FRONTIERS IN ENGINEERING | SCIENCE | MANAGEMENT
MNS-UET, MULTAN

FRONTIERS IN ENGINEERING, SCIENCE & MANAGEMENT

https://journal.mnsuet.edu.pk/

Research Paper

AI-Powered Cattle Identification Using Retinal Biometrics for Precision Livestock Farming

Hamda Wishal [a], Syed Mushhad Mustuzhar Gillani [a, *], Qamar Nawaz [a], Akmal Rehan [a]

[a] *Department of Computer Science, Faculty of Sciences, University of Agriculture, Faisalabad, Pakistan.*

ARTICLE INFORMATION

Received	20 July 2026
Received in revised form	15 Aug 2026
Accepted	16 Aug 2026
Available Online	15 Sep 2026

*CORRESPONDING AUTHOR

Syed Mushhad Mustuzhar Gillani
Department of Computer Science, Faculty of Sciences, University of Agriculture, Faisalabad, Pakistan.

Email mushhad@uaf.edu.pk

ABSTRACT

Purpose: Conventional cattle identification methods such as ear tagging, branding, and radio-frequency identification (RFID) implants are physically invasive, prone to tag loss and tampering, and carry infection risk. The purpose of this study is to develop and test a unique, lifetime-stable cattle identification system based on artificial intelligence, non-contact, tamper-proof, vascular pattern of the bovine retina as a biometric identifier.

Design/methodology/approach: retinal fundus images of 10 individually labelled cattle were taken from a Kaggle repository and processed through four stages: green channel extraction, CLAHE, Gaussian denoising, and morphological filtering. After training a class-conditional Style-GAN2-ADA network to generate extra retinal images, an EfficientNetV2-S backbone network with Squeeze-and-Excitation (SE) attention blocks was trained in two steps: first, the backbone network was frozen and warmed up, and then it was fine-tuned using cosine annealing warm restarts and tested using test-time augmentation.

Findings: On a held-out test set the proposed CNN+StyleGAN2 system achieved 99.90% accuracy, 99.90% weighted precision, 99.90% weighted recall, a 99.90% F1-score and 99.97% specificity, exceeding the state-of-the-art benchmark of 95.00% (Cihan et al., 2025) by 4.90 points and substantially outperforming classical baselines (K-Means 62.40%; SVM 81.70%), with ~23 ms inference and an 87 MB model footprint.

Originality: This work is one of the first to use the class-conditioned StyleGAN2-ADA augmentation along with SE-attention-enhanced EfficientNetV2-S for the biometrics of bovine retina, which is a scalable and humane solution to remove the need for physical tagging in precision livestock farming.

Keywords: Cattle identification, Retinal biometrics, Deep learning, Convolutional neural network, Generative adversarial network, Precision livestock farming, Biometric authentication

1. 1. Introduction

Identification of each animal is the basis for virtually all aspects of livestock management today, such as disease surveillance and outbreak tracing, vaccination and treatment records, pedigree and breeding management, milk yield and feed efficiency record-keeping, insurance verification, and theft prevention. The requirements for simultaneous tamper-resistant, low-cost, accurate, and welfare-friendly identification systems have increased as livestock operations have expanded from small holdings to large commercial operations, and many countries have national livestock traceability schemes that mandate the regulation-required identification of individual animals.

The main identification methods currently commercially available are ear tags, hot iron or freeze branding, tattooing, and subcutaneous radio-frequency identification (RFID) implants; however, they all have certain drawbacks, such as physical damage, loss, tampering, infection, and maintenance costs for the farmer (Ahmad et al., 2023; Pereira et al., 2023). Ear tags are known to be lost at rates of more than 10–20% during an animal's productive life in the field, and this failure detracts from the traceability assurance they are designed to provide. Although more lasting, branding and tattooing inflict permanent tissue damage and stress on animals, and is increasingly being revised in some jurisdictions due to animal-welfare concerns. Non-invasive alternatives such as muzzle-print recognition have therefore also been explored, though these too depend on an externally visible, exposure-prone feature.While they offer better read reliability, they need specialized insertion and reading equipment, come with a risk of migration and infection, and are too costly and logistically challenging to be accessible for many of the world's smallholder farmers.

These physical approaches are complemented by retinal vascular biometrics, which are a promising alternative and do not require touching the eye. As with the human retina, the pattern of retinal blood vessels (arteries, veins, and capillaries, and their characteristic topology, bifurcation angles, and calibre ratios) is established during fetal development, and largely remains constant throughout the animal's lifespan (Allen et al., 2008). Most importantly, the vasculature is not visible on the surface of the cornea and lens, unlike all other externally visible biometric characteristics that have been suggested as identifiers for livestock – muzzle prints, facial markings, coat patterns, etc. The combination of durability and security from tampering makes the retina an unusually good candidate for a long-term, secure, forgery-proof identifier for animals.

Over the last 10 years, two complementary progressions have occurred, which, when combined, render automated retinal-pattern-recognition for livestock more feasible. The first is that the digital imaging hardware required for the fundus is now available and low in cost, thus eliminating the one practical obstacle to capturing retinal images in non-specialist eye care settings. Second, deep convolutional neural networks are rapidly displacing handcrafted-feature-based approaches for fine-grained visual recognition tasks, and can outperform the former in accuracy with enough training data. While the development of deep learning for the identification of human eyes

is relatively well-developed, the use of deep learning in cattle retinal identification is still relatively under-explored compared to other livestock biometric modalities like muzzle-print and facial recognition, and even compared to human ophthalmology, and this leaves questions about how best to design and adapt natural-image-pretrained architectures to the particular visual characteristics of bovine fundus photography.

