Kafkas Universitesi Veteriner Fakultesi Dergisi
Journal Home-Page: http://vetdergikafkas.org
E-ISSN: 1309-2251

Kafkas Univ Vet Fak Derg
32 (1): 85-90, 2026
DOI: 10.9775/kvfd.2025.35393

Research Article

Cattle Identification with CLIP-Based Biometric Features

Yücel DEMİREL [1] Afşin KOCAKAYA [2 (*)] Orhan YAMAN [3]

[1] Fırat University, Graduate School of Natural and Applied Sciences, Digital Forensics Engineering Department, TR-23119 Elazığ - TÜRKİYE
[2] Ankara University, Faculty of Veterinary Medicine, Animal Husbandry and Breeding Department, TR-06070 Ankara - TÜRKİYE
[3] Fırat University, Faculty of Technology, Digital Forensics Engineering Department, TR-23119 Elazığ- TÜRKİYE

(*) **Corresponding author:**
Afşin Kocakaya
Phone: +90 312 3170315
Cellular phone: +90 505 8818005
Fax: +90 312 3164472
E-mail: akocakaya@ankara.edu.tr

How to cite this article?

Demirel Y, Kocakaya A, Yaman O: Cattle Identification with CLIP-Based Biometric Feature. *Kafkas Univ Vet Fak Derg*, 32 (1): 85-90, 2026.
DOI: 10.9775/kvfd.2025.35393

Article ID: KVFD-2025-35393
Received: 30.09.2025
Accepted: 03.02.2026
Published Online: 04.02.2026

Abstract

The individual identification of cattle is crucial for herd management and food safety, as well as for complying with the demands of export markets, particularly those within the European Union. In addition, traditional identification methods such as ear tagging, tattooing, or hot-cold branding have significant limitations in terms of reliability, loss rates, and animal welfare. The study proposes and evaluates a non-invasive biometric identification method using the analysis of distinctive patterns in cow coat colours. The approach we use is the CLIP deep learning model (ViT-L-14) to derive a feature vector, or "biometric signature," from a picture of each cow's coat colour pattern. This method was evaluated on a large dataset (Cows2021) containing 23.350 images representing 301 unique individuals. Utilizing a cross-validation technique (80% training/20% testing), the system exhibits better performance with an accuracy of 94.28%. Additionally, performance metrics revealed precision at 94.67%, recall at 94.28%, and an F1-score at 94.27%; this result confirms the robustness of the model in the face of class imbalances. Consequently, it is believed that the extensive adoption of this method will reduce labour in herd management and improve automatic, reliable, and animal welfare-oriented identification and traceability within the livestock sector, thereby facilitating substantial advancements in precision livestock farming practices.

Keywords: Biometric identification, Cattle welfare, Cattle management, CLIP model, Deep learning, Precision livestock farming

Introduction

Identification and traceability in livestock is crucial for achieving accuracy in sustainable animal husbandry and ensuring food security. This entails monitoring the whole lifetime of animals, from birth to processing, enabling farmers to successfully oversee health, welfare, and environmental implications. Implementing comprehensive traceability systems enables stakeholders to improve transparency and foster customer confidence in the food supply chain [1,2]. Besides simple herd management, it becomes a strategic concern for biosecurity, epidemiological surveillance, product certification, and customer trust, embodying the notion of "from farm to fork". Traditional animal identification methods such as branding, tattooing, ear notching, collar id, ear-tagging, and even electrical identification methods such as Radio Frequency Identification (RFID) are not reliable enough for cattle identification due to ear-tag loss, label fading, physical damage to tags due to harsh climates, damage to ears, animal welfare concerns, theft, fraud, and duplication [3-12]. Ear tag identification is also performed manually and is consequently susceptible to human error [13]. These strategies do not produce successful results. However, there are operational, economic, and management challenges with large-scale monitoring of livestock animals [3,5]. Therefore, non-invasive approaches for the identifying of livestock on farms are required [5].

