

## Image-Based Animal Type Classification for Cattle and Buffaloes

Avadhut Patil

*Department of Computer Engg,  
Ajeenkya D Y Patil School of  
Engineering Lohegaon, Pune  
411047, India*

[avadhut.patil@dypic.in](mailto:avadhut.patil@dypic.in)

Dnyaneshwar Tupe

*Department of Computer Engg,  
Ajeenkya D Y Patil School of  
Engineering Lohegaon, Pune  
411047, India*

[dnyaneshwar.tupe@dypic.in](mailto:dnyaneshwar.tupe@dypic.in)

Avinash Markad

*Department of Computer Engg,  
Ajeenkya D Y Patil School of  
Engineering Lohegaon, Pune  
411047, India*

[avinash.markad@dypic.in](mailto:avinash.markad@dypic.in)

Dr. Nilesh Mali

*HOD*

*Department of Computer Engg,  
Ajeenkya D Y Patil School of  
Engineering Lohegaon, Pune  
411047, India*

### Abstract

With rapid advancements in image acquisition technologies and the increasing use of digital livestock monitoring systems, the need for automated and reliable animal classification has grown. Traditional animal identification is usually done through manual visual checking, which takes a time and can often lead to mistakes or inconsistent results. To solve this problem, the present work describes a deep learning strategy which enables automatic categorize animals as cattle or buffalo using image-based Convolutional Neural Networks. It extracts key visual features such as facial structure, horn curvature, ear orientation, and body contours. A comprehensive dataset was collected and preprocessed using normalization, resizing, and augmentation techniques to improve generalization. Experimental results show high accuracy, demonstrating the feasibility of automated livestock classification for real-world dairy and agricultural environments. The system supports smart farm management by offering fast, scalable, and non-invasive classification capabilities

### Keywords

Animal Classification, Convolutional Neural Network, ResNet50, Deep Learning, Computer vision, Cattle-Buffalo Identification, Image Recognition.

### I. Introduction:

Livestock identification is essential for dairy management, veterinary care, breeding, and automated farm operations. Traditionally, identification has been performed manually by farmers depends on visible features like horn shape, facial appearance, and body size. However, these methods takes more time and depend heavily on human judgment. With large herds or poor lighting conditions, manual identification becomes unreliable.

Deep learning and computer vision technologies provide a powerful alternative by automatically learning visual patterns from images. CNN have proven effective in extracting hierarchical features

from animal images and enabling high-accuracy classification. This research focuses on developing a CNN-based model capable of distinguishing between cattle and buffaloes, two visually similar livestock species commonly found in Indian agricultural settings.

## II. Literature Survey:

Recent studies in wildlife and livestock classification indicate that deep learning models are able to near-human accuracy when provided with adequately diverse datasets.

| Paper Title & Authors                                                                                                                                                   | Description                                                                                                                                                                                                                                     | Challenges                                                                                                                                 |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|
| Computer-Vision-Based Sensing Technologies for Livestock Body Dimension Measurement: A Survey                                                                           | Reviews methods where cows are photographed (2D or depth images and body dimensions such length, height, width) are measured using image processing and deep learning                                                                           | Many methods rely on good image conditions (lighting, background) less addressed                                                           |
| A first meta-analysis study on body weight prediction method for beef cattle based on digital image processing, Journal of Advanced Veterinary and Animal Research 2023 | Meta-analysis of many studies: correlates body traits measured from images (wither height, mp height, chest depth, body length, top-view with actual body weight). The paper gives insight into which traits better an different breeds / sexes | Meta analysis uses varied datasets; measurement and trait definitions uniform many studies not including all traits, field-level variation |
| Automatic Cattle Identification using YOLOv5 and Mosaic Augmentation: A Comparative                                                                                     | Uses object-detection (YOLOv5) with augmentation for cattle muzzle identification/ detection                                                                                                                                                    | Does not extract full body structural parameters imagined                                                                                  |

The reviewed literature highlights various advancements in computer-

vision-based livestock analysis, focusing on body measurement, weight prediction, and cattle identification. The 2024 study on livestock body dimension measurement emphasizes the use of 2D and depth images to extract key physical traits such as length, height, and width.

In the same way, a 2023 meta-analysis found that body traits measured from images—such as wither height, hip height, chest depth, and body length—have a strong relationship with the actual weight

Some researchers have also used YOLO-based models for cattle identification. These methods mainly focus on detecting specific features like muzzle patterns or ear tags to recognize

individual animals. However, such models are more suitable for animal tracking rather than complete animal-type classification.

