

Cattle Breed Identification Using Deep Learning: A Comparative Study of ResNet50, EfficientNet-B0 and Vision Transformer

Dr. Sheik Imran¹, Namratha V², Mahima Danappa Kasambi³, Nagashree C N⁴,
Meghana Venkatesh Telkar⁵

¹Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India
Corresponding Author Email: [sheikimran81\[at\]gmail.com](mailto:sheikimran81[at]gmail.com)

²Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India
Email: [namratha48v\[at\]gmail.com](mailto:namratha48v[at]gmail.com)

³Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India
Email: [nagashreecn1611\[at\]gmail.com](mailto:nagashreecn1611[at]gmail.com)

⁴Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India
Email: [861mahima\[at\]gmail.com](mailto:861mahima[at]gmail.com)

⁵Department of Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India
Email: [meghanavtelkar\[at\]gmail.com](mailto:meghanavtelkar[at]gmail.com)

Abstract: *Accurate cattle breed identification is important for livestock management, breeding programs, genetic conservation, and agricultural applications. Manual identification based on physical characteristics can be time-consuming and requires expert knowledge. This study presents a deep learning-based approach for automated cattle breed identification using images of 41 cattle breeds. A dataset containing 5,946 images was prepared and divided into training, validation, and testing sets. Images were resized to 224 × 224 pixels and appropriate preprocessing and augmentation techniques were applied. Three pretrained deep learning architectures, ResNet50, EfficientNet-B0, and Vision Transformer (ViT), were trained using transfer learning and evaluated on the test dataset. The obtained test accuracies were 87.20% for ResNet50, 90.55% for EfficientNet-B0, and 94.87% for Vision Transformer. Among the evaluated models, ViT achieved the highest classification accuracy. A Gradio-based graphical user interface was also developed to allow users to upload cattle images and obtain breed predictions. The results demonstrate the potential of deep learning-based approaches for automated cattle breed identification while highlighting the challenges associated with fine-grained classification across visually similar breeds.*

Keywords: Cattle Breed Identification, Deep Learning, Image Classification, Vision Transformer, Transfer Learning

1. Introduction

India maintains one of the world's most genetically diverse populations of indigenous cattle with over 40 recognized cattle breeds and more than a dozen buffalo breeds, each adapted to distinct agro-climatic regions and valued for traits such as milk yield, draught capability, and disease resistance. Reliable breed identification underpins genetic conservation efforts, targeted breeding strategies, government subsidy and insurance verification, and dairy supply-chain traceability. In practice, breed assessment is typically performed by veterinarians or livestock experts who visually evaluate morphological cues- horn shape, coat color and pattern, hump size, and ear structure- a process that is slow, requires scarce expertise, and does not scale to India's large and geographically dispersed cattle population.

Deep learning has substantially advanced fine-grained visual classification in adjacent domains such as bird and plant species identification, and a growing body of work has begun applying convolutional and, more recently, transformer-based architectures to livestock identification and breed classification tasks. However, as this review shows, most existing studies target a small number of breeds (often four to ten classes), non-Indian breed sets, or related tasks such as individual animal identification and disease detection rather

than large-scale, fine-grained breed classification across dozens of visually similar indigenous breeds.

This paper makes two contributions. First, we synthesize and compare reported results from recent literature applying ResNet, EfficientNet, and Vision Transformer architectures to cattle, buffalo, and related livestock classification tasks, highlighting trends in architecture choice, dataset scale, and reported accuracy. Second, we introduce a curated 41-class Indian cattle and buffalo breed image dataset- substantially larger in class count than prior work in this specific domain- and propose it as a benchmark to support future comparative and reproducible research, alongside a detailed experimental protocol intended to guide such studies.

2. Literature Survey

A. CNN-based approaches

Residual networks (ResNet) [1] remain a widely used backbone for transfer learning in agricultural image classification due to their strong ImageNet-pretrained features and stable training dynamics. A study applying ResNet-50 with hyperparameter-optimized transfer learning to a 5-class cattle breed dataset (Ayrshire, Brown Swiss, Holstein Friesian, Jersey, and Red Dane; 1,251 images) reported a 14-percentage-point improvement using transfer

learning (95% accuracy) over training from scratch (81% accuracy), underscoring the importance of pretrained initialization for cattle breed tasks with limited data [6]. In contrast, a study using ResNet50 and EfficientNetB3 for a harder, more imbalanced 4-class species-level task involving buffalo alongside visually distinct wildlife (elephant, rhino, zebra) reported substantially lower validation accuracy (47% and 42%, respectively), with particularly poor buffalo recall, illustrating that reported CNN performance is highly sensitive to class balance and task framing rather than architecture alone [7].