Conventional identifying systems, have been used for an extended period [9,10]. However, traditional methods are being questioned due to their low reliability, risk of loss, wear and tear, forgery, and the damage they cause to animals' ears and bodies. From an ethical and animal welfare perspective, these practices are a demonstrated cause of pain, distress, and infection risk for animals [3-6,8,11,12,14]. Addressing these concerns requires a thorough examination of current practices and a commitment to implementing more humane alternatives. By prioritizing the well-being of animals, we can create a more compassionate approach that aligns with societal values. These challenges have paved the way for the emergence of a new paradigm based on biometric identification, which aims to utilize unique and unchangeable physiological characteristics as natural identifiers. The literature explores many of these biometric data types. For examples, the analysis of nose prints (the tip of the nose) and retina images has demonstrated a high capacity for discrimination [15,16]. However, its implementation encounters a significant operational constraint: the need for stringent immobilization of the animal and the method's sensitivity to the cleanliness of the nasal surface. Other techniques, such as face recognition or eye analysis, although promising, have obstacles due to differences in brightness, capture angles, age-related morphological changes, and, most importantly, phenotypic uniformity within certain cow breeds, which might confuse the algorithms [17-19]. In this context, the coat color pattern of cattle stands out as particularly robust biometric data: it is unique to each individual, stable over time, covers a large area, and can be captured remotely without direct intervention on the animal [20-23]. However, the potential of this marker has long been underutilized due to the inherent complexity of its patterns and the poor performance of traditional pattern recognition algorithms in the face of such variation. The emergence of deep learning architectures, and particularly pre-trained baseline models on billions of images, offers an unprecedented opportunity to overcome these technological hurdles [24].

This research introduces and substantiates a comprehensive cattle identification approach using the CLIP (Contrastive Language-Image Pre-Training, ViT-L-14) model [25]. The primary contribution of this work is the application of an advanced vision model, not for traditional classification tasks, but for its capacity to encode intricate covering patterns into dense and semantically rich vector representations. These vectors function as a secure digital signature, or "biometric data value," for each animal. This study aims to demonstrate the feasibility and high performance of the proposed approach by developing a system capable of generating signatures, storing them in a database, and utilizing them for identification purposes.

In this respect, the main contributions of this study are threefold: first, it proposes a method of contactless biometric identification that eliminates the stress and risks of infection associated with invasive devices (tags, implants). Secondly, it validates the use of overlay patterns as a permanent and unique biometric feature, offering a reliable alternative to artificial methods that are susceptible to loss or interference. Third, it demonstrates the effectiveness of transfer learning through the CLIP model; its use ensures accurate identification without requiring large annotated datasets and computationally intensive resources typical of traditional deep learning models.

Material and Methods

Ethical Statement

This study does not present any ethical concerns.

Dataset

This study used 23.350 images from the dataset published on *https://datasetninja.com/cows2021#download.* The dataset was specifically designed to improve and evaluate the performance of our identification system. It consists of two separate directories located in the same folder as the execution code: a reference directory called images and a benchmark directory called tests. The reference database was constructed from a directory of images, each containing four images representing unique cattle (*Fig. 1*); these images were used to generate biometric signatures stored in the animal_biometric.db database, which forms our repository of known individuals. At the same time, the test set in the test directory consists of 23.350 different images used to query the system and quantitatively evaluate its performance by comparing it with four reference individuals. All images in the dataset are pre-processed, sized at a standard resolution of 224 x 224 pixels, in accordance with the requirements of the CLIP model. To provide a rigorous and statistically significant assessment, the "Identification" dataset of 301 individuals and 23.350 images was used. The data was divided according to a strict protocol: 80% of the images were set as training and the remaining 20% as test sets.

For each image, a 14-dimensional embedding vector was extracted using the CLIP ViT-L-768 model. Then the cosine similarity between the test vector and the set of training vectors was calculated. In contrast to the use of fixed empirical thresholds, the definition is defined as "1. Level Matching" approach, where the identity corresponding to the highest similarity score is preserved. The soundness of the decision is confirmed by the analysis of the ROC curve and the calculation of the AUC score."

Fig 1. Images in the dataset

Proposed Approach

This section summarizes the design and functionality of our proposed identification technique. The proposed methodology uses the CLIP (Adversarial Language-Image Pre-Training) deep learning model, specifically its ViT-L-14 variant, to extract features from photographs of cow coat colours [28].

System Architecture

The proposed identification system has an architecture with two primary stages *(Fig. 2)*:

1. Mining and Documenting Biometric Signatures: In this initial phase, a photograph is taken of each cow so that it can be identified. The CLIP model takes the photo of its coat and makes a rich, high-dimensional feature vector from it. This vector, which is the animal's "biometric signature," is stored in a relational database that gives each species a unique ID and the name of the photo file that goes with it. People may then utilize this database to find out who they are.

