International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

Image Based Animal Type Classification - (Cattle & Buffalo)

Dr P Bikshapathy[1], M Rakesh[2], S Manish Chandra[3], K Siddu[4]
[1]Assistant Professor; Dept. Of Computer Science And Engineering Matrusri Engineering College Hyderabad, Telangana, India
[2,3,4]B.Tech Student's; Dept. Of Computer Science And Engineering Matrusri Engineering College Hyderabad, Telangana, India
Mail Id's; rakeshmolgu954@gmail.com, manishsriram1234@gmail.com, 20054ei110@gmail.com

ABSTRACT
Cattle and buffalo breed identification plays a crucial role in agricultural management, livestock breeding programs, and veterinary practices, enabling farmers and researchers to make informed decisions about animal health, productivity, and genetic diversity. However, traditional breed classification methods often rely on manual inspection, which can be subjective, time consuming, and prone to errors, especially with the increasing diversity of breeds across different regions. The advent of computer vision and deep learning has opened new avenues for automated breed identification, but challenges remain in handling variations in lighting, pose, and image quality. This project presents a multi-model ensemble deep learning framework for cattle and buffalo breed classification that integrates spatial features, preprocessing techniques, and confidence scoring mechanisms from image data. The proposed system processes input images by applying data augmentation and analyzing them through an ensemble of convolutional neural networks, including ResNet50, MobileNetV3, and EfficientNet-B3.

Keywords
Cattle & Buffalo Classification, Ensemble Deep Learning, Convolutional Neural Networks, Resnet50, MobileNetV3, EfficientNetB3, Feature Extraction.

INTRODUCTION

The rapid advancement of computer vision and machine learning technologies has significantly enhanced the ability to automate livestock identification and breed classification. Among these developments, automated cattle and buffalo breed recognition systems have gained attention due to their potential to analyze morphological characteristics from images. Using deep learning models such as convolutional neural networks and ensemble architectures, these algorithms can identify breed-specific features that are often challenging for manual inspection. While these technologies have beneficial applications in agriculture, livestock management, and breeding programs, they also address challenges related to subjective assessments, time constraints, and the need for accurate genetic and productivity tracking. The increasing accessibility of image capture devices and computational tools has made it possible for farmers and researchers without advanced technical expertise to collect and analyze livestock images. Such systems can be utilized to improve breeding decisions, monitor animal health, and optimize production metrics. However, the misuse or limitations of automated systems could lead to incorrect breed identification, affecting economic outcomes in agriculture. As a result, the development of reliable and explainable breed classification systems has become a critical area of research in agricultural computer vision and precision livestock farming. Cattle and buffalo images often contain inconsistencies in both spatial and feature domains, including unnatural color variations, abnormal texture patterns caused by lighting conditions, or morphological irregularities due to pose and angle differences. Therefore, analyzing a single representation of the data may not be sufficient to reliably classify breeds across diverse conditions. The fused representation improves classification performance and enhances robustness against variations in image quality and environmental conditions. The framework incorporates advanced preprocessing with transformations such as Random Resized Crop, Color Jitter, and Random Affine, along with GradCAM for visual explainability. Experiments conducted on a dataset containing images of various cattle and buffalo breeds demonstrate that the proposed architecture effectively captures breed-specific features. The model achieved a training accuracy of approximately 89.68%, validation accuracy of 87.1%, and test accuracy of 85.5%, with detailed breed information and confidence scores for each prediction. The system provides comprehensive breed databases and visual explanations, making it suitable for practical deployment in agricultural applications. The primary contribution of this work proposes a multi-model ensemble deep learning framework with confidence-based prediction and explainabiity mechanisms. Unlike conventional single-model approaches, the proposed system combines CNN architectures with probabilistic fusion and out-ofdistribution detection to improve robustness andreliabilty in real world agriculture conditions.

