

Review

Not peer-reviewed version

Socializing AI: Integrating Social Network Analysis and Deep Learning for Precision Dairy Cow Monitoring—A Critical Review

[Sibi Chakravathy Parivendan](#) , [Kashfia Sailunaz](#) , [Suresh Neethirajan](#) *

Posted Date: 30 May 2025

doi: 10.20944/preprints202505.2441.v1

Keywords: dairy monitoring; social network analysis; precision livestock farming; computer vision; deep learning; cattle identification; behavior detection; animal welfare



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Review

Socializing AI: Integrating Social Network Analysis and Deep Learning for Precision Dairy Cow Monitoring—A Critical Review

Sibi Chakravathy Parivendan ¹, Kashfia Sailunaz ¹ and Suresh Neethirajan ^{1,2,*}

¹ Faculty of Computer Science, 6050 University Avenue, Dalhousie University, Halifax, NS B3H 4R2, Canada

² Faculty of Agriculture, Dalhousie University, Truro, NS B3H 4R2, Canada

* Correspondence: sneethir@gmail.com

Simple Summary: Dairy cows are social animals that form relationships, compete for food, and interact in ways that affect their health and well-being. Traditionally, monitoring these behaviors required humans observing cows directly, which could be stressful for the animals. Now, using artificial intelligence (AI), cameras, and sensors, researchers can precisely track and analyze cows' movements and social interactions without disturbing them. This review explores how advanced technologies, specifically deep learning and social network analysis, help us understand cow behavior in real-time barn conditions. We explain how tools like object detection, behavior tracking, and facial recognition identify individual cows and monitor interactions. However, current technology still struggles to clearly distinguish friendly from aggressive behaviors. We also discuss ethical considerations, including potential stress on cows from continuous monitoring and the need to help farmers understand these advanced systems. Overall, this work supports the dairy industry's move towards smarter, kinder, and less invasive farming practices.

Abstract: Integrating artificial intelligence (AI) with social network analysis (SNA) offers transformative opportunities for improving dairy cattle welfare, but current applications remain limited. This review critically analyzes recent advancements in dairy cow behavior recognition, highlighting novel methodological contributions through the integration of advanced AI techniques such as transformer models and multi-view tracking with SNA. We describe the transition from manual, observer-based assessments to automated, scalable methods using convolutional neural networks (CNNs), spatio-temporal models, and attention mechanisms. Although models like YOLO, EfficientDet, and BiLSTM have improved detection and classification, significant challenges remain, including occlusions, annotation bottlenecks, dataset diversity, and limited generalizability. Existing interaction inference methods rely heavily on proximity heuristics, lacking the semantic depth essential for comprehensive SNA. To address this, we propose innovative methodological intersections such as pose-aware SNA frameworks and multi-camera fusion techniques. Moreover, we explicitly discuss ethical considerations, emphasizing data transparency and animal welfare concerns within precision livestock contexts. We clarify how these methodological innovations directly impact practical farming by enhancing monitoring precision, herd management, and welfare outcomes. Ultimately, this synthesis advocates for strategic, empathetic, and ethically responsible precision dairy farming practices, significantly advancing both dairy cow welfare and operational effectiveness.

Keywords: dairy monitoring; social network analysis; precision livestock farming; computer vision; deep learning; cattle identification; behavior detection; animal welfare

1. Introduction

Dairy cows are inherently social animals. They actively participate in both affiliative and agonistic social interactions with other cows within the barn. Affiliative behaviors such as grooming and resting in proximity to others serve to reduce stress while strengthening social bonds. On the other hand, agonistic behaviors such as headbutting and displacement tend to result from competition and the establishment of dominance. Social interactions are important because they influence animal welfare, health, and productivity [1].

Social Network Analysis (SNA) provides a robust framework to explore and quantify these social behaviors. It offers a computational analysis of interaction maps to examine individual social preferences and group level cohesions. SNA is becoming more common in the study of dairy cows alongside other research elements, such as computer vision, deep learning, and artificial intelligence (AI). These advancements help with the unobtrusive monitoring of dairy cows, enabling the collection of interaction data and analysis of social behaviors to be conducted on a larger scale [2].

Therefore, understanding cow sociality through SNA methods can support the development of practical welfare-oriented strategies for barn management, in turn, improving cow health, longevity, and productivity. However, Hosseininoorbin et al. [3] points out that existing AI-driven SNA frameworks prioritize proximity metrics over behavioral intentionality (e.g., distinguishing grooming from forced displacement), limiting their capacity to model true social agency in dairy herds. Considering such constraints and the growing importance of AI in monitoring livestock, a comprehensive synthesis of current approaches is necessary. This review focuses on three main areas: (1) understanding affiliative, agonistic, and dominance behaviors of dairy cows and the application of SNA as a metric to capture these behaviors; (2) discussing the impact of computer vision, deep learning, and AI technology on monitoring behavior, predicting actions, and constructing social networks; and (3) outlining gaps in the research pertaining to precision livestock farming, specifically those relating to monitoring dairy cows.

1.1. Literature Search and Selection Methodology

A systematic literature review was conducted following PRISMA 2020 guidelines [4] to maintain methodological transparency, reproducibility, and academic rigor. The review sought to consolidate literature on SNA of dairy cows, paying particular attention to studies that integrate computer vision, deep learning, and AI technologies for monitoring and behavioral analysis of cattle.

The primary search was conducted using Web of Science, Scopus, Google Scholar and ScienceDirect. To enhance discovery and mitigate publication bias, Litmaps was used for citation chaining and cluster mapping to find papers that are semantically related but might not be accessible in traditional search listings. The keywords strategy consisted of combining and permuting terms including "Social Network Analysis", "Social Network Analysis of Dairy Cows", "Precision Livestock Farming", "Cow Identification", "Cow Tracking", "Cattle Keypoint Detection", "Cow 3D Tracking", "Proximity Interactions", "Herd Health Monitoring", and "Cow Pose Estimation".

The review focused on literature published from 2019 to 2025 to capture recent advancements like transformer-based architectures, multi-modal sensing technologies, and edge inference platforms with real-time capabilities. However, there was a selective add of literature outside the temporal scope prior to 2019 under well-defined inclusion criteria. Such older literature was retained only when it was shown to offer foundational concepts in machine learning applicable to current animal monitoring technologies or provided novel AI-based cattle monitoring technologies that have been substantively validated or extensively referenced in contemporary research. This mitigates temporal bias while still meeting the needs of the review's theoretical underpinnings.

Eligible studies were those conducted in English, published by a peer-reviewed journal, and focused on monitoring dairy cattle behavior, social interactions, or welfare. The selected studies employed SNA methods to quantify behaviors, including affiliative grooming, agonistic displacements, dominance, and proximity-based interactions and implementation of computer

vision or deep learning algorithms or sensor technologies such as Radio Frequency Identification (RFID), Ultra-wideband (UWB), and Automated Milking Systems (AMS) for monitoring cattle. This review analyzed empirical simulations as well as conceptual frameworks relevant to SNA and AI in livestock systems. The inclusion criteria extended to theoretical sources only when they served a foundational explanatory role like Alpaydin's Introduction to Machine Learning [5] that served to explain supervised learning paradigms that are essential for several of the models present in the included research.

Studies were excluded if they were not original research, had no relation to dairy cattle, lacked any application of AI or social network techniques, or failed to include animal science and animal-level behavioral modeling. Additional exclusions were made for papers that lacked methodological clarity or could not be reproduced due to insufficient reporting.

The screening followed a three stage process. First, a total of 311 records were obtained: 280 from databases, and 31 from manual and Litmaps-aided citation discovery. After removing three duplicates, 308 records remained for title and abstract screening. In this phase, 68 papers were excluded for lack of relevance, originality, or not being in English. Two hundred forty papers proceeded to full-text review. During this stage, 112 additional papers were exclusion candidates: 41 were on topics unrelated to AI and cattle monitoring, 38 did not focus on animal science; 33 were deemed methodologically unsound and unfit.

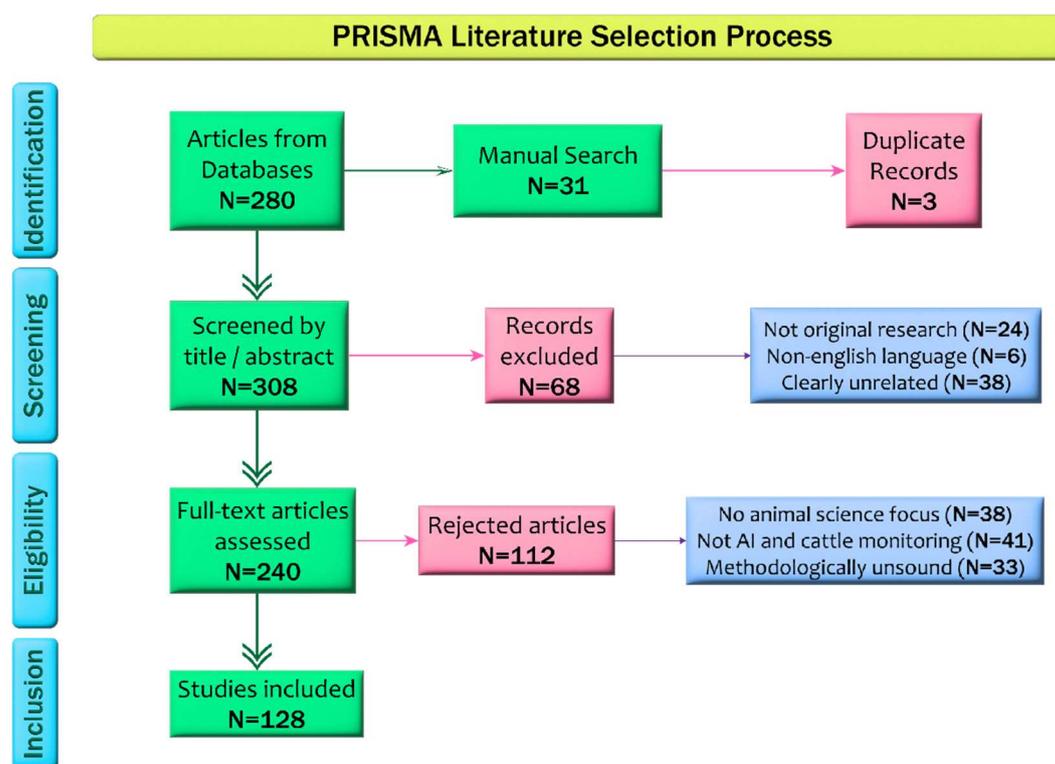


Figure 1. Flowchart illustrating the systematic literature search process following PRISMA guidelines. It outlines the methodical approach taken to select relevant studies on dairy cow behavior, social network analysis, and AI-based monitoring technologies.

In total, 128 papers were deemed eligible for inclusion. These papers cover a range of subtopics such as behavioral analysis with SNA, identity tracking through object detection models, proximity sensing, and inference of social structure through network metrics. A number of included studies utilized modern architectures such as YOLO (You Only Look Once), EfficientDet, CNN-LSTM (Convolutional Neural Networks – Long Short Term Memory) hybrids, and attention modules like CBAM (Convolutional Block Attention Module). The quality assessment was performed based on reproducibility and clarity benchmarks that involved data transparency, availability of source code

or algorithms, and strength of model validation methods (eg: cross-validation, multi-farm validation).

Several strategies were used to minimize bias. No studies were excluded based on result positivity or statistical outcomes, which along with multiple sensing modalities and geographic locations included, ensured the review was not narrowly tailored to a specific context. Also, citation tracing with Litmaps allowed inclusion of strong studies that are often ignored —to construct a more comprehensive corpus.

This approach provided a well-defined yet comprehensive collection of literature regarding the application of SNA in dairy cow monitoring systems in relation to AI and practical farming applications.

1.2. Review Scope and Structure

This review aims to summarize the recent findings and advancements that have been made in the field of SNA of Dairy cows with particular emphasis on Computer Vision, Deep Learning, Neural Networks and A. It starts by explaining social behaviors in dairy cows and the development of sociality. It then explains the application of SNA for quantification of social interactions, followed by detailing the technological improvements in dairy cattle monitoring systems. Further sections provide a detailed account of deep learning, object recognition, identity tracking, and interaction inference. Finally, the review examines the ongoing issues in classifying behaviors and synthesizes the identified research gaps alongside future scopes of work.

2. Social Network Analysis

Understanding social interactions in dairy cattle starts from looking at the basic social behaviors that define their social lives. This section discusses some SNA results retrieved from analyzing affiliative and agonistic behaviors of dairy cows, the development of social roles, as well as network-level patterns such as dominance structures, stability, and the impacts of regrouping. It also explains how A, and especially vision-based systems, improves the monitoring and modeling of these behaviors in real time and connects social behaviors to welfare and productivity outcomes.

2.1. Grooming Relations and Affiliative Behavior

Grooming among cows is a type of interaction which is systematic, structured and non-random. Foris et al. [6] reported that grooming behaviors are often asymmetrical. However, Freslon et al. [7] observed that reciprocal grooming was common—suggesting a tendency for cows to groom those who had previously groomed them. It is also interesting to note that cows who heavily invested in grooming others were less likely to be groomed themselves; suggesting high social spenders may bear some costs.

Rocha et al. [8] drew attention to the existence of stable preferential partnerships as a form of social organization at the herd level, where cows maintained contact with specific partners, sometimes referred to as 'affinity pairs' in the context of their social groups. In addition, Machado et al. [1] observed social licking behavior in 95% of the cows, noting that it reached its greatest intensity around 10:00 a.m and was often accompanied by feeding. In addition, proximity and licking events were positively correlated, albeit weakly, which indicates social proximity and contributes to the bonds supporting social cohesion among the group [9]. Familiarity is important since familiar cows had a higher likelihood of grooming each other than unfamiliar ones [10], and social grouping influenced their access to resources and general behavioral patterns [11].

2.2. Development of Sociability During Weaning

The foundation for the development of adult cattle social behavior is laid very early in their lives, during their pre-weaning and weaning phases as calves. Diosdado et al. [12] observed that calves formed stronger bonds with familiar peers, although these associations proved to be rather unstable

over time. Burke et al. [13] found that weaning increased social centralization, with some calves possessing relatively strong social roles that spanned several weeks.

Heifers raised socially exhibited more expansive and cohesive networks compared to those raised in isolation, as reported by Clein et al. [14]. This illustrates that early social exposure enables strong social integration post-weaning. Also, Marina et al. [15] showed that cows that were born close in time to each other, or were otherwise related, were more likely to form long lasting social ties.

2.3. Impact of Grooming and Affiliative Bonds

Affiliative interactions do not serve only for social comfort. They entail a number of more complex effects in cows' behavior. As an example, Gutmann et al. [16] remarked that cows placed into unfamiliar groups postpartum demonstrated lower lying times and dyadic synchrony—both of which are considered markers of social stress—which underlines the buffering impact of companionship. Likewise, de Sousa et al. [9] demonstrated that subordinate individuals were enabled to more efficiently access food due to Egest-Helper relationships with dominants.

2.4. Dominance and Hierarchy Structures

Most researchers agree that dairy cows form a dominance hierarchy which is not strictly linear. Krahn et al. [17] stated environmental conditions, hunger, reproductive status, and personality traits tend to influence resource access more than the rank of domination. Burke et al. [13] also showed that heavier and male calves were more central in the networks which indicates their greater social value. Some researchers employ various dominance scoring techniques, but they all seem to face difficulty in establishing consistent rankings due to clear differences in measurement methods and observational contexts [17]. Current SNA metrics, while descriptively rich, lack validated links to productivity biomarkers—a critical gap for translational precision livestock farming. This illustrates the need for future studies to combine structural network metrics such as centrality or dominance metrics with measurable indicators of welfare, including but not limited to, milk production, lameness, or lying time.

2.5. Influence of Parity, Age, and Health

Social behaviors are slightly influenced by age and reproductive condition. Older cows seem to participate in grooming more often, which may reflect their social rank [7]. Pregnant cows tended to receive more licking while older individuals both gave and received more licking as noted by Machado et al. [1], though these behaviors were not significantly associated with the hierarchy of dominance.

Social avoidance behavior due to lameness, parity and lactation stage, as reported by Chopra et al. [18], did not appear to be consistent, although there was some emerging preference on the socially active individual level. Additionally, Marumo et al. [19] reported that multiparous cows outperformed primiparous cows in relative social association strength and milk production, though their maximum association strength did not directly correlate with milk yield. Health as a factor impacting sociability was supported by Burke et al. [20] and Diosdado et al. [12], who independently reported that sick or socially challenged calves, despite being more socially active, exhibited lower centrality and association strength.

2.6. Individual and Spatial Sociability Patterns

Social behavior differs on an individual basis. Foris et al. [6] and Rocha et al. [8] documented individual differences in social behavior that were consistent over time across many individuals. Chopra et al. [18] also noted that some cows demonstrated stable behavioral patterns over time, indicating underlying sociability traits. Marina et al. [2] observed that individual positions in a social network, such as centrality, were more stable over time than group level dynamics. Spatial placement is essential as well. Rocha et al. [8] and Chopra et al. [18] showed that cows exhibit greater

individual differences in resting areas over feeding areas. Marina et al. [15] substantiated that there are location-dependent differences in the interaction patterns with other animals in the barn.

2.7. Network Stability

Social networks among cows are not random. Freslon et al. [7] noted that the social networks of cows are structured and orderly. Pacheco et al. [21] reported herds of cattle would form affinity relationships and maintain them over an extended period. Burke et al. [13], as well as Marina et al. [15], documented some repeatability in the aforementioned roles, together with degree and centrality, which suggests some level of stability within the network over shorter periods.

2.8. Consequences of Regrouping

Regrouping disrupts already existing social systems. Rocha et al. [8] noted that the addition of new cows to existing networks weakened them for at least two weeks. Even the resident-resident ties began to diminish weakening network strength. However, the underlying cause of this destabilization remains unclear. Longitudinal studies are needed to determine if post-regrouping network fragmentation reflects transient stress or permanent social memory impairment in dairy cows. Pacheco et al. [21] demonstrated that the separation of affinity pairs led to increased variability in milk yield by three-fold, underscoring the extent of impacts in productivity. Smith et al. [22] noted that unfamiliar cows possessed lower centrality, often remaining on the periphery of social structures even days after introduction, while familiar cows offered little interaction to newcomers, indicating passive rejection.

2.9. Agonistic vs. Affiliative Interactions

Agonistic and affiliative networks are distinct and as such uncorrelated. These relationships remain relatively stable over time, according to Foris et al. [6]. In contrast, the cows receiving the most preferred mates were shown to display significantly higher rates of both affiliate (3× more licking) and agonistic (1.3× more displacements) behaviors, suggesting emotionally charged bonds, as noted by Machado et al. [1]. Foris et al. [10] noted that while grooming networks were sparse and were stable compared to the more volatile displacement networks. Familiarity appeared to have an impact on affiliative behaviors, but agonistic actions were largely unaffected which suggests that competition for resources might be more uniform as opposed to preference for affinity pairs.

The prediction of cow social roles became possible with the rise of computational modeling. Marina et al. [2] showed STERGM's (Separable Temporal Exponential Random Graph Models) ability to estimate centrality using network features with moderate predictive power ($r = 0.22-0.49$) and improved accuracy with triangle-based features. This underscores the growing possibility of short-term behavioral forecasting using graph-based models. As discussed above, SNA illuminates rather sophisticated and subtle social interactions of dairy cows that can be studied and interpreted through quantification. From early-life bonding, affiliative grooming, and the disruptive influence of regrouping, spatial positioning and hierarchy, bovine behavior is intricate yet remarkably individualistic. Individual traits such as centrality, association strength, and closeness as social traits remain stable across varying times and contexts, providing reliable behavioral identifiers, or "fingerprints," for each animal [6,8].

2.10. Bridging AI and Animal Ethology

We have come a long way from when AI and cows were rarely spoken together in the same sentence, until technology brought them together in the dairy industry. Now, with the use of AI such as computer vision and deep learning models, monitoring dairy farms has become more efficient and accurate. Instead of needing field staff to constantly walk about the farm with clipboards collecting data, cows can now be monitored through unobtrusive cameras that have been placed around the

barn, feeding time, licking time and even idling can all be tracked giving remarkable insights into the health of the herd as a whole [23].