2. Identification of a New Cow: A photograph of the cow's coat pattern is taken so that it may be identified. To get a feature vector from the query image, the CLIP model is employed. After that, a similarity measure is used to compare the vector to all the feature vectors in the database. The algorithm finds the vector that is most like the query vector and calls it the identity if its score is high enough.

Extracting Visual Features with the CLIP Model

The proposed methodology basically utilizes the CLIP (Contrastive Language-Image Pre-Training) framework developed by OpenAI [26]. CLIP learned from a large number of image-text pairs, which helped it learn how to connect visual representations with natural language. This contrastive learning process equips CLIP with a significant ability to understand and express the semantic substance of images, beyond basic object recognition. The ViT-L-14 variation of the CLIP was selected because it worked better, had more complicated architecture generating an integration space of 768 dimensions, and was more accurate and resilient in visual representations [27].

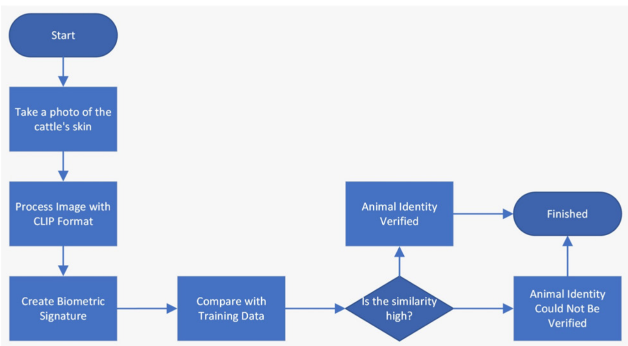

Fig 2. Flowchart of the proposed method

This is how to get visual characteristics from a picture of a cow's coat colour pattern:

1. Image Loading & Preprocessing: The texture photo is taken from the right folder and saved in RGB format. The image is resized to a common resolution of 224x224 pixels so that the CLIP model can get the same input every time. This resize standardizes the dimensions of the model's input tensors.

2. Using the CLIP Model to Code: The preprocessed picture is then used with the CLIP ViT-L-14 model. The model uses its image encoder to get a vector representation of the picture. This vector shows the coat's typical patterns, textures, and color distributions. It is a high-level abstraction of the image's visual content.

3. Normalization of the Feature Vector: The L2 norm is used to normalize the feature vector that the CLIP model created. This normalization ensures that all vectors have a unit magnitude, which is crucial for subsequent comparison based on cosine similarity. Normalization allows us to focus on the orientation of the vectors rather than their magnitude.

The resultant feature vector is a compact digital representation of the coating pattern, prepared for storage in the database or comparison with other vectors for identification purposes.

Similarity Measurement: Cosine Similarity

Cosine similarity quantifies the resemblance between the feature vector of the query picture and each feature vector (Vbase) inside the database. Cosine similarity quantifies the cosine of the angle between two vectors in a multidimensional space. It is delineated by the accompanying formula:

$$Cosine\ Similarity(Vquery, Vdatabase) = \frac{Vquery \times Vdatabase}{\|Vquery\| \times \|Vdatabase\|}$$

a- Vquery×Vdatabase, this expression represents the scalar product of two vectors.

b- ‖Vquery‖×‖Vdatabase‖, this represents the euclidean norms of the involved vectors.

Cosine similarity produces a value between -1 and 1, where 1 indicates perfect similarity (the two vectors are directed in the same direction), 0 indicates no linear correlation, and -1 indicates perfect contrast.

Identification and Similarity Thresholds

The identification process relies on calculating the cosine similarity between the query vector and the set of vectors in the database. To ensure maximum resilience in the face of environmental changes, this study opts for a "matching" approach rather than an arbitrary cut-off threshold.

According to this method, the identification decision is determined by the highest similarity score: the queried image is assigned to the identity of the reference vector (Similarity_{max}), which maximizes cosine similarity. The discriminating capacity of the model is assessed by analysis of the ROC curve with a sub curve area (AUC) of 0.723. This result confirms that the system consistently gives higher scores to positive pairs (in the same individual) than to negative pairs; thus, without the need for manual calibration of a certain threshold, it verifies the reliability of the decision.

