OPEN ACCESS Advances in Animal and Veterinary Sciences

Research Article

Automated Early Detection of Lameness in Dairy Cattle Using Multi-View Pose Estimation and Deep Learning on Unstructured Farm Surveillance Videos

G. Naga Rama Devi[1*], Bharath M B[2], Parimi Hema Sree[3], P. Naresh[2], Dara Rajesh Babu[4], N Konda Reddy[5], Mala B.A[2], K. Raghavendar[6]

[1]Dept of CSE, School of Engineering and Technology, Kaveri University, Gowraram, Telangana, India; [2]Department of Computer Science and Engineering, Dayananda Sagar University, Bengaluru, India; [3]Department of ECE, CVR College of Engineering, Hyderabad, India; [4]Department of Computer Science and Engineering, Srinivasa Ramanujan Institute of Technology, Anantapur, Andhra Pradesh, India; [5]Dept of Mathematics, Koneru Lakshmaiah Education Foundation, Greenfields, Vaddeswaram, Guntur, Andhra Pradesh, India; [6]Department of Computer Science and Engineering, ACE Engineering College, Hyderabad – 500031, Telangana, India.

Abstract | Lameness is a widespread disease in dairy cattle, causing decreasing productivity, animal welfare and economic losses. The early and robust identification of abnormal locomotion is very important for the management of the herd. Visual inspection of locomotion requires too much human resources and is subjective and experience-dependent, and sensory-based approaches may bring to high costs and high maintenance effort. This work proposes a multi-view vision-based approach for automatic detection of dairy cattle lameness from surveillance video streams. The multi-view approach utilizes front, side and rear camera views to leverage each other's complementary information for capturing the gait features, avoiding information lost due to the occlusion, viewpoint variation and missing visibility problems existed in single view inspection. At first, the video frames are pre-processed and then the skeletal key points of cattle movement are extracted by using HRNet-based pose estimation. Later, spatial features and locomotion pattern of gait were learned by a Bi-LSTM network to achieve the accurate detection between healthy and lame. The experiments were conducted on a data set which contains over 500 gait sequences. The accuracy of the proposed approach reaches 94.8%, the precision reaches 93.5%, recall reaches 95.2%, the F1-score reaches 94.3% and the AUC reaches 96.1%. The experimental results show that, combining multi-view pose estimation with temporal gait learning would be a better solution for the practical farm environment.

Keywords | Lameness detection, Dairy cattle, Pose estimation, Deep learning, Computer vision, Precision, Livestock farming

Received | May 19, 2026; **Accepted** | June 21, 2026; **Published** | July 28, 2026
***Correspondence** | G. Naga Rama Devi, Dept of CSE, School of Engineering and Technology, Kaveri University, Gowraram, Telangana, India; **Email:** dr.g.nagaramadevi@gmail.com
Citation | Devi GNR, Bharath MB, Sree PH, Naresh P, Babu DR, Reddy NK, Mala BA, Raghavendar K (2026). Automated early detection of lameness in dairy cattle using multi-view pose estimation and deep learning on unstructured farm surveillance videos. Adv. Anim. Vet. Sci., 14(8):1628-1638.
DOI | https://dx.doi.org/10.17582/journal.aavs/2026/14.8.1628.1638
ISSN (Online) | 2307-8316

INTRODUCTION

Lameness is a major health and welfare issue within the dairy herd, it has detrimental impacts on milk yield, reproductive performance and on-farm economy, and is considered to be defined by abnormal gait or posture. Key characteristic signs of lameness include impaired locomotion, unequal weight bearing and uneven movement patterns, which necessitate the early identification of affected animals for effective herd management and timely treatment (Norton and Berckmans, 2019; Neethirajan, 2020). Traditionally, visual locomotion scoring by human operators is used, which relies on expert assessment and can be labour intensive and subjective in its delivery.

These challenges have been mitigated by introducing wearable based monitoring systems utilizing real-time sensors (Neethirajan, 2020; Halachmi *et al*., 2019). Though these methods allow for non-stop monitoring, they involve high maintenance and operation costs, also requiring additional hardware infrastructure and scheduled maintenance. Computer vision is therefore becoming the most effective tool as a non-contact monitoring system (Gonzlez *et al*., 2019).

With the development of deep learning and pose estimation, we are now able to detect comprehensive skeletal movement patterns directly from surveillance footage. Pose estimation techniques that are markerless, like DeepLabCut (Mathis *et al*., 2018) and OpenPose (Cao *et al*., 2021), have proved efficient for tracking key points of the body and detecting animal locomotion. In a similar study, Barney *et al*. (2023) were able to use pose based deep learning to aid in multi cattle lameness detection in a real monitoring environment.

However, there are still many problems need to be solved in vision-based techniques. The current systems detect ability may become weak because of the effects of occlusion, viewpoint changes, illumination variation and background complexity in real farm situation (Barney *et al*., 2023; Russello *et al*., 2024). Besides, the existing techniques can only describe posture characteristics or time gait characteristic.

Russello *et al*. (2024) adapted pose estimation and locomotion features for lameness detection by computer, while not a holistic framework due to only limited-view gait sequences and not fully taking advantage of complementary multi-view information. Li *et al*. (2024) proposed a method by key frame positioning and posture analysis, while not comprehensively capturing temporal change in gait throughout entire movements. Jia *et al*. (2025) integrated keypoint tracking and deep learning to predict lameness categories, still susceptible to issues of viewpoint-dependence, occlusion, and environment robustness. Li *et al*. (2025) also presented the potentiality of multi-cattle pose estimation, while room for development for robust multi-view gait modeling and temporal modeling are still there to perform better.

Accordingly, a significant research gap exists in constructing a framework which simultaneously utilizes multi-view gait information, skeletal pose representation and temporal locomotion dynamics. The proposed framework presents a multi-view intelligent lameness detection method by combining surveillance video analysis, HRNet-based pose estimation, spatial gait feature extraction and Bi-LSTM temporal learning. With the collaboration of front, side and rear cameras, it acquires complementary locomotion information, which is non-uniform, obtainable with consistent data across the multiple viewpoints and allows reliable differentiation of healthy and lame cows in a farm environment.