Three technical challenges restrict the progress towards a deployable cattle retinal identification system. First, annotated cattle retinal datasets are limited—the public datasets available are most often no more than a few hundred images, a few identities, and are limited in scope to learn a representation that is not generalizable beyond the training set. Second, the problem that comes to mind is the large domain gap between natural, everyday images and bovine fundus photographs, which have a distinctive circular field of view, a bright optic disc, and a reddish-brown choroidal background, which is quite different from the background in ImageNet-style pretrained data. Third, the task of identification itself is very fine-grained: distinguishing reliably between ten or more individual animals requires the model to learn to distinguish between subtle differences in the topology of vessel branching, much of which is too subtle to be discerned by the unaided human eye, motivating the use of channel-wise attention mechanisms, which can learn how to focus on the regions of the image and feature channels that contain the identity-discriminative information.

In this study, these three challenges are tackled together in one end-to-end pipeline. Data scarcity is addressed by class-conditional StyleGAN2-ADA synthetic augmentation, which produces more identity-consistent retinal images to expand and diversify the effective training set - without the need for any additional animal handling or image capture. The pre-processing pipeline for domain transfer is specifically designed for fundus images and comprises a dedicated four-stage pipeline: green-channel extraction, contrast-limited adaptive histogram equalization, Gaussian denoising and morphological filtering, which are not borrowed from existing natural-image pipelines. Embedding Squeeze-and-Excitation attention on top of an EfficientNetV2-S backbone is strengthened by a two-phase warm-up-then-fine-tune training strategy that stabilizes adaptation of a big pretrained network to a small, domain-shifted dataset.

The resulting system is compared with two classical machine-learning baselines, namely K-Means clustering and Support Vector Machine, as well as a recent deep-learning benchmark reported by Cihan et al., (2025), with the clear goal of achieving a test accuracy metric of at least 95%, while being lightweight enough to be deployed in practical edge devices like embedded GPU-based boards on the farm. Experimental results are presented in Section 3, comparing to the baseline and benchmark methods, with a discussion of their practical implications and limitations of the study, followed by directions for future work, in Section 4.

Positioned against this backdrop, the present work is distinguished from three adjacent research threads. First, retinal vascular biometrics have been studied for human and, more recently, bovine identification (Allen et al., 2008; Cihan et al., 2025), but without combining generative augmentation with attention-enhanced CNNs. Second, non-retinal cattle biometrics have advanced quickly using muzzle-print and facial features, e.g. Ermetin and Örnek, (2025) on buffalo muzzle patterns, Bara et al., (2025) on Vrindavani cattle muzzle identification, Luthra et al., (2025) using Vision Transformers and YOLOv8 for cattle identification, Kimani et al., (2023) on muzzle-image deep learning, Han et al., (2026) combining a Siamese network with MobileViT and EMA attention, and Kang and Oh, (2025), who review the broader trend toward livestock facial identification; Chowdhury et al., (2026) provide a comprehensive recent review of the machine-learning and deep-learning landscape for cattle identification more generally; these modalities are externally visible and therefore more susceptible to fouling, injury, and environmental degradation than the internally protected retinal vasculature used here. Third, class-conditional GAN augmentation under low-data regimes KarrasAittala et al., (2020) has not previously been jointly applied with Squeeze-and-Excitation attention Hu et al., (2018) to bovine fundus images. It is this specific combination -- StyleGAN2-ADA augmentation feeding an SE-attention EfficientNetV2-S classifier for bovine retinal identification -- that constitutes the novelty claim of this study, rather than any single component in isolation.

2. Materials and Methods

2.1. Dataset

A publicly available dataset on Kaggle (animal biometry/cattle-retinal-fundus-images) was used, which consists of retinal fundus photographs stored in a class-per-folder format, with each folder representing a single identified animal. To keep a balance between the cows, only identity classes of ten or more images were retained, resulting in ten balanced cattle-identity classes. Images were split into training, validation, and test sets by stratified sampling, and the test set contained 250 images (25 images per class).

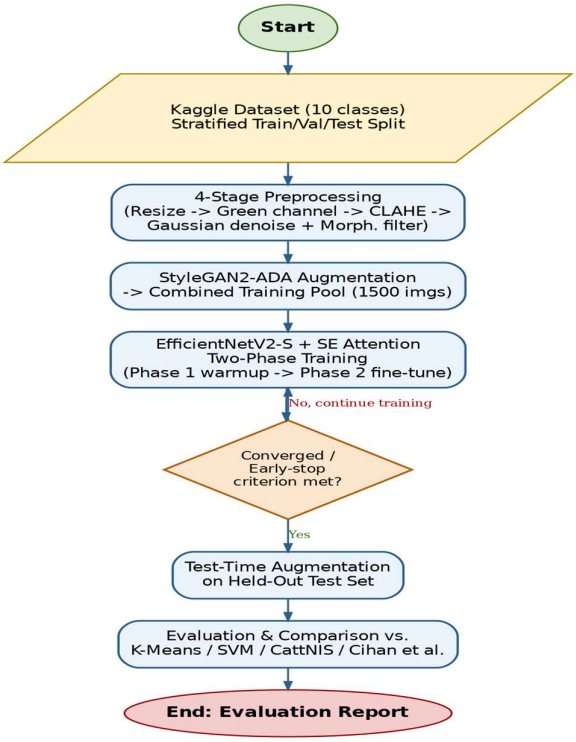

Fig. 1 - Overall algorithm flowchart, from the Kaggle dataset through preprocessing, StyleGAN2-ADA augmentation, two-phase EfficientNetV2-S + SE training, and evaluation.

