Indian Journal of Animal Sciences **96** (5): 391–397, May 2026/Article
https://doi.org/10.56093/ijans.v96i5.175242

Automated dairy cattle identification: A YOLOv5-Siamese framework for Deoni breed

RAJAS CHAVHAN[1], ANUJ MAHAJAN[1], ADITI DHENGE[1], NETRA BATWE[1], POONAM SHARMA[2], M M VAIDYA[3] and V B DONGRE[4]✉

Maharashtra Animal and Fishery Sciences University, Nagpur 440 006 Maharashtra India

Received: 16 January 2026; Accepted: 02 July 2026

ABSTRACT

Accurate identification of indigenous cattle breeds is important for livestock management, selective breeding, and conservation of valuable genetic resources. Deoni cattle, a recognized indigenous breed of India, exhibit considerable phenotypic variation that can make visual breed identification challenging and subjective. The present study investigated the feasibility of using a deep learning-based approach for automated Deoni cattle identification from images. A hybrid framework combining object detection for feature localization and a Siamese neural network for similarity learning was developed. A dataset comprising images of Deoni cattle, other cattle breeds, and non-cattle animal species was assembled and evaluated under varying field conditions, including differences in pose, illumination, and background. The model extracted discriminative visual features from facial and morphological characteristics and mapped them into an embedding space to distinguish similar and dissimilar images. Performance was assessed using standard classification and verification metrics. The proposed framework achieved effective discrimination between Deoni cattle and non-target animals within the available dataset and demonstrated consistent performance across different image conditions. However, the study was conducted using a relatively limited dataset, and further validation on larger and more diverse populations is required to establish broader generalizability. The findings indicated that deep learning-based image analysis can serve as a promising non-invasive tool for cattle breed identification and may support digital livestock record systems, breed characterization, and conservation initiatives.

Key words: Breed identification, Deoni cattle, Machine learning, YOLOv5

Cattle farming and animal husbandry play an important role in the agricultural sector and contribute significantly to rural livelihoods, particularly through dairy farming, which serves as a major supplementary income source for farming households. Accurate identification of cattle breeds is essential for efficient livestock management, genetic improvement programs, and conservation of indigenous breeds (Jogi *et al.* 2024). Deoni cattle, a prominent indigenous breed primarily distributed in the drought-prone regions of the Marathwada area of Maharashtra and parts of Karnataka, are valued for their milk production and adaptability to harsh climatic conditions. The presence of three distinct color strains viz. Wanera, Balankya, and Shevera often makes visual identification difficult. Traditional identification methods such as visual inspection and physical marking are often subjective, labor-intensive, and may compromise animal welfare (Dongre *et al.* 2017). Although identification tools like ear tags and rumen boluses are widely used, they are prone to loss, damage, or manipulation (Weber *et al.* 2020). Automated identification using computer vision and machine learning provides a non-invasive and scalable alternative. However, existing studies often rely on controlled datasets or focus mainly on commercial breeds, limiting their applicability to indigenous Indian cattle. To address these limitations, the present study proposed a deep learning framework integrating YOLOv5-based feature extraction with a Siamese neural network trained using triplet loss for robust similarity learning, enabling reliable identification of Deoni cattle under diverse field conditions and supporting digital livestock management and breed conservation.

Present address:[1]UG Student, Visvesvaraya National Institute of Technology, Nagpur, [2]Associate Professor, Visvesvaraya National Institute of Technology, Nagpur (poonamsharma@cse.vnit.ac.in),[3]Assistant Professor, College of Veterinary and Animal Sciences, Akola; Maharashtra Animal and Fishery Science University, Nagpur (mangeshvaidya@mafsu.ac.in), [4]Subject Matter Specialist, College of Veterinary and Animal Sciences, Udgir; Maharashtra Animal and Fishery Science University, Nagpur (vilasdongre@mafsu.ac.in). ✉Corresponding author email: vilasdongre@mafsu.ac.in

MATERIALS AND METHODS

The proposed framework used YOLOv5 architecture for feature extraction and Siamese neural network trained with triplet loss function for discriminative learning in an embedding space. The aim was to create a system able to precisely classify deoni animals and verify their breeding accuracy. Initially, a standard Resnet-50 convolutional neural network was used for the same classification process but the model faced some problems, mainly in capturing some very fine-grained details and visuals between the two classes, Deoni and Non-Deoni and this also led to overfitting and unstable validation accuracy.