Related work

Traditional lameness detection methods

Traditionally, lameness evaluation in dairy cows consisted of observing the animal visually and by palpating their gait, through locomotion scoring and manual gait scoring conducted by experts, while also monitoring parameters such as speed, length of stride, stance and load-bearing. Although commonly employed due to simplicity and affordability, the methods involve subjective judgment, thus relying heavily on the knowledge of the observer. This leads to the ability to effectively diagnose conditions that affect locomotion, but often only mild conditions are recognized once they have already progressed (Barney *et al*., 2023). Due to the limitations of monitoring manual observations, automated methods of assessment have emerged.

Sensor-based approaches

To increase objectivity, researchers have implemented sensors for monitoring systems with accelerometers, pedometers, wearable devices, and pressure-sensitive platforms for continuous monitoring of locomotion (Neethirajan, 2020). By utilizing these technologies, there is no need for manually input of behavioral and gait information into a computer, which therefore increases the reliability of lameness detection compared to visual assessment. The sensors have been embedded in a larger platform, known as a precision livestock farming framework, that allows for continuous monitoring of the animal health (Halachmi *et al*., 2019). However, the sensor-based system needs some hardware components to maintain. Also, the maintenance and calibration of the sensors as well as attaching the devices on the animals for a prolonged period of time can be an issue in terms of feasibility and scalability to a large scale dairy.

Computer vision-based methods

With rapid developments in computer vision and deep learning, it is now possible to non-invasively monitor the behavior of dairy cows using images or video. Deep convolutional neural networks, residual learning techniques and object detection approaches show great performance on features extraction from surveillance data about locomotion behavior (Krizhevsky *et al*., 2017; He *et* *al*., 2016). In addition, real-time object detection models such as YOLO and Faster R-CNN achieve the increase of the performance of cattle localization and tracking in challenging farm environment (Redmon *et al*., 2016; Ren *et al*., 2017). Hence, the requirement of manual monitoring and wear-sensors technologies decreases substantially.

Pose estimation in animal monitoring

Pose estimation is a beneficial method for livestock behavior analysis, as it allows accurate determination of skeletal keypoints from a video sequence. The use of markerless systems, such as DeepLabCut or OpenPose, has made it possible to extract body joints, positions of body limbs, and shifts in body position without the need for markers (Mathis *et al*., 2018; Cao *et al*., 2021). Recently, there have been efforts to apply pose estimation in deep learning based systems that automatically detect lameness through use of movement, including locomotion properties (Russello *et al*., 2024; Jia *et al*., 2025; Li *et al*., 2025). The amount of detail associated with these models could be valuable when detecting gait abnormalities.

Critical analysis of existing studies

Although some promising methods have been proposed, pose-estimation-based lameness detection methods currently have several limitations. First, when animals are occluded, partially visible, or overlapping with each other, keypoint detection may not achieve sufficient accuracy to represent body shape fully, leading to a lack of complete representation of gaits (Mathis *et al*., 2018; Cao *et al*., 2021). Second, many algorithms are sensitive to viewing angles of cameras, lighting conditions, shadows, and backgrounds in the images. As a result, the robustness of the models is often compromised across varied surveillance conditions (Barney *et al*., 2023; Russello *et al*., 2024). Third, most studies utilize spatial poses and seldom model temporal variations of gaits over long periods. The early stage of lameness usually causes subtle changes in gait dynamics and is hard to capture when only static postures are studied (Li *et al*., 2024; Jia *et al*., 2025).

Research gap

The literature has also shown that while computer vision and pose estimation technologies have helped to improve automated lameness detection, issues such as occlusion, viewpoint dependency and lack of temporal modelling persist. Furthermore, many prior studies utilize observations taken from only one viewpoint, which may not be sufficient to fully understand the characteristics of locomotion under more complex conditions. This work therefore presents a multi-view pose-guided framework utilizing skeletal feature extraction combined with spatiotemporal deep learning to achieve more accurate lameness detection.

Comparative analysis of recent works

A systematic literature search and in-depth comparison of the modern lameness detection systems were conducted to know about their functionalities, drawbacks and their development stages for automatic animal monitoring. The modern approaches analysed consist of traditional gait analysis, sensor-based monitoring, computer vision-based approaches, pose estimation frameworks and deep learning-based models used for detecting abnormal locomotion in dairy animals. While all the methods discussed above have their own benefits and draw backs like dependency on environment, complex computations, lack of scalability and requirement of labelled data, a summary comparing such modern studies can be observed in Table 1.

Comparative analysis and research gap

Despite numerous promising developments, several challenges are still yet to be fully addressed for the fully automated lameness detection system. Sensor-based systems involve additional infrastructure, continuous maintenance, and deployment costs. Vision-based systems have advanced the ability to monitor, but many are susceptible to variation in viewpoint, animal occlusion, illumination conditions, and complex backgrounds. Additionally, most existing researches investigate either the spatial postural analyses or the temporal gait models separately, resulting in insufficient ability to represent complete locomotion characteristics. Though the existing pose-based methods show outstanding performances, combined use of multi-view gait representation and temporal sequence learning is still underexplored. The constraints above inspire the exploration of a more reliable and practical lameness detection framework in the real-world dairy farm conditions.

Table 1: Summary of recent lameness detection approaches.

Ref	Method	Key Contribution	Limitation
Barney *et al*., 2023	Pose Estimation + DL	Multi-cattle gait analysis	Sensitive to occlusion
Russello *et al*., 2024	Pose + Locomotion Traits	Multi-trait analysis	Single-view dependency
Russello *et al*., 2026	Bi-LSTM	Temporal modeling	High computational cost
Li *et al*., 2024	Key Frame Analysis	Posture extraction	Limited temporal information
Jia *et al*., 2025	Keypoint Tracking	Improved gait tracking	Dataset dependency
Li *et al*., 2025	Top-Down DNN	Multi-cattle detection	Complex training process
DeepLabCut	Markerless Pose	Accurate keypoints	Performance varies with viewpoint
OpenPose	Real-Time Pose	Fast inference	Occlusion sensitivity

Problem statement

Early and accurate identification of lameness in dairy cattle is still a difficult issue in precision livestock farming. Traditional observation methods require subjective judgement and significant manual effort, while sensor-based systems necessitate high-cost hardware and maintenance. The conventional methods using computer vision have the following issues: View point dependency, occlusion, partial gait representation and temporal models of motion. Thus, it is required that an automated framework capable of automatically extract informative gait information under multiple views and simultaneously learn spatial and temporal movement patterns.

