A Computer Vision Driven Ecosystem for Cattle Monitoring: Multi-Disease Classification with Severity Grading, Multi-View Individual Identification, and Weight Estimation

1. Aabu Yousuf Raj[a] | Corresponding Author | aabuyousufraj@gmail.com | +88.01.973.632.826
2. Md. Rakibul Hasan[a] | md.rakibul.hasan5@g.bracu.ac.bd
3. Abdur Rahman Arafat[a] | abdur.rahman.arafat@g.bracu.ac.bd
4. S.M.Sadman Rahman[a] | sm.sadman.rahman@g.bracu.ac.bd
5. Md. Roman Bin Jalal[a] | roman.bin.jalal@g.bracu.ac.bd
6. Dr. Jannatun Noor Mukta[b] | jannatun@cse.uiu.ac.bd

[a]Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh
[b]Computing for Sustainability and Social Good (C2SG) Research Group, Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh

A Computer Vision Driven Ecosystem for Cattle Monitoring: Multi-Disease Classification with Severity Grading, Multi-View Individual Identification, and Weight Estimation

Aabu Yousuf Raj[a], Md. Rakibul Hasan[a], Abdur Rahman Arafat[a], S.M.Sadman Rahman[a], Md. Roman Bin Jalal[a] and Dr. Jannatun Noor Mukta[b]

[a] *Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh*

[b] *Computing for Sustainability and Social Good (C2SG) Research Group, Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh*

ARTICLE INFO

Keywords:
Computer Vision
Cattle Disease Classification
Severity Grading
Multi-view Identification
Weight Estimation
Deep Learning
Precision Livestock Farming

ABSTRACT

This work introduces a computer vision based cattle monitoring ecosystem that will (i) create a multi-disease classification and severity grading, (ii) create a multi-view unique cattle identification, and (iii) create a body weight estimation with the use of four view RGB images, based on a custom-collected, curated dataset supplemented with publicly available sources. For disease analysis, a hierarchical deep learning architecture classifies Lumpy Skin Disease (LSD), Foot-and-Mouth Disease (FMD), Infectious Bovine Keratoconjunctivitis (IBK), and Healthy cattle, and grades diseased cases into Stage-1 (mild), Stage-2 (moderate), and Stage-3 (severe) with veterinarian supervision. The cross-attentional multi-task model achieves 89.96% disease classification accuracy, 83.75% severity staging accuracy, and 85.88% hierarchical accuracy (disease → severity). For individual recognition and weight estimation, a multi-view appearance-based pipeline is built using left, right, front, back images where YOLOv8s localizes cattle regions with IoU = 0.93 and mAP@0.5 = 0.97. Using ConvNeXt-Tiny, identification reaches Rank-1 accuracy of 96.40% (Rank-5: 100%) under Leave-One-View-Out and 74.56% (Rank-5: 89.80%) under cross-view-angle and domain shift testing. For body weight estimation, regression over multi-view images plus metadata (ID, sex, breed, age) achieves an MAE of 36.99 kg with a DN121-tuned regressor which improved to 35.10 kg with ensemble learning. Overall, the results demonstrate that an affordable, low-shot, non-invasive RGB-based system can support disease diagnosis with severity grading, reliable multi-view cattle identification, and practical live-weight estimation without invasive sensors.

1. Introduction

The health control of livestock is a crucial problem in the production of dairy and beef as they take up a considerable portion of the agricultural economy and rural livelihoods. In many resource-limited settings, cattle diseases lead to major losses from early death, lower output, and treatment expenses [1]. Veterinary diagnostic methods such as manual examination and laboratory tests are time consuming, costly and not easily accessible in rural locations, so diagnosis is delayed and results in disease transmission [2].

Besides, two other pillars of cattle management include individual identification and body-weight monitoring, which are major aspects for small and rural farms because reliable identity tracking supports vaccination and treatment follow-ups, and breeding records, while accurate weight estimates are essential for correct drug and vaccine dosage determination, feed planning, market pricing, and monitoring growth and health status to reduce losses and improve overall profitability. Tag-based approaches (ear tags or RFID) are prone to loss and require handling that increases stress and labor, while manual identity logging is unreliable at scale, motivating vision-based re-identification that can track the

*Corresponding author
aabu.yousuf.raj@g.bracu.ac.bd (A.Y. Raj); md.rakibul.hasan5@g.bracu.ac.bd (Md.R. Hasan); abdur.rahman.arafat@g.bracu.ac.bd (A.R. Arafat); sm.sadman.rahman@g.bracu.ac.bd (S.M.Sadman Rahman); roman.bin.jalal@g.bracu.ac.bd (Md.R.B. Jalal); jannatun@cse.uiu.ac.bd (Dr.J.N. Mukta)
ORCID(s): 0009-0000-1394-1545 (A.Y. Raj); 0009-0009-1904-8614 (Md.R. Hasan); 0009-0002-4263-459X (A.R. Arafat); 0009-0008-5530-6820 (S.M.Sadman Rahman); 0009-0000-8253-488X (Md.R.B. Jalal); 0000-0001-9669-151X (Dr.J.N. Mukta)

same animal over time; however, many existing methods are constrained to a single viewpoint and degrade under pose, illumination, and occlusion changes. Likewise, accurate weight measurement typically depends on weighing platforms or specialized sensors (e.g., depth devices) that are expensive and impractical in low-resource settings, creating demand for RGB-only weight estimation that leverages easily captured images, ideally enriched with basic animal metadata, without reference objects or controlled environments.

The last upgrades in artificial intelligence and computer vision have enabled automated livestock monitoring systems. The first methods were based on classical machine learning with systematic observation of symptoms, which had not been robust under various farm conditions [3, 4]. Deep learning models and, specifically, Convolutional Neural Networks (CNNs) have shown better performance in identifying the existence of visually manifest diseases based on RGB images since 2021 [5]. Nevertheless, a large majority of current systems focused on single-disease detection, muzzle-based recognition, or weight estimation of the controlled-environment, and did not offer any solutions in the form of integrated monitoring [6, 7].

There are a number of acute problems of modern livestock management that the given research solves. The diagnosis of diseases is highly reactive, and symptoms are only observed when the diseases are in an advanced stage. Lumpy Skin Disease (LSD), Foot-and-Mouth Disease (FMD), and Infectious Bovine Keratoconjunctivitis (IBK) are the most significant risks, although most detection systems detect only in binary mode without determining the severity [8, 9]. Cow tracking (individual) with conventional systems such as ear tags and RFID have the problem of loss, breakage, and handling stress [10], whereas the use of mechanical platforms to monitor weight is costly and impractical to the local or small farmers [11]. The current computer vision systems have weaknesses: (1) the disease classifiers do not include the severity of diseases clinically [12], (2) identification methods rely on a single viewpoint and fail under pose and viewpoint variations in real farm conditions [13], (3) weight estimation models are based on depth sensors or reference objects [14, 15], and (4) the majority of the datasets are farm specific or breed specific [16].

This paper has filled these gaps by creating a complete computer vision-based cattle monitoring system with three contributions: (1) a hierarchical disease classification system with clinically-validated levels of severity controlled by a District Livestock Officer, (2) a protocol-based multi-view identification framework that is resistant to variations in viewpoint with four standard views (left, right, front, back), and (3) a metadata-enhanced RGB-based weight estimation system without depth sensors or reference object. System results are validated on a hierarchical custom collected and merged disease dataset covering three major diseases (LSD, FMD, IBK) along with three clinically defined severity stages of Stage-1 (mild), Stage-2 (moderate), Stage-3 (severe) and on custom dataset that contains 215 cattle, where identification uses only four standard-view images (left, right, front, back) and weight estimation uses the same four-view RGB images combined with metadata where performance is reported using accuracy, precision, recall, F1 for disease classification and severity grading, Rank-k and mAP for identification, and MAE, Root Mean Squared Error (RMSE), R^2 for weight estimation. Lastly, although disease classification, cattle identification, and weight estimation have largely been studied as independent tasks, their shared reliance on RGB images enables a unified deployment scenario and paves the way to integrate all three modules into a single end-to-end monitoring pipeline.

2. Related Work

2.1. Livestock Disease Detection

The use of ML and AI to predict the diseases of livestock has been in use in the previous years [17, 4]. Early work by Motohashi et al. [18] created subclinical mastitis detection in an automated milking system by utilizing ML-based models which yielded 81% percent with Random Forest. Phulu et al. [2] put forward a two-model AI system that combines symptom-based Random Forest (90.62% accuracy) with image-based MobileNetV2 (91.1% accuracy) in FMD detection. Recent transfer learning research reported MobileNetV2 with a 92.8% accuracy in FMD classification. [5], while Saqib et al. [8] recorded 95% detection of LSD with MobileRMSNet. Girmaw [12] reported EfficientNetB7 that attained 99.01% accuracy in multi-class skin disease classification. These systems, however, are not severity graded, and detect multiple classes or binary disease, so that clinical utility in terms of prioritizing treatment is limited.

2.2. Cattle Identification Systems

Computer vision techniques used in identifying cattle non-invasively have become an option to RFID and ear tags. Qiao et al. [19] produced hybrid CNN+LSTM with 91% identification accuracy on video sequences. Li et al. [6] had 95.74% accuracy with multiple feature decision layer fusion networks with spatial and temporal cues. The biometric identification through the muzzle using VGG16-BN had 98.7% accuracy. [10] Specifically exhibiting strong discriminative ability of muzzle patterns. Multi-view approaches by Bergamini et al. [13] obtained 81.7% Top-1 using a combination of frontal and side-views. Nevertheless, single or limited views are used in the majority of systems with reliance on many images per view, and a cross-angle generalization or domain shift is hardly tested, so effective multi-view identification systems remain in need.