B. Efficientnet-Based Approaches

EfficientNet architectures [2], which use compound scaling of depth, width, and resolution, have been applied specifically to cattle and buffalo breed classification. A study using EfficientNetV2-S on a 10-class cattle-and-buffalo dataset (2,427 images), combined with focal loss, class re-weighting, and test-time augmentation to address data scarcity and imbalance, reported 90.68% test accuracy with macro and weighted F1-scores of 0.91 [4]. Another study targeting phenotypically similar tropical *Bos indicus* breeds compared EfficientNet- and MobileNet-based backbones, finding EfficientNet variants achieved 84–86% accuracy with F1-scores above 83%, outperforming MobileNet variants (80–82% accuracy)- evidence that backbone choice measurably affects performance even when the classification task (visually similar breeds) is held constant [5].

C. Vision Transformer-Based Approaches

The Vision Transformer (ViT) [3] has been increasingly applied to livestock recognition tasks, frequently reporting strong results relative to CNN baselines. SheepFormers, a ViT-based framework evaluated across several transformer variants (ViT-Google, ViT-MAE, ViT-VAN, BEiT, ViT-ResNet50, DiT) for 4-class sheep breed classification on a small dataset (1,680 images), found the ViT-Google variant achieved the best performance at 98.21% validation accuracy, outperforming both CNN architectures and ensemble

baselines reported in prior work, despite the small dataset size typical of livestock imaging studies [9]. Similarly, ViT-Sheep, incorporating a LayerScale module into a ViT-B/16 encoder for individual sheep face recognition, achieved 97.9% recognition accuracy on a 160-sheep dataset [11]. In a directly cattle-related application, a ViT-based model for lumpy skin disease detection in cattle achieved 98.12% accuracy, 98.5% precision, and 98.5% recall, exceeding traditional CNN approaches on the same task and images [12]. More broadly, a general-purpose ViT model evaluated on multi-species animal image classification achieved 89% accuracy [10], and independent comparative work on a 90-class general animal dataset found ViT achieving 97% test accuracy within just seven training epochs, versus a plateauing 86% for ResNet with signs of overfitting at higher epoch counts [15]. Collectively, these results suggest ViT-based architectures are competitive with, and in several livestock-specific studies superior to, CNN baselines — though sample sizes in this literature remain small and standardized comparisons on large, fine-grained bovine breed sets are lacking.

D. Related Detection and Identification Approaches

Beyond breed classification, deep learning has been applied to adjacent cattle-related tasks. YOLO-family detectors have been used for cattle breed detection directly from unconstrained imagery, with YOLOv4 reported to achieve 92.85% accuracy in one study [8]. Attention-augmented ResNet variants have also been applied to individual cattle facial recognition (SK-ResNet, 98.42% accuracy on 5,677 images) and muzzle-pattern-based identification (YOLOv3 with ResNet50, 99.11% accuracy) [13], reflecting a broader trend of combining CNN backbones with attention mechanisms for fine-grained livestock recognition tasks.

E. Summary of Reported Literature Results

Table I summarizes architecture choices, task framing, dataset scale, and reported outcomes across the studies discussed above.

Table I: Summary of architectures, tasks, and reported results in recent cattle/buffalo/livestock classification literature

Study	Architecture	Task	Dataset Size	Reported Result
Priyanka et al. [4]	EfficientNetV2-S	10-class cattle/buffalo breed classification	2,427 images	90.68% test acc.
Warhade et al. [5]	EfficientNet (various)	Phenotypically similar <i>Bos indicus</i> breed classification	Not specified	84–86% acc.
JUSIFOR study [6]	ResNet50	5-class cattle breed classification	1,251 images	95% (transfer learning) vs. 81% (scratch)
Naik & Ramu [7]	ResNet50 / EfficientNetB3	Species-level buffalo vs. wildlife classification	4-class, balanced	47% / 42% val. acc.
Yilmaz et al. [8]	YOLOv4	Cattle breed detection	Not specified	92.85% acc.
SheepFormers [9]	ViT-Google (ViT-B variant)	4-class sheep breed classification	1,680 images	98.21% val. acc.
Animal ViT study [10]	ViT (base)	General multi-species animal classification	Not specified	89% acc.
ViT-Sheep [11]	ViT-B/16 + LayerScale	Individual sheep face recognition	160 sheep	97.9% acc.
LSD detection [12]	ViT vs. CNN baselines	Cattle disease (lumpy skin disease) detection	8,000 images	98.12% acc. (ViT)

