front Vet Sci*. 2025;12:1589462 https://doi.org/10.3389/fvets.2025.1589462.

[125] Stygar AH, Gómez Y, Berteselli GV, Dalla Costa E, Canali E, Niemi JK, et al. A Systematic Review on Commercially Available and Validated Sensor Technologies for Welfare Assessment of Dairy Cattle. *Front Vet Sci*. 2021;8 https://doi.org/10.3389/fvets.2021.634338.

Figures

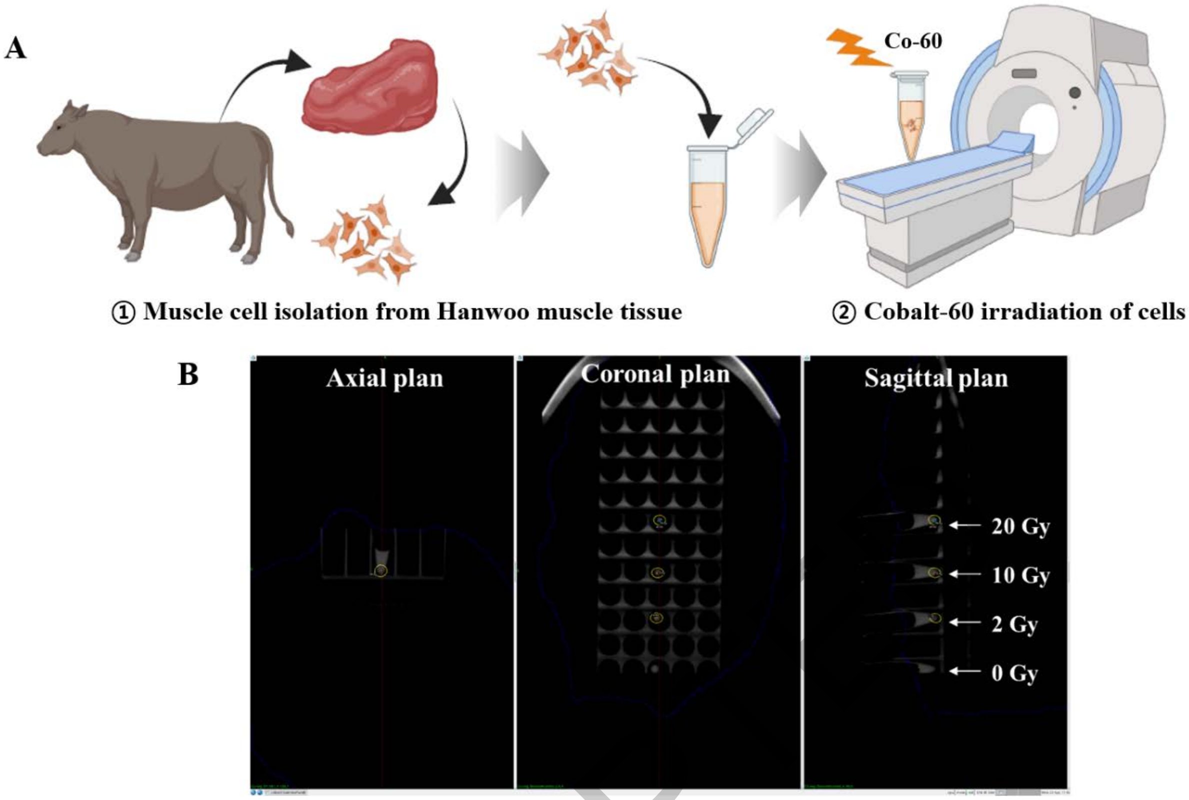

Figure 1. (A) Schematic illustration of Cobalt-60 gamma irradiation applied to primary muscle cells isolated from Hanwoo muscle tissue. (B) Representative axial, coronal, and sagittal plane images showing the targeted irradiation field using the Gamma Knife system.