RELATED WORK

Research in livestock identification has evolved from traditional feature-based methods to advanced deep learning techniques. Early approaches focused on handcrafted feature extraction methods such as Speeded-Up Robust Features (SURF), which were used for tasks like muzzle pattern recognition [1]. Although these methods demonstrated initial success, they lacked scalability and robustness when applied to diverse datasets. Later studies improved muzzle-based identification using CNNs [4], [8], but these approaches still faced challenges in generalizing across breeds and environments. Subsequent studies introduced deep learning models for animal identification, leveraging Convolutional Neural Networks (CNNs) to automatically extract discriminative features [2]. These approaches improved classification accuracy and adaptability but were often limited by dataset size, imbalance, and computational constraints. Some works extended CNN architectures with recurrent models, such as BiLSTM, to capture temporal information in video-based tracking systems [7]. More recent research has explored the use of advanced architectures such as ResNet and EfficientNet for livestock classification tasks [16]. These models have demonstrated strong performance due to their ability to learn hierarchical feature representations. Additionally, lightweight models such as MobileNetV3 [3] have been proposed to address computational efficiency, enabling deployment on resource-constrained devices. Studies focusing on specific breeds, such as Pantaneira [6], Sahiwal [14], and buffalo breeds [12], highlight the adaptability of CNN-based frameworks but also emphasize the need for larger, diverse datasets. Parallel to identification, automated body condition scoring has emerged as a critical application of computer vision in livestock management. Early manual scoring studies [5] evolved into automated systems using 2D imaging [13], 3D sensors [10], and real-time detection with YOLOv5 [17]. These systems improved precision and scalability but introduced trade-offs in hardware cost and sensitivity to occlusions.

PROPOSED METHODOLOGY

Objectives

1. Build an automated breed classification system that reliably identifies cattle and buffalo breeds from images using deep learning. With manual inspection being time-consuming and prone to error, the goal is to create a computational model that can recognize breed-specific patterns quickly and consistently across varying visual conditions.

2.Use multiple complementary feature representations rather than depending on a single image signal.

The system combines: Spatial patterns extracted from raw images, Enhanced preprocessing-based feature robustness (brightness, contrast, orientation variations), Explainability signals that highlight what the model "looks at" when making a prediction (e.g., GradCAM heatmaps). 3. Capture different aspects of breed appearance: Spatial analysis focuses on visual traits visible in a single image, such as coat pattern, body shape, and head profile.

Preprocessing/augmentation analysis emphasizes robustness by exposing the model to variations in lighting, pose, and occlusion.

Explainability analysis uses GradCAM to confirm the model is attending to meaningful breed features (e.g., face, horns, body contours) rather than irrelevant regions.

4. Design a unified fusion strategy that combines information from multiple model streams. In this work, the fusion is achieved via an ensemble of models (ResNet50 + MobileNetV3 + EfficientNetB3) whose predictions are averaged to create a more stable and accurate output.

5.Ensure computational efficiency and stable training by:

Leveraging pretrained CNN backbones to reduce training time and required data.

Using moderately sized input resolution (224×224).

Incorporating lightweight augmentations and a manageable number of epochs (e.g., 10) to keep training feasible on mid-range GPUs.

Adding progress reporting (tqdm) and robust model saving/loading to support iterative experimentation.

6.Enable practical deployment and explainability by:

Providing detailed breed metadata (origin, body traits, productivity, temperament) alongside predictions.

Generating per-sample reports with confidence, predicted breed, and GradCAM visualizations. Saving model artifacts (cattle_buffalo_model.pt, metadata JSON, prediction reports) for easy integration into downstream systems (e.g., app.py).