A contemporary computerized dashboard enables instant access to numerous performance indicators such as grazing time, feeding, and standing duration, which are very useful KPIs (Key Performance Indicators) to determine comfort and productivity. Withstanding time in alleys being unproductive, lying time directly represents the feed to milk conversion ratio which significantly impacts profitability as well as welfare [16]. With this technology, assessment of welfare has shifted from periodic manual checks, which were very labor intensive, to something done on a daily basis in real time, enabling automated evaluations that would have been impractical before [24].

In addition, the cow comfort index, which was only occasionally applied in academic work, has turned into a measurable standard on commercial farms, where it is tracked daily by AI sensors and cameras [25]. What the Automated milking systems and cameras provide for cattle is often misunderstood as an unwanted intrusion of privacy and the natural setting of the animals. Along with reducing the manual labor needed on the farm, these modern systems give supervisors and operators the ability to intervene less often, but more strategically. Also, modernized barns enable cows to behave freely and naturally, and they have the flexibility to decide when they want to rest, eat, or be milked which bolsters welfare and overall productivity. However, this technological shift also raises important ethical considerations. The pervasive deployment of vision systems in barns necessitates explicit discussion of farmer-cow data consent frameworks and algorithmic transparency to avoid “digital paternalism” in livestock management.

These visual information streams are mapped by SNA into interpretable social metrics. Indicators like degree centrality, betweenness, or association strength, as listed in the table, quantify the level of connectivity and interaction each cow has with the other members of the group, and how central it is to the cohesion of the herd. These are not ungrounded academic abstractions—they constitute practical measures of stress, social isolation, or declining health.

Table 1. Sociability Metrics for Evaluating Dairy Cow Social Interactions.

Metric	Definition	Behavioral Interpretation	Calculation Method	Data Requirements	Limitations	References
Degree Centrality	Number of direct connections	Measures social popularity; high = frequent interactions	\sum edges per node	Interaction logs	Ignores interaction quality	[2,20,26]
Betweenness	Role as a social bridge	Identifies gatekeepers controlling resource access	Paths passing through node	Network topology	Computationally intensive	[2]
Closeness	Average path length to others	Reflects social integration; low = isolated individuals	$1 / \sum$ shortest paths	Full network data	Sensitive to network size	[12]
Eigenvector Centrality	Influence within network	Highlights cows central to cohesive subgroups	Adjacency matrix eigenvectors	Weighted interactions	Favors high-degree nodes	[12,14,15]
Association Strength	Frequency of pairwise interactions	Indicates affinity bonds or avoidance	Interaction count / time	Continuous tracking	Context-dependent	[12]
Reciprocity	Mutual grooming /	Measures social balance; high =	Mutual interactions / total	Directed interactions	Fails in dominance hierarchies	[7,14]

	displacement	reciprocal relationships				
Network Density	Proportion of realized connections	Group cohesion; high = tightly knit social structure	Actual edges / possible edges	Complete interaction data	Biased by group size	[10]
Dominance Index	Asymmetry in agonistic interactions	Hierarchy stability; high = clear dominance order	Wins / (wins + losses)	Agonistic event logs	Misses subtle competition	[17]
Clustering Coefficient	Tendency to form triangles	Subgroup formation; high = cliquish behavior	Triples of connected nodes	Local network structure	Less meaningful in small networks	[8,18]
Reachability	Access to others via indirect paths	Social integration; low = marginalized cows	Binary reachability matrix	Full network	Binary simplification	[22]
Synchrony Index	Temporal alignment of behaviors	Social bonding; high = coordinated resting/feeding	Cross-correlation of timelines	High-resolution tracking	Requires timestamped data	[20]
Social Differentiation	Variation in interaction rates	Individual sociability traits; high = diverse social roles	Standard deviation of interactions	Longitudinal data	Sensitive to observation duration	[18]
Edge Persistence	Stability of pairwise ties	Long-term social preferences	Interactions over time windows	Multi-session tracking	Requires repeated measures	[15]
Affinity Pair Score	Strength of preferential partnerships	“Friendship” bonds; high = stable grooming/resting pairs	Dyadic interaction frequency	Individual-level tracking	Environment-dependent	[21]
Isolation Index	Proportion of time alone	Welfare risk; high = social withdrawal	Solo time / total time	Location + interaction data	Confounded by barn layout	[18]

As technology advances, so does intelligence in interpretation of camera footage. Modern computer vision systems can now recognize lying, feeding, and even affiliative behaviors such as grooming—all of which integrate seamlessly into SNA frameworks that aid in daily management. To be precise, SNA has transformed from merely a research tool into a real-time welfare system monitoring the health and well-being of animals. It bridges biology to technology to help answer the fundamental question in dairy science: What makes a cow happy? Because, as experts put it, a happier cow is a healthier cow.

To transform social interactions into actionable insights, one must first build a functional social network for the observed cattle group. Within this framework, social networks constitute graph-structured representations where each individual cow is a node, while social weight interactions between them are denoted as edges. Although directed edges theoretically offer richer relational interplay, most studies today employ undirected networks due to the practical difficulties of determining interaction direction in uncontrolled real-world dairy settings.

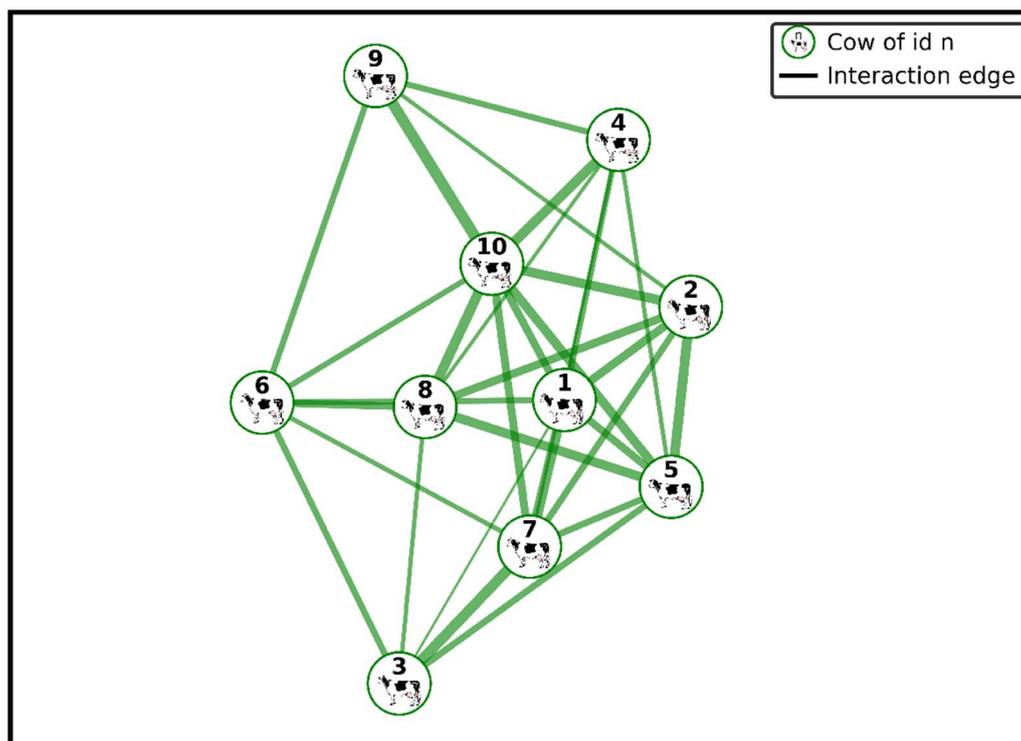


Figure 2. Example of an undirected social network graph depicting interactions among dairy cows. Each node represents an individual cow, and thicker lines indicate stronger or more frequent social interactions, providing insights into herd social dynamics.

3. Cattle Monitoring Systems: Enabling Precision Livestock Farming

The ability to evaluate livestock welfare, behavior, and productivity at scale and automatically is provided by automated Cattle monitoring systems which form a part of Precision Livestock Farming (PLF). Wearable and vision-based systems capable of tracking a multitude of behaviors including feeding, lying, locomotion, and social interactions have become more widespread because of technological advancements [27]. These systems mark a drastic movement from observation based on manual methods to analytics based on rich data collected in real-time.

3.1. From Manual Observation to Semi-Automation

In the past, behavioral data collection in cattle studies involved a lot of fieldwork, requiring lots of time and personnel trained to observe the subjects. In Machado et al. [1], a number of trained observers employed a scan sampling method to note licking and agonistic behaviors, checking on them every six minutes and noting their spatial proximity (within four meters). This approach was replicated by Freslon et al. [7], where human observers performed daily evaluations over a period of six weeks.

Nevertheless, such protocols that are intensive in nature posed issues with accuracy and scaling. This is the reason semi-automated systems came into existence. In Jo et al. [28], video analysis softwares allowed trained observers to annotate social interactions in terms of the instigator, receiver, and the location which improved data collection in terms of reliability and efficiency.

3.2. Rise of Smart Farms and Automated Data Acquisition

The cattle monitoring industry has changed very quickly because of the need for real-time, automatic, and situationally aware systems. Xu et al. [29] stressed the possibility of animal detection, identification, and behavior monitoring under ever-changing environments. In this regard, camera traps, drones, and RGB (Red Green Blue)/thermal imaging along with RFID and GPS (Global

Positioning System) have made it possible to acquire behavior-rich datasets from across farms and other settings.

Sort gates demonstrate the level of automation that has been achieved in barn logistics. These gates identified cows through RFID and assisted in their movement for milking, feeding, or health checks. In Pacheco et al. [21], sort gate passage logs were used for affinity pair analysis in order to estimate movement-based affinity pair analysis which revealed hidden social relationships captured through logistical data.

3.3. Sensor-Based Monitoring and Network Inference

Technologies that utilize sensors such as GPS, ultra-wideband RTLS (Real Time Locating System), accelerometers, and pedometers have become critical for continuous high resolution positional tracking. Using radio collar tags for triangulation enabled Rocha et al. [8] to perform SNA over a herd of one hundred fifty-eight individuals. Spatial coordinates were estimated by Chopra et al. [18] using weighted neck collars equipped with accelerometers. Marina et al. [2,15] also utilized ultra-wideband RTLS tags that can monitor cow positions at 1 Hz.

Proximity-based metrics are widely accepted as social interactions, especially within the realm of sensor-based SNA [21]. Although interaction types cannot be defined due to lack of proximity, cows being close to one another indicates that there is likely an affinity bond forming [1]. Also note that agonistic behaviors of affinity pairs are likely due to locational resource competition due to resource scarcity, which is why proximity is often regarded as an important social metric. Wearable sensors come with numerous advantages; however, data drift, calibration, external noise slow the sensor operations down, affecting its long-term reliability in outdoor conditions. Additionally, while RTLS tags excel in low-light or crowded barn environments, their inability to distinguish affiliative behaviors such as licking from agonistic ones like headbutting renders them inferior to pose-aware vision systems in behaviorally complex contexts. This behavioral ambiguity has driven the development of vision-based systems that offer richer contextual interpretation of social interactions.

3.4. Advancing Non-Contact and Vision-Based Monitoring

Recent innovations have led to the development of non-invasive methods camera-based methods for cattle monitoring. These systems allow for greater insight in behavioral studies without physical contact. Shorten et al. [30] developed a hybrid system integrating computer vision with a weigh scale to estimate milk production and assess udder attributes. Their model achieved 94% accuracy using 3D imaging for estimation of teat configuration and $R^2=0.92$ for yield estimation, showcasing efficiency and scalability in performance monitoring. Similarly, Fuentes et al. [31] developed the first contactless physiological system based on vital signs estimation using RGB and thermal imaging for heart rate and respiration counting. Their system correctly estimated milk yield and fat% and protein%, $R=0.96$, with ANNs (Artificial Neural Network) offering a persuasive option to invasive methods especially in uncontrolled real world farm settings.

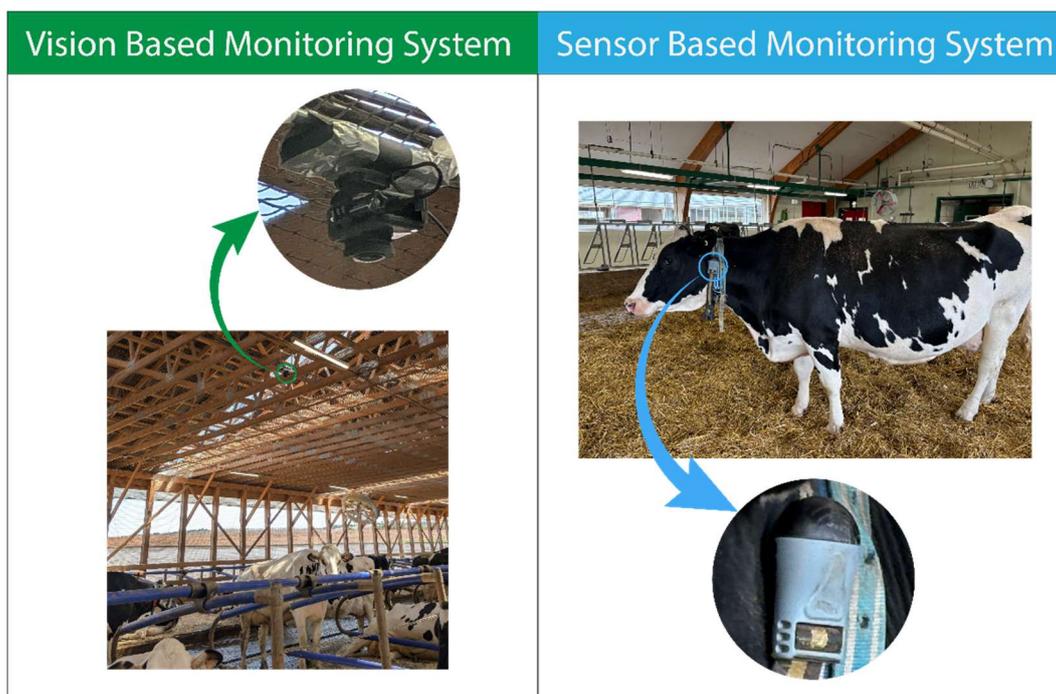


Figure 3. Visual comparison of typical equipment setups in dairy barns contrasting vision-based monitoring systems (using cameras) and sensor-based systems (using wearable sensors). Highlights the differences in complexity, invasiveness, and setup requirements.

3.5. Computer Vision and Deep Learning for Behavior Detection

With recent advances in deep learning, models like YOLOv5, YOLOv7 and Faster R-CNN (Region-based Convolutional Neural Network) are now available for real-time cattle detection and tracking [29]. These systems outperform traditional detection tools in identification of interaction behaviors, behavioral deviations, and spatial distribution. Real-world applications of vision systems are hindered by problems like occlusion, lighting conditions, and even the varying breeds of animals. To add onto this, the absence of labeled datasets is troublesome for supervised learning approaches [27]. For behavioral inference, pose estimation, mounting, and grooming detection have been implemented using 3D CNNs and LSTMs [32]. However, many of these systems are focused on precise data collection which makes them prone to errors in uncontrolled and unfamiliar barn conditions.

3.6. Sensor Fusion and Systemic Challenges

The integration of vision, audio, and even wearable sensors into a single framework enhances behavioral inference — these systems are known as multi-modality fusion systems [33]. More sensors which enable recording of cattle behavior must be integrated into wearable devices. Even though the employment of tracking algorithms such as DeepSORT in combination with Kalman filters has advanced tracking robustness in high-density cattle areas, merging multi-sensors is still in its early development stages, and field ready deployments are limited [34]. Additionally, the processing of some data from sensors (such as Interpolation, smoothing and filtering) is much easier than computer vision which requires a lot of processing to be done (like image preprocessing, object detection, identification and tracking). This sets the computational goal of having efficient and scalable pipelines capable of real-time operation within large farms on-site.

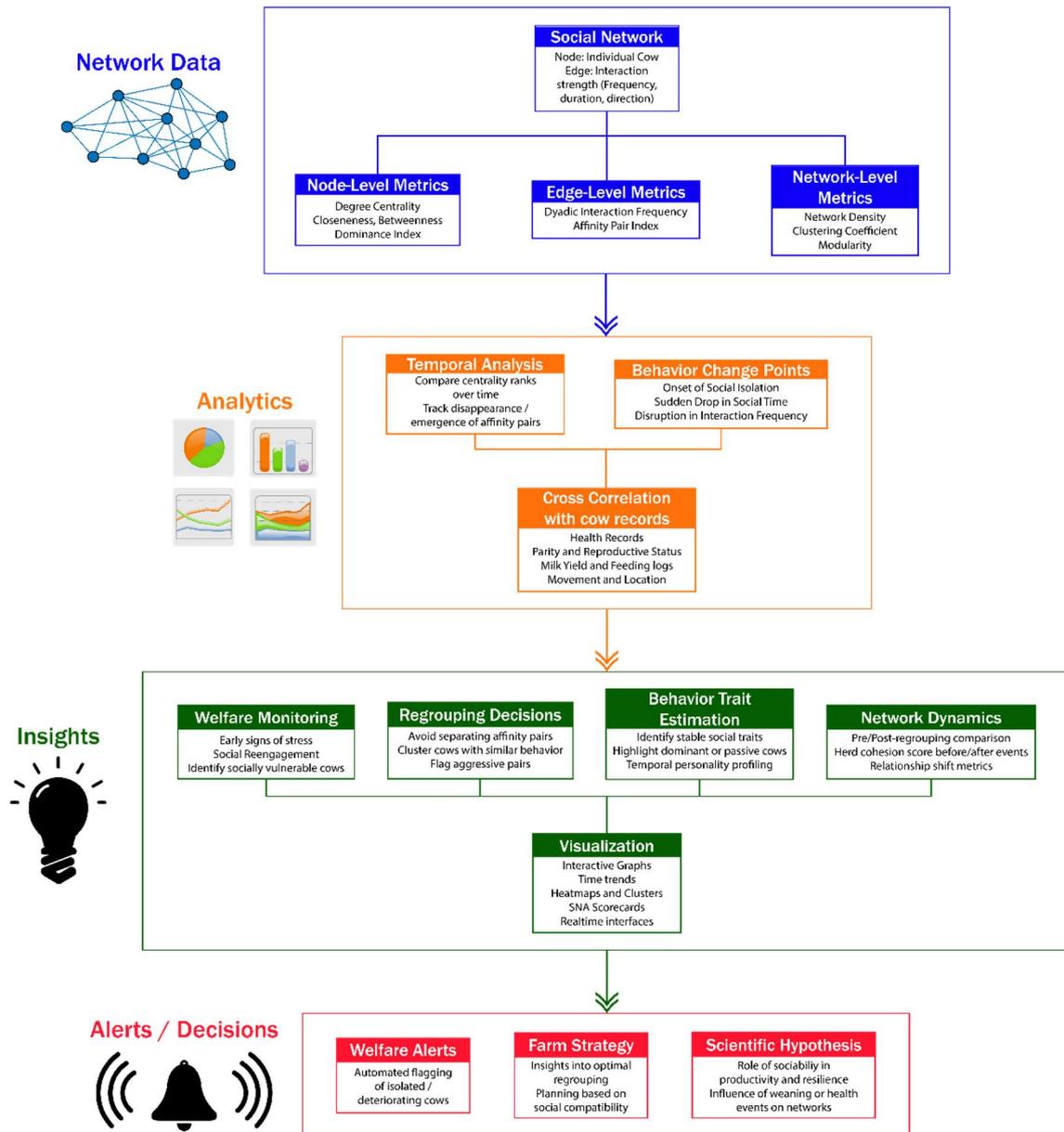


Figure 4. Illustration of the workflow used to construct and interpret dairy cow social networks from collected data. It shows the step-by-step transformation from raw sensor and visual data into actionable insights for herd management and animal welfare.