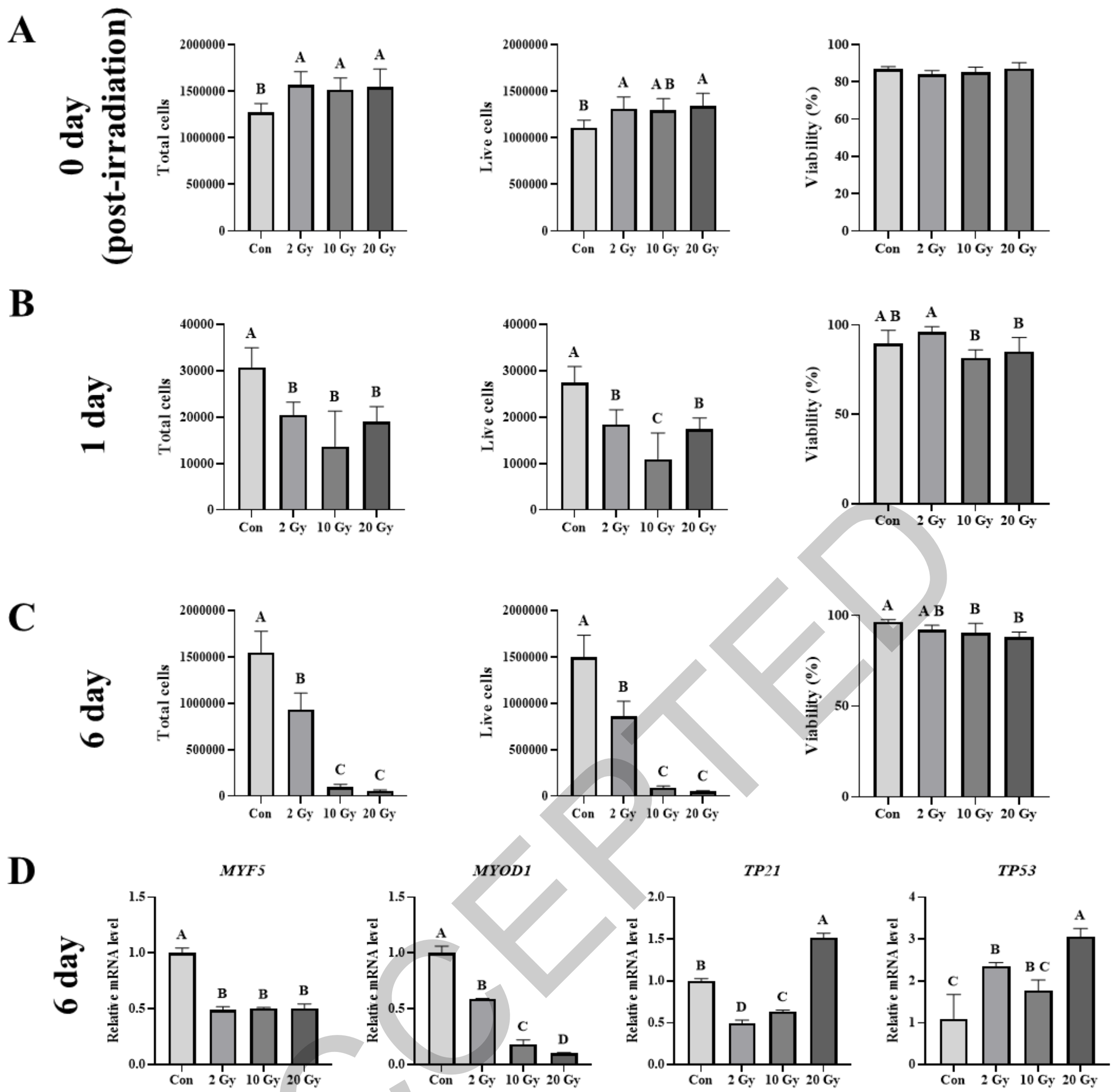

Figure 2. (A) Total cell number, live cell number, and viability of Hanwoo muscle-derived cells irradiated with different doses (Con, 2, 10, and 20 Gy) under suspension conditions. (B) Total cell number, live cell number, and viability of irradiated Hanwoo muscle-derived cells after 1 day of culture. (C) Total cell number, live cell number, viability of irradiated Hanwoo muscle-derived cells after 6 day of culture. (D) Relative mRNA expression levels of *MYF5, MYOD1*, *TP21*, and *TP53* in Hanwoo muscle-derived cells after 6 days of culture under each irradiation dose (Con, 2, 10, and 20 Gy). All mRNA expression levels were normalized to the housekeeping gene *GAPDH*. Different letters indicate statistically significant differences ($p <$ 0.05, one-way ANOVA followed by Tukey's post hoc test).