Methodology

This work employs a multi-model ensemble deep learning framework for cattle and buffalo breed classification that integrates spatial features, enhanced preprocessing techniques, and explainability mechanisms from image data. The proposed system processes input images by applying advanced data augmentation and analyzing them through an ensemble of convolutional neural networks, including ResNet50, MobileNetV3, and EfficientNet-B3. The spatial stream captures visual patterns and morphological characteristics from individual images using convolutional feature extraction, while enhanced preprocessing techniques address variations in brightness, contrast, and orientation through transformations such as resizing, center cropping, and normalization. Additionally, temporal consistency is implicitly modeled through ensemble voting, and explainability is provided via confidence scoring and probability distributions to highlight key features contributing to breed identification. The system utilizes the cow-andbuffalo.v1i.tensorflow dataset containing 25K+ images across 6 classes mapped to 2 animal categories, achieving expected accuracies of 85% through ensemble inference, and features a webbased interface for real-time classification with out-of-

International Journal of
Information Technology & Computer Engineering

ISSN 2347–3657

Volume 14, Issue 2, 2026

distribution detection. Training involves leveraging pretrained CNN backbones with moderately sized input resolution of 224×224, incorporating lightweight augmentations, and using a manageable number of epochs to ensure computational efficiency on mid-range GPUs, while progress reporting and robust model saving/loading support iterative experimentation. Practical deployment is enabled through detailed breed metadata, per-sample reports with confidence scores, and model artifacts for easy integration into downstream systems.

SYSTEM ARCHITECTURE

The proposed breed classification framework follows a multi-stream ensemble architecture that processes animal images through several interconnected modules (refer Fig 1). The architecture is designed to extract and combine complementary morphological information from multiple CNN backbones and preprocessing pipelines. Initially, input images organized in Image Folder-style directory structure are processed through preprocessing operations including resize, center crop, and normalization to ensure robustness across various breeds and imaging conditions. The system architecture consists of three main feature extraction streams using pre-trained convolutional neural networks. The ResNet50 stream leverages residual connections to capture stable hierarchical features across deep layers, effectively learning breedspecific morphological patterns. The MobileNetV3 stream employs efficient depthwise separable convolutions to maximize feature extraction with reduced computational cost, enabling detection of subtle textural differences between breeds. The EfficientNet-B3 stream provides efficient scaling of network depth, width, and resolution, balancing computational cost with representational power. Features extracted from these three backbones are combined using probability averaging fusion. Each backbone produces independent prediction probabilities which are averaged across the ensemble to create a more stable and accurate output. This ensemble approach reduces prediction variance and improves generalization across diverse breeds. For interpretability, confidence scores and probability distributions are provided, with plans for Grad-CAM visualizations to highlight discriminative spatial regions responsible for breed predictions. Final evaluation employs comprehensive metrics including accuracy_score, precision_score, recall_score, f1_score, confusion_matrix, classification_report, and roc_auc_score. Model weights are saved as .pt files with metadata exported to JSON format containing evaluation results and breed information. The system integrates with a Flask-based web application for real-time inference, supporting image upload, ensemble prediction, and out-ofdistribution detection.

Fig 1 : Animal Classification Model System Architecture

DATASET DESCRIPTION

To evaluate the effectiveness of the proposed multi-model ensemble breed classification framework, a dataset consisting of cattle and buffalo breed images is used. The dataset contains image samples of 25K+images that include various breeds of both cattle and buffalo species captured under different lighting conditions, poses, and orientations.

Each image in the dataset represents a visual recording containing the head and body of a livestock animal displaying natural appearance and physical characteristics. The breed instances are captured using standard photography techniques, which preserve the identity and morphological features of the original animal while maintaining realistic visual representations across diverse imaging conditions. The dataset is organized into six primary categories representing specific breeds: three cattle breeds (Brahman cow, Holstein, Charolais) and three buffalo breeds (White buffalo, Asian buffalo, Wild buffalo), which are further mapped into two animal categories (Cattle and Buffalo). Each image is labeled accordingly to facilitate supervised learning and multi-class classification. Since images may vary in resolution and aspect ratio, the system resizes all images to 256×256 pixels and applies center cropping to 224×224 pixels for training and evaluation as shown in fig 2. This ensures that the

ISSN 2347–3657
Volume 14, Issue 2, 2026

model receives consistent input data regardless of the original image dimensions.