Results

To evaluate the effectiveness of the proposed approach on a representative scale, rigorous experiments were conducted on the full "Identification" subset of the dataset; this subset included 301 cattle, and a total of 23.350 images were found. In contrast to limited pre-tests, this assessment was conducted using the 80% (Training)/20% (Test) section protocol.

Feature vectors were extracted using the CLIP ViT-L-14 model. Identification was performed by the matching method, where the predicted ID matched the vector in the gallery with the highest cosine similarity to the test image *(Table 1)*. In the test set, the system achieved an overall accuracy rate of 94.28%. This performance is justified by an F1 score of 94.27%, indicating an optimal balance between precision and recall, as detailed in *Table 2*.

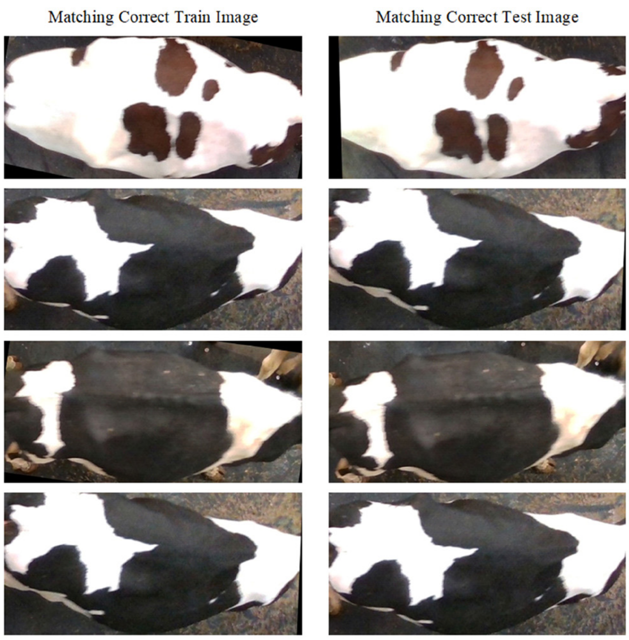

Fig 3. Correctly matched sample images in train and test datasets

Table 1. *Recognition percentages of the test group (20%)*

Image	Matches	Mismatches	Total
Number	4402	268	4670
Percentage (%)	94.28	5.72	100

Table 2. *Performance Metrics of CLIP ViT-L-14*

Criteria	Percentage
Accuracy	94.28%
Precision	94.67%
Recall	94.28%
F1-Score	94.27%

A ROC (Receiver Operating Characteristic) curve was generated to analyze the stability of the system in the verification task *(Fig. 4)*. The area under the curve (AUC) is 0.723; this value demonstrates the model's ability to accurately distinguish identities regardless of threshold variations.

In addition, a detailed analysis by class was carried out to assess the consistency of recognition across the 301 individuals. The distribution of accuracy *(Fig. 5)* reveals a strong asymmetry towards maximum performance: 146 individuals (48.5%) were identified with a perfect accuracy of 100%. Conversely, only 23 individuals have a recognition rate of less than 80%. This confirms that the CLIP model is able to extract visual representations that are unique enough for the vast majority of the herd, despite environmental challenges.

Furthermore, in a group of 23 individuals, the recognition rate having an accuracy rate less than 80% indicates that severe clogging, excessive lighting changes, and motion blur are the main causes of misidentification *(Fig. 6)*.

Consequently, the experimental results strongly confirm the potential of the proposed approach while identifying

