Anomaly detection approach for cow behavior without relying on a predefined gold standard

Arnas Nakrošis[1], Jad Kaedbey[1], Dmytro Klepachevskyi[1], Marius Klumbys[1], Vusal Aliyev[1], Agnė Paulauskaitė-Taraseviciene[1,2]

[1]*Department of Applied Informatics, Faculty of Informatics, Kaunas University of Technology, 51368 Kaunas, Lithuania*
[2] *Artificial Intelligence Excellence Centre, Kaunas University of Technology, 51423 Kaunas, Lithuania*

Abstract

This study investigates an unsupervised anomaly detection approach for monitoring cow behavior using thermal video data. Detecting abnormal behavior in livestock is particularly challenging due to the absence of labeled anomalies, high intra-animal variability, and environmental noise in thermal recordings. Four unsupervised learning models – K-means clustering, Isolation Forest, Convolutional Autoencoder, and Generative Adversarial Network, were implemented, evaluated, and comparatively analyzed with respect to their effectiveness in detecting abnormal behavioral patterns. The results show that unsupervised methods combined with thermal imaging can provide an effective solution for automated livestock monitoring. Despite the inherent complexity of the task, the Convolutional Autoencoder achieved the highest performance, attaining an accuracy of 86.53%, thereby underscoring the potential of deep unsupervised models and the need for further investigation.

Keywords

Livestock Monitoring; Thermal Imaging; Deep Learning; Unsupervised Learning; Anomaly Detection

1. Introduction

The health and welfare of livestock are paramount in modern dairy farming, directly impacting both ethical standards and economic viability. Early and accurate detection of health issues, such as lameness, mastitis, or other infections, is critical for timely intervention, minimizing animal suffering, and preventing significant economic losses.

Continuous, individualized observation by human observers has thus become impractical as the scale and automation of modern dairy farms continue to increase. Manual inspection faces limitations in terms of time and subjectivity [1, 2]. Given this background, automated monitoring solutions have recently received increased interest within the framework of precision livestock farming, where sensor-based and data-driven approaches support more objective and timely decision-making [3, 4].

The techniques of deep learning have been transferred to posture estimation, activity recognition, and behavior classification on different livestock species [5, 6, 7]. Yet most of these existing approaches depend on supervised learning and large, carefully pre-annotated datasets; the creation of such gold-standard labels is labor-intensive, expensive, and often restricted to expert interpretation. This limits scalability and generalization at real farm environments [8, 9].

Of all the available sensing modalities, Thermal imaging or Infrared Thermography (IRT) has shown potential as a non-invasive method for health monitoring in livestock. Since it offers physiological data, which cannot be easily captured through visible images, several studies have shown the ability of IRT in monitoring conditions like inflammation, infection, or stress in cattle, specifically mastitis and hoof lesions. [10, 11, 12, 13], while several recent reviews have shown its potential integration for automated health monitoring through Machine Learning in PLF systems [4, 14].

IVUS 2026: Information Society and University Studies 2026, May 12, Kaunas, Lithuania

arnas.nakrosis@ktu.lt; jad.kaedbey@ktu.lt; dmykle@ktu.lt; marius.klumbys@ktu.lt; vusal.aliyev@ktu.lt; agne.paulauskaite-taraseviciene@ktu.lt

0000-0002-2168-9981 (A. 1); 0009-0004-3527-7461 (A. 3); 0000-0002-8787-3343 (A. 6)

Copyright © 2026 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-06

CEUR-WS.org/Vol-4317/paper7.pdf

The unsupervised and semi-supervised learning approaches have received extensive attention in computer vision, specifically for applications involving rare and not easily defined anomalies in a priori knowledge settings [15, 16, 17, 18]. Applying anomaly detection in animal monitoring has been proposed for identifying unusual behavioural or thermal patterns; however, its use for continuous thermal analysis in real-world livestock settings has received limited attention in agricultural research.