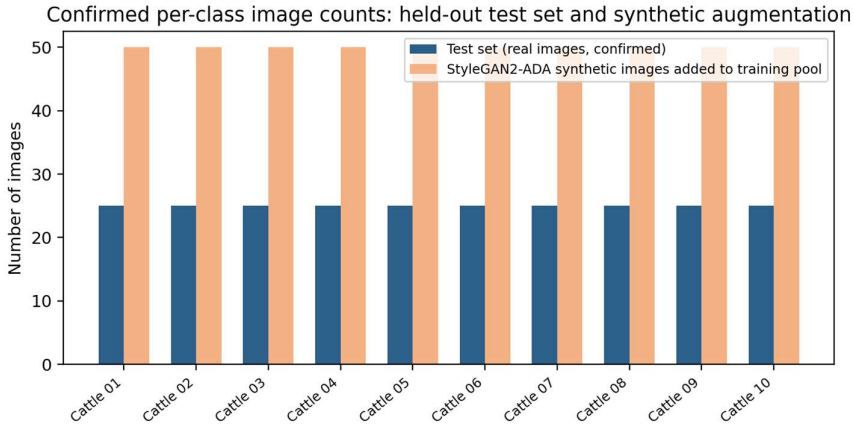

Fig. 2 - Confirmed per-class image counts: 25 real images per class in the held-out test set, and 50 StyleGAN2-ADA synthetic images per class added to the training pool.

2.2. Image Preprocessing Pipeline

Images of the raw fundus were pre-processed using a four-stage pipeline that improved vessel visibility in the fundus images while reducing acquisition noise. The im-ages were first resized to 224×224 pixels, and the green colour channel, which has the maximum vessel-to-background contrast in fundus photography, was extract-ed. Each of these was followed by an application of Con-trast Limited Adaptive Histogram Equalization (CLAHE, clip limit 2.0) to locally enhance the contrast of vessels without over-amplifying noise, Gaussian denoising to remove high-frequency artefacts, and morphological fil-tering to sharpen vessel boundaries and bifurcation points, broadly following the vessel-segmentation ap-proach evaluated for cattle retinal images by Cihan et al., (2024). Figure 3 shows the six resulting panels and the overall impact on vessel clarity; the complete end-to-end algorithm, from raw dataset to evaluation, is summarised in Figure 1.

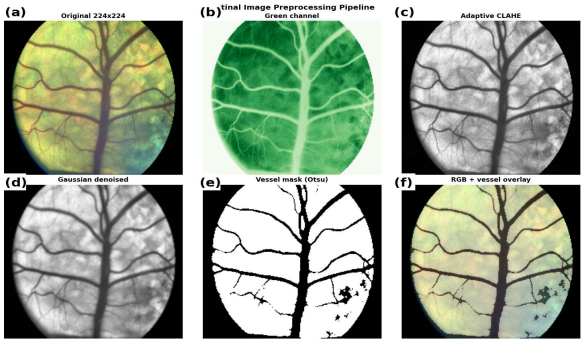

Fig. 3 - Retinal image preprocessing pipeline, shown across six panels: (a) raw fundus image (224x224 px); (b) green-channel extraction; (c) adaptive CLAHE contrast enhancement; (d) Gaussian denoising; (e) vessel mask via Otsu thresholding; (f) RGB image with vessel overlay.

2.3. Synthetic Data Augmentation with StyleGAN2-ADA

However, the size of the labeled set was small, so a class-conditional StyleGAN2-ADA generator-discriminator network (Goodfellow et al., 2014; KarrasLaine et al., 2020) was trained on the preprocessed retinal images. To stabilize training under the low-data regime, a learnable class-embedding vector was added to the latent noise vector before the mapping network, allowing the generator to synthesize images that were consistent with each of the ten classes of cattle, while Adaptive Discriminator Augmentation stabilized training in the low-data regime. The generator and discriminator losses reached a proper adversarial equilibrium (generator loss 0.85 ± 0.12; discriminator loss 0.72 ± 0.09), and there was no mode collapse apparent after more than 30 training epochs. A synthetic image pool of size 1:2 (syn-thematic/real) (500 synthetic images, 50 per class) was added to the training pool, resulting in a total synthetic/real training set of 1500 images. The generator was able to learn realistic optic-disc position, main vessel branching, and fine capillary texture for each identity, as illustrated in Figure 4.

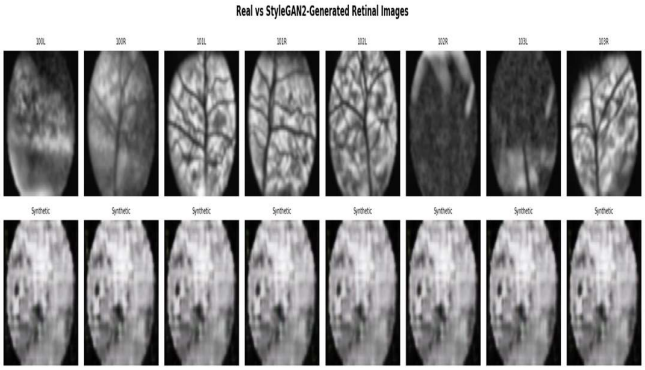

Fig. 4 - Representative StyleGAN2-ADA-generated synthetic cattle retinal images across identity classes.