Two patterns emerge from Table I. First, reported accuracy is strongly influenced by class count and inter-class visual similarity rather than architecture family alone- tasks with fewer, more visually distinct classes tend to report higher accuracy regardless of whether a CNN or transformer backbone is used. Second, no identified study benchmarks ResNet, EfficientNet, and ViT architectures together on a single large-scale, fine-grained Indian cattle and buffalo breed

dataset, despite India's exceptionally high bovine breed diversity- indicating a clear gap that a standardized benchmark dataset could address.

3. Methodology

- 1) Dataset Collection:** A publicly available cattle image dataset containing 5,946 images from 41 cattle breeds

was used for the study. The dataset includes breeds such as Alambadi, Amritmahal, Gir, Hallikar, Khillari, Sahiwal, Vechur, and other Indian and commonly reared cattle breeds. The images were organized into separate classes according to their breed names.

- 2) **Dataset Splitting:** The dataset was divided into training, validation, and testing sets. The training data was used to train the models, the validation data was used to monitor model performance during training, and the test data was used for final evaluation. The final training/validation division was maintained at 85:15, while the overall dataset followed an 80:20 training/testing split.
- 3) **Image Preprocessing:** Before model training, the cattle images were preprocessed to provide a consistent input format. All images were resized to 224×224 pixels and converted to JPG format where required. Image normalization was also applied using the preprocessing convention associated with the pretrained models. These steps ensure that the input images have a uniform size and suitable numerical representation for deep learning models.
- 4) **Data Augmentation:** Data augmentation was applied to the training images to improve model generalization and reduce the effect of limited samples and class imbalance. Augmentation techniques were used to introduce variations in the training images while preserving the important visual characteristics required for breed identification.
- 5) **Transfer Learning:** Transfer learning was used to reduce the training requirements and utilize features learned from large-scale image datasets. Three pretrained architectures were selected: ResNet50, EfficientNet-B0,

and Vision Transformer (ViT-B/16). The final classification layer of each model was modified to produce predictions for the 41 cattle breed classes.

- 6) **Model Architecture:** Three deep learning architectures were implemented and compared:
 - **ResNet50:** A convolutional neural network using residual connections for effective deep feature learning.
 - **EfficientNet-B0:** A CNN architecture designed to achieve efficient feature extraction with a balanced network structure.
 - **Vision Transformer (ViT):** A transformer-based architecture that processes image patches and learns relationships between different regions of an image.
- 7) **Model Training and Evaluation:** Each model was trained using the prepared training dataset and evaluated using the validation dataset during training. The models were trained for 15 epochs with a learning rate of 1×10^{-4} . The best-performing model checkpoint was saved based on validation performance. After training, the models were evaluated on the 1,202-image test set. Test accuracy and classification metrics were used to compare the performance of the three architectures. The final test accuracies obtained were 87.20% for ResNet50, 90.55% for EfficientNet-B0, and 94.87% for ViT. This methodology provides a consistent experimental pipeline for comparing CNN-based and transformer-based approaches for 41-class cattle breed identification.

