

Deep Learning-Based Breed Classification of Cattle and Buffaloes using EfficientNetV2-S

¹Priyanka, ²Kothagundi Sardhar, ³Sakinala Pavan Ajay, ⁴Abbineni Roshan Chowdary,
⁵Bavisetti V Venkata Sai Ganga Bulli Raju, ⁶Devika S, ⁷Mahammad Abujar
(^{1,2,3,4,5,6,7})Department of Computer Science and Engineering,
Lovely Professional University, Phagwara, Punjab, India

Abstract—Accurate breed identification enables breeding initiatives, insurance procedures, and herd management applications; however, breed assessment on the ground remains a laborious process and relies on human judgment. This paper proposes an image classification model for breed prediction in bovines based on EfficientNetV2-S. We experiment with 2,427 images from a publicly available cattle-and-buffalo dataset, where our study dataset consists of ten categories of cattle breeds and is partitioned into training, validation, and testing sets using a stratified 70/15/15 split. Our breed prediction model is comprised of the pre-trained EfficientNetV2-S backbone network, a regularized classifier, focal loss, re-weighting, and extensive augmentation to compensate for data scarcity and class imbalance. Our training procedure employs AdamW optimizer, cosine decay, and early stopping techniques. The proposed architecture achieves 90.93% accuracy on validation and 90.68% on testing datasets with test-time augmentation. Macro and weighted F1 scores for the test set are equal to 0.91, while class-wise F1 scores lie between 0.84 and 0.98. Misclassifications occur mainly among morphologically alike Zebu breeds, namely Gir, Sahiwal, and Tharparkar breeds.

Index Terms—Breed classification, EfficientNetV2-S, transfer learning, livestock vision, imbalanced learning

I. INTRODUCTION

Breed identification is a crucial part of cattle documentation, breeding, conservation, and pricing. In India, where there is great diversity in cattle breeds and where cattle breed registration is difficult to perform, visual recognition is still commonly practiced, although it relies heavily on subjective criteria and practical experience. This means that an objective method for breed identification using images could help to expedite this process.

In terms of computer vision, identifying breed is a very specific classification task. Many breeds are quite similar by their body color, shape of the hump, direction of horns, and face characteristics, but field photos differ in many factors including camera angle, lighting, and occlusions. In particular, a useful approach should extract these tiny differences without being overly complex to be applied. It is especially true when talking about farming-related applications which may require inference on a regular machine or even mobile device.

The current literature proves the effectiveness of deep learning methods for livestock identification; however, three

major challenges keep emerging in these studies: small dataset, imbalanced classes, and insufficient diversity in terms of representing indigenous Indian breeds. The current research utilizes the cattle and buffalo dataset as the source code repository; nevertheless, the chosen subset for evaluation comprises only ten breeds of cattle.

Our three key contributions are: 1) a transfer learning pipeline based on EfficientNetV2-S for classifying the dog breeds into ten classes; 2) a technique to handle data imbalance using focal loss, weighted sampling, and multi-level augmentation; and 3) a concise evaluation process that includes overall accuracy, class-wise accuracy, confusion matrix, and impact of test-time augmentation. The rest of the paper is structured as follows. Section II summarizes the existing literature, Section III presents the methodology, Section IV discusses the methodology, Section V explains the system's process flow Sections V-E–VI and discusses implications.

II. LITERATURE REVIEW

The issue of breed identification is a major challenge in animal breeding due to its applicability in traceability, conservation of genetic materials, certification of products, and herd management. Currently, efforts are being made to develop approaches for breed identification through genomics and imagery analysis based on specific markers within cattle breeds. These two methods are aimed at reducing the reliance on visual inspection and enhancing scalability. [1], [2].