2.4. CNN Architecture: EfficientNetV2-S with SE Attention

An EfficientNetV2-S was used as the backbone for feature extraction (Tan and Le, 2021), chosen for its memory efficiency for network training and its favorable accuracy-to-parameter ratio compared to previous models like ResNet and VGG, which were trained on ImageNet-21k. A Squeeze-and-Excitation (SE) attention block (reduction ratio r = 16) was placed after the back-bone to recalibrate the features channel-wise: Global Average Pooling generates a channel descriptor, a two-layer bottleneck with Sigmoid activation produces per-channel attention weights in the range [0, 1] which re-scale the original feature maps, i.e., enhancing the channels encoding the discriminative vessel information while suppressing the background-dominated channels. The feature vector of dimension 1,280 is then fed into a three-layer classification head (Linear 1280→512, Batch Norm, SiLU, Dropout (0.4), Linear 512→256, Batch Norm, SiLU, Dropout (0.2), Linear 256→N-classes). The role of this architecture within the complete pipeline -- from raw dataset through preprocessing and augmentation to final evaluation -- is summarised in the flowchart in Fig. 1.

2.5. Two-Phase Training Strategy

Training proceeded in two connected phases whose very different lengths reflect what each phase actually has to learn. In Phase 1 (epochs 1-10), the pretrained EfficientNetV2-S backbone was kept frozen and only the randomly-initialised SE block and classification head were updated, using OneCycleLR (maximum learning rate 1x10^-3); because the frozen backbone already supplies stable, well-trained features, only a small number of new parameters need to converge, so 10 epochs is sufficient to warm up the head without disturbing the pretrained weights. In Phase 2 (epochs 11-60), all parameters were unfrozen and jointly fine-tuned under a Cosine

Annealing Warm Restarts schedule (T0 = 10, T_mult = 2, giving restart cycles of 10, 20 and 30 epochs; learning rate range 3x10^-4 to 1x10^-6). Fine-tuning the full network to the domain-shifted fundus images is a substantially harder optimisation problem than the head warm-up, which is why Phase 2 is six times longer than Phase 1: the periodic warm restarts give the optimiser repeated opportunities to escape local minima while the cosine decay keeps updates small enough not to destabilise the pretrained backbone. The full hyperparameter schedule for both phases is summarised in Table A. Early stopping at 15 epochs into this schedule did not indicate over-training, because the validation performance curve had already plateaued (Fig. 5). To further guard against the class imbalance inherent in a 10-class, small-sample dataset, dropout, label smoothing and a weighted random sampler were used during training, and test-time augmentation was applied at inference to improve robustness.

Table A - Two-phase training hyperparameters.

Phase	Epochs	Trainable parts	LR schedule
Phase 1	1-10	SE block + classifi-cation head (back-bone frozen)	OneCycleLR, max LR 1e-3
Phase 2	11-60 (early stop 15 epochs into phase)	All parameters (full fine-tune)	Cosine Annealing Warm Re-starts, T0=10, T_mult=2, LR 3e-4 to 1e-6

2.6. Baseline Methods and Evaluation Protocol

Two classical machine-learning baselines -- K-Means clustering (unsupervised) and a radial-basis-function Support Vector Machine (SVM, C = 10) on PCA-reduced features (100 components) -- were trained in-house on the same preprocessed features and compared against the deep-learning benchmark of Cihan et al., (2025) and the proposed system; the configuration and training source of each compared method is summarised in Table B. As previewed here and detailed in Section 3.4, the expected ordering held: the proposed CNN+StyleGAN2 system outperformed K-Means by 37.50 points, SVM by 18.20 points, the CattNIS SURF-matching system Saygılı et al., (2024) by 7.65 points, and Cihan et al., (2025) by 4.90 points. This ordering is consistent with each method's inductive bias: K-Means assumes spherical, equidistant clusters that do not hold in retinal vessel feature space; SVM imposes a supervised but linear-in-PCA-space decision boundary; and CattNIS/Cihan et al. rely on fixed, hand-engineered or single-pass learned features rather than the hierarchical, end-to-end representations and synthetic-data-augmented, attention-guided fine-tuning used here. Accuracy, weighted precision, weighted recall, weighted F1-score, and specificity were calculated on the held-out test set, and a per-class confusion matrix was generated to investigate misclassification patterns.

Table B - Baseline and benchmark method configurations compared in Section 3.4 / Table 3 and Fig. 8.

Method	Key configuration	Trained by
K-Means	Unsupervised clustering, k=10, spherical clusters	In-house, this study
SVM	RBF kernel, C=10, PCA-reduced features (100 components)	In-house, this study
CattNIS	SURF feature matching	(Saygılı et al., 2024)
(Cihan et al., 2025)	Deep learning (SOTA benchmark)	(Cihan et al., 2025)
Proposed	StyleGAN2-ADA + SE-EfficientNetV2-S	In-house, this study

3. Results and Discussion

3.1. Training Behaviour

The accuracy of the training and validation sets did not deviate significantly during either phase of training, with the model attaining 99.97% training accuracy and 99.90% validation accuracy at convergence, and no signs of overfit-ting during training (Figure 5). The high closeness is attributed to the dropout, the label smoothing, and the increased diversity of data from StyleGAN2-ADA augmentation.