4. System Architecture

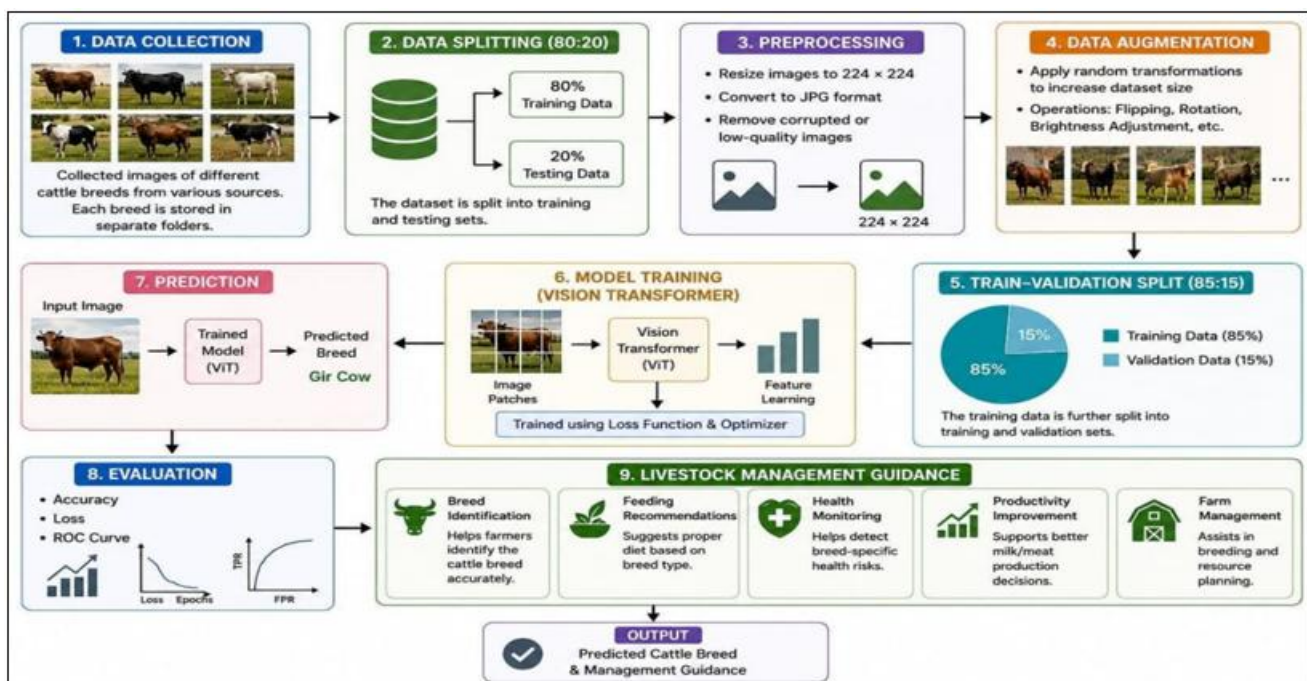


Figure 1: System architecture of the proposed deep learning-based cattle breed identification system

The proposed system architecture illustrates the workflow of the cattle breed identification system using deep learning. The system consists of data collection, data splitting, image preprocessing, data augmentation, training and validation, model training, breed prediction, model evaluation, and livestock management guidance.

Initially, images of 41 cattle breeds are collected from the dataset and organized into separate breed-specific folders. The dataset is divided into training and testing sets using an 80:20 ratio. During preprocessing, the images are resized to 224×224 pixels and converted into JPG format wherever

required. Image normalization is also performed to ensure consistency in the input data. Data augmentation techniques, such as flipping, rotation, and brightness adjustment, are applied to increase image diversity and improve the model's ability to generalize. The training data is further divided into training and validation sets using an 85:15 ratio. Three pretrained deep learning architectures, namely ResNet50, EfficientNet-B0, and Vision Transformer (ViT), are trained using transfer learning and evaluated for cattle breed classification.

Their performance is compared using evaluation metrics such as accuracy, loss, and ROC curves. Based on the experimental results, Vision Transformer achieves the highest test accuracy of 94.87% and is selected for the final prediction interface. The trained model is integrated into a Gradio-based graphical user interface, where users can upload a cattle image and obtain the predicted breed, confidence score, and top-three predictions. The system also provides breed-specific livestock management guidance, including feeding recommendations, health monitoring, productivity improvement, and farm management information. Thus, the proposed architecture combines automated cattle breed identification with supporting information to assist livestock management.

5. Results and Discussion

Performance Analysis of the Model

Model	Accuracy	Precision	Recall	F1-Score
ResNet50	87.20%	0.872	0.870	0.868
EfficientNet-B0	90.55%	0.905	0.900	0.899
Vision Transformer	94.87%	0.948	0.940	0.939

The experimental results show that the Vision Transformer achieved the highest test accuracy among the three evaluated architectures, with an accuracy of 94.87%. EfficientNet-B0 achieved 90.55%, while ResNet50 achieved 87.20%. The results indicate that the evaluated models face challenges in distinguishing visually similar cattle breeds across the 41-class dataset.

Discussion

The experimental comparison indicates that transfer learning is effective for the considered cattle breed identification task. All three models were initialized using pretrained weights, allowing them to utilize previously learned visual representations and adapt them to the cattle breed classification problem.

The performance ranking observed in the experiment was ViT-B/16, followed by EfficientNet-B0 and ResNet50. The higher performance of ViT suggests that modelling long-range relationships between image regions can be beneficial for fine-grained cattle breed identification. EfficientNet-B0 also demonstrated strong performance, while ResNet50 provided a competitive baseline.