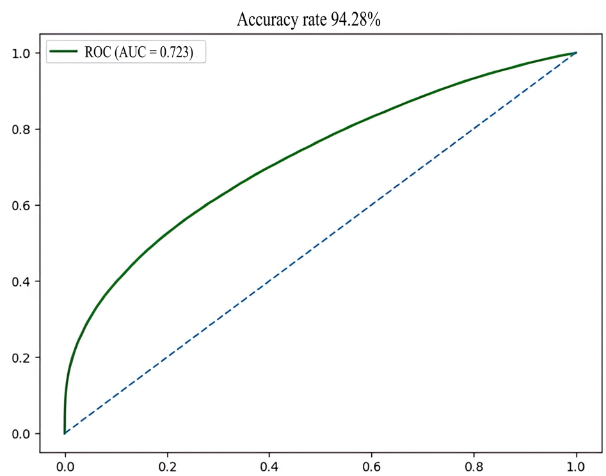

Fig 4. Roc curve

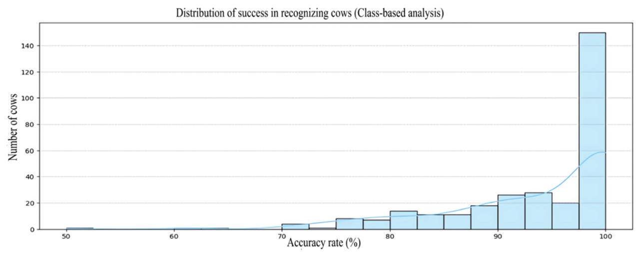

Fig 5. Distribution of recognition success of cows

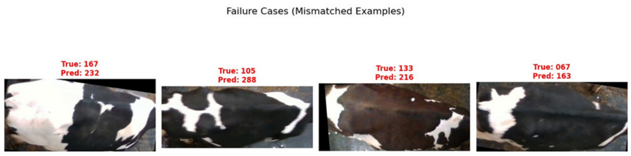

Fig 6. Incorrectly matched images

clear areas for improvement, particularly regarding robustness to capture conditions and optimization of the decision threshold.

Discussion

The experimental results, obtained in a large dataset containing 301 cows and 23.350 images, demonstrate the robustness of the proposed approach. In contrast to limited preliminary studies, this large-scale analysis achieved a 94.28% Accuracy rate and a 94.27% F1-Score. These performances convincingly confirm the effectiveness of the CLIP model for biometric identification of cattle that does not compromise their body integrity.

The main contribution of this study lies in the successful implementation of zero-shot learning without the need for costly retraining. While the seminal work of Martinez and Kak [29] addressed the limitations of classical linear methods, recently, deep learning-based approaches and specialized architectures for livestock identification have been proposed by Sharma et al.[22] and Wang et al.[23] Despite their high accuracy, these methods necessitate significant computational resources and intricate training processes. Similarly, Jing et al.[25] explored vision and language models (Animal-CLIP) for action recognition. Our study complements this emerging literature by showing that the CLIP Standard Model can be used efficiently with minimal computational cost for individual identification; thus, it offers a more accessible and scalable solution for the daily management of herds.

Analysis of errors related to 5.72% of unidentified cases reveals that failures are not random but linked to specific circumstances. Visual inspection of cases with an accuracy rate of less than 80% shows that severe clogging (sludge, equipment) and excessive lighting changes are the main causes, confirming the challenges noted by Andrew et al.[13] in their pioneering work on RGB-D imaging. The current ROC-AUC analysis (AUC = 0.72) provides a solid statistical basis for the reliability of the system.

It is important to highlight that the recommended method against misidentification in herd management provides 94.28%. While no biometric system is infallible, this high success rate positions the proposed method as a reliable decision support system. When used in conjunction with human oversight during critical operations, it offers a better alternative to traditional methods and minimizes risks.

In conclusion, the images have environmental challenges that lead to identification errors: loss of emphasis and detail at high light exposure, blurring of motion affecting the sharpness of the pattern, excessive shooting angle limiting visibility, and shadows that alter the appearance of biometric features. These examples validate the primary relationship between system performance and image acquisition quality. It is believed that these challenges encountered during field imaging can be overcome with more advanced camera and imaging methods.

In terms of practical application, the proposed approach considerably simplifies herd management. Unlike traditional models that require expensive retraining at each birth, our system allows for instant database updates simply by adding a reference photo of the new animal, making the technology accessible via standard surveillance cameras.

Despite the findings demonstrate the robustness of the CLIP methodology with a Rank-1 accuracy of 94.28% over 301 subjects, this research presents several limits that should be acknowledged for practical implementation. The examination was conducted post hoc on static pictures. This method models visual fluctuation but fails to replicate the complexities of real-time deployment on a continuous video stream, where optimal frame selection is essential. Secondly, while the model demonstrates resilience to partial occlusions and moderate motion blur, its efficacy under extreme environmental conditions—such as complete obscuration of the coat pattern by excessive mud or near-total darkness—has not been assessed and may necessitate supplementary infrared sensors. Ultimately, an ethical and operational analysis reveals a residual error rate of around 5%. Consequently, this system need to be regarded as a management tool (for instance, for observing behavior or feeding) rather than an unequivocal authority for irrevocable choices (such as slaughter or the implementation of medical treatments). Human validation is advised for these essential procedures to provide comprehensive security.