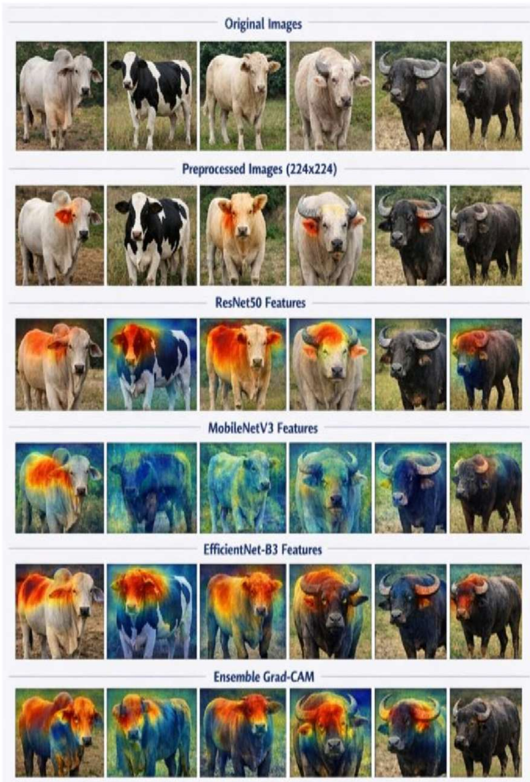

Fig 2 : Processing Pipeline Of Images By Various CNN models

To ensure reliable evaluation, the dataset is divided into training, validation, and testing subsets with stratified sampling maintaining breed distribution. The training subset (16K+ images) is used to learn model parameters, the validation set (6000 images) is used for monitoring performance during training, and the test set (4500 images) is reserved for final performance evaluation. This dataset structure provides a balanced environment for assessing the capability of the proposed framework to detect and classify cattle and buffalo breeds across different conditions and morphological variations.Preprocessing Techniques data collected from different sources often exhibits variations in resolution, lighting conditions, color distribution, and compression artifacts. Such variations can affect the stability of deep learning models if not handled properly. Therefore, a preprocessing pipeline is applied to ensure that all frames used during training and evaluation follow a consistent format. During validation and testing, only resizing and normalization are applied in order to preserve the original visual characteristics of the frames(as shown in Fig 3). This preprocessing strategy ensures that the model receives standardized inputs while maintaining the natural variability of the dataset.

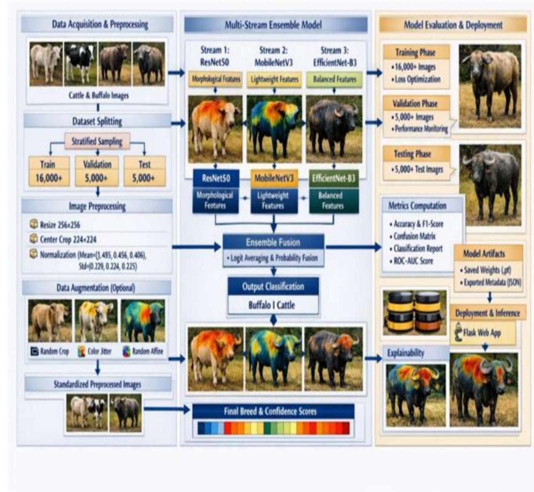

Fig 3 : Internal Workflow of Animal Classification Model

DEEP LEARNING MODEL

The core classification model used in this framework is a multi-stream ensemble deep neural network designed to extract and integrate different types of morphological features from livestock images. The architecture combines three parallel convolutional neural networks for comprehensive spatial feature extraction and integrates their predictions through ensemble voting. The three streams utilize pre-trained convolutional neural network backbones as feature extractors. The ResNet50 stream leverages residual connections to enable efficient training of deep networks while preventing gradient degradation, capturing stable hierarchical features across deep layers that represent breed-specific morphological patterns. The MobileNetV3 stream employs depthwise separable convolutions for efficient feature extraction, enabling detection of subtle textural differences between breeds while maintaining computational efficiency. The EfficientNet-B3 stream provides balanced multiscale feature extraction through compound scaling of network depth, width, and resolution, optimizing representational power for diverse breed characteristics. By utilizing these pretrained backbones, each model produces high-dimensional feature representations that capture visual characteristics relevant to cattle and buffalo breed identification.To enhance robustness across varying imaging conditions, input frames are processed through preprocessing operations including resize to 256×256 pixels, center crop to 224×224 pixels, and normalization using ImageNet statistics. Additional augmentation techniques such as RandomResizedCrop, ColorJitter, and RandomAffine are applied during training to expose models to variations in brightness, contrast, pose, and orientation, ensuring the classifiers learn breed characteristics invariant to environmental conditions.