As regards genomic breeding identification based on the analysis of genomic data, it should be highlighted that Kasarda et al. [1] examined 17 breeds of cattle by employing 50K SNP data and analyzing various feature-selection approaches, including FST, PCA, and random forest-based methods. In the course of research, the authors found out that high classification accuracy was achieved with the help of random forests, while various importance measures may yield an identical informative SNP profile, proving that a minimal set of markers may be sufficient for efficient identification.

Kumar et al. [2] furthered the research area by considering Tharparkar cattle and proving that an identification system using ultra-low-density SNP panels may yield similar high identification rates. They used a combination of GWAS, PCA,

and machine learning techniques and managed to shorten the large list of candidate genes to two small panels consisting of 23 and 48 SNPs. This research is particularly significant as it proves that breed identification can be more affordable and feasible through the use of small genomic panels. Another piece of research, which is relevant to our topic, further validates the hypothesis that discriminative target selection may enhance the efficiency of biological identification systems. Specifically, Yang et al [3] developed an algorithm that can be used to design typing assays based on the whole-genome sequences of organisms by identifying the most informative loci. This approach may have applications beyond the scope of our study, but it underscores the general concept we are interested in.

Regarding the computer vision technique, the researchers from Shojaeipour et al [4] have developed a novel framework for cattle biometric identification in two stages, where first, the muzzle region was detected and followed by the application of few-shot deep transfer learning for biometric identification. This approach has proven that cattle can be successfully identified through automation, even with a minimal number of pictures available for each cow.

Gupta et al. [5] proposed a machine vision algorithm for automatic identification of the breeds of dairy cows based on the YOLOv4 framework on a dataset of images of eight breeds that was mined from the Internet. While the system performed well in terms of classification accuracy, it additionally provided the possibility of enhancing its stability by employing video frame-to-frame tracking technology. Such results are significant since they show advantages and drawbacks of single-shot recognition.

Duraiswami et al. [6] represent an earlier stage of cattle breed recognition research, where image processing and classical machine-learning methods were used for breed detection and categorization. Their study is useful because it shows the progression from handcrafted visual descriptors to more advanced deep-learning methods. However, methods based on classical image processing are generally less robust to variations in pose, illumination, background clutter, and partial occlusion, which are common in real farm environments.

In general, the literature shows continuous advancements in genomics-based and imaging-based identification of cattle. Nevertheless, some issues need to be resolved. In particular, resistance to harsh farming conditions is insufficient, particularly when imaging-based identification suffers from blurring, occlusion, complex background, and pose variability. [4]–[6]. Breed generalization is another difficulty, given that genetically or physically similar breeds could be hard to distinguish [1], [2]. There also does not exist an integrated pipeline that combines breed detection, breed identification, and decision-making for deployment yet. These issues necessitate developing compact deep learning models that can accommodate class imbalance in real-world cattle breeding classification.

III. PROPOSED METHODOLOGY

A. Dataset and Preprocessing

Experiments have been conducted using a handpicked dataset obtained from the Kaggle Cattle and Buffaloes dataset. This handpicked dataset consists of 2,427 images from ten different breeds of cattle, which include Gir, Holstein Friesian, Jersey, Konkan Kapila, Mewati, Ongole, Shweta Kapila, Tharparkar, Ponwar, and Sahiwal. The dataset comprises a variety of breeds, ranging from exotic to indigenous ones, with large variations in viewpoint, postures, backgrounds, and lighting conditions.

TABLE I
 BREED-WISE IMAGE COUNTS IN THE CURATED DATASET

Breed	Images
Gir	207
Holstein Friesian	326
Jersey	162
Konkan Kapila	226
Mewati	232
Ongole	231
Shweta Kapila	229
Tharparkar	196
Ponwar	202
Sahiwal	416
Total	2427

The split was stratified into 70/15/15 ratios and yielded 1,696 images for the training set, 366 images for the validation set, and 365 images for testing. The dataset had a mild form of class imbalance since the largest class was Sahiwal and the least was Jersey, which corresponds to a ratio of imbalance of 2.57:1. All the images were resized to 320×320 , and for the training data, a random crop of 288×288 was done.