2.3. Weight Estimation Approaches

The weight estimation through computer vision has developed in three paradigms of methodology: measurement based through anatomical landmarks. [16, 20], end-to-end regression [21, 22], and hybrid segmentation-assisted approaches [23, 11]. Ruchay et al. [14] obtained MAE of 18.5 kg on 3D depth cameras with Microsoft Kinect. In the case of RGB-D fusion methods, MAE was 24.3 kg [15], and

RGB only methods recorded 45.2 kg MAE by single-view [21] and 33.2 kg with depth estimation [24]. Dang et al. [23] tested 3D point cloud segmentation on Random Forest with MAE of 25.2 kg. There are also some RGB image based systems that got MAE of 18.02 kg and 43.44 kg but it completely rely on reference object and many images per cow which is not actually practical [25, 22]. Nevertheless, although such improvements have been made, RGB-only methods are still not as precise as systems based on depth, and most of them lack metadata on cattle (age, sex, breed) to make predictions more stable.

3. Methodology

The proposed methodology independently works on three interrelated activities, including hierarchical disease classification and severity grading, multi-view cattle identification, and RGB-based weight estimation. All the components utilize transfer learning based on CNN architectures and share a common data pre-processing pipeline and have task-specific model architectures and evaluation protocols. Figure 1 illustrates the overall methodological pipeline.

3.1. Dataset Collection and Preparation

3.1.1. Disease Classification Dataset

A detailed dataset of diseases was formed by incorporating publicly available files in Kaggle [26, 27] and Roboflow [28] along with personal-collected IBK images. The sample consists of the photographs of cattle with LSD (nodular skin lesions), FMD (vesicles on the mouth and hooves), IBK (opacities and discharge of the eyes), and healthy cattle. Because original sources did not have severity annotations, the clinical criteria used to classify the diseases severity of each diseased image resulted in Stage-1 (mild), Stage-2 (moderate), and Stage-3 (severe), based on its symptoms. The stage severity grading was classified by Dr. S.M. Mahbubur Rahman (District Livestock Officer, Naogaon, Rajshahi, Bangladesh) by validating visual cues of from the diseased images.

3.1.2. Identification and Weight Estimation Dataset

With the volunteering permission from farm owners, four-view images which includes left, right, front, back view (one image per view) were photographed using smartphones of 215 individual cattle from 15 local farms in Naogaon, Rajshahi, Bangladesh to identify them and estimate their weight. In each sample, there is metadata that includes, cattle ID, age (0.5-6 years), sex, breed (Australian Friesian Sahiwal, Australian Holstein Friesian, Brahma, Local and Cross breeds), and live weight (59-654 kg, mean 265 kg) which has been taken after manually weighing each cow on weight scale. Identification uses only images, while the weight estimation uses images and metadata. Tight manual bounding box of the cattle area was done using Computer Vision Annotation Tool (CVAT)[29] to train YOLOv8s object detector for localizing, cropping cattle region and remove background cues to initiate automatic determination of cattle region and enhance practical feasible situation. For identification cross angle and domain shift generalization test, protocol B consists of images from different view and angle from the train dataset (e.g., image from top view). Figure 4 shows representative four-view samples.

3.2. Data Preprocessing and Augmentation

For disease classification, images were preprocessed to ensure consistent EfficientNet-B1 [30] model input quality and transfer learning compatibility. All disease samples were resized to 240×240 using bicubic interpolation and normalized by scaling pixel values to [0,1] followed by ImageNet [31] standardization ($\mu = [0.485, 0.456, 0.406]$, $\sigma = [0.229, 0.224, 0.225]$). For evaluation consistency, we applied a fixed stratified split by the 10-class label, reserving 15% as a held-out test set, and performed stratified 5-fold cross-validation on the remaining 85%. For training only data augmentation was done by random cropping (scale 0.85-1.0), horizontal flipping ($p = 0.5$), rotation (±10▪), and color jittering to enhance the robustness of the models and stop overfitting without altering the morphology of the lesions. Figure 3 presents the distribution of the classes by the stages of severity demonstrating class imbalance which has been handled by oversampling or balanced sampling with on the fly augmentation and weighted loss.

For identification, all images underwent automatic cattle detection and localization using YOLOv8s, followed by cropping to the predicted bounding box region to reduce background noise. Cropped regions were resized to 224×224 using bicubic interpolation, normalized to [0,1], and standardized with ImageNet statistics ($\mu = [0.485, 0.456, 0.406]$, $\sigma = [0.229, 0.224, 0.225]$) for transfer learning compatibility. The dataset was split using an ID-based stratified random sampling scheme into 80:20 train-validation to ensure that images of the same cattle ID never appear in both splits and prevent identity leakage. Identification-specific augmentation emphasized viewpoint and appearance variability by applying random flip ($p = 0.5$), rotation (±15°), perspective transformation (scale 0.2), and random erasing ($p = 0.1$) to simulate partial occlusion and real farm conditions.

For weight estimation, YOLOv8s was similarly used for cattle detection and localization, and the cropped regions were resized to 224×224 and standardized using the same normalization scheme. The dataset was split using stratified random sampling into 80:20 (172 training, 43 validation) to preserve representative weight distributions (training mean: 264.8 kg, validation mean: 266.2 kg). To preserve body proportions that directly affect weight correlation, conservative augmentation was applied (flip $p = 0.3$, rotation ±5°, brightness ±10%). Four-view images were concatenated into a 12-channel tensor: $\mathbf{X} = \text{Concat}(\mathbf{I} * \text{front}, \mathbf{I} * \text{back}, \mathbf{I} * \text{left}, \mathbf{I} * \text{right}) \in \mathbb{R}^{224\times224\times12}$, and A custom PyTorch layer that splits the tensor into four views for shared-backbone processing and aggregation via global average pooling before the regression head.

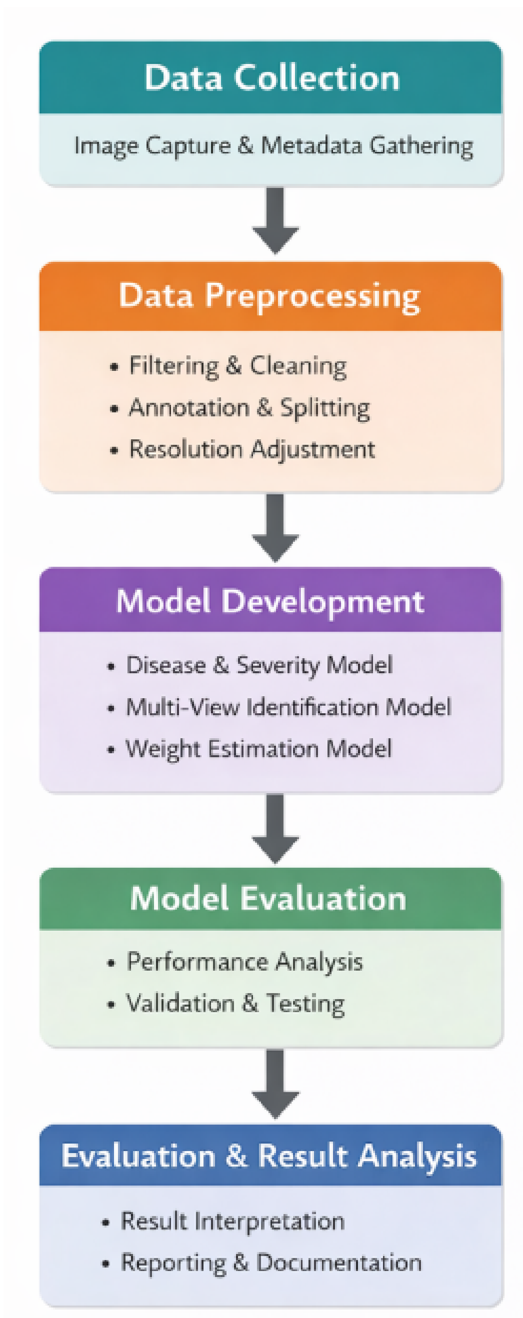

Figure 1: Methodology Plan and Development Workflow

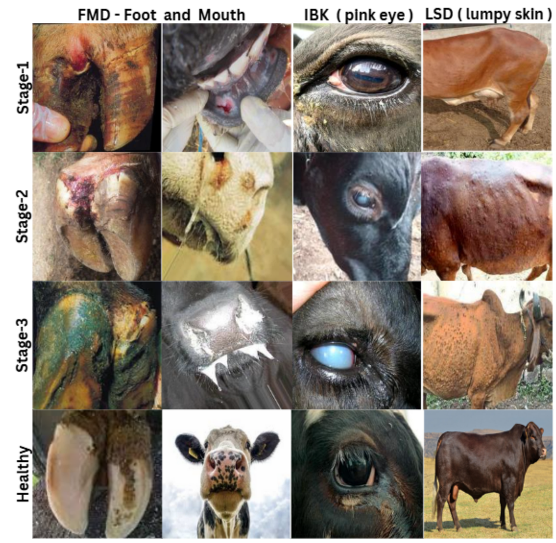

Figure 2: Disease dataset example after severity labeling

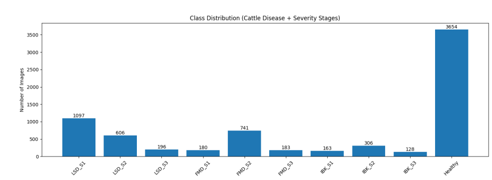

Figure 3: Disease dataset class distribution per stage

Figure 4: Locally acquired dataset for identification and weight estimation

3.3. Disease Classification with Severity Grading

The problem of disease classification was defined as a hierarchical two-level problem: first of all, classical disease classification Healthy, LSD, FMD, IBK and then the disease cases are subject to severity staging Stage-1, Stage-2, Stage-3. The figure 5 shows methodology overview for cattle disease classification with severity grading. Evaluation was done on five modeling architectures:

Option A - Flat 10-class Classifier: One single softmax classifier with the different combinations of disease-severity as individual classes. Nonhierarchical, enforcement not performed. Training minimized class-weighted cross-entropy over the 10-class label:

$$\mathbf{L} = \mathbf{CE}(\mathbf{y}_{\text{pred}}, \mathbf{y}^{\mathbf{10}}_{\text{true}})$$

Option B - Cascaded Two-stage Pipeline: Sequential architecture with independent disease classifier then conditional severity classifier. Conforms to the clinical reasoning, but it has a problem with errors propagation.