3.7. Data Annotation, Quality, and Reproducibility

The dataset's annotation's availability and quality pose a significant bottleneck in model development. Very little research data is available for benchmarking due to the possibility of human errors in manual annotation [29]. To address this annotation inconsistency, Ramesh et al. [35] designed and introduced the SURABHI (Self-Training Using Rectified Annotations-Based Hard Instances) framework, which uses self-training and label rectification to correct annotation inconsistencies using spatial logic combined with confidence thresholds. Their model demonstrated an 8.5% increase in keypoint detection accuracy which proves that temporal self-correction and attention-based filtering can enhance label robustness in complex frames. AI and sensors are now being used to fully automate monitoring of cattle, gradually shifting from outdated manual and semi-automated systems. Although vision-based systems and sensor-based systems each have their own unique advantages, their integration along with better data infrastructure and better annotation holds

significant promise for constructing sophisticated, intelligent, and welfare-oriented farm management systems.

Now that the foundational network behaviors and sociability patterns are mapped and established, it is necessary to explore the computational backbone driving these systems. The next section investigates self-organized neural networks which observe, interpret, predict, and quantify cow behaviors, positions, and interactions at a fundamental level vital for the formation of social networks.

4. Deep Learning Algorithms for Computer Vision Tasks

Transforming from manual observation of cattle behaviors to employing computer vision represents a remarkable advancement in monitoring animal welfare. Deep learning (DL) has shown significant promise in cattle detection, posture recognition, and social interaction analysis, particularly using convolutional and recurrent neural architectures. However, implementing deep learning poses significant challenges due to barn environments which are obstructed with complex lighting and unstructured movement combined with confined spaces, causing occlusions and limiting the scope of scalable solutions.

4.1. Convolutional Neural Networks (CNNs)

For image-based cattle data, CNNs are the most commonly utilized deep learning models. CNNs process data with grid-like structure, such as images. They use convolutional layers, which apply filters to input data to detect patterns like edges and textures. Tasks that have spatial detection patterns, such as detection, posture classification, and ID recognition, hinge on CNNs.

Qiao et al. [36] used an Inception-V3 CNN model on cows' rear-view video frames, classifying them on a per-cow basis. The backbone model was pretrained on ImageNet. The CNN-only model captured dynamic behavior features without temporal context, leading it to attain a low accuracy of 57%. Considering alongside the findings by Oliveira et al. [33], it's evident that most baseline CNN pipelines are unusable under barn conditions, where occlusion, dirt, and lighting variation interfere with visual input. Li et al. al. [32] noted the effectiveness of CNNs in detecting static postures (lying down or sitting, standing). However, they fall short in recognition of transitions or multiple overlapping behaviors. This shows that CNNs are effective in spatial computation, but are naive for temporal problems.

To address this issue, Qiao et al. [37,38] recommend augmenting CNN features with temporal models and alternative SVM (Support Vector Machine) based classifiers to improve performance. Notably, [37] utilized 2048-dimensional feature vectors extracted from the pooling layers of the CNN, as feature representation which consists of high-dimensional, information rich embeddings. While this improves feature representation, it increases the computational cost and may constrain real-time applications. Chen et al. [39] modified the Mask R-CNN to lower the computational overhead and achieve precise back segmentation even under occlusion. Mask R-CNN architectures have also been adopted for effective pose-variant segmentation. At the same time, [40,41] observed the higher detection accuracy achieved by the two-stage CNNs like Faster R-CNN, however they pose difficulties for use in real-time applications due to their trade-offs in speed and scalability.

The extensive use of CNNs in livestock applications is corroborated by [42,43], and [44], who reported high accuracies in animal detection and ID classification with CNN-based systems, utilizing YOLOv8 and VGG16 (Visual Geometry Group). Still, pure CNN systems tend to focus on visually distinct features and struggle with tracking social behaviors as they are poor in identifying subtle behavioral cues, and temporal continuity as mentioned earlier.

4.2. Spatio-Temporal Modeling and Attention Mechanisms

Perplexingly, CNNs processes each frame independently, neglecting the temporal behavior dynamics. This deficiency is addressed by LSTM and BiLSTM (Bi-directional Long Short Term

Memory) networks which establish dependencies from one frame to the next. They extend beyond the static spatial feature extraction by learning how spatial patterns evolve over time. As a great example, Qiao et al. [36] improved identification accuracy from 57% to 91% by using 20-frame sequences, attributed to incorporating LSTM layers to process a series of CNN features, strongly endorsing that temporal modeling is critical. Similarly, Gao et al. [45] applied a CNN-BiLSTM architecture in real-world barn settings, achieving greater than 93% accuracy despite heavy occlusions. Their high video frame rate processing led to enhanced motion detail but poses a worrying issue for constrained edge devices regarding bandwidth due to limited scalability.

BiLSTM networks, which analyze sequences in both temporal directions, were used by Qiao et al. [37], who attained 91% accuracy on 30-frame sequences. Their model outperformed both CNN-only and unidirectional LSTM baselines, suggesting motion-sensitive identity cues (e.g., gait, coat pattern flow) can only be captured via bidirectional sequence modeling. Similar approaches by [46,43], and [38] showed that BiLSTM improves behavior classification in cluttered environments, especially when fused with spatial CNN features. To address the issue of noisy frames and partial occlusion, attention mechanisms were introduced. Qiao et al. [46] embedded an attention layer after BiLSTM, allowing the model to focus on clear, identity-rich frames, achieving up to 96.67% accuracy. Notably, even short clips (1–2 seconds) were sufficient, underscoring attention's efficiency in low-data settings.

Fuentes et al. [47] extended the sequence learning paradigm by using ConvLSTM (Convolutional Long Short-Term Memory), which retains spatial structures during temporal modeling. Their system could detect 15 hierarchical behaviors across multiple farms, demonstrating high scalability and robustness in uncontrolled settings. Collectively, these studies suggest that spatio-temporal models with attention are state-of-the-art for livestock behavior detection—but at the cost of higher computational requirements, increased latency, and reduced suitability for real-time deployment unless optimized.

Qiao et al. [37] applied BiLSTM networks, which analyze sequences in both forward and backward directions, achieving 91% accuracy on 30-frame sequences. Their model surpassed both CNN-only and unidirectional LSTM baselines, indicating that identity motion cues (e.g., gait and coat pattern motion) are effectively captured through bidirectional sequence modeling. Approaches from [46,43], and [38] also reported similar findings where BiLSTM exhibited enhanced behavior classification in cluttered environments, especially in integration with spatial CNN features. To address the problems posed by noisy frames and partial occlusion, attention mechanisms were introduced. An attention mechanism embedding after BiLSTM allows the model to focus on clear, unambiguous identity frames, achieving 96.67% accuracy as stated by Qiao et al. [46]. Notably, short clips, even as brief as one to two seconds, were efficient, showcasing data efficiency that attention mechanisms provide. Fuentes et al. [47] expanded the sequence learning framework by utilising ConvLSTM, which preserves spatial structures during temporal modeling. This system was capable of detecting 15 hierarchical behaviors from multiple farms, illustrating strong scalability and robustness in uncontrolled environments.

On the whole, these investigations propose that spatio-temporal attentional models are the best for precision livestock farming behavioral analysis concerning livestock detection — though they make the system computationally more expensive, increase latency, and hence, are not well suited for real-time deployment unless optimized.

4.3. Transfer Learning and Pretraining

To enhance training efficiency and reduce data dependency, numerous researchers utilize transfer learning from models which are pretrained using large-scale datasets such as ImageNet. Qiao et al. [36] and Qiao et al. [37] reported significant improvement on the performance of Inception-V3 after fine-tuning it on rear-view cow images. Nevertheless, these pretrained models tend to underperform because of domain mismatch: urban-trained models cannot identify farm-specific features such as tail movement and muddy texture [38,33]. Xu et al. [29] demonstrated an increase in

YOLOv7 performance by augmenting with barn-specific images for retraining, showing how crucial domain adaptation is. However, pretraining also poses challenges—specifically the possibility of overfitting to unnatural augmentations that do not accurately represent real barn variability. As such, although transfer learning can be helpful in accelerating the development of a model, it cannot replace the thorough farm data collection and additional model training needed to fully optimize performance.

4.4. YOLO Frameworks for Livestock Applications

The YOLO-based architectures are often preferred due to their real-time detection capabilities enabling on-farm deployments. Their performance, however, commonly suffers under occlusion, lighting variation, and animal overlap unless specially tailored.

Xu et al. [29] modified YOLOv5/v7 with custom anchor calibration and augmentation strategies enhancing detection performance in cluttered contexts. Qiao et al. [46] observed how pretrained YOLO models from urban or clear farm scenes were unable to accurately localize cattle in occluded contexts. YOLOv8-CBAM, which Jo et al. [28] tested, added a Convolutional Block Attention Module which emphasizes important features and surpassed Mask R-CNN as well as YOLOv5 in performance. This architecture reached mAP@0.5 96.8% (Mean Average Precision) while precision stood at 95.2%, proving superior in heavily cluttered real world scenes as compared to Mask R-CNN and YOLOv5.

Other researches [48–50] focused on the new variants of Yolo, such as: YOLOv5x, YOLOv4, YOLOv8, all believing in the accuracy-speed ratio offered. Even YOLOv7 however, required attention-engagements (such as Coordinate Attention and ACmix (Attention Convolution Mix)) to process dense, object-rich datasets like VisDrone [51].

Although versatile, the YOLO models maintain a competitive edge only with an extensive amount of anchor and attention modular tuning. Moreover, unless combined with a robust tracking system, they heavily struggle with identity switch issue, which need to be addressed for social network tracking. The tracking system unless lightweight would hinder the real-time capability of YOLO. In practice, deployment feasibility also hinges on energy efficiency: YOLOv8-CBAM's 40W power draw per camera significantly limits its scalability in solar-powered or low-resource barns, especially when compared to EfficientDet's 12W baseline [52].

4.5. Capabilities, Challenges, and Future Directions

Neural network models have demonstrated strong effectiveness in the following areas:

- Detection of Static Postures with CNNs [32,33];
- Behavior Recognition with LSTM/BiLSTM and ConvLSTM [36,43,45];
- Real Time Detection with YOLO and EfficientDet [53,54];
- Multimodal fusion of RGB, thermal, and spatial data [28,31,55];
- Attention-based frame selection to reduce noise [46,48].

End-to-end models such as EfficientDet preprocess streams of video, and true to their name, achieving real-time inference with fewer FLOPs (Floating Point Operation)—paving the way for edge device and mobile GPU (Graphics Processing Unit) deployment [40,41,48]. The new self-training framework, SURABHI [35], strengthens pose estimation by improving annotations produced by machines, automating an important step in low-annotation data situations. Nonetheless, development of efficient neural networks and deploying them for real-time, real-world use has its own set of challenges.

4.5.1. Key Limitations

Some key bottleneck regarding neural networks include:

- Absence of public datasets greatly limit reproducibility [33,44,56];

- Farm-specific retraining is needed to improve breed, lighting, and occlusion generalization.[32,57];
- Real Time Detection with YOLO and EfficientDet [53,54];
- The cost of computation for attention-augmented or multi-camera systems still limits deployment in real-time [43,53,58];
- Applicability for complex spatial behaviors is limited as many models are tested using single-view data [36–38].

4.5.2. Recommendations

The immediate focus regarding the neural network's use in monitoring dairy cattle should emphasize the following,

- Standardization and sharing of the dataset for benchmarking [32,33,56];
- Semi-supervised learning and label revision (for example, SURABHI [35]) to minimize manual tagging;
- Increased[43,53,58] generalizability with Multimodal integration (thermal, RGB, depth) [28,46,55];
- Compound scaling and modular attention for edge-optimized architectures [48,51,53];
- Crossover in-breed and cross-layout validation to test and validate robustness [38,46].

Table 2. Comparison of Deep Learning Architectures Used in Dairy Cow Monitoring.

Model	Application	Performance Metrics	Computational Cost (TFLOPs)	Strength	Limitations	References
YOLOv8-CBAM	Detection	mAP@0.5: 96.8%, 95.2% P	40 W/camera	Occlusion robustness	High energy use	[28]
EfficientDet-D4	Detection	mAP@0.5: 94.1%, 12W	5.6	Edge-device optimized	Struggles with small objects	[53]
BiLSTM + Attention	Identification	96.67% accuracy	28	Temporal context modeling	Requires video sequence	[59]
Mask R-CNN	Segmentation/ID	94% IoU, 98.67% ID	22	Precise instance segmentation	Slow for real-time	[40,60]
Vision Transformer	Open-set ID	99.79% CMC@1	45	Scale-invariant features	Needs large datasets	[61]
DeepSORT + YOLOv5	Tracking	MOTA: 82.6%, IDF1: 89.4%	18	Occlusion handling	ID switches in dense groups	[42,57]
ResNet-50 + ArcFace	Facial ID	93.14% CMC@1	8.2	Lightweight embeddings	Frontal view required	[62]
ConvLSTM	Hierarchical behavior	84.4% F1-score	33	Spatio-temporal modeling	Computationally heavy	[63]
ByteTrackV2	Multi-object tracking	HOTA: 68.9%, IDF1: 76.2%	14	Balances speed/accuracy	Struggles with erratic motion	[34]
PointNet++	3D ID	99.36% accuracy	21	Depth-invariant features	Requires RGB-D sensors	[64]

DenseNet-121	Facial ID	97% accuracy	6.7	Feature reuse efficiency	Overfits small datasets	[44]
STERGM	Network prediction	$r = 0.49$ (centrality)	N/A	Dynamic network modeling	Requires historical data	[2,15]
SURABHI (Self-train)	Pose estimation	+8.5% keypoint accuracy	9.1	Reduces annotation effort	Initial manual labels needed	[65]
Graph Neural Network	Multi-object tracking	89% precision	19	Reduces computational cost	Detection, tracking tradeoff	[66]

4.5.3. Emerging Directions

The capabilities of precision livestock farming have been enhanced with new automated systems for behavioral understanding, automated tracking, and scalable health diagnostics benefitting from the innovations in deep learning. But as Arulprakash et al. [40] shows, no single architecture is proven yet to satisfy generalizability across multiple domains and problems, speed, robustness, and scalability simultaneously in the context of dairy cow monitoring. Advancements in Deep Learning models will stem from innovation in architecture combined with improved infrastructure, including better datasets, smart annotations, and multimodal sensing. Moreover, in practical scenarios, the system's backbone determines model performance, and dataset quality dramatically compounds the impact. Deep learning models used for dairy welfare and management will gradually shift from analytical tools to components of real-time autonomous AI decision-maker systems, enabling rapid response to monitoring and analytical challenges.

5. Object Detection in Cattle Monitoring

The analysis of cattle social behavior is sequential and hierarchical in nature. It starts with a video capture and ends with an interpretation of interactions among the cattle. The figure below (to be inserted) shows in broad terms the cattle monitoring steps within precision livestock farming.

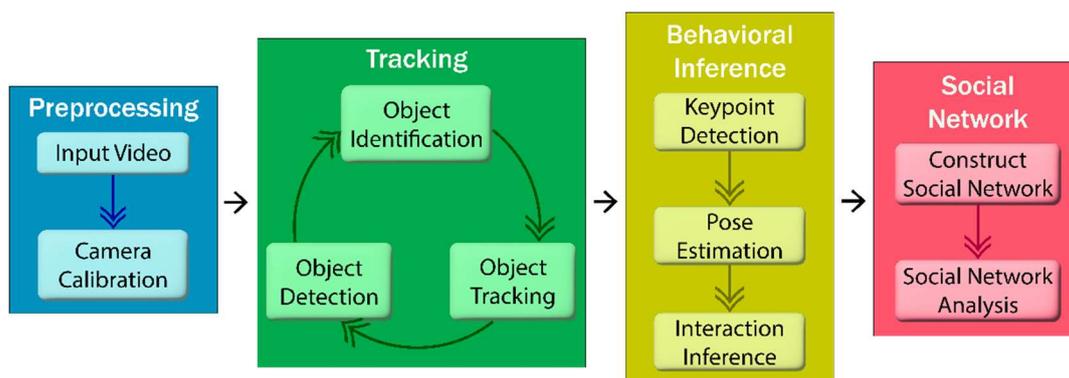


Figure 5. Detailed pipeline overview illustrating the end-to-end process of cattle monitoring using computer vision. This pipeline covers object detection, cow identification, tracking, pose estimation, behavior inference, and subsequent analysis for herd management.

Although the complete pipeline is defined as the best way to tackle the problem, some of the studies might choose to only a smaller subset, such as omitting explicit identification, keypoint detection, or pose estimation based on the data they have, or the goals defined for their problem. Object detection is the first and most foundational in this pipeline. In this regard, it is essential to explore applied computer vision through the lens of object detection for cattle in barn environments facing occlusion, lighting changes, and movement.

5.1. Object Detection: From Static Identification to Context-Aware Sensing

Accurate and robust object detection systems are not only crucial for identifying the presence of cattle but are also fundamental enablers of ID, tracking, pose estimation, and behavior inference. It is desirable for the cattle detection to perform well across a variety of visual distortions including occlusions, low contrast (e.g. black cattle against dark background), and irregular movements. The evolution of object detection architectures from early CNN-based models to more accurate, transformer-enhanced hybrid systems has also seen an increased complexity.

5.1.1. Early CNN-Based Detection and Two-Stage Architectures

Cattle monitoring initially relied on traditional object detection techniques using CNN-based models along with two-stage detectors such as Faster R-CNN and SSD (Single Shot Detector). These models claimed to achieve reasonable baseline accuracy with simple and controlled scenarios. Li et al. [32] offered a benchmark of YOLOv3, ResNet, VGG16, and R-CNN in barn settings, capturing strong accuracy but noted significant limitation regarding real-time inference, occlusion generalization, and farm-wide applicability and generalization. Huang et al. [55] applied an enhanced tail detection model incorporating Inception-v4 and DenseNet block SSDs, reporting outstanding detection metrics (Precision: 96.97%, Recall: 99.47%, IoU: 89.63%). This model bested YOLOv3 and R-CNN's accuracy and speed while harnessing over 8,000 annotated images. However, tail-based systems are occlusion-sensitive, severely limiting their application for behavior tracking when animals are densely packed or in motion. In the same way, Tassinari et al. [58] utilized YOLOv4 to detect general presence of cows and classify basic behavioral patterns such as lying, feeding, and walking, providing early indications that computer vision could replace sensor-based systems for continuous cattle monitoring.

Around the same time, Andrew et al. [67] tested RetinaNet, another two-stage model, for cow detection. It achieved a mAP of 97.7%, with ID accuracies reaching 94%. RetinaNet did outclass YOLOv3 in classification accuracy, but was slower—showing the speed versus accuracy compromise of two-stage detectors.

5.1.2. The Rise of Real-Time Detection: YOLO and Efficiency

Single-stage detectors, particularly variants of YOLO, were adopted to tackle practical considerations around real-time, low-latency inference. Noe et al. [57] showed that YOLOv5 models provided the best balance of accuracy and speed, with YOLOv7 surpassing detections even more but at a higher cost of computation. Wang et al. [51] incorporated attention mechanisms such as Coordinate Attention (CA), ACmix, and SPPCSPC (Spatial Pyramid Pooling—Cross Stage Partial Connections) into YOLOv7. These updates increased the model's mAP by 3–5% while maintaining a real-time frame rate of 30–40 FPS. However, this added computational burden creates difficulties for edge deployment in farms with limited resources or no dedicated GPUs. Dulal et al. [50] acknowledged that YOLOv5 outperformed Faster R-CNN and DETR (Detection Transformer) in terms of inference time and accuracy, thus further validating YOLO as a real-time cattle detector. This was advanced by Jo et al. [28] who incorporated CBAM into YOLOv8. This attention-based detector was able to achieve a mAP@0.5 of 96.8%, surpassing Mask R-CNN and YOLOv5 in highly cluttered scenes. However, the energy efficiency and hardware adaptations needed for optimized energy savings would have to be addressed before any practical use of the system.