To address this, ResNet-Siamese network was implemented where two parallel ResNet encoders were trained using contrastive loss function. While this improved the model's ability to learn relative similarity between images, the high-level features extracted by ResNet were still not sufficiently localized in capturing cattle-specific identity traits (e.g., head patterns, ear shape, and spot alignment).

So in the next part, YOLOv5s feature extractor was used instead of Res Net encoder because of YOLO's capability of capturing the object level features. But the model still did not improve its performance as it used contrastive loss while training which penalized both similar and dissimilar pair of images and in comparison, to triplets, it lacked the structured separation of the images. Finally, a hybrid model was created that combined:

- YOLOv5s as a pretrained feature extractor.
- A fully connected Siamese network to embed features into a lower-dimensional space.
- A triplet loss function enforcing sensible distance restrictions between positive, negative, and anchor samples.

The following subsections explain in detail the technical architecture and procedures used throughout the process.

Data collection and preprocessing: A custom dataset consisting of 450 photos of Deoni cattle was developed for this research, mostly with an eye on the head section of the calf since it is a significant characteristic for identification. These images were captured across various locations and under different lighting conditions in order to ensure variability. To supplement the dataset, an additional 450 non-Deoni images were sourced from the publicly available Kaggle Cow and Buffalo/ Cattle Dataset, which contains images of cattle and buffalo captured under diverse field conditions. For this study, only images belonging to non-Deoni cattle and visually distinct non-target animals were selected, while irrelevant or poor-quality samples (blurred, heavily occluded, or low-resolution images) were excluded. The selected images included variations in pose, illumination, background, and viewing angle, thereby introducing heterogeneity into the negative class and improving the model's ability to distinguish Deoni cattle from visually similar and dissimilar animals. The images underwent preprocessing with the following transformations to enhance generalization and robustness:

- They were resized to dimensions of 224 × 224 pixels.
- Augmentation techniques such as *Random Horizontal Flip*, *Affine Transformation*, and *Color Jittering* were applied.
- Normalization was performed using ImageNet statistics, with a mean of [0.485*,* 0.456*,* 0.406] and a standard deviation of [0.229*,* 0.224*,* 0.225].

The dataset was divided into 80% for training and 20% for testing. For scalability purposes, DataLoader instances were established with a batch size of 32 for both the training and evaluation phases. Although the Deoni dataset size (450 images) is relatively limited compared to large-scale computer vision datasets, this limitation is partly mitigated by transfer learning through a pretrained YOLOv5 backbone, extensive data augmentation, and metric-learning-based training using triplet loss. These techniques reduced the amount of labeled data required by enabling the model to learn robust feature representations from limited samples. Similar deep metric learning approaches have shown effective performance even under small to medium-sized datasets in animal biometric applications.

Due to limited data availability, a separate validation set was not maintained. Instead, model convergence was monitored using training loss trends and repeated experimental runs. This limitation was acknowledged and future work need to incorporate dedicated validation sets or k-fold cross-validation for stronger statistical evaluation.

YOLOv5 feature extraction: Convolutional neural networks (CNNs) as ResNet, VGG, or DenseNet are often employed in conventional computer vision pipelines to extract hierarchical picture features for classification or embedding applications. However, the spatial localisation cues that are essential for differentiating region-specific distinctions among fine-grained categories, like cow breeds, are absent from these networks, which are usually built for whole-image classification. To address this short- coming, YOLOv5s, a real-time object detection model was incorporated into the framework as a feature extractor. YOLO represents a group of single-stage object detectors recognized for their rapidity, spatial awareness, and rich feature representation. In particular, YOLOv5s (the lighter version) was chosenfor its optimal combination of inference speed and feature detail. In contrast to ResNet, which produces one feature vector for every image, YOLO evaluates an image in multiple stages of spatial convolution and attention, preserving local features like ear patterns, forehead shape, and spot distribution factors that are essential for recognizing Deoni cattle. Each image (batch of shape [B, 3, 224, 224]) was processed through the YOLOv5 backbone to produce a flattened high-dimensional feature vector. This resulted in a feature vector of size [B, 86436] per image batch, where 86,436 represents the number of flattened neurons from YOLO's feature maps. While the feature vector size of 86,436 may appear large, such high-dimensional embeddings are rich in spatial and semantic content. This is particularly valuable for similarity learning, where even subtle inter-class differences must be preserved and that is what our dataset demands.