However, the results should be interpreted in the context of the dataset and experimental configuration used in this study. Performance on a particular dataset does not necessarily guarantee equivalent performance under field conditions. Differences in camera devices, geographical regions, lighting,

animal pose, and breed representation may affect classification accuracy.

Overall, the comparison demonstrates the potential of transfer-learning-based architectures for automated cattle breed identification. The results also indicate that transformer-based image classification can provide a useful alternative to conventional convolutional architectures for large multi-class livestock image classification problems.

6. Conclusion

This study developed and evaluated a deep learning-based cattle breed identification system covering 41 cattle breeds and 5,946 images. Three pretrained architectures, ResNet50, EfficientNet-B0, and Vision Transformer, were comparatively evaluated using transfer learning. The obtained test accuracies were 87.20%, 90.55%, and 94.87%, respectively. Vision Transformer achieved the highest accuracy among the evaluated models. A Gradio-based interface was also developed to provide image-based breed prediction in a user-friendly manner. The results demonstrate the feasibility of automated cattle breed identification while also highlighting the difficulty of fine-grained classification among visually similar breeds.

7. Future Scope

Future work can focus on increasing the dataset size and improving class balance to enhance the robustness of the classification system. Further improvements can be achieved through hyperparameter fine-tuning, advanced data augmentation techniques, and the use of improved loss functions. Ensemble approaches combining multiple deep learning models can also be explored to improve classification performance. In addition, lightweight model architectures can be developed to support practical field deployment on mobile and resource-constrained devices. These improvements can help make the system more reliable and suitable for real-world cattle breed identification applications.

References

- [1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
- [2] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. Int. Conf. Machine Learning (ICML), 2019, pp. 6105–6114.
- [3] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in Proc. Int. Conf. Learning Representations (ICLR), 2021.
- [4] Priyanka, K. Sardhar, S. P. Ajay, A. R. Chowdary, B. V. V. S. G. B. Raju, Devika S, and M. Abujar, "Deep Learning-Based Breed Classification of Cattle and Buffaloes using EfficientNetV2-S," Int. J. Engineering Research & Technology (IJERT), vol. 15, no. 4, April 2026.

- [5] R. Warhade, I. Devi, N. Singh, et al., "Attention module incorporated transfer learning empowered deep learning-based models for classification of phenotypically similar tropical cattle breeds (*Bos indicus*)," *Tropical Animal Health and Production*, vol. 56, art. 192, 2024.
- [6] "Cattle Breed Classification Using ResNet-50 Hyperparameter Optimization Based on Transfer Learning," *JUSIFOR: Jurnal Sistem Informasi dan Informatika*, 2026.
- [7] N. Naik and S. Ramu, "Precise AI-Driven Cattle Identification and Classification System," preprint.
- [8] A. Yılmaz, G. N. Uzun, M. Z. Gürbüz, and O. Kıvrak, "Detection and breed classification of cattle using YOLO v4 algorithm," in *Proc. Int. Conf. Innovations in Intelligent Systems and Applications (INISTA)*, IEEE, 2021, pp. 1–4.
- [9] "SheepFormers: vision transformer-based sheep breed classification on small datasets for smart farming," *Journal of Electrical Systems and Information Technology*, 2026.
- [10] "Animal Image Classification Using the Vision Transformer Model," *Authorea preprint*, 2024.
- [11] "A high-precision facial recognition method for small-tailed Han sheep based on an optimised Vision Transformer," *ScienceDirect*, 2023.
- [12] "A vision transformer model-integrated mobile application for early and accurate detection of lumpy skin disease in cattle," *Scientific Reports*, 2025.
- [13] "AI in Agriculture: A Survey of Deep Learning Techniques for Crops, Fisheries and Livestock," *arXiv preprint arXiv:2507.22101*, 2025.
- [14] W. Kusakunniran, A. Wiratsudakul, U. Chuachan, S. Kanchanapreechakorn, and T. Imaromkul, "Automatic cattle identification based on fusion of texture features extracted from muzzle images," in *Proc. IEEE Int. Conf. Industrial Technology*, 2018, pp. 1484–1489.
- [15] S. Shabbir, "97% Accuracy with ViT on 90 Animal Dataset: A comparative study Vision Transformers vs. ResNet in Multi-Class Classification," *Medium*, 2023.