Guzhva et al. [68] developed a method for the detection of cow heads and torsos using a rotated bounding box approach increasing their orientation and pose detection capabilities. Spatial clustering facilitated more accurate orientation identification. The use of watchdog mechanisms to prune irrelevant frames showcased a novel method for mitigating computation waste—an essential system design for battery-powered edge devices. Regardless, the influence of lighting and shadow artifacts significantly impacted performance, and thus, the generalizability is limited. Still, this method provides an important extension of moving from presence-based detection towards pose-informed

spatial modeling. EfficientDet serves as a scalable detector, leveraging compound scaling with BiFPN (Bidirectional Feature Pyramid Network) alongside others, and is therefore considered less heavy as compared with YOLO. Tan et al. [53] claims that EfficientDet surpasses YOLOv3 and RetinaNet in FLOPs Per accuracy making it exceedingly fit for mobile and embedded systems. Its effectiveness in human detection propelled its use beyond the scope of reliance, to non-human domains, farm animals included. Nonetheless, the accuracy it needs to perform under the specific visual constraints of a barn remains largely untested, hence, needing dedicated benchmarking against livestock datasets. Efforts toward 3D cattle detection remain in their infancy, while 2D object detection remains the dominant focus. Zhang et al. [34] compared PETRv2 (Position Embedding Transformation) and TransFusion-L, both 3D object detectors, with Faster R-CNN for cattle use cases. While these models are pose attentive and offer good depth perception, their exceedingly high resource demand and structured environment dependence renders them unsuitable for large scale deployment in dairy barns.

One limitation that many studies on object detection share is the use of homogeneous datasets, as they tend to overstate precision while underestimating the variability of real-world conditions. Arulprakash et al. [40] and Zaidi et al. [41] strongly advocate for cross-domain testing, especially the transition from pristine datasets to unstructured barn settings. They also strongly advocate for hybrid pipelines that combine different detection strategies (e.g., 2D YOLO + 3D PETR + attention modules) and completely cross-determine across breed and illumination and camera angles and views. Mon et al. [49] presented an elementary multi-stage pipeline consisting of YOLOv5x, VGG16x, and SVM/Random Forest classifiers. However, they also pointed out that their system was highly sensitive to unknown cattle and suffered from ID drift during occlusions, which is a fundamental problem for social network inference that requires reliable identity continuity.

5.2. Rise of Smart Farms and Automated Data Acquisition

Although object detection forms the computational core of cattle monitoring systems, it is clear from recent literature that detection alone is far from being sufficient for enabling intelligent behavioral or socially aware models. Object detectors declare presence of “who is here” but, in the absence of tracking, continuity, pose semantics, and social interactions remain highly opaque. Wang et al. [51] highlights this drawback by incorporating YOLOv7 detection modules with tracking and behavior recognition systems, thus deepening the understanding of cattle motion and interactions. Jo et al. [28] builds further on this by adding keypoint detection immediately after YOLOv8, which is followed by pose estimation and behavior analysis, marking an extension of object detection to behavior cognition.

Tracking systems should maintain the cattle’s identity through occlusion and reappearance scenarios across frames. Yi et al. [69] proposed a single object tracking framework that dynamically combines detection and tracking, where the detector is used to reset identity after a tracking failure. Lyu et al. [70] builds on this approach to emphasize instance-level re-ID (re-identification) for ensuring continuity in groups for SNA. Unsurprisingly, the focus on integrated and automated systems, is a common thread across different works and multiple. Dendorfer et al. [63] support the integration of tracking, detection, and pose estimation toward a single predictive engine, where modeling of appearance and trajectory works in parallel. Zaidi et al. [41] and Arulprakash et al. [40] make similar claims for hybrid object detection pipelines that combine object detection with scene understanding, pose interpretation and multi-object tracking, thereby enriching system-level intelligence by making it spatial and subject aware. From a behavioral perspective, Gupta et al. [71] and Wang et al. [72] advocate the integration of video-based identification systems with open-set re-ID frameworks and tracking, which allows continuous behavior monitoring of both familiar and unfamiliar subjects. Yu et al. [73] and Mon et al. [42] have explored identity tracking and extended it further into pose estimation and behavior classification, thereby creating a full pipeline from visual presence to social action inference.

As noted by Tan et al. [53], even efficiently scalable detectors such as EfficientDet can be extended to pose estimation and multi-object tracking which implies a modular framework for deployment in real world scenarios. Furthermore, Qiao et al. [36], Qiao et al. [37] propose the fusion of identification with activity recognition—linking who the cow is and what the cow is doing—which is fundamental for the automated construction of social networks. In sum, the literature strongly supports the view that object detection is not be treated as an isolated task, but rather a first layer and a starting point of a more complex system for automated reasoning, which encompasses the integration with identity persistence, motion continuity, keypoint extraction, and labeling behaviors.

6. Tracking and Identity Integration Based on Prediction

Cattle often disappear behind obstacles, move through crowded barns, and interact closely with one another, resulting in occlusions. A combination of tracking and detection is necessary in order to achieve continuity across space and time. Thus, detection must evolve into systems that can predict, interpolate, and re-identify animals in sophisticated scenarios. As such, the integration of models based on prediction tracking has surfaced as a focal solution for maintaining behavioral interpretation and long-term identity preservation. Notable contributions were made by Chen et al. [74] with their work on template-matching trackers which update object appearance models via weighted interpolation between two frames. By brute force computing the pairs of all nearby bounding boxes, they severely limited robust tracking under motion blur and partial occlusion. This approach illustrates how ineffective static frame-wise detection is for sustaining identity across varying scenes.

Building on this, Lyu et al. [70] incorporated pretrained detectors into regression-based trackers with correlation layers and achieved more stable box updates over time. Their model performance improved by ~5% mAP, strongly supporting the value of learned temporal features over simple sequential detection.

In a more structurally sophisticated framework, Wang et al. [66] incorporated a Graph Neural Network (GNN) to unify detection in multi-object tracking. This system, which was trained in an end-to-end fashion with both classification and contrastive loss, showed improvement in IDF1 and HOTA scores meaning identity association and consistency was better. Noteworthy, GNN's message-passing paradigm which captures object relationships across time robustly provides a key advantage for this model in interaction-rich barn settings. Notably, Tassinari et al. [58] came up with a YOLOv4-based displacement tracker. While it offers a functional lower bound, its lack of predictive mechanisms highlights the problem with naive frame-to-frame tracking, especially in action-dense settings.

To conclude, these models help illustrate that moving from detection to predicting tracking is not simply a matter of performance improvement, but rather a shift of paradigm — a fundamental approach and a conceptual necessity. A key shortcoming of systems relying solely on detection is the absence of behavioral modeling in longitudinal approaches, particularly in cluttered or occluded farm environments. However, after considering the latter, it is obvious the models of Wang et al. [66] and Lyu et al. [70] demonstrate the great need for and possibility of integrated detection-tracking pipelines.

7. Object Tracking in Cattle Monitoring

Tracking an object is particularly important in collecting data for SNA since it allows accurate identification of cow movements based on time stamps throughout the period of observation. Maintaining a single identity over the course of time is important on detecting proximities, affiliative bonds and even dominance structures. Advances in visual tracking recently experienced a shift from rule-based trackers and sensor fusion approaches to end-to-end trained architectures capable of continuous and real-time monitoring.

7.1. Vision-Based Tracking Systems for SNA

Most traditional approaches to tracking in cattle monitoring systems relied to focus on sensor-vision fusion. For example Ren et al. [75] used UWB tags integrated with computer vision for cow localization, while interaction detection at feeding points was done through a camera. This approach well served its purpose, but the infrastructure-based nature of its implementation restricts scalability.

On the other hand, Ozella et al. [76] removed the sensors: object detection was performed at top-down views with EfficientDet and cow identity was maintained through Euclidean tracking. Importantly, lost track re-identification was done through trajectory synchronization with milking parlor exits. This illustrates the potential of vision-only systems for automated long-term monitoring of large herds (240 cows in this case) in real time. Though the method relied on predefined infrastructure (milking parlor exit times) for track reidentification, this is something intrinsic with modernized farms and is thus not impractical.

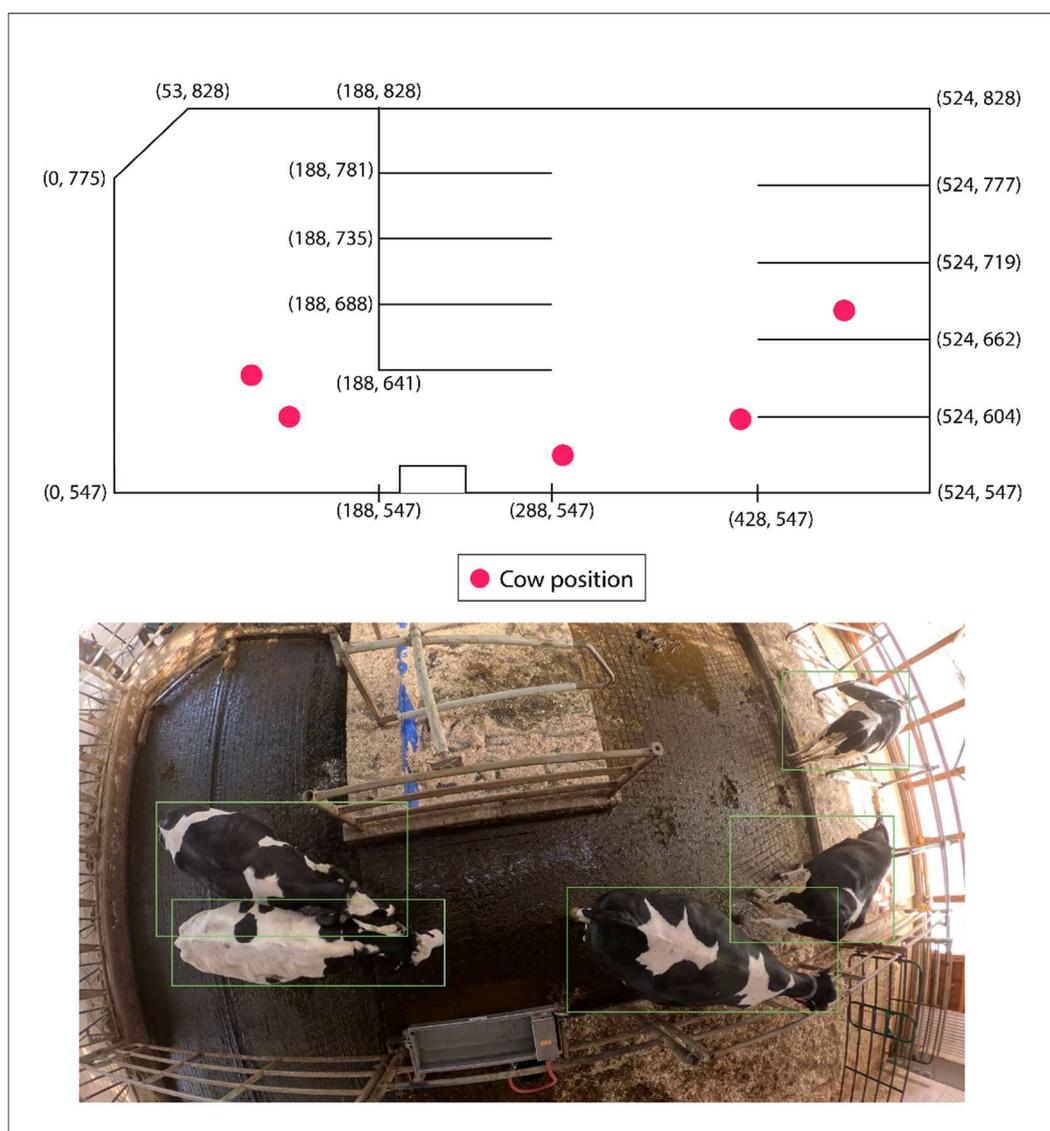


Figure 6. Example barn layout depicting real-time tracking of a small group of cows. The figure shows how cows' locations and movements within the barn are monitored continuously, aiding in assessing their social interactions and daily activities.

Mar et al. [77] enhanced vision-only systems using a multi-feature tracker that integrated spatial location, appearance features (color, texture), and CNN embeddings. In their pipeline, detection was achieved using YOLOv5 and ID tracking was performed with multi-feature association, gaining

95.6% detection accuracy alongside estimated tracking accuracy of 90%. Nonetheless, performance suffered greatly due to severe occlusion, underscoring a major challenge of MOT (Multi-Object Tracking) systems: identity fragmentation within cluttered scenes.

7.2. Orientation, Keypoints, and Interaction-Aware Tracking

Keypoint-guided tracking has been particularly useful for estimating cow posture, social interactions as well as direction of movement in relation to herd. Guzhva et al. [78] proposed a rotated bounding box detector based on head, tail, and torso localization. From the probabilistic model and orientation information derived from keypoints, next frame locations were predicted. Identity tracking was solved using a greedy NMS algorithm. Moreover, their watchdog filtering logic provided a means to for up to 50% irrelevant footage cut while losing only 4% of the meaningful interactions. This shows that intelligent pre-filtering significantly streamlines the entire annotation process without compromising behavioral data.

7.3. Deep Affinity Networks and Graph-Based Association

Track-by-detection models frequently suffer from fragmented identities caused by occlusions or missed detections. Liu et al.'s [79] work introduced a design of a Deep Affinity Network (DAN) which learns feature embeddings for detected objects and calculates pairwise affinities for object association across successive frames. The system managed entry, reentry and exit of objects robustly as well, which enabled its use in the crowded barns scenario with complicated trajectory movements.

Wang et al. [80] built upon this by introducing a graph-based formulation of tracking with min-cost flow optimization. Their innovation, muSSP (Minimum-Update Successive Shortest Path), applied a graph matching approach paired with a minimum-path graph finding algorithm for bounding box position alignment across two successive frames. By avoiding recalculations in parts of the tracking graph with stable associations, a lot of unnecessary computation reduced acting similar to a high-level graph-based optimization on DAN. It achieved between 5–337-fold acceleration over previous methods while still ensuring optimal association quality. This greatly enhances the real-time computational feasibility of large-scale herd monitoring.

7.4. Hybrid Tracking Models: Motion + Detection Fusion

Guzhva et al. [68] designed and used CNN-based tracking with visual markers in top-down views of barns as he managed to track 23 out of 26 cows for an average of 225 seconds per session, even in mildly crowded scenes. However, occlusions and visual ambiguity posed as major limiting factors particularly in dense scenarios.

Yi et al. [69] further validated the use of CNN-based tracking by creating a hybrid Single-Object Tracker (SOT) which combined CNN-based correlation filter and optical flow motion compensation. Regular motion was dealt with by the tracker, while a cascade classifier embedded detector dealt with more complex scenarios involving drifts or occlusion events. Their design enhances recovery from tracking failures quite well, and it was tested on standard benchmarks OTB-2013/2015 (Object Tracking Benchmark) and VOT2016 (Visual Object Tracking Challenge).

Complementary to this, Tan et al. [53] pointed out real time detectors like EfficientDet can be adapted for multi-object tracking and pose estimation, which shows how multi-purpose detector frameworks can be redefined to end-to-end tracking systems for track-and-vision integration.

7.5. End-to-End Architectures for Joint Detection and Tracking

Several modern tracking systems tend to prefer joint tackling of detection and association-based problems, also known as tracking-by-deep-learning. Wang et al. [48] proposed the Joint Detection and Association Network (JDAN) comprised of: An anchor-free detection head An association head with attention-based feature matching JDAN, having been trained on MOT16 and MOT17, outperformed dual-stage baselines on both MOTA and IDF1, offering greater consistency in identity

retention and fewer ID switches, even in occluded scenarios. Nonetheless, testing on actual footage from the barn wasn't done for this model, marking a gap in practical verification as a key limitation.

7.6. State-of-the-Art Models: ByteTrackV2 and Beyond

In scenes of high density which contributes to the failure case scenario of most tracking systems, ByteTrackV2 [34] has emerged as a benchmark. It retains low-confidence detections that are usually discarded by other pipelines, thus enabling continuity of trajectories under occlusion and motion blur. With both 2D and 3D tracking capabilities, it leads the nuScenes and HiEve benchmarks for performance, effectiveness and accuracy, and even stands as a candidate for realtime deployment in barns. ByteTrackV2's performance could be enhanced with spatio-temporal modeling by employing transformers.

Also, its active focus on edge devices with low power requirements fits the constraints of on-farm usage. Furthermore, its combination of motion and Kalman filtering provides better continuity for smoother long-term tracking. To resolve identity switches in occluded or crowded barn scenarios, the use of multi-camera fusion with epipolar geometry constraints offers a promising solution. Using the knowledge of barn layout and camera geometry, it is possible to perform preprocessing and consistently triangulate identities across views which helps eliminate fragmentation.

7.7. Tracking Benchmarks: A Caution on Generalizability

Some of the tracking systems tested under more academic conditions used datasets designed for humans, such as MOT17, MOT20, and OTB2015. As pointed out by Dendorfer et al. [63] and Wang et al. [48], these seemingly robust trackers, which perform well in their specific benchmarks, perform poorly in barn environments due to: Higher levels of occlusion, Irregular trajectories, Lower contrast (e.g., black cattle [57]), Non-rigid body deformation. As an example, trackers that relatively excelled on MOT17 suffered a 10-15% accuracy drop on the more dense MOT20 benchmark [81]. This difference is problematic because it suggests that models trained in a lab aren't necessarily robust enough to be freely adapted for use in agriculture.

To conclude, AI and sensors are now being used to fully automate monitoring of cattle, gradually shifting from outdated manual and semi-automated systems. Although vision-based systems and sensor-based systems each have their own unique advantages, their integration along with better data infrastructure and better annotation holds significant promise for constructing sophisticated, intelligent, and welfare-oriented farm management systems.

Now that the foundational network behaviors and sociability patterns are mapped and established, it is necessary to explore the computational backbone driving these systems. The next section investigates self-organized neural networks which observe, interpret, predict, and quantify cow behaviors, positions, and interactions at a fundamental level vital for the formation of social networks.

The development of cattle tracking systems indicates a clear shift towards multimodal, predictive, and identity-aware systems. Systems relying on traditional feature matching have been replaced by more advanced and robust architectures like DAN [79], JDAN [48], and ByteTrackV2 [34] that perform better under occlusion and crowding. Models still struggle with identity persistence over time, pose-informed tracking, and cross-farm generalization. Since social behaviour of cattle is highly context dependent, tracking systems need to further integrate: Keypoint and pose estimation [78], Behavioral prediction [43], Multi-camera fusion [46], And end device optimization for barn deployment in real-time.

8. Object Identification in Dairy Cows: Approaches, Architectures, and Advances

Seamless and reliable identification of individual cows is fundamental to intelligent monitoring systems for cattle. Identification is critical not only for tracking and behavior analysis, but also to link

interactions temporally, which is pivotal for Social Network Analysis (SNA). For tracking, verification is a fail-safe step, which ensures the absence of track and id switches. The progress of identification methods in cow monitoring shows an increasing tendency to multi-view. The advancement of identification methods pertaining to cow monitoring suggests an increasing tendency towards multi-view, multi-modal, real-time, and open-set capable architectures. This section reviews the most important changes in cow identification technologies and classify them according to their technological lineage and imaging techniques.

8.1. From AlexNet to Contemporary Pipelines: The CNN Foundation

Modern models for identification of cows are traced to Alexnet [82], which laid the foundation and revolutionized vision systems as it built a robust multi-layered convolution design and demonstrated a breakthrough performance in the 2012 ImageNet challenge. The architecture of Alexnet comprising five convolutional and three fully connected layers set the trend for a new era of deep learning systems in pattern recognition. Other designs which came later including ZFNet, OverFeat and VGG did further work on AlexNet by introducing smaller filters, multi-tasking, and sliding-window detection [83]. These backbone designs can be regarded as the base from which livestock models were developed.

8.2. Identification Pipelines: From Pattern-Based to Re-ID Systems

Pattern-based identification systems in cows depend on cow body coat patterns and require that the patterns are sufficiently unique and temporally invariant. Bello et al. [84] observed that a CNN trained on merely 1000 top-down images of 100 cows (10 different species) achieved 89.95 percent accuracy, thereby validating the claim that body patterns could serve as viable biometric markers. The precision of ResNet-based models can be further enhanced by integrating transformer components (CMT (Convolutional Neural Networks Meet Vision Transformers) modules, coordinate attention). Li et al. [71] achieved 96.84% mAP by multi-scale and semantic features integration, surpassing the performance of conventional CNN pipelines. Ramesh et al. [85] applied Keypoint R-CNN to identify key anatomical markers, transform body regions into bitmaps and matched using CNNs. This approach was effective and efficient for identification, but requires highly structured scenarios where geometric consistency across datasets was essential.