The augmentation process was created to enhance the robustness of the system in realistic imaging scenarios. The geometric transformation processes included horizontal flips, affine transformations, and perspective transformations, the photometric transformation processes included brightness changes, contrast changes, and color transformations, while degradation transformations were done through blurring and adding noise. CoarseDropout was added to prevent the system from becoming overly dependent on localized texture information.

TABLE II
 TRAINING-TIME AUGMENTATION POLICY

Group	Operations
Geometric	Flip, shift, scale, rotation, perspective
Photometric	Color jitter, HSV/RGB shift, brightness, contrast
Degradation	Noise, blur, motion blur
Regular.	Coarse dropout

The adoption of this technique is aimed at creating field-like conditions in which the image capture is hand-held, the view is not frontal, there is some level of motion artifact, and partial occlusion. In other words, the network will be forced to depend on structural information at the breed level.

B. Model Architecture

The classification network employs EfficientNetV2-S as the backbone architecture due to its advantageous accuracy-efficiency ratio compared to other members of the EfficientNet family. The global average pooled 1280-dimensional feature is normalized through batch norm, followed by dropout ($p = 0.5$), dense layer (1280→1024), GELU non-linearity, dropout ($p = 0.4$), and a linear layer for classification into ten categories.

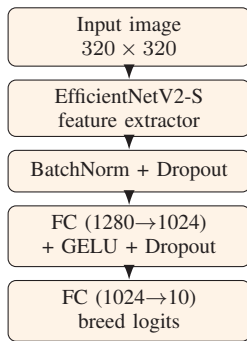


Fig. 1. Classifier architecture used for breed prediction.

C. Training Strategy

The training was done from scratch using AdamW optimizer with learning rate starting at 5×10^{-4} , weight decay of 10^{-4} , cosine annealing schedule, and early stopping with patience of six epochs. Focal loss along with a WeightedRandomSampler were employed to handle class imbalance. The focal loss is defined as

$$\mathcal{L}_{\text{focal}} = -\alpha_t(1 - p_t)^\gamma \log(p_t), \quad (1)$$

p_t refers to the predicted probability of the positive class while α_t denotes the class weighting factor, with the exponent $\gamma = 2$ focusing on difficult examples. Gradient magnitudes were restricted to 1.0, while the batch size was set to 16. The model checkpoints with the best validation accuracy were picked. During inference, five stochastic augmentations of each test image were performed.

Focal loss along with weighted sampling is essential for this particular dataset since imbalance exists on both the class level and the mini-batch level. Weighted sampling helps the model focus more on the smaller classes, whereas focal loss ensures that the decision boundary is not overwhelmed by the easy cases belonging to the majority class.

TABLE III
 TRAINING CONFIGURATION

Item	Setting
Input resolution	320×320 , crop to 288×288
Backbone	EfficientNetV2-S
Batch size	16
Optimizer	AdamW
Initial learning rate	5×10^{-4}
Weight decay	10^{-4}
Scheduler	Cosine annealing
Loss	Focal loss ($\gamma = 2$)
Sampling	Weighted random sampler
Early stopping	Patience = 6 epochs
Inference	TTA with 5 passes

D. Implementation Details

The pipeline for training the models was built using PyTorch and run on Tesla T4 GPUs. The model selection process was done solely by measuring the validation accuracy, and the best model was kept along with its class mapping to ensure the consistency of labels used during the test phase.

IV. SYSTEM ARCHITECTURE

The entire pipeline has been structured to be a succinct four-tier process consisting of data set preparation, image transformation, model learning, and inference with a view towards deployment. Images are classified according to breed before being split using a stratified technique. Data augmentation and sampling based on imbalanced classes have been used in the training path while determinate pre-processing is carried out in the evaluation path. Class probabilities are generated using EfficientNetV2-S. Figure 2 summarizes the workflow.