Option C - Shared-backbone Multi-task: Single CNN backbone with dual prediction heads for disease and severity. Efficient inference with shared feature learning. Both heads followed the same structure:

$$GAP \rightarrow Dropout(0.25) \rightarrow Linear(1280 \rightarrow C) \rightarrow Softmax$$

The training objective combined disease loss with a masked severity loss to prevent healthy samples from contributing severity gradients:

$$\begin{aligned} \mathbf{L}_{\text{total}} &= \mathbf{L}_{\text{disease}} + \lambda, \quad \lambda = 1.0 \\ \mathbf{L}_{\text{disease}} &= \text{WCE}(\text{disease}_{\text{pred}}, \text{disease}_{\text{true}}) \\ \text{disease}_{\text{mask}} &= (\text{disease}_{\text{label}} \neq \text{healthy}) \end{aligned}$$

Option D - Multi-task with Consistency Constraints: This is an extension of Option C where inference time validation is performed with hierarchical consistency.

Option E - Cross-attentional Multi-task: Improved multi-task architecture with a two-way flow of information (disease and severity branches) with the help of attention. Chosen as the best architecture depending on preliminary consideration. Disease and severity representations were fused as:

$$\begin{aligned} \text{disease}_{\text{fused}} &= \alpha \cdot \text{disease}_{\text{feat}} + (1-\alpha) \cdot \text{Attn} \\ \text{severity}_{\text{fused}} &= \beta \cdot \text{severity}_{\text{feat}} + (1-\beta) \cdot \text{Attn} \end{aligned}$$

where α and β were learned gating parameters. The training objective preserved masked severity learning and added a small regularization term for attention and gate weights:

$$\mathbf{L}_{\text{total}} = \mathbf{L}_{\text{disease}} + \lambda \cdot \mathbf{L}^{\text{masked}}_{\text{severity}} + \mathbf{L}_{\text{attention\_reg}}, \quad \lambda = 1.0$$

EfficientNet-B1 [30] was used in all models as backbone. Training used AdamW optimizer with differential learning rates: 1×10^{-3} for classification heads and 5×10^{-5} for backbone, with 0.25 dropout and weight decay 1×10^{-4}. A 5-epoch warmup froze the backbone to stabilize head initialization. The models were trained with batch size 32 with 25 epochs using ReduceLROnPlateau scheduler (factor 0.1, patience 5).

3.4. Multi-view Cattle Identification

The concept of individual cattle identification was developed as a deep metric learning problem instead of closed-set classification, allowing the flexibility of identity matching based on the similarity of embedding. The identification pipeline consists of YOLOv8s based cattle detection to focus on region of interest and prevent model focusing on background cues, feature extraction, and similarity-based retrieval. The figure 6 shows methodology overview for unique cattle identification system.

Two CNN backbones were evaluated: ResNet-50 [32] and ConvNeXt-Tiny [33], both pre-trained on ImageNet. Both of them are trained on ImageNet prior to use. The architecture comprises of: (1) CNN backbone extracting visual features, (2) projection head projecting visual features to 512-dimensional embedding space and (3) combined ArcFace [34] and SupCon [35] metric learning losses.

ArcFace introduces angular margin penalty to enhance inter-class separability:

$$\mathcal{L}_{\text{ArcFace}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j\neq y_i} e^{s\cos\theta_j}} \quad \text{(I)}$$

where s is scale factor, m is angular margin, θ_j is angle between embedding and class center.

SupCon encourages embeddings of the same identity to cluster together:

$$\mathcal{L}_{\text{SupCon}} = -\sum_i \frac{1}{|P(i)|} \sum_{p\in P(i)} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_p/\tau)}{\sum_{a\neq i} \exp(\mathbf{z}_i \cdot \mathbf{z}_a/\tau)} \quad \text{(II)}$$

where $P(i)$ are positive pairs (same identity), τ is temperature parameter, $\mathbf{z}$ are normalized embeddings.

Models were trained with systematic hyperparameter tuning with backbone learning rate 1×10^{-4}, embedding head learning rate 3×10^{-4}, weight decay 3×10^{-4}, batch size was 32 with training for 50 epochs using early stopping (patience 10).

There were two complementary protocols, which would measure various aspects of model performance:

Protocol A (Leave-One-View-Out): A 3-view training, and the last view testing to compares view-invariant representation learning where 5-fold cross-validation enables identity-level splits to avoid information leakage.

Protocol B (Cross-angle Domain Shift): Testing on different viewing angles that the images were not observed during the training as is typical in the real world where reference and query images might be different in perspective. This protocol evaluates cross view and domain shift generalization which is important in practical farm deployment.

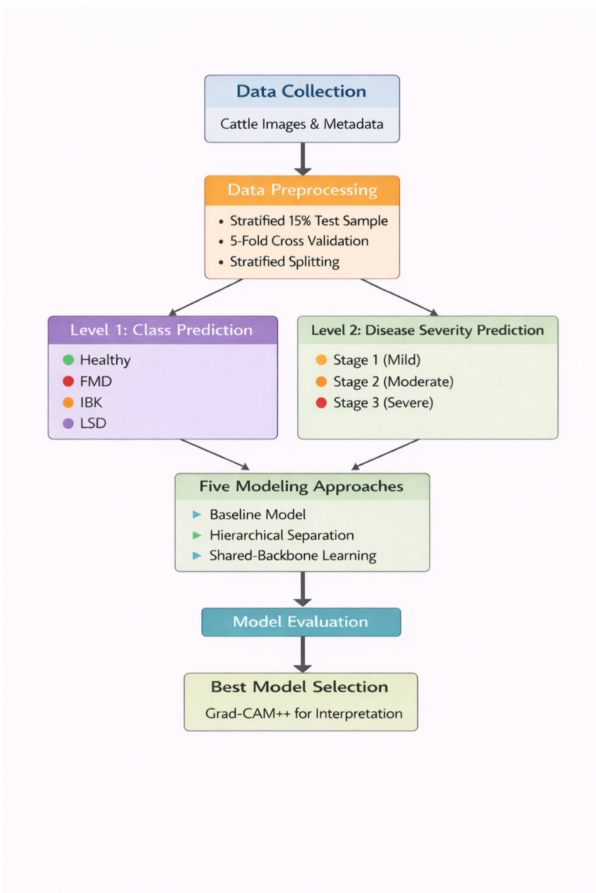

Figure 5: Methodology overview for cattle disease classification with severity grading system

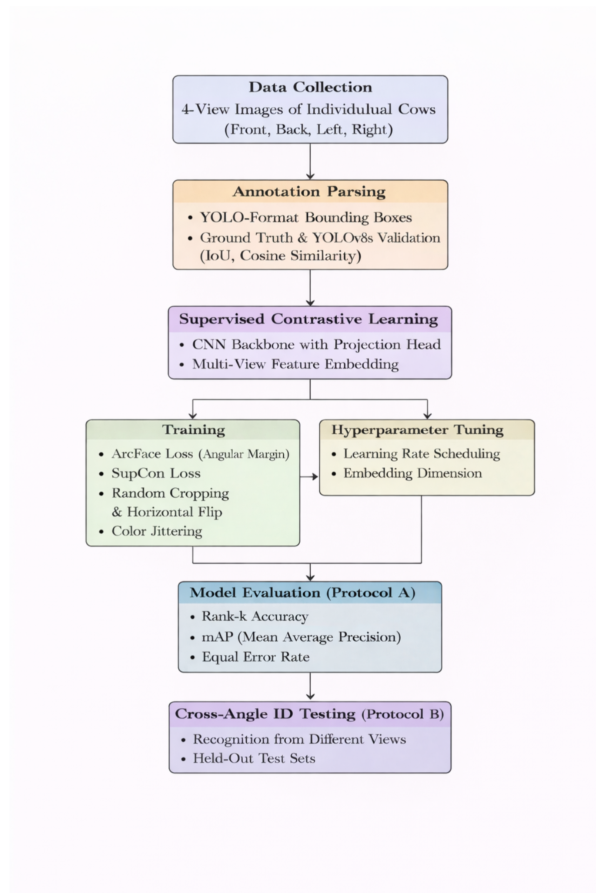

Figure 6: Methodology overview for unique cattle identification system

3.5. Body Weight Estimation

The estimation of weights was developed as a multi-view regression formulation comprising visual features with cattle metadata where the first aspect was YOLOv8s based cattle detection to focus on region of interest and prevent model focusing on background cues. The evaluation of the CNN architectures consisted of six CNN architectures: DenseNet121 [36], ResNet50 [32], EfficientNet-B1 [30], InceptionV3 [37], MobileNetV2 [38], and ConvNeXt-Tiny [33]. The two architectures were then tested under two training conditions (1) frozen transfer learning where backbone weights are fixed and regression head is only fine-tuned, (2) fine-tuned transfer learning where the entire network is fine-tuned with differential learning rates in domain adaptation. The figure 7 shows methodology overview for cattle disease classification with severity grading. Evaluation was done on five modeling architectures:

Models for weight estimation were trained at firsst with all layers frozen and trained only the regression head using Adam (learning rate = 0.001), Mean Squared Error (MSE) loss, batch size 32, and 50 epochs with early stopping (patience = 15) and fine-tuned with unfroze higher layers for domain adaptation with a reduced learning rate of 0.0005, specifically unfreezing from Layer 100 (DN121, 327 layers), Layer 80 (MobileNetV2, 74 layers), and Layer 200 (InceptionV3, 111 layers).