Siamese network design: Siamese Neural Network was developed and applied to enable strong identification of Deoni cattle by learnt similarity instead of explicit class-based categorization. The Siamese network enabled the system to generalize unseen data and perform pairwise comparison, which is more suitable for tasks involving fine-grained visual similarity. The high dimensional feature vectors created by the YOLOv5 cannot be used directly for similarity comparison so the Siamese network mapped these feature vectors into a lower-dimensional space, where the similar images (of the same cattle breed) are projected close together, while the dissimilar (different breeds) ones are projected farther apart.

Architecture overview: Comprising a sequence of fully connected (dense) layers each followed by a nonlinear activation function (ReLU), the Siamese network design generated an embedding of dimension 64. While enforcing generalization and regularisation, this incremental dimensionality reduction maintained the discriminative qualities. The architecture was defined as follows:

Input Layer: Receives flattened YOLOv5 features (dimension = 86,436)

Hidden Layer 1:
Fully Connected: 86,436 → 1024
Activation: ReLU

Hidden Layer 2:
Fully Connected: 1024 → 256
Activation: ReLU

Hidden Layer 3:
Fully Connected: 256 → 128
Activation: ReLU

Output Layer:
Fully Connected: 128 → 64

No activation function (used directly for distance computations)

The output of the Siamese network was a 64-dimensional vector for each image.

This embedding space was trained such that:

$$D(f(x_{\text{anchor}}), f(x_{\text{positive}})) \rightarrow \text{minimized}$$

$$D(f(x_{\text{anchor}}), f(x_{\text{negative}})) \rightarrow \text{maximized}$$

where $f(x)$ was the 64-dimensional embedding of input image x, and $D(\cdot)$ denoted the Euclidean distance.

Metric learning via triplet loss: Learning an embedding space where semantically related samples are near together but dissimilar samples are further away being the fundamental concept underlying metric learning (Dong *et al.* 2021). Classification occurs in this embedding space through proximity; that is, the similarity of samples is determined by the distance between them. Triplet Loss is especially effective for this purpose, as it can enforce relative distance constraints, making it suitable for scenarios where class boundaries are nuanced and inter-class variability is low—as frequently observed in the dataset of deoni images.

Each training example was represented as a triplet: Anchor (A): The reference image. Positive (P): An image of the same class as the anchor. Negative (N): An image of a different class. Let $f(x)$ be the output embedding of the Siamese Network for input image x. The loss for a single triplet was given by:

The triplet loss function was defined as:

$$L_{\text{triplet}} = \max 0, \|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha$$

Where:

$\|\cdot\|_2$: Euclidean distance (L2 norm)

α: Margin hyperparameter (typically set to 1.0)

max(0, ·): Ensures only violating triplets contribute to the loss

This formulation ensured that the distance between anchor and positive embeddings is at least α units less than the distance between anchor and negative embeddings.

Intuition behind the loss: If $D(A, P) + \alpha < D(A, N)$: The triplet is well-separated, and the loss becomes 0.

If $D(A, P) + \alpha \geq D(A, N)$: The embeddings are too close, and the loss increases proportionally.

This forced the network to learn embeddings where intra-class distances were minimized and inter-class distances were maximized, with at least a safety margin. In our implementation, we used online random mining to dynamically generate triplets from each training batch. The procedure was:

1. For each anchor image in the batch:
2. Positive sample was randomly selected from the same class (excluding the anchor itself).
3. Negative sample was randomly selected from a different class.

Training Configuration: Given the architecture of the model where the YOLOv5 was used for creating feature vectors and the Siamese network was responsible for learning the embeddings, the training process focuses exclusively on optimizing the Siamese network weights using the triplet loss function (Khosla *et al.* 2020). The YOLOv5 model remained frozen throughout training to retain the benefits of its pretrained weights and avoid unnecessary gradient computation.