This paper presents a study into the application of unsupervised machine learning techniques for anomaly detection in dairy cows using thermal video data captured in a real-world farm environment. We propose and evaluate a comprehensive processing pipeline, encompassing data preprocessing, feature extraction, and anomaly identification. Multiple unsupervised models are implemented and comparatively assessed to determine their suitability for this challenging task. The primary contributions of this study are as follows:

1. An efficient video processing pipeline that automatically segments relevant footage of individual cows from continuous, long-duration recordings, significantly reducing the data volume required for analysis.
2. The implementation and evaluation of a K-Means clustering algorithm to identify visually anomalous frames based on holistic image features, capturing atypical postures or behaviours.
3. The development of a deep learning convolutional autoencoder specifically trained on thermal images of cow legs to detect subtle anomalies based on reconstruction error, targeting a common site of injury and infection.
4. An analysis using an Isolation Forest model to detect anomalous frames based on thermal texture and statistical features, offering an alternative to purely visual methods.
5. The application of a Generative Adversarial Network (GAN), trained exclusively on normal thermal data, to identify anomalies. Inputs that the model's discriminator flags as 'fake' or that result in high reconstruction error are considered anomalous, indicating a deviation from healthy patterns.

By comparing these diverse methodologies, this study aims to validate the feasibility of a fully automated, non-invasive system capable of reliably flagging individual cows that exhibit potential health anomalies.

2. Related Works

Continuous monitoring of animal behavior and physiological responses enables earlier detection of health deviations than periodic visual inspection [3, 19]. Advances in precision livestock farming have increasingly emphasized behavior-based monitoring supported by sensor technologies and artificial intelligence [3]. However, substantial inter- and intra-animal variability complicates the definition of universal behavioral baselines, making herd-level reference models prone to misclassification even within homogeneous farm environments [1].

Many existing animal monitoring systems rely on supervised learning approaches that require extensive expert annotation. The acquisition and maintenance of such labeled datasets are costly and difficult to scale in real farm settings [19]. Vision-based methods, including head, ear, and pose estimation, have shown promise as non-invasive indicators of stress and affective states in cattle [5], but they remain heavily dependent on carefully curated training data. Reviews consistently report practical deployment challenges such as occlusion, background complexity, class imbalance, and limited availability of high-quality annotations [6].

Fixed labels and global thresholds further limit model robustness, as behavioral patterns that are normal for one individual may be abnormal for another. As a result, supervised models often exhibit reduced generalization when applied across animals or farms [1]. Numerous studies have employed supervised machine learning and deep learning techniques—such as support vector machines, random forests, convolutional neural networks, and recurrent neural networks—for detecting specific conditions including lameness, mastitis, and metabolic disorders [19]. Nevertheless, these approaches remain fundamentally constrained by the availability and consistency of labeled training data.

Unsupervised anomaly detection offers an alternative paradigm that does not require explicit labeling of abnormal events. Surveys of the field classify these methods into clustering-based,

isolation-based, reconstruction-based, and generative categories [15]. Within livestock research, unsupervised approaches have primarily been applied to wearable sensor data, such as accelerometer and gait measurements. Several studies demonstrate that learning individualized behavioral baselines and identifying deviations enables early detection of health-related events, including lameness and metabolic disorders, without the need for explicit anomaly labels [1, 2]. Comparative evaluations further indicate that animal-specific models generally outperform herd-level approaches, particularly for applications such as estrus detection using activity sensors [8].

Thermal imaging represents a complementary, contactless modality capable of capturing physiological changes associated with inflammation, infection, or stress through surface temperature distributions [20]. Previous studies have shown that thermal responses exhibit pronounced time-of-day variation and are strongly influenced by environmental conditions, complicating the definition of stable reference patterns [11]. Despite these challenges, infrared thermography has been repeatedly identified as a promising tool for early detection of inflammatory conditions, including mastitis, highlighting its potential role in non-invasive health monitoring systems [13].

Nevertheless, the reliability of thermal-based monitoring remains limited by the lack of standardized acquisition protocols, consistent region-of-interest definitions, and robust analytical methodologies [12, 20]. Many existing approaches rely on manually selected regions or empirically defined temperature thresholds, making them sensitive to environmental variability and reducing generalizability across farms [10].

Beyond the agricultural domain, unsupervised and weakly supervised anomaly detection has been extensively studied in related areas such as video surveillance and industrial inspection. These fields face similar challenges, including complex spatiotemporal patterns, contextual variability, the rarity of anomalous events, and the absence of reliable ground-truth annotations in realistic video data [16, 17, 18]. Insights from these domains further motivate the development of robust, scalable anomaly detection frameworks applicable to continuous thermal video analysis in livestock monitoring.

3. Dataset and Preprocessing

A comprehensive dataset of thermal video was collected and processed to train and evaluate the unsupervised anomaly detection models.

