

Review

An Operation-Centered Review of Deep-Learning Computer Vision for Dairy Cow Management

Chi Zhang ¹, Ziwen Yu ^{1,*} and Jin Wang ²¹ Agricultural and Biological Engineering Department, University of Florida, Gainesville, FL 32603, USA; chi.zhang1@ufl.edu² Animal Sciences Department, University of Florida, Gainesville, FL 32608, USA; jin.wang@ufl.edu

* Correspondence: ziwen.yu@ufl.edu

Abstract

Dairy cow management depends on repeated observations of behavior and physical condition to support health, welfare, and operational decisions, but these observations remain labor-intensive. Deep learning (DL)-based computer vision can automate parts of this work, although deployment requirements differ among management operations. This operation-centered structured review retained 97 distinct publications. Five operations were compared: lameness detection, behavior recognition, individual identification and tracking, body condition score estimation, and body weight prediction. The comparison addresses management tasks, inputs, outputs, label requirements, validation designs, and deployment constraints. It identifies shared concerns involving reference-standard reliability, data-capture protocols, multi-animal scenes, identity continuity, cross-farm generalization, and data access. A qualitative three-part research agenda is proposed: candidate foundations in reference standards and identity continuity; transfer and data infrastructure; and conditional extensions involving multimodal and longitudinal evaluation.

Keywords: dairy farming; deep learning; computer vision; precision livestock farming; behavior recognition; lameness detection; individual identification

1. Introduction

Dairy farming is highly dependent on signals of cow locomotion, posture, feeding, drinking, lying behavior, and other daily activities to assess animal health, welfare, and management status. However, interpretation of these signals still depends largely on human observation, which restricts timely and appropriate intervention. To automate this process, advanced technologies in data collection and image processing, including high-resolution sensing and computer vision, have begun to transform animal signals into actionable information for management decisions [1,2]. Because these technologies are non-contact, compatible with existing barn infrastructure, and suitable for repeated monitoring [3,4], they have been applied in dairy farming to tasks such as individual identification and tracking, behavior monitoring, body condition assessment, lameness detection, mastitis-related observation, milk yield estimation, and behavior analysis [5–7].

From a technical perspective, these sensing applications have created data infrastructure for more advanced artificial intelligence (AI) technologies, especially deep learning (DL)-based methods, in dairy farming. AI is used here as the broader term for computational systems that support perception, prediction, or decision support, whereas DL refers more specifically to neural-network methods that are commonly embedded in AI systems



Academic Editor: Angela Salzano

Received: 31 May 2026

Revised: 16 July 2026

Accepted: 17 July 2026

Published: 19 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

for computer-vision applications. More broadly, deep learning has been applied across agricultural image analysis tasks [8].

In current computer-vision pipelines for dairy farming, common model families include CNN-based image classifiers such as ResNet [9] and transformer-based image classifiers such as ViT [10]. One-stage object detectors include YOLO variants [11–13] and SSD [14]. Instance-segmentation models include Mask R-CNN [15], while a dairy-specific Open Pose Mask R-CNN network has been proposed for individual cattle recognition [16]. General pose-estimation frameworks include OpenPose and DeepLabCut [17,18], which can be adapted to extract posture or movement features. Tracking models, including BoT-SORT, SORT, DeepSORT, and ByteTrack [19–22], maintain animal identity over time. Temporal models, such as CNN-LSTM/LRCN and 3D CNN, classify behaviors or locomotion patterns from video sequences [23,24].

The YOLO family [11–13] is highlighted as a representative detection family rather than as the only important model, because many dairy computer-vision systems require real-time cow localization before behavior recognition, individual tracking, mounting-event detection, or downstream phenotyping can be performed. Recent dairy and cattle studies illustrate this operational linkage. Improved YOLO models have been used for standing, lying, walking, and mounting recognition; YOLO-BoT has combined cattle behavior recognition with multi-object tracking; and Cow-YOLO has targeted mounting-event detection [25–27]. Other model families also appear in dairy applications, including CNN/depth-image models for body condition score estimation, Mask R-CNN-based segmentation for body weight or identification pipelines, keypoint-based lameness analysis, and spatiotemporal behavior recognition models [23,28–31]. Overall, algorithm choice is closely tied to operation: detection provides localization, segmentation provides body-region extraction, tracking provides identity continuity, and temporal or pose-based models support behavior and lameness classification.

Existing reviews [32–34] have provided valuable summaries of DL in dairy and live-stock systems, while other reviews focus on operations such as lameness or behavior detection [35,36]. The present review adds an operation-centered comparison of management purposes, evidence conditions, and deployment constraints. Its intended contribution is a qualitative synthesis of how recurring barriers relate across operations, not an estimate of field-wide prevalence or combine effect sizes.

Accordingly, this operation-informed structured narrative review has three objectives: (i) to organize the retained literature around animal-level dairy-management operations; (ii) to compare the operations by management task, required input, targeted output, label requirement, validation design, and deployment constraint; and (iii) to translate the cross-operation comparison into a qualitative, deployment-oriented research agenda.

2. Materials and Methods

2.1. Review Scope and Literature Organization

This article is an operation-informed structured narrative review. It is not a systematic review or meta-analysis, and no statistical pooling of effect sizes was undertaken. The review focuses on DL-based computer-vision applications at the individual-animal level in dairy cows. General-purpose AI/CV/ML method publications were retained solely to define technical context and were not treated as operation-specific dairy-management evidence. Non-vision sensing approaches, including wearable precision technologies and accelerometer-based behavior monitoring [37,38], as well as decision-support systems, robotics, and farm-management software, are acknowledged as complementary technologies but remain outside the operation-specific synthesis. Because the cross-operation

comparison and associated priorities are derived qualitatively, they are presented as a research agenda rather than as empirically validated conclusions or a quantitative ranking.

Following the record-identification process summarized in Figure 1, an operation-informed literature search was conducted across Web of Science, Scopus, ScienceDirect, and Google Scholar for English-language literature published from 2016 through 8 July 2026 on DL-based computer vision in dairy cow management. Full-text assessment yielded 97 publications. The final database-specific strings and the date of the retrieval-count recheck are reported in Appendix B Table A2. OpenAI tools assisted with candidate search-term expansion, database-syntax adaptation and language editing. The first and second authors jointly reviewed and classified all retained full texts and completed the detailed extraction used for the operation-level synthesis. The third author contributed to preliminary scoping and search development. The prompt provided to ChatGPT was: “I am preparing a literature review on artificial intelligence and deep learning applications in dairy-cow management. Please suggest additional English search terms and Boolean combinations related to dairy cows, computer vision, deep learning, artificial intelligence, monitoring, detection, recognition, estimation, prediction, tracking, and management. Format the terms for Web of Science, Scopus, ScienceDirect, and Google Scholar where their syntax differs. Do not select papers, judge eligibility, extract data, or make inclusion decisions.”

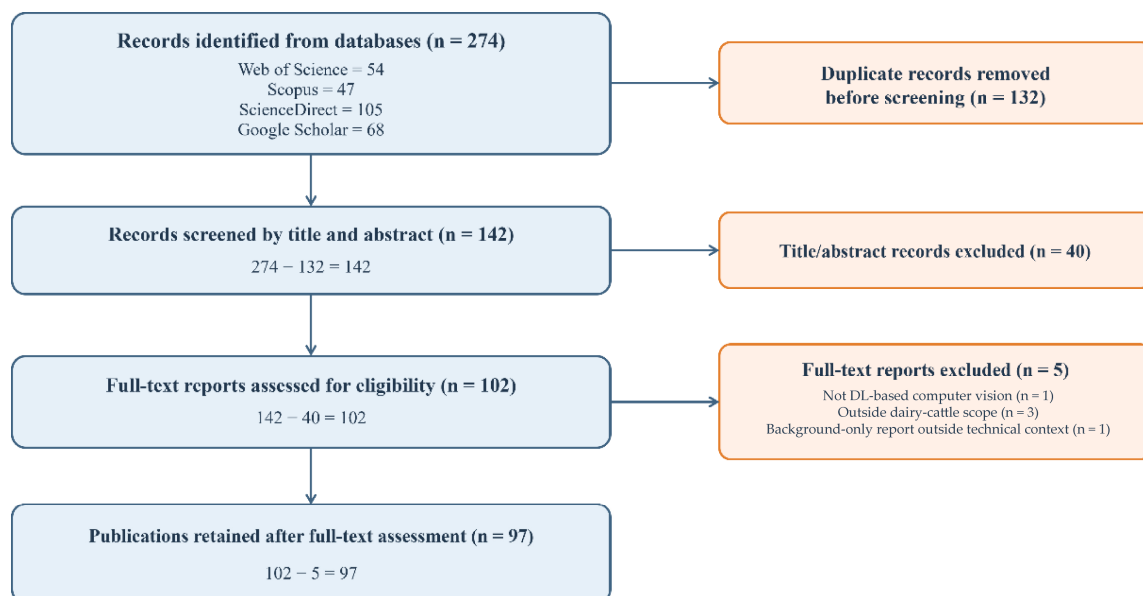


Figure 1. Record-identification and screening summary for the operation-informed structured narrative review. Counts describe retrieval and screening through 8 July 2026. Full-text assessment yielded 97 retained publications.

The AI-suggested candidate terms were manually checked, edited, and combined with author-defined terms. Search development was iterative and operation-informed: broad dairy, AI/DL, computer-vision, monitoring, and management concepts were refined with terms for the operations. The final strings in Appendix B Table A2 therefore document the search state used for the 8 July 2026 recheck rather than a preregistered systematic-review protocol. Operation labels structure the narrative synthesis, and their assignment counts are not interpreted as database-wide prevalence, popularity, practical importance, or economic value.

Database-specific syntax was adapted to the capabilities of each platform. Web of Science used TS fields, Scopus used TITLE-ABS-KEY fields, ScienceDirect used platform-compatible keyword combinations, and Google Scholar used simplified phrase combina-

tions. Appendix B Table A2 reports the final strings used for the 8 July 2026 retrieval-count recheck. Because the platforms differ in field tags and Boolean behavior, the queries are operationally documented but are not strictly equivalent across databases.

Retrieved records were screened in stages. Duplicate records were checked manually, primarily using DOI information, after which titles and abstracts were assessed for relevance. Full texts were then examined for either a clear contribution to the operation-specific synthesis or a necessary technical or contextual role. After full-text assessment, 97 publications were retained. Each retained publication was counted once.

Primary operation-specific evidence was eligible when it (i) was a peer-reviewed journal article or full peer-reviewed conference paper published between 2016 and 8 July 2026, (ii) concerned dairy cows, (iii) used DL-based image or video computer vision, and (iv) addressed animal-level monitoring, recognition, detection, estimation, or prediction function. Works were excluded from the primary operation-specific synthesis when their evidence was predominantly sensor-, biosensor-, or milk-record-based rather than vision-based. Operations supported by too little retained evidence for a comparable synthesis, including feeding and rumination monitoring and reproductive monitoring, could remain visible as secondary categories but were not developed as review sections. Figure 1 summarizes retrieval and screening, whereas Appendix A Table A1 is a manuscript citation map rather than a one-to-one inventory of the screened publications.

ChatGPT assisted only with candidate search-term expansion, database-syntax adaptation, and language editing. It did not retrieve or screen records, determine eligibility, retain or exclude works, extract study characteristics, assign operation categories, or establish research priorities. All tool outputs were reviewed and edited by the authors, who made the final search, screening, coding, and interpretive decisions. A Generative AI Disclosure is provided at the end of the paper.

2.2. Analysis Procedure

The narrative analysis organized the selected literature and supporting citations by farming operation. For the operation-specific synthesis, sources were characterized by management task, required input, targeted output, label or reference requirement, deployment constraint, validation design, and main reported metric. For counting purposes, each retained publication was assigned to one exclusive primary category based on its principal management objective.

Management task denotes the practical farming purpose addressed by DL. Required input denotes the sensing modality and capture conditions used to develop or evaluate a model. Targeted output denotes the predicted variable or class, including animal identity, behavior or locomotion class, body condition, and body weight. Label or reference requirement denotes the annotation or measurement used for training or evaluation. Deployment constraint denotes a practical condition that may limit the intended use, such as camera position, capture geometry, barn environment, occlusion, identity continuity, or controlled animal posture.

The first author coded all retained publication records, and the second author subsequently reviewed and cross-checked every record using the same predefined eligibility criteria and extraction fields. Any discrepancies concerning eligibility, extracted study characteristics, or operation-category assignments were resolved through discussion between the first two authors until consensus was reached. Because this procedure involved full-record cross-checking rather than independent duplicate coding before consensus, neither Cohen's kappa nor a pre-consensus percentage agreement was calculated. The third author contributed to the preliminary literature scoping and development of the search framework.

Table 1 summarized five reviewed DL operations in dairy farming and Table 2 aggregates the operation assignments descriptively. Because study definitions, metrics, and validation designs differ, the table does not provide a meta-analytic comparison or a validated deployment-readiness scale. Barriers were synthesized at the operation level rather than coded as binary variables for each individual study. Accordingly, Table 3 presents qualitative task-level relationships and does not report the number or proportion of studies affected by each barrier.

Table 1. Summary of the five reviewed DL operations in dairy farming.

DL Operation	Management Task	Input	Output	Label	Main Deployment Constraint
Lameness detection	Early welfare and health intervention for gait abnormality.	RGB video, pose/key points, depth or back-posture data.	Lameness class, mobility score, gait/posture alert.	Expert locomotion or mobility score.	Mild-case separation, scorer variation, walking-view setup, camera position.
Behavior recognition	Daily activity and time budget monitoring.	Continuous RGB or multi-view video.	Behavior class, event sequence, time-budget summary.	Frame-, clip-, or event-level behavior annotation.	Occlusion, overlapping cows, similar behaviors, rare events, long sequences.
Individual identification and tracking	Link observations to the correct cow over time.	RGB image or video including body, face, or muzzle regions, geometric cues.	Cow ID, re-ID match, tracking trajectory.	Individual ID and track labels across time.	Open herds, appearance change, dirt, lighting, overlap, herd turnover.
Body condition score estimation	Nutrition, energy balance, reproduction, and management.	Top or rear depth images, RGB-D, point clouds and rump/body region features.	BCS class or continuous score.	Expert BCS score or scorer consensus.	Controlled posture, camera placement, body-region segmentation, scorer inconsistency.
Body weight prediction	Growth, nutrition, production, and health planning.	RGB/depth/3D images, point clouds, body measurements.	Body weight estimate or growth trend.	Scale-measured body weight paired with image or time point.	Structured capture, clean body shape, accurate reference weight, limited data sharing.

The complete reference list contains 126 entries, including the retained publications and 29 additional references used for methodological framing or discussion.

Table 2. Descriptive summary of technical and validation characteristics, organized by operation. Here, *n* denotes the number of distinct publications assigned primarily to each operation. The five rows summarize 79 publications assigned to the five principal operations. Four additional operation-specific publications were assigned to reproductive monitoring or feeding and rumination monitoring, while 14 general AI/CV/ML publications were retained only for technical context. These 18 publications are documented in Appendix C Tables A3 and A4 but are not summarized in Table 2. Dataset ranges are the observed minima and maxima in the coded sources. Reported performance ranges are task-specific and are not directly comparable across operations; they may also not be directly comparable within an operation because outcome definitions, data splits, units of analysis, and validation designs differ. The final column is a provisional qualitative interpretation rather than a validated readiness score.

Operation (<i>n</i>)	Input	Model Families/Methods	Dataset Scale	Validation Design	External Validation	Metric	Provisional Deployment Interpretation
Lameness detection (<i>n</i> = 18)	RGB-only video (12), RGB-D video (3), depth/3D-only video or images (3); often with pose keypoints	Pose keypoints with a classifier, YOLO detection, spatiotemporal CNN or LSTM	22 to at least 2515 cows; up to 40,116 paired scores; 1 to 11 farms	Training/test split, k-fold, grouped validation, or external multi-farm comparison	1 direct; 4 partial/unclear or multi-setting components (5 of 18 with any cross-setting evidence)	AP/mAP approximately 64.1–99.4%	Feasibility shown mainly under controlled walking views; transfer across farms and scoring protocols remains uncertain.
Behavior recognition (<i>n</i> = 22)	RGB or infrared-capable video/images (22); no non-visual sensor inputs	YOLO detection with tracking, CNN-LSTM or 3D-CNN, vision-language or temporal models	15 to 107 cows among studies reporting a complete cohort; more than 1.7 million still images or up to 206,100 video frames; 1 to 7 farms/sites	Mostly image-, frame-, or clip-level train/validation/test splits; occasional day/location or later-video tests	0 held-out-farm tests; 3 pooled multi-farm/multi-site datasets without site-held-out validation	Accuracy or mAP approximately 75 to 99%	Feasibility shown in RGB video; occlusion, rare events, and individual-level continuity remain limiting.

Table 2. Cont.