- The training proceeded as follows for each epoch:
- A mini-batch (32 images) of image-label pairs was sampled from the training data loader.
- The images were passed through the YOLOv5 feature extractor to obtain high-dimensional feature vectors.
- These features, along with labels, were used to dynamically construct triplets: anchor, positive, and negative.
- Each image in the triplet was independently forwarded through the Siamese network to produce 64-dimensional embedding.
- The triplet loss was calculated using Euclidean distances between the result- ing embeddings.
- The loss was back propagated and the Siamese network weights were updated using the Adam optimizer.

Key Training Parameters:

- Optimizer: Adam
- Learning Rate: 0.001
- Beta Values: β1 = 0.9, β2 = 0.999
- Weight Decay (L2 Regularization): 1 × 10−4
- Batch Size and Epochs:

Batch Size: 32

Each batch allowed multiple anchor samples to form distinct triplets with positives and negatives, enriching the training signal.

Epochs: 30

Convergence in the triplet loss curve was observed around epoch 20–25, with minimal improvements beyond 30.

The average triplet loss was logged at each epoch. The goal during training was to minimize this loss, which implied that embeddings of the same class were being brought closer and dissimilar ones pushed apart. A steady decline in loss values was taken as a sign of embedding space convergence.

Testing procedure and similarity-based classification: Our testing procedure leveraged a similarity-based approach, which performed classification by comparing the embedding of a test image with those of labeled training samples in the embedding space learned through triplet loss. Before testing, all samples in the training dataset were used to create a reference embedding database. Each image in the training set was:

- Passed through the YOLOv5 feature extractor to obtain a high-dimensional feature vector. Forwarded through the trained Siamese network to obtain a 64-dimensional embedding.
- Paired with its corresponding class label (Deoni = 0, Non-Deoni = 1).
- All embedding was computed and stored in memory before test-time inference for efficient distance computation.

Test-Time Inference

For each image in the test set, the following steps were executed:

YOLO Feature extraction: The test image was passed through the YOLOv5s model to obtain a feature map, which was then flattened into a high-dimensional vector.

Siamese Embedding Projection: This vector was input to the trained Siamese network to compute its 64-dimensional embedding $f(x_{\text{test}})$.

Distance calculation: The Euclidean distance between the test embedding and all reference embed- dings was calculated as:

$$d_i = \| f(x_{\text{test}}) - f(x_i) \| 2 , \forall (f(x_i), y_i) \in \text{R}$$

where R is the reference embedding set.

Nearest neighbor classification: The label y_k of the closest reference embedding (i.e., the one with the minimum distance) was assigned as the predicted class:

$$y\hat{} = y\text{arg min}i\ di$$

Generalization to unseen data: To evaluate model robustness, the model was tested on a domain-shifted dataset composed of images from a different source (animal dataset), which included images of animals from other breeds and species.

The same similarity-based strategy was applied, using the original reference embeddings from the training set. Performance on this external dataset was used to assess the model's:

- Transferability to new domains,
- Robustness to image noise and variations,
- Ability to distinguish Deoni cattle from non-Deoni and unrelated animal classes.

RESULTS AND DISCUSSIONS

The model was evaluated on multiple metrics such as accuracy, precision, recall, F1-score and also False rejection rate i.e. FRR. Extensive cross validation experiments were conducted to ensure robustness and generalizability of the model.

Cross-Validation Setup

To evaluate the model's performance, three distinct test datasets were used:

- Dataset A: Contains 450 images of Deoni cattle.
- Dataset B: Contains 450 images of other cattle breeds.
- Dataset C: Contains 450 images of other animal breeds.

The model was tested separately on these datasets and the performance was recorded.

Accuracy analysis: The use of triplet loss along with fine-tuning parameters and additional architectural and preprocessing enhancements significantly improved model performance compared to contrastive loss.

Table 1. Accuracy comparison between different loss functions

Loss Function Used	Accuracy Range
Contrastive Loss	55% – 60%
Triplet Loss (with other enhancements)	94% – 96.5%

Evaluation Metrics: Evaluation metrics of precision and recall were calculated using the following formulas

$$\text{Precision}=\text{TP}/(\text{TP}+\text{FP})$$

$$\text{Recall}=\text{TP}/(\text{TP}+\text{FN})$$

where:

TP (True Positive): Deoni cattle was identified correctly.

FP (False Positive): Irrelevant identification the model wrongly identified non-Deoni cattle as Deoni.

FN (False Negative): The model failed to identify Deoni cattle even though it exists in the provided data.