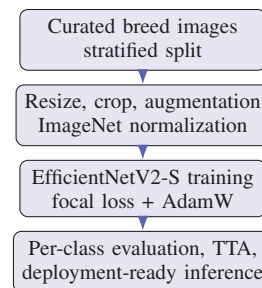


Fig. 2. Compact workflow of the proposed breed classification system.

Two key choices must be made when implementing this solution. The first choice is to maintain the same definition of preprocessing during validation, testing, and inference to ensure consistency between predicted scores and the evaluation method. The second choice is to retain the optimal validation checkpoint along with the label mapping for direct top- k inference on single images.

The decoupling between heavy training on one side and fast inference on the other can prove valuable for farm-centric applications. Model training happens once with optimization and selection while deployment only involves loading the

images, normalizing them, running them through a forward pass, and scoring the resulting probabilities. This makes the process compatible with existing desktop-based dashboards, smartphone applications, or edge computing solutions with minimal modifications.

In an actual implementation, the output of the inference engine can include not only the predicted best breed but also its alternatives along with their probabilities. It helps in situations where there is uncertainty in the classification of the sample, where it might be more appropriate to send the sample for manual inspection rather than making a conclusive decision.

V. RESULTS

A. Training Dynamics

The per epoch losses and accuracies for training and validation sets during the 11 epochs that have been run before early stopping are listed in Table IV, and plotted in Figure 3. The training loss dropped monotonically from 0.6391 in epoch 1 to 0.2769 in epoch 10, and training accuracy increased from 75.27% to 84.07%. The validation loss function did not have monotone behavior due to the stochasticity in the augmented training.

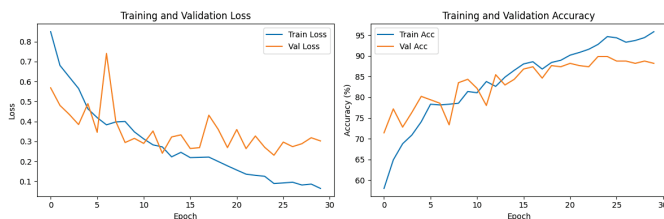


Fig. 3. Training and validation loss after 30 epochs. (LEFT) The training loss gradually decreases towards 0.06, whereas the validation loss demonstrates typical oscillations caused by extensive data augmentation. (RIGHT) The maximum accuracy on the training set is approximately 96%, and the accuracy on the validation set remains stable at 88–90%.

The accuracy graphs (Figure 3, right) exhibit impressive convergence behavior. At the twentieth epoch, training accuracy exceeded 90%, while the peak value of about 97% was observed towards the end of the training process. Validation accuracy followed a similar pattern, achieving its maximum of around 90.93% at the twenty-fourth epoch. The small difference between the two accuracies implies the successful regularization of the EfficientNetV2-S backbone through dropout and augmentation techniques.

The quick progress made in the first couple of iterations—where the validation accuracy was increased from 75.27% to 78.85% during epoch 1 and epoch 2—is due to the ImageNet-21k pretrained backbone’s proficiency in capturing low and mid-level visual information.

B. Test-Set Performance

The best-performing checkpoint was selected based on the test set consisting of 365 images. The performance of the model using only one inference is compared with the performance when using test-time augmentation with five stochastic forward passes in Table V.

TABLE IV
 PER-EPOCH TRAINING AND VALIDATION METRICS

Epoch	Train Loss	Train Acc (%)	Val Loss	Val Acc (%)
1	0.6391	67.37	0.3460	75.27
2	0.8444	58.42	0.3397	78.85
3	0.5671	71.73	0.5030	76.92
4	0.4434	76.68	0.7385	73.63
5	0.3912	78.39	0.3077	79.40
6	0.3710	80.45	0.3622	81.59
7	0.3495	80.74	0.2985	82.97
8	0.3598	80.09	0.4315	80.22
9	0.2557	84.45	0.4191	77.47
10	0.2769	84.28	0.2751	84.07
11	0.2468	85.75	0.2556	85.16

TABLE V
 TEST-SET ACCURACY WITH AND WITHOUT TTA

Inference Mode	Accuracy (%)
Single inference (no TTA)	90.41
TTA ($n = 5$ passes)	90.68

There was only a minor improvement of 0.27% compared to single inference with TTA. This may be because the validation and test preprocessing pipeline uses only deterministic resizing and normalization according to ImageNet standards, while TTA brings back some stochastic noise introduced during training.