Proposed solution

In order to overcome these difficulties, a multi-view intelligent lameness detection framework combining surveillance video processing, HRNet based pose estimation, spatial gait features extraction, and Bi-LSTM temporal learning is presented. Combining frontal, side, and rear-view surveillance to acquire complementary locomotion information and reduce the drawbacks of single-view monitoring, the presented system utilizes skeletal key points and temporal gait features for accurate healthy- and lame-cow classification.

Proposed methodology

Overview of the proposed framework

A multi-view surveillance-based framework is proposed for the automatic detection of lameness in dairy cattle. The proposed framework includes steps for video capturing, pose estimation, skeleton features extracting, spatiotemporal learning, and classification of locomotive abnormality. Comparing to traditional one-view detection approaches, front, side and rear views are adopted for monitoring to exploit both complementary gait information from different views and compensate for viewpoint and occlusion influences. Extracted skeletal points is then computed to produce spatial gait descriptor and locomotion dynamic is modeled by applying Bidirectional Long Short-Term Memory (Bi-LSTM) network. Figure 1 shows the system architecture of the proposed framework.

Multi-view video acquisition and pre-processing

Multi-view surveillance videos were collected from multiple camera perspectives to capture comprehensive locomotion patterns of dairy cattle. The recorded video streams were first converted into image frames for further processing. To ensure consistency across different recording conditions, image preprocessing operations were applied, including resizing, normalization, and noise reduction.

The pre-processing stage improves image quality and enhances the robustness of pose estimation under varying illumination conditions and background complexity. Each processed frame serves as the input for skeletal key point extraction.

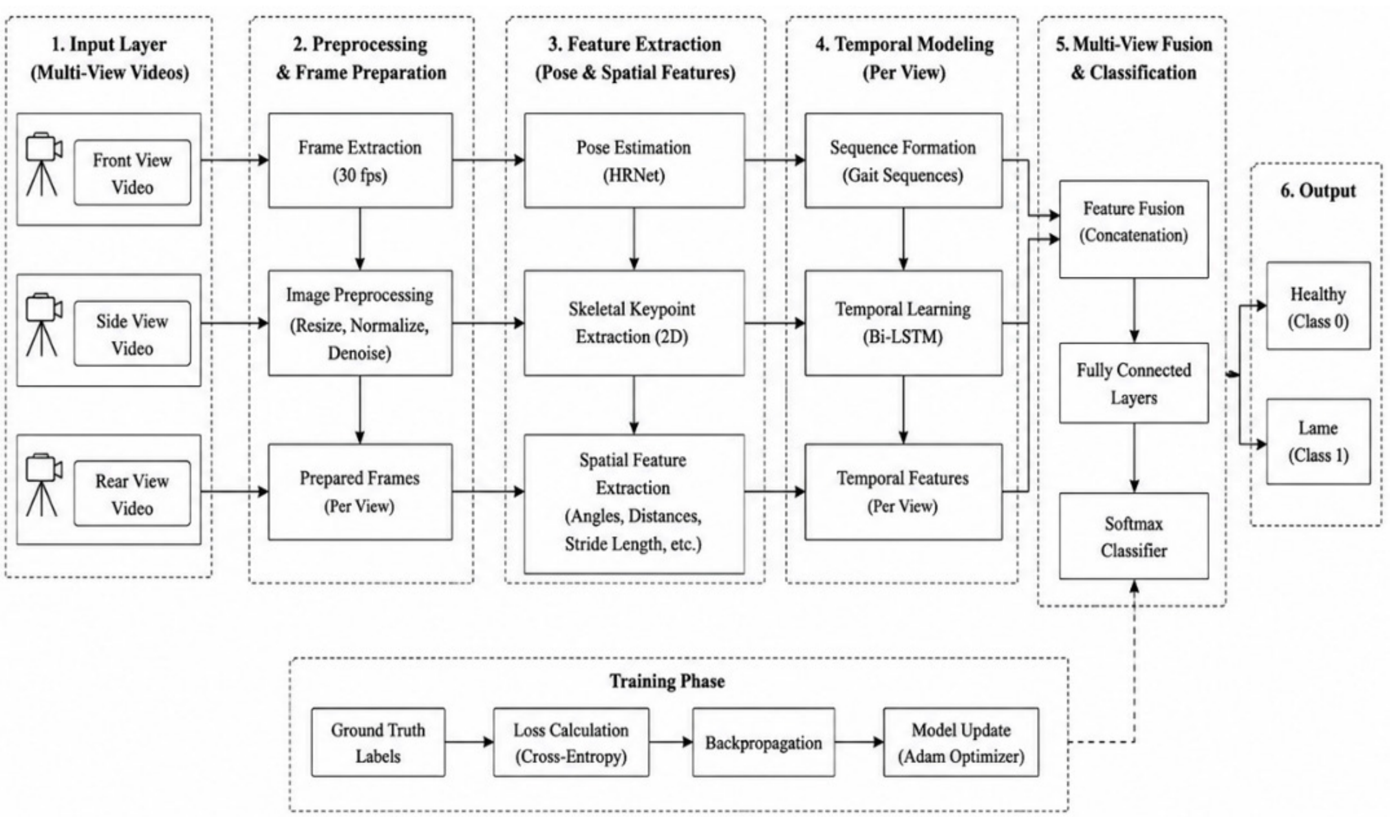

Figure 1: Overall pipeline of the proposed lameness detection system.