Table 2. Classification performance for Deoni and Non-Deoni classes

Class	Precision	Recall	F1-score
Deoni	0.96	0.96	0.96
Non-Deoni	0.97	0.97	0.97

Confusion Matrix and FAR Evaluation: FAR: The rate at which negative class instances were incorrectly accepted as positive class instances

$$\text{FAR}=\text{FP}/(\text{FP}+\text{TN})$$

FAR: The rate at which positive class instances were incorrectly accepted as negative class instances

$$FRR=FN/(FN+TP)$$

Test on Deoni-Only Dataset: Confusion Matrix

Table 3: Confusion matrix for Deoni-only dataset

True Class	Predicted Non-Deoni	Predicted Deoni
Deoni	3	447

False Rejection Rate (FRR)=3/(3+447)=0.006

Test on Mixed Cattle Dataset: Confusion Matrix

Table 4: Confusion matrix for mixed cattle dataset

True Class	Predicted Non-Deoni	Predicted Deoni
Non-Deoni	434	16

False Rejection Rate (FRR)=16/(16+434)=0.0356

*Test on other animals breeds Dataset:*Confusion Matrix

Table 5. Confusion matrix for mixed cattle dataset

True Class	Predicted Non-Deoni	Predicted Deoni
Non-Deoni	414	16

False Rejection Rate (FRR)=16/(16+414)=0.08

The YOLOv5-Siamese architecture achieved an accuracy ranging from 94.0% to 96.5%, which was comparable to or slightly better than several recent deep learning-based cattle identification studies. Unlike conventional CNN-based classifiers that rely primarily on global image features, the proposed framework combined object-level feature extraction with metric learning, enabling more effective discrimination of visually similar cattle breeds. Furthermore, the use of triplet loss facilitated the formation of a structured embedding space, improving inter-class separation while reducing intra-class variation (Table 6).

This showed that the model exhibited a low FAR, maintaining strong discrimination ability between Deoni

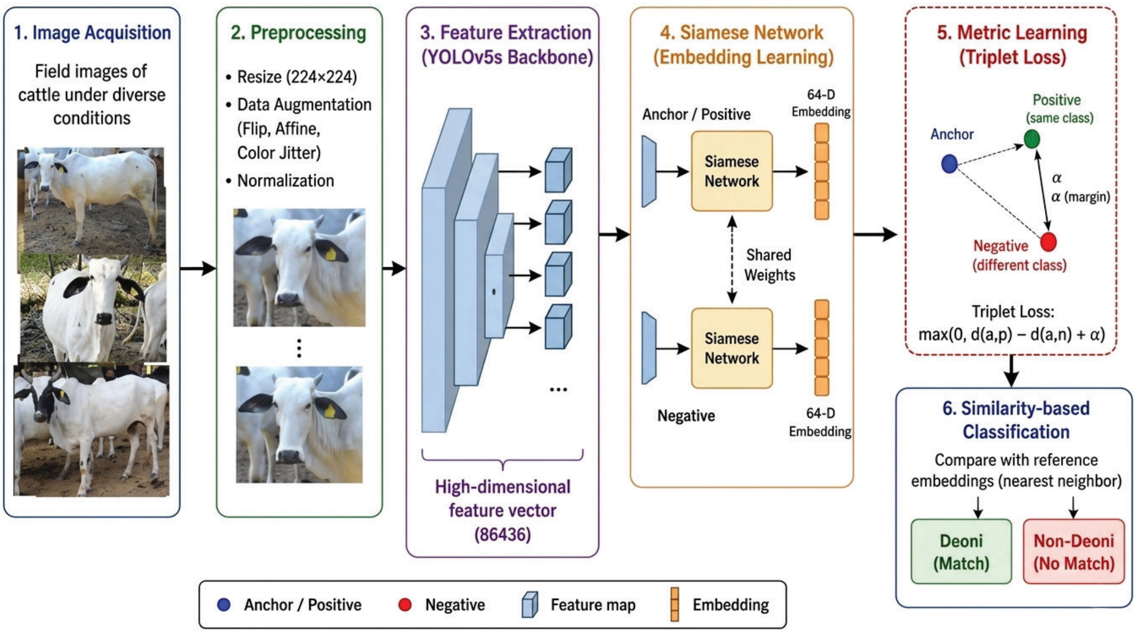

Fig. 2. Architecture pipeline of the YONOv5 Siamese framework

Table 6. Comparison of the proposed method with previous cattle identification studies