Operation (<i>n</i>)	Input	Model Families/Methods	Dataset Scale	Validation Design	External Validation	Metric	Provisional Deployment Interpretation
Individual identification and tracking (<i>n</i> = 18)	RGB face, muzzle, body, rear, or ear-tag imagery (17); depth/3D (1)	Mask R-CNN, lightweight CNNs, face/muzzle encoders, contrastive or few-shot learning, and trackers	13 to 686 cows among studies reporting a complete cohort; up to 33,019 identification images; mostly 1 farm	Fixed-roster image/video splits; several cow-disjoint or same-farm open-set tests; one cross-farm adapted tracking test	0 strict zero-shot external-farm tests; 1 cross-farm adapted test; 7 animal-disjoint/open-set evaluations	Identification accuracy approximately 83.6 to 99.7%;	Performance is strongest for fixed rosters; open-set recognition and long-term identity stability remain uncertain.
Body condition score estimation (<i>n</i> = 12)	2D RGB (7); depth or 3D (5)	CNN classifiers/regressors; YOLOv5/8/11; Faster R-CNN; 3D point-cloud nets	56 to at least 3817 cows; up to 34,150 scored observations; 1 to 20 farms/sites	Image-level splits predominated; cow-level separation in 4 studies; held-out-farm tests in 2	2 held-out-farm tests; no other quantified external validation; qualitative second-farm examples were not counted	Within 0.25 BCS 69.89 to 91.21%; within 0.5 BCS 89.96 to 97.60%	Feasibility depends on controlled capture and a consistent scoring reference.
Body weight prediction (<i>n</i> = 9)	Depth or 3D (4), RGB or stereo (3), LiDAR or RGB-D fusion (2)	Segmentation with regression, point-cloud or 3D reconstruction, CNN regression	12 to 914 animals; up to 83,011 records; 1 to 3 farms	k-fold, leave-N-cows-out, cow-level split, or comparison with scale weight	No unseen-farm external validation was reported; three studies included multi-location or multi-herd data, but these were not evaluated as independent external test sets	MAPE 2.03–4.53%; MRE 1.49–1.75%; R ² 0.863–0.99; RMSE 2.49–33 kg	Feasibility depends on structured capture and calibrated reference weights; cross-farm transfer remains uncertain.

Table 3. Provisional operation-level mapping of cross-cutting barriers across five dairy DL operations. Entries describe task-level relationships and are not study counts, prevalence estimates, or a validated ranking. Direct indicates that a barrier is directly relevant to routine task evaluation or deployment; conditional indicates relevance to workflows or time horizons, and potential identifies a candidate methodological extension whose incremental value remains to be tested. The three agenda groups are hypothesis-generating.

Provisional Agenda Group	Barrier or Theme	Lameness Detection	Behavior Recognition	ID/Tracking	Body Condition Score	Body Weight	Interpretation
Candidate foundations	Reference-standard and annotation reliability	Direct	Direct	Direct	Direct	Direct	Reference labels or measurements require task-specific quality control.
Candidate foundations	Individual identity continuity	Conditional	Conditional	Direct	Conditional	Conditional	Applies when outputs must be linked to an individual across time.
Transfer and data infrastructure	Cross-farm generalization	Direct	Direct	Direct	Direct	Direct	Relevant when a model is used beyond its development setting.
Transfer and data infrastructure	Data access and interoperability	Direct	Direct	Direct	Direct	Direct	Affects representative training and independent external validation.
Conditional extensions	Multimodal integration	Potential	Potential	Potential	Potential	Potential	A candidate extension whose incremental value must be demonstrated.
Conditional extensions	Longitudinal, end-to-end evaluation	Conditional	Conditional	Direct	Conditional	Conditional	Depends on repeated measurements and stable identity linkage.

3. Results

3.1. Publication Trend

Figure 2 shows the publication-year distribution of the retained publications. The annual number of publications remained at approximately five from 2016 to 2021 and increased after 2022. This temporal association coincided with wider access to object-detection and tracking frameworks, but it does not establish that any framework caused the increase. Because the search was operation-informed and 2026 coverage ends on

8 July 2026, the distribution is descriptive of the publications retained in this review rather than an unbiased estimate of field-wide publication growth.

Publication timeline of retained papers

Bars: annual retained papers; line: cumulative retained papers; 2026 through 8 July

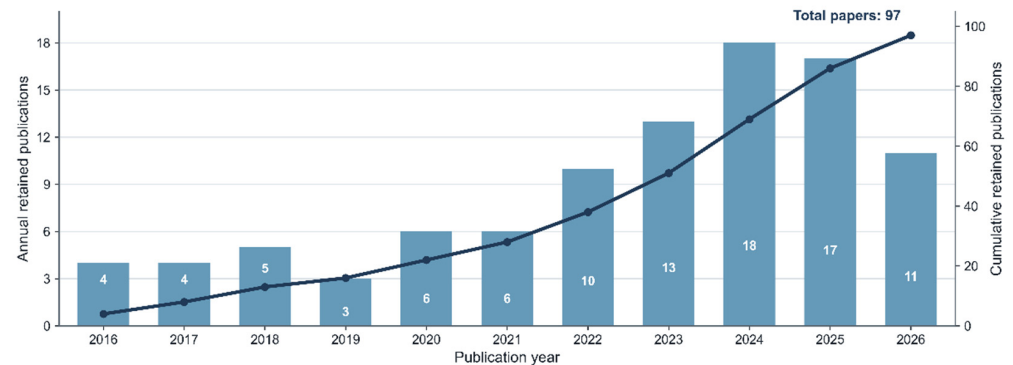


Figure 2. Publication-year distribution of the 97 publications retained after full-text assessment from 2016 through 8 July 2026. Bars show the annual number of retained publications, and the line shows the cumulative total. The 2026 interval is partial through 8 July.

3.2. Main Operation Identification

Figure 3 summarizes manuscript-level category assignments. Five operations were selected for synthesis: behavior recognition, lameness detection, individual identification and tracking, body condition score estimation, and body weight prediction. Their selection used three combined considerations: each represents a distinct animal-level management function, each has sufficient cited DL-based computer-vision evidence for cross-operation comparison, and together they span routine observation, identity continuity, and capture-constrained phenotyping. Assignment frequency describes coverage in this manuscript and is not evidence of practical or economic importance. Reproductive monitoring and feeding and rumination monitoring remain visible as secondary operation-specific categories but were not developed as sections because the cited evidence was too limited for comparable synthesis. The general AI/CV/ML method/background category supplies technical context and, for a different reason, is also excluded from the five-operation comparative synthesis. Accordingly, the grey bars in Figure 3 denote categories not included in that five-operation synthesis, but only the general AI/CV/ML category is context-only.

The rationale for synthesis is operation-informed rather than count-based. The five operations correspond to connected animal-level management decisions but differ in biological signal, label type, time scale, and acceptable error [32,39,40]. Behavior recognition describes what an animal is doing, whereas lameness detection evaluates whether gait or posture indicates abnormal locomotion. Identification and tracking provide identity continuity when observations must be linked to an animal over time. Body condition score estimation and body weight prediction add slower phenotypic signals related to welfare, nutrition, and production. Together, the operations support comparison across routine monitoring, identity continuity, and structured phenotyping; their assignment frequencies do not establish relative practical or economic importance.

Across the selected literature, several recurring methodological shifts can be identified, although they did not occur uniformly across all operations. Earlier work often combined engineered geometric, morphological, or gait features with classical learning methods. Later studies increasingly used CNN-based classifiers, real-time detectors, and multi-object trackers, followed by growing interest in pose estimation and transformer-based representations. These shifts coincided with wider access to annotated datasets. The

association is descriptive and does not establish that any single model family caused the post-2022 increase in retained publications.

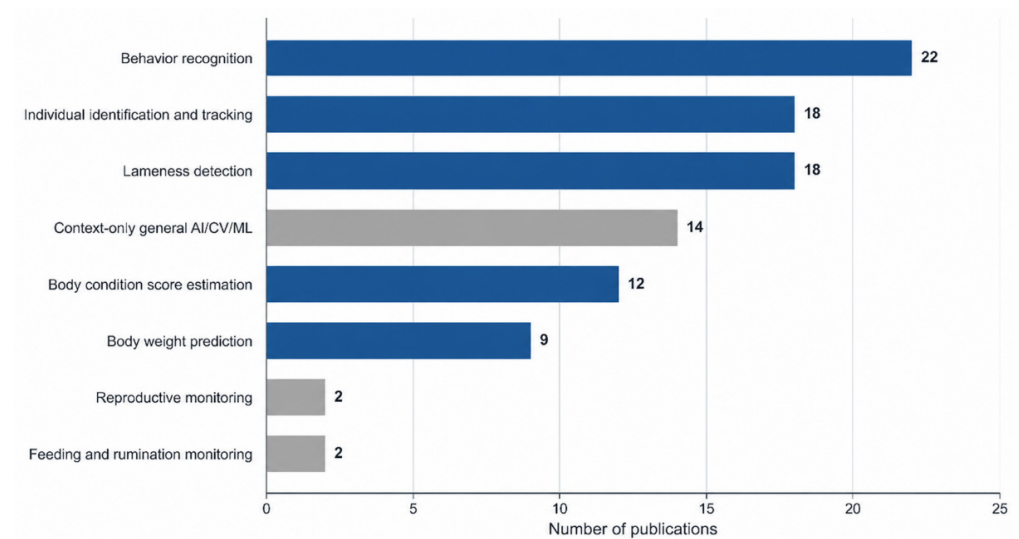


Figure 3. Distribution of the 97 retained publications by category. Blue bars represent the five principal operation categories included in the cross-operation synthesis. Grey bars represent categories not included in that five-operation synthesis: two secondary operation-specific categories with limited evidence and 14 general AI/CV/ML publications retained only for technical context. The two secondary categories remain operation-specific, only the general AI/CV/ML category is context-only. Each publication is counted once. Appendix A Table A1 provides the citation map, and Appendix C Tables A3 and A4 provides the corresponding study-level fields.

3.3. Deep Review of Lameness Detection and Behavior Recognition

Lameness detection and behavior recognition are examined in greater detail because both address repeated animal observation, have sufficiently broad cited evidence for within-operation synthesis, and illustrate different temporal and labeling requirements. This analytical emphasis does not imply that the other operations have lower practical value.

3.3.1. Lameness Detection

Lameness detection is an important dairy computer-vision application because locomotion abnormalities can inform welfare assessment, treatment timing, production management, and culling decisions. Visual scoring is labor-intensive and may vary among observers, which motivates automated support while also complicating the reference standard. The cited lameness literature includes 3D video and back-posture measurement, optical-flow or action-recognition networks, YOLO-based gait features, depth-camera top-view classification, and attention- or keypoint-based models [41–49].

Many lameness studies use controlled walking settings, such as alleys, exit lanes, or post-milking passages [50]. These settings make stride, back posture, limb movement, and gait asymmetry easier to observe, but they also bound the conditions under which performance is evaluated. The use of biologically interpretable gait signals gives lameness detection a clearer connection to biomechanics than many other dairy vision tasks. However, capture conditions are not standardized across farms, which limits transfer from a single controlled setting.

Methodologically, DL applications in recent lameness studies [29,50–52] can be grouped into three broad directions by different inputs and targeted outputs. The first direction uses pose or key point extraction followed by kinematic or postural features for classification. This route is common because lameness is closely related to visible changes (e.g., weight-bearing pattern), so a feature-based pipeline preserves a clearer link between

movement abnormality and the final decision [35]. The second direction uses multi-stream learning to combine appearance, motion, or skeleton information. This strategy is often adopted when subtle lameness-related changes are difficult to capture from a single visual cue alone [53]. The third direction moves toward end-to-end video classification, where the goal is to reduce manual feature design and simplify the inference pipeline for practical deployment [54]. These routes therefore reflect not only model preference, but also different ways of matching the learning strategy to the biological signal, the available data, and the intended use of the system.

However, the main limitations also remain across the literature. Detecting mild lameness remains a core challenge, as subtle gait changes are often difficult to separate from normal or borderline locomotion, particularly in early-stage cases where timely intervention would be most beneficial [55]. Label quality is another recurring concern, since ground truth annotations depend on subjective visual scoring by human observers, and inter-scorer agreement is often inconsistent, introducing noise into training data and undermining model reliability [56]. Beyond the data itself, environmental and practical factors also limit real-world performance. Variations in floor type, lighting conditions, camera angle, surface contamination such as dirt or manure, and differences in the animal's walking pace can all degrade detection accuracy, making it difficult to generalize models trained under controlled conditions to the diversity of on-farm settings [57]. These issues explain why strong results from one farm or one camera setup often do not transfer cleanly to another. For this reason, the future value of lameness detection may depend less on reporting another high within-dataset accuracy and more on improving cross-farm robustness, linking outputs to clinical action thresholds, and combining visual signals with other contextual information that improves practical use.

3.3.2. Behavior Recognition

Behavior analysis is a major application of DL-based vision in dairy systems, with targets including feeding, locomotion, social behavior, and estrus-related activity [23,36,58–61].

Compared with lameness detection, behavior recognition covers a broader range of operational targets, including feeding, drinking, lying, standing, walking, and mounting [23,26,60], as well as rumination [62] and respiratory behavior [63]. Such systems may support feeding-time analysis [58], estrus-related monitoring [61,64], and broader activity or welfare assessment [59,65]. Cited studies use CNN-, pose-, YOLO-, attention-, and spatiotemporal approaches for feeding, ingestive, daily, and multi-animal behavior recognition [66–77]. Long video streams available in barns make this application attractive, but availability of video does not by itself resolve annotation, occlusion, or identity limitations [78,79].

The cited studies generally follow two broad routes [23,27,59,64]. One route uses short video clips and spatiotemporal models that learn from motion over time. The other first detects or tracks cows and then classifies behavior from clips, shorter segments, or events. The appropriate route depends on whether the intended output is group-level or animal-level, whether identity continuity matters, and whether the target behavior is common or rare. Group summaries may not require persistent identity, whereas an individual time budget or rare-event history requires localization and association of the relevant animal.

Behavior recognition remains incomplete as a farm tool [59]. Visually similar or co-occurring behaviors can be confused; for example, standing and feeding may overlap from some camera angles [27]. Brief events also create class imbalance, as illustrated by mounting behavior [26]. Within-dataset accuracy may therefore overstate field performance for rare but management-relevant events. General methodological approaches that may be relevant include class re-weighting [80], loss [81], targeted augmentation or synthetic

and simulated video [82], few-shot learning [83], and anomaly-detection formulations [84]. These references provide candidate methods rather than dairy-specific evidence that the problem has been solved. Recent dairy studies continue to report sensitivity to occlusion, environmental variation, and single-farm evaluation [58,60,64]. Reviews also identify inconsistent benchmarks and the need for stronger temporal and contextual evaluation [36].

This point links behavior recognition to identification and tracking. A group-level behavior summary can be useful without individual identity, but a persistent record of one cow's behavior over days or weeks is more informative for animal-level management. Such histories may also be interpreted alongside slower phenotypic signals such as body condition or body weight. Behavior recognition can therefore function either as a stand-alone group-monitoring task or as part of a longitudinal individual-cow system.

3.4. Other Important Operations

The three remaining operations are also important for dairy management, although their deployment requirements differ from those of lameness detection and behavior recognition. Their treatment here is shorter because the review uses them primarily for cross-operation comparison, not because assignment counts determine farm value.

3.4.1. Individual Identification and Tracking

Identification and tracking constitute important supporting functions in dairy management because many decisions depend on repeated observations of the same animal [85,86]. Without stable identity, behavior histories can fragment, lameness alerts are harder to attribute, and longitudinal phenotyping becomes less reliable. The broader literature on livestock DL identifies individual recognition as a component of precision farming [39,87] and notes that long-term re-identification in real herds remains difficult [39,87]. Related studies have used muzzle images, facial or body appearance, multimodal identification, transfer learning, few-shot learning, and multi-cattle tracking to maintain identity across views and time [39,88–96].

Data collection remains challenging because commercial dairy herds are neither closed nor visually uniform environments [16,31]. Animals overlap, dirt changes visible appearance, lighting varies, and herd composition changes over time. A method developed for a fixed roster and a single camera view may lose reliability when new animals enter or appearance changes over months. Identification and tracking therefore function both as an operation and as a source of uncertainty for downstream animal-level longitudinal tasks [39,97].

3.4.2. Body Condition Score Estimation

Body condition score estimation has progressed steadily because it targets a phenotype that is visually tied to body shape, especially around the rump and hindquarter region [98,99]. This operation is important for nutrition, energy balance, reproduction, and transition-cow management. Recent DL studies have applied image- and video-based models to automated BCS estimation [100–102], including an improved YOLOv5 pipeline that detects and segments individual cows and assigns body condition scores from group-state video [100]. In many studies, depth images, top-view images, or point-cloud representations are used because geometry is more informative than color or texture alone [30,98,101–103]. Additional BCS studies have tested 2D or 3D imaging, regression-based scoring, score-class design, and annotation-region effects for automated body condition prediction [104–106].

Body condition score estimation has shown feasibility under structured capture, but continuous deployment remains dependent on imaging position, animal posture, and reliable segmentation of the relevant body region [99,102]. It also depends on expert

scoring for labeling and validation, and variation among scorers can affect both training and performance interpretation [101,102]. The cited evidence therefore supports task feasibility under defined protocols, not a general conclusion that routine deployment is already mature.