3.1. Data Acquisition

The data was sourced from a single commercial dairy farm located in Lithuania. The recordings were captured over the months of September and October, 2025. A TOPDON's TC001 portable thermal imaging camera was used for all recordings. The camera was positioned to overlook the entrance and exit of an automated milking station. This setup ensured a consistent field of view, capturing the side profile and legs of each cow as it passed through the area. The initial collection comprises approximately 30 hours of footage. All videos were recorded in the MP4 format. The footage contains a mixture of relevant and irrelevant activity. Relevant activity includes cows walking into, standing in, and exiting the milking station (see Figure 1). Irrelevant periods include long stretches of inactivity, motion from non-bovine sources (e.g., farm personnel, cleaning equipment), and environmental artifacts.

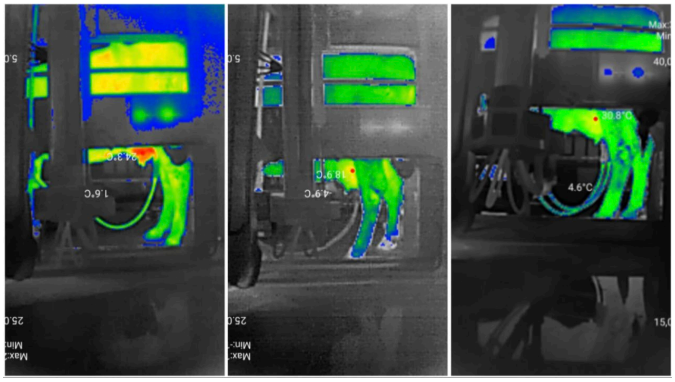


Figure 1. Frames of healthy cows from raw video dataset

Given the volume of the raw data, a multi-stage preprocessing pipeline was used to extract a smaller, more manageable dataset for analysis. The automated video pruning process, detailed in the Data Augmentation section, reduced the 30 hours of footage to approximately 12 hours of relevant motion segments.

To facilitate a quantitative evaluation of our models, a ground truth dataset was established in collaboration with the farm veterinarian. Over the data collection period, 16 cows were officially recorded with health issues, primarily mastitis, and leg infections (see Figure 2). The video segments corresponding to the dates when these cows were identified as ill were manually reviewed. Specific frames or sequences showing visible signs of distress or containing the labeled cows were marked as the "ground truth" for anomalous events. This labeled subset was used to calculate the precision, recall, and F1-score for the relevant algorithms.

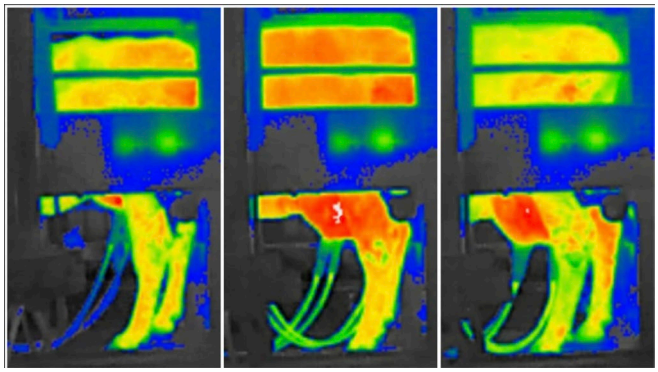


Figure 2. Focused frames of ill cows

Raw thermal video streams are characterized by large data volumes and prolonged intervals of inactivity or irrelevant motion. To construct a tractable and informative dataset suitable for training anomaly detection models, a multi-stage data preprocessing and augmentation pipeline was implemented. This pipeline first focuses on isolating relevant events from continuous recordings and subsequently applies data augmentation strategies to artificially increase the dataset size, thereby enhancing model robustness and generalization capability.

3.2. Motion-Based Video Pruning

The initial and most critical preprocessing step consists of pruning multi-hour thermal video recordings into short, relevant clips that contain actual cow activity. Manual annotation of such large-scale data would be prohibitively time-consuming and impractical. Consequently, an automated filtering pipeline was developed to identify and extract informative segments from continuous recordings.

The pipeline begins with the definition of a fixed Region of Interest (ROI) corresponding to the area in which cows are present during the milking process. Within this ROI, motion detection is performed using OpenCV's MOG2 background subtraction algorithm. However, motion-based detection alone is insufficient, as it may be activated by irrelevant events such as water splashes or moving mechanical components. To mitigate this issue, an additional color based filtering step is incorporated. Specifically, the motion mask produced by the background subtraction algorithm is combined with a thermal color mask generated by thresholding video frames in the HSV (Hue, Saturation, Value) color space to isolate the characteristic color signature associated with thermal camera measurements.