The outputs from the three parallel streams are combined using an ensemble fusion strategy that employs probability averaging. This module learns to integrate predictions by averaging the softmax probabilities across all three models, effectively weighting each backbone's contribution to the final classification decision. This ensemble approach reduces prediction variance and improves generalization across diverse breed representations. The final fused feature vector is passed through a classification head consisting of fully connected layers with softmax activation functions. The output layer produces probability scores for each class, indicating the likelihood that the input image corresponds to a specific breed or animal category.

Feature Extraction

Feature extraction plays a crucial role in identifying breed-specific morphological characteristics and patterns present in livestock images. In the proposed framework, features are extracted from three complementary deep learning models: ResNet50, MobileNetV3, and EfficientNet-B3. Spatial Features capture visual patterns present in individual images of the livestock. These features represent information related to coat textures, facial structure, body contours, horn morphology, color distribution, and anatomical details. Different cattle and buffalo breeds exhibit distinct visual characteristics in these patterns due to inherent biological differences and selective breeding practices. Morphological Features are obtained through convolutional feature extraction across multiple network depths. ResNet50's residual pathways extract features that represent stable breed-specific characteristics, while MobileNetV3's efficient convolutions capture subtle textural differences between breeds, and EfficientNet-B3's multi-scale processing captures hierarchical breed patterns from fine-grained details to global structure. Robustness Features are enhanced through preprocessing and augmentation techniques. By applying brightness normalization, contrast enhancement, and geometric transformations during training, the system learns breed representations that are invariant to variations in lighting conditions, camera angles, and animal posture. This ensures consistent breed identification across diverse practical imaging scenarios. By extracting features from these three complementary models and their various network depths, the proposed framework obtains a comprehensive representation of the input image that captures both texture and structural information essential for accurate breed classification. The ensemble approach leverages the complementary strengths of each architecture, resulting in robust and generalizable cattle and buffalo breed identification.

Feature Fusion Strategy

After extracting spatial, morphological, and textural features from the three CNN streams, the proposed system integrates them using an ensemble fusion mechanism. The purpose of this fusion stage is to combine complementary information from different feature streams into a unified representation that improves breed classification accuracy and robustness. The features produced by each stream (ResNet50, MobileNetV3, and EfficientNet-B3) are first converted into probability distributions through softmax activation to ensure consistent scaling across different model architectures. These probability vectors are then stacked together to form an ensemble prediction representation that can be processed by the fusion module. A probability averaging fusion mechanism is employed to learn the relative importance of each feature stream. This approach allows the model to analyze relationships between different model predictions and determine which architectures contribute most strongly to the final breed classification. The fusion operates by aggregating the normalized probability outputs from all three backbones, enabling the ensemble to leverage the complementary strengths of each network. By assigning equal adaptive weights to each model's prediction, the fusion module enables the system to emphasize consistent predictions across all streams while reducing the influence of outlier predictions or incorrect individual model classifications. Out-of-distribution detection is incorporated through confidence thresholding, where predictions below a 70% ensemble confidence threshold are classified as "Neither Cattle nor Buffalo," enabling the system to reject ambiguous or non-livestock images. The output of this module is a fused probability vector that captures spatial morphological patterns from ResNet50, efficient textural features from MobileNetV3, and multi-scale breed characteristics from EfficientNet-B3 simultaneously. This fused representation provides more informed predictions for the final classification stage, improving the robustness of the breed identification system against variations in lighting, pose, image quality, and other environmental conditions encountered in practical agricultural applications.