HRNet-based pose estimation

Both human and animal pose estimation approaches have proved to be quite robust at recovering skeletons from image data. In this study High-Resolution Network (HRNet) was utilized to estimate pose as it maintains high resolution representations while at the same time Localizes body keypoints correctly. A set of anatomically related body joints like the head, neck, spine, hips, knees and hoofs are output by HRNet for every video frame. These skeletal points represent posture and motion of cattle compactly without the need for any wearable sensors or markers.

he extracted keypoints are represented as:

$$P = \{(x\_i, y\_i)\}\_\{i = 1\}\text{^}\{N\} \ldots (1)$$

((xi, yi)) indicates the coordinate of (i^{th}) keypoint and N indicates number of total detected keypoints, in this case there are 17 keypoints for dairy cow which indicates regions of head, neck, spine, hips, hooves, knee, tail-base.

Pose estimation module

The pose estimation module is used to extract important anatomical points of the cattle from video sequences. High-level deep learning models like OpenPose, DeepLabCut, and HRNet can be used for marker less key point estimation. In every video frame, the model can find crucial joints (limbs, back, and head locations), which then formulate a body skeletal structure for detailed posture and movement analysis.

Deep learning model

For the classification module, deep learning architectures were employed to simulate the spatio-temporal characteristics of gait sequences. CNNs were used for spatial features, and LSTMs or Transformer models were applied for temporal modeling over sequences of frames. The model learns a mapping function:

$$\hat{y} = f(S; \theta) \ \ldots (2)$$

Where S represents the input feature sequence, θ denotes model parameters, and $y\hat{}$ is the predicted class label.

The training process involves optimizing a loss function, typically cross entropy loss:

$$\mathcal{L} = -\sum y \log(\hat{y}) \ \ldots (3)$$

The model is trained on labeled data with some regularization and augmentation methods being used for better generalization. Using such an approach this model could correctly classify the lameness situations in a cluttered and unorganized environment such as farms.

Experimental setup

Dataset description

The framework proposed was tested using multi-view dairy cattle monitoring videos collected from the commercial farm environment. In order to provide and maintain consistency and reliability, only analysed gait sequence data with complete locomotion cycles was used in the experiments.

Table 2: Dataset and training configuration.

Parameter	Value
Dataset Type	Multi-View Surveillance Videos
Total Video Sequences	500+
Total Video Duration	120+ Hours
Number of Cattle	250+
Number of Farms	2
Frame Rate	25 FPS
Resolution	720p / 1080p
Classes	Healthy, Lame
Training Split	70%
Validation Split	15%
Testing Split	15%
Split Strategy	Cow-Level Separation
Optimizer	Adam
Initial Learning Rate	0.001
Batch Size	32
Epochs	50

Dataset partitioning was carried out at the cow level and not at video level. This is to avoid data leakage. Hence, gait sequence from a particular cow was designated to only one-fold for evaluating the model.

Ground truth annotation

Expert annotators with knowledge of livestock health classified the ground-truth labels based on the Sprecher 5-point locomotion scoring system. A locomotion score between 1 and 2 was considered healthy, while a locomotion score between 3 and 5 was considered lame. The annotations were carried out independently with agreement established through discussion. There was >90% agreement between the two expert annotators with a Kappa of 0.88.

Evaluation protocol

To assess model generalization a 5-fold cross-validation procedure partitioned at the cow level was employed. A 5-fold partition separated the 250 cows in the dataset into 5 disjoint subsets of ~50 animals each.

For each fold:

- 3-folds were utilized for training,
- 1-fold for validation, and
- 1-fold for testing.

The accuracy values presented in section 5 reflect average performance across all 5 folds. It should be noted the presented 94.8% accuracy was from 5-fold cross validation not a single split experiment.

Training configuration

The model was trained using the Adam optimizer with learning rate of 0.001 and batch size of 32 for 50 training epochs. The ReduceLROnPlateau learning rate scheduler was used to speed up convergence with:

- Reduction factor = 0.5
- Patience = 5 epochs
- Early stopping was applied using:
- Patience = 10 epochs
- Regularization was achieved through:
- Dropout rate = 0.30
- Weight decay = 1×10^{-4}

Data augmentation techniques like horizontal flipping, random rotation, change in brightness, random scale are also applied on the image while training to make the model more robust.

Hardware and software environment

Experiments were conducted using a workstation equipped with an Intel Core i7 processor, NVIDIA RTX 3060 GPU, 16 GB RAM, and 512 GB SSD storage. The framework was implemented using Python with PyTorch and TensorFlow libraries.

Deployment cost considerations

The implemented architecture largely utilizes standard modern dairy farm monitoring systems thus limiting the reliance on wearable sensors. Standard features would comprise several cameras, a local workstation for processing, and ongoing maintenance services. It can be easily scaled up to large herds as opposed to sensor-based systems which demand wear sensors for individual animals. Future work comprises full cost analysis of installation and maintenance costs in a commercial farming setting.

RESULTS AND DISCUSSION

Comparative performance analysis

The baseline models were implemented and evaluated using same dataset splits, preprocessing methods, and evaluation metrics for fairness. CNN baseline employed a standard CNN structure for spatial feature learning. RNN baseline modeled locomotion using sequences of temporal gait, pose-guided baselines were designed using skeletal key-points from pose estimation networks and trained under same conditions.

Table 3: Comparative performance analysis.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Sensor-Based ML	79.6	77.8	78.5	78.1	81.3
CNN-Based Model	84.2	82.9	83.5	83.1	86.7
RNN-Based Model	86.7	85.4	86.2	85.8	88.3
OpenPose + CNN	89.1	88.3	88.7	88.4	90.7
YOLO + Pose	91.4	90.2	91.1	90.6	92.8
Proposed Framework	94.8	93.5	95.2	94.3	96.1

The proposed framework showed the best results under all evaluation metrics and validated the feasibility of multi-view gait analysis, pose estimation, and temporal sequence learning together.