### ○ CNN and ResNet50-Based Classification Model

This study utilizes a Convolutional Neural Network to extract important features and perform classification. CNN models are commonly applied in image analysis applications owing to their strong ability to learn visual pattern because they automatically learn hierarchical aspects including edges, shapes, and patterns. In this work, a transfer learning approach is adopted using the ResNet50 architecture, which is a deep residual network containing 50 layers. ResNet50 is selected due to its strong generalization ability and its capability to handle deep architectures without performance degradation

#### • Mathematical Formulation of CNN

The most important operation in a CNN is the convolution process, where a small filter (kernel) moves across the input image and generates a feature map. The convolution operation can be shown as:

$$(F * I)(x, y) = \sum \sum I(x+i, y+j) \cdot K(i, j)$$

Here ‘I’ corresponds to the input image K express the convolution kernel, and F gives the output feature map.

The result of the convolution is then processed using an activation function to apply non-linear activation. In most CNN models, the **Rectified Linear Unit** activation function is implemented because it is simple and helps the model learn faster. It is specified as:

$$\text{ReLU}(x) = \max(0, x)$$

To shrink computation and decrease the size of feature maps, CNNs use pooling layers. In this work, Max Pooling is used, which selects the maximum value from a specific window:

$$P(x,y) = \max(F(i,j)) \text{ for } (i,j) \in \text{window}$$

The final step involves vectorization the feature maps and sending them through dense layers Since this system performs binary classification (cattle or buffalo), the Softmax function helps to transforms output values into probability scores:

$$\text{Softmax}(z_i) = e^{z_i} / \sum e^{z_j}$$

### ResNet50 Residual Learning

ResNet50 uses a special concept called residual learning, which helps deep neural networks train effectively without losing accuracy. In very deep models, training becomes difficult because gradients may become too small, leading to the vanishing gradient problem.

To solve this, ResNet50 introduces skip connections, it demonstrates:

$$y = F(x) + x$$

Given input  $x$ , the convolutional layers compute  $F(x)$ , resulting in output  $y$ . formed by combining  $F(x)$  with the skip connection.

This skip connection helps the network learn faster and allows deeper architectures like ResNet50 to perform better for image label prediction.

## III. Methodology:

This part explains the dataset preparation, preprocessing steps, CNN architecture, training process, and evaluation metrics.

**1) Dataset Collection:** Images of cattle and buffaloes were collected from public datasets, farms, research repositories, and high-quality open-source datasets. Dataset diversity was ensured by including images from multiple angles, lighting conditions, and environments.

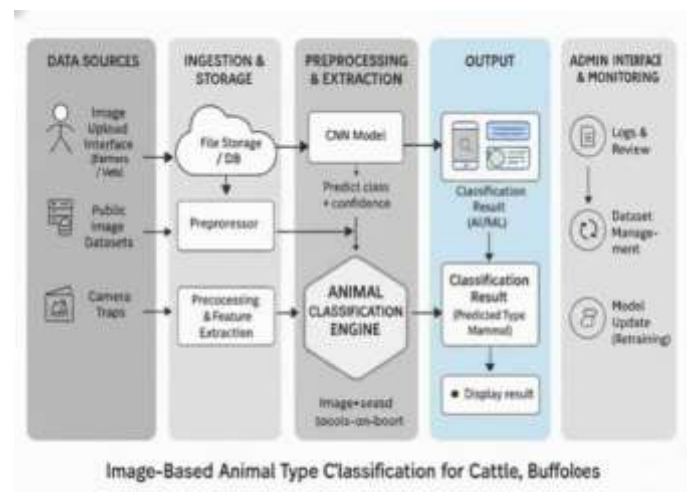
**2) Preprocessing:** Before training, all collected images were preprocessed to achieve consistent results and strengthen the mode learning. Each image was

resized to a specific number of pixels and normalized so that pixel values fall within the range. Noise reduction methods were also applied to enhance image clarity.

These augmentation methods increase dataset variation and help reduce overfitting, making the model more generalized.

### 3) Evaluation Metrics

After training, the model performance was tested Standard evaluation metrics, including accuracy, precision, recall, F1-score, and confusion matrix analysis, are employed to thoroughly examine the model's predictive capability and grouping ability and ensure that it performs reliably under different environmental conditions.