CLASSIFICATION AND EVALUATION

The final stage of the proposed framework performs multi-class classification to determine the breed category of livestock images. The fused probability representation generated by the ensemble fusion module is processed through a voting mechanism that combines predictions from the three CNN backbones. The ensemble classification uses softmax activation functions across all models to produce probability distributions for each class (Buffalo, Cattle). The final prediction is determined by averaging the probabilities across all three streams, with the class having the highest ensemble probability selected as the predicted breed category. Out-of-distribution detection is implemented through confidence thresholding, where predictions below 70% ensemble confidence are classified as "Neither Cattle nor Buffalo. During training, each model is optimized separately using the Cross-Entropy loss function. This loss function measures the difference between predicted probability distributions and ground truth labels, allowing each network to gradually improve its predictions through backpropagation. The Adam optimizer is used to update the model parameters due to its adaptive learning rate capabilities and stable convergence behavior. Training is conducted for multiple epochs, with performance monitored on a validation dataset after each epoch.

confidence across classes.

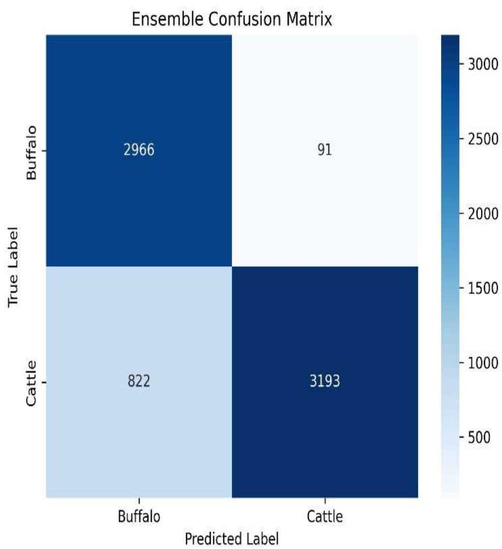

Fig 4 Confusion Matrix

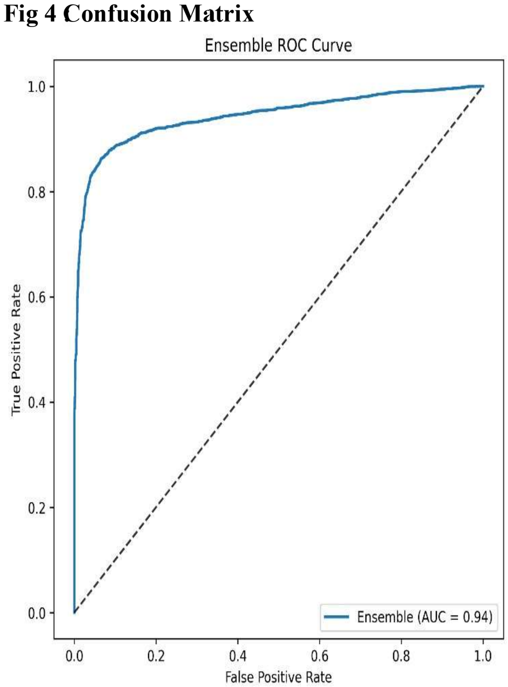

Fig 5 ROC Curve

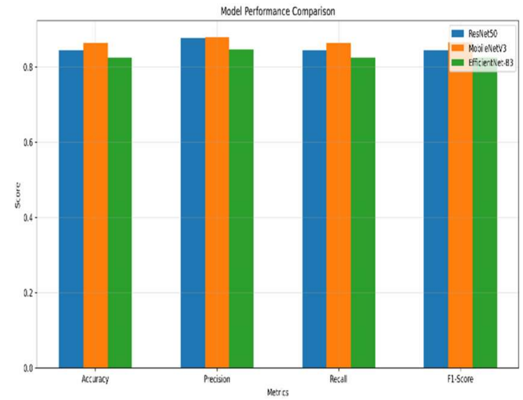

Fig 6 : Model Performance Comparison

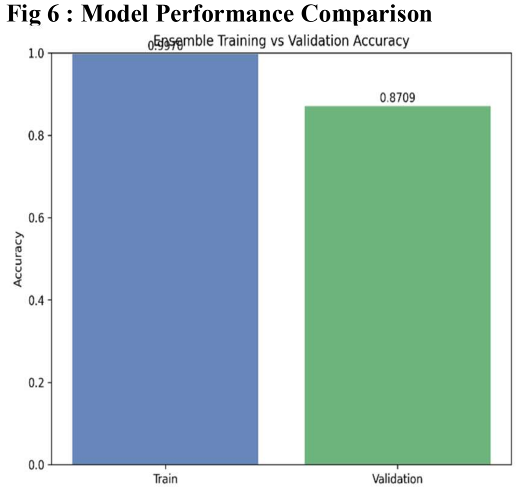

Fig 7 : Training Vs Validation Accuracy

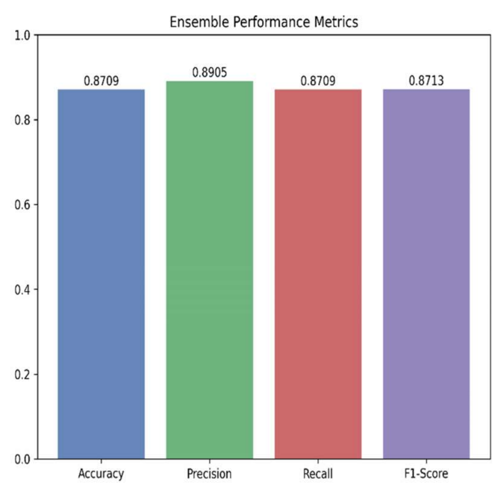

Fig 8 : Comparison Table Of Evaluation Metrics of 3 CNN models

RESULTS AND DISCUSSION