Declarations

Availability of Data and Materials: The data supporting this study's findings are available from the corresponding author upon reasonable request. Also the pictures of the 301 individual cattle was published on *https://datasetninja.com/cows2021#download*

Acknowledgements: We would like to thank Neill Campbell, Tilo Burghardt, Jing Gao, William Andrew, and Andrew Dowsey for sharing their datasets and allowing us to test and develop new artificial intelligence algorithms for identification in cattle.

Funding Support: Nope

Competing Interests: The authors declared that there is no conflict of interest.

Declaration of Generative Artificial Intelligence (AI): The authors declare that the article, tables and figures were not written/created by AI and AI-assisted Technologies. Quilbot was used for English.

Author Contributions: Idea, concept and design: YD, AK, OY; Data collection and analysis: YD; Drafting of the manuscript: YD, AK, OY; Critical review: YD, AK, OY

References

1. **Zhang M, Wang X, Feng H, Huang Q, Xiao X, Zhang X:** Wearable internet of things enabled precision livestock farming in smart farms: A review of technical solutions for precise perception, biocompatibility, and sustainability monitoring. *J Clean Prod*, 312:127712, 2021. DOI: 10.1016/j.jclepro.2021.127712
2. **Dayoub M, Shnaigat S, Tarawneh RA, Al-Yacoub AN, Al-Barakeh F, Al-Najjar, K:** Enhancing animal production through smart agriculture: Possibilities, hurdles, resolutions, and advantages. *Ruminants*, 4 (1): 22-46, 2024. DOI: 10.3390/ruminants4010003
3. **Awad AI, Zawbaa HM, Mahmoud HA, Nabi EHHA, Fayed RH, Hassanien AE:** A robust cattle identification scheme using muzzle print images. **In,** *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems.* 08-11 September, Krakow, Poland, 2013.
4. **Awad AI:** From classical methods to animal biometrics: A review on cattle identification and tracking. *Comput Electron Agric*, 123, 423-435. DOI: 10.1016/j.compag.2016.03.014
5. **Bhole A, Falzon O, Biehl M, Azzopardi G:** A computer vision pipeline that uses thermal and RGB images for the recognition of Holstein cattle. **In**, *Computer Analysis of Images and Patterns: 18th International Conference,* CAIP. 02-06 September, Salerno, Italy, 2019.
6. **Bello RW, Olubummo DA, Seiyaboh Z, Enuma OC, Talib AZ, Mohamed ASA:** Cattle identification: The history of nose prints approach in brief. *IOP Conf Ser: Earth Environ Sci,* 594:012026, 2020. DOI: 10.1088/1755-1315/594/1/012026
7. **Cihan P, Saygılı A, Ozmen NE, Akyuzlu M:** Identification and recognition of animals from biometric markers using computer vision approaches: A review. *Kafkas Univ Vet Fak Derg,* 29 (6): 581-593, 2023. DOI: 10.9775/kvfd.2023.30265
8. **Gaber T, Tharwat A, Hassanien AE, Snasel V:** Biometric cattle identification approach based on Weber's Local Descriptor and AdaBoost classifier. *Comput Electron Agric,* 122, 55-66, 2016. DOI: 10.1016/j.compag.2015.12.022
9. **Gobbo Oliveira Erünlü N, Rieder S, Kuntzer T, Bérard J, Wellnitz O:** Effects of sensor ear tags with twin pin fixing system on health and well-being of cattle. *Schweizer Archiv fur Tierheilkunde*, 165 (7): 512-523, 2023. DOI: 10.17236/sat00399
10. **Özbeyaz C:** Sığır Yetiştiriciliği. Medisan Yayın Serisi: 95, Ankara, 2025.
11. **Qiao Y, Su D, Kong H, Sukkarieh H, Lomax, S, Clark C:** BiLSTM-based individual cattle identification for automated precision livestock farming. *IEEE 16th International Conference on Automation Science and Engineering (CASE)*, Hong Kong, China, 2020.
12. **Qiao Y, Clark C, Lomax S, Kong H, Su D, Sukkarieh S:** Automated individual cattle identification using video data: A unified deep learning architecture approach. *Front Anim Sci,* 2:759147, 2021. DOI: 10.3389/fanim.2021.759147