8.3. Rear View and Lateral Image Based Identification

Rear View Images Qiao et al. [36] implemented Inception-V3 for CNN feature extraction in video-based identification systems and employed LSTM, BiLSTM for spatio-temporal modeling. This yielded a 91% accuracy using 20-frame clips, outperforming standalone CNNs by a wide margin. Fu et al. [86] provided several architectural modifications to ResNet50 that increased the rear-view classification accuracy to 98.58% while decreasing model parameters by 24.85×. Various architectural choices, including dilated convolutions and Ghost Modules, as well as CBAMs, showcased the impact of design decisions on edge deployment feasibility.

8.4. Top-Down (Dorsal) Views and 3D Identification

Dorsal perspectives are preferable for consistency and reduced occlusion when the animal is walking through the barn passages. With augmentation, Jowett et al. [67] applied esNet-50 embeddings from dorsal images and generalization was made robust through data augmentation. Using SVMs trained on back images, Chen et al. [39] reported 98.67% accuracy across 48 cows. Xiao et al. [87] further improved dorsal detection by using a custom Mask R-CNN with Light_ResNet101 and SE (Squeeze-and-Excitation) blocks, along with features selected using Hu moments and Fisher Score. Their dual pipeline (segmentation + SVM classification) achieved over 98% precision. The efficiency claim was supported by real-time processing of 1.02s/frame.

Mon et al. [88] integrated YOLOv8 for detection and VGG for feature extraction, the output of which was fed into an SVM with majority voting across frames. The system exhibited robust scalability and ease-of-use, however it is to be noted that the tracking zones were narrow limiting the area covered for monitoring, and occlusions remained an issue. Ramesh et al. [64] went even further with CowDepth2023 incorporating RGB+depth and point cloud data. Their open-set deep metric learning achieved 99.97% using ResNet and 99.36% with PointNet. These are compelling results in support of depth-based identification, especially considering the lighting and occlusion challenges.

8.5. Multi-View and Free-View Identification

Multi-view systems assign a special focus on identification problems without a fixed camera perspective. Yu et al. [73] presented a multi-view re-ID system built with contrastive and supervised learning. Using 101,329 images of 90 cows, their pipeline attained single image accuracy of over 96%. Mon et al. [42] studied free-view recognition employing YOLOv3 on images taken from variable angles and reported 92.2% accuracy in ideal or clean conditions. However, accuracy suffered significantly with overlapping cows, highlighting the need for pose-robust embeddings and occlusion-aware architectures.

8.6. Facial Recognition, Keypoints and 3D Biometrics

The ease of mounting cameras and the uniqueness of faces as biometric attributes has fueled interest in facial identification. Dac et al. [62] applied a full facial ID pipeline deploying YOLOv5s6 for detection, ResNet101+ArcFace for embeddings, and performed cosine similarity search achieving 93.14% CMC@R1. Gunda et al. [89] used autoencoders and GANs (Generative Adversarial Network) to denoise and enhance facial images, with classification performed by Xception and 1D CNNs. Mahato et al. [90] pointed out the effectiveness of metric learning (e.g., Siamese Networks) and transfer learning from human face recognition for sparse datasets.

3D keypoints transform identity recognition to pose-aware recognition. Okura et al. [91] integrated gait and texture information using 3D joint data and grayscale images, achieving 96.5% accuracy. Menezes et al. [92] utilized 7 dorsal keypoints and the Euclidean distances in Random Forests, confirming that body geometry alone can enable accurate identification using infrared imaging, even under varying poses and BCS (Body Condition Scores).

8.7. Identification with Open-Set and No Supervision

Open-set and label-free identification models address real-world herd dynamics. Wang et al. [93] proposed an unsupervised model through CMCL (Cross-Model Contrastive Learning) and AGS (Adaptive Graph Sampling) with ResNet encoders. Despite the absence of labels, their model remained resilient to occlusion and variation in lighting. Wang et al. [94] apply ResSTN (ResNet + Spatial Transformer), reaching 94.58% open-set accuracy while incorporating four loss functions for embedding robustness. Burke et al. [72] applied MobileNetV2 with triplet loss and achieved 93.8% Top-1 and 98.3% Top-5 accuracy, demonstrating lightweight models can scale efficiently.

Table 3. AI-Based Methods for Individual Cow Identification.

Identification Feature	Model	Camera / View	Strengths	Limitations	References
Coat Pattern	CNN Identification	Top-down body photos	Coat patterns were shown to be viable biometric fingerprints	Sensitive to image quality, pose variation and lighting conditions	[38,84]
Coat Pattern	RetinaNet	Top-view	Efficient One-Stage Detection, robust to	Low tolerance to occlusion, bounding box threshold, training	[67]

			torso images	lighting, viewpoint, class imbalance	data quality, limited real-time scalability	
Coat pattern	FAST + SIFT + FLANN	Side view images	High accuracy, scalable and efficient for real time use	Vulnerable to visually similar cows, asymmetric coat patterns, lighting, environment variation	[94]	
Coat pattern	Resnet-18	Multi Top-down body photos	Good performance in confined environments without manual annotation	Not robust to occlusions, varied lighting conditions, texture invariant herd	[73]	
Coat pattern	YOLOv3	Non-fixed point of view images	Flexible over data sources, multiple angles and effective for real-time use	Poor performance with occlusions and group images	[95]	
Cow back pattern	Mask R-CNN + SVM	Top-view images	Accuracy > 80% for behaviors including licking, headbutt	Limited to the feed bunk area in AMS context	[40]	
Video-Based ID with temporal motion	Inception-V3 + LSTM / BiLSTM (+attention)	Rear view video	Temporal modeling greatly improved ID accuracy vs single-frame CNN	Accuracy decreases when the cows are static or have minimal movement which is common in dairy barns	[36,59,96]	
Facial ID	Resnet101 + ArcFace	Frontal Face Images + Thermal	Recognizes cattle facial biometrics accurately akin to human face-ID	Affected by lighting / angles	[62]	
Muzzle pattern	YOLOv5+Transformer	Muzzle images	Scalable, one stage, real time, robust to partial occlusion	Needs quality images, high training cost and complexity due to transformer, feature loss due to cropping	[50]	
Body anatomical keypoint geometry	Random Forest Classifier	Top-down view IR imaging	Robust to similar coat patterns, varying BCS, poses and lighting variation	Heavily reliant on manual annotations, limited real-time capability	[97]	
3D motion + Coat pattern	RGB depth maps + SIFT	Lateral RGB-D videos from both sides	Robust across viewpoints, lighting conditions, text invariant herds	Requires RGB-D infrastructure and sensitive to occlusion	[91]	

8.8. Challenges for Cattle Identification

Real-time performance of most models is less than their test accuracies when faced with occlusion, cow pose variations, and overlapping cows [98,42], despite their promising results. Most systems operate under controlled static top-down views and fixed lighting which seldom exists in

commercial environments. The combination of tracking with identification is an unexplored area. Tracking based ID correction is proposed by only a handful of studies including [36,42,88], most of whom seem to struggle with even moderately dynamic and dense crowds.

8.8.1. Data Scarcity and Dataset Demands

In addition, there are insufficient empirical datasets available for public access. Mahato et al. [90] and Mon et al. [42] talk about the need for larger, annotated, heterogeneous datasets that capture realistic conditions of barns. Li et al. [98] suggests creation of benchmark datasets across multiple modalities such as RGB, depth, and IR in order to enable cross-research compatibility and a fair basis for evaluation.

8.8.2. Future Directions: Multimodal, Re-ID, and Real-Time Systems

It is not hard to infer that there is a notable push for expansion of identification systems towards integrating multimodal biometric data including face, body, depth and IR. Integrating identification with behavioral analysis and health tracking is recommended by multiple studies [71,88,98].

Cross-view and multi-camera ID systems are potential solutions for occlusion issues [73,46]. Openset and real-time systems (e.g., [72,94]) need to be optimized and tested extensively for edge device deployment [90,71].

A clear pathway forward for identification could include these strategies: Self-supervised learning to reduce annotation costs [93]. Depth + IR imaging for robust ID under barn noise [64,92]. Joint ID-Tracking-Behavior architectures for full SNA integration [42,46,88].

From simple CNN classifiers on rear view images and coat patterns of cows, Cow Identification research has advanced to complex hybrid and open-set models trained on depth, keypoints, and top-down trajectories. ID systems today integrate YOLO, ResNet, BiLSTM, transformers and advanced loss functions for sophisticated fine-grained classification.

However, data and model generalization, occlusion robustness, and real-time deployment still remain as focal bottlenecks. Integrating robust identification models with behavior-aware tracking systems alongside multimodal sensory inputs could be the potential key to scalable PLF and SNA.

As Bello et al. [84] aptly illustrated, even the coat of a cow isn't just a coat anymore, when seen through modern AI lens, it's a biometric signature with alphanumeric characteristics that needs to be processed, decoded and identified.

9. Keypoint Detection and Pose Estimation Techniques in Dairy Cow Monitoring

Within cattle monitoring systems, keypoint identification resides as a vital intermediary step for Posture-Aware Monitoring. Pose estimation utilizes these keypoints to construct the body configuration of the subject of interest, cows in this case. Keypoint detection also provides standard methods and abstraction for identification, and behavior analysis. By identifying anatomical markers such as shoulders, hips, and spine across frames, these model enable robust biometric identification and temporal tracking under varying orientation and occlusion scenes.

However, uniform texture, deformable body structure, recurrent occlusions, and variation in lighting within barns poses significant challenges in keypoint detection.

9.1. 2D Keypoint Detection and Structural Geometry Modeling

Menezes et al. [92] identified dairy cows using infrared imaging and geometric keypoints. Seven dorsal landmarks like the hips and tailhead were manually annotated, and pairwise Euclidean distances were calculated which were then used to train a Random Forest classifier. This geometric representation resulted in high identification accuracy even with occlusion. Especially, the approach was robust to coat color variability and prevailing light conditions which showcased the importance

structural keypoints bring. Still, the reliance on manual annotation lacks scalability, particularly in large herds.

To reduce dependence on annotation, Lu et al. [99] developed a “Locating Key Area” (LKA) model that looks at the head, torso, and whole body. Though segmentation was performed manually during the training stage, the model demonstrated remarkable generalizability after deployment. Yet, this remains a semi-automated solution which emphasizes the need for a single-step, end-to-end keypoint localization system.

Ramesh et al. [85] proposed a zero-shot keypoint-based re-identification pipeline, where triangular dorsal patches formed by the joints keypoints were binarized and stored using a boolean vector. This method proved useful for identity database expansion without retraining because of its efficiency. Yet, it was too sensitive to slight keypoint inaccuracies which would cascade through the vector serialization system, greatly diminishing reliability for large, visually homogeneous cohorts.

9.2. 3D Pose Estimation and Point Cloud-Based Approaches

Realizing the constraints of 2D planar geometry, a number of works progressed toward 3D representations. Yang et al. [100] developed a Structure-from-Motion (SfM) approach that utilized cow RGB images to estimate cow body dimensions. It accurately captured withers height and chest girth and dwarfed depth-sensor accuracy. The method, however, suffered with direct lighting and reflective surfaces, greatly restricting its applicability in real-life barns.

Li et al. [101] used Kinect DK sensors with IR grating to carry out a point-cloud based anatomical slicing. The study applied DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering combined with logistic regression for anatomical region segmentation, precise identification of the legs and trunk without the use of deep networks. The system’s generalization to pigs and sheep indicates a robust multi-species applicability. However, the real-time scalability is impeded by the hardware complexity and computational expense.

9.3. Keypoint Detection for Behavior and Identity Fusion

Both Menezes et al. [92] and Ramesh et al. [85] demonstrated the dual functionality of keypoints in biometric identity and behavior inference. Particularly, the variability in postures was implicitly managed in the geometric relationships between landmarks. This is an emerging stream of research which has a strong focus on considering keypoints as more than spatial markers, but as behavioral expressions having meaning—this shift is essential for modeling social interactions.

Also, Qiao et al. [43] integrated the keypoint and the pose modules with the temporal Bi-LSTM networks to improve the granularity of behavior recognition. Such pose-temporal integration represents the next advancement towards holistic monitoring systems where the change in posture over time serves as an indicator for detection of grooming, lying, or agonistic interactions.

9.4. Limitations and Future Directions

Regardless of the considerable work that has been accomplished towards keypoint detection and pose estimation, several limitations persist: Real-time scalability has yet to be fully developed. The current 3D models, such as those used by Li et al. [101], lack optimization for barn deployment due to their high computational cost. Continued occlusion management remains an important limitation in high density herds, as body regions are frequently only partially visible. In some form, Manual annotation remains the backbone of numerous systems [92,99], which is a hindrance for extensive scaling up. Automated frameworks are of course a welcome addition, as evidenced by the work of Menezes et al. [92] who implemented CNN-based keypoint regression network. Further, as reported in [92], real-time implementation was more of a priority which is why behavior tracking and lightweight integration required more focus.

9.5. Necessity for Behavior Inference

In addition to serving as biometric references, identification keypoints and pose estimation are also vital to understanding posture, interactions and animal health. The landscape of keypoint-based cattle analysis is undergoing rapid transformation, ranging from basic 2D landmark matching to complex 3D reconstruction and zero-shot biometric encoding. However, for these models to be useful in commercial dairy operations, focus on achieving real-time responsiveness, unsupervised scalability, and integration with downstream behavior inference systems is vital.

In the next sections, the focus shifts to the modeling of interactions and behaviors, elevating keypoints and pose trajectories from anatomic markers to social movement metrics.

10. Interaction Inference: Evaluating Contactless Approaches for Social Behavior Mapping

Social networks in dairy cattle are built on social interactions, hence identifying and understanding these interactions are vital for mapping them. Though manual observation followed by video review (manual observation paired with video analysis, termed post hoc analysis) allows accurate measurement, it is very resource demanding and lack scalability. Because of this, more recent research adopt contactless, automatic systems. Most of these systems, however, depend entirely on proximity-based interactions logic, which, while efficient, but fails to recognize incidence of spatial presence as opposed to social interaction.

Ren et al. [75] provide a case of this shift with the hybrid system that integrates RTLS with computer vision for interaction detection in feeding areas. This method allows spatial proximity detection among cows and demonstrates viability for large-scale real-time monitoring; however, it lacks semantic richness. The system cannot assess the nature of interactions (affiliative versus agonistic) and who plays the initiating or receiving role, thus the system provides undirected social network graphs which may be considered as falsely directed under the right assumptions. Proximity methods such as those in [75] advance the social constructs of Networks but limit the network's integrity especially in cases when the behavioral quality of the interactions is very valuable.

To improve precision, Foris et al. [102] developed a rule-based algorithm for detecting displacement events at feeding bins, utilizing timestamps of entry and exit timestamps to infer agonistic behavior. This method was able to measure dominance and determine social hierarchy through asymmetry in interactions. Still, the method is spatially constrained to the external vicinity of feeding bins. Its extrapolative potential to other contexts within the barn or behavioral scope remains unexamined, indicating limitations in behavioral and environmental scope.

A significant improvement is noted with Ozella et al. [76] where a vision-based system with YOLOv5 and top-down video was implemented. Their system successfully classifies interactions as head-to-head (presumably affiliative) and head-to-body (presumably agonistic). Most importantly, it employs temporal smoothing and ID tracking, making it less sensitive to frame detection noise. On the other hand, the use of distance heuristics as rules for interaction classification lacks behavioral depth. The system works but not for the fine detail needed for interpretation of complex actions such as grooming, mounting, and subtle displacements. In addition, the robustness of the model across different barn layouts, lighting conditions, and herd composition variability has not been investigated. The requirement of pose-aware representations with anatomical keypoints and joints in motion is recognized but lacks practical implementation in the interaction classification stage. This is quite a critical gap because interaction type tends to highly depend on particular body positions and motion patterns. Bounding volumes and proximity measures on their own are insufficient for classifying interactions.

Table 4. Methods of Inferring Social Interactions Among Dairy Cows.

Method	Definition	Strengths	Limitations	References
Manual Observation	Scan sampling method used to observe and record interactions in the herd by human observers	Highly accurate affiliative, agonistic interaction data	Labor intensive, involves human errors, human limitation to observe the whole area leading to missed interaction observation	[1,7,9]
Continuous video-based observation	Affiliative, agonistic interactions identified with aid of analysis software	High-resolution behavior identification, with varied interaction types	Labor intensive, limited scope for automated system integration	[6,10]
UWB RLTS	Proximity based interaction inference	Suitable for PLF	Cannot identify interaction type	[2,8]
UWB RLTS + Accelerometer	Proximity based interaction inference	Automated and scalable, can detect spatiotemporal patterns	Cannot differentiate affiliative and agonistic interactions	[18]
Rule based using video at feed bunk	Detects Displacements via feed bunk entry and exit times	Infers dominance hierarchy	Limited to feeding context	[102]
AMS infrastructure record-based analysis	Affinity pairs identified by association in sort gate, milking parlor passing times	Simple, efficient, non-invasive, scalable way to identify affinity pairs	Cannot distinguish interactions, and may conflate dominance, friendship, avoidance	[19,21]
EdgeNeXt + IMU	Behavior classification using acceleration, angular velocity, and magnetometer data from IMU devices	High accuracy (95.85%), includes social licking detection	Licking, neck and leg rubbing were identified, but primarily for skin disease detection	[39]
LSTM + IMU	LSTM based RNN trained on time-series IMU data to predict behavior	Good accuracy and includes Social licking: 80.3%; Head butt: 81.9%	Frequent misclassification of Licking, headbutt due to short duration, similar movement, fluctuation IMU patterns	[103]
RLTS tags + LSTM (vision)	Detects interactions at feeders using RLTS proximity, and LSTM vision	Accuracy > 80% for behaviors including licking, headbutt	Limited to the feed bunk area in AMS context	[75]
CNN detection	Object Detection and proximity threshold-based interaction detection	Non-contact method of interaction inference	Doesn't differentiate interaction type	[68,104]
Vision-based (YOLO + tracking)	Classifies head-to-head vs head-to-body interactions	Adds semantic info and temporal smoothing	Uses distance heuristics and misses subtle behavior	[76]

In sum, while recent studies such as [76,75], and [102] attempt towards automating interaction inference demonstrating significant progress, they still lack deeply semantic or spatial reasoning or are computationally fragile. Systems purely relying on spatial proximity or pre-defined threshold

heuristics risk oversimplifying social complexity, particularly in crowded barns. Integrating keypoint detection, pose estimation, and temporal context represents the next step for directionally sensitive behavior-specific interaction mapping and constructing socially valid rich networks for complex analytical social systems. To truly distinguish active social engagement from passive spatial co-location, future systems must integrate gaze estimation as well as ear and head pose tracking to identify intentional interactions instead of relying solely on proximity thresholds. Such refinements would enable systems to interpret not merely whether animals are located in each other's vicinity, but how they are attending to each other, enabling a better classification of behaviors like affiliative, agonistic or avoidant interactions.

11. Behavior Analysis

The analysis of dairy cattle behavior is critical in enabling automated systems for welfare evaluation, productivity optimization, and early-stage disease detection. While traditional methods used wearable sensors like accelerometers and IMUs, more recent studies are leaning toward vision-based deep learning frameworks because of their high scalability, non-intrusive nature, and ability to capture behavioral subtleties. Unfortunately, there is still some work to be done especially regarding occlusion, pose variability, and generalization across farms.

One of the pioneering works toward integrated spatio-temporal modeling is Wang et al.'s [105] introduction of Temporal Segment Network (TSN). Though designed initially for generic video-level activity recognition, TSN's ability to incorporate long-term temporal context was crucial for modeling cattle behaviors, and highlighted the limitations of CNN-only models, which are heavily reliant on spatial appearance over temporal dynamics, which is most commonly the changes over multiple time frames.

On the other hand, the sensor-based system outlined by Peng et al. [103] utilized LSTM units trained on IMU data to classify several behaviors in cattle. The model achieved better recognition accuracy than older CNN-based approaches, particularly for lying and walking, but had challenges with physically similar actions such as social licking and headbutting. Most notably, the use of wearable sensors poses a limitation for its implementation in large scale or low-resource farm environments, where maintenance and cost become an issue.

Tassinari et al. [58] tried to bridge this gap by incorporating stance, spatiotemporal context, and other time-based features into behavior recognition systems. In their study, they tracked cattle and contextually auto-labeled activities such as feeding, walking, or lying only with video data, proving that behavior recognition does not necessitate the use of wearable sensors. However, extending this system for more behavior classes and intra-animal variability continues to be an unsolved problem.