An event is registered only when enough pixels simultaneously satisfy both the motion and thermal color criteria over a sustained number of consecutive frames. Once this condition is met, the corresponding video segment is automatically extracted and stored.

Building upon the pruned video segments, the dataset is further refined through the extraction of specific sequences of interest. The system continuously monitors the retained clips for instances in which the thermal pixel count exceeds a predefined high threshold, indicating that a cow is directly

within the camera's field of view. Upon detection, a 10-second sequence of frames is captured. From each sequence, a grid-like composite image is subsequently generated for downstream analysis.

4. Unsupervised Anomaly Detection Methodology

To identify potential health issues in dairy cows from thermal imaging data without relying on pre-existing labels, we evaluated four distinct unsupervised anomaly detection algorithms. Each method uses a different approach to learn the patterns of "normal" data and flag deviations as anomalies.

4.1. K-Means Clustering

K-Means is a foundational clustering algorithm that partitions a dataset into a predefined number of clusters (k). In our approach, we applied K-Means for visual anomaly detection in the following way [21]:

1. Individual frames were extracted from the pre-processed video sequences.
2. Each frame was resized to a standard dimension (256x256 pixels) and flattened into a one-dimensional feature vector based on its raw pixel values.
3. The K-Means algorithm (with k=8 clusters) was trained on the collection of these feature vectors, grouping visually similar frames together.

The anomaly score for each frame is its Euclidean distance to the centroid of its assigned cluster.

4.2. Isolation Forest

The Isolation Forest is an ensemble-based algorithm designed specifically for anomaly detection. It operates by building a forest of random trees. To isolate any given data point, the algorithm recursively and randomly partitions the data by selecting a feature and then a random split value for that feature. Anomalies, being "few and different," are typically easier to isolate and will therefore have a much shorter average path length from the root of a tree to the terminating node. Implementation: This method employs a feature-based approach instead of raw pixel values [22].

For each frame, a set of statistical and texture-based features is calculated from the thermal data.

4. These features include the mean, standard deviation, min/max pixel values, entropy, and texture properties derived from a Grey-level Co-occurrence Matrix (GLCM), such as contrast and homogeneity.
5. The Isolation Forest model is then trained on this structured feature set.

The model directly calculates an anomaly score for each frame based on its average path length.

4.3. Convolutional Autoencoder (CAE)

An autoencoder is a class of neural network architectures designed to learn an efficient and compact representation (encoding) of input data by minimizing the reconstruction error between the original input and its reconstructed output. Convolutional Autoencoders (CAEs) extend this paradigm to image-based data by employing convolutional and pooling layers in the encoder, and corresponding upsampling or transposed convolution layers in the decoder, enabling the model to effectively learn spatial hierarchies and localized feature patterns inherent in visual data [23].

In this study, the CAE was trained exclusively on thermal images representing normal physiological conditions and subsequently evaluated on a dedicated test set of cropped thermal images of cow legs containing both Normal and Ill samples. By restricting training to normal data, the model was encouraged to learn a robust representation of healthy thermal patterns, while abnormal samples were expected to yield higher reconstruction errors. An anomaly detection threshold was defined dynamically as the mean reconstruction error plus two standard deviations computed from the normal test samples. Images exceeding this threshold were classified as anomalous.

4.4. Generative Adversarial Network (GAN)

A Generative Adversarial Network (GAN) uses two competing neural networks - a Generator and a Discriminator - to learn the underlying distribution of a dataset. This implementation includes an Encoder, a Generator (Decoder), and a Discriminator [24].

1. The GAN is trained exclusively on normal image frames. The Generator learns to produce realistic thermal images of healthy cows, and the Discriminator learns to identify them as real.
2. The Encoder learns to map input images to a latent space representation that the Generator can use for reconstruction.

5. Results

This section presents the performance of the four unsupervised anomaly detection algorithms on the thermal video dataset. The evaluation is based on a combination of quantitative metrics, where a ground truth is available, and qualitative analysis of the models' ability to identify meaningful deviations.

5.1. K-Means Clustering

The K-Means algorithm was evaluated on its ability to identify visually distinct frames from the captured sequences. The clustering performance itself, measured by internal validation metrics, yielded a Silhouette Score of 0.4023 and a Davies-Bouldin Index of 0.9979. The Silhouette Score, ranging from -1 (poor) to 1 (excellent), suggests that the clusters are only moderately well-separated. The Davies-Bouldin Index, where lower values are better, supports this finding, with a score near 1.0 suggesting that the clusters have a fair amount of overlap and are not exceptionally compact. This is an expected result in a real-world farm environment, where cow movements are fluid and do not fall into neatly separated visual categories.