Moreover, they were evaluted using Mean Squared Error (MSE) loss: $\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$, where y_i is true weight and $\hat{y}_i$ is predicted weight. An ensemble model combining top-performing architectures was constructed using weighted averaging: $\hat{y}_{\text{ensemble}} = \frac{1}{K}\sum_{k=1}^{K}\hat{y}_k$, where K is number of models.

3.6. Performance Evaluation Metrics

Models were evaluated using task-specific metrics. Disease classification used accuracy, precision, recall, and F1-score at three levels: disease classification (4-class), severity staging (3-class on diseased samples), and hierarchical joint prediction (10-class), computed as mean ± standard deviation across 5 folds. For identification and weight estimation YOLOv8s localization was evaluated by comparing YOLO-predicted bounding boxes against ground-truth annotations and reporting IoU and mAP@0.5 which reports an IoU range of 0.93, with mAP@0.5 of 0.97 which indicates that YOLO crops closely match GT crops. Identification performance measured using Rank-1, Rank-5 accuracy and mean Average Precision (mAP). Weight estimation evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and coefficient of determination R^2.

4. Results and Analysis

In this section, extensive experimental findings concerning the three combined tasks, which are: disease classification and severity grading, multi-view cattle recognition, and body weight estimation, are provided.

4.1. Disease Classification Performance

Table I compares the results of five modeling strategies in the disease classification case (4-class), the severity staging case (3-class), and the hierarchical joint prediction case. All values are given as the mean ± standard deviation of 5-fold cross-validation with stratification.

Cross-attentional multi-task model had a superior outcome in all levels of evaluation with disease classification accuracy of 89.96% severity grading accuracy of 83.75% and hierarchical joint prediction accuracy of 85.88%. The standard deviations are low which means that the performance is stable when the data is divided into various splits. Table II offers all-inclusive performance indicators such as precision, recall, and F1-score for the best performing case of cross-attentional multi-task.

The balanced precision-recall trade-off indicates the model performs equally well at minimizing false positives and false negatives, which is critical for clinical decision support. In the figure 8 analysis of the confusion matrix showed that adjacent-stage confusion was the most common cause of severity errors, while cross-disease misclassifications were the least common and misclassifying healthy samples as diseased was also a notable error pattern and is important from a deployment perspective due to unnecessary alerts. Lastly, the figure 9 in the loss curve demonstrated stable convergence with no major divergence between training and validation, indicating good generalization to unseen samples.

4.2. Cattle Identification Results

Table III presents identification performance for ResNet-50 and ConvNeXt-Tiny under Protocol A (leave-one-view-out) and Protocol B (cross-angle domain shift evaluation).

with an IoU of 0.93 and mAP@0.5 of 0.97.

When using Protocol A, both architectures demonstrated high accuracy of Rank-1 which indicated high view-invariant representation learning. ConvNeXt-Tiny slightly outperformed ResNet-50 (97.12 in comparison against 96.40), and the better feature normalization and depthwise convolutional design. The results of Protocol B indicate that the performance would decline with the domain shift, and Rank-1 accuracy will decrease to 72-75%. But Rank-5 accuracy was still quite good (87-90%), meaning that correct identities are very likely to be found in the top-5 retrievals, which implies that meaningful similarity structure is preserved in the embedding space even with a change of viewpoint.

The figure 10 in the loss curve demonstrated stable convergence with no major divergence between training and validation, indicating good generalization to unseen samples.

Table IV compares identification performance using ground-truth versus YOLO-detected crops. YOLO-based cropping introduces minimal performance degradation (< 0.3% for Protocol A, < 2% for Protocol B), confirming the detection pipeline's suitability for end-to-end automated deployment.

Table I
Disease classification and severity grading performance (mean ± std across 5 folds)

Model	Disease Acc. (%)	Severity Acc. (%)	Hierarchical Acc. (%)
Flat 10-class	82.45 ± 1.32	78.21 ± 2.15	79.12 ± 1.87
Cascaded 2-stage	86.73 ± 1.08	81.34 ± 1.92	82.56 ± 1.45
Shared-backbone	87.89 ± 0.95	82.47 ± 1.68	84.23 ± 1.21
Cross-attentional	**89.96 ± 0.87**	**83.75 ± 1.54**	**85.88 ± 1.09**
Ordinal regression	88.21 ± 1.12	82.93 ± 1.76	84.67 ± 1.33

Table II
Detailed metrics for cross-attentional model (mean ± std)

Task	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Disease (4-class)	89.96 ± 0.87	89.23 ± 0.94	88.76 ± 1.02	88.98 ± 0.91
Severity (3-class)	83.75 ± 1.54	82.94 ± 1.67	82.13 ± 1.89	82.52 ± 1.71
Hierarchical (10-class)	85.88 ± 1.09	84.67 ± 1.24	83.92 ± 1.36	84.28 ± 1.19

4.3. Body Weight Estimation Performance

Table V presents comprehensive weight estimation performance for six CNN architectures under frozen and fine-tuned training strategies.

The individual model performance of DenseNet121-tuned was the best with a MAE of 36.99 kg (7.78% MAPE) because it had a dense connectivity to facilitate effective multi-view feature fusion. Also, the performance of all the architectures was always increased with fine-tuning, but the most significant reduction (10.60 kg MAE) was observed with InceptionV3, which means that its multi-scale characteristics need domain-specific adjustment. The ensemble model that used DenseNet121, EfficientNet-B1 and ConvNeXt-Tiny with MAE of 35.10 kg illustrated that these architectures complement each other in their error patterns.

Table VI compares the performance of the various weight categories to determine model bias. The highest performance is observed in the range of 150-450 kg which has the highest data density. The higher error in heavy bulls (450+ kg) is due to breaking the dataset balance and morphological diversity whereas the higher error of calves is attributed to the error being proportionately larger among smaller animals.

5. Comparisons and Relationships

This section depicts a comparison analysis of the existing works with our work to evaluate the actual aspects where our solutions might be effective or not with reasoning.

5.1. Comparative Performance Analysis

Table VII relates the suggested system to the current cattle disease detection research. The suggested system provides competitive levels of disease classification, and it is the only system to offer clinically-validated levels of severity grading for multiple diseases, capable of doing so and not existing in other methods. This facilitates priority of treatment and monitoring of progression of the disease.

Table VIII relates the suggested multi-view identification system to the latest cattle recognition research. The proposed system has maximum Rank-1 accuracy (96.40%) in controlled multi-view conditions, which is better than the past multi-view systems. [13]. The results of protocol B (74.56%) are realistic cross-angle performance, which is not considered in the existing literature under many situations.

Table IX compares the proposed weight estimation approach against recent RGB-based and depth-based methods.

Table III
Multi-view cattle identification performance comparison

Protocol	Model	R-1	R-5	R-10	mAP
Protocol A (LOVO)	ResNet-50	96.40	100.00	100.00	98.73
	ConvNeXt-Tiny	97.12	100.00	100.00	99.05
Protocol B (Cross-angle)	ResNet-50	72.34	87.45	93.67	80.21
	ConvNeXt-Tiny	74.56	89.80	95.12	82.45

Table IV
Impact of detection method on identification accuracy

Crop Method	Protocol A Rank-1 (%)	Protocol B Rank-1 (%)	mAP A (%)	mAP B (%)
Ground Truth	97.12	76.23	99.05	83.78
YOLO Detected	96.85	74.56	98.67	82.45
Performance Gap	0.27	1.67	0.38	1.33

Table V
Comprehensive performance metrics for all models (sorted by validation MAE)

Model	Type	MAE	MAPE	RMSE	R^2
ResNet50	Frozen	42.35	8.92	54.21	0.854
ResNet50	Tuned	39.87	8.41	51.34	0.871
EfficientNet-B1	Frozen	40.12	8.45	52.18	0.868
EfficientNet-B1	Tuned	38.23	8.05	49.56	0.883
DenseNet121	Frozen	38.76	8.12	49.87	0.879
DenseNet121	Tuned	**36.99**	**7.78**	**47.23**	**0.892**
InceptionV3	Frozen	50.66	10.67	63.45	0.792
InceptionV3	Tuned	40.06	8.44	51.89	0.865
MobileNetV2	Frozen	43.82	9.23	56.12	0.842
MobileNetV2	Tuned	41.54	8.75	53.67	0.858
ConvNeXt-T	Frozen	39.45	8.31	50.78	0.873
ConvNeXt-T	Tuned	37.82	7.96	48.91	0.886
Ensemble	**Avg**	**35.10**	**7.39**	**45.67**	**0.901**

Table VI
Weight estimation performance by weight range (DenseNet121-tuned)

Weight Range (kg)	Samples	MAE (kg)	MAPE (%)	R^2
50-150 (Calves)	42	28.34	12.45	0.7634
150-300 (Young adults)	98	32.67	6.89	0.8912
300-450 (Adults)	58	38.91	6.23	0.9134
450+ (Heavy bulls)	17	52.78	8.67	0.8456
Overall	**215**	**36.99**	**7.78**	**0.8921**

The proposed system achieves competitive performance using only RGB images without depth sensors or reference objects. The ensemble approach (MAE 35.10 kg) outperforms most RGB-only methods and approaches depth-assisted performance demonstrating the effectiveness of multi-view fusion and metadata integration.