Similarly, Fuentes et al. [47] proposed a fully vision-based approach with feature extraction based on CNNs, spatio-temporal modeling, and hierarchical behavior classification. Their system was able to recognize 15 distinct behaviors in real-time which was an impressive milestone. Still, lack of validation data and highly controlled testing conditions are major limitations for its validity under ambient and unpredictable farm conditions.

Achour et al. [38] designed a similar system focusing on feeder behaviors employing posture-classification CNNs and feeder-state detectors. This research, however, very optimistically, only focused on the feeder regions and didn't consider more holistic behavioral contexts such as lying or aggressive interactions, limiting its range and generalizability. This lack of operational scope is a common problem in many behavior detection pipelines, particularly those constrained single-zone camera setups or sparse class labels and annotations.

Work towards classification based on posture has shown promise. For instance, Avanzato et al. [106] worked on implementing a YOLOv5-based CNN for the differentiation of standing and lying postures in multi-camera barn footage. Although posture detection was indeed proved to be effective using CNNs, the absence of multi-class behavioral differentiation (e.g., feeding, walking, or mounting), and lack posture-aware modeling greatly diminishes its utility in rich behavior settings.

Expanding into physiological behavioral monitoring, Wu et al. [107] used LK optical flow and Phase-Based Video Magnification (PBVM) for respiration monitoring in cows. This method achieved 93.56% accuracy and has the potential for early disease detection. However, the high-fidelity video requirements for this approach precludes real-time use in environments with variable illumination and resolution, making widespread use difficult.

To conclude, behavior analysis systems for cattle have advanced from classification models that relied on wearables to multifunctional, contactless, vision-based systems that cover multiple behaviors. However, the majority of today's models have restrictive boundaries in real-world generalization, system scalability, and behavioral granularity. For complex farming settings, pose-informed modeling, context-aware classification, and multi-modal fusion (like vision with sound or thermal) need to be implemented to achieve robust and actionable behavior inference.

12. Limitations and Future Directions in Keypoint Detection and Pose Estimation

While substantial advancements have been made in vision-based cattle monitoring and social behavior analysis, several core limitations continue to hinder methodological consistency, scope of implementation, and feasibility of real-world deployment. The following gaps outline persisting issues which were substantiated by the literature analyzed in this study.

12.1. Incomplete Pipelines Without Comparative Evaluation

While several studies feature partial monitoring pipelines, such as detection and identification ([88,42]) or detection and tracking ([57,48]), very few have shown complete integration from object detection to behavioral categorization to social network graph construction. Not many systematic comparisons have been done where partially implemented pipelines are set against complete ones by filling the missing modules with standardized models and measuring the SNA output dependencies.

Future studies should prioritize developing methodological benchmarks. Systems with partial architectures in past studies should be extended using standardised, modular and reproducible techniques (e.g., ResNet for re-ID, BiLSTM for behavior) and benchmarked as end-to-end systems. This sandbox approach, however, requires publicly available open code and datasets for the established studies which are both lacking. This brings us to the next limitation towards the development of a cattle monitoring pipeline.

12.2. Dataset Bottlenecks and the Need for Open, Multi-Modal Benchmarks

A number of high-performing models from the literature are built from proprietary datasets ([39,64,84]) under controlled settings. Moreover, the behaviors and interactions also lacks standardization for labeling, rendering them impossible to cross compare for different studies. Very few datasets contain multi-modality (e.g., RGB + depth + thermal) or necessitate long duration tracking for temporal stability analysis within a network. Among the publicly available datasets referenced by Bhujel et al. [108], only a few support the implementation of a robust continuous monitoring pipeline, and none provide sufficient data for a complete end-to-end pipeline.

The absence of open multi-modal datasets containing various types of barns, their cameras, and labeled social behaviors poses a problem. Without these, claimed model performance is context bound, and studies in the dairy field risk stagnation due to lack of reproducibility.

Future efforts should aim at comprehensive creation of open, standardized, multimodal datasets annotated with rich and relevant social behaviors. Such initiatives can adopt the approach taken by the autonomous driving domain where progress has been rapid because researchers and developers use shared publicly benchmarked datasets like KITTI and WaymoOpen which are openly available, well-annotated, and used as standard benchmarks for evaluating and comparing models.

12.3. Interaction Inference Still Relies on Heuristics

Despite some vision-based systems like Ozella et al. [76] extend beyond raw proximity, most techniques for interaction inference still operate using static spatial thresholds and are bound by proximity. There is little to no motion or posture cue disambiguation for instigator-receiver role determination. Key behavioral actions such as grooming, butting, or displacement are often oversimplified, and the undirected edges are very commonly considered.

Implementing an interaction classification model based on pose with keypoint trajectory and temporal calibration would fill this gap. Labeling interactions at the frame level with real validation will serve to achieve the demands of a holistic interaction detection system which are currently lacking across most datasets.

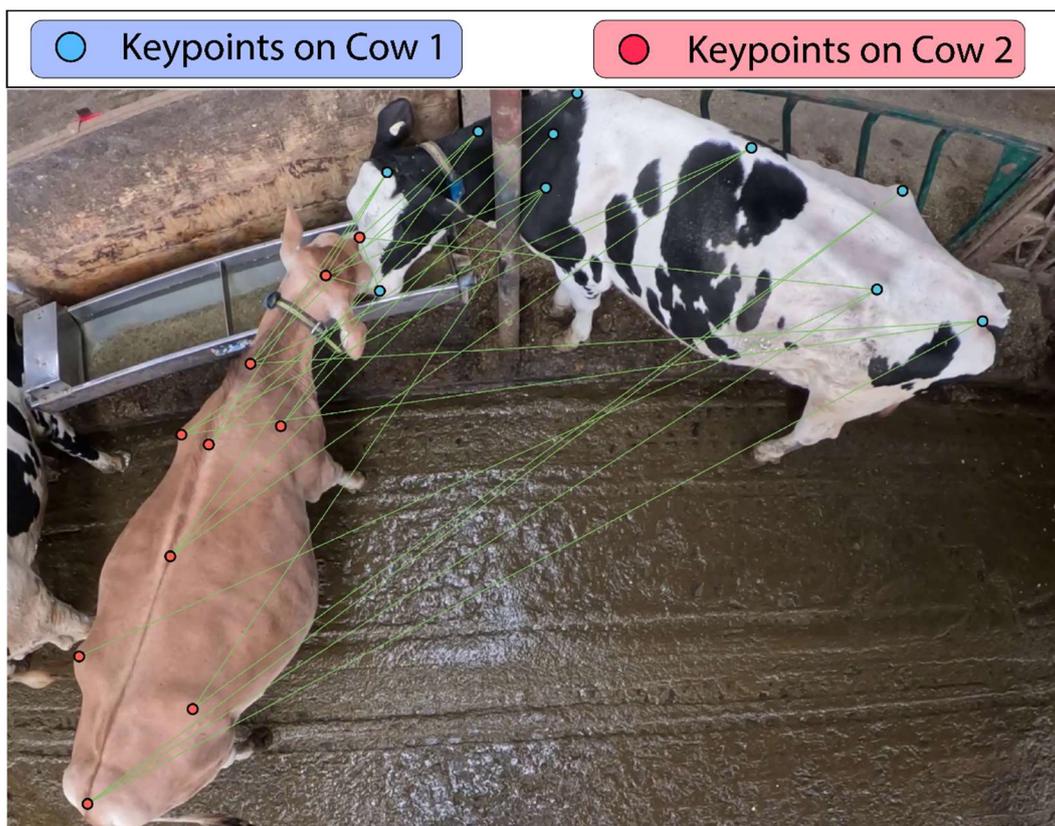


Figure 7. Graphical depiction of the distance between key body points (keypoints) of two cows during a headbutting event. This illustrates how pose-aware tracking methods can distinguish between aggressive interactions and general proximity.

Future research should prioritize developing standard baseline benchmarked pose-based models trained on annotated temporal sequences of interactions, leveraging spatio-temporal graph networks or 3D CNNs. Designing a standardized annotation protocol could guide development of datasets containing affiliative and agonistic interactions and cues related to them.

12.4. Inability to Generalize in Complex Barn Environments

While many models report high performance under controlled test setups, generalization to free-moving, multi-breed, and occlusion-heavy barns are seldom demonstrated. Studies like [42,67] observe severely falling accuracy over uncontrolled viewpoints or overlapping animals. There is a significant gap in cross-barn, cross-breed, and cross-season generalization of algorithms, models and datasets.

This requires systematic, multiple and varied benchmarks compounded with external evaluation—these have been remarkably overlooked in existing literature. Datasets benchmarked

under such varied conditions must be established. Models developed for interaction detection and cattle monitoring in future studies should be pretrained on these datasets to establish a baseline upon which continuous progress can be observed in this field.

To further fortify the solutions, cross-barn and cross-breed benchmark challenges should be complimented by collecting and organizing datasets that are multi-seasonal and multi-location as well. Researchers should focus on designing domain adaptation and transfer learning strategies tailored to livestock monitoring contexts.

12.5. Towards Comprehensive Behavioral Taxonomies for Welfare Assessment

Most behavior recognition systems concentrate on 2–3 high-level categories such as feeding, lying, or walking ([38,58,106]), and very little work has been done to expand taxonomies, for example, on social behaviors, stress indicators, or reproductive cues. Studies exist that were able to recognize as many as 15 behaviors [47], but there is no established benchmark for behavioral taxonomy hierarchies that are deemed important for health and welfare analytics.

Future work should focus on designing a developed behavior taxonomy that is both layered, structured, and encompasses social and individual actions. The taxonomy can be complemented by using interdisciplinary consensus involving ethologists, veterinarians, and AI researchers. These should be incorporated into open and standard datasets for training and testing universal behavior recognition models.

13. Ethical Considerations in Precision Livestock Farming

The increasing integration of AI and automated surveillance systems in dairy barns has created new ethical predicaments regarding data ownership, algorithmic transparency, and the digital rights of both farmers and animals.

As data is one of the most valuable resources in PLF, question arises regarding who gets to own the data that captures the behavioral and biometric features of dairy cows—the farmers or the third-party monitoring system service providers? Farmers spend capital on sensor systems and data infrastructure, but the processing and insights are often handled by proprietary algorithms from third-party vendors and developers. In the absence of clear frameworks, this can result in unfettered and asymmetric control over access to livestock data by the service providers. The models interpreting the livestock aren't usually simple and understandable for the farmers, and in addition they might not have the proper knowledge and training to access the dairy cow monitoring data on their farm.

Just as troubling are the black box decision making algorithms of cameras and tracking systems. Modern AI-driven analytics increasingly influence management actions ranging from selective milking to automated health alerts. The absence of explainable AI and transparent model documentation leaves farmers dependent on trusting the opaque systems, risking overdependence on tools they don't have the ability to audit and have control over. This situation raises the threat of digital paternalism where algorithmic control on the barn decision supersedes human oversight without recourse.

In addition, although the concept of animal privacy might seem like an overly humanized perspective, continuous video monitoring and biometric surveillance of individual cows indeed require ethical framing. Although cows cannot consent for being monitored, it is the responsibility of researchers, engineers, and the dairy monitoring system designers to implement systems that would avoid causing stress, behavior distortion, unreasonable intrusion, and respect animal welfare.

For these reasons, the pervasive deployment of vision systems in barns necessitates proactive attention to the farmer-cow data consent policies, and algorithmic transparency so as not to incur 'digital paternism' in livestock management. On a broader scale, AI in agriculture requires an ethics-first approach prioritizing participatory design, equitable data governance, and regular ethical auditing to ensure that PLF technologies serve as a tool of empowerment in the field, rather than instruments to control, invade and subjugate.

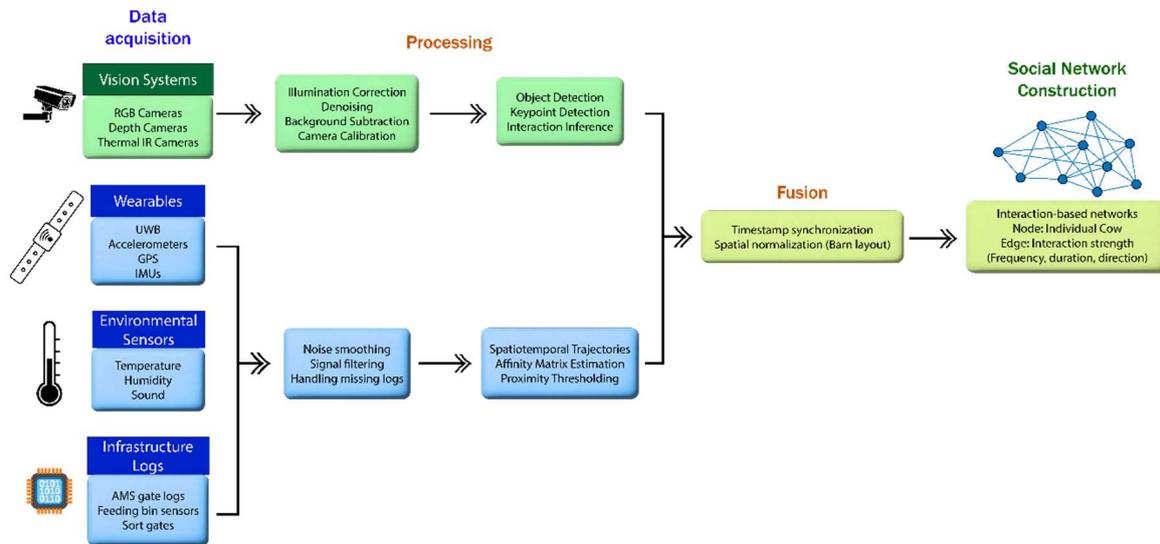


Figure 8. Flowchart representing a multimodal, integrated cattle monitoring system combining visual data, sensor data, and advanced analytics. Highlights how different data streams merge into a unified decision-making tool for enhancing dairy cow welfare and productivity.

14. Conclusions

The integration of advanced artificial intelligence into the dairy barn has lifted the curtain on the previously unseen social complexities of cows, challenging conventional views and opening new avenues for compassionate herd management. Traditionally, the social lives of dairy cows were simplified and observed through basic human observation, leaving critical details overlooked. However, the emerging combination of social network analysis (SNA) with cutting-edge AI technologies provides a powerful lens to decode the subtle social cues, affiliations, and hierarchies within dairy herds. The transformative potential of this approach reaches far beyond mere academic curiosity; it stands to significantly enhance animal welfare, farm productivity, and ethical farming practices.

At the heart of this review lies a profound novelty: the explicit integration of sophisticated AI methodologies such as transformer architectures and multi-view tracking into traditional SNA frameworks. These advanced AI techniques offer a significant methodological leap forward, enabling precise tracking and interpretation of individual and group interactions with a depth previously unattainable. Although systems like convolutional neural networks (CNNs), recurrent architectures such as BiLSTM, and detection models including YOLO and EfficientDet have achieved notable progress in behavior classification, detection, and identification, they still fall short of fully capturing the intricacies of cow interactions. Current methods rely heavily on simple proximity metrics, inadequate for accurately differentiating intentional social interactions like grooming, aggression, or social affiliation from mere passive proximity. Such limitations hinder our ability to accurately interpret herd dynamics and consequently restrict practical applications in farm management.

To directly confront these shortcomings, this review explicitly underscores innovative methodological intersections. Pose-aware frameworks and multi-camera fusion methods represent novel avenues to enhance semantic richness and improve the granularity of interaction interpretation. Pose-aware tracking leverages keypoint detection to recognize specific behaviors and intentional social signals between cows, moving well beyond basic proximity measures. Similarly, multi-camera fusion addresses critical problems like occlusion and lost animal identity, common hurdles in practical barn settings. These methodological innovations significantly strengthen the robustness and accuracy of AI-driven cattle monitoring systems, thus elevating their utility in real-world farming scenarios.

Moreover, this review uniquely emphasizes ethical dimensions within precision livestock farming, raising compelling questions about animal welfare, data governance, and algorithm transparency. Continuous surveillance methods, although technically promising, carry the risk of inducing stress and anxiety among cows, sensitive animals capable of complex emotional responses. The adoption of advanced AI technologies must, therefore, be accompanied by a thoughtful ethical framework prioritizing animal well-being and transparent data management. Such ethical considerations not only bolster the credibility of these technological systems but also foster acceptance and trust among farmers and consumers alike. This explicit acknowledgment of ethical responsibility represents an essential novelty and provides a solid foundation for sustainable long-term deployment of AI in dairy farms.

In terms of practical, real-world applications, the implications of integrating advanced AI methods with SNA are substantial and immediate. These methodologies empower farmers with precise, real-time insights into herd health, behavioral changes, and social dynamics, directly translating into informed management decisions. For instance, the ability to detect subtle shifts in social interactions can serve as an early warning sign for stress or illness, enabling timely intervention and reducing the economic impact of health-related disruptions. Additionally, recognizing stable social affiliations and hierarchies informs grouping strategies, mitigating stress during regrouping and optimizing milk production. Hence, these technological advancements hold significant promise not just theoretically, but practically, offering tangible benefits to daily farm operations.

Nevertheless, achieving the full potential of AI-driven cattle monitoring systems depends on addressing several critical challenges. The lack of standardized, openly accessible multimodal datasets with comprehensive behavioral annotations remains a major barrier to reproducibility and generalizability across different farm contexts. Without robust, standardized benchmarks, progress in this field risks becoming fragmented, slowing the advancement of universally applicable AI solutions. Future research must thus prioritize the creation of openly available datasets, comprehensive annotations, and modular validation protocols applicable to real farm environments.

Ultimately, the convergence of AI, animal ethology, and SNA embodies more than a mere technical evolution—it signifies a philosophical transformation in dairy farming practices. This shift calls for viewing dairy cows not merely as producers of milk but as social beings deserving of nuanced understanding and care. Integrating advanced AI methodologies with SNA promotes an empathetic, ethically responsible approach to animal management, profoundly enhancing welfare and productivity. By embracing this compassionate technological ethos, the dairy industry can achieve operational excellence while fostering a deeper respect and understanding of animal life, setting a powerful precedent for future agricultural practices.

Author Contributions: Conceptualization, S.N.; methodology, S.P.; formal analysis, S.P.; investigation, S.P.; S.N.; resources, S.N.; data curation, S.P.; writing—original draft preparation, S.P.; writing—review and editing, S.N.; K.S.; visualization, S.P.; supervision, S.N.; project administration, S.N.; funding acquisition, S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This work is kindly sponsored by the Natural Sciences and Engineering Research Council of Canada (RGPIN 2024-04450), and the Department of NB Agriculture (NB2425-0025).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank the Dairy Farmers of New Brunswick, Canada for the access to over 6 farms for our data collection and for consultation and advise on the on-farm daily operations of dairy farming.