3.4.3. Body Weight Prediction

Body weight has been identified as one of the main target traits for DL and image analysis in cattle because it supports growth monitoring, nutrition management, and production decisions [107]. However, this area contains a wide range of approaches, including classical regressors on engineered shape features as well as DL models based on images, depth data, or point clouds [28,108]. Additional body weight studies use deep neural networks, semantic segmentation, modality comparison, image-based heifer mass prediction, and 3D reconstruction to infer live weight from visual measurements [109–113]. This makes the category methodologically less settled than lameness or behavior recognition.

Compared with body condition score estimation, body weight prediction usually benefits from structured capture geometry and accurately timed reference measurements [107,114]. It is often framed as scheduled phenotyping rather than continuous monitoring, and the visual observation must be paired correctly with scale weight. Transfer learning and flexible image-based methods may reduce local data requirements, but cross-farm usefulness still depends on capture compatibility, population differences, and access to representative validation data. Body weight prediction is therefore best interpreted here as a phenotyping application whose deployment conditions must be specified, rather than as a universal real-time monitoring solution.

3.5. Cross Comparison

The five dairy-management operations are compared across management purposes, required inputs, targeted outputs, label or reference requirements, validation designs, and deployment constraints.

For management tasks, identification and tracking are mainly built around detection, tracking, and re-identifying individual animals. Behavior recognition and lameness detection are more focused on classification of temporal or visual patterns, although they may also include detection and tracking as upstream steps. Body condition score estimation and body weight prediction are closer to estimation or prediction tasks because they aim to infer physiological state or body mass from visual information [115,116].

For input type, lameness detection and behavior recognition often rely on video and RGB data because motion is an important part of the signal [23,29,57]. Identification and tracking are also commonly based on RGB and video data, although some studies use additional geometric information. By contrast, body condition score estimation and body weight prediction often benefit more from depth images or point-cloud data because body geometry is closely related to the targeted output for these two operations [30,117].

For output type, lameness detection produces categorical or ordinal outputs such as lameness class, mobility score, or gait-abnormality alert. Behavior recognition outputs behavior classes, event sequences, or time-budget summaries. Identification and tracking output cow identity, re-identification matches, or tracking trajectories over time. Body condition score estimation outputs a BCS class or continuous score, whereas body weight prediction outputs a body weight estimate or growth trend.

Labeling requirement is where many practical bottlenecks arise. Lameness detection requires reliable expert locomotion or mobility scores, but label quality, mild-case separation, and farm-to-farm variation remain limiting factors [52]. Behavior recognition requires frame-, clip-, or event-level annotations [23,58,59,78]. Occlusion and overlapping animals

complicate behavior-label assignment [27,58,60], while rare events such as mounting create substantial class imbalance [26]. Individual-level behavior histories additionally require reliable identity continuity [27,79]. Identification and tracking require persistent identity and track labels across time and views, which become unstable under open-set conditions [118]. Body condition score estimation depends on expert scoring consistency and controlled body positioning, while body weight prediction depends on accurate reference measurements paired with clean body-shape extraction [107,119].

For deployment, lameness and behavior systems can be linked to repeated barn observation, but their practical validity remains conditional on the represented camera and animal-flow conditions [54,57,58,60]. Identification and tracking are most important when outputs require animal-level continuity over time. Body condition and body weight systems more often use controlled camera placement or structured imaging locations, so conclusions should remain bounded to those workflows [120,121].

The differences across these operations are summarized in Table 1.

Table 1 provides a qualitative comparison of the five operations in terms of management purpose, typical inputs and outputs, reference requirements, and deployment constraints. Table 2 complements it with a descriptive summary of input modalities, model families, dataset scales, validation designs, external or multi-farm evaluations, reported metrics, and provisional deployment interpretations. Because the cited studies differ in task definition, dataset composition, unit of analysis, and validation design, Table 2 is neither a meta-analysis nor a ranking of operations. Its counts and ranges summarize the 79 publications assigned to the five principal operations. The remaining 18 retained publications comprise four studies in the two secondary operations and 14 context-only technical-method publications, all documented in Appendix C Tables A3 and A4.

Table 2 describes substantial variation in sensing configurations, model families, dataset scales, and validation designs. Appendix C Table A3 provides study-level extraction data for the publications summarized in Table 2 and for those assigned to the two secondary operation categories. Appendix C Table A4 lists the context-only general AI/CV/ML publications. The operation-level counts and ranges reported in Table 2 can therefore be verified against the corresponding records in Appendix C Table A3.

The reported ranges combine non-equivalent outcomes, including accuracy, average precision, body condition tolerance, mean absolute percentage error, coefficient of determination, and root mean square error. A numerically higher value in one row therefore cannot be interpreted as stronger evidence than a value in another row. The lameness literature includes an example of evaluation beyond the development setting [122], but the operation-level external-validation counts and ranges are author-compiled summaries that can be checked against the study-level records in Appendix C Table A3. Inference speed, computing requirements, and benchmark-dataset identifiers were not reported consistently enough for reliable aggregation. The table identifies validation and reporting gaps; it does not establish that current accuracy is sufficient or that further model development is unimportant.

4. Discussion

Tables 1 and 2 indicate that several deployment barriers recur across the five operations, although their form and importance differ by task and intended use. The following sections distinguish reference-related limitations, capture-dependent constraints, multi-animal and temporal variation, and the role of individual identity continuity. The discussion then uses these themes to formulate a provisional research agenda derived from qualitative synthesis rather than an empirically validated ranking.

4.1. Label Reliability

Label reliability differs in form across the five operations and cannot be represented by a simple objective-to-subjective hierarchy. Body weight can be paired with a physical scale, but calibration error, measurement timing, and incorrect image-to-weight matching can still affect the reference value. Lameness detection and body condition score estimation depend more strongly on expert assessment, which may vary among observers and near category boundaries. Behavior categories may be defined explicitly, but continuous activity must be segmented into frames, clips, or events, creating uncertain transition boundaries. Posture and activity can also co-occur, so single-label schemes may conceal overlap. Identification and tracking likewise require reliable identity and track references during occlusion or herd change. Inter-observer agreement is not a universal numerical ceiling on model performance; it is evidence of reference uncertainty against which model results should be interpreted.

4.2. Data Capture Protocol

Data-capture protocol affects both deployment scope and interpretation of reported performance. The protocol defines the visual domain in which a model was developed and evaluated. A lameness model trained on a fixed side-view passage addresses a different operating condition from a model intended for continuous observation in an open pen. Likewise, a body condition model trained on standardized posture and camera geometry should not be assumed to perform under arbitrary views. Controlled capture is not inherently a weakness when it is a practical component of the intended workflow. The limitation arises when conclusions extend beyond the represented conditions.

Controlled protocols can reduce occlusion, background variation, and pose diversity, whereas less constrained settings introduce additional failure modes. Studies should therefore report camera placement, field of view, lighting, animal flow, posture requirements, the number of simultaneously visible animals, and exclusions of poor-quality observations. These details define the conditions under which performance claims remain valid. Consequently, the same metric name can represent substantially different levels of task difficulty across capture settings [120,123,124].

4.3. Multi-Animal Scenes and Changing Conditions

Beyond capture protocols, multi-animal scenes and changing conditions create additional processing challenges. Multi-animal scenes introduce occlusion, overlapping detections, ambiguous animal-to-behavior assignments, and identity switches. Changing conditions create distribution shifts across time, farms, camera systems, seasons, and herd compositions. Both problems may affect several operations, but their importance depends on the intended capture workflow and management output.

4.3.1. Scenes with Many Animals

In free-stall barns, cameras often observe several animals simultaneously. Occlusion and overlap can obscure body regions, confuse behavior labels, and cause identity switches. Some lameness, body condition, and body weight systems deliberately reduce these difficulties by observing one animal at a time in a passage or structured capture area [54,120,121]. Such configurations are valid deployment designs, but their results should remain bound to the corresponding workflow. Systems intended for open-pen use require separate evaluation under representative crowding and occlusion.

4.3.2. Conditions That Change over Time

Deployment conditions change after model development. Animals enter and leave the herd, and appearance varies with growth, coat condition, dirt, lactation stage, camera view,

lighting, and season. Models trained on fixed rosters or narrow environmental conditions may lose accuracy when new animals appear or the visual domain changes. Temporal and cross-farm distribution shifts should be distinguished from limited data access: the former is a model-transfer problem, whereas the latter limits the evidence available to study and validate transfer.

4.4. Identification and Tracking

Individual identity continuity is required when observations must be linked to the same animal over time. It is not a universal prerequisite for every computer-vision output: group-level summaries or per-observation classifications may remain useful when identity is unnecessary or supplied by another farm component. By contrast, individual time budgets, mobility trajectories, condition trends, and growth curves require reliable association across observations. Identification and tracking should therefore be treated as enabling infrastructure for animal-level longitudinal applications, not as an unavoidable visual module for every operation.

Within a vision-based pipeline, association errors can propagate into downstream records. A mis-association attributes an observation to the wrong cow, whereas fragmentation divides one animal's history among several identities. The downstream effect is not described by a universal numerical bound because it depends on the error type, event frequency, and aggregation rule. When identity is part of the proposed pipeline, studies should evaluate end-to-end performance using estimated rather than ground-truth identity and report identity switches, track fragmentation, open-set recognition, and the resulting change in downstream outcomes.

4.5. Implications and Future Research

The qualitative synthesis suggests a provisional three-part research agenda. The first group concerns candidate foundations for interpretable animal-level outputs: reference-standard reliability and identity continuity. The second concerns transfer and data infrastructure, including cross-farm generalization and data access. The third concerns conditional extensions, including multimodal integration and longitudinal evaluation. This grouping reflects dependency logic rather than measured effect size, prevalence, expected return, or a validated priority order. Different operations may reasonably follow different sequences.

Table 3 maps each barrier to an operation at the task level. Its entries are qualitative judgments derived from the operation definitions and constraints summarized in Tables 1 and 2; they are not independently coded study frequencies. The table should therefore be used to visualize candidate relationships and dependencies, not to estimate how many studies documented a barrier or to prove that one barrier has greater impact than another.

The candidate foundation group includes reference-standard reliability. Reference uncertainty differs by task: lameness and body condition depend heavily on expert scores, behavior depends on annotation unit and transition boundaries, identification depends on identity and track references, and body weight depends on calibrated and correctly paired measurements. Human agreement is not a strict numerical ceiling on model performance, but it changes how apparent performance should be interpreted. Studies should report observer agreement where appropriate, document measurement and adjudication procedures, preserve uncertainty near class boundaries when useful, and evaluate objective biomechanical measurements as complements to expert lameness scores.

Identity continuity is the second candidate foundation for applications that require animal-level histories. It is the direct output of identification and tracking and an enabling input for longitudinal uses of the other operations. Evaluation should include crowded

scenes, open-set conditions, track fragmentation, and identity switches, and downstream tasks should be tested with estimated identity when visual association is part of the system. An external farm identification component may be appropriate when visual identity is not itself the research objective.

The transfer and data infrastructure group includes cross-farm generalization. Each operation may encounter changes in population, housing, camera configuration, capture protocol, season, and management practice beyond the development setting. Studies should state these conditions explicitly and distinguish within-farm evaluation from temporal, cross-camera, and cross-farm validation. Comparisons between controlled and operational settings may help identify which performance losses arise from capture complexity rather than from the target task itself.

Table 2 supports only a descriptive version of this argument. Reported metrics are often strong within individual datasets, but they are based on different sensing modalities, outcomes, and validation designs and cannot be compared directly. The dataset and validation columns suggest that evaluation beyond a development farm is less common than within-dataset testing. The evidence therefore establishes feasibility under conditions more clearly than transfer across farms, seasons, camera setups, or herd compositions.

Data access and interoperability are enabling constraints rather than operation-specific failure modes. Representative cross-farm evaluation requires compatible metadata, annotation definitions, reference measurements, and governance arrangements. Collaborative datasets, privacy-preserving analysis, and agreed reporting standards may expand the conditions under which transfer can be evaluated. These approaches should be judged by whether they improve external validation and reproducibility, not simply by the volume of pooled data.

The conditional extension group includes multimodal integration and longitudinal evaluation. Additional modalities may address a defined visual failure mode, but they also increase calibration, synchronization, cost, and maintenance requirements; fusion is therefore not inherently superior. Longitudinal evaluation is central when the intended output is an animal-level history or trajectory, but less central for some per-observation or group-level tasks. It should be defined at the level of the management decision and evaluated end to end, including missing observations and identity errors.

The proposed grouping is a hypothesis-generating research agenda, not a demonstrated universal sequence. It suggests that stronger reference standards and identity continuity may make subsequent transfer, fusion, and longitudinal analyses more interpretable, but this review does not test that prediction. Prospective cross-farm and end-to-end comparisons are needed to determine which research sequence provides the greatest deployment benefit for each operation.

4.6. Deployment Considerations and Reporting Priorities

Hardware and operating requirements influence whether a dairy computer-vision system can be used routinely on farms. This review did not systematically extract equipment prices, energy use, inference latency, maintenance requirements, staff time, or return-on-investment estimates, so it does not compare the economic performance of individual approaches. These variables should instead be treated as priorities for future reporting. Deployment-oriented studies should describe camera and computing hardware, inference speed on the reported device, installation and maintenance needs, staff training, expected service life, and assumptions used in cost-benefit analysis.

4.7. Methodological Directions for Further Evaluation

Several methodological directions may be relevant to the limitations identified in this review, although their value in dairy computer vision remains to be established. Class

re-weighting, loss, targeted augmentation, few-shot learning, and anomaly-detection formulations provide candidate tools for imbalanced labels and scarce examples [80–84]. Transferable visual representations, privacy-preserving multi-farm learning, and integrated simulation or decision-support systems are also plausible research directions rather than established trends within the reviewed dairy literature. Their usefulness should be evaluated through rare-event sensitivity, cross-farm transfer, identity continuity, robustness to changing capture conditions, and improvement beyond a single curated dataset.

4.8. Ethical, Privacy, and Reliability Considerations

Deploying AI in dairy operations raises governance, welfare, and reliability questions that this review did not systematically evaluate. Continuous video may create obligations concerning farm-data ownership, access, retention, and security. Automated welfare monitoring may influence care decisions, particularly when model failures are not readily visible to farm staff. Deployed systems should therefore support human oversight, documented failure handling, and periodic evaluation under changing herd and environmental conditions. These considerations are implementation requirements, not evidence that the reviewed systems have resolved privacy, welfare, or reliability risks.

4.9. Limitations of This Review

This review has several limitations. The search was restricted to four databases, to English-language peer-reviewed journal articles, and to the period 2016 to 2026, and the 2026 count is partial because the search covered publications only through 8 July 2026. Although the first and second authors jointly reviewed and classified all retained publications and completed the detailed extraction used for the operation-level synthesis, the coding process involved cross-checking and consensus rather than independent duplicate coding before consensus. Consequently, inter-reviewer agreement could not be quantified. In addition, the barriers were not coded as binary study-level variables. Therefore, this review does not estimate their frequency or prevalence across the included studies. The scope was limited to DL-based computer vision for dairy cows at the animal level, so sensor-based machine learning, decision-support systems, and non-cattle dairy livestock are not covered. The five operation categories were derived from the selected literature and the authors' operation-centered coding rather than from a formal protocol, and studies spanning multiple operations required judgment to classify. The citation counts reported in Figure 3 reflect the papers selected and discussed in this manuscript, not the entire dairy DL literature. Finally, the three-tier framework is a qualitative prioritization derived from the cross-operation comparison in Table 1 and is not a formal meta-analysis; it should be read as a research agenda rather than as a quantitatively validated ranking.

5. Conclusions

This review organizes selected literature on DL-based computer vision for dairy cow management around five operations. Across the manuscript's evidence map, recurring deployment concerns involve reference labels, capture protocols, multi-animal scenes, changing farm conditions, data access, and identity continuity. The synthesis supports treating reference-standard quality and individual identity as candidate foundations for animal-level longitudinal applications, while recognizing that identity may be supplied by non-visual farm infrastructure and is not required for every group-level or per-observation output. Cross-farm validation, interoperable data, multimodal integration, and longitudinal assessment are additional needs, but their relative order cannot be established from this narrative review. The proposed three-part agenda is therefore a qualitative organizing scheme rather than a validated ranking or meta-analysis. Its value requires con-

firmation through transparent study-level extraction, reproducible coding, and evaluation across farms, seasons, herds, and camera configurations. Progress will require coordination among animal scientists, computer-vision researchers, veterinarians, producers, and technology providers.

Author Contributions: Conceptualization, C.Z., Z.Y. and J.W.; methodology, C.Z. and Z.Y.; literature search and organization, C.Z. and J.W.; formal analysis, C.Z.; writing—original draft preparation, C.Z. and Z.Y.; writing—review and editing, C.Z. and Z.Y.; supervision, Z.Y.; project administration, Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: The study was supported by the Hatch project (FLA-ABE-006557).

Institutional Review Board Statement: This review did not involve human or animal participants.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new experimental data were generated in this review. The coded literature dataset created during the review is provided in Appendix C Tables A3 and A4. Table A3 contains study-level extraction fields for the 79 publications summarized in Table 2 and four additional operation-specific publications assigned to reproductive monitoring or feeding and rumination monitoring. Table A4 contains the 14 context-only general AI/CV/ML technical-method publications; operation-specific fields are marked as not applicable for these records. Together, Tables A3 and A4 account for all 97 retained publications. Table 3 is an operation-level qualitative mapping and was not derived from study-level barrier-frequency coding.