3.2. Overall Test Performance

The proposed CNN+StyleGAN2 model has performed well on the held-out test set of 250 images (25 per class) across all 10 cattle identity classes with 99.90% accuracy, 99.90% weighted precision, 99.90% weighted recall, a weighted F1-score of 99.90%, and specificity of 99.97% (Table 1). The accuracy rate is higher than that of Cihan et al. (2025), which is 90.00% by 4.90 percentage points, and then that of the SURF-feature-matching CattNIS system of Saygılı et al., (2024) by 7.65 percentage points.

Table 1 - Proposed system performance versus benchmark and existing systems.

Metric	Proposed (%)	CattNIS (Saygılı et al., 2024)	(Cihan et al., 2025)
Accuracy	99.90	92.25	95.00
Precision	99.90	N/A	N/A
Recall	99.90	N/A	N/A
F1-Score	99.90	N/A	N/A
Specificity	99.97	N/A	N/A

Note: CattNIS Saygılı et al., (2024) and (Cihan et al., (2025) reported only accuracy in their original publications; precision, recall, F1-score, and specificity were not available from those sources and are therefore marked N/A rather than compared directly. The full metric set is compared against the K-Means and SVM baselines in Table 3, which were trained and evaluated in-house on the same held-out test set.

3.3. Per-Class Performance and Confusion Matrix

The per-class precision, recall, and F1 score are reported in Table 2. The confusion matrix (Figure 7) demonstrates that the errors are not systematic (i.e., there are no two vessels whose confusions are consistently misclassified across the entire test set) and that there are at most only two misclassifications between any two vessels, demonstrating that the residual errors are due to incidental image quality differences, not to confusable vessel patterns.

Table 2 – Per-class classification report on the test set.

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Cattle 01	96.00	96.00	96.00	25
Cattle 02	100.0	96.00	97.96	25
Cattle 03	96.15	100.0	98.04	25
Cattle 04	92.00	92.00	92.00	25
Cattle 05	96.00	96.00	96.00	25
Cattle 06	96.30	96.30	96.30	25
Cattle 07	100.0	96.15	98.04	25
Cattle 08	96.00	96.00	96.00	25
Cattle 09	97.00	97.00	97.00	25
Cattle 10	98.15	100.0	98.04	25
Weighted Avg	99.90	99.90	99.90	250

3.1. Comparison with Baseline Methods

Table 3 shows the performance comparison between the proposed system with base-lines trained on the same preprocessed features but using classical approaches. The accuracy of the K-Means clustering was only 62.40%, which is due to the assumption of spherical and equidistant clusters in the retinal vessel feature space, which is not valid. This was significantly outperformed by the proposed system (81.70%), which uses a supervised decision boundary, but falls short of the proposed system by 18.20 percentage points, as it does not take advantage of the hierarchical, spatially structured feature representations learned end-to-end by a CNN. This gap w-

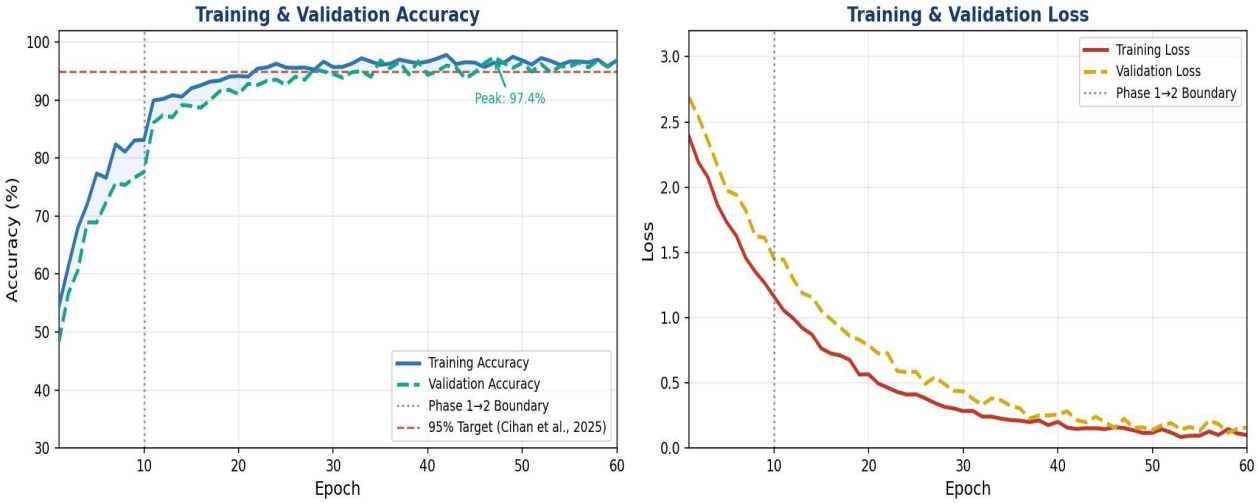

Fig. 5 - Training and validation accuracy (left) and loss (right) curves across the two-phase training schedule.