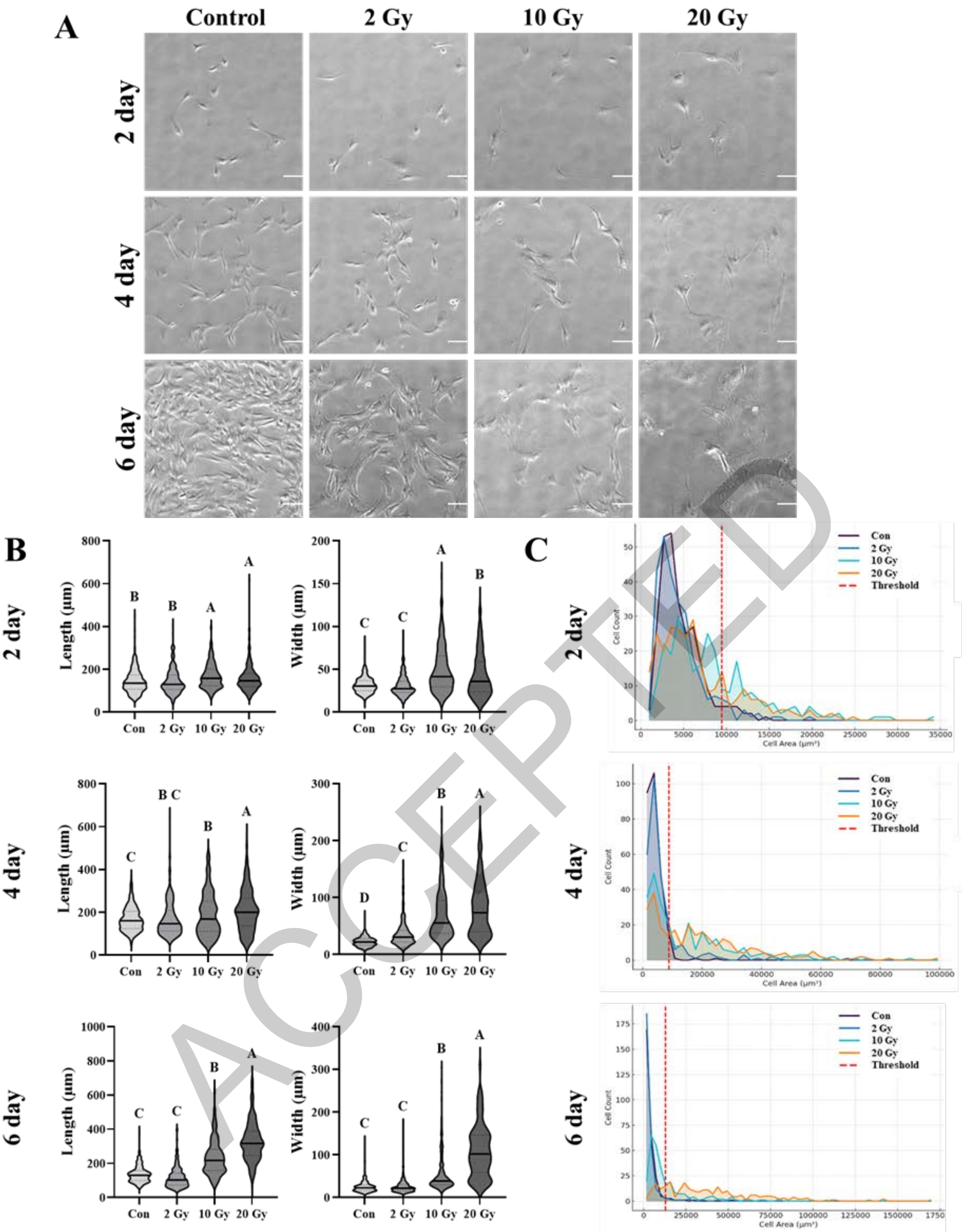

Figure 3. (A) Representative phase-contrast images of Hanwoo muscle-derived cells at 2, 4, and 6 days of culture following gamma irradiation at different doses (Con, 2, 10, and 20 Gy). Magnification: 40×, Scale bars = 100 μm. (B) Quantitative analysis of cell morphology showing cell length (μm) and width (μm) at 2, 4, and 6 days post-irradiation in each treatment group. (C) Distribution histograms of calculated cell area (length × width) at 2, 4, and 6 days of culture under each irradiation condition. Threshold values were defined as the mean + 2 standard deviations (2SD) of the control (Con) group at each time point. Different letters indicate statistically significant differences (*p* < 0.05, one-way ANOVA followed by Tukey's post hoc test).