Acknowledgments: The authors thank colleagues in the Animal Sciences Department at the University of Florida for helpful discussions during the preparation of this review. During manuscript preparation, the authors used ChatGPT 5.5 (OpenAI) to suggest candidate search synonyms, assist with database-specific Boolean syntax, and improve language and readability. The authors reviewed and revised the resulting suggestions before use. ChatGPT was not used to retrieve or screen records, make eligibility decisions, extract study data, perform analyses, assign categories, or determine conclusions. The authors are responsible for the final search strategy, interpretation, and manuscript content.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Mapping of Retained Publications by Synthesis Status and Primary Category

This appendix maps all 97 retained publications. The seven operation-specific categories contain 83 publications, while 14 general AI/CV/ML technical-method publications were retained only for context. Each publication appears once.

Table A1. Mapping of the 97 retained publications by synthesis status and primary category.

Synthesis Status/Primary Category	Distinct Publications	Citations
Behavior recognition	22	Achour et al. (2020) [66]; Bai et al. (2024) [25]; Li et al. (2024) [65]; Fuentes et al. (2020) [23]; Hua et al. (2023) [67]; Koskela et al. (2022) [78]; Li et al. (2025) [68]; Liu et al. (2026) [69]; Long et al. (2025) [70]; Ma et al. (2022) [71]; McDonagh et al. (2021) [59]; Rao et al. (2026) [72]; Shang et al. (2022) [73]; Bai et al. (2023) [123]; Li et al. (2026) [74]; Tong et al. (2024) [27]; Wu et al. (2021) [24]; Wu et al. (2023) [63]; Wu et al. (2025) [75]; Yu et al. (2022) [76]; Yu et al. (2024) [60]; Zong et al. (2025) [77]

Table A1. Cont.

Synthesis Status/Primary Category	Distinct Publications	Citations
Individual identification and tracking	18	Hooker et al. (2026) [87]; Wang et al. (2024) [118]; Dac et al. (2022) [88]; Guzhva et al. (2018) [79]; Wang et al. (2023) [16]; Kumar et al. (2018) [89]; Li et al. (2022) [90]; Pathak and Prakash (2025) [91]; Qiao et al. (2021) [39]; Ruchay et al. (2024) [92]; Shu et al. (2026) [93]; Smink et al. (2024) [85]; Xiao et al. (2022) [31]; Xu et al. (2025) [94]; Yukun et al. (2019) [86]; Han et al. (2023) [95]; Mg et al. (2024) [96]; Li et al. (2021) [125]
Lameness detection	18	Abdul Jabbar et al. (2017) [41]; Russello et al. (2024) [57]; Czirok et al. (2026) [42]; Duan et al. (2025) [53]; Higaki et al. (2025) [29]; Barney et al. (2023) [43]; Jia et al. (2025) [51]; Jiang et al. (2020) [44]; Kang et al. (2020) [50]; Kang et al. (2022) [45]; Kang et al. (2025) [46]; Myint et al. (2024) [54]; Narli et al. (2025) [52]; Siachos et al. (2025) [122]; Szyc et al. (2025) [124]; Tun et al. (2024) [47]; Tun et al. (2026) [48]; Wu et al. (2020) [49]
Body condition score estimation	12	Lewis et al. (2025) [104]; Li et al. (2025) [100]; Nagy et al. (2023) [105]; Rodriguez Alvarez et al. (2018) [30]; Rodriguez Alvarez et al. (2019) [98]; Shi et al. (2023) [101]; Siachos et al. (2024) [119]; Yao et al. (2026) [120]; Feng et al. (2024) [99]; Summerfield et al. (2023) [102]; Zhao et al. (2023) [103]; Dandil et al. (2024) [106]
Body weight prediction	9	Afridi et al. (2024) [109]; Bi et al. (2023) [28]; Dohmen et al. (2021) [110]; Gebreyesus et al. (2023) [107]; Guvenoglu (2023) [111]; Guzhva et al. (2026) [108]; Liao et al. (2025) [121]; Xu et al. (2024) [112]; Zhang et al. (2026) [113]
Reproductive monitoring	2	Hanpinitasak et al. (2026) [64]; Li et al. (2024) [26]
Feeding and rumination monitoring	2	Giannone et al. (2025) [58]; Wang et al. (2024) [62]
General AI/CV/ML technical method/background	14	Aharon et al. (2022) [19]; Amershi et al. (2019) [126]; Bewley et al. (2016) [20]; Bochkovski et al. (2020) [11]; Cao et al. (2017) [17]; Dosovitskiy et al. (2021) [10]; He et al. (2016) [9]; He et al. (2017) [15]; Liu et al. (2016) [14]; Mathis et al. (2018) [18]; Redmon et al. (2016) [12]; Redmon and Farhadi (2018) [13]; Wojke et al. (2017) [21]; Zhang et al. (2022) [22]

Appendix B. Database-Specific Search Strings Used for Recheck

The operation-informed search was adapted to each database because field tags and Boolean handling differ across platforms. Searches covered English-language literature published from 2016 through 8 July 2026. The strings below are the final refined queries used to recheck the per-database retrieval counts on 8 July 2026. They document the endpoint of an iterative structured-narrative search rather than a preregistered systematic-review protocol. Resulting counts are specific to these strings and date and do not measure field-wide publication prevalence, popularity, practical importance, or economic value.

Table A2. Database-specific search fields, syntax, and final refined search strings used for the literature recheck conducted on 8 July 2026.

Database	Search Field/Syntax	Search String
Web of Science	TS field	TS = (("dairy cow*" OR "dairy cattle") AND "deep learning" AND "computer vision" AND (lameness OR behavior OR behaviour OR identification OR tracking OR "body condition" OR "body weight" OR "body condition score")) AND PY = (2016–2026) AND DT = (Article) AND LA = (English)
Scopus	TITLE-ABS-KEY field	TITLE-ABS-KEY(("dairy cow*" OR "dairy cattle") AND "deep learning" AND "computer vision" AND (lameness OR behavior OR behaviour OR identification OR tracking OR "body condition" OR "body weight" OR "body condition score")) AND PUBYEAR > 2015 AND PUBYEAR < 2027 AND DOCTYPE(ar) AND LANGUAGE(english)
ScienceDirect	keyword search	"dairy cow" AND "deep learning" AND "computer vision" AND (lameness OR "body condition" OR "body weight"), date filter 2016–2026, research articles
Google Scholar	simplified phrase search	"dairy cow" "dairy cattle" "deep learning" "computer vision" lameness behavior "body condition" "body weight", year range 2016–2026

* is a truncation symbol in a query to represent multiple characters.

Appendix C. Coded Literature Dataset for the 97 Retained Publications

Table A3. Study-level extraction for the 83 operation-specific publications.

Study	Operation Category	Sensor/Input Modality	Dataset Size	No. of Animals	No. of Farms/Sites	Model Family	Validation Method	External Validation Status	Main Metric
Fuentes et al. (2020) [23]	Behavior	RGB video; optical flow; keyframes	350 videos (7 farms); 2714 keyframes; 15 behaviors	NR	7	YOLOv3-Darknet53 + 3D CNN/optical flow; Faster R-CNN/VGG16 comparator	80/10/10 train/validation/test; split unit NR; no farm/animal holdout	No	Spatiotemporal mAP 0.856 vs. frame-level mAP 0.788; YOLOv3: 30 fps
Wu et al. (2021) [24]	Behavior	Infrared-capable surveillance RGB video; cow clips	4566 videos (~63 h; 10–55 s); 5 behaviors; 3196 train/1370 test	31	1	VGG16 CNN feature extractor + bidirectional LSTM	Video-level 70/30 train/test; same 31-cow cohort; no cow holdout	No	Mean accuracy 0.976, precision 0.971, recall 0.965, specificity 0.983
Bai et al. (2024) [25]	Behavior	RGB video frames/images	1295 images; 1033 train/262 test; 4 behaviors	NR	1	GC_Res2-YOLOv3 with Res2 backbone and global-context head	Image-level random 80/20; new-scene examples were qualitative only	No	P 0.912, R 0.857, mAP@0.5 0.904, mAP@0.5: 0.95 0.685; mean class AP 90.4%
Tong et al. (2024) [27]	Behavior	RGB surveillance, handheld and smartphone images/video	3484 images/16,673 labels (2440/696/348); 10 60 s clips; 16,288 re-ID images/15 cattle	15 (re-ID); behavior-detection count NR	1 named farm; public-source sites NR	YOLOv8n detector + BoT-SORT tracking + ResNet50 re-identification	Image-level 5-fold CV on 3136 train/validation images, then held test; site/animal holdouts NR	No	Detector P/R/mAP/F1: 90.9/84.7/91.7/87.7%; HOTA/MOTA/MOTP/IDF1: 73.1/80.4/84.4/87.5%
Achour et al. (2020) [66]	Behavior	Top-of-feeder RGB images/video	5 days; 7801 images; feeding set 4875; ID train/test 2931/1944	17	1	Multi-stage CNNs for presence, behavior, segmentation, feed, and identification	Day-level: days 1–2 train, days 3–5 validation at other feeders; same cows/farm	No	Feeding acc 92.61%, P 94.18%; ID acc 97.06%; feed assessment 100%
Li et al. (2024) [65]	Behavior	Multi-camera RGB video; daylight and night	>96 h; 687 annotated segments; 206,100 frames; CBVD-5	107	1 ranch	SlowFast video classifier with detection/cropping preprocessing	Random 70/20/10 train/test/validation; split unit NR; no cow/ranch holdout	No	SlowFast test error 21.28% (derived accuracy 78.72%)
Hua et al. (2023) [67]	Behavior	RGB video/images; pose-skeleton sequences	400 ~10 s videos (100/action); 1800 labeled images; 4 actions	NR	1	YOLOX-Pose keypoint extraction + PoseC3D 3D CNN	Video- and image-level random 60/20/20 train/validation/test; no cow holdout	No	YOLOX-Pose AP 96.3% at 33.1 fps; PoseC3D Top-1 accuracy 92.6%
Koskela et al. (2022) [78]	Behavior	RGB security video; extracted still images	~19 h; >1.7 M images; 10 behaviors; experiments: 100,000 unbalanced and 49,900 balanced	NR	1 barn	VGG-based CNN image classifier	Image-level shuffled splits: 80/10/10 (100,000) and 50/25/25 (49,900); animal independence NR	No	Unbalanced weighted P 86%/R 85%; balanced P 78%/R 79%
Li et al. (2025) [68]	Behavior	Outdoor surveillance/smartphone RGB video frames	78 surveillance + 116 phone videos; 146 10–20 s segments; 9460 frames/4728 images/13,224 labels	NR	1	RFR-YOLO (YOLOv11n with RsiConv, FMSDA and RepGRFPN)	Image-level random 60/20/20: 2838/945/945 train/validation/test	No	P 95.9%, R 91.2%, mAP@0.5 94.9%, mAP@0.5: 0.95 85.2%; 147.6 fps

Table A3. Cont.

Study	Operation Category	Sensor/Input Modality	Dataset Size	No. of Animals	No. of Farms/Sites	Model Family	Validation Method	External Validation Status	Main Metric
Liu et al. (2026) [69]	Behavior	Public and self-collected RGB video frames/images	1777 pooled + 194 added images; 1226 augmented; 3197 total	NR	≥2 sources; independent site count NR	BFDet-YOLO (YOLOv10), dual forage-density and feeding-behavior detection	Pooled image-level random 6:2:2: 1918/639/640 train/validation/test; no site holdout	No	mAP@0.5 95.7%; mAP@0.5: 0.95 70.7%; 1.85 M parameters
Long et al. (2025) [70]	Behavior	Indoor/outdoor RGB and nighttime IR video frames	35 one-hour indoor videos; 3360 images; 4 behaviors	NR	1	FSCA-YOLO (YOLOv11 with FEM-SCAM, P2 head, CoordAtt and Siou)	Internal development/validation/test; counts, ratio, split unit and cow/source-video separation NR	No	P 95.7%, R 92.1%, mAP@0.5 94.5%
Ma et al. (2022) [71]	Behavior	RGB surveillance video	286 videos/181,500 frames (228/58); 120 validation clips; 10 unedited (1722 s/43,065 frames)	30	1	Improved RexNet3D CNN + Softmax	Video-level 80/20 (228/58); separate 120 clips + 10 unedited videos; cow independence NR	No	120 clips: acc 95%, 76.52 fps; unedited videos: acc 91.02%, 101.02 fps
McDonagh et al. (2021) [59]	Behavior	Infrared-capable surveillance RGB video clips	10 h annotated; 19,191 records; 3969 clips of 3–10 s (3186 train/783 validation)	46	1 site	Non-local video classifier pretrained on Kinetics	Clip-level 3186/783 train/validation; cow independence NR	No	Per-class acc ≥ 80%; contraction 83%; standing 84%; lying 90%; eating 95%; drinking 93%
Rao et al. (2026) [72]	Behavior	Continuous multi-camera RGB video, images, 10 s clips	Detection: 2308 images (1923/193/192); behavior: 4964 clips, augmented to >9600	Model datasets NR; timing test: 16	1	YOLOv11 + ByteTrack + TimeFormer; SlowFast comparator	Image-level 1923/193/192 and clip-level 70/15/15 splits; cow independence NR; one 24 h case cow	No	TimeFormer acc 85.0%; macro-F1 0.84, 22.6 fps; detector mAP@0.5 0.994; pipeline 5.5 fps/cow
Shang et al. (2022) [73]	Behavior	Farm RGB images/video; public action datasets	Detection: 1400 images (1000/400); behavior: 61 videos/4320 images (3600/720)	NR	1 cattle farm (target task)	YOLOv4 + Siamese selection + attention-enhanced MobileNetV3 + trajectory analysis	Image-level train/test: detection 1000/400, behavior 3600/720; 6-fold CV	No for target cattle operation	Behavior-recognition accuracy up to 95.17%
Bai et al. (2023) [123]	Behavior	Top-down oblique RGB video clips	Text: 4366 videos; Table 1 class counts total 5650; 4 behaviors	NR	1	X3DFast two-pathway 3D CNN (X3D + fast pathway with attention and R(2+1)D)	Video-level stratified random 80/20 train/validation; animal independence NR	No	Top-1 behavior-classification accuracy 98.49%
Li et al. (2026) [74]	Behavior	Three-second RGB video clips	11,684 3 s clips (7391/2224/2069); source: 687 segments, 206,100 frames, >96 h	107	1 ranch	CLIP visual encoder + temporal adapters/pooling + behavior-prior loss	Source-video-level 160/45/46 train/validation/test; clips generated after split; 3 random seeds	No	Accuracy 75.45%; macro-F1 0.7246
Wu et al. (2023) [63]	Behavior	Fixed-camera RGB video and extracted images	4000 annotated images from 34 videos; 60 full-pipeline test videos	NR	1	YOLACT + VGG16/CNN-BiLSTM + optical-flow/PBVM respiratory algorithms	Image-level 70/30 on 4000; pipeline on 60 videos; overlap with 34 development videos NR	No	Respiratory-rate accuracy 93.56%; MAE 3.42 and RMSE 3.74 breaths/min
Wu et al. (2025) [75]	Behavior	RGB video frames/images	4632 augmented images (3234/918/480; 70/20/10), every 100th frame; 31,824 labels	NR	2	DMSF-YOLO (YOLOv11n): MSFConv, C2BRA, DyHead	Random pooled 70/20/10 image split across 2 farms; no farm-held-out test	No	Precision 86.7%, recall 84.0%, mAP@0.5 88.8%, F1 85.3%, 66.7 fps
Yu et al. (2022) [76]	Behavior	Front/top RGB video frames; depth camera (monocular)	20 sets × 4 videos; 10,288 images; front 4484/758 and top 4320/726 train/test; 3 behaviors	17	1	DRN-YOLO (YOLOv4-derived): DRNet, 4 scales, SPP, CioU	Fixed image-level train/test sets by view; selection procedure and source-video separation NR	No	Front P/R/mAP/F1 97.16/96.51/96.91/96.83%; top 96.84/96.25/96.49/96.55%; 22.65 ms/image
Yu et al. (2024) [60]	Behavior	Day RGB + night infrared surveillance video frames	35 two-minute clips; every 12th frame; 5516 images; 4 behaviors	90	1	Res-DenseYOLO (YOLOv5): dense/residual backbone, CoordAtt, Siou, 4 heads	Internal validation set; allocation counts/ratio and cow/source-video separation NR	No	Precision 94.7%, recall 91.2%, mAP@0.5 96.3%
Zong et al. (2025) [77]	Behavior	Indoor/outdoor RGB surveillance video frames	1044 videos (1008 ~5 min; 36 of 0.5–3 h); 3458 images; 14,754 boxes; 480 qualitative	38	1	Improved YOLOv5: Shuffle Attention, DCNv3, DyHead	Random image-level 80/10/10 split (2766/346/346); 480 extra images qualitative only	No	P/R/F1/mAP 93.8/96.7/95.2/97.7%; class AP stand/lie/eat/drink 98.1/98.9/97.7/96.1%
Wang et al. (2023) [16]	ID/tracking	High-resolution RGB cattle images	2012 captured/2000 selected images; augmented to 8000 (4800/1600/1600 train/val/test)	100	1	OP-Mask R-CNN: OpenPose skeleton features + CBAM; Mask R-CNN/YOLOv5/Attention-BiLSTM	Animal-level: 80/10/10 cows train/val/test; images 3:1:1; robustness perturbation	No; unseen-cow holdout was same-farm	Cross-subject accuracy 90.2%; cross-view accuracy 92.1%
Xiao et al. (2022) [31]	ID/tracking	Top-view RGB video and extracted images	12,000 segmentation images (7:2:1); 8640 identity images (7:3)	48	1	Improved Mask R-CNN (Light-ResNet101, SE, FPN); grid/shape/Hu + Fisher + RBF-SVM ID	Random image-level splits: segmentation 7:2:1, ID 7:3; animal-disjoint separation NR	No	Segmentation AP/F1 97.39/97.34%, 1.02 s/image; ID accuracy 98.67%
Hooker et al. (2026) [87]	ID/tracking	Ceiling-mounted RGB video; cow crops/track pairs	3120 behavior images; 1059 ID crops; test 1620 images/8490 labels; 20,000 pairs from ID crops	21	1	YOLOv7 behavior detector; MTCN Siamese MobileNetV3 ID; baseline CNN	Detector image 80/20; ID image 90/10; later-time same-cow, same-farm test	No; later-time test used same cows/farm	Behavior accuracy 90.47%, kappa 0.859; ID accuracy 83.6%, kappa 0.827
Wang et al. (2024) [118]	ID/tracking	Overhead RGB images; natural/artificial lighting	9974 images; 62 known cows; 7069 train/1801 val; 8 unseen cows; 1104 open-set test	70 (62 known + 8 unseen)	1	ResSTN + SimAM + ArcFace metric learning/open-set recognition	62-cow images 80/20 train/val; animal-level open-set test on 8 unseen cows; 10-fold threshold analysis	No; 8 unseen cows were from same farm	Known-cow val top-1 99.72%; unseen/open-set cosine accuracy 94.58%

Table A3. Cont.