Study	Method	Dataset/Target	Accuracy (%)	Key Findings
Weber de Lima *et al.* (2020)	CNN-based classification	Pantaneira cattle breed	93.2	Demonstrated feasibility of breed recognition using deep learning and computer vision.
Gupta *et al.* (2022)	Computer vision-based breed classification	Dairy cattle breeds	92.0–95.0	Effective breed discrimination under controlled image conditions.
Li *et al.* (2024)	Multi-feature decision fusion	Cattle identification	95.4	Combined multiple facial and morphological traits for improved identification accuracy.
Sharma *et al.* (2024)	Deep metric learning with depth images	Universal bovine identification	94.8	Demonstrated robust animal identification using embedding-based representations.
Present Study	YOLOv5 + Siamese Network + Triplet Loss	Deoni cattle identification	94.0–96.5	Accurate identification of Deoni cattle under varying field conditions using similarity-based metric learning.

and non-Deoni cattle, even when tested against diverse breeds. Triplet loss enforced relative distance constraints among anchor, positive, and negative samples, leading to a more structured and discriminative embedding space. This was reflected in the improved precision, recall, and F1-scores for both Deoni and non-Deoni classes. Similar advantages of triplet loss have been reported in animal biometric and facial recognition studies (Porto *et al.* 2024; Dong and Shen, 2021), supporting its suitability for livestock identification under limited and heterogeneous datasets. The model's low false acceptance and false rejection rates, even when tested against other cattle breeds and unrelated animal species, indicated strong generalization capability and robustness to domain shift. This is particularly important for practical deployment, where variations in lighting, pose, background, and partial occlusion are unavoidable. In contrast, several previous studies relied on curated or web-sourced datasets captured under ideal conditions, which limits their real-world applicability (Gupta *et al.* 2022). Similarly, recent research highlighted advanced deep metric learning approaches for cattle identification, demonstrating that embedding-based methods can effectively distinguish individual animals without relying solely on visible coat patterns or physical markings (Sharma *et al.* 2024). In addition, multi-feature identification frameworks that combine facial characteristics, muzzle patterns, and ear features have been shown to further improve identification accuracy in complex farm environments. These findings aligned with the present study, which emphasized localized feature extraction and similarity-based classification for robust breed recognition (Li *et al.* 2024). Overall, the results indicated that the integration of object detection models with metric learning frameworks provides a powerful approach for automated livestock identification. The developed framework offers a reliable, non-invasive, and scalable method for Deoni cattle identification, contributing to digital livestock management, breed documentation, and conservation of indigenous cattle genetic resources. Furthermore, the approach may support emerging applications in precision livestock farming, including automated herd monitoring, traceability systems, and data-driven breeding programs.

To further investigate the discriminative ability of the proposed framework, the learned 64-dimensional embeddings were visualized using t-distributed Stochastic Neighbor Embedding (t-SNE). As shown in Figure 3, images belonging to Deoni cattle formed a compact cluster that was clearly separated from non-Deoni cattle and other animal species. The observed clustering pattern confirmed that the Siamese network successfully learned meaningful feature representations and that triplet loss effectively enhanced class separability within the embedding space. These findings supported the quantitative performance metrics and provided visual evidence of the model's capability for breed identification (Fig.2).

Despite promising results, the study has certain limitations. First, the dataset size remains modest and

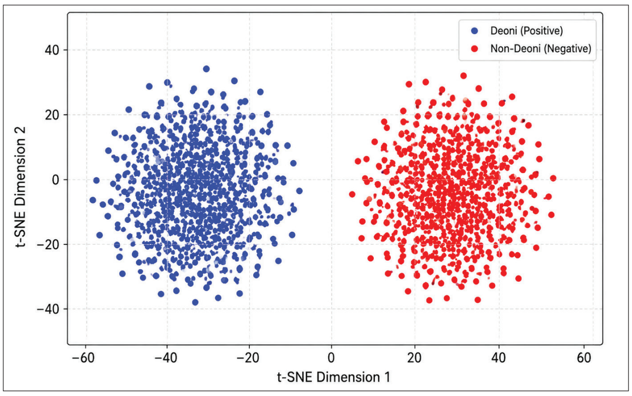

Fig. 3. t-SNE visualization of the 64-dimensional embeddings generated by the YOLOv5-Siamese

primarily focused on Deoni cattle from limited geographical locations. Second, the external non-Deoni dataset included publicly available Kaggle images that may not fully represent real-world field variability. Third, absence of k-fold cross-validation limits statistical confidence in generalization. Future work will address these limitations through larger multi-institutional datasets, cross-validation, and real-time deployment benchmarking.