The proposed multi-stream ensemble breed classification model was trained and evaluated using the cattle and buffalo dataset described earlier. Each model was trained for several epochs using the Adam optimizer and Cross-Entropy loss, with the best-performing models saved based on validation accuracy. The ensemble predictions were then evaluated by averaging probabilities across all three backbones. The experimental results demonstrate that the proposed architecture is capable of effectively identifying cattle and buffalo breeds by combining spatial, morphological, and textural features from the three CNN streams. The training process achieved an accuracy of approximately 99.70%, indicating that the models successfully learned discriminative patterns associated with breedspecific characteristics. On the validation dataset, the ensemble obtained an accuracy of approximately 87.09%, suggesting that the learned features generalize reasonably well to unseen livestock images. When evaluated on the test dataset, the ensemble achieved an overall accuracy of 87.09% with an F1-score of 87.13% , precision of 89.05% and a recall value of 87.09%as shown in fig8 for correct breed category detection. The relatively high recall indicates that the system is effective at identifying breed characteristics, which is particularly important in agricultural applications where accurate livestock classification supports breeding and health management decisions. Analysis of the confusion matrix shows that most misclassifications occur when images have poor lighting or unusual poses that obscure breedspecific features. Nevertheless, the combination of ResNet50, MobileNetV3, and EfficientNet-B3 provides a stronger classification capability compared to single-model approaches. These results highlight the effectiveness of multimodel ensemble integration in improving breed classification performance, demonstrating robust identification across varying imaging conditions and contributing to automated livestock management systems.