5.2. Key Findings

Disease Module: Cross-attentional architecture showed higher performance over cascaded and flat architecture where feature learning in a task-specific manner was achieved and hierarchical consistency maintained. Patterns of errors were dominated by adjacent-stage confusion (68% of severity errors) implying that future research undertakings would be better directed towards more delicate visual cues to differentiate severity levels. In successful cases figure 11, Grad-CAM++ heatmaps concentrate on disease-specific pathology regions (e.g., vesicular lesions, nodules, or the eye area), while in unsuccessful cases figure 12 the model typically attends to the correct region but misgrades severity

Work	Input	Dataset details	Accuracy
Our work	Multi-task with cross-task attention and gated fusion (Option-5); RGB images	(Healthy + FMD, IBK, LSD); 7,254 images	89.96±0.23% (Disease Acc)
O. Andurkar et al. [39]	Images + IoT/physio sensors	3 classes (FMD, LSD, IBK)	99.13%
Lavanya et al. [40]	Images	3 classes (Normal, mild LSD, severe for LSD only); 450 images	95.7%
Phulu et al. [2]	Symptom ontology data + labeled cow images	2 classes (Healthy/Infected)	91.1%
R. M. D. S. M. Chandrarathna et al. [41]	Images + videos	2 classes (FMD, Bovine Johne's); 454 images	94% (FMD), 99% (Bovine Johne's)
S. M. Saqib et al. [8]	Images	2 classes (Healthy vs LSD); 793 images (464 healthy + 329 LSD)	95%

Table VII
Comparison with state-of-the-art cattle disease classification methods

Method	Dataset	Input	Protocol A Rank-1 (%)	Protocol B Rank-1 (%)
Our Work (ConvNeXt-Tiny)	215 cattle (4 views)	Multi-view RGB (YOLO)	**96.40**	**74.56**
Our Work (ResNet-50)	215 cattle (4 views)	Multi-view RGB (YOLO)	92.38	69.06
Bergamini et al. [13] (Custom DCNN)	439 cattle	Multi-view RGB	81.7	–
Qiao et al. [19] (CNN+LSTM)	41 cattle	Single-view RGB	91	–
Li et al. [6] PP-LCNetV3,FaceNet	176 cattle	Face, Muzzle, Eartag	95.74	–
Li et al. [10]	268 cattle	Muzzle RGB	98.7	–
Zhao and Lian [42] YOLOV8sdetect cows	200 data-points	RGB images	90.3	–
Mon et al. [43] VGG16	1,263 cattle (3 farms)	RGB video	96.34	–

Table VIII
Comparison with state-of-the-art cattle identification methods

Work	Method	Dataset	Input	MAE
Our Work	Ensemble	215 Cattles	4-view RGB Images	35.10
Our Work	DN121-Tuned	215 Cattles	4-view RGB Images	36.99
Afridi et al. [44]	Stereo 3D mesh; PointNet segmentation; regression	Real-farm incomplete 3D shapes	RGB images (2D)	25.2
Dang et al. [23]	PointNet segmentation; CatBoost/LightGBM/Polynomial Regression/RF/XGBoost	1,190 point cloud meshes (270 cattle, various postures)	Top-view + Side-view RGB	25.2
M. J. Hossain et al. [22]	Custom CNN vs EfficientNetB3 + ML baselines; YOLOv5 detection; RFE; LIME (XAI)	2D RGB cattle images	Side-view images	18.02
P. Nilchuen et al. [25]	YOLOv11m for hip depth & body length extraction + linear/multivariate regression	Brahman cattle; Dataset-1: 12,660 side-view images; Dataset-2: 523 cattle	Side-view 2D RGB smartphone images	43.44

Table IX
Comparison with state-of-the-art cattle weight estimation methods

by confusing adjacent stages due to subtle visual differences and progression boundaries.

Identification Module: Identification Module: Metric learning using combined ArcFace+SupCon losses: This achieved better embedding discrimination. Multiview training was much more resistant to changes in viewpoint with perfect Rank-5 accuracy (100%) in Protocol A Protocol-based evaluation showed that cross-angle generalization is still a problem (Rank-1: 74.56%) and needs the increased diversity of view in training data. The automated cropping based on the YOLO algorithm added the minimal degradation (>2%), which confirms the possibility of end-to-end deployment.

Weight Module: DenseNet121 has dense connectivity that enabled easy integration of multi view features, which was better than other architectures. Fine-tuning yielded uniform gains when compared to frozen transfer learning and InceptionV3 was the most benefiting domain adaptation (10.60 kg MAE reduction). Ensemble learning minimized the variance in prediction and recorded the highest overall performance (35.10 kg MAE). Extreme weight range performance degradation indicates the necessity to use balanced data collection, and metadata integration enhanced prediction consistency over morphological differences.

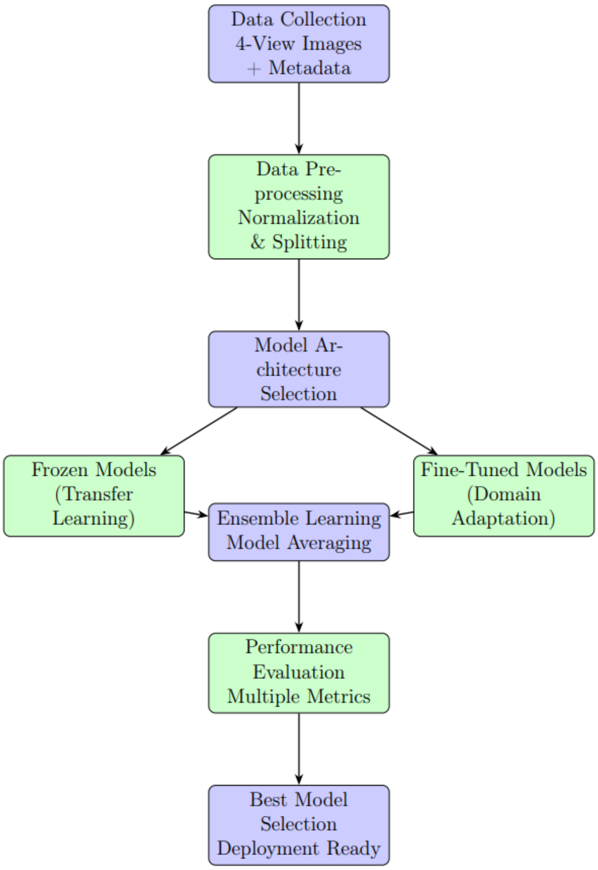

Figure 7: Methodology overview for cattle weight estimation system

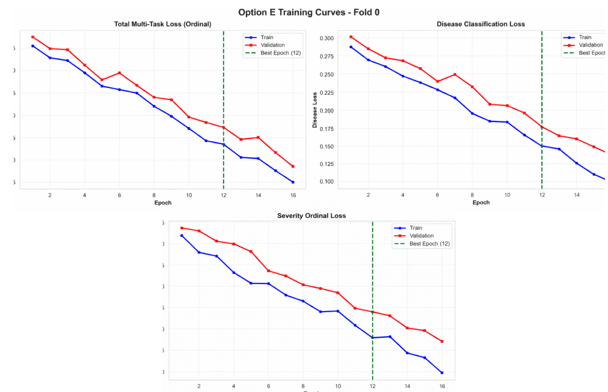

Figure 9: Loss curve of disease best performing model (Option-E)