The high recall of 0.50 is particularly noteworthy. It demonstrates that the model successfully identified 50% of the true anomalous events within the dataset.

5.2. Isolation Forest

The Isolation Forest model was selected as the primary method for frame-level anomaly classification due to its superior results for outlier detection in high-dimensional feature spaces. Isolation Forest directly models anomalous behavior by isolating samples through random partitioning. In this study, the Isolation Forest method was applied to a feature set derived from each frame, including statistical and texture-based thermal properties. The model was configured with an expected contamination rate of 2%, meaning it was tasked with identifying the top 2% of most unusual frames.

A plot of the mean thermal value per frame (see Figure 3) shows the algorithm successfully flagging frames that correspond to sharp spikes or significant deviations from the temporal trend. This indicates that the model is effective at identifying sudden changes in the thermal environment, which could correspond to a cow entering the frame or a rapid temperature change on the animal's body. The output is a list of frame indices flagged as anomalous, providing timestamps for further veterinary review.

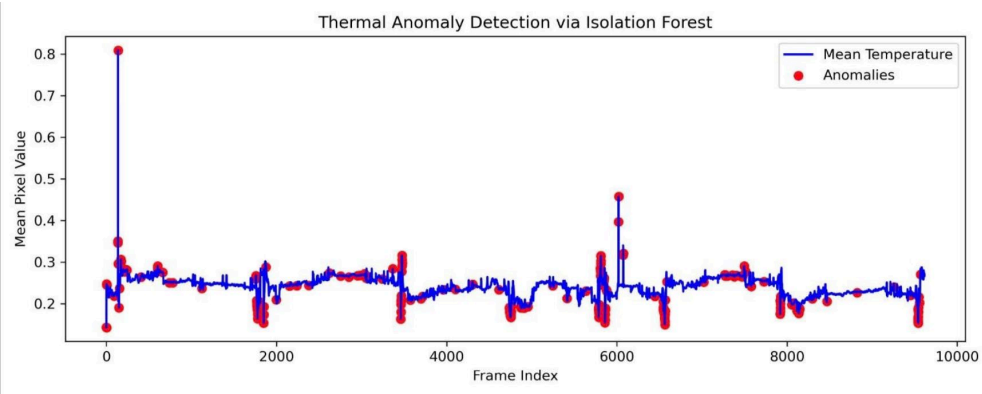


Figure 3. Thermal Anomaly Detection via Isolation Forest

The final quantitative evaluation achieved decent results on 4 metrics, highlighting the potential of the Isolation Forest method for this task. Quantitative evaluation confirms the suitability of Isolation Forest as a classification-based anomaly detector, achieving 73.54% of precision. The balanced precision-recall results indicate that the model achieves a reasonable trade-off between sensitivity and false-positive rate (FPR).

5.3. Convolutional Autoencoder

The evaluation results are illustrated through a histogram of reconstruction errors, which provides a clear visual confirmation of the separability between normal and ill cow legs (see Figure 4). As shown in the distribution, samples corresponding to ill legs consistently exhibit higher reconstruction error values compared to normal legs, resulting in a distinct rightward shift of the anomalous class. In contrast, reconstruction errors for normal samples are tightly clustered around lower values.

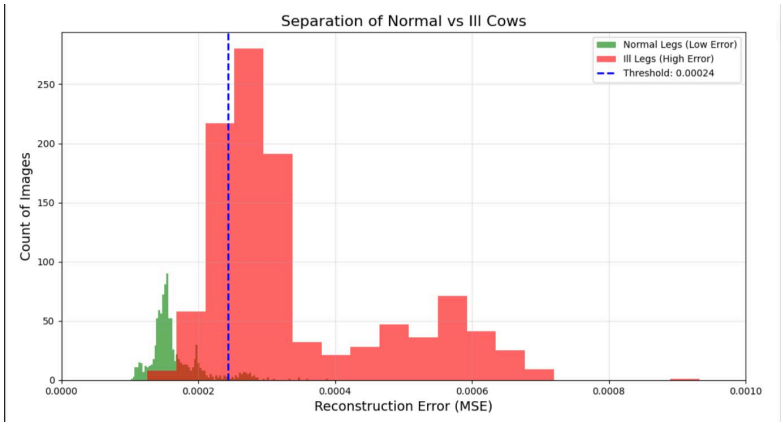


Figure 4. Auto-Encoder model prediction diagram

Mean Squared Error (MSE) is employed as the reconstruction error metric to quantify the pixel-wise deviation between the original input images and their reconstructed counterparts. The observed error separation in Figure 4 confirms that threshold-based anomaly detection using the CAE is both feasible and effective, supporting its suitability for identifying subtle anomalies.