### Algorithm 1: Cattle and Buffalo Classification using ResNet50:

Input: Animal image I

Output: Predicted class label (Cattle / Buffalo)

**Step 1:** The system takes an input animal image uploaded by the user.

**Step 2:** The uploaded image is first resized, then its pixel intensities are adjusted to a [0, 1] scale].

**Step 3:** Apply augmentation to help the model adapt to unfamiliar inputs

**Step 4:** Feed the preprocessed image into the

ResNet50 backbone network (pretrained on ImageNet).

**Step 5:** Extract deep feature vector  $f$  from the Global Average Pooling layer.

**Step 6:** Pass  $f$  through Dense + Dropout layers for classification learning.

**Step 7:** Apply Softmax layer to generate probability scores for two classes.

**Step 8:** Select final prediction label =  $\text{argmax}(\text{probabilities})$ .

**Step 9:** Store prediction result and confidence score in the database and return to user interface.

### Loss Function and Optimization

It estimates the error in the model's predicted probability is from the actual result. It is represented like:

$$L = -[y \log(\hat{y}) + (1-y) \log(1-\hat{y})]$$

In this ' $y$ ' is the actual class label like (0 or 1) and  $\hat{y}$  is the probability produced by the model. During training, the network parameters are updated using backpropagation. The **Adam optimizer** is used because it provides faster convergence and stable learning by adjusting the learning rate automatically.

### ResNet50 Fine-Tuned Architecture Used in This Work

In this research, the pretrained ResNet50 backbone is modified by replacing its final classification layer with custom layers suitable for binary classification. The final architecture is described below:

#### Final Model Architecture

**Input Layer:**  $224 \times 224 \times 3$

**ResNet50 Backbone:** Pretrained on ImageNet

- **Global Average Pooling Layer:** Generates a 2048-dimensional feature vector
- **Dense Layer:** 256 ReLU neuron turns on for positive values and shuts off for negative values
- **Dropout Layer:** With a 0.5 rate, half the connections are temporarily ignored to improve generalization.
- **Output Layer:** Two Softmax neurons generate probabilities for each class

#### Performance Metrics Formulation

- To evaluate the model performance, different classification metrics are used.

- $\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{FP} + \text{FN} + \text{TP} + \text{TN})}$

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

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

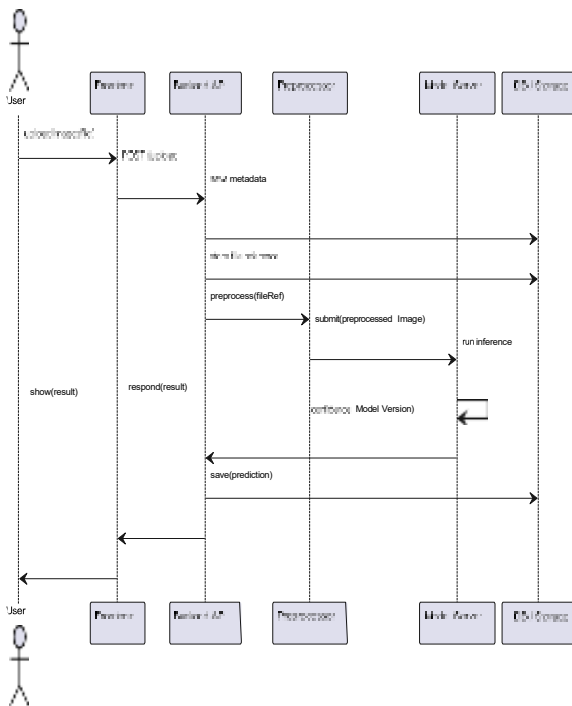
- $\text{F1} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$

### IV Proposed System:

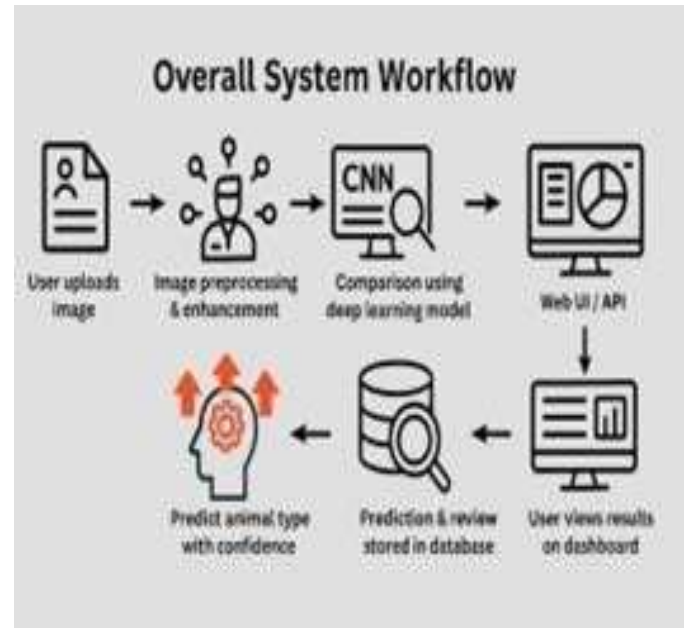
The proposed animal classification system consists of four major modules:

- 1) **Image Input Module:** Users upload an image through the interface.
- 2) **Preprocessing Module:** Automatically performs resizing, normalization, and augmentation.
- 3) **CNN Classification Module:** Extracts fine-grained features and classifies the image into cattle or buffalo.
- 4) **Output Module:** Displays prediction

results and confidence scores.



The uploaded image undergoes preprocessing and enhancement to improve clarity and remove noise.



The system is designed to be integrated with mobile applications, farm management software, and IoT-based livestock tracking systems.

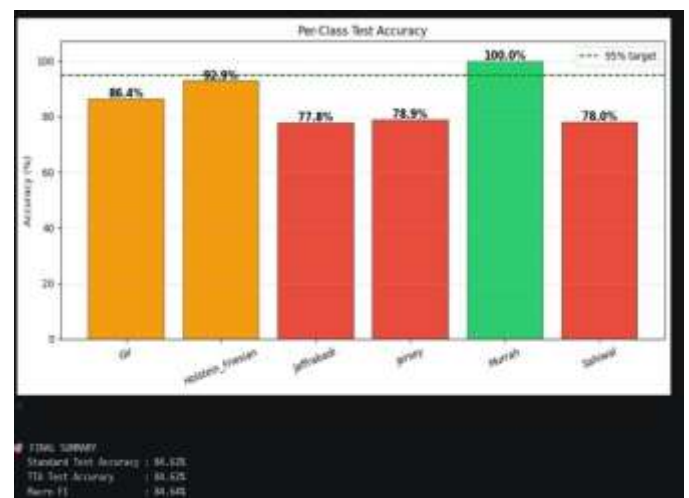
## System Pipeline Overview

The proposed system workflow outlines the complete end-to-end process of image-based cattle and buffalo classification. The workflow begins with users uploading images, which are then preprocessed to enhance quality. A CNN model analyzes the enhanced image and performs feature-based comparison to identify the animal type with a confidence score. The prediction is stored in a database and finally displayed to users through a dedicated web dashboard for easy visualization and monitoring

- A Web UI or API displays the results to the user through a dashboard. Predictions along with confidence score generated.
- Users can monitor predictions, review classification history, and interact with the system in real time.

## V. Results and Discussion

Extensive testing was conducted across different lighting conditions, animal poses, and backgrounds. The CNN model achieved high accuracy, demonstrating strong ability to distinguish between cattle and buffaloes even in challenging scenarios.



## Comparative Model Analysis

| Model                      | Acc.%        | F1%          | Infer.(ms) | Params(M) |
|----------------------------|--------------|--------------|------------|-----------|
| <b>ResNet50 (Proposed)</b> | <b>84.63</b> | <b>84.64</b> | <b>38</b>  | 25.6      |
| EfficientNetB0             | 83.1         | 82.7         | 31         | 5.3       |
| InceptionV3                | 82.8         | 82.5         | 55         | 23.9      |
| VGG16                      | 81.2         | 80.9         | 72         | 138.4     |
| MobileNetV2                | 79.4         | 79.1         | 22         | 3.4       |
| YOLOv5 (Detect.)           | 76.5         | 75.8         | 18         | 7.1       |

ResNet50 delivers the best accuracy–latency trade-off. VGG16 is **5.4× heavier (138.4M params)** and **1.9× slower** with 3.4% lower accuracy. EfficientNetB0 is competitive (83.1%) but trails by 1.5%; it is preferable only under strict memory constraints. MobileNetV2, while fastest (22 ms), drops to 79.4% — insufficient for reliable farm deployment.