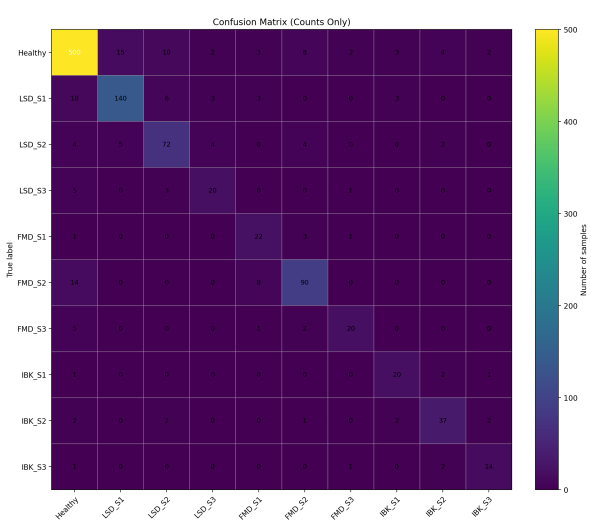

Figure 8: Confusion matrix of disease classification

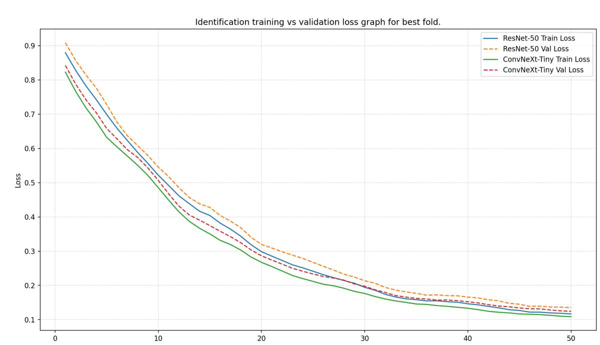

Figure 10: Train vs Validation loss curve of cattle identifcaition

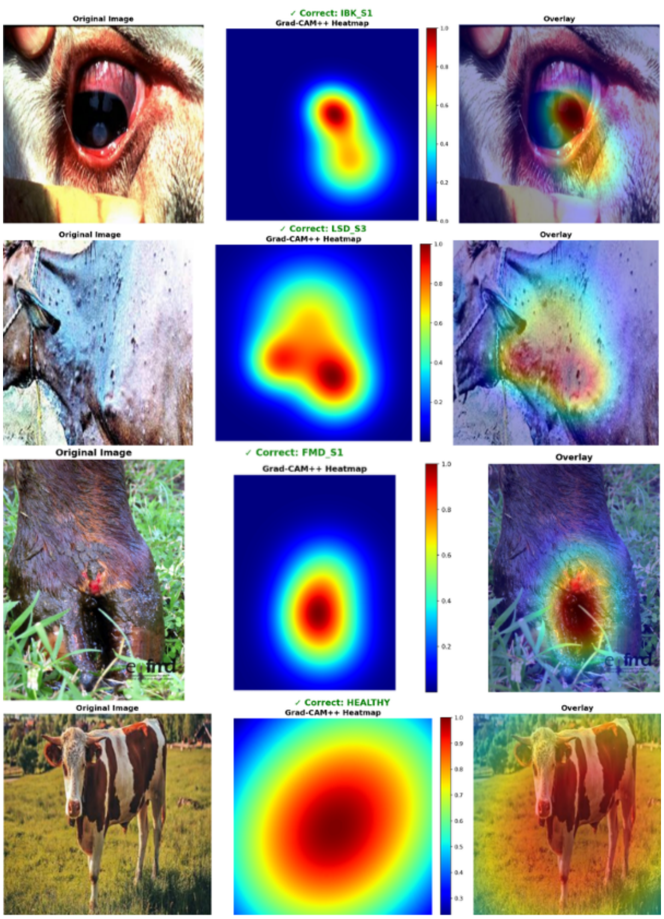

Figure 11: Grad-CAM++ successful prediction example of disease classification

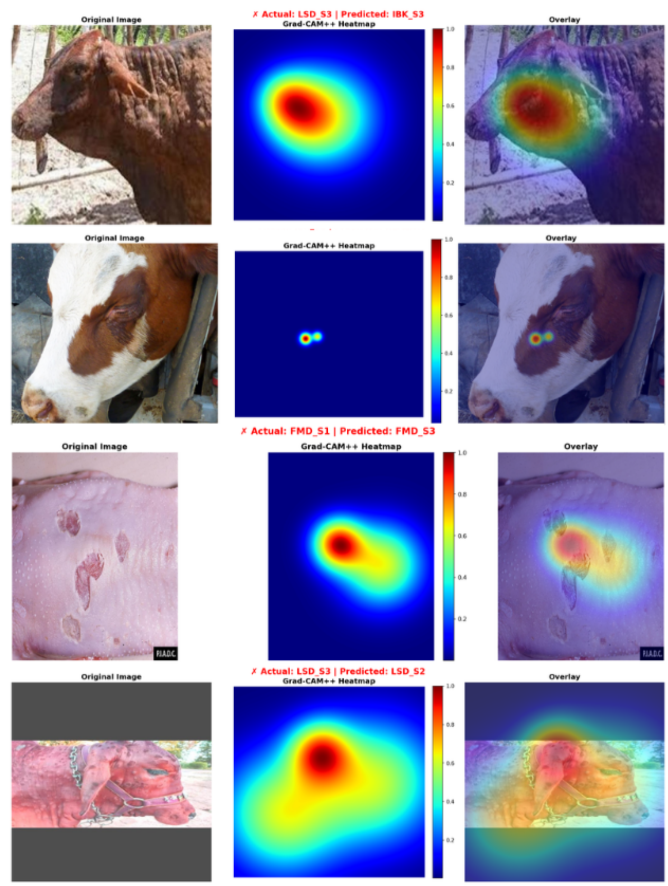

Figure 12: Grad-CAM++ unsuccessful prediction example of disease classification

6. Discussion

6.1. System Performance

The suggested cattle monitoring ecosystem based on computer vision proves that the RGB imaging can be used to offer a viable, economical and efficient solution to precision livestock farming under resource-restricted conditions. The combination of three complementary functions: disease classification and severity grading, multi-view identification, and weight estimation in a single framework is of great benefits compared to single-task systems.

The hierarchical disease classification methodology fulfills a very significant gap in the literature as it offers clinically-tested severity grading in addition to detection of the disease. The 89.96% accuracy of the disease classification is competitive with single-disease specialized systems, and most importantly the 83.75% severity staging feature provides the ability to prioritize the treatment and allocate resources. The hypothesis that the superior performance of the cross-attentional architecture compared to independent or cascaded techniques occurs because of the bi-directional exchange of information between disease and severity prediction tasks is supported by the superior performance of the cross architectural design in feature learning.

Under controlled conditions, the multi-view identification system delivered the state of art performance (96.40% Rank-1) that was superior to the performance of the multi-view systems in the past. The protocol-based systematic evaluation indicates that important issues are related to deployment: leave-one-view-out testing is clearly shown to be the best view-invariant learner, but the cross-angle domain shift protocol (74.56% Rank-1) reminds that there is still room to be desired when the camera view between enrollment and recognition is vastly different. Such performance disparity indicates that when it comes to practical deployments, the concept of standardizing the camera angles needs to be considered, or alternatively, different viewing angles are to be introduced in the course of training.

The weight estimation results show that a four-shot, multi-view RGB setup enriched with metadata can approach the accuracy range of depth or 3D-based systems without incurring the added hardware and capture complexity. Our best ensemble achieves 35.10 kg MAE under a strict low-shot constraint, whereas many RGB-based multi-view estimators that report lower MAE typically depend on collecting many more images per cow or repeated captures from similar views to reduce noise and average out pose variability. This positions our work as a complementary contribution that emphasizes the practical trade-off between error minimization and deployability.

6.2. Practical Implications and Deployment Considerations

The computational efficiency analysis evidences that the system is feasible to execute edge deployment in the case of agricultural environments. The system can fit on low-cost edge devices (NVIDIA Jetson, Raspberry Pi with accelerators) with a total model size of less than 180 MB, and inference time of less than 150 ms per animal without the need to have to be always connected to the cloud. This deals with a major challenge to the implementation in rural settings with low internet connectivity.

The slight drop in performance in automated YOLO detection (below 2 percent in all tasks) over automated detection justifies the end-to-end automation, where no manual annotation is needed to deploy the system. The scaling to large herds and continuous monitoring that require manual intervention is considered impractical and so this automation is necessary.

Nevertheless, several deployment issues should be highlighted. For the disease module, the adjacent-stage confusion rate indicates that the model often struggles to separate consecutive severity levels. Besides, its dominant error trend was confusing healthy samples as diseased rather than confusing between different diseases, which can be clinically critical due to unnecessary alarms and treatment actions. In practice, the system is therefore most reliable for detecting disease presence and flagging clearly severe cases, while intermediate stages should be verified by a veterinarian. For identification, the observed cross-angle performance degradation emphasizes the need for standardized imaging protocols during both enrollment and recognition to reduce viewpoint mismatch. Finally, the weight module shows limited accuracy in deployment-oriented settings, with MAE still relatively high and performance degrading for outlier cattle, suggesting that broader and more balanced data coverage is necessary to improve robustness at the extremes.

7. Limitations

In **disease classification**, The system is constrained by data factors. Many samples are **single-view** images, so lesions outside the captured region may be missed, which can reduce sensitivity and make severity estimation unreliable when disease extent is not fully visible. The dataset also reflects relatively **curated capture conditions**, so real field noise (low light, blur, mud, occlusion, pose changes) may degrade deployment performance. **Class imbalance** (healthy-dominant) can bias predictions toward negative cases. Finally, **severity staging is inherently ambiguous** because progression is continuous but labels are discrete (Stage 1/2/3), so adjacent-stage confusion is expected even when attention is on the correct region, and visual cues alone may not fully capture clinical severity without additional signs/tests.

The **cattle identification** module still has several limitations: firstly, the identification dataset is relatively small (only 215 cattle, with 4 views per cow), which can limit generalization as herd size and visual diversity grow. Secondly, the current recognition setup is effectively closed-set that is optimized to match a query against a fixed gallery of known identities which results in unknown reliability in handling unseen cows. Moreover, performance drops under realistic domain shift, where changes in viewpoint, lighting, pose, and partial occlusion alter appearance (evidenced by the Protocol A → Protocol B gap and the noted sensitivity). In addition, the pipeline is not designed to identify multiple cows in a single image simultaneously, and it mostly struggle to discriminate very similar-looking cattle (e.g., full black or full reddish coats) during cross-angle-view or domain shift especially under harsh lighting, shadowing, and occlusions, which can suppress fine-grained texture cues needed for reliable re-identification.

A key limitation of the proposed **cattle weight estimation** module is the small dataset size (215 samples), which increases the risk of overfitting and weak generalization to unseen cattle and farm settings. The dataset also has imbalanced weight coverage (most samples concentrated in the 200–400 kg range, with very few extreme weight animals), which can bias the regressor toward "typical" weights and reduce accuracy for calves or very heavy cattle. In addition, outliers and noisy measurements can inflate error and create unstable learning especially when rare cases are present but underrepresented. Finally, real-farm image issues such as variable lighting, occlusion, pose changes, and inconsistent capture distance or angle can degrade feature quality and prediction reliability, particularly because imaging conditions are not fully standardized.