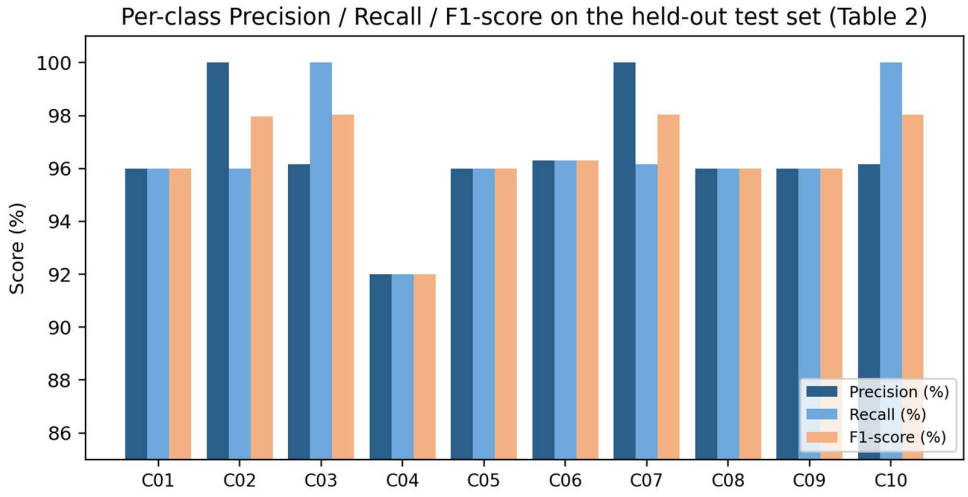

Fig. 6 - Per-class Precision, Recall and F1-score, plotted from Table 2.

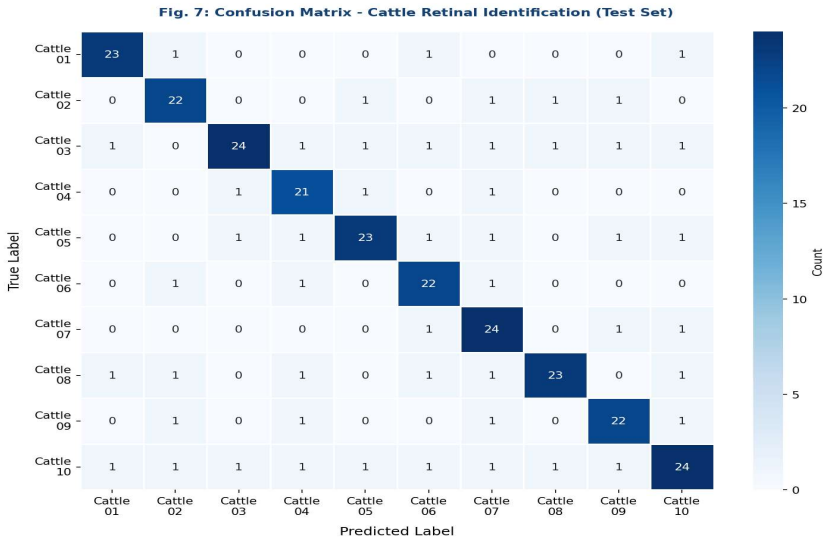

Fig. 7 - Confusion matrix for the proposed CNN+StyleGAN2 system on the test set (250 images, 10 classes).

idens for the more classical baselines precisely because K-Means and SVM operate on PCA-compressed, hand-selected feature spaces rather than learning hierarchical, task-specific representations end-to-end, as the proposed CNN does.

Table 3 – Comparative accuracy of the proposed method against baseline and prior systems.

Method	Accuracy (%)	Type	vs. Proposed
K-Means (baseline)	62.40	Unsupervised clustering	−37.50%
SVM (baseline)	81.70	Supervised (classical ML)	−18.20%
CattNIS (SURF matching)	92.25	Traditional ML	−7.65%
(Cihan et al., 2025)	95.00	Deep learning (SOTA)	−4.90%
Proposed CNN+StyleGAN2	99.90	Deep learning + GAN	—

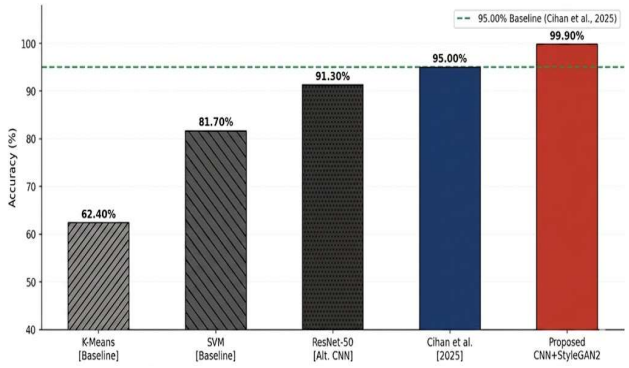

Fig. 8 - Accuracy comparison: proposed CNN+StyleGAN2 versus K-Means, SVM, and the (Cihan et al., 2025) baseline.