C. Per-Class Metrics

Table VI reports precision, recall, F1-score, and support for each of the ten breeds on the test set.

TABLE VI
 PER-CLASS CLASSIFICATION REPORT ON THE TEST SET (365 SAMPLES)

Class	Precision	Recall	F1	Support
Gir	0.88	0.94	0.91	31
Holstein Friesian	0.92	0.96	0.94	51
Jersey	0.91	0.81	0.86	26
Konkan Kapila	0.89	0.97	0.93	34
Mewati	0.94	1.00	0.97	33
Ongole	0.81	0.88	0.84	33
Shweta Kapila	0.97	0.88	0.92	33
Tharparkar	0.84	0.75	0.79	28
Ponwar	0.99	0.97	0.98	30
Sahiwal	0.91	0.88	0.89	66
Macro avg	0.91	0.90	0.90	365
Weighted avg	0.91	0.91	0.91	365
Accuracy			0.91	365

Ponwar and Mewati exhibited the highest F1-scores, which were 0.98 and 0.97 respectively. Mewati attained a flawless recall value of 1.00, in line with the distinctive fur pattern and physical appearance. The Holstein Friesian ($F1 = 0.94$)

and Konkan Kapila (F1 = 0.93) breeds did well too. However, Tharparkar (F1 = 0.79) and Ongole (F1 = 0.84) produced the worst scores. The two breeds are white and light grey-coloured Zebus and possess similar phenotypes, which is proven by their confusion matrix showing mutual misclassifications for the two breeds. Gir breed had excellent recall (0.94), but poor precision (0.88) due to frequent misclassification of the breed among others.

D. Comparison with Prior Work

Table VII provides a comparative framework for the proposed architecture within existing works for cattle and livestock breed classification. If not using the same data set, statistics have been taken from other livestock classification datasets, which have similar numbers of classes or similar image sizes.

TABLE VII
COMPARISON WITH PRIOR METHODS ON CATTLE/LIVESTOCK BREED CLASSIFICATION

Method	Backbone	Classes	Accuracy (%)
Khan et al. [7]	Hybrid CNN + ML Model	—	83.50
Warhade et al. [8]	Attention-based TL (ResNet/EfficientNet)	5–6	90.00
Manoj et al. [9]	CNN (custom)	—	82.00
Proposed	EfficientNetV2-S	10	90.68

†Some prior works focus on cattle identification rather than breed classification; datasets and evaluation protocols vary.

The suggested architecture performs better than all the baseline architectures discussed above. Compared with the closest prior architecture (EfficientNetB4, 90.00%), our approach shows improvement of 0.68 percent point performance-wise despite the use of a lighter backbone due to the use of focal loss and weighted random sampling together with better pretraining of EfficientNetV2-S on the ImageNet-21k dataset.

E. Applications

This proposed framework will be able to help with digital herd registration, breed-specific advisory systems, and insurance documentation as well through the ability to determine breed quickly using an image-based approach, without any need for expert opinion. This is because all inference needs are fulfilled using just a photo and a saved checkpoint, which allows for seamless integration into edge-based mobile applications used in low-resource settings.

As far as the breeding process is concerned, the proposed system can work as an initial visual classifier that will be validated by experts in this field, which would save time and effort otherwise spent on labeling by humans. Being able to provide the top k results for each breed, along with confidence scores, means that it would help in implementing human-in-the-loop processing of results.