Study	Operation Category	Sensor/Input Modality	Dataset Size	No. of Animals	No. of Farms/Sites	Model Family	Validation Method	External Validation Status	Main Metric
Dac et al. (2022) [88]	ID/tracking	RGB face video and extracted face chips	281 one-min videos (~505,800 frames); detector 1780 frames (80/20); encoder 3560/1780 chips	102 (89 closed-set + 13 open-set)	1	YOLOv5s6 face detector; ResNet-18 landmarks; MobileNetV2 embeddings + cosine lookup	89 cows: 1 video train/1 test; 13 single-video open set; 89 train/102 test videos; component image splits	No; 13 unseen/open-set cows were same-farm	Closed-set rank-1/CMC 84%; temporal buffering ~90–91%; encoder val lookup 93.14%
Guzhva et al. (2018) [79]	ID/tracking	Stitched multi-camera RGB panoramic barn video	4 months; 2200 frames; 9279 cow instances; 64,953 landmarks; 26 evaluation tracks	252 farm cows; 26 evaluated tracks	1	Two-stage CNN cow detector + probabilistic multi-object tracker	Annotated-frame detector test; 26 manually initialized tracks; split unit: frame/track; no holdout	No	23/26 tracks correct; 101.29 min tracked; reported mean 225 s/start (calculated ~234 s)
Kumar et al. (2018) [89]	ID/tracking	RGB muzzle-print images	>5000 images; >500 cattle; 100 train/400 test cattle; galleries 1–4 images	>500 (100 train; 400 test)	NR	CNN, SDAE, DBN/RBM and one-shot similarity framework	Animal-level: 100 train vs. 400 test cattle; 5 × random subsampling; gallery-size tests	Unclear: source private; farm independence unreported	DBN accuracy 95.99%; up to 98.99% with 400 features
Li et al. (2022) [90]	ID/tracking	RGB cattle-face video frames/images	68,260 raw images; 10,239 face images from 103 cattle	103	2 collection sites	Lightweight 6-conv CNN with global pooling; Raspberry Pi 4 deployment	Ambiguous random 70/30 train/validation/test; split unit: image; identities shared across splits	No held-out-site or held-out-animal test	Accuracy 98.37% (macro ~98%); 2.01 MB, 0.17 M parameters; ~300 ms/image on Raspberry Pi
Pathak et al. (2025) [91]	ID/tracking	RGB cattle face and muzzle images	CMPD300: 2632 images/300 cattle; CMPD268: 4923/268; combined >7000/568	568 across two public datasets	NR	YOLOv8n + EfficientNetV2-B0 attention; confidence-based face/muzzle fusion	Random splits: CMPD300 58/21/21; CMPD268 65/15/20; combined 61/18/21; split unit: image	No: no untouched cross-dataset or unseen-farm test	CMPD300 muzzle/face fusion 99.47/98.58/99.64%; CMPD268 98.57%; combined 98.92%
Qiao et al. (2021) [39]	ID/tracking	Rear-view RGB video at three monthly dates	363 videos from 50 cattle; 288 train, 75 test	50	1	InceptionV3 frame features + BiLSTM + self-attention	288/75 video train/test split; split unit: video; same 50-cow cohort	No	Accuracy 93.3%; precision 89.3%; recall 91.0%; F1 90.2%
Ruchay et al. (2024) [92]	ID/tracking	RGB cattle-face photographs	315 originals from 91 cattle; augmented to 3150 images	91	1	YOLOv8 face preprocessing + VGGFace/VGGFace2 ResNet-50 transfer learning	Random 80/20 image split with augmentation; split unit: image; same cattle across splits	No	VGGFace2-based identification accuracy 97.1%
Shu et al. (2026) [93]	ID/tracking	Fixed side-view RGB surveillance face images/video	ID: 3915 images/102 cows (82 dev; 20 unseen); detector: 1250 images/3623 instances/6 pens	102	1	YOLO11n-OBb + ConvNeXt-Tiny; improved ArcFace (IAAL/ICCL)	Detector pens 1–4/5/6 train/val/test; ID gallery/probe on 20 unseen cows; split unit: pen/cow	No: pen and unseen-cow tests were same-farm	AP 0.995 (11.2 ms); closed-set acc. 0.97 (0.6 ms); open-set F1/R/F1/acc. 0.98/0.61/0.75/0.68
Smink et al. (2024) [85]	ID/tracking	RTSP RGB security-camera video; ear-tag crops	1453 detector images; 1581 tag crops; 1 h field test: 550 cows, 646 readable tags	550 field-test cows; training unique NR	≥2 training + 1 field farm; overlap NR	YOLOv5s + DeepStream + WhenToRead + TRBA tag reader (TPS/ResNet/BiLSTM/attention)	Random image/tag splits; 600-frame ablation; separate 1 h field test; split unit: image/tag/field stream	Unclear: field-farm independence unreported	Detector mAP near/far 0.937/0.676; printed-tag accuracy 96.1%; all-cow 71.3%; 24.69 fps
Xu et al. (2025) [94]	ID/tracking	Multi-camera RGB cow head/back images	33,019 ID images: 26,897/195 cows train, 3311/24 val, 2811/24 test; detector 3782 images	243	1	YOLOv5n-Ghost + task-driven autoencoder + MAML few-shot identifier	Cow-disjoint 195/24/24 meta-train/val/test; detector early-day train vs. later-day val/test; split unit: cow/time	No: cow-disjoint and temporal tests were same-farm	Detector mAP 99.5%; ID accuracy 90.43% (20-way 5-shot), 95.62% (5-way 5-shot)
Yukun et al. (2019) [86]	ID/tracking	Overhead 3D video, depth/grayscale and phase features	3430 images (5/cow): 2400/480 cows train; 1030/206 cows test	686 (480 train; 206 disjoint test)	1	CNN detection/back-pattern ID + ultrasound-backfat linear regression for BCS	Animal-level 70/30 split; 480 train vs. 206 disjoint test cows; split unit: cow	No: disjoint test cows were from same farm	ID 0.937 online (0.985 model); BCS accuracy within 0/0.25/0.5; 0.45/0.77/0.98; detection 100%
Han et al. (2023) [95]	ID/tracking	RGB CCTV video	2250 detector images; 570 appearance/9690 instances; Farm 2: 2000 fine-tune; test: 6 clips/3800	17 test cattle; training unique NR	2	YOLOv5 + wide ResNet/SPP appearance model + Kalman/bench-matching tracker	Farm 1 transfer; Farm 2 detector fine-tuned on 2000 images, then six-clip test; split unit: farm/image/clip	No untouched external test: target farm used for fine-tuning	IDF1 90.22; AssA 84.49; HOTA 77.64; MOTA 82.75; ~3 fps
Mg et al. (2024) [96]	ID/tracking	360° fisheye RGB video; four camera-view regions	1320 images (1080 train/240 val); day and night tests: 72,000 frames each (144,000 total)	Up to 7 in temporal tests; training unique NR	1	Detectron2 detector + customized cattle tracker (CTA); eight tracker comparators	Date-separated 1080/240 image train/val; same-farm 5 h day and 5 h night tests; unit: image/time	No	Detector AP/AP50 91.12/97.67; detection night/day 99.91/99.81%; CTA MOTA up to 99.99%
Li et al. (2021) [125]	ID/tracking	Side-view RGB photos; varied angles/backgrounds	3772 originals; 1485 selected; 80/20 split; training images expanded 3 × by rotation	13	1	Lightweight AlexNet CNN with multiscale blocks, attention and global pooling	Random 80/20 train/validation; split unit: image; no cow- or farm-held-out test	No	Validation accuracy 97.95%; 1,707,592 parameters; 6.51 MB
Higaki et al. (2025) [29]	Lameness	Top-view RGB video; side-view RGB reference	116 raw videos; 803 sequences; 660 candidate and 256 selected unique-cow clips (6.5 s mean)	256	1	AiPEC proprietary pose/keypoints + 25 mobility variables + random forest	Repeated 10 × cow-level 80/20 holdout, no cow overlap; split unit: cow	No	κ 0.72; AUC 0.89; score-0 sens./spec./F1 accuracy 87.2/94.7/88.2/92.1%
Abdul Jabbar et al. (2017) [41]	Lameness	Overhead 3D depth video; RFID identity only	23 sessions from 22 cows; ~70 frames/cow/session	22	1	Depth curvedness features + Hilbert transform + linear SVM	Confusion matrix on all 23 sessions; no holdout/CV; split unit: session (apparent resubstitution)	No	Accuracy 95.7%; sensitivity 100%; specificity 75% (n = 23 sessions)
Russello et al. (2024) [57]	Lameness	Side-view RGB-D walking video	1101 collected; 272 selected videos/98 cows; pose: 28 cow videos, 968 two-frame sequences	98	1	Faster R-CNN + T-LEAP pose; 6 gait traits; SVM-R/RF/gradient boosting	5-fold stratified group CV by cow for lameness; pose 774/194 samples; split unit: cow vs. frame-sequence	No	SVM-R accuracy/F1/sens./spec. 80.07/78.70/76.78/81.15%; pose PCKh@0.2 99.60%

Table A3. Cont.

Study	Operation Category	Sensor/Input Modality	Dataset Size	No. of Animals	No. of Farms/Sites	Model Family	Validation Method	External Validation Status	Main Metric
Czirok et al. (2026) [42]	Lameness	Side-view RGB post-milking video; still frames	699 cow videos; 12,182 images; >50,000 boxes	699	3	Faster R-CNN, SSD Inception v2 or YOLOv5m detectors + rule-based lameness score	Random pooled multi-farm 80/20 train/val; split unit: image	No farm-held-out validation	YOLOv5m mAP@0.5 73.2%; body-part AP ~64.1–78.6%; CPU 14 fps, GPU > 30 fps
Duan et al. (2025) [53]	Lameness	Overhead RGB-D walking sequences; RGB keypoints	741 sequences (260 sound/237 mild/244 severe); 593/148 split; 2520 keypoint images (2016/504)	NR (unique cows unreported)	1	AGC-SimCC HR-ACRNet keypoints + 6 gait features + PIMP-selected random forest	Random 80/20 sequence and keypoint-image splits; units: walking sequence/image; cow independence NR	No	Keypoints: PCK@0.02 100%; AP 95.89%; lameness accuracy 91%, macro-F1 0.85
Barney et al. (2023) [43]	Lameness	Side-view RGB video; Google Images for pose training	25 recordings/250 cow passages; 500 Google images; pose: 189 farm cows train, 61 held out	250 farm cows (189 train; 61 held out)	1 study farm + Google image sources	Modified Mask R-CNN pose + SORT tracking + CatBoost classifier	Pose held out 61 cows; CatBoost 3-fold CV; split unit: cow	No external farm	4-grade accuracy 94 ± 0.5%; κ 0.8782; precision 0.865; recall 0.9209; binary up to 100%/98%
Jia et al. (2025) [51]	Lameness	Side-view RGB walking video	127 videos/15,165 frames; pose 300 images/8 cows (90/10); pool 143 videos incl. 16 other-farm	45 self-collected; other-farm count NR	≥2	DeepLabCut dlcnnet_ms keypoints + temporal features + rule-based scoring	Pose 90/10 image split; final 20 videos randomly sampled from 143; units: image/video	Partial/unclear: other-farm test composition unreported	Accuracy 90.21%; keypoint error 4.679 px; class accuracies 89.0/85.3/92.6/100%
Jiang et al. (2020) [44]	Lameness	Side-view RGB walking video + optical flow	1080 videos (12/cow); 756 train, 324 test	90	1	Single-stream long-term optical-flow CNN with temporal modeling	Random 70/30 train/test; split unit: video; cow independence NR	No	Augmented clips: accuracy 95.13%, mAP 98.24%; best optical-flow configuration ~98.32%
Kang et al. (2020) [50]	Lameness	Side-view RGB walking video	100 cows: 75 threshold development, 25 independent validation; detector 1000 train/500 test images	100 (75 development; 25 validation)	1	RFB-Net/SSD leg detector + supporting-phase/lame-hoof rules	Independent 25-cow endpoint holdout; detector image train/test; units: cow (endpoint), image (detector)	No external farm; independent 25-cow same-farm holdout	Lameness accuracy 96%; lame-hoof localization 93%; detector mAP 87% at 83.3 fps
Kang et al. (2022) [45]	Lameness	Side-view RGB walking video; derived gait maps	456 videos; 3045 hoof images (70/20/10); 23,994 augmented gait maps (60/20/20)	456 usable (500 initially selected)	1	YOLOv4 hoof detector + reduced gait maps + DenseNet classifier	Image/video/map train/val/test splits; units: image/video/gait map; cow grouping of augmented maps unclear	No	Hoof detector mAP 92.39% at 19.96 fps; lameness accuracy 98.50%
Kang et al. (2025) [46]	Lameness	Side-view RGB video; gait/back feature maps	3150 gait maps/175 cows at Farm 1; 5309 hoof images; external Farm-2 test: 30 cows	205 total (175 development; 30 external)	2	SAM + CLFM gait/back maps + DenseNet with DCAM/DCAC attention	Primary 80/20 split plus independent 30-cow Farm-2 test; split unit: cow/map; external unit: farm	Yes: independent second-farm test	Primary severity acc./sens./spec. 92.80/89.21/94.60%; binary 99.05%; external accuracy 89.20%
Myint et al. (2024) [54]	Lameness	Side-view RGB video and extracted images	5458 detection images; Jan test 247,250 instances; classifier 1031 images (690/341)	116 classifier unique count NR	2; only 1 in quantitative classification	Mask R-CNN/YOLOv8 + IoU tracking + back-curvature features + classical classifiers	January detection holdout; classifier 690/341 image split; Farm-2 qualitative detection; units: time/image	Partial: second-farm detection qualitative only	AdaBoost lameness accuracy 77.9%; detection/tracking ~99%
Narli et al. (2025) [52]	Lameness	Side-view RGB-D video; annotated RGB frames	337 videos (~78 frames each); 7650 annotated frames from 260 cows	260	4	DeepLabCut 12 back keypoints + fully connected DNN lameness classifier	Pose 90/10 frame split; classifier 5-fold CV on pooled frames; unit: frame/sample; cow/farm holdout NR	No explicit held-out-farm test	Binary accuracy/sens./spec. 97.0/94.1/100; adjacent-score accuracies 0.88/0.83/0.93
Szyc et al. (2025) [124]	Lameness	Side-view RGB walking video, 960 × 540, 30 fps	832 videos; binary subset 638 (540 healthy, 98 lame)	832 reported cow videos/cows	5	YOLOv5 + MMPose; anatomical features; rules, RF, XGBoost and neural networks	Leave-one-out on pooled 5-barn videos; split: video/sample; no barn holdout	No barn-held-out validation	XGBoost acc./recall/precision 0.821/0.704/0.670; DL acc. 0.813; 7-grade RF relaxed/exact 0.736/0.237
Tun et al. (2024) [47]	Lameness	Top-view depth images/video	4944 frames; det. train/val 4302/915; classifier dev 45/31 and test 229/116 non-lame/lame	NR; 56–64 cows/session	1	Detection2 + IoU tracking; back features; RF, KNN, and decision tree	Random 80/20 frame split for detection; later-period same-farm classifier test; split: frame/time	No; temporal same-farm holdout	RF/KNN/tree test acc. 81.1/78.2/69.2%; detection/tracking 99.94/99.92%
Tun et al. (2026) [48]	Lameness	Overhead ToF depth images; 7-frame sequences	Detection 9550 cow + 988 background/290 human/255 noise; lameness 2056 seq. (1644/412)	93 Farm-A cows; broader/Farm-B unique NR	2	YOLOv11m/v8m/Mask R-CNN; PTAV3; height processing; EfficientNet-B7 + LSTM	Methods: 80/20 by cow; limitations: sequence split/not cow-independent; split unit unresolved; Farm A endpoint	Partial: 2-farm component tests; no unseen-farm endpoint	Box/mask AP50 99.38/99.26% (75.49 fps); lameness acc./F1 95.95/96.06%; tracking 99.94%
Wu et al. (2020) [49]	Lameness	Side-view RGB walking video	750 videos (15/cow); 700 model samples; 5000 labeled leg-detection images	50	1	YOLOv3 leg detector; step-size vectors; LSTM and bidirectional LSTM	Detector images 50/50; 700 videos 70/30 + 2–10-fold CV; split: image/video; cow grouping NR	No	YOLOv3 mAP/P/R 93.73/99.18/97.51% at 21 fps; 10-fold LSTM/BiLSTM acc. 98.57/99.05%
Siachos et al. (2025) [122]	Lameness	Overhead RGB video; keypoint trajectories and ID	40,116 paired scores/42 sessions; lesion 2515 cows (758 human-scored); pretrain 1986; follow-up 143	NR overall; 2515 in lesion subset	11	Commercial CNN keypoint/tracking pipeline; 0–100 score; mixed models and ROC	11-farm field evaluation vs. blinded vet scores/lesions; no new split; unit: cow/lesion; overlap NR	Unclear: 11-farm evaluation; training-farm overlap NR	Human agreement 81.5–86.3%; κ 0.23–0.38; AC1 0.76–0.83; severe lesions sens./spec./acc. 0.53/0.74/0.71
Rodríguez Alvarez et al. (2018) [30]	BCS	Depth images	1661 images; primary split 1158/503; farm holdouts 1000/411/250	NR	3	SqueezeNet CNN	Random 70/30 image split plus 3 leave-one-farm-out tests; split: image (primary), farm (cross-site)	Yes; unseen-farm internal cross-site validation	Exact 40%; within ±0.25/±0.50: 78/94%; farm-holdout results reported