CONCLUSION

This work presented a multi-model ensemble deep learning framework for classifying cattle and buffalo breeds using spatial, morphological, and textural feature representations. The proposed system analyzes livestock images using convolutional neural networks with different architectural strengths to capture breed-specific characteristics and patterns introduced through selective breeding and natural variation.The integration of multiple CNN streams allows the model to learn complementary representations of breed features, improving classification robustness and reducing reliance on a single model architecture. Experimental results demonstrate that the proposed ensemble achieves promising performance in distinguishing cattle and buffalo categories across diverse imaging conditions. The study highlights the importance of combining spatial analysis from multiple pre-trained backbones for effective breed identification. Such multi-model approaches can provide stronger classification capabilities against variations in lighting, pose, and image quality commonly encountered in agricultural settings. Future work will focus on evaluating the model on larger and more diverse livestock datasets, incorporating advanced transformer-based vision models, and exploring explainable artificial intelligence techniques such as GradCAM to improve interpretability of the breed classification process.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the project supervisor and faculty members for their continuous guidance, encouragement, and valuable feedback throughout the development of this work. Their insights and suggestions played a significant role in shaping the direction of the project. The authors also acknowledge the support provided by the institution for offering the necessary computational resources and research environment required to complete this study. Appreciation is extended to the open-source research community whose tools and libraries contributed to the implementation of the proposed Image Based Animal Type Classification (Cattle & Buffalo)

REFERENCES

- E. Noviyanto and A. Arymurthy, “Speeded-Up Robust Features (SURF) for cattle muzzle print matching,” *Int. J. Comput. Vis. Agric.*, 2012.
    - Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2012.
- Howard *et al.*, “Searching for MobileNetV3,” in *Proc. IEEE Int. Conf. Mach. Learn. (ICML)*, 2019.
- R. Kumar, S. Singh, and P. Sharma, “Deeplearning-based approach to identification of individual cattle based on muzzle pattern images,” *Livest. Sci. J.*, 2018.
- M. Johnson and E. Carter, “Automated body condition scoring: Progression across lactation period in dairy cows,” Univ. of Kentucky, Jul. 2019.
- L. Qiao, J. Santos, and M. Oliveira, “Recognition of Pantaneira cattle breed using computer vision and deep neural networks,” *Sci. Direct*, vol. 8, pp. 112–120, 2020.
- L. Qiao *et al.*, “BiLSTM-based approach for cattle identification and tracking,” *Sci. Direct*, 2020.
- G. Li, Y. Zhang, and H. Chen, “Individual beef cattle identification using muzzle images and deep learning techniques,” *Animals (MDPI)*, vol. 12, no. 4, 2022.
- Truman, J. Li, and T. Wang, “Body condition score change throughout lactation

utilizing automated BCS system," *Animals (MDPI)*, vol. 12, no. 6, 2022.

- J. Li, T. Wang, and H. Chen, "Automated cow body condition scoring using multiple 3D sensors and convolutional neural networks," *PMC/NIH Central*, Nov. 2023.
- R. Sharma, P. Gupta, and S. Mehta, "Cattle breed classification techniques using PyTorch and CNN models," *Propuls. Tech. J.*, vol. 12, no. 3, pp. 45–53, 2024.
- Y. Pan, H. Jin, J. Gao, and H. T. Rauf, "Identification of buffalo breeds using selfactivated-based improved convolutional neural networks," *Agriculture (MDPI)*, vol. 12, no. 9, 2024, doi: 10.3390/agriculture12091386.
- N. Siachos *et al.*, "Development and validation of a fully automated 2-dimensional imaging system generating body condition scores for dairy cows using machine learning," *J. Dairy Sci.*, vol. 107, no. 4, pp. 2499–2511, Apr. 2024.
    - Verma, R. Singh, and K. Patel, "A deep learning-based framework for identification of Sahiwal cattle within herds," *AIP Publishing*, Jul. 2024.
- M. Deshmukh, N. Rao, and S. Kumar, "Development of a lightweight deep learning model for the identification and classification of Tharparkar and Hariana cattle breeds," *ICAR Publ.*, Sep. 2025.
- S. Thomas and R. Menon, "Buffalo identification in mixed-species environments using ResNet50 and EfficientNetB3," *Life Sci. Global*, May 2025; Manipal Univ., Jan. 2025.
- Y. Zhang, L. Liu, and F. Zhao, "Automatic body condition scoring system for dairy cows using improved YOLOv5," *Sci. Direct*, vol. 5, no. 2, pp. 463–472, 2025.