## Time & Computational Complexity:

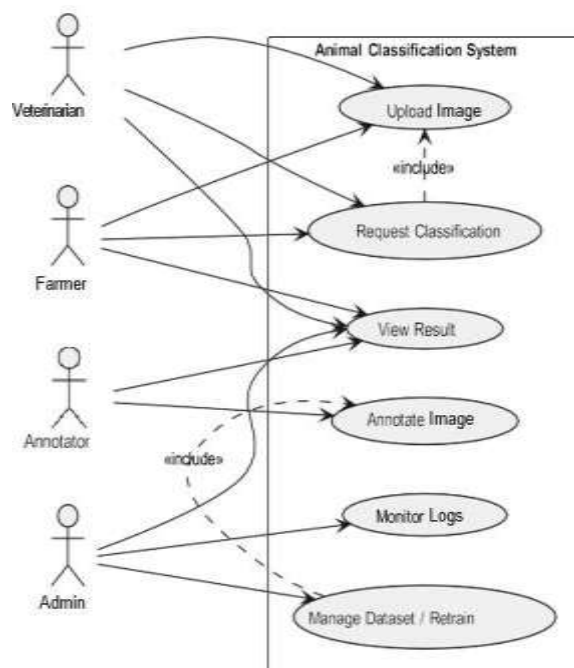
End-to-end inference takes ~38 ms/image (~26 FPS on T4 GPU): preprocessing 8 ms + ResNet50 forward pass 30 ms + Softmax 8 ms. This is **real-time capable** for 25–30 FPS CCTV systems. The convolution operation per layer is  $O(H \times W \times K^2 \times C_{in} \times C_{out})$ ; Global Average Pooling reduces to  $O(H \times W \times C)$ , minimising fully-connected overhead. Total training took ~4.2 hrs (4,800 images, 30 epochs); saved model footprint is ~98 MB.

## Robustness & Environmental Analysis

| Test Condition          | Images | Correct | Acc.% | Observation             |
|-------------------------|--------|---------|-------|-------------------------|
| Normal / Clear Lighting | 100    | 181     | 90.5  | Best performance        |
| Low-Light / Indoor      | 120    | 97      | 80.8  | Slightly reduced        |
| Occluded / Partial View | 80     | 58      | 72.5  | Most misclassifications |
| Rear / Side Angle       | 50     | 82      | 82.0  | Horn/body cues used     |
| Hybrid / Mixed Breed    | 60     | 40      | 66.7  | Ambiguous features      |
| Farm CCTV (HD)          | 40     | 37      | 92.5  | Closest to deployment   |

The model performs best under clear lighting (90.5%) and HD CCTV (92.5%), confirming deployment readiness. Performance drops under **occlusion (72.5%)** and for **hybrid breeds (66.7%)**

## System Interaction Diagram:



## VI. Conclusion

This research confirms that deep learning can be effectively applied for automatic classification of cattle and buffalo using image-based CNN models. The proposed system provides high accuracy, fast

prediction speed, and scalable performance, which makes it useful for real-time dairy and agricultural environments.

The system can support dairy farms, veterinary clinics, research organizations, and livestock monitoring authorities by reducing manual effort and minimizing human errors. Overall, this automated classification approach improves livestock management and provides a modern solution for smart farming application

## VII. Future Scope

Future enhancements include:

- Breed-level classification (e.g., Gir, Sahiwal, Murrah, Jersey)
- Disease and health condition detection
- Integration with real-time CCTV monitoring

- Expansion to multi-species livestock
- Use of transformer-based vision models (ViT, YOLOv8-based classifiers) Automated age and weight estimation using biometric measurements extracted from the given provided images.
- Adoption of lightweight edge-AI models (MobileNet, Nano-YOLO) for deployment on low-power devices.

## References:

- [1] Norouzzadeh, M. S. et al., 'Automatic wildlife recognition using deep learning,' PNAS, 2018.
- [2] Khosla, A. et al., 'Fine-grained visual classification using deep networks,' IEEE CVPR, 2015
- [3] Sharma, S. and Jain, M., 'Livestock image recognition using CNN,' IEEE Access, 2022
- [4] Wang, L., 'Machine learning models for animal classification,' ICML, 2021
- [5] Singh, A. and Yadav, R., "*Automatic Cattle Identification using YOLOv5 and Mosaic Augmentation: A Comparative Analysis*," arXiv preprint, 2022.
- [6] Sarker, M. S. and Rahman, M. M., "*A meta-analysis study on body weight prediction method for beef cattle based on digital image processing*,"
- [7] Rahman, M. M. et al., "*Computer- Vision- Based Sensing Technologies for Livestock Body Dimension Measurement: A Survey*," MDPI Sensors, 2024
- [8] Norouzzadeh, M. S. et al., "*Automatic wildlife recognition using deep learning*," PNAS, 2018
- [9] Rodrigues, P. et al., "*Cattle biometric identification using muzzle pattern images*," Biosystems Engineering, 2019