Table A3. Cont.

Study	Operation Category	Sensor/Input Modality	Dataset Size	No. of Animals	No. of Farms/Sites	Model Family	Validation Method	External Validation Status	Main Metric
Lewis et al. (2025) [104]	BCS	2D RGB images: front, rear and top views	1764 images (521/648/595 by BCS); later final dataset 705 (210/315/180): internal conflict	56	1	YOLOv8 view-specific detection/classification models	Stratified 70/20/10 image split (1234/353/177); 3 repeated runs, not k-fold; split: image	No	Best mAP@.5 rear/front/top 0.439/0.432/0.421; mAP@.5: 0.95 0.41
Li et al. (2025) [100]	BCS	Overhead-oblique RGB video; hindquarter frames	1611 original images (6444 augmented); 10 videos/240 cow passages	213 development; video-test unique NR	1 quantified + 1 qualitative-only	Improved YOLOv5s (DW-CDPN) + spatiotemporal tracking/segmentation	Random 70/30 image split after augmentation + 10-video test; split: image/video; cow holdout NR	No quantified validation; second-farm examples only	mAP@0.5 91.8%; overall video BCS accuracy 87.1%
Nagy et al. (2023) [105]	BCS	Rear/rump-view RGB images	4061 images: 2922 train, 732 validation, 407 held-out Farm-3 test	NR	3	COCO-pretrained Detectron2 Faster R-CNN; 12-class and aggregated 3-class BCS	Farms 1–2 stratified 80/20 by image; Farm 3 held out (407 images); split: image, then farm	Yes; held-out Farm 3	Accuracy and κ ; 3-class models $\geq$ moderate agreement; some $\kappa \geq 0.8$; no single headline value
Rodríguez Alvarez et al. (2019) [98]	BCS	Top-view depth images	1661 images: 1158 train, 503 test/validation	NR	NR	SqueezeNet CNNs, VGG16 transfer learning and weighted ensembles	Single 70/30 image-level train/test split; split: image	No	Best ensemble exact 41.15%; within ± 0.25 82%; within ± 0.50 97.42%
Shi et al. (2023) [101]	BCS	Top-view depth-derived 3D point clouds + normals	3660 depth/point-cloud samples from 512 cows	512	2	Attention-guided PointNet++ with normals; 9-class softmax	Train/test evaluation; ratio and split unit NR; no farm holdout	No	Exact accuracy 0.49; within ± 0.25 0.80; within ± 0.50 0.96
Siachos et al. Alvarez et al. (2024) [119]	BCS	Overhead RGB video/images; pose and silhouette	34,150 manual BCS labels/35 visits; later 9657 pairs, including 3817 cows scored twice	NR development; 3817 repeated-score test	7 listed; caption says 8 (conflict)	EfficientNetV2 ordinal regressor + CattleEye detection/pose	Development: within-farm 80/20 cow split; temporal test reused 4 development farms; split: cow/time	No; temporal test reused development farms	Temporal κ w 0.69; exact 44.4%; within ± 0.25 ± 0.50 : 84.6/94.8%
Yao et al. (2026) [120]	BCS	Side-view 2D RGB surveillance images	3208 labeled images/211 cows; ~24 h source video	211	1	YOLOv11n detector + ConvNeXt-Tiny continuous BCS regressor	Stratified 5-fold cow-level CV; all images of each cow stayed together; split: cow	No	MAE 0.41 $\pm$ 0.02; RMSE 0.51 $\pm$ 0.02; Pearson r 0.62 $\pm$ 0.03; detector mAP 0.995
Feng et al. (2024) [99]	BCS	Rear/tail-region 2D RGB video frames	8972 cow-tail images from 24 videos (~90 min)	NR	1	ShuffleNet-SEH: ShuffleNetV2 + SE attention + H-Swish; 5.64 MB	Random 9:1 split stratified by BCS; no cow separation reported; split: image	No	Standard acc./precision/recall/F1 98.2/98.5/98.3/98.3%; blur/dim/bright acc. 95.5/96.1/83.1%
Summerfield et al. (2023) [102]	BCS	Three synchronized depth views: top, rear, angled	Frames: 2053 angled/2211 rear/2016 top (≤ 7 /cow); test 137 cows, augmented to 548	462	Origin NR; 1 collection facility	Small CNNs; single-view and early/middle/late multi-camera fusion	70/30 cow-level split; test cows disjoint; split: cow	No; one collection site	All-camera exact 35.77%; within ± 0.25 69.89%; within ± 0.50 89.96%
Zhao et al. (2023) [103]	BCS	Top-view depth converted to point clouds/convex-hull maps	5119 stated; train 2627/39 cows + test 2504/38 cows = 5131 (conflict)	77	1	Two-level EfficientNet + convex-hull maps; whale-optimized scaling	~1:1 cow-level split: 39 train, 38 test, no crossover; split: cow	No	Exact 45.01%; within ± 0.25 91.21%; within ± 0.50 97.60%; 3.446 frames/s
Dandl et al. (2024) [106]	BCS	Rear/pelvic-region 2D RGB photographs	1270 images (925 Holstein, 345 Simmental); 1144 train, 126 test	NR	20	YOLOv8x five-class BCS detector/classifier	Random 90/10 image split; no cow/farm grouping reported; split: image	No	102/126 correct (acc. 0.81); mean PR-AUC ~0.87; Holstein/Simmental acc. 0.82/0.80
Bi et al. (2023) [28]	Body weight	Top-view depth video/maps	40,405 depth images from 10–12 s recordings; ~50.1 GB raw data	12	1	Mask R-CNN + geometric measures; OLS, ridge, LASSO, and LMM regressions	Forecast time splits 90/10–50/50; exhaustive leave-3-cows-out (9/3); split: time and cow	No	Forecast LMM R^2 0.98, MAPE 2.03%; leave-3-cows-out OLS mean R^2 ~0.87, MAPE 4.53%
Afridi et al. (2024) [109]	Body weight	RGB, depth, RGB-D, and SAM-segmented variants	1289 paired observations: Farm 1 613, Farm 2 676	NR	2	SAM masking; InceptionV3, MobileNet and EfficientNetB1 regressors	80/20 + 5-fold CV on pooled observations; split: image pair; animal grouping NR; no farm holdout	No	Best: InceptionV3 RGB; MAE 15.06 kg, RMSE 20.01 kg, MAPE 3.57%, R^2 0.97
Dohmen et al. (2021) [110]	Body weight	Side/top RGB frames $\rightarrow$ segmented binary images	63 enrolled; 61 analyzed; 1 final image per cow/view	63 enrolled; 61 analyzed	1	Mask R-CNN foreground segmentation + custom CNN mass regressor	60/20/20 train/test/validation; one image per cow/view, so sample = animal; split: animal/image	No	Top R^2 0.96, RMSE 20 kg; side R^2 0.91, RMSE 27 kg
Gebreyesus et al. (2023) [107]	Body weight	3D contour features + RFID/walk-over weights	83,011 matched 3D-contour/weight records from 914 cows	914	3	CatBoost/AdaBoost/RF/tree; linear, ridge, LASSO, and SV regressions	Time and stricter-time splits; random-animal holdout kept 20% cows out; split: time/cow; no farm holdout	No; no unseen-farm test	Best Pearson r 0.94; MAPE 4.0%; RMSE 33 kg
Guvencoglu (2023) [111]	Body weight	Side/rear stereo RGB; mask-pixel and distance features	85 animals; 170 stereo pairs/340 images; one 85-row feature dataset	85	NR	DeepLabv3+ segmentation; 3 fully connected ANNs (side, rear, combined)	Random 90/10 animal-row split + 10-fold CV of combined ANN; split: animal row	No	Generally within ± 20 kg; no MAE/RMSE; CV mean actual/predicted 411.95/416.403 kg
Guzhva et al. (2026) [108]	Body weight	Handheld RGB + LiDAR range; silhouette/body measures	2847 images/9 rounds; 567 manual masks; weight set 1080 Batch-9 images, 216 grouped holdout	270 stated (248 + 22); later says 248 unique	2 research facilities	Mask R-CNN/ResNet-101; GMM weight classes; Extra Trees with fold-internal SMOTE	Segmentation cow split 75/15/10; weight grouped 80/20 + grouped CV; untouched 216-image holdout; split: cow	No; two-site development, no unseen-site test	Extra Trees macro-F1 0.936 (95% CI 0.860–0.983), weighted-F1 0.957, acc. 0.958; segmentation F1 0.98

Table A3. Cont.

Study	Operation Category	Sensor/Input Modality	Dataset Size	No. of Animals	No. of Farms/Sites	Model Family	Validation Method	External Validation Status	Main Metric
Liao et al. (2025) [121]	Body weight	Top-view depth video/images + geometric features	Trial 1: 196 records / 20 calves; Trial 2: 3286 videos / 48 calves, 5751 images (4794/957)	68 total (20 Trial 1; 48 Trial 2)	1	YOLOv8 vs. threshold segmentation; LR/XGBoost; longitudinal LMM	Segmentation held out 13/48 calves; weight 5-fold calf CV $\times 100$; longitudinal 90/10–50/50 $\times 100$; units: calf/time	No	IoU ~ 0.98 ; XGBoost R^2 0.91, RMSE 7.86 lb, MAE 6.22 lb, MAPE 4.37%; LMM R^2 0.99, MAPE 2.39%
Xu et al. (2024) [112]	Body weight	Top/rear RGB + masks + body area, age and range	550 top-view + 550 rear-view images from 55 cows	55	NR (private farms, plural)	Attention ResNet-101-D segmenter; MLR/tree/SVM/GPR/BP neural-net regressors	Random 80/20 image split; split: image; cow/farm grouping NR	No	BP test R^2 0.98; RMSE 22.73 lb; MSE 516.59 lb ² ; MAE 13.11 lb
Zhang et al. (2026) [113]	Body weight	Top/side depth $\rightarrow$ 3D point clouds + semantic labels	600 dual-view sets/300 cattle; segmentation 1029 point clouds, augmented to 2058	300	1 breeding center	CSS-PointNet++/CBAM; semantic reconstruction; BPNN/RBFNN/RF/MLR regressors	Segmentation 3:1:1 sample split, no overlap; weight sets 2:1; split: point-cloud/sample; animal grouping NR	No	Calf/adult MAE 1.84/12.58 kg; RMSE 2.49/14.45 kg; MRE 1.49/1.75%; R^2 0.920/0.863; OA/mIoU 98.35/95.61%
Hanpinitasak et al. (2026) [64]	Reproductive	Synchronized top/front RGB CCTV video (570 $\times$ 960, 10 fps)	6671 labeled boxes (5532 behavior; 1139 identity); continuous test 17 h 16 min	3	1	View-specific YOLOv8 ID/behavior detectors + IoU association + rule-based two-view fusion	Internal train/validation/test split (proportions/grouping NR); continuous same-site, same-cow test	No; same farm and three-cow cohort	Estrus F1: top 0.133, front 0.435, two-view ensemble 0.481
Li et al. (2024) [26]	Reproductive	Internet-sourced RGB images/video frames of cow mounting	6483 images; base split 4048/1012/1187; training augmented to 10,120; challenge test 647	NR	NR	Cow-YOLO; YOLOv5s + GCNet/Swin blocks, BiFPN/SPPF and coordinate attention	Internal random image split; occlusion-focused tests derived from test pool plus 236 images	No; no independent farm/site test	Precision 99.7%, recall 99.5%, mAP 99.5%, 156.3 fps; challenge mAP 99.5%
Giannone et al. (2025) [58]	Feeding monitoring	Fixed 2D RGB video; feeding-lane ROI	7235 frames; 70/20/10 split	19	1	Fine-tuned YOLOv8n individual-cow detector and ROI occupancy analysis	Frame-level 70/20/10 split of consecutive video images; same-farm evaluation; no cow- or farm-held-out test	No; same-farm evaluation only	$p = 0.85$; $R = 0.62$; $F1 = 0.72$; feeding-time accuracy ≈ 0.88
Wang et al. (2024) [62]	Feeding monitoring	RGB video from three camera positions (1280 $\times$ 720, 30 fps)	30 videos; 4015 images (3209/401/405 train/validation/test); 9 chew-count videos	NR; 1–3 visible/video	1	Improved YOLOv5s (C3_CA, Slim_Neck, WIoU) + DeepSORT jaw tracking	Internal frame split; detector/tracker tests; manual chew-count comparison on 9 videos	No; single farm/date, no independent herd test	Upper/lower jaw mAP@0.5 97.5/97.9%; chew-count false detection 3.08%

Notes: Study labels use manuscript citations only. BCS = body condition score; ID = individual identification; NR = not reported; N/A = not applicable. External validation requires an independent farm/site or external source without retraining or fine-tuning.

Table A4. Context-only general AI/CV/ML technical-method publications ($n = 14$).

Study	Operation Category	Sensor/Input Modality	Dataset Size	No. of Animals	No. of Farms/Sites	Model Family/Technical Role	Validation Method	External Validation Status	Main Metric
He et al. (2016) [9]	Context only	General benchmark images	context-only	N/A	N/A	Residual CNN image-classification backbone (ResNet)	N/A—method/context paper	N/A	context-only
Dosovitskiy et al. (2021) [10]	Context only	General benchmark images	context-only	N/A	N/A	Vision Transformer image-classification backbone (ViT)	N/A—method/context paper	N/A	context-only
Bochkovskiy et al. (2020) [11]	Context only	General images/video frames	context-only	N/A	N/A	Real-time one-stage object detector (YOLOv4)	N/A—method/context paper	N/A	context-only
Liu et al. (2016) [14]	Context only	General benchmark images	context-only	N/A	N/A	One-stage object detector (SSD)	N/A—method/context paper	N/A	context-only
Redmon et al. (2016) [12]	Context only	General images/video frames	context-only	N/A	N/A	Real-time one-stage object detector (original YOLO)	N/A—method/context paper	N/A	context-only
Redmon and Farhadi (2018) [13]	Context only	General images/video frames	context-only	N/A	N/A	One-stage object detector (YOLOv3)	N/A—method/context paper	N/A	context-only
He et al. (2017) [15]	Context only	General images/video frames	context-only	N/A	N/A	Instance detection and segmentation (Mask R-CNN)	N/A—method/context paper	N/A	context-only
Cao et al. (2017) [17]	Context only	General images/video frames	context-only	N/A	N/A	Multi-subject 2D pose/keypoint estimation (OpenPose)	N/A—method/context paper	N/A	context-only
Mathis et al. (2018) [18]	Context only	User-labeled animal images/video frames	context-only	N/A	N/A	Markerless pose estimation of user-defined body parts (DeepLabCut)	N/A—method/context paper	N/A	context-only
Aharon et al. (2022) [19]	Context only	Video detections and tracks	context-only	N/A	N/A	Multi-object tracking with association and re-identification (BoT-SORT)	N/A—method/context paper	N/A	context-only
Bewley et al. (2016) [20]	Context only	Video detections and tracks	context-only	N/A	N/A	Simple online real-time multi-object tracking (SORT)	N/A—method/context paper	N/A	context-only
Wojke et al. (2017) [21]	Context only	Video detections and tracks	context-only	N/A	N/A	Multi-object tracking with deep appearance association (Deep SORT)	N/A—method/context paper	N/A	context-only

Table A4. Cont.