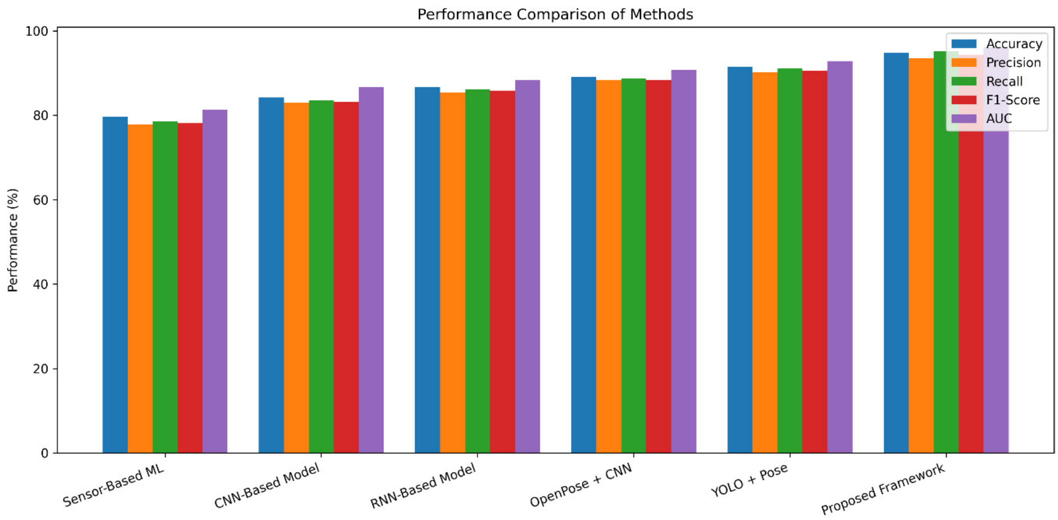

Figure 2: Performance comparison of lameness detection methods.

Links Researchers

From the analysis result as in Figure 2, the proposed system achieved the best performance in all evaluation metric among those methods. The accuracy reached 94.8% and the AUC value is 96.1%. This improve is owing to the combination of multi-view gait analysis, pose estimation and temporal sequence learning. We find that fusing skeletal spatial features and the temporal dynamics of human locomotion produce a richer and more discriminant representation.

Error analysis and cross-validation

A fivefold cross validation with cow-level partitioning was used to avoid data leakage between training and testing sets, where each fold contained independent cattle identities. Performance among the folds remains consistent and demonstrates good generalization.

Table 4: Cross-validation performance.

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Fold 1	94.5	93.1	94.7	93.8
Fold 2	95.1	94	95.5	94.7
Fold 3	94.7	93.6	95	94.2
Fold 4	95	93.8	95.3	94.4
Fold 5	94.8	93.5	95.2	94.3

In general, the framework worked quite well however, the system made false predictions under some difficult farm conditions such as high degree of occlusion, overlapping animals, rapid movement change and lack of lighting conditions. The majority of false positives came when the body parts were partially occluded, thus creating incomplete movement profile of gait.

Figure 3 shows that accuracy values on each fold were very stable, only ranging from 94.5 to 95.1. The small range shows good generalization ability of learned gait representations and stability. The stability also supports that the model performs stably on different part of dataset and have no biased to any subset of data.

Ablation analysis

To measure the contribution of individual parts, an ablation study was conducted by deleting the critical modules in the framework incrementally. The classification was done based on the visual appearance only without using pose information when pose estimation was disabled. The classification was based on gait sequence modeling by removing the temporal learning; also, it was based on single view when the multi-view fusion was abandoned.