The present study demonstrated the effectiveness of a deep learning–based hybrid framework for automated identification of Deoni cattle using facial characteristics. The integration of YOLOv5 for feature extraction with a Siamese network trained using triplet loss effectively overcomes the limitations of traditional visual inspection and conventional classifiers. The proposed model achieved high accuracy (94–96.5%) along with strong precision, recall, and F1-scores for both Deoni and non-Deoni classes. Triplet loss significantly outperformed contrastive loss by learning a well-structured embedding space that enhances inter-class separation while minimizing intra-class variation. The system exhibited low false acceptance and rejection rates, even under domain-shifted testing conditions involving diverse breeds and species. The proposed framework demonstrates promising performance as a non-invasive approach for Deoni cattle identification. However, further validation using larger and more diverse multi-location datasets is necessary before establishing large-scale deployment and broader generalization. Overall, the methodology provides a reliable, non-invasive, and scalable solution for indigenous cattle breed identification and supports digital livestock management and conservation efforts.

ACKNOWLEDGMENT

We acknowledge funding provided by Rajiv Gandhi Science and Technology Commission (RGSTC), Government of Maharashtra (India) vide order No. Sanction order: Letter No: RGSTC/File-2021/DPP-256/CR-85/149; Dt. 05/04/2023.

REFERENCES

Dong X and Shen J. 2021. Triplet loss in Siamese network for object tracking. Beijing Lab of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology; Inception Institute of Artificial Intelligence. Pp. 101–105.

Dong X and Shen J. 2021. Triplet loss in Siamese network for object tracking. Beijing Lab of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, China and Inception Institute of Artificial Intelligence, Abu Dhabi, UAE. Pp. 256–265.

Dongre VB, Gandhi RS, Salunke VM, Kokate LS, Durge SM, Khandait VN and Patil PV. 2017. Present status and future prospects of Deoni cattle. *Indian Journal of Animal Sciences* **87**(7): 800–803.

Gupta H, Jindal P, Verma OP, Arya RK, Ateya AA, Soliman NF and Mohan V. 2022. Computer vision-based approach for automatic detection of dairy cow breed. *Electronics* **11**(22): 3791.

Jogi R, Temburnikar G, Jadhav A, Biradar A, Gajbhiv S and Malge A. 2024. Cattle breed classification techniques: Framework and algorithm evaluation. *Journal of Propulsion Technology* **45**(1): 1739–1745.

Khosla P, Teterwak P, Wang C, Sarna A, Tian Y, Isola P, Maschinot A, Liu C and Krishnan D. 2020. Supervised contrastive learning. *Advances in Neural Information Processing Systems* **33**: 18661–18673.

Kumar S, Singh SK. 2020. Cattle recognition: A new frontier in visual animal biometrics research. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences* **90**(4): 689–708.

Li D, Li B, Li Q, Wang Y, Yang M and Han M. 2024. Cattle identification based on multiple feature decision layer fusion. *Scientific Reports* **14**: 26631.

Sharma A, Randewich L, Andrew W, Hannuna SL, Campbell NW, Mullan S, Dowsey AW, Smith M, Hansen M and Burghardt T. 2024. Universal bovine identification via depth data and deep metric learning. arXiv preprint abs/2404.00172.

Weber de Lima F, Magalhães LP, Pereira DRA, Monteiro PAJ, Pereira LAA, Oliveira LFS and Salgado MCA. 2020. Recognition of Pantaneira cattle breed using computer vision and convolutional neural networks. *Computers and Electronics in Agriculture* **175**: 105548.

Zhao C, Wang H, Zhang Y, Zhao Y, Gao J. 2023. Breed identification using breed-informative SNPs and machine learning based on whole genome sequence data and SNP chip data. *Journal of Animal Science and Biotechnology* **14**(85): 1–13.