

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

by

Aabu Yousuf Raj

24241222

MD Rakibul Hasan

22101326

Abdur Rahman Arafat

22101061

S.M.Sadman Rahman

22101008

Md. Roman Bin Jalal

22201973

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Aabu Yousuf Raj
24241222

MD Rakibul Hasan
22101326

Abdur Rahman Arafat
22101061

S.M.Sadman Rahman
22101008

Md. Roman Bin Jalal
22201973

Approval

The thesis titled “A Computer Vision driven Ecosystem for Cattle Monitoring: Multi-Disease Classification with Severity Grading, Multi-View Individual Identification, and Weight Estimation” submitted by