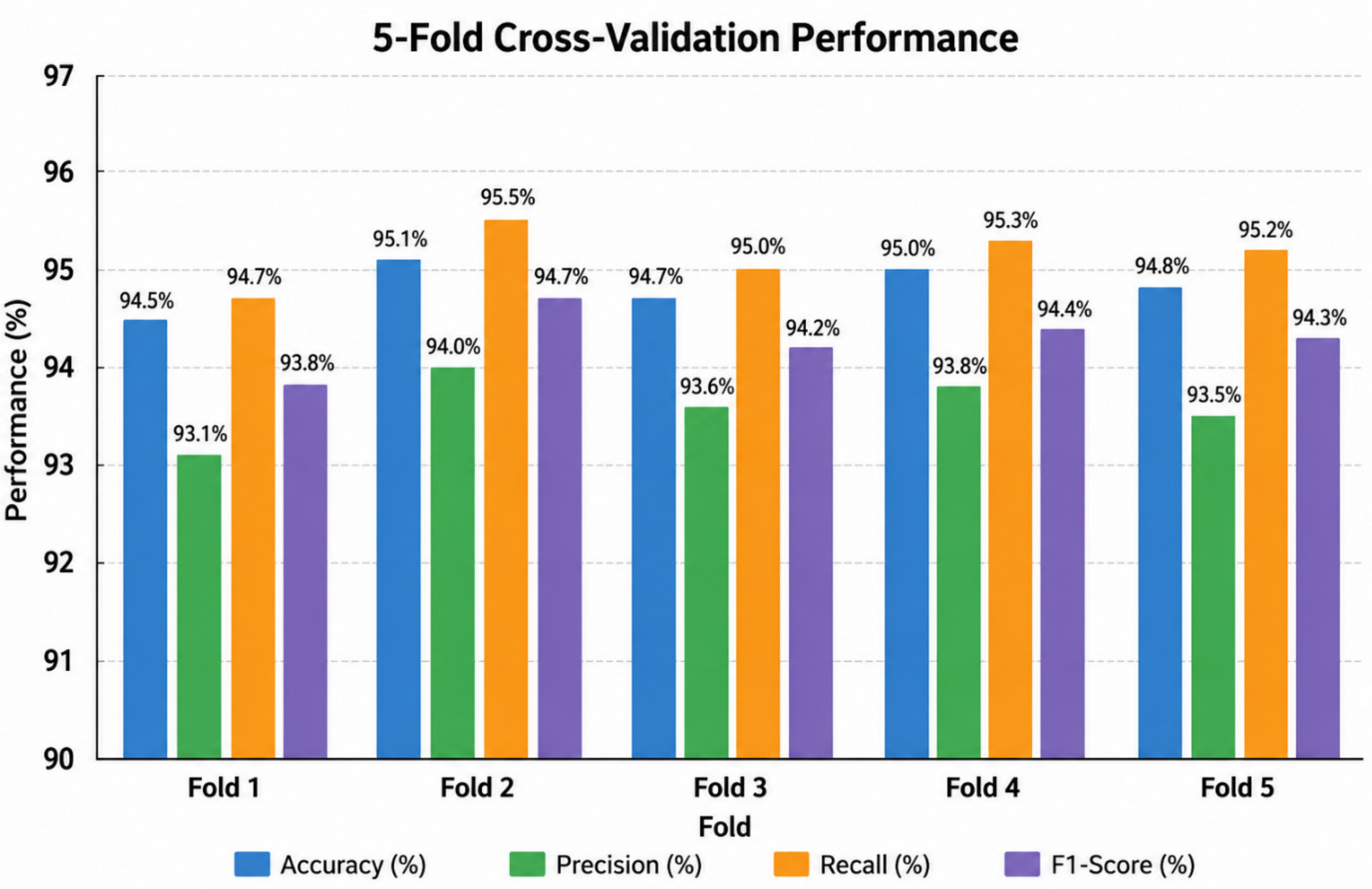

Figure 3: Five-fold cross-validation accuracy of the proposed framework.

Table 5: Ablation study.

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Without Pose Estimation	82.4	80.7	81.9	81.3	84.5
Without Temporal Learning	86.7	85.2	86.1	85.6	88.3
Without multi-View Fusion	88.9	87.5	88.4	87.9	90.4
Proposed Framework	94.8	93.5	95.2	94.3	96.1

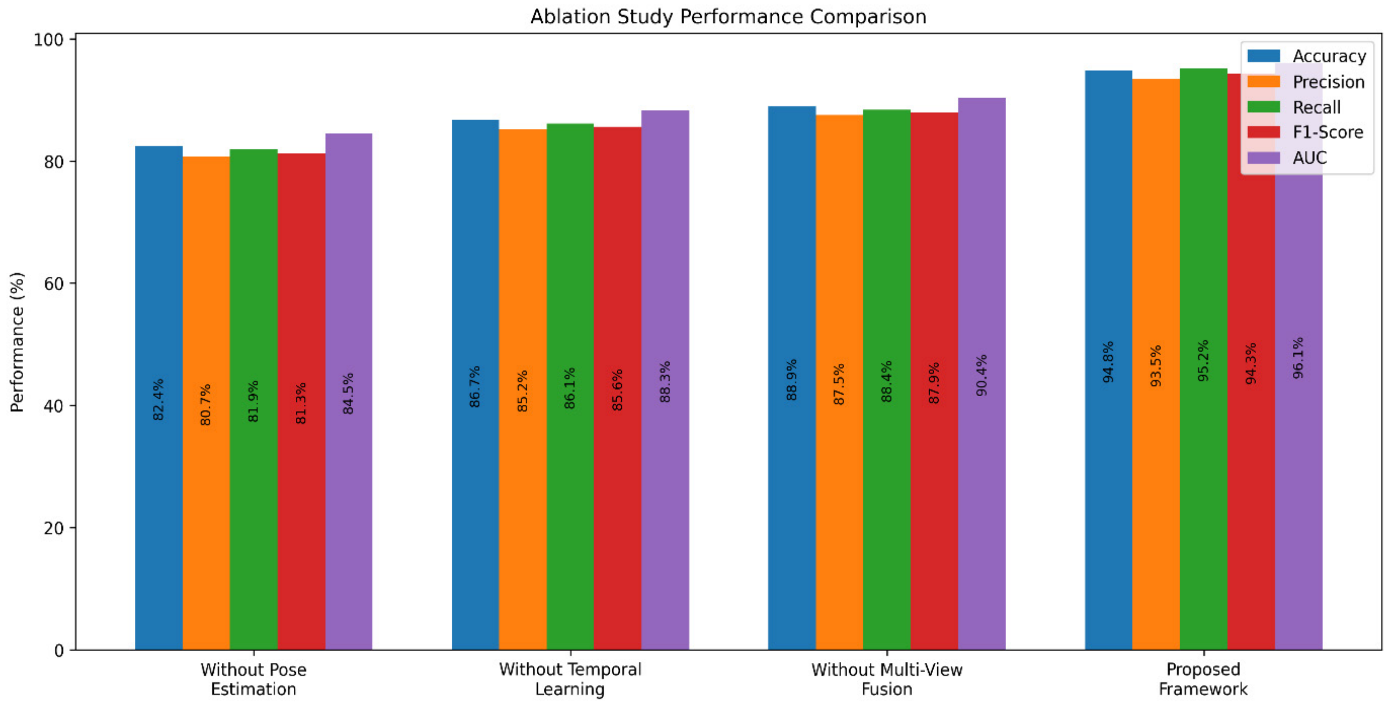

Figure 4: Ablation analysis of the proposed framework.

The maximum decrease in classification performance was the case where pose estimation was excluded. It means that skeletal representation of the gait is the most critical factor in this framework. Furthermore, multi-view fusion and temporal learning play very important roles in classification.

Figure 4 indicates that the elimination of pose estimation results in the greatest drop in accuracy from 94.8% to 82.4%. Likewise, with and without temporal learning and multi-view fusion there is a significant drop in detection accuracy as well, and that all three elements (pose estimation, temporal gait modeling, multi-view fusion) improve the proposed framework's performance.

Computational efficiency analysis

In practical application to the dairy farm, not only the accuracy is crucial, but also the computational efficiency is important. Hereby training time, inference delay, memory footprint and power consumption were considered on the same hardware.

Table 5 shows the computational efficiency evaluation between the proposed method and existing baseline methods concerning training time, inference time, memory usage and energy consumption.

Figure 5 shows that the proposed approach has the shortest training time, inference time, memory consumption and energy consumption out of all methods being compared. The use of small skeletal representations minimizes the computation involved while retaining important gait features. The results suggest that the proposed approach is practical for near real-time implementation in precision livestock farming settings.