13. **Andrew W, Hannuna S, Campbell N, Burghardt T:** Automatic individual holstein friesian cattle identification via selective local coat pattern matching in RGB-D imagery. *IEEE International Conference on Image Processing (ICIP)*, 25-28 September, Phoenix, AZ, USA, 2016.
14. **Canan S:** İşletmelerde hayvan refahının ekonomik yönlerinin değerlendirilmesi. *Iğdır Üniv Fen Bil Enst Der,* 13 (4): 3021-3029, 2023. DOI: 10.21597/jist.1361329
15. **Kaur A, Kumar M Jindal MK:** Cattle identification with muzzle pattern using computer vision technology: A critical review and prospective. *Soft Comput,* 26, 4771-4795, 2022. DOI: 10.1007/s00500-022-06935-x
16. **Cihan P, Saygili A, Akyüzlü M, Özmen NE, Ermutlu CŞ, Aydin U, Yilmaz A, Aksoy Ö:** U-Net-Based approaches for biometric identification and recognition in cattle. *Kafkas Univ Vet Fak Derg*, 31 (3): 425-436, 2025. DOI: 10.9775/kvfd.2025.34130
17. **Mahato S, Neethirajan S:** Integrating Artificial Intelligence in dairy farm management-biometric facial recognition for cows. *Inf Process Agric,* 12 (3): 312-325, 2024. DOI: 10.1016/j.inpa.2024.10.001
18. **Myint BB, Onizuka T, Tin P, Aikawa M, Kobayashi I, Zin TT:** Development of a real-time cattle lameness detection system using a single side-view camera. *Sci Rep*, 14:13734, 2024. DOI: 10.1038/s41598-024-64664-7
19. **Ruchay A, Kolpakov V, Guo H, Pezzuolo A:** On-barn cattle facial recognition using deep transfer learning and data augmentation. *Comput Electron Agric*, 225:109306, 2024. DOI: 10.1016/j.compag.2024.109306
20. **Mon SL, Onizuka T, Tin P, Aikawa M, Kobayashi I, Zin TT:** AI-enhanced real-time cattle identification system through tracking across various environments. *Sci Rep,* 14:17779, 2024. DOI: 10.1038/s41598-024-68418-3
21. **Meng H, Zhang L, Yang F, Hai L, Wei Y, Zhu L, Zhang J:** Livestock biometrics identification using computer vision approaches: A review. *Agriculture*, 15 (1):102, 2024. DOI: 10.3390/agriculture15010102
22. **Sharma A, Randewich L, Andrew W, Hannuna S, Campbell N, Mullan S, Dowsey AW, Smith M, Hansen M, Burghardt T:** Universal bovine identification via depth data and deep metric learning. *Comput Electron Agric,* 229:109657, 2025. DOI: 10.1016/j.compag.2024.109657
23. **Wang Y, Mücher S, Wang K, Wang W, Kooistra L:** Deep metric learning for individual cattle identification using coat patterns: Proposal for a best practice. *Comput Electron Agric,* 238:110754, 2025. DOI: 10.1016/j.compag.2025.110754
24. **Yıldız Bİ:** Intelligent approaches in livestock farming: Using deep learning models. *Turkish JAF Sci Tech*, 12 (11): 1959-1967, 2024. DOI: 10.24925/turjaf.v12i11.1959-1967.6861
25. **Jing Y, Liang K, Zhang R, Sun H, Li Y, He Z, Ma Z:** Animal-CLIP: A dual-prompt enhanced vision-language model for animal action recognition. *Int J Comput Vis,* 133, 5062-5082, 2025. DOI: 10.1007/s11263-025-02408-4
26. **Zhang H, Shao H:** Exploring the latest applications of OpenAI and ChatGPT: An in-depth survey. *Comput Model Eng Sci*, 138 (3): 2061-2102, 2024. DOI: 10.32604/cmes.2023.030649
27. **Sam D, Willmott D, Semedo DJ, Kolter JZ:** Finetuning CLIP to Reason about Pairwise Differences. http://www. https://arxiv.org/abs/2409.09721; *Accessed:* 05.08.2025.
28. **Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I:** Learning transferable visual models from natural language supervision. *Proceedings of the 38th International Conference on Machine Learning*, Jul 18th-24th 2021, Virtual Congress, San Diego, CA, 92109. PMLR 139:8748-8763
29. **Martinez AM, Kak AC:** Pca versus lda. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 23 (2): 228-233, 2001.