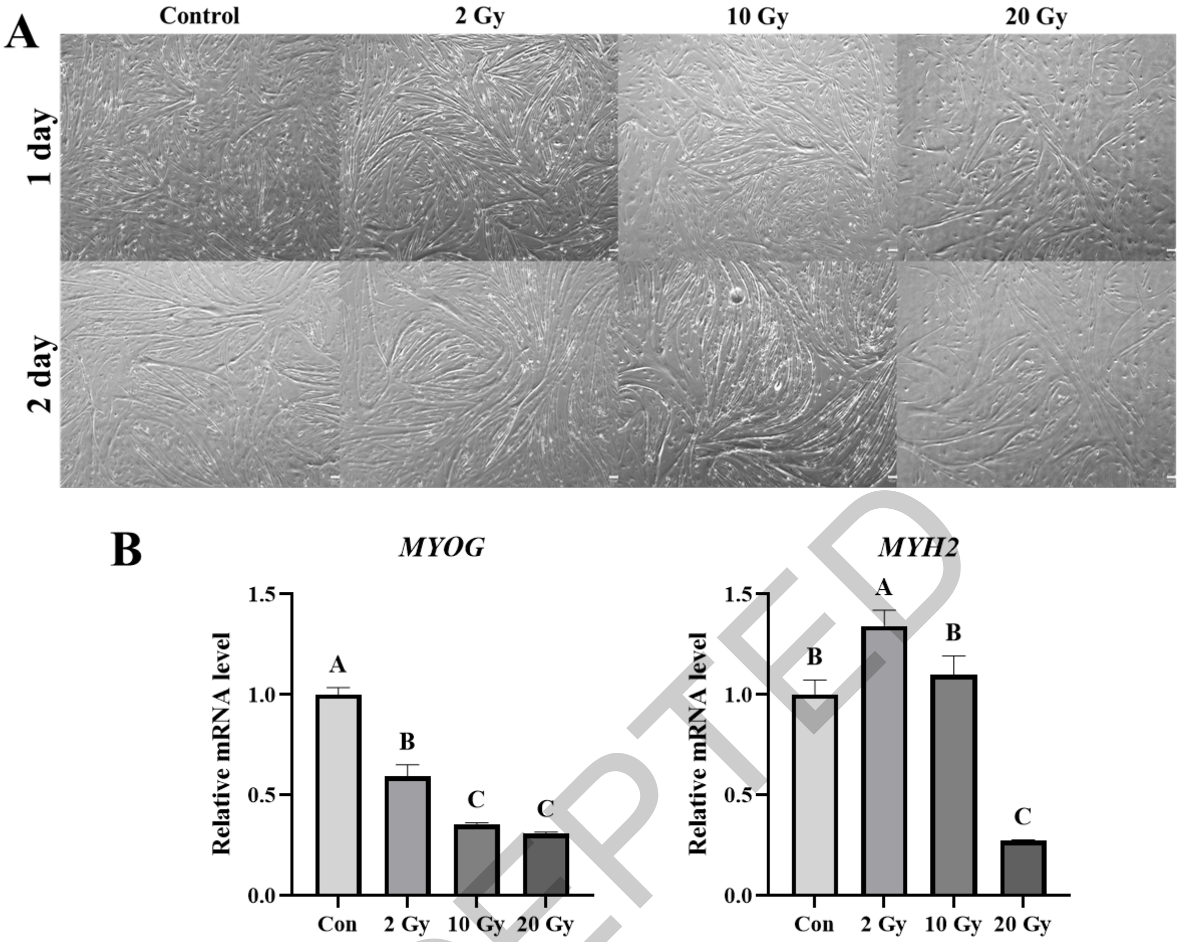

Figure 4. (A) Representative phase-contrast images of Hanwoo muscle-derived cells cultured under differentiation conditions for 1 and 2 days following gamma irradiation at different doses (Con, 2, 10, and 20 Gy). Magnification: 40×, Scale bars = 100 µm. (B) Relative mRNA expression levels of *MYOG* and *MYH2* in irradiated Hanwoo muscle-derived cells after 2 days of differentiation culture. All mRNA expression levels were normalized to the housekeeping gene *GAPDH*. Different letters indicate statistically significant differences ($p < 0.05$, one-way ANOVA followed by Tukey's post hoc test).