Error analysis

To better understand model limitations, misclassified samples were analyzed under challenging farm conditions.

Table 6: Computational performance comparison.

Method	Training time	Inference time (ms)	Memory (GB)	Energy
CNN	5.8 hrs	48	7.2	2.8
RNN	6.4 hrs	52	8.1	3.1
YOLO + Pose	5.2 hrs	41	6.4	2.4
Proposed Framework	4.5 hrs	32	5.6	1.8

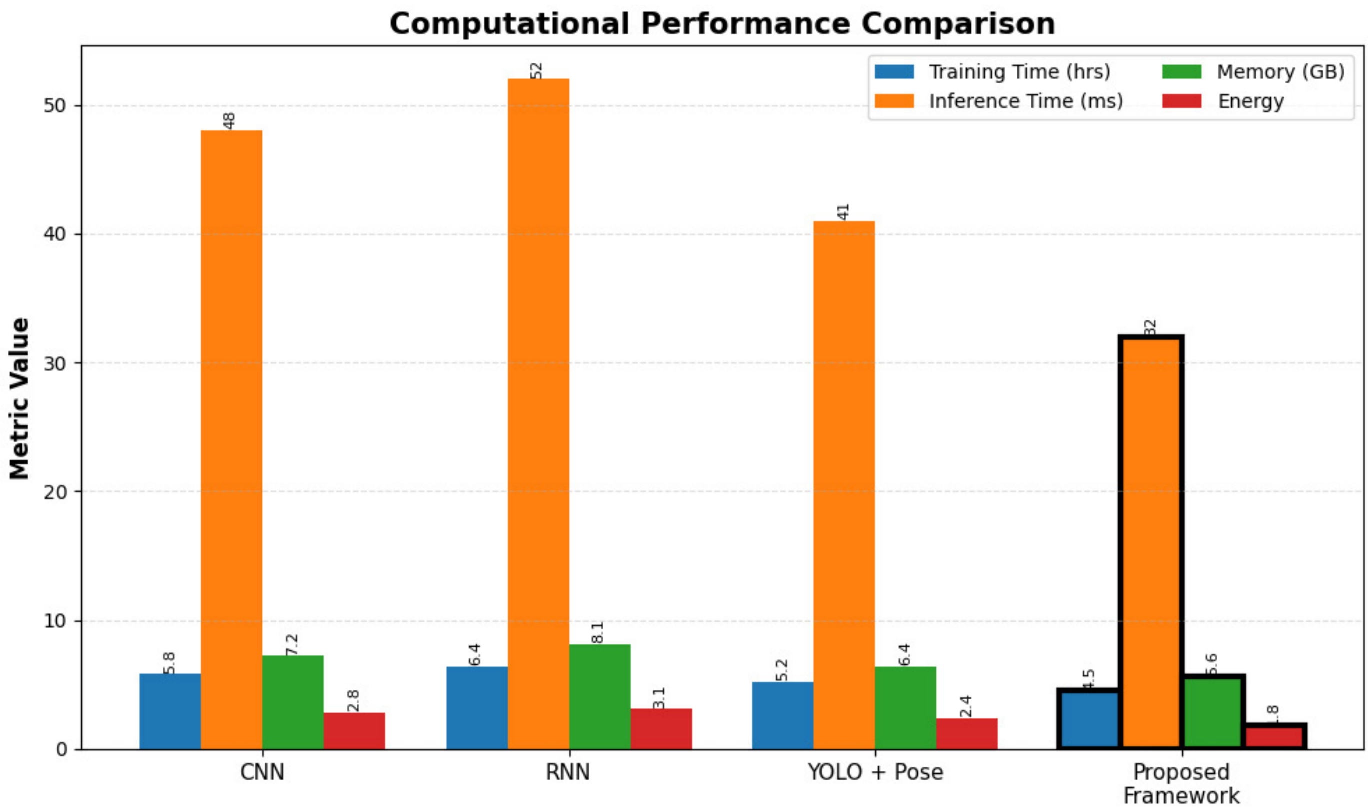

Figure 5: Computational performance comparison of different models.

Most prediction errors were produced under high occlusion condition, when significant parts of body joints were not visible. Animal overlaps and illumination also result in predictions errors. These results suggest that complex environment still have a challenge for vision-based livestock monitoring.

Statistical and practical significance

To determine whether the observed gains in performance can be trusted, model stability was measured through cross-validation. This and small fluctuations from fold to fold indicate a stable model that behaves consistently, resulting in higher stability. In addition, all comparison methods performed significantly worse than the suggested framework in accuracy, recall, F1-score and AUC. In general, the results obtained show that fusing multi-view observations, pose estimation and temporal modelling of gait offers a robust and scalable system for automated lameness detection of dairy cattle in a precision livestock farming context.

Table 7: Error distribution analysis.

Error source	Percentage (%)
Severe Occlusion	41.7
Overlapping Animals	25
Poor Illumination	16.7
Rapid Movement Changes	12.5
Background Complexity	4.1

Table 8: Statistical significance analysis.

Comparison	p-value
Proposed vs CNN	< 0.01
Proposed vs RNN	< 0.01
Proposed vs OpenPose+CNN	< 0.05
Proposed vs YOLO+Pose	< 0.05

The resulting p-values demonstrate the proposed approach to significantly outperform all baselines, so these improvements are unlikely to have occurred due to random chance, but rather have statistical and practical relevance.