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

Abbreviations

The following abbreviations are used in this manuscript:

ACmix	Attention Convolution Mix
AGS	Adaptive Graph Sampling
AI	Artificial Intelligence
AMS	Automated Milking System
ANN	Artificial Neural Network
BCS	Body Condition Scores
BiFPN	Bidirectional Feature Pyramid Network
BiLSTM	Bi-directional Long Short-Term Memory
CA	Coordinate Attention
CBAM	Convolutional Block Attention Module
CMC	Cumulative Matching Characteristic
CMCL	Cross-Model Contrastive Learning
CMT	Convolutional Neural Networks Meet Vision Transformers
CNN	Convolutional Neural Networks
ConvLSTM	Convolutional Long Short-Term Memory
DAN	Deep Affinity Network
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DETR	Detection Transformer
FAST	Features from Accelerated Segment Test
FLANN	Fast Library for Approximate Nearest Neighbors
FLOP	Floating Point Operation
GNN	Graph Neural Network
GPS	Global Positioning System
GPU	Graphics Processing Unit
HOTA	Higher Order Tracking Accuracy
ID	Identification / Identification Score
IDF1	Identification F1 Score
IMU	Inertial Measurement Unit
IR	Infrared
JDAN	Joint Detection and Association Network
KPI	Key Performance Indicator
LKA	Locating Key Area
LSTM	Long Short-Term Memory
mAP	Mean Average Precision
MOT	Multi-Object Tracking
MOTA	Multiple Object Tracking Accuracy
OTB	Object Tracking Benchmark
PBVM	Phase-Based Video Magnification
PETR	Position Embedding Transformation
PLF	Precision Livestock Farming
R-CNN	Region-based Convolutional Neural Network
re-ID	Re Identification
RFID	Radio Frequency Identification
RGB	Red-Green-Blue
RGB-D	Red-Green-Blue + Depth
RTLS	Real Time Locating System
SfM	Structure-from-Motion
SIFT	Scale-Invariant Feature Transform
SNA	Social Network Analysis
SOT	Single-Object Tracker / Tracking

SPPCSPC	Spatial Pyramid Pooling—Cross Stage Partial Connections
SSD	Single Shot Detector
STERGM	Separable Temporal Exponential Random Graph Models
SURABHI	Self-Training Using Rectified Annotations-Based Hard Instances
SVM	Support Vector Machine
TSN	Temporal Segment Network
UWB	Ultra-wide Band
VGG	Visual Geometry Group
VOT	Visual Object Tracking
YOLO	You Only Look Once

References

1. Machado, T.M.P., Machado Filho, L.C.P., Daros, R.R., Machado, G.T.B.P. and Hötzel, M.J., 2020. Licking and agonistic interactions in grazing dairy cows as indicators of preferential companies. *Applied Animal Behaviour Science*, 227, p.104994. DOI: <https://doi.org/10.1016/j.applanim.2020.104994>
2. Marina, H., Fikse, W.F. and Rönnegård, L., 2024. Social network analysis to predict social behavior in dairy cattle. *JDS communications*, 5(6), pp.608-612. DOI: <https://doi.org/10.3168/jdsc.2023-0507>
3. Hosseiniorbin, S., Layeghy, S., Kusy, B., Jurdak, R., Bishop-Hurley, G.J., Greenwood, P.L. and Portmann, M., 2021. Deep learning-based cattle behaviour classification using joint time-frequency data representation. *Computers and electronics in agriculture*, 187, p.106241. DOI: <https://doi.org/10.1016/j.compag.2021.106241>
4. Page, M.J., Moher, D., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E. and Chou, R., 2021. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *bmj*, 372. DOI: <https://doi.org/10.1136/bmj.n160>
5. Alpaydin, E., 2021. "4 NEURAL NETWORKS AND DEEP LEARNING," in *Machine Learning*. MIT Press, pp.105-141.
6. Foris, B., Zebunke, M., Langbein, J. and Melzer, N., 2019. Comprehensive analysis of affiliative and agonistic social networks in lactating dairy cattle groups. *Applied Animal Behaviour Science*, 210, pp.60-67. DOI: <https://doi.org/10.1016/j.applanim.2018.10.016>
7. De Freslon, I., Peralta, J.M., Strappini, A.C. and Monti, G., 2020. Understanding allogrooming through a dynamic social network approach: an example in a group of dairy cows. *Frontiers in Veterinary Science*, 7, p.535. DOI: <https://doi.org/10.3389/fvets.2020.00535>
8. Rocha, L.E., Terenius, O., Veissier, I., Meunier, B. and Nielsen, P.P., 2020. Persistence of sociality in group dynamics of dairy cattle. *Applied Animal Behaviour Science*, 223, p.104921. DOI: <https://doi.org/10.1016/j.applanim.2019.104921>
9. de Sousa, K.T., Machado Filho, L.C.P., Bica, G.S., Deniz, M. and Hötzel, M.J., 2021. Degree of affinity among dairy heifers affects access to feed supplementation. *Applied Animal Behaviour Science*, 234, p.105172. DOI: <https://doi.org/10.1016/j.applanim.2020.105172>
10. Foris, B., Haas, H.G., Langbein, J. and Melzer, N., 2021. Familiarity influences social networks in dairy cows after regrouping. *Journal of Dairy Science*, 104(3), pp.3485-3494. DOI: <https://doi.org/10.3168/jds.2020-18896>
11. Reyes, F.S., White, H.M., Weigel, K.A. and Van Os, J.M., 2023. Social interactions, feeding patterns, and feed efficiency of same-and mixed-parity groups of lactating cows. *Journal of Dairy Science*, 106(12), pp.9410-9425. DOI: <https://doi.org/10.3168/jds.2023-23295>
12. Vázquez-Diosdado, J.A., Occhiuto, F., Carslake, C. and Kaler, J., 2023. Familiarity, age, weaning and health status impact social proximity networks in dairy calves. *Scientific Reports*, 13(1), p.2275. DOI: <https://doi.org/10.1038/s41598-023-29309-1>
13. Burke, K.C., Gingerich, K. and Miller-Cushon, E.K., 2024. Factors associated with the variation and consistency of social network position in group-housed calves. *Applied Animal Behaviour Science*, 271, p.106169. DOI: <https://doi.org/10.1016/j.applanim.2024.106169>

14. Clein, D., Burke, K.C. and Miller-Cushon, E.K., 2024. Characterizing social networks and influence of early-life social housing in weaned heifers on pasture. *JDS communications*, 5(5), pp.441-446. DOI: <https://doi.org/10.3168/jdsc.2023-0473>
15. Marina, H., Ren, K., Hansson, I., Fikse, F., Nielsen, P.P. and Rönnegård, L., 2024. New insight into social relationships in dairy cows and how time of birth, parity, and relatedness affect spatial interactions later in life. *Journal of Dairy Science*, 107(2), pp.1110-1123. DOI: <https://doi.org/10.3168/jds.2023-23483>
16. Gutmann, A.K., Špinka, M. and Winckler, C., 2020. Do familiar group mates facilitate integration into the milking group after calving in dairy cows?. *Applied Animal Behaviour Science*, 229, p.105033. DOI: <https://doi.org/10.1016/j.applanim.2020.105033>
17. Krahn, J., Foris, B., Weary, D.M. and von Keyserlingk, M.A., 2023. Invited review: Social dominance in dairy cattle: A critical review with guidelines for future research. *Journal of Dairy Science*, 106(3), pp.1489-1501. DOI: <https://doi.org/10.3168/jds.2022-22534>
18. Chopra, K., Hodges, H.R., Barker, Z.E., Vázquez Diosdado, J.A., Amory, J.R., Cameron, T.C., Croft, D.P., Bell, N.J. and Codling, E.A., 2020. Proximity interactions in a permanently housed dairy herd: Network structure, consistency, and individual differences. *Frontiers in Veterinary Science*, 7, p.583715. DOI: <https://doi.org/10.3389/fvets.2020.583715>
19. Marumo, J.L., Fisher, D.N., Lusseau, D., Mackie, M., Speakman, J.R. and Hambly, C., 2022. Social associations in lactating dairy cows housed in a robotic milking system. *Applied Animal Behaviour Science*, 249, p.105589. DOI: <https://doi.org/10.1016/j.applanim.2022.105589>
20. Burke, K.C., do Nascimento-Emond, S., Hixson, C.L. and Miller-Cushon, E.K., 2022. Social networks respond to a disease challenge in calves. *Scientific Reports*, 12(1), p.9119. DOI: <https://doi.org/10.1038/s41598-022-13088-2>
21. Fadul-Pacheco, L., Liou, M., Reinemann, D.J. and Cabrera, V.E., 2021. A preliminary investigation of social network analysis applied to dairy cow behavior in automatic milking system environments. *Animals*, 11(5), p.1229. DOI: <https://doi.org/10.3390/ani11051229>
22. Smith, L.A., Swain, D.L., Innocent, G.T. and Hutchings, M.R., 2023. Social isolation of unfamiliar cattle by groups of familiar cattle. *Behavioural Processes*, 207, p.104847. DOI: <https://doi.org/10.1016/j.beproc.2023.104847>
23. Liu, M., Wu, Y., Li, G., Liu, M., Hu, R., Zou, H., Wang, Z. and Peng, Y., 2023. Classification of cow behavior patterns using inertial measurement units and a fully convolutional network model. *Journal of Dairy Science*, 106(2), pp.1351-1359. DOI: <https://doi.org/10.3168/jds.2022-22350>
24. Deepak, D., D'Mello, D.A. and Divakarla, U., 2024, March. Advancements in Automated Livestock Monitoring: A Concise Review of Deep Learning-Based Cattle Activity Recognition. In *2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)* (Vol. 1, pp. 321-327). IEEE. DOI: <https://doi.org/10.1109/ICACCS60874.2024.10717327>
25. Melzer, N., Foris, B. and Langbein, J., 2021. Validation of a real-time location system for zone assignment and neighbor detection in dairy cow groups. *Computers and Electronics in Agriculture*, 187, p.106280. DOI: <https://doi.org/10.1016/j.compag.2021.106280>
26. Jowett, S., Barker, Z. and Amory, J., 2022. The structure and temporal changes in brokerage typologies applied to a dynamic sow herd. *Applied Animal Behaviour Science*, 246, p.105509. DOI: <https://doi.org/10.1016/j.applanim.2021.105509>
27. Besler, B.C., Mojabi, P., Lasemiimani, Z., Murphy, J.E., Wang, Z., Baker, R., Pearson, J.M. and Fear, E.C., 2024. Scoping review of precision technologies for cattle monitoring. *Smart Agricultural Technology*, 9, p.100596. DOI: <https://doi.org/10.1016/j.atech.2024.100596>
28. Araújo, V.M., Rili, I., Gisiger, T., Gambs, S., Vasseur, E., Cellier, M. and Diallo, A.B., 2025. AI-Powered Cow Detection in Complex Farm Environments. *Smart Agricultural Technology*, 10, p.100770. DOI: <https://doi.org/10.1016/j.atech.2025.100770>
29. Xu, P., Zhang, Y., Ji, M., Guo, S., Tang, Z., Wang, X., Guo, J., Zhang, J. and Guan, Z., 2024. Advanced intelligent monitoring technologies for animals: A survey. *Neurocomputing*, 585, p.127640. DOI: <https://doi.org/10.1016/j.neucom.2024.127640>

30. Fatoki, O., Du, C., Hans, R. and Bello, R.W., 2024. Role of computer vision and deep learning algorithms in livestock behavioural recognition: A state-of-the-art-review. DOI: <https://dx.doi.org/10.2139/ssrn.5029111>
31. Shorten, P.R., 2021. Computer vision and weigh scale-based prediction of milk yield and udder traits for individual cows. *Computers and Electronics in Agriculture*, 188, p.106364. DOI: <https://doi.org/10.1016/j.compag.2021.106364>
32. Mg, W.H.E., Tin, P., Aikawa, M., Kobayashi, I., Horii, Y., Honkawa, K. and Zin, T.T., 2024. Customized Tracking Algorithm for Robust Cattle Detection and Tracking in Occlusion Environments. *Sensors (Basel, Switzerland)*, 24(4). DOI: <https://doi.org/10.3390/s24041181>
33. Li, G., Huang, Y., Chen, Z., Chesser Jr, G.D., Purswell, J.L., Linhoss, J. and Zhao, Y., 2021. Practices and applications of convolutional neural network-based computer vision systems in animal farming: A review. *Sensors*, 21(4), p.1492. DOI: <https://doi.org/10.3390/s21041492>
34. Zhang, Y., Wang, X., Ye, X., Zhang, W., Lu, J., Tan, X., Ding, E., Sun, P. and Wang, J., 2023. ByteTrackV2: 2D and 3D multi-object tracking by associating every detection box. *arXiv preprint arXiv:2303.15334*. DOI: <https://doi.org/10.48550/arXiv.2303.15334>
35. Zhang, W., Wang, Y., Guo, L., Falzon, G., Kwan, P., Jin, Z., Li, Y. and Wang, W., 2024. Analysis and Comparison of New-Born Calf Standing and Lying Time Based on Deep Learning. *Animals*, 14(9), p.1324. DOI: <https://doi.org/10.3390/ani14091324>
36. Qiao, Y., Su, D., Kong, H., Sukkarieh, S., Lomax, S. and Clark, C., 2019. Individual cattle identification using a deep learning based framework. *IFAC-PapersOnLine*, 52(30), pp.318-323. DOI: <https://doi.org/10.1016/j.ifacol.2019.12.558>
37. Papa, M., de Medeiros Oliveira, S.R. and Bergier, I., 2024. Technologies in cattle traceability: A bibliometric analysis. *Computers and Electronics in Agriculture*, 227, p.109459. DOI: <https://doi.org/10.1016/j.compag.2024.109459>
38. Achour, B., Belkadi, M., Filali, I., Laghrouche, M. and Lahdir, M., 2020. Image analysis for individual identification and feeding behaviour monitoring of dairy cows based on Convolutional Neural Networks (CNN). *Biosystems Engineering*, 198, pp.31-49. DOI: <https://doi.org/10.1016/j.biosystemseng.2020.07.019>
39. Peng, Y., Chen, Y., Yang, Y., Liu, M., Hu, R., Zou, H., Xiao, J., Jiang, Y., Wang, Z. and Xu, L., 2024. A multimodal classification method: Cow behavior pattern classification with improved EdgeNeXt using an inertial measurement unit. *Computers and Electronics in Agriculture*, 226, p.109453. DOI: <https://doi.org/10.1016/j.compag.2024.109453>
40. Chen, X., Yang, T., Mai, K., Liu, C., Xiong, J., Kuang, Y. and Gao, Y., 2022. Holstein cattle face re-identification unifying global and part feature deep network with attention mechanism. *Animals*, 12(8), p.1047. DOI: <https://doi.org/10.3390/ani12081047>
41. Zaidi, S.S.A., Ansari, M.S., Aslam, A., Kanwal, N., Asghar, M. and Lee, B., 2022. A survey of modern deep learning based object detection models. *Digital Signal Processing*, 126, p.103514. DOI: <https://doi.org/10.1016/j.dsp.2022.103514>
42. Zheng, Z., Li, J. and Qin, L., 2023. YOLO-BYTE: An efficient multi-object tracking algorithm for automatic monitoring of dairy cows. *Computers and Electronics in Agriculture*, 209, p.107857. DOI: <https://doi.org/10.1016/j.compag.2023.107857>
43. Mar, C.C., Zin, T.T., Tin, P., Honkawa, K., Kobayashi, I. and Horii, Y., 2023. Cow detection and tracking system utilizing multi-feature tracking algorithm. *Scientific reports*, 13(1), p.17423. DOI: <https://doi.org/10.1038/s41598-023-44669-4>
44. Hao, W., Zhang, K., Han, M., Hao, W., Wang, J., Li, F. and Liu, Z., 2023. A novel Jinnan individual cattle recognition approach based on mutual attention learning scheme. *Expert Systems with Applications*, 230, p.120551. DOI: <https://doi.org/10.1016/j.eswa.2023.120551>
45. Gao, G., Wang, C., Wang, J., Lv, Y., Li, Q., Ma, Y., Zhang, X., Li, Z. and Chen, G., 2023. CNN-Bi-LSTM: A complex environment-oriented cattle behavior classification network based on the fusion of CNN and Bi-LSTM. *Sensors*, 23(18), p.7714. DOI: <https://doi.org/10.3390/s23187714>
46. Wang, Y., Xu, X., Wang, Z., Li, R., Hua, Z. and Song, H., 2023. ShuffleNet-Triplet: A lightweight RE-identification network for dairy cows in natural scenes. *Computers and Electronics in Agriculture*, 205, p.107632. DOI: <https://doi.org/10.1016/j.compag.2023.107632>

47. Fuentes, A., Yoon, S., Park, J. and Park, D.S., 2020. Deep learning-based hierarchical cattle behavior recognition with spatio-temporal information. *Computers and Electronics in Agriculture*, 177, p.105627. DOI: <https://doi.org/10.1016/j.compag.2020.105627>
48. Amjoud, A.B. and Amrouch, M., 2023. Object detection using deep learning, CNNs and vision transformers: A review. *IEEE Access*, 11, pp.35479-35516. DOI: <https://doi.org/10.1109/ACCESS.2023.3266093>
49. Mon, S.L., Zin, T.T., Tin, P. and Kobayashi, I., 2022, October. Video-based automatic cattle identification system. In *2022 IEEE 11th Global Conference on Consumer Electronics (GCCE)* (pp. 490-491). IEEE. DOI: <https://doi.org/10.1109/GCCE56475.2022.10014109>
50. Dulal, R., Zheng, L., Kabir, M.A., McGrath, S., Medway, J., Swain, D. and Swain, W., 2022, November. Automatic cattle identification using yolov5 and mosaic augmentation: A comparative analysis. In *2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1-8). IEEE. DOI: <https://doi.org/10.1109/DICTA56598.2022.10034585>
51. Yousra, T., Afridi, H., Tarekegn, A.N., Ullah, M., Beghdadi, A. and Cheikh, F.A., 2023, October. Self-supervised Animal Detection in Indoor Environment. In *2023 Twelfth International Conference on Image Processing Theory, Tools and Applications (IPTA)* (pp. 1-6). IEEE. DOI: <https://doi.org/10.1109/IPTA59101.2023.10320047>
52. McDonagh, J., Tzimiropoulos, G., Slinger, K.R., Huggett, Z.J., Down, P.M. and Bell, M.J., 2021. Detecting dairy cow behavior using vision technology. *Agriculture*, 11(7), p.675. DOI: <https://doi.org/10.3390/agriculture11070675>
53. Tan, M., Pang, R. and Le, Q.V., 2020. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10781-10790). DOI: <https://doi.org/10.48550/arXiv.1911.09070>
54. Fuentes, A., Han, S., Nasir, M.F., Park, J., Yoon, S. and Park, D.S., 2023. Multiview monitoring of individual cattle behavior based on action recognition in closed barns using deep learning. *Animals*, 13(12), p.2020. DOI: <https://doi.org/10.3390/ani13122020>
55. Huang, X., Hu, Z., Qiao, Y. and Sukkarieh, S., 2022. Deep learning-based cow tail detection and tracking for precision livestock farming. *IEEE/ASME Transactions on Mechatronics*, 28(3), pp.1213-1221. DOI: <https://doi.org/10.1109/TMECH.2022.3175377>
56. Zhang, Y., Tian, Q., Liu, T. and Kong, J., 2022. Dynamic trajectory quantification strategy for multiple object tracking with feature rearrangement. *Journal of Electronic Imaging*, 31(6), pp.063025-063025. DOI: <https://doi.org/10.1117/1.JEL.31.6.063025>
57. Myat Noe, S., Zin, T.T., Tin, P. and Kobayashi, I., 2023. Comparing state-of-the-art deep learning algorithms for the automated detection and tracking of black cattle. *Sensors*, 23(1), p.532. DOI: <https://doi.org/10.3390/s23010532>
58. Tassinari, P., Bovo, M., Benni, S., Franzoni, S., Poggi, M., Mammi, L.M.E., Mattoccia, S., Di Stefano, L., Bonora, F., Barbaresi, A. and Santolini, E., 2021. A computer vision approach based on deep learning for the detection of dairy cows in free stall barn. *Computers and Electronics in Agriculture*, 182, p.106030. DOI: <https://doi.org/10.1016/j.compag.2021.106030>
59. Qiao, Y., Clark, C., Lomax, S., Kong, H., Su, D. and Sukkarieh, S., 2021. Automated individual cattle identification using video data: a unified deep learning architecture approach. *Frontiers in Animal Science*, 2, p.759147. DOI: <https://doi.org/10.3389/fanim.2021.759147>
60. Salau, J. and Krieter, J., 2020. Instance segmentation with Mask R-CNN applied to loose-housed dairy cows in a multi-camera setting. *Animals*, 10(12), p.2402. DOI: <https://doi.org/10.3390/ani10122402>
61. Wang, B., Li, X., An, X., Duan, W., Wang, Y., Wang, D. and Qi, J., 2024. Open-Set Recognition of Individual Cows Based on Spatial Feature Transformation and Metric Learning. *Animals*, 14(8), p.1175. DOI: <https://doi.org/10.3390/ani14081175>
62. Dac, H.H., Gonzalez Viejo, C., Lipovetzky, N., Tongson, E., Dunshea, F.R. and Fuentes, S., 2022. Livestock identification using deep learning for traceability. *Sensors*, 22(21), p.8256. DOI: <https://doi.org/10.3390/s22218256>