5.4. Generative Adversarial Network

The GAN-based model was trained to learn the distribution of normal thermal frames and detect anomalies via reconstruction error. Frames that the model could not reconstruct faithfully were flagged as anomalous, with the threshold set at the 99th percentile of reconstruction scores (see Figure 5). The output of this model is a set of images visually highlighting the regions of high reconstruction error.

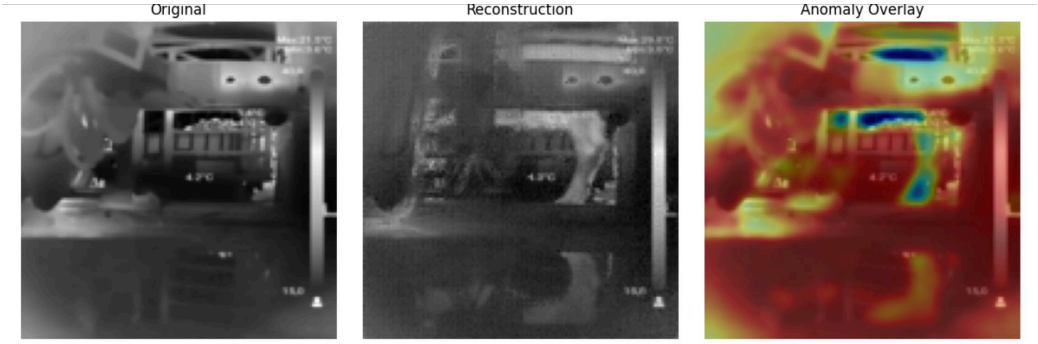


Figure 5. Example of reconstruction-based anomaly visualization produced by the GAN-based model

For each cow, reconstruction error values were computed over the corresponding time interval, allowing the model to capture both brief abnormal events and longer-lasting thermal irregularities.

5.5. Experimental results comparison

Table 1 presents a comparative evaluation of the four unsupervised models investigated in this study, using accuracy, precision, recall, and F1-score as performance metrics. Among the evaluated approaches, the CAE clearly outperforms the other models, achieving the highest accuracy of 86.53%, alongside a precision of 92.71%, recall of 79.85%, and an F1-score of 85.80%.

The Isolation Forest model demonstrates moderate and relatively balanced performance, with an accuracy of 72.78% and comparable precision (73.54%) and recall (71.82%), resulting in an F1-score of 72.66%. This suggests that Isolation Forest can capture certain anomalous characteristics in the data, but lacks the representational capacity required to model the complex spatial and thermal structures.

Table 1
Comparative results across 4 models

Model	Accuracy	Precision	Recall	F1-Score
K-Means	N/A	6.45%	50.00%	11.43%
CAE	**86.53%**	**92.71%**	79.85%	**85.80%**
Isolation Forest	72.78%	73.54%	71.82%	72.66%
GAN	57.1%	25.0%	**100.0%**	40.0%

In contrast, the K-means clustering approach exhibits poor discriminative capability, reflected in a very low precision of 6.45% and an F1-score of 11.43%, despite achieving a recall of 50.00%. This indicates a high false-positive rate and highlights the limitations of simple clustering methods when applied to high-dimensional thermal image data. The accuracy metric is not reported for K-means due to its unsupervised nature and the absence of a direct class assignment framework.

The Generative Adversarial Network (GAN) achieves a recall of 100.0%, indicating that all anomalous samples were successfully identified. However, this comes at the cost of extremely low precision (25.0%), resulting in an overall accuracy of 57.1% and an F1-score of 40.0%.

Overall, the results summarized in Table 1 demonstrate that deep learning–based unsupervised models, particularly the Convolutional Autoencoder, provide superior performance for thermal anomaly detection in livestock compared to classical clustering and tree-based approaches.