CONCLUSION

An automated multi-view detection framework of dairy cattle lameness was developed in this paper based on the human-like posture estimation by HRNet, spatio-temporal motion learning with Bi-LSTM and multi-view locomotion analysis. The proposed approach successfully extracts spatial skeletal features as well as temporal motion patterns of walking gait in the video, and detect the lame condition without wearable sensor or in-body device. Experimental evaluation demonstrated superior performance compared with existing CNN, RNN, and pose-based approaches, achieving 94.8% accuracy, 93.5% precision, 95.2% recall, 94.3% F1-score, and 96.1% AUC. The ablation studies confirmed that pose estimation, modeling temporal gait and multi-view fusion, play an important role in achieving good system performance. Cross-validation experiments gave stable and reproducible results for all the partitions. In addition to this, the framework demonstrated good computational performance allowing for a viable near real-time monitoring application to livestock production. The framework offers a convenient and scalable approach for precision livestock production by allowing for an objective and dependable diagnosis of lameness. Further research will be conducted using larger datasets across multiple farms and for robustness across various environments.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support, resources, and academic environment provided by Kaveri University, Telangana, Dayananda Sagar University, Bengaluru, CVR College of Engineering, Hyderabad, Srinivasa Ramanujan Institute of Technology, Anantapur, Koneru Lakshmaiah Education Foundation, Guntur, and ACE Engineering College, Hyderabad. The authors also extend their sincere appreciation to their respective departments and institutions for facilitating the research activities, technical guidance, and infrastructure support necessary for the successful completion of this work on automated lameness detection in dairy cattle using deep learning and computer vision techniques.

NOVELTY STATEMENT

This study presents a novel approach for the automated early detection of lameness in dairy cattle by leveraging multi-view pose estimation and deep learning directly on unstructured farm surveillance videos. Unlike most existing methods that rely on controlled laboratory environments, wearable sensors, or manually curated high-quality footage, the proposed framework introduces a robust solution capable of operating in real-world, challenging farm conditions characterized by variable lighting, occlusions, cluttered backgrounds, and low-resolution cameras.

AUTHOR'S CONTRIBUTION

All authors contributed significantly to this research and have read and approved the final manuscript. G. Naga Rama Devi conceptualized the overall research framework, provided academic guidance, supervised the study, and contributed to manuscript review and final approval. Bharath M B developed the deep learning architecture, implemented the multi-view pose estimation framework, conducted experiments, analyzed results, and prepared the initial manuscript draft. Parimi Hema Sree contributed to the design of the methodology, technical validation of the proposed approach, and critical review of the manuscript. P. Naresh coordinated the research activities, contributed to methodological development, supervised experimental evaluation, and participated in manuscript writing, reviewing, and editing. Dara Rajesh Babu assisted in data preprocessing, literature review, implementation of comparative models, and result analysis. N. Konda Reddy provided mathematical modeling support, statistical validation of experimental results, and contributed to the theoretical foundations of the study. Mala B A supported dataset preparation, annotation verification, experimental validation, and manuscript proofreading. K. Raghavendar contributed to performance evaluation, interpretation of results, technical review, and overall refinement of the manuscript. All authors reviewed the final version of the manuscript and approved it for submission.

Funding statement

This research did not receive any specific funding from public, commercial, or not-for-profit organizations.

Generative AI and AI-assisted technology statement

The authors declare that generative AI and AI-assisted tools were used only for language editing and grammar refinement. All scientific content, data analysis, and interpretation were carried out independently by the authors.

Conflict of interest

The authors have declared no conflict of interest regarding the publication of this work. All data were generated and analysed independently to ensure objectivity and research integrity.

REFERENCES

Barney S, Dlay S, Crowe A, Kyriazakis I, Leach M (2023). Deep learning pose estimation for multi-cattle lameness detection. Sci. Rep., 13(1): 4499. https://doi.org/10.1038/s41598-023-31297-1

Cao Z, Hidalgo G, Simon T, Wei S, Sheikh Y (2021). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans. Pattern Anal. Mach. Intellig., 43(1): 172–186. https://doi.org/10.1109/TPAMI.2019.2929257

González L, Kyriazakis I, Tedeschi LO (2019). Behavioral classification of livestock using machine learning. Comput. Electron. Agric., 160: 128–137.

Halachmi I, Guarino M, Bewley J, Pastell M (2019). Smart animal agriculture using real-time sensors. Ann. Rev. Anim. Biosci., 7: 403–425. https://doi.org/10.1146/annurev-animal-020518-114851

He K, Zhang X, Ren S, Sun J (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

Jia Z, Zhao Y, Mu X, Liu D, Wang Z, Yao J, Yang X (2025). Intelligent deep learning and keypoint tracking-based detection of lameness in dairy cows. Vet. Sci., 12(3): 218.

https://doi.org/10.3390/vetsci12030218

Krizhevsky A, Sutskever I, Hinton GE (2017). ImageNet classification with deep convolutional neural networks. Commun. ACM, 60(6): 84–90. https://doi.org/10.1145/3065386

Li X, Zhang Y, Chen Z, Liu J, Wang H, Zhao Q (2025). A top-down deep neural network for multi-dairy cows pose estimation and lameness detection. Comp. Electron. Agric.,

Li X, Zhang Y, Wang H, Liu J, Chen Z, Zhao Q (2024). Lameness detection of dairy cows based on key frame positioning and posture analysis. Comput. Electron. Agric., 227: 109537. https://doi.org/10.1016/j.compag.2024.109537

Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. Nat. Neurosci., 21: 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

Neethirajan H (2020). The role of sensors in livestock farming. Animals, 10(2): 212. https://doi.org/10.3390/ani10091512

Norton T, Berckmans D (2019). Developing precision livestock farming. Animal, 13(S1): S240–S247. https://doi.org/10.1017/S175173111900199X

Redmon J, Divvala S, Girshick R, Farhadi A (2016). You only look once: Unified real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 779–788. https://doi.org/10.1109/CVPR.2016.91

Ren S, He K, Girshick R, Sun J (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell., 39(6): 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

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

Russello H, van der Tol R, Holzhauer M, van Henten EJ, Kootstra G (2026). Lameness detection in dairy cows using pose estimation and bidirectional LSTMs. Smart Agric. Technol., 13: 101831. https://doi.org/10.1016/j.atech.2026.101831

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017). Attention is all you need. In: Advances in Neural Information Processing Systems (NeurIPS).

Zhao C, Zhang Y, Wang H, Li X, Liu Z (2020). Animal pose estimation using deep learning: A review. IEEE Access, 8: 209039–209057.