63. Qiao, Y., Guo, Y., Yu, K. and He, D., 2022. C3D-ConvLSTM based cow behaviour classification using video data for precision livestock farming. *Computers and electronics in agriculture*, 193, p.106650. DOI: <https://doi.org/10.1016/j.compag.2021.106650>
64. Sharma, A., Randewich, L., Andrew, W., Hannuna, S., Campbell, N., Mullan, S., Dowsey, A.W., Smith, M., Hansen, M. and Burghardt, T., 2025. Universal bovine identification via depth data and deep metric learning. *Computers and Electronics in Agriculture*, 229, p.109657. DOI: <https://doi.org/10.1016/j.compag.2024.109657>
65. Ramesh, M. and Reibman, A.R., 2024. SURABHI: Self-Training Using Rectified Annotations-Based Hard Instances for Eidetic Cattle Recognition. *Sensors (Basel, Switzerland)*, 24(23), p.7680. DOI: <https://doi.org/10.3390/s24237680>
66. Wang, Y., Kitani, K. and Weng, X., 2021, May. Joint object detection and multi-object tracking with graph neural networks. In *2021 IEEE international conference on robotics and automation (ICRA)* (pp. 13708-13715). IEEE. DOI: <https://doi.org/10.1109/ICRA48506.2021.9561110>
67. Andrew, W., Gao, J., Mullan, S., Campbell, N., Dowsey, A.W. and Burghardt, T., 2021. Visual identification of individual Holstein-Friesian cattle via deep metric learning. *Computers and Electronics in Agriculture*, 185, p.106133. DOI: <https://doi.org/10.1016/j.compag.2021.106133>
68. Ardö, H., Guzhva, O., Nilsson, M. and Herlin, A.H., 2018. Convolutional neural network-based cow interaction watchdog. *IET Computer Vision*, 12(2), pp.171-177. DOI: <https://doi.org/10.1049/iet-cvi.2017.0077>
69. Shakeel, P.M., bin Mohd Aboobaidar, B. and Salahuddin, L.B., 2022. A deep learning-based cow behavior recognition scheme for improving cattle behavior modeling in smart farming. *Internet of Things*, 19, p.100539. DOI: <https://doi.org/10.1016/j.iot.2022.100539>
70. Lyu, Y., Yang, M.Y., Vosselman, G. and Xia, G.S., 2021. Video object detection with a convolutional regression tracker. *ISPRS journal of photogrammetry and remote sensing*, 176, pp.139-150. DOI: <https://doi.org/10.1016/j.isprsjprs.2021.04.004>
71. Li, Y., Gou, X., Zuo, H. and Zhang, M., 2024, July. A Multi-scale Cattle Individual Identification Method Based on CMT Module and Attention Mechanism. In *2024 7th International Conference on Computer Information Science and Application Technology (CISAT)* (pp. 336-341). IEEE. DOI: <https://doi.org/10.1109/CISAT62382.2024.10695246>
72. Neethirajan, S. and Kemp, B., 2021. Social network analysis in farm animals: Sensor-based approaches. *Animals*, 11(2), p.434. DOI: <https://doi.org/10.3390/ani11020434>
73. Yu, P., Burghardt, T., Dowsey, A.W. and Campbell, N.W., 2024. MultiCamCows2024--A Multi-view Image Dataset for AI-driven Holstein-Friesian Cattle Re-Identification on a Working Farm. *arXiv preprint arXiv:2410.12695*. DOI: <https://doi.org/10.48550/arXiv.2410.12695>
74. Chen, C. and Li, D., 2021. [Retracted] Research on the Detection and Tracking Algorithm of Moving Object in Image Based on Computer Vision Technology. *Wireless Communications and Mobile Computing*, 2021(1), p.1127017. DOI: <https://doi.org/10.1155/2021/1127017>
75. Ren, K., Bernes, G., Hetta, M. and Karlsson, J., 2021. Tracking and analysing social interactions in dairy cattle with real-time locating system and machine learning. *Journal of Systems Architecture*, 116, p.102139. DOI: <https://doi.org/10.1016/j.sysarc.2021.102139>
76. Ozella, L., Magliola, A., Vernengo, S., Ghigo, M., Bartoli, F., Grangetto, M., Forte, C., Montrucchio, G., BROTTI REBULI, K. and Giacobini, M., 2023. A computer vision approach for the automatic detection of social interactions of dairy cows in automatic milking systems. In *Proceedings of 2023 IEEE International Workshop on Measurements and Applications in Veterinary and Animal Sciences* (pp. 267-272). IEEE. DOI: <http://dx.doi.org/10.21014/actaimeko.v13i3.1628>
77. Yangyang, G.U.O., Shuzeng, D.U., Yongliang, Q.I.A.O. and Dong, L.I.A.N.G., 2023. Advances in the applications of deep learning technology for livestock smart farming. *Smart Agriculture*, 5(1), p.52. DOI: <https://doi.org/10.12133/j.smartag.SA202205009>
78. Xiang, X., Ren, W., Qiu, Y., Zhang, K. and Lv, N., 2021. Multi-object tracking method based on efficient channel attention and switchable atrous convolution. *Neural Processing Letters*, 53(4), pp.2747-2763. DOI: <https://doi.org/10.1007/s11063-021-10519-5>

79. Sun, S., Akhtar, N., Song, H., Mian, A. and Shah, M., 2019. Deep affinity network for multiple object tracking. *IEEE transactions on pattern analysis and machine intelligence*, 43(1), pp.104-119. DOI: <https://doi.org/10.1109/TPAMI.2019.2929520>
80. Wang, C., Wang, Y., Wang, Y., Wu, C.T. and Yu, G., 2019. muSSP: Efficient min-cost flow algorithm for multi-object tracking. *Advances in neural information processing systems*, 32. DOI: <https://dl.acm.org/doi/10.5555/3454287.3454326>
81. Dendorfer, P., 2020. Mot20: A benchmark for multi object tracking in crowded scenes. *arXiv preprint arXiv:2003.09003*. DOI: <https://doi.org/10.48550/arXiv.2003.09003>
82. Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), pp.84-90. DOI: <http://dx.doi.org/10.1145/3065386>
83. Himabindu, D.D. and Kumar, S.P., 2021. A survey on computer vision architectures for large scale image classification using deep learning. *International Journal of Advanced Computer Science and Applications*, 12(10). DOI: <https://dx.doi.org/10.14569/IJACSA.2021.0121013>
84. Bello, R.W., Talib, A.Z., Mohamed, A.S.A., Olubummo, D.A. and Ootob, F.N., 2020. Image-based individual cow recognition using body patterns. *Image*, 11(3), pp.92-98. DOI: <http://dx.doi.org/10.14569/IJACSA.2020.0110311>
85. Ramesh, M., Reibman, A.R. and Boerman, J.P., 2023. Eidetic recognition of cattle using keypoint alignment. *Electronic Imaging*, 35, pp.279-1. DOI: <https://doi.org/10.2352/EI.2023.35.7.IMAGE-279>
86. Fu, L., Li, S., Kong, S., Ni, R., Pang, H., Sun, Y., Hu, T., Mu, Y., Guo, Y. and Gong, H., 2022. Lightweight individual cow identification based on Ghost combined with attention mechanism. *Plos one*, 17(10), p.e0275435. DOI: <https://doi.org/10.1371/journal.pone.0275435>
87. Xiao, J., Liu, G., Wang, K. and Si, Y., 2022. Cow identification in free-stall barns based on an improved Mask R-CNN and an SVM. *Computers and Electronics in Agriculture*, 194, p.106738. DOI: <https://doi.org/10.1016/j.compag.2022.106738>
88. Mon, S.L., Onizuka, T., Tin, P., Aikawa, M., Kobayashi, I. and Zin, T.T., 2024. AI-enhanced real-time cattle identification system through tracking across various environments. *Scientific Reports*, 14(1), p.17779. DOI: <https://doi.org/10.1038/s41598-024-68418-3>
89. Gunda, V.S.P., Gulla, H., Kosana, V. and Janapati, S., 2022, November. A hybrid deep learning based robust framework for cattle identification. In *2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)* (pp. 1-5). IEEE. DOI: <https://doi.org/10.1109/ASSIC55218.2022.10088414>
90. Mahato, S. and Neethirajan, S., 2024. Integrating artificial intelligence in dairy farm management—biometric facial recognition for cows. *Information Processing in Agriculture*. DOI: <https://doi.org/10.1016/j.inpa.2024.10.001>
91. Okura, F., Ikuma, S., Makihara, Y., Muramatsu, D., Nakada, K. and Yagi, Y., 2019. RGB-D video-based individual identification of dairy cows using gait and texture analyses. *Computers and Electronics in Agriculture*, 165, p.104944. DOI: <https://doi.org/10.1016/j.compag.2019.104944>
92. Chen, J., Xi, Z., Wei, C., Lu, J., Niu, Y. and Li, Z., 2020. Multiple object tracking using edge multi-channel gradient model with ORB feature. *IEEE Access*, 9, pp.2294-2309. DOI: <https://doi.org/10.1109/ACCESS.2020.3046763>
93. Wang, Y., Xu, X., Zhang, S., Wen, Y., Pu, L., Zhao, Y. and Song, H., 2024. Adaptive group sample with central momentum contrast loss for unsupervised individual identification of cows in changeable conditions. *Applied Soft Computing*, 167, p.112340. DOI: <https://doi.org/10.1016/j.asoc.2024.112340>
94. Zhao, K., Jin, X., Ji, J., Wang, J., Ma, H. and Zhu, X., 2019. Individual identification of Holstein dairy cows based on detecting and matching feature points in body images. *Biosystems Engineering*, 181, pp.128-139. DOI: <https://doi.org/10.1016/j.biosystemseng.2019.03.004>
95. Kalmukov, Y., Evstatiev, B. and Kadirova, S., 2024. Individual Cow Identification Using Non-Fixed Point-of-View Images and Deep Learning. *International Journal of Advanced Computer Science & Applications*, 15(10). DOI: <https://doi.org/10.14569/IJACSA.2024.0151066>
96. Qiao, Y., Su, D., Kong, H., Sukkarieh, S., Lomax, S. and Clark, C., 2020, August. BiLSTM-based individual cattle identification for automated precision livestock farming. In *2020 IEEE 16th International Conference*

- on Automation Science and Engineering (CASE) (pp. 967-972). IEEE. DOI: <https://doi.org/10.1109/CASE48305.2020.9217026>
97. Menezes, G.L., Negreiro, A., Ferreira, R. and Dórea, J.R.R., 2023, May. Identifying dairy cows using body surface keypoints through supervised machine learning. In Proceedings of the 2ND US Precision Livestock Farming Conference (USPLF 2023). DOI: <https://api.semanticscholar.org/CorpusID:266759896>
 98. Meng, H., Zhang, L., Yang, F., Hai, L., Wei, Y., Zhu, L. and Zhang, J., 2025. Livestock Biometrics Identification Using Computer Vision Approaches: A Review. *Agriculture*, 15(1), p.102. DOI: <https://doi.org/10.3390/agriculture15010102>
 99. Lu, Y., Weng, Z., Zheng, Z., Zhang, Y. and Gong, C., 2023. Algorithm for cattle identification based on locating key area. *Expert Systems with Applications*, 228, p.120365. DOI: <https://doi.org/10.1016/j.eswa.2023.120365>
 100. Yang, G., Xu, X., Song, L., Zhang, Q., Duan, Y. and Song, H., 2022. Automated measurement of dairy cows body size via 3D point cloud data analysis. *Computers and electronics in agriculture*, 200, p.107218. DOI: <https://doi.org/10.1016/j.compag.2022.107218>
 101. Li, J., Ma, W., Zhao, C., Li, Q., Tulpan, D., Wang, Z., Yang, S.X., Ding, L., Gao, R. and Yu, L., 2022. Extraction of key regions of beef cattle based on bidirectional tomographic slice features from point cloud data. *Computers and Electronics in Agriculture*, 199, p.107190. DOI: <https://doi.org/10.1016/j.compag.2022.107190>
 102. Foris, B., Thompson, A.J., Von Keyserlingk, M.A.G., Melzer, N. and Weary, D.M., 2019. Automatic detection of feeding-and drinking-related agonistic behavior and dominance in dairy cows. *Journal of dairy science*, 102(10), pp.9176-9186. DOI: <https://doi.org/10.3168/jds.2019-16697>
 103. Peng, Y., Kondo, N., Fujiura, T., Suzuki, T., Yoshioka, H. and Itoyama, E., 2019. Classification of multiple cattle behavior patterns using a recurrent neural network with long short-term memory and inertial measurement units. *Computers and electronics in agriculture*, 157, pp.247-253. DOI: <https://doi.org/10.1016/j.compag.2018.12.023>
 104. Guzhva, O., Ardö, H., Nilsson, M., Herlin, A. and Tufvesson, L., 2018. Now you see me: Convolutional neural network based tracker for dairy cows. *Frontiers in Robotics and AI*, 5, p.107. DOI: <https://doi.org/10.3389/frobt.2018.00107>
 105. Wang, L., Xiong, Y., Wang, Z., Qiao, Y., Lin, D., Tang, X. and Van Gool, L., 2018. Temporal segment networks for action recognition in videos. *IEEE transactions on pattern analysis and machine intelligence*, 41(11), pp.2740-2755. DOI: <https://doi.org/10.1109/TPAMI.2018.2868668>
 106. Avanzato, R., Beritelli, F. and Puglisi, V.F., 2022, November. Dairy cow behavior recognition using computer vision techniques and CNN networks. In 2022 IEEE International Conference on Internet of Things and Intelligence Systems (IoT&IS) (pp. 122-128). IEEE. DOI: <https://doi.org/10.1109/IoT&IS56727.2022.9975979>
 107. Wu, D., Han, M., Song, H., Song, L. and Duan, Y., 2023. Monitoring the respiratory behavior of multiple cows based on computer vision and deep learning. *Journal of Dairy Science*, 106(4), pp.2963-2979. DOI: <https://doi.org/10.3168/jds.2022-22501>
 108. Bhujel, A., Wang, Y., Lu, Y., Morris, D. and Dangol, M., 2024. Public Computer Vision Datasets for Precision Livestock Farming: A Systematic Survey. *arXiv preprint arXiv:2406.10628*. DOI: <https://doi.org/10.48550/arXiv.2406.10628>
 109. Wang, R., Gao, R., Li, Q., Zhao, C., Ru, L., Ding, L., Yu, L. and Ma, W., 2024. An ultra-lightweight method for individual identification of cow-back pattern images in an open image set. *Expert Systems with Applications*, 249, p.123529. DOI: <https://doi.org/10.1016/j.eswa.2024.123529>
 110. Wang, J., Wu, J., Wu, J., Wang, J. and Wang, J., 2023. YOLOv7 optimization model based on attention mechanism applied in dense scenes. *Applied Sciences*, 13(16), p.9173. DOI: <https://doi.org/10.3390/app13169173>
 111. Wang, H., He, X., Li, Z., Yuan, J. and Li, S., 2023. JDAN: Joint detection and association network for real-time online multi-object tracking. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(1s), pp.1-17. DOI: <https://doi.org/10.1145/3533253>

112. Hossain, M.E., Kabir, M.A., Zheng, L., Swain, D.L., McGrath, S. and Medway, J., 2022. A systematic review of machine learning techniques for cattle identification: Datasets, methods and future directions. *Artificial Intelligence in Agriculture*, 6, pp.138-155. DOI: <https://doi.org/10.1016/j.aiia.2022.09.002>
113. Arulprakash, E. and Aruldoss, M., 2022. A study on generic object detection with emphasis on future research directions. *Journal of King Saud University-Computer and Information Sciences*, 34(9), pp.7347-7365. DOI: <https://doi.org/10.1016/j.jksuci.2021.08.001>
114. Wang, G., Song, M. and Hwang, J.N., 2022. Recent advances in embedding methods for multi-object tracking: a survey. *arXiv preprint arXiv:2205.10766*. DOI: <https://doi.org/10.48550/arXiv.2205.10766>
115. Gupta, H., Jindal, P., Verma, O.P., Arya, R.K., Ateya, A.A., Soliman, N.F. and Mohan, V., 2022. Computer vision-based approach for automatic detection of dairy cow breed. *Electronics*, 11(22), p.3791. DOI: <https://doi.org/10.3390/electronics11223791>
116. Chandra, M.A. and Bedi, S.S., 2021. Survey on SVM and their application in image classification. *International Journal of Information Technology*, 13(5), pp.1-11. DOI: <https://doi.org/10.1007/s41870-017-0080-1>
117. Qiao, Y., Kong, H., Clark, C., Lomax, S., Su, D., Eiffert, S. and Sukkarieh, S., 2021. Intelligent perception for cattle monitoring: A review for cattle identification, body condition score evaluation, and weight estimation. *Computers and electronics in agriculture*, 185, p.106143. DOI: <https://doi.org/10.1016/j.compag.2021.106143>
118. Oliveira, D.A.B., Pereira, L.G.R., Bresolin, T., Ferreira, R.E.P. and Dorea, J.R.R., 2021. A review of deep learning algorithms for computer vision systems in livestock. *Livestock Science*, 253, p.104700. DOI: <https://doi.org/10.1016/j.livsci.2021.104700>
119. Zhang, Y., Wang, C., Wang, X., Zeng, W. and Fairmot, W.L., 2021. On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision*, 129(11), pp.3069-3087. DOI: <https://doi.org/10.48550/arXiv.2004.01888>
120. Ren, K., Nielsen, P.P., Alam, M. and Rønnegård, L., 2021. Where do we find missing data in a commercial real-time location system? Evidence from 2 dairy farms. *JDS communications*, 2(6), pp.345-350. DOI: <https://doi.org/10.3168/jdsc.2020-0064>
121. Jia, N., Kootstra, G., Koerkamp, P.G., Shi, Z. and Du, S., 2021. Segmentation of body parts of cows in RGB-depth images based on template matching. *Computers and Electronics in Agriculture*, 180, p.105897. DOI: <https://doi.org/10.1016/j.compag.2020.105897>
122. Fuentes, S., Gonzalez Viejo, C., Tongson, E., Lipovetzky, N. and Dunshea, F.R., 2021. Biometric physiological responses from dairy cows measured by visible remote sensing are good predictors of milk productivity and quality through artificial intelligence. *Sensors*, 21(20), p.6844. DOI: <https://doi.org/10.3390/s21206844>
123. Liu, H., Reibman, A.R. and Boerman, J.P., 2020. Video analytic system for detecting cow structure. *Computers and Electronics in Agriculture*, 178, p.105761. DOI: <https://doi.org/10.1016/j.compag.2020.105761>
124. Karthik, S., Prabhu, A. and Gandhi, V., 2020. Simple unsupervised multi-object tracking. *arXiv preprint arXiv:2006.02609*. DOI: <https://doi.org/10.48550/arXiv.2006.02609>
125. Yi, Y., Luo, L. and Zheng, Z., 2019. Single online visual object tracking with enhanced tracking and detection learning. *Multimedia Tools and Applications*, 78, pp.12333-12351. DOI: <https://doi.org/10.1007/s11042-018-6787-6>
126. Dendorfer, P., Rezatofighi, H., Milan, A., Shi, J., Cremers, D., Reid, I., Roth, S., Schindler, K. and Leal-Taixe, L., 2019. CVPR19 tracking and detection challenge: How crowded can it get?. *arXiv preprint arXiv:1906.04567*. DOI: <https://doi.org/10.48550/arXiv.1906.04567>
127. Wurtz, K., Camerlink, I., D'Eath, R.B., Fernández, A.P., Norton, T., Steibel, J. and Siegford, J., 2019. Recording behaviour of indoor-housed farm animals automatically using machine vision technology: A systematic review. *PloS one*, 14(12), p.e0226669. DOI: <https://doi.org/10.1371/journal.pone.0226669>

128. Zhan, Y., Wang, C., Wang, X., Zeng, W. and Liu, W., 2020. A simple baseline for multi-object tracking. arXiv preprint, arXiv:2004.01888. DOI: <http://dx.doi.org/10.48550/arXiv.2004.01888>
129. Yang, Y., Komatsu, M., Ohkawa, T. and Oyama, K., 2022, December. Real-Time Cattle Interaction Recognition via Triple-stream Network. In 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 61-68). IEEE. DOI: <https://doi.org/10.1109/ICMLA55696.2022.00016>

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.