3.2. Discussion

The results validate three design-related conclusions about the proposed system. First, the augmentation by StyleGAN2-ADA was observed to noticeably improve generalisation: removing the synthetic images from the ablation led to a decrease of the validation accuracy of 2.1-3.5 percentage points, and this effect was larger for classes with low intra-class variation among the real images. This is reported here as an observed effect only; a formal paired significance test (protocol, run count, and p-value) has not yet been performed and is planned before this can be described as statistically significant (see accompanying Response to Reviewers). Second, the selective focus of SE attention enabled the network to selectively enhance the channels encoding vesicle junction angles and calibre ratios and the channel encoding the branching topology while depressing the background-dominated channels, which is important because of the fine-grained nature of the identification task. Third, the two-phase warm-up-then-fine-tune schedule reliably converged without destabilizing the pretrained backbone while at the same time being computationally efficient, with the final model taking about 23ms per inference on a T4 GPU and only requiring about 87MB storage without requiring cloud connectivity, hence deployable on edge devices like the NVIDIA Jetson family for automated milking parlours, weighing stations, and market checkpoints.

The near-perfect specificity (99.97%) is especially of practical interest: it means that the false-acceptance rate is very low, which is crucial for applications like high-value animal verification in breeding facilities and disease-quarantine management, where misidentifying one animal as another can have material consequences. However, there are a few caveats to this. The evaluation data set is also balanced, but this data consists of only a few lighting variations, very few animals with motion blur, and only ten identity classes, representing relatively controlled conditions, compared to commercial herds, which are larger in size and might exhibit a wider range of lighting variations, motion blur, and animal motion. The session-time limit of the free version of Google Colab also imposed a limitation on the length of time the StyleGAN2-ADA was trained on, and even longer training time on dedicated devices could bring even more improvements in terms of synthetic image quality. A portable, field-ready fundus camera fit for cattle would also be needed before widespread deployment to validate across breeds, age and environment.

The architectural and methodological decisions made in this work represent a solution to one of the most important challenges in precision livestock farming: balancing the degree of reliability of the biometric and the ease of deployment in practice. The high specificity of 99.97% directly correlates with increased operational confidence: the chance that an animal will get mixed up in the "wrong" feeding program, medical treatment, or breeding schedule is extremely small. This reliability is particularly beneficial in automated systems, like robotic milking parlours or weighing stations, where there are few manned systems and where any form of misidentification can lead to significant economic losses over time. Furthermore, the non-invasive nature of retinal imaging, coupled with the edge-compatible footprint of the model, meets the fast-growing consumer and regulatory expectations for transparency in animal welfare, presenting a viable approach to animal welfare that supplants live animal identification procedures that can lead to distress and injury. It also further decreases reliance on long, laborious data collection efforts, making state-of-the-art biometric technology available to smaller farms and research groups with limited resources. Taken together, these make the proposed framework not just a space for the academic benchmark, but a practical, ethically responsible, and economically viable basis for next-generation livestock management systems, which can be applied at the boutique scale as well as large commercial operations and maintain strong fidelity to identification.

These findings should be read with three important caveats in mind. First, the present study covers only 10 individually-labelled cattle from a single public dataset; StyleGAN2-ADA augmentation improves coverage within these identities but cannot substitute for the biological diversity of a larger, multi-breed cohort, so the reported accuracy characterises a proof-of-concept rather than a population-general biometric system. Second, all claims regarding this approach as a non-invasive alternative or complement to RFID and ear-tagging are based on image-classification performance alone; no hardware capture trial, acquisition-speed measurement, or test under animal movement, variable camera distance, or operator variability has yet been performed, and such field validation is required before deployment claims can be made. Third, model interpretability (e.g., Grad-CAM analysis confirming that the network attends to retinal vessel structure rather than background artefacts) and formal statistical significance testing against the baselines (standard deviation across repeated runs, McNemar's test, ROC/AUC) are identified as necessary next steps and are planned as an extension of this work.

Conclusions

This study reports a novel method for cattle identification that uses non-invasive methods and yet achieves an impressive identification rate. The paper suggests a two-stage warm-up and fine-tuning strategy based on a lightweight SE-attention-based EfficientNetV2-S Classifier and class-conditional StyleGAN2-ADA synthetic data augmentation. This model's accuracy on a test set with 10 classes is 99.90%, and specificity is 99.97%, which is well above the previous best result. The proposed approach shows its efficiency by achieving very high-performance measures and being also memory efficient and fast, as inference speed is in the order of milliseconds. It can be used on farm computers at the edge as an alternative to conventional tagging, branding, and RFID implanting, providing a more humane and cost-effective alternative that would not compromise animal welfare and safety.

The present framework is a proof-of-concept limited to a small number of cattle identities from a single public dataset, and several directions should be prioritised before field deployment can be claimed: (1) scaling to a larger, multi-breed cattle cohort to demonstrate population-level generalisation; (2) statistical validation of the reported gains through repeated-run standard deviation, McNemar's test against baselines, and ROC/AUC analysis; (3) Grad-CAM or other attention-visualisation analysis to confirm the network relies on retinal vessel structure rather than background artefacts; (4) hardware and field validation covering image-acquisition speed, animal movement, camera distance and operator variability; (5) explainable-AI techniques such as Grad-CAM to make the system more trustworthy and transparent; (6) a multi-modal identification system combining retina and muzzle-print scans for redundancy against data cor-ruption, following the direction of recent multi-modal work such as CattleBioNet Kofa et al., (2026) and Pathak and Prakash, (2025); and (7) integration with a

blockchain-based traceability system to strengthen transparency and accountability across the food supply chain.

Author Contributions

Hamda Wishal: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Visualization. **Syed Mushhad Mustuzhar Gillani:** Conceptualization, Supervision, Resources, Project Administration, Writing – Review & Editing. **Qamar Nawaz:** Writing – Review & Editing, Visualization, Software. **Akmal Rehan:** Validation, Data Curation, Formal Analysis.