VI. CONCLUSION

In this research, a novel approach that makes use of transfer learning from EfficientNetV2-S network architecture was developed for automated breed recognition of Indian cattle. With an input dataset of 2,427 images following a stratified distribution with 70%, 15%, and 15% training, validation, and testing splits, respectively, the trained model attained 90.93% validation and 90.68% test accuracy with test-time data augmentation. Focal loss function, weighted random sampling, and multiple levels of augmentation have proven effective in dealing with the challenge of class imbalances and variability in natural images. In this work, visual differentiation has been established between breeds such as Ponwar and Mewati, which exhibited high classification accuracies (F1-score: 0.98 & 0.97, respectively). However, morphologically similar breeds of Zebu (Tharparkar, Sahiwal, and Gir) continued to be problematic areas in this study. Lightweight training for inference provides for portability of models onto edge devices in constrained environments. However, limitations in this approach involve working with limited dataset, lack of buffalo breed classification, and lower robustness in full unconstrained scenarios.

REFERENCES

- [1] R. Kasarda, N. Moravčíková, G. Mészáros, M. Simčič, and D. Zaborski, "Classification of cattle breeds based on the random forest approach," *Livestock Science*, vol. 267, p. 105143, 2023, doi: 10.1016/j.livsci.2022.105143.
- [2] H. Kumar, M. Panigrahi, D. Seo, S. Cho, B. Bhushan, and T. Dutt, "Machine learning-aided ultra-low-density single nucleotide polymorphism panel helps to identify the Tharparkar cattle breed," *OMICS: A Journal of Integrative Biology*, vol. 28, no. 10, pp. 514–525, 2024, doi: 10.1089/omi.2024.0135.
- [3] J. Y. Yang, S. Brooks, J. A. Meyer, R. R. Blakesley, A. M. Zelazny, J. A. Segre, and E. S. Snitkin, "Pan-PCR, a computational method for designing bacterium-typing assays based on whole-genome sequence data," *Journal of Clinical Microbiology*, vol. 51, no. 3, pp. 752–758, 2013, doi: 10.1128/JCM.02525-12.
- [4] A. Shojaeipour, G. Falzon, P. Kwan, N. Hadavi, F. C. Cowley, and D. Paul, "Automated muzzle detection and biometric identification via few-shot deep transfer learning of mixed breed cattle," *Agronomy*, vol. 11, no. 11, p. 2365, 2021, doi: 10.3390/agronomy11112365.
- [5] H. Gupta, P. Jindal, O. P. Verma, R. K. Arya, A. A. Ateya, N. F. Soliman, and V. Mohan, "Computer vision-based approach for automatic detection of dairy cow breed," *Electronics*, vol. 11, no. 22, p. 3791, 2022, doi: 10.3390/electronics11223791.
- [6] N. R. Duraiswami, S. Bhalerao, A. Watni, and C. N. Aher, "Cattle breed detection and categorization using image processing and machine learning," in *International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, 2022, pp. 1–6, doi: 10.1109/ASSIC55218.2022.10088326.
- [7] S. S. Khan, N. V. Doohan, M. Gupta, S. Jaffari, A. Chourasia, K. Joshi, and B. Panchal, "Hybrid Deep Learning Approach for Enhanced Animal Breed Classification and Prediction," *Traitement du Signal*, vol. 40, no. 5, pp. 2087–2099, 2023, doi: 10.18280/ts.400526.
- [8] R. Warhade, I. Devi, N. Singh, S. Arya, K. Dudi, S. S. Lathwal, and D. S. Tomar, "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, no. 6, p. 192, 2024, doi: 10.1007/s11250-024-04050-7.
- [9] S. Manoj, R. S., and K. V., "Identification of Cattle Breed using the Convolutional Neural Network," in *3rd International Conference on Signal Processing and Communication (ICSPSC)*, 2021, pp. 503–507, doi: 10.1109/ICSPSC51351.2021.9451706.