Study	Operation Category	Sensor/Input Modality	Dataset Size	No. of Animals	No. of Farms/Sites	Model Family/Technical Role	Validation Method	External Validation Status	Main Metric
Zhang et al. (2022) [22]	Context only	Video detections and tracks	context-only	N/A	N/A	Multi-object tracking using high- and low-confidence detections (ByteTrack)	N/A—method/context paper	N/A	context-only
Amershi et al. (2019) [126]	Context only	No sensor input	context-only	N/A	N/A	Qualitative case study of the ML software-engineering lifecycle, covering development, testing, deployment, and maintenance	N/A—method/context paper	N/A	context-only

Note: These publications were retained solely for technical or contextual positioning and were not assigned to a dairy cow operation or included in the operation-level synthesis. Operation-specific fields are marked N/A because they do not apply to these context-only records.

References

1. Stygar, A.H.; Gómez, Y.; Berteselli, G.V.; Dalla Costa, E.; Canali, E.; Niemi, J.K.; Llonch, P.; Pastell, M. A Systematic Review on Commercially Available and Validated Sensor Technologies for Welfare Assessment of Dairy Cattle. *Front. Vet. Sci.* **2021**, *8*, 634338. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Tucker, C.B.; Jensen, M.B.; de Passillé, A.M.; Hänninen, L.; Rushen, J. Invited review: Lying time and the welfare of dairy cows. *J. Dairy Sci.* **2021**, *104*, 20–46. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Distant, D.; Albanello, C.; Zaffar, H.; Faralli, S.; Amalfitano, D. Artificial intelligence applied to precision livestock farming: A tertiary study. *Smart Agric. Technol.* **2025**, *11*, 100889. [\[CrossRef\]](#)
4. Lovarelli, D.; Bacenetti, J.; Guarino, M. A review on dairy cattle farming: Is precision livestock farming the compromise for an environmental, economic and social sustainable production? *J. Clean. Prod.* **2020**, *262*, 121409. [\[CrossRef\]](#)
5. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Guarnido-Lopez, P.; Pi, Y.; Tao, J.; Mendes, E.D.M.; Tedeschi, L.O. Computer vision algorithms to help decision-making in cattle production. *Anim. Front.* **2024**, *14*, 11–22. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Palma, O.; Plà-Aragón, L.M.; Mac Cawley, A.; Alborno, V.M. AI and Data Analytics in the Dairy Farms: A Scoping Review. *Animals* **2025**, *15*, 1291. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [\[CrossRef\]](#)
9. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [\[CrossRef\]](#)
10. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. In Proceedings of the International Conference on Learning Representations 2021, Vienna, Austria, 3–7 May 2021.
11. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [\[CrossRef\]](#)
12. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [\[CrossRef\]](#)
13. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767. [\[CrossRef\]](#)
14. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016*; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; pp. 21–37. [\[CrossRef\]](#)
15. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988. [\[CrossRef\]](#)
16. Wang, J.; Zhang, X.; Gao, G.; Lv, Y.; Li, Q.; Li, Z.; Wang, C.; Chen, G. Open Pose Mask R-CNN Network for Individual Cattle Recognition. *IEEE Access* **2023**, *11*, 113752–113768. [\[CrossRef\]](#)
17. Cao, Z.; Simon, T.; Wei, S.-E.; Sheikh, Y. Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 22–25 July 2017; pp. 7291–7299. [\[CrossRef\]](#)
18. Mathis, A.; Mamidanna, P.; Cury, K.M.; Abe, T.; Murthy, V.N.; Mathis, M.W.; Bethge, M. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **2018**, *21*, 1281–1289. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Aharon, N.; Orfaig, R.; Bobrovsky, B.-Z. BoT-SORT: Robust Associations Multi-Pedestrian Tracking. *arXiv* **2022**, arXiv:2206.14651. [\[CrossRef\]](#)
20. Bewley, A.; Ge, Z.; Ott, L.; Ramos, F.; Upcroft, B. Simple Online and Realtime Tracking. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3464–3468. [\[CrossRef\]](#)

21. Wojke, N.; Bewley, A.; Paulus, D. Simple Online and Realtime Tracking with a Deep Association Metric. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 3645–3649. [\[CrossRef\]](#)
22. Zhang, Y.; Sun, P.; Jiang, Y.; Yu, D.; Weng, F.; Yuan, Z.; Luo, P.; Liu, W.; Wang, X. ByteTrack: Multi-Object Tracking by Associating Every Detection Box. In *Computer Vision—ECCV 2022*; Springer Nature: Cham, Switzerland, 2022; pp. 1–21. [\[CrossRef\]](#)
23. Fuentes, A.; Yoon, S.; Park, J.; Park, D.S. Deep learning-based hierarchical cattle behavior recognition with spatio-temporal information. *Comput. Electron. Agric.* **2020**, *177*, 105627. [\[CrossRef\]](#)
24. Wu, D.; Wang, Y.; Han, M.; Song, L.; Shang, Y.; Zhang, X.; Song, H. Using a CNN-LSTM for basic behaviors detection of a single dairy cow in a complex environment. *Comput. Electron. Agric.* **2021**, *182*, 106016. [\[CrossRef\]](#)
25. Bai, Q.; Gao, R.; Li, Q.; Wang, R.; Zhang, H. Recognition of the behaviors of dairy cows by an improved YOLO. *Intell. Robot.* **2024**, *4*, 1–19. [\[CrossRef\]](#)
26. Li, D.; Wang, J.; Zhang, Z.; Dai, B.; Zhao, K.; Shen, W.; Yin, Y.; Li, Y. Cow-YOLO: Automatic cow mounting detection based on non-local CSPDarknet53 and multiscale neck. *Int. J. Agric. Biol. Eng.* **2024**, *17*, 193–202. [\[CrossRef\]](#)
27. Tong, L.; Fang, J.; Wang, X.; Zhao, Y. Research on cattle behavior recognition and multi-object tracking algorithm based on YOLO-BoT. *Animals* **2024**, *14*, 2993. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Bi, Y.; Campos, L.M.; Wang, J.; Yu, H.; Hanigan, M.D.; Morota, G. Depth video data-enabled predictions of longitudinal dairy cow body weight using thresholding and Mask R-CNN algorithms. *Smart Agric. Technol.* **2023**, *6*, 100352. [\[CrossRef\]](#)
29. Higaki, S.; Menezes, G.L.; Ferreira, R.E.P.; Negreiro, A.; Cabrera, V.E.; Dórea, J.R.R. Objective dairy cow mobility analysis and scoring system using computer vision-based keypoint detection technique from top-view 2-dimensional videos. *J. Dairy Sci.* **2025**, *108*, 3942–3955. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Rodríguez Alvarez, J.; Arroqui, M.; Mangudo, P.; Toloza, J.; Jatip, D.; Rodríguez, J.M.; Teyseyre, A.; Sanz, C.; Zunino, A.; Machado, C.; et al. Body condition estimation on cows from depth images using Convolutional Neural Networks. *Comput. Electron. Agric.* **2018**, *155*, 12–22. [\[CrossRef\]](#)
31. Xiao, J.; Liu, G.; Wang, K.; Si, Y. Cow identification in free-stall barns based on an improved Mask R-CNN and an SVM. *Comput. Electron. Agric.* **2022**, *194*, 106738. [\[CrossRef\]](#)
32. Besler, B.C.; Mojabi, P.; Lasemiemi, Z.; Murphy, J.E.; Wang, Z.; Baker, R.; Pearson, J.M.; Fear, E.C. Scoping review of precision technologies for cattle monitoring. *Smart Agric. Technol.* **2024**, *9*, 100596. [\[CrossRef\]](#)
33. Espinoza-Sandoval, O.R.; Angeles-Hernandez, J.C.; Gonzalez-Ronquillo, M.; Ghavipanje, N.; Zhang, N.; Bayat, A.R.; Hervás, G.; Kholif, A.E.; Mele, M.; Loo, J.J.; et al. Dairy farming in the era of artificial intelligence: Trend or a real game changer? *J. Dairy Res.* **2024**, *91*, 139–145. [\[CrossRef\]](#)
34. Mahmud, M.S.; Zahid, A.; Das, A.K.; Muzammil, M.; Khan, M.U. A systematic literature review on deep learning applications for precision cattle farming. *Comput. Electron. Agric.* **2021**, *187*, 106313. [\[CrossRef\]](#)
35. Kang, X.; Zhang, X.D.; Liu, G. A Review: Development of Computer Vision-Based Lameness Detection for Dairy Cows and Discussion of the Practical Applications. *Sensors* **2021**, *21*, 753. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Yuan, L.; Li, W.; Yue, J.; Li, Z.; Zhao, Y.; Li, Q.; Feng, W.; Kou, G. A review of cattle behavior recognition methods based on computer vision and deep learning. *Smart Agric. Technol.* **2026**, *14*, 101867. [\[CrossRef\]](#)
37. Eckelkamp, E.A. Invited Review: Current state of wearable precision dairy technologies in disease detection. *Appl. Anim. Sci.* **2019**, *35*, 209–220. [\[CrossRef\]](#)
38. Riaboff, L.; Shalloo, L.; Smeaton, A.F.; Couvreur, S.; Madouasse, A.; Keane, M.T. Predicting livestock behaviour using accelerometers: A systematic review of processing techniques for ruminant behaviour prediction from raw accelerometer data. *Comput. Electron. Agric.* **2022**, *192*, 106610. [\[CrossRef\]](#)
39. Qiao, Y.; Clark, C.; Lomax, S.; Kong, H.; Su, D.; Sukkarieh, S. Automated Individual Cattle Identification Using Video Data: A Unified Deep Learning Architecture Approach. *Front. Anim. Sci.* **2021**, *2*, 759147. [\[CrossRef\]](#)
40. Ferreira, R.E.P.; Dórea, J.R.R. International Symposium on Ruminant Physiology: Leveraging computer vision, large language models, and multimodal machine learning for optimal decision making in dairy farming. *J. Dairy Sci.* **2025**, *108*, 7493–7510. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Abdul Jabbar, K.; Hansen, M.F.; Smith, M.L.; Smith, L.N. Early and non-intrusive lameness detection in dairy cows using 3-dimensional video. *Biosyst. Eng.* **2017**, *153*, 63–69. [\[CrossRef\]](#)
42. Czirik, M.; Kovács, L.; Póti, P.; Lecsek, N.; Lengyel, J.; Jurkovich, V. Development of a computer vision-aided lameness detection system for dairy cattle. *Smart Agric. Technol.* **2026**, *14*, 101952. [\[CrossRef\]](#)
43. Barney, S.; Dlay, S.; Crowe, A.; Kyriazakis, I.; Leach, M. Deep learning pose estimation for multi-cattle lameness detection. *Sci. Rep.* **2023**, *13*, 4499. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Jiang, B.; Yin, X.; Song, H. Single-stream long-term optical flow convolution network for action recognition of lameness dairy cow. *Comput. Electron. Agric.* **2020**, *175*, 105536. [\[CrossRef\]](#)
45. Kang, X.; Li, S.; Li, Q.; Liu, G. Dimension-reduced spatiotemporal network for lameness detection in dairy cows. *Comput. Electron. Agric.* **2022**, *197*, 106922. [\[CrossRef\]](#)

46. Kang, X.; Liang, J.; Li, Q.; Liu, G. Detecting Lameness in Dairy Cows Based on Gait Feature Mapping and Attention Mechanisms. *Agriculture* **2025**, *15*, 1276. [\[CrossRef\]](#)
47. Tun, S.C.; Onizuka, T.; Tin, P.; Aikawa, M.; Kobayashi, I.; Zin, T.T. Revolutionizing Cow Welfare Monitoring: A Novel Top-View Perspective with Depth Camera-Based Lameness Classification. *J. Imaging* **2024**, *10*, 67. [\[CrossRef\]](#)
48. Tun, S.C.; Tin, P.; Aikawa, M.; Kobayashi, I.; Zin, T.T. Cattle lameness detection using depth image and deep learning. *Sci. Rep.* **2026**, *16*, 13456. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Wu, D.; Wu, Q.; Yin, X.; Jiang, B.; Wang, H.; He, D.; Song, H. Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector. *Biosyst. Eng.* **2020**, *189*, 150–163. [\[CrossRef\]](#)
50. Kang, X.; Zhang, X.D.; Liu, G. Accurate detection of lameness in dairy cattle with computer vision: A new and individualized detection strategy based on the analysis of the supporting phase. *J. Dairy Sci.* **2020**, *103*, 10628–10638. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Jia, Z.; Zhao, Y.; Mu, X.; Liu, D.; Wang, Z.; Yao, J.; Yang, X. Intelligent Deep Learning and Keypoint Tracking-Based Detection of Lameness in Dairy Cows. *Vet. Sci.* **2025**, *12*, 218. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Narli, S.S.; Schmidt, H.; Firouzabadi, A.; Schönnagel, L.; Reich, M.S.; Reitmaier, S. Automated detection of lameness in dairy cattle through computer vision analysis of back shape characteristics. *Comput. Biol. Med.* **2025**, *197*, 111038. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Duan, W.; Wang, F.; Li, H.; Liu, N.; Fu, X. Lameness detection in dairy cows from overhead view: High-precision keypoint localization and multi-feature fusion classification. *Front. Vet. Sci.* **2025**, *12*, 1675181. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Myint, B.B.; Onizuka, T.; Tin, P.; Aikawa, M.; Kobayashi, I.; Zin, T.T. Development of a real-time cattle lameness detection system using a single side-view camera. *Sci. Rep.* **2024**, *14*, 13734. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Thomsen, P.T.; Shearer, J.K.; Houe, H. Prevalence of lameness in dairy cows: A literature review. *Vet. J.* **2023**, *295*, 105975. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Roche, S.M.; Renaud, D.L.; Saraceni, J.; Kelton, D.F.; DeVries, T.J. Invited review: Prevalence, risk factors, treatment, and barriers to best practice adoption for lameness and injuries in dairy cattle—A narrative review. *J. Dairy Sci.* **2024**, *107*, 3347–3366. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Russello, H.; van der Tol, R.; Holzhauer, M.; van Henten, E.J.; Kootstra, G. Video-based automatic lameness detection of dairy cows using pose estimation and multiple locomotion traits. *Comput. Electron. Agric.* **2024**, *223*, 109040. [\[CrossRef\]](#)
58. Giannone, C.; Sahraeibeverdy, M.; Lamanna, M.; Cavallini, D.; Formigoni, A.; Tassinari, P.; Torreggiani, D.; Bovo, M. Automated dairy cow identification and feeding behaviour analysis using a computer vision model based on YOLOv8. *Smart Agric. Technol.* **2025**, *12*, 101304. [\[CrossRef\]](#)
59. McDonagh, J.; Tzimiropoulos, G.; Slinger, K.R.; Huggett, Z.J.; Down, P.M.; Bell, M.J. Detecting Dairy Cow Behavior Using Vision Technology. *Agriculture* **2021**, *11*, 675. [\[CrossRef\]](#)
60. Yu, R.; Wei, X.; Liu, Y.; Yang, F.; Shen, W.; Gu, Z. Research on Automatic Recognition of Dairy Cow Daily Behaviors Based on Deep Learning. *Animals* **2024**, *14*, 458. [\[CrossRef\]](#) [\[PubMed\]](#)
61. Reith, S.; Hoy, S. Review: Behavioral signs of estrus and the potential of fully automated systems for detection of estrus in dairy cattle. *Animal* **2018**, *12*, 398–407. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Wang, N.; Li, X.; Shang, S.; Yun, Y.; Liu, Z.; Lyu, D. Monitoring Dairy Cow Rumination Behavior Based on Upper and Lower Jaw Tracking. *Agriculture* **2024**, *14*, 2006. [\[CrossRef\]](#)
63. Wu, D.; Han, M.; Song, H.; Song, L.; Duan, Y. Monitoring the respiratory behavior of multiple cows based on computer vision and deep learning. *J. Dairy Sci.* **2023**, *106*, 2963–2979. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Hanpinitsak, P.; Katanyukul, T.; Tonmitr, N.; Suntra, C.; Tanusilp, S.; Phuphaphud, A. Artificial intelligence-based dairy cattle behavior recognition for estrus detection via ensemble fusion of two camera views. *PLoS ONE* **2026**, *21*, e0340999. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Li, K.; Fan, D.; Wu, H.; Zhao, A. A new dataset for video-based cow behavior recognition. *Sci. Rep.* **2024**, *14*, 18702. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Achour, B.; Belkadi, M.; Filali, I.; Laghrouche, M.; Lahdir, M. Image analysis for individual identification and feeding behaviour monitoring of dairy cows based on Convolutional Neural Networks (CNN). *Biosyst. Eng.* **2020**, *198*, 31–49. [\[CrossRef\]](#)
67. Hua, Z.; Wang, Z.; Xu, X.; Kong, X.; Song, H. An effective PoseC3D model for typical action recognition of dairy cows based on skeleton features. *Comput. Electron. Agric.* **2023**, *212*, 108152. [\[CrossRef\]](#)
68. Li, C.; Ma, J.; Cao, S.; Guo, L. RFR-YOLO-Based Recognition Method for Dairy Cow Behavior in Farming Environments. *Agriculture* **2025**, *15*, 1952. [\[CrossRef\]](#)
69. Liu, Z.; Fang, J.; Zhao, Y. Multi-Object Detection of Forage Density and Dairy Cow Feeding Behavior Based on an Improved YOLOv10 Model for Smart Pasture Applications. *Sensors* **2026**, *26*, 1273. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Long, T.; Yu, R.; You, X.; Shen, W.; Wei, X.; Gu, Z. FSCA-YOLO: An Enhanced YOLO-Based Model for Multi-Target Dairy Cow Behavior Recognition. *Animals* **2025**, *15*, 2631. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Ma, S.; Zhang, Q.; Li, T.; Song, H. Basic motion behavior recognition of single dairy cow based on improved Rexnet 3D network. *Comput. Electron. Agric.* **2022**, *194*, 106772. [\[CrossRef\]](#)