8. Future Work

Although this study provides a solid base upon which the Computer Vision-based cattle monitoring can be based, a number of areas in which future research should be conducted are identified:

Disease module: The framework could be expanded to other cattle illnesses and parasite infections to enhance applicability and strength of the system in a wide range of farming conditions along with a enlarged dataset and class balance. Future work should reduce single-view blind spots by collecting multi-view or short video samples so lesions across the body are captured, improving sensitivity and severity reliability.

Identification module: Expanding the dataset with more cows, more views, and more variation (time gaps, lighting, occlusion) to improve generalization can be very beneficial. Moving from closed-set to open-set recognition by adding unknown cattle detection so unseen identities are not forced to match. Include explainable AI by showing similarity-based evidence (top-k matches) and attention or heatmaps highlighting the coat regions used for the decision.

Weight module: Increasing dataset size and balance weight ranges (more calves and heavy cattle) to reduce bias toward typical weights and prevent overfitting is necessary to reduce bias. To improve robustness with pose or scale, angle aware modeling (keypoints or segmentation or standardized capture cues) and uncertainty aware regression to handle noisy measurements might be crucial. Addition of explainable AI by providing region evidence for visualization plus an uncertainty score or error bar when images are low quality or occluded can be beneficial.

The system can be extended into a complete ecosystem by adding modules like breed classification, pose estimation, body condition scoring, age estimation, activity or lameness detection, estrus detection, and multi-animal tracking, enabling richer, end-to-end farm decision support.

9. Conclusion

One of the key elements of livestock management is the possibility of an integrated RGB-based monitoring pipeline that combines multi-disease classification with severity grading, multi-view individual identification, and weight estimation which can be used to form a complete computer vision driven ecosystem for cattle monitoring. This research has shown that low-shot, non-invasive monitoring can be made cost-effective using solely RGB images and deep learning to offer realistic solutions to resource-constrained agricultural settings.

The hierarchical disease classification system recognizes 89.96% of the diseases (Healthy, LSD, FMD, IBK) and 83.75% of the severity grades (Stage-1, Stage-2, Stage-3), and annotations of the clinical relevance are validated by experts. The cross-attentional multi-task architecture performed better than the flat, cascaded, and standard multi-task designs in bidirectional feature interaction between tasks. Multi-view identification has a high performance of 96.40% Rank-1 accuracy when implemented in leave-one-view-out condition and 74.56% when tested in cross-angle conditions. ArcFace and SupCon losses make the combination highly discriminative embeddings with little degradation on automated YOLO detection. Body weight estimation with ensemble learning with metadata integration has a MAE of 35.10 kg which is competitive with depth-based approaches given the use of RGB images.

Some of the important contributions includes, exploring the possibility of combined cattle monitoring ecosystem through an unified RGB-based system, hierarchical classification with annotated clinical severity by experts, protocol based evaluation framework to simulate realistic deployment, metadata enriched dataset of cattle to facilitates reproducible research, and ensemble learning to show usefulness in weight estimation in RGB-based system without any special sensors. The efficiency due to low shot RGB images in the form of computational supports the deployment of edges, which means that it even has the potential of operating offline in places where connectivity is limited. The future enlarged coverage of the diseases, switched to open-set recognition, introduced uncertainty quantification, created more monitoring modules, and practical implemented studies of multi-farms validation.

10. Ethical Statement

The study is devoted to the design of a computer vision-based decision-support system to monitor cattle health, and the area of work is closely associated with the ethical issues because of the direct influence on the welfare of animals, farm economies, and the profession of veterinarians. Everything in the proposed system is intended to focus on responsible utilization, openness, and assistance in making decisions based on knowledge instead of the automatic substitution of professional judgment. The system is aimed at helping farmers and veterinarians, giving early signs of illness, levels of severity, each patient, and estimating weight, but the ultimate diagnosis and treatment should be left under qualified veterinary care.

The study design is based on the principles of applying the concept of artificial intelligence to agriculture and animal health monitoring ethically. Specifically, it is a system that is considered a supportive tool that complements the current livestock management practices and not a way to replace the expertise of the veterinarian. Automated reports are supposed to present possible health issues and trends, but they should not prescribe or override human judgment. This practice fits into the responsible AI practices which focus on human supervision and responsibility in safety-critical applications.

All image data utilized in the study are only cattle imagery gathered through public data and controlled field data collection with proper consent of farm owners. No human subjects are used and no personally identifiable information is gathered or processed. RGB images eliminate the need to use invasive sensing methods and reduce the amount of stress or damage that can be caused to animals in the process of data acquisition. A qualified veterinary professional is required to do the disease severity annotations in order to be confident that labels indicate clinically significant and ethically defensible interpretations.

11. Credit Authorship Contribution Statement

Aabu Yousuf Raj: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project Administration. **Md. Rakibul Hasan**: Data Curation, Software, Validation, Investigation, Writing - Review & Editing. **Abdur Rahman Arafat**: Data Curation, Software, Validation, Investigation. **S.M.Sadman Rahman**: Data Curation, Validation, Investigation. **Md. Roman Bin Jalal**: Data Curation, Validation, Investigation. **Jannatun Noor**: Supervision, Resources, Writing - Review & Editing, Funding Acquisition.

12. Declaration of Competing Interest

The authors declare that no known competing financial interests or personal relationships that may have been construed to have influenced work reported herein.

13. Acknowledgements

First and foremost, we would like to thank our Creator, the Almighty Allah, for making our journey possible and removing all obstacles from our paths during our research. Secondly, we are grateful to our research supervisors, Professor Dr. Jannatun Noor Mukta Ma'am, and Dr. S.M. Mahbubur Rahman Sir for their support and advice during the research. Both of them shared their wisdom and guidelines whenever we reached out to them without any hesitation. We would also like to express our sincere gratitude to the farm owners for their voluntary support and cooperation in allowing us to collect the data for this thesis. Finally, to our parents, friends, and acquaintances. It wouldn't have been possible for us to continue this journey without their prayer and kind support.

14. Data Availability

The datasets used in this study include both custom-collected and publicly available sources. Custom-collected datasets may be made available upon reasonable request and subject to approval from participating farms and ethical review. Code and model weights will be shared through appropriate repositories to support reproducibility and further research.

References

[1] Zayyin Dinana, Fedik Abdul Rantam, Suwarno Suwarno, Imam Mustofa, Jola Rahmahani, and Kusnoto Kusnoto. Detection of foot and mouth disease virus in cattle in lamongan and surabaya, indonesia using RT-PCR method. *Jurnal Medik Veteriner*, 6(2):191–196, 2023. doi: 10.20473/jmv.vol6.iss2.2023.191-196. URL https://doi.org/10.20473/jmv.vol6.iss2.2023.191-196.

[2] Shwele Phulu, Mateus N. Natanael, Naftali N. Indongo, Angeline Mukamana, Metumo Shifidi, and Joseph Musenge. A system for early detection of foot and mouth disease in cattle using machine learning. In *2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, pages 1–6, 2024. doi: 10.1109/ETNCC63262.2024.10767515.

[3] Bhuiyan Mustafa Tawheed, Syed Tahmid Masud, MD. Shajedul Islam, Hossain Arif, and Samiul Islam. Application of machine learning techniques in the context of livestock. In *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, pages 2029–2033, 2019. doi: 10.1109/TENCON.2019.8929721.

[4] Priya Bhardwaj, S J K Jagadeesh Kumar, G. Prabu Kanna, and A. Mithila. Machine learning based approaches for livestock symptoms and diseases prediction and classification. In *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)*, pages 1–6, 2024. doi: 10.1109/IC3SE62002.2024.10593244.

[5] R. Newlin Shebiah and S. Arivazhagan. Deep learning based image analysis for classification of foot and mouth disease in cattle. In *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, pages 701–705, 2023. doi: 10.1109/ICIRCA57980.2023.10220765.

[6] Dongxu Li, Baoshan Li, Qi Li, Yueming Wang, Mei Yang, and Mingshuo Han. Cattle identification based on multiple feature decision layer fusion. *Scientific Reports*, 14(1):26631, 2024. doi: 10.1038/s41598-024-76718-x. URL https://doi.org/10.1038/s41598-024-76718-x.

[7] Weihong Ma, Xiangyu Qi, Yi Sun, Ronghua Gao, Luyu Ding, Rong Wang, Cheng Peng, Jun Zhang, Jianwei Wu, Zhankang Xu, Mingyu Li, Hongyan Zhao, Shudong Huang, and Qifeng Li. Computer vision-based measurement techniques for livestock body dimension and weight: A review. *Agriculture*, 14(2), 2024. ISSN 2077-0472. doi: 10.3390/agriculture14020306. URL https://www.mdpi.com/2077-0472/14/2/306.

[8] Syed Muhammad Saqib, Muhammad Iqbal, Mohamed Tahar Ben Othman, Tehreem Shahazad, Yazeed Yasin Ghadi, Shamsul Haque, Salman A. AlQahtani, and Nadia Gul. Lumpy skin disease diagnosis in cattle: A deep learning approach optimized with RMSProp and MobileNetV2. *PLOS ONE*, 19(8):e0302862, 2024. doi: 10.1371/journal.pone.0302862. URL https://doi.org/10.1371/journal.pone.0302862.