6. Conclusions

This study confirms the feasibility of using unsupervised machine learning methods for automated, non-invasive detection of health-related anomalies in dairy cows using thermal video data. An end-to-end pipeline was developed and evaluated using four unsupervised approaches, demonstrating that problem-specific model selection is critical for effective anomaly detection. Among the evaluated models, the Convolutional Autoencoder (CAE) achieved the strongest overall

performance, with an F1-score of 0.8580 and high precision (0.9271). The Isolation Forest model exhibited balanced performance (precision: 0.7354, recall: 0.7182) and proved effective at identifying global statistical deviations in thermal features, making it suitable for detecting broader thermal anomalies. In contrast, K-means clustering and GAN-based approaches were less reliable for direct anomaly detection. K-means showed high sensitivity but poor specificity, resulting in a high false-positive rate, while the GAN approach suffered from training instability and indirect anomaly scoring despite offering interpretable anomaly localization.

7. Discussion and Limitations

The complementary strengths of the Isolation Forest and CAE suggest that a hybrid, two-stage framework could provide the most practical solution for real-world deployment. The computationally efficient Isolation Forest can act as a continuous screening tool to detect frames with anomalous global thermal characteristics, while the CAE can subsequently perform detailed analysis of specific regions. This hierarchical approach supports the development of a decision-support system rather than a fully autonomous diagnostic tool, generating concise and context rich alerts that allow veterinarians and farm managers to efficiently prioritize inspections. The study is limited by the relatively small amount of expertly annotated ground truth data, which restricts comprehensive model validation and limits conclusions about generalizability. Future research should therefore focus on expanding the dataset with larger and more diverse thermal recordings collected across multiple farms and under varying environmental conditions. Incorporating ambient temperature, humidity, time of day, and other contextual variables could improve the robustness of thermal anomaly detection, since these factors may influence surface temperature patterns. In addition, richer veterinarian labeled datasets would allow more reliable evaluation of model performance and support stronger comparisons between anomaly detection approaches. Further work will also explore longitudinal tracking of individual cows to establish personalized thermal and behavioral baselines, as deviations from an animal specific norm may be more informative than deviations from a herd level average. Beyond thermal imaging alone, multimodal approaches combining thermal data with other sources such as RGB video, accelerometers, or activity sensors could improve detection accuracy and reduce false positives. Finally, future studies should investigate real-time deployment of the proposed framework in operational farm environments, evaluating not only model performance but also computational efficiency, alert quality, and practical usefulness for livestock health monitoring. Overall, the results highlight the potential of unsupervised learning applied to thermal imaging for scalable and proactive livestock health monitoring, with hybrid feature-based and deep learning approaches offering a promising pathway toward practical deployment.

8. Declaration on Generative AI

The authors have not employed any Generative AI tools.

9. References

[1] N. Wagner *et al.*, "Machine learning to detect behavioural anomalies in dairy cows under subacute ruminal acidosis," *Comput. Electron. Agric.*, vol. 170, p. 105233, Mar. 2020, https://doi.org/10.1016/j.compag.2020.105233.

[2] J. Haladjian, J. Haug, S. Nüske, and B. Bruegge, "A Wearable Sensor System for Lameness Detection in Dairy Cattle," *Multimodal Technol. Interact.*, vol. 2, no. 2, p. 27, Jun. 2018, https://doi.org/10.3390/mti2020027.

[3] B. Jiang, W. Tang, L. Cui, and X. Deng, "Precision Livestock Farming Research: A Global Scientometric Review," *Animals*, vol. 13, no. 13, p. 2096, Jan. 2023, https://doi.org/10.3390/ani13132096.

[4] F. G. Silva, C. Conceição, A. M. F. Pereira, J. L. Cerqueira, and S. R. Silva, "Literature Review on Technological Applications to Monitor and Evaluate Calves' Health and Welfare," *Animals*, vol. 13, no. 7, p. 1148, Jan. 2023, https://doi.org/10.3390/ani13071148.