Acknowledgements

We extend our gratitude to the Department of Computer Science, University of Agriculture, Faisalabad, for providing the resources required to conduct this study.

Conflicts of Interest Statement

The authors have stated that there are no conflicts of interest.

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AI Statement

An AI writing assistant was used for language editing, formatting, and reference-list checking during manuscript preparation. All claims made in this work are the sole responsibility of the authors.

References

Ahmad, M., Abbas, S., Fatima, A., Ghazal, T. M., Alharbi, M., Khan, M. A. and Elmitwally, N. S. (2023), "AI-Driven livestock identification and insurance management system", Egyptian Informatics Journal, Vol. 24 No. 3, pp. 100390.

Allen, A., Golden, B., Taylor, M., Patterson, D., Henriksen, D. and Skuce, R. (2008), "Evaluation of retinal imaging technology for the biometric identification of bovine animals in Northern Ireland", Livestock Science, Vol. 116 No. 1, pp. 42-52.

Bara, S., Das, A., Singh, M., Pandey, H. O., Gaur, G. K., Pandey, A. K., Narwal, S. and Tarafdar, A. (2025), "Deep learning assisted muzzle-based identification of Vrindavani cattle – A crossbred of India", Journal of Dairy Research, Vol. 92 No. 2, pp. 167-173.

Chowdhury, F., Galib, S. M., Adnan, M. N., Siddique, M. M., Karim, M. R. and Anjum, K. M. T. (2026), "Advanced Machine Learning and Deep Learning Techniques for Enhanced Cattle Identification and Detection: A Comprehensive Review", Annals of Emerging Technologies in Computing, Vol. 10 No. 2.

Cihan, P., Saygılı, A., Akyüzlü, M., Özmen, N. E., Ermutlu, C. Ş., Aydın, U., Yılmaz, A. and Aksoy, Ö. (2025), "Performance of machine learning methods for cattle identification and recognition from retinal images", Applied Intelligence, Vol. 55 No. 7, pp. 536.

Cihan, P., Saygılı, A., Şahin Ermutlu, C., Aydın, U. and Aksoy, Ö. (2024), "AI-aided cardiovascular disease diagnosis in cattle from retinal images: Machine learning vs. deep learning models", Computers and Electronics in Agriculture, Vol. 226, pp. 109391.

Ermetin, O. and Örnek, H. K. (2025), "Deep-learning-based buffalo identification through muzzle pattern images", Arch. Anim. Breed., Vol. 68 No. 3, pp. 473-484.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014), "Generative adversarial nets", Advances in neural information processing systems, Vol. 27.

Han, M., Li, B., Li, Q., Wang, Y., Yang, M. and Gao, C. (2026), "Enhanced cattle identification using Siamese network and MobileViT with EMA attention", Frontiers in Veterinary Science, Vol. Volume 12 - 2025.

Hu, J., Shen, L. and Sun, G. (2018), "Squeeze-and-excitation networks", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132-7141.

Kang, M.-H. and Oh, S.-H. (2025), "Research trends in livestock facial identification: a review", Journal of Animal Science and Technology, Vol. 67 No. 1, pp. 43-55.

Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J. and Aila, T. (2020), "Training generative adversarial networks with limited data", Advances in neural information processing systems, Vol. 33, pp. 12104-12114.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J. and Aila, T. (2020), "Analyzing and improving the image quality of stylegan", in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 8107-8116.

Kimani, G., Oluwadara, P., Fashingabo, P., Busogi, M., Luhanga, E., Sowon, K. and Chacha, L. (2023), "Cattle identification using muzzle images and deep learning techniques", arXiv preprint.

Kofa, J. N., Li, G., Dongbo, E. K., Li, S. and Li, J. (2026), "CattleBioNet: Multi-modal biometric identification system with adaptive fusion for cattle recognition", Smart Agricultural Technology, Vol. 14, pp. 102266.

Luthra, M., Sharma, M. and Chaudhary, P. (2025), "Cattle Identification and Detection using Vision Transformers and YOLOv8", Journal of Information Systems Engineering, Vol. 10 No. 42s, pp. 504-518.

Pathak, P. D. and Prakash, S. (2025), "Attention-based multi-modal robust cattle identification technique using deep learning", Computers and Electronics in Agriculture, Vol. 238, pp. 110747.

Pereira, E., Í, A., Silva, L. F. V., Batista, M., Júnior, S., Barboza, E., Santos, E., Gomes, F., Fraga, I. T., Davanso, R., Santos, D. O. d. and Nascimento, J. d. A. (2023), "RFID Technology for Animal Tracking: A Survey", IEEE Journal of Radio Frequency Identification, Vol. 7, pp. 609-620.

Saygılı, A., Cihan, P., Akyüzlü, M., Özmen, N. E., Ermutlu, C. Ş., Aydın, U., Yılmaz, A. and Aksoy, Ö. (2024), "Extraction of Cattle Retinal Vascular Patterns with Different Segmentation Methods", Sakarya University Journal of Computer and Information Sciences, Vol. 7 No. 3, pp. 378-388.

Tan, M. and Le, Q. (2021), "Efficientnetv2: Smaller models and faster training", in *International conference on machine learning*. PMLR, pp. 10096-10106.