[9] Asim Iqbal, Narit Hnoohom, Sakorn Mekraksavanich, and Morakot Kaewthamasorn. A deep contrastive learning-based image retrieval system for automatic detection of infectious cattle diseases. *Journal of Big Data*, 12(1), 2025. doi: 10.1186/s40537-024-01057-7. URL https://doi.org/10.1186/s40537-024-01057-7.

[10] G. Li, G. E. Erickson, and Y. Xiong. Individual beef cattle identification using muzzle images and deep learning techniques. *Animals*, 12 (11):1–17, 2022. doi: 10.3390/ani12111453.

[11] Beibei Xu, Yifan Mao, Wensheng Wang, and Guipeng Chen. Intelligent weight prediction of cows based on semantic segmentation and back propagation neural network. *Frontiers in Artificial Intelligence*, Volume 7 - 2024:56789–56798, 2024. doi: 10.3389/frai.2024.1299169.

[12] A. Girmaw. Livestock animal skin disease detection and classification using deep learning approaches. *IEEE Trans. Agri. Food Electron.*, 3 (1):67–76, 2025. doi: 10.1016/j.bspc.2024.107334.

[13] Luca Bergamini, Angelo Porrello, Andrea Capobianco Dondona, Ercole Del Negro, Mauro Mattioli, Nicola D'alterio, and Simone Calderara. Multi-views embedding for cattle re-identification. In *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 184–191, 2018. doi: 10.1109/SITIS.2018.00036.

[14] Alexey Ruchay, Vitaly Kober, Konstantin Dorofeev, Vladimir Kolpakov, and Sergei Miroshnikov. Accurate body measurement of live cattle using three depth cameras and non-rigid 3-d shape recovery. *Computers and Electronics in Agriculture*, 179:105821, 2020. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2020.105821.

[15] Alexey Ruchay, Vitaly Kober, Konstantin Dorofeev, Vladimir Kolpakov, Alexey Gladkov, and Hao Guo. Live weight prediction of cattle based on deep regression of rgb-d images. *Agriculture*, 12(11), 2022. ISSN 2077-0472. doi: 10.3390/agriculture12111794. URL https://www.mdpi.com/2077-0472/12/11/1794.

[16] M. Matvieiev, H. Bila, I. Ugnivenko, D. Nosevych, M. Loban, and T. L. Tyasi. Application of data mining algorithms for estimating live body weight from linear body measurements of Ukrainian beef cattle breed. *Applied Ecology and Environmental Research*, 23(2): 1853–1864, 2025. doi: 10.15666/aeer/2302_18531864. URL https://doi.org/10.15666/aeer/2302_18531864.

[17] Harsh J. Shah, Chirag Sharma, and Chirag Joshi. Cattle disease prediction using artificial intelligence. *International Journal for Research in Applied Science and Engineering Technology (IJRASET)*, 11(4), 2023. doi: 10.22214/ijraset.2023.50535. URL https://doi.org/10.22214/ijraset.2023.50535.

[18] M. Motohashi, T. Nakano, and H. Saito. Early detection method for subclinical mastitis in auto milking systems using machine learning. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 13:4567–4575, 2020. doi: 10.1109/ICCICC50026.2020.9450258.

[19] Y. Qiao, D. Su, and H. Kong. Individual cattle identification using a deep learning based framework. In *2019 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 9–16. IEEE, 2019. doi: 10.1016/j.ifacol.2019.12.558.

[20] Anjar Setiawan, Ema Utami, and Dhani Ariatmanto. Cattle body weight prediction using regression machine learning. *Jurnal Teknik Informatika (Jutif)*, 5(2):509–518, Apr. 2024. doi: 10.52436/1.jutif.2024.5.2.1521. URL https://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/1521.

[21] Roel Dohmen, Cagatay Catal, and Qingzhi Liu. Image-based body mass prediction of heifers using deep neural networks. *Biosystems Engineering*, 204:283–293, 2021. ISSN 1537-5110. doi: https://doi.org/10.1016/j.biosystemseng.2021.02.001.

[22] Md Junayed Hossain, Jannatul Ferdaus, Ashraful Islam, and M. Ashraful Amin. CattleNet-XAI: An explainable CNN framework for efficient cattle weight estimation. *PLOS ONE*, 20(11):e0336434, 2025. doi: 10.1371/journal.pone.0336434. URL https://doi.org/10.1371/journal.pone.0336434.

[23] Changgyu Dang, Seungwook Son, Jaegu Jang, Anhvu Nguyen, Thi Hai Yen Nguyen, Yooil Suh, and Daihee Park. A Korean cattle weight prediction approach using 3D segmentation-based feature extraction and regression machine learning from incomplete 3D shapes acquired from real farm environments. *Agriculture*, 13(12):2266, 2023. doi: 10.3390/agriculture13122266. URL https://doi.org/10.3390/agriculture13122266.

[24] Guilherme Botazzo Rozendo, Maichol Dadi, Marco Bovo, Daniele Torreggiani, Patrizia Tassinari, Stefano Benni, and Annalisa Franco. Cattle weight estimation using 2D side-view images and estimated depth-based 3D modeling. *Smart Agricultural Technology*, 12: 101099, 2025. doi: 10.1016/j.atech.2025.101099. URL https://doi.org/10.1016/j.atech.2025.101099.

[25] Peerayut Nilchuen, Thanathip Suwanasopee, and Skorn Koonawootrittriron. Integrating deep learning and mobile imaging for assessment of automated conformational indices and weight prediction in brahman cattle. *Smart Agricultural Technology*, 12:101079, December 2025. ISSN 2772-3755. doi: 10.1016/j.atech.2025.101079. URL https://www.sciencedirect.com/science/article/pii/S2772375525003120.

[26] shivamagarwal29. Cow lumpy disease dataset. Kaggle dataset. URL https://www.kaggle.com/datasets/shivamagarwal29/cow-lumpy-disease-dataset. Accessed: 2026-02-04.

[27] devang03mgr. Cattle diseases datasets. Kaggle dataset. URL https://www.kaggle.com/datasets/devang03mgr/cattle-diseases-datasets. Accessed: 2026-02-04.

[28] sliit. cattle diseases dataset. https://universe.roboflow.com/sliit-kuemd/cattle-diseases, jul 2023. URL https://universe.roboflow.com/sliit-kuemd/cattle-diseases. visited on 2026-01-27.

[29] CVAT.ai Corporation. Computer vision annotation tool (cvat). Zenodo (Version v2.4.8), June 2023. URL https://doi.org/10.5281/zenodo.8070041. Published 2023-06-22.

[30] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks, 2020. URL https://arxiv.org/abs/1905.11946.

[31] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[33] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s, 2022. URL https://arxiv.org/abs/2201.03545.

[34] Beibei Xu, Wensheng Wang, Leifeng Guo, Guipeng Chen, Yongfeng Li, Zhen Cao, and Saisai Wu. Cattlefacenet: A cattle face identification approach based on retinaface and arcface loss. *Computers and Electronics in Agriculture*, 193:106675, 2022. ISSN 0168-1699. doi: https://doi.org/10.1016/j.compag.2021.106675. URL https://www.sciencedirect.com/science/article/pii/S016816992100692X.

[35] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 18661–18673. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/d89a66c7c80a29b1bdbab0f2a1a94af8-Paper.pdf.

[36] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2018. URL https://arxiv.org/abs/1608.06993.

[37] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision, 2015. URL https://arxiv.org/abs/1512.00567.

[38] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks, 2019. URL https://arxiv.org/abs/1801.04381.

[39] Onkar Andurkar, Krishna Khandelwal, Akshat Kotpalliwar, Subodh Palaspagar, and Vinayak Musale. AI-driven classification of external diseases in cattle. *International Journal of Engineering Development and Research (IJEDR)*, 13(2), 2025. URL https://rjwave.org/ijedr/papers/IJEDR2502137.pdf.

[40] V. Lavanya, A.Dinesh Kannan, S. Kamaleshwaran, R. Nishanth, and S.Muthu Kamatchi. Detecting disease in livestock and recommending medicines using machine learning. In *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, pages 517–522, 2023. doi: 10.1109/ICECAA58104.2023.10212385.

[41] R.M.D.S.M.Chandrarathna, T.W.M.S.A.Weerasinghe, N.S.Madhuranga, T.M.L.S.Thennakoon, Anjalie Gamage, and Erandika Gamage. "the taurus": Cattle breeds & diseases identification mobile application using machine learning. *International Journal of Engineering and Management Research*, 12 (6):198–205, 2022. doi: 10.31033/ijemr.12.6.27.

[42] Ancai Zhao, Huijuan Wu, Jie Zhou, Daoerji Fan, Huijie Lu, and Xiaoyu Li. Individual cow recognition based on ultra-wideband and computer vision. *Animals*, 15(3):456, 2025. doi: 10.3390/ani15030456. URL https://doi.org/10.3390/ani15030456.

[43] Su Larb Mon, Tsubasa Onizuka, Pyke Tin, Masaru Aikawa, Ikuo Kobayashi, and Thi Thi Zin. Ai-enhanced real-time cattle identification system through tracking across various environments. *Scientific Reports*, 14(1):17779, 2024. doi: 10.1038/s41598-024-68418-3. URL https://doi.org/10.1038/s41598-024-68418-3.

[44] Hina Afridi, Mohib Ullah, Øyvind Nordbø, Solvei Cottis Hoff, Siri Furre, and Anne Guro Larsgard. Analyzing data modalities for cattle weight estimation using deep learning models. *Journal of Imaging*, 10(3):72, 2024. doi: 10.3390/jimaging10030072. URL https://doi.org/10.3390/jimaging10030072.