[5] S. Kim, "Automated Cattle Head and Ear Pose Estimation Using Deep Learning for Animal Welfare Research," *Vet. Sci.*, vol. 12, no. 7, p. 664, Jul. 2025, https://doi.org/10.3390/vetsci12070664.
[6] A. Rohan, M. S. Rafaq, Md. J. Hasan, F. Asghar, A. K. Bashir, and T. Dottorini, "Application of deep learning for livestock behaviour recognition: A systematic literature review," *Comput. Electron. Agric.*, vol. 224, p. 109115, Sep. 2024, https://doi.org/10.1016/j.compag.2024.109115.
[7] G. Calderone, M. Molins, P. Catania, J. Boney, and E. Casella, "Aerial monitoring of turkey behavior using unmanned aerial vehicles and computer vision," *Poult. Sci.*, vol. 104, no. 12, p. 106103, Dec. 2025, https://doi.org/10.1016/j.psj.2025.106103.
[8] Á. Michelena, A. Díaz-Longueira, P. Novais, D. Simić, Ó. Fontenla-Romero, and J. L. Calvo-Rolle, "Comparative analysis of unsupervised anomaly detection techniques for heat detection in dairy cattle," *Neurocomputing*, vol. 618, p. 129088, Feb. 2025, https://doi.org/10.1016/j.neucom.2024.129088.
[9] Y. Deng, H. Qu, A. Leng, X. Tang, and S. Zhai, "Methods and challenges in computer vision-based livestock anomaly detection, a systematic review," *Biosyst. Eng.*, vol. 253, p. 104135, May 2025, https://doi.org/10.1016/j.biosystemseng.2025.104135.
[10] M. Alsaaod and W. Büscher, "Detection of hoof lesions using digital infrared thermography in dairy cows," *J. Dairy Sci.*, vol. 95, no. 2, pp. 735–742, Feb. 2012, https://doi.org/10.3168/jds.2011-4762.
[11] R. J. Berry, A. D. Kennedy, S. L. Scott, B. L. Kyle, and A. L. Schaefer, "Daily variation in the udder surface temperature of dairy cows measured by infrared thermography: Potential for mastitis detection," *Can. J. Anim. Sci.*, vol. 83, no. 4, pp. 687–693, Dec. 2003, https://doi.org/10.4141/A03-012.
[12] I. Novotna, L. Langova, and Z. Havlicek, "Risk Factors and Detection of Lameness Using Infrared Thermography in Dairy Cows – A Review," *Ann. Anim. Sci.*, vol. 19, no. 3, pp. 563–578, Jul. 2019, https://doi.org/10.2478/aoas-2019-0008.
[13] Y. Wang, X. Kang, Z. He, Y. Feng, and G. Liu, "Accurate detection of dairy cow mastitis with deep learning technology: a new and comprehensive detection method based on infrared thermal images," *animal*, vol. 16, no. 10, p. 100646, Oct. 2022, https://doi.org/10.1016/j.animal.2022.100646.
[14] Y. Wang, Q. Li, M. Chu, X. Kang, and G. Liu, "Application of infrared thermography and machine learning techniques in cattle health assessments: A review," *Biosyst. Eng.*, vol. 230, pp. 361–387, Jun. 2023, https://doi.org/10.1016/j.biosystemseng.2023.05.002.
[15] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep Learning for Anomaly Detection: A Review," *ACM Comput Surv*, vol. 54, no. 2, p. 38:1-38:38, Mar. 2021, https://doi.org/10.1145/3439950.
[16] G. A. Noghre, A. D. Pazho, and H. Tabkhi, "A Survey on Video Anomaly Detection via Deep Learning: Human, Vehicle, and Environment," Aug. 19, 2025, *arXiv*: arXiv:2508.14203. https://doi.org/10.48550/arXiv.2508.14203.
[17] Y. Liu *et al.*, "Generalized Video Anomaly Event Detection: Systematic Taxonomy and Comparison of Deep Models," *ACM Comput Surv*, vol. 56, no. 7, p. 189:1-189:38, Apr. 2024, https://doi.org/10.1145/3645101.
[18] M. Abdalla, S. Javed, M. Al Radi, A. Ulhaq, and N. Werghi, "Video anomaly detection in 10 years: a survey and outlook," *Neural Comput. Appl.*, vol. 37, no. 32, pp. 26321–26364, Nov. 2025, https://doi.org/10.1007/s00521-025-11659-8.
[19] E. Fazzari, D. Romano, F. Falchi, and C. Stefanini, "Animal behavior analysis methods using deep learning: A survey," *Expert Syst. Appl.*, vol. 289, p. 128330, Sep. 2025, https://doi.org/10.1016/j.eswa.2025.128330.
[20] R. McManus *et al.*, "Thermography for disease detection in livestock: A scoping review," *Front. Vet. Sci.*, vol. 9, Aug. 2022, https://doi.org/10.3389/fvets.2022.965622.
[21] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010, https://doi.org/10.1016/j.patrec.2009.09.011.
[22] S. Hariri, M. C. Kind, and R. J. Brunner, "Extended Isolation Forest," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 4, pp. 1479–1489, Apr. 2021, https://doi.org/10.1109/TKDE.2019.2947676.
[23] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," Apr. 03, 2021, *arXiv*: arXiv:2003.05991. https://doi.org/10.48550/arXiv.2003.05991.

[24] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018, https://doi.org/10.1109/MSP.2017.2765202.