72. Rao, S.; Garcia, E.; Neethirajan, S. Video-based cattle behaviour detection for digital twin development in precision dairy systems. *npj Vet. Sci.* **2026**, *1*, 3. [CrossRef] [PubMed]
73. Shang, C.; Wu, F.; Wang, M.; Gao, Q. Cattle behavior recognition based on feature fusion under a dual attention mechanism. *J. Vis. Commun. Image Represent.* **2022**, *85*, 103524. [CrossRef]
74. Li, X.; Wu, H.; Fan, D.; Bai, J.; Wang, C.; Liu, Y. A Clip-Based Dairy Cow Behavior Recognition Method Integrating Temporal Modeling and Behavioral Priors. *Animals* **2026**, *16*, 2087. [CrossRef] [PubMed]
75. Wu, C.; Fang, J.; Wang, X.; Zhao, Y. DMSF-YOLO: Cow Behavior Recognition Algorithm Based on Dynamic Mechanism and Multi-Scale Feature Fusion. *Sensors* **2025**, *25*, 3479. [CrossRef] [PubMed]
76. Yu, Z.; Liu, Y.; Yu, S.; Wang, R.; Song, Z.; Yan, Y.; Li, F.; Wang, Z.; Tian, F. Automatic Detection Method of Dairy Cow Feeding Behaviour Based on YOLO Improved Model and Edge Computing. *Sensors* **2022**, *22*, 3271. [CrossRef] [PubMed]
77. Zong, Z.; Ban, Z.; Wang, C.; Wang, S.; Yuan, W.; Zhang, C.; Su, L.; Yuan, Z. A Study on Multi-Scale Behavior Recognition of Dairy Cows in Complex Background Based on Improved YOLOv5. *Agriculture* **2025**, *15*, 213. [CrossRef]
78. Koskela, O.; Benitez Pereira, L.S.; Pölönen, I.; Aronen, I.; Kunttu, I. Deep learning image recognition of cow behavior and an open data set acquired near an automatic milking robot. *Agric. Food Sci.* **2022**, *31*, 89–103. [CrossRef]
79. Guzhva, O.; Ardo, H.; Nilsson, M.; Herlin, A.; Tufvesson, L. Now You See Me: Convolutional Neural Network Based Tracker for Dairy Cows. *Front. Robot. AI* **2018**, *5*, 107. [CrossRef] [PubMed]
80. Cui, Y.; Jia, M.; Lin, T.-Y.; Song, Y.; Belongie, S. Class-Balanced Loss Based on Effective Number of Samples. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 9260–9269. [CrossRef]
81. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [CrossRef] [PubMed]
82. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
83. Wang, Y.; Yao, Q.; Kwok, J.T.; Ni, L.M. Generalizing from a Few Examples: A Survey on Few-Shot Learning. *ACM Comput. Surv.* **2020**, *53*, 1–34. [CrossRef] [PubMed]
84. Pang, G.; Shen, C.; Cao, L.; van den Hengel, A. Deep Learning for Anomaly Detection: A Review. *ACM Comput. Surv.* **2021**, *54*, 1–38. [CrossRef]
85. Smink, M.; Liu, H.; Döpfer, D.; Lee, Y.J. Computer Vision on the Edge: Individual Cattle Identification in Real-Time with Read My Cow System. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2024; pp. 7056–7065. Available online: https://openaccess.thecvf.com/content/WACV2024/html/Smink_Computer_Vision_on_the_Edge_Individual_Cattle_Identification_in_Real-Time_WACV_2024_paper.html (accessed on 1 July 2026).
86. Sun, Y.; Huo, P.; Wang, Y.; Cui, Z.; Li, Y.; Dai, B.; Li, R.; Zhang, Y. Automatic monitoring system for individual dairy cows based on a deep learning framework that provides identification via body parts and estimation of body condition score. *J. Dairy Sci.* **2019**, *102*, 10140–10151. [CrossRef] [PubMed]
87. Hooker, J.; de Medeiros, B.; Saha, C.; Abdulrahman, T.; Alves, A. Multitask contrastive learning for individual dairy cow recognition across different behavior classes based on small image sets. *J. Dairy Sci.* **2026**, *109*, 1800–1815. [CrossRef] [PubMed]
88. Dac, H.H.; Gonzalez Viejo, C.; Lipovetzky, N.; Tongson, E.; Dunshea, F.R.; Fuentes, S. Livestock Identification Using Deep Learning for Traceability. *Sensors* **2022**, *22*, 8256. [CrossRef] [PubMed]
89. Kumar, S.; Pandey, A.; Sai Ram Satwik, K.; Kumar, S.; Singh, S.K.; Singh, A.K.; Mohan, A. Deep learning framework for recognition of cattle using muzzle point image pattern. *Measurement* **2018**, *116*, 1–17. [CrossRef]
90. Li, Z.; Lei, X.; Liu, S. A lightweight deep learning model for cattle face recognition. *Comput. Electron. Agric.* **2022**, *195*, 106848. [CrossRef]
91. Pathak, P.D.; Prakash, S. Attention-based multi-modal robust cattle identification technique using deep learning. *Comput. Electron. Agric.* **2025**, *238*, 110747. [CrossRef]
92. Ruchay, A.; Kolpakov, V.; Guo, H.; Pezzuolo, A. On-barn cattle facial recognition using deep transfer learning and data augmentation. *Comput. Electron. Agric.* **2024**, *225*, 109306. [CrossRef]
93. Shu, H.; Jin, Z.; Guo, G. Deep learning-based holstein face recognition in real-world farming conditions. *Smart Agric. Technol.* **2026**, *13*, 101690. [CrossRef]
94. Xu, X.; Wang, Y.; Shang, Y.; Yang, G.; Hua, Z.; Wang, Z.; Song, H. Few-shot cow identification via meta-learning. *Inf. Process. Agric.* **2025**, *12*, 80–90. [CrossRef]
95. Han, S.; Fuentes, A.; Yoon, S.; Jeong, Y.; Kim, H.; Park, D.S. Deep learning-based multi-cattle tracking in crowded livestock farming using video. *Comput. Electron. Agric.* **2023**, *212*, 108044. [CrossRef]
96. Mg, W.H.E.; Tin, P.; Aikawa, M.; Kobayashi, I.; Horii, Y.; Honkawa, K.; Zin, T.T. Customized Tracking Algorithm for Robust Cattle Detection and Tracking in Occlusion Environments. *Sensors* **2024**, *24*, 1181. [CrossRef] [PubMed]

97. Hossain, M.E.; Kabir, M.A.; Zheng, L.; Swain, D.L.; McGrath, S.; Medway, J. A systematic review of machine learning techniques for cattle identification: Datasets, methods and future directions. *Artif. Intell. Agric.* **2022**, *6*, 138–155. [\[CrossRef\]](#)
98. Rodríguez Alvarez, J.; Arroqui, M.; Mangudo, P.; Toloza, J.; Jatip, D.; Rodríguez, J.M.; Teyseyre, A.; Sanz, C.; Zunino, A.; Machado, C.; et al. Estimating Body Condition Score in Dairy Cows From Depth Images Using Convolutional Neural Networks, Transfer Learning and Model Ensembling Techniques. *Agronomy* **2019**, *9*, 90. [\[CrossRef\]](#)
99. Feng, T.; Guo, Y.; Huang, X.; Wu, J.; Cheng, C. A Method of Body Condition Scoring for Dairy Cows Based on Lightweight Convolution Neural Network. *J. ASABE* **2024**, *67*, 409–420. [\[CrossRef\]](#)
100. Li, J.; Zeng, P.; Yue, S.; Zheng, Z.; Qin, L.; Song, H. Automatic body condition scoring system for dairy cows in group state based on improved YOLOv5 and video analysis. *Artif. Intell. Agric.* **2025**, *15*, 350–362. [\[CrossRef\]](#)
101. Shi, W.; Dai, B.; Shen, W.; Sun, Y.; Zhao, K.; Zhang, Y. Automatic estimation of dairy cow body condition score based on attention-guided 3D point cloud feature extraction. *Comput. Electron. Agric.* **2023**, *206*, 107666. [\[CrossRef\]](#)
102. Summerfield, G.I.; De Freitas, A.; van Marle-Koster, E.; Myburgh, H.C. Automated Cow Body Condition Scoring Using Multiple 3D Cameras and Convolutional Neural Networks. *Sensors* **2023**, *23*, 9051. [\[CrossRef\]](#) [\[PubMed\]](#)
103. Zhao, K.; Zhang, M.; Shen, W.; Liu, X.; Ji, J.; Dai, B.; Zhang, R. Automatic body condition scoring for dairy cows based on EfficientNet and convex hull features of point clouds. *Comput. Electron. Agric.* **2023**, *205*, 107588. [\[CrossRef\]](#)
104. Lewis, R.; Kostermans, T.; Brovold, J.W.; Laique, T.; Ocepek, M. Automated Body Condition Scoring in Dairy Cows Using 2D Imaging and Deep Learning. *AgriEngineering* **2025**, *7*, 241. [\[CrossRef\]](#)
105. Nagy, S.Á.; Kilim, O.; Csabai, I.; Gábor, G.; Solymosi, N. Impact Evaluation of Score Classes and Annotation Regions in Deep Learning-Based Dairy Cow Body Condition Prediction. *Animals* **2023**, *13*, 194. [\[CrossRef\]](#) [\[PubMed\]](#)
106. Dandil, E.; Çevik, K.K.; Boğa, M. Automated Classification System Based on YOLO Architecture for Body Condition Score in Dairy Cows. *Vet. Sci.* **2024**, *11*, 399. [\[CrossRef\]](#) [\[PubMed\]](#)
107. Gebreyesus, G.; Milkevych, V.; Lassen, J.; Sahana, G. Supervised learning techniques for dairy cattle body weight prediction from 3D digital images. *Front. Genet.* **2023**, *13*, 947176. [\[CrossRef\]](#) [\[PubMed\]](#)
108. Guzhva, O.; Ternman, E.; Lindberg, M.; Telezhenko, E.; Kronqvist, C. PickAMoo: LIDAR-enhanced mask R-CNN segmentation for precision weight estimation in dairy cattle using smartphone imaging. *Sci. Rep.* **2026**, *16*, 16020. [\[CrossRef\]](#) [\[PubMed\]](#)
109. Afridi, H.; Ullah, M.; Nordbø, Ø.; Hoff, S.C.; Furre, S.; Larsgard, A.G.; Cheikh, F.A. Analyzing Data Modalities for Cattle Weight Estimation Using Deep Learning Models. *J. Imaging* **2024**, *10*, 72. [\[CrossRef\]](#) [\[PubMed\]](#)
110. Dohmen, R.; Catal, C.; Liu, Q. Image-based body mass prediction of heifers using deep neural networks. *Biosyst. Eng.* **2021**, *204*, 283–293. [\[CrossRef\]](#)
111. Guvenoglu, E. Determination of the Live Weight of Farm Animals with Deep Learning and Semantic Segmentation Techniques. *Appl. Sci.* **2023**, *13*, 6944. [\[CrossRef\]](#)
112. Xu, B.; Mao, Y.; Wang, W.; Chen, G. Intelligent weight prediction of cows based on semantic segmentation and back propagation neural network. *Front. Artif. Intell.* **2024**, *7*, 1299169. [\[CrossRef\]](#) [\[PubMed\]](#)
113. Zhang, J.; Wang, X.; Meng, H.; Huang, J.; Zhang, X.; Zhou, K.; Li, Y.; Peng, H. Semantically Guided 3D Reconstruction and Body Weight Estimation Method for Dairy Cows. *Agriculture* **2026**, *16*, 182. [\[CrossRef\]](#)
114. Wang, Z.; Shadpour, S.; Chan, E.; Rotondo, V.; Wood, K.M.; Tulpan, D. ASAS-NANP SYMPOSIUM: Applications of machine learning for livestock body weight prediction from digital images. *J. Anim. Sci.* **2021**, *99*, skab022. [\[CrossRef\]](#) [\[PubMed\]](#)
115. Qiao, Y.; Kong, H.; Clark, C.; Lomax, S.; Su, D.; Eiffert, S.; Sukkarieh, S. Intelligent perception for cattle monitoring: A review for cattle identification, body condition score evaluation, and weight estimation. *Comput. Electron. Agric.* **2021**, *185*, 106143. [\[CrossRef\]](#)
116. Qiao, Y.; Kong, H.; Clark, C.; Lomax, S.; Su, D.; Eiffert, S.; Sukkarieh, S. Intelligent Perception-Based Cattle Lameness Detection and Behaviour Recognition: A Review. *Animals* **2021**, *11*, 3033. [\[CrossRef\]](#) [\[PubMed\]](#)
117. Dohmen, R.; Catal, C.; Liu, Q. Computer vision-based weight estimation of livestock: A systematic literature review. *N. Z. J. Agric. Res.* **2022**, *65*, 227–247. [\[CrossRef\]](#)
118. Wang, B.; Li, X.; An, X.; Duan, W.; Wang, Y.; Wang, D.; Qi, J. Open-Set Recognition of Individual Cows Based on Spatial Feature Transformation and Metric Learning. *Animals* **2024**, *14*, 1175. [\[CrossRef\]](#) [\[PubMed\]](#)
119. Siachos, N.; Lennox, M.; Anagnostopoulos, A.; Griffiths, B.E.; Neary, J.M.; Smith, R.F.; Oikonomou, G. Development and validation of a fully automated 2-dimensional imaging system generating body condition scores for dairy cows using machine learning. *J. Dairy Sci.* **2024**, *107*, 2499–2511. [\[CrossRef\]](#) [\[PubMed\]](#)
120. Yao, L.; Kong, F.; Hong, W.; Liu, J.; Li, X.; Fan, Z.; Lei, L. Automated dairy cattle body condition score using side-view images and deep learning. *J. Dairy Sci.* **2026**; *in press*. [\[CrossRef\]](#) [\[PubMed\]](#)
121. Liao, M.; Morota, G.; Bi, Y.; Cockrum, R.R. Predicting Dairy Calf Body Weight from Depth Images Using Deep Learning (YOLOv8) and Threshold Segmentation with Cross-Validation and Longitudinal Analysis. *Animals* **2025**, *15*, 868. [\[CrossRef\]](#) [\[PubMed\]](#)

122. Siachos, N.; Griffiths, B.E.; Wilson, J.P.; Bedford, C.; Anagnostopoulos, A.; Neary, J.M.; Smith, R.F.; Oikonomou, G. Evaluation of a fully automated 2-dimensional imaging system for real-time cattle lameness detection using machine learning. *J. Dairy Sci.* **2025**, *108*, 4206–4224. [[CrossRef](#)] [[PubMed](#)]
123. Bai, Q.; Gao, R.; Wang, R.; Li, Q.; Yu, Q.; Zhao, C.; Li, S. X3DFast model for classifying dairy cow behaviors based on a two-pathway architecture. *Sci. Rep.* **2023**, *13*, 20519. [[CrossRef](#)] [[PubMed](#)]
124. Szyc, K.; Hebda, M.; Dembiński, K.; Zdunek, M.; Unold, O. Video-Based Automated Lameness Detection for Dairy Cows. *Sensors* **2025**, *25*, 5771. [[CrossRef](#)] [[PubMed](#)]
125. Li, S.; Fu, L.; Sun, Y.; Mu, Y.; Chen, L.; Li, J.; Gong, H. Individual dairy cow identification based on lightweight convolutional neural network. *PLoS ONE* **2021**, *16*, e0260510. [[CrossRef](#)] [[PubMed](#)]
126. Amershi, S.; Begel, A.; Bird, C.; DeLine, R.; Gall, H.; Kamar, E.; Nagappan, N.; Nushi, B.; Zimmermann, T. Software Engineering for Machine Learning: A Case Study. In Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP), Montréal, QC, Canada, 25–31 May 2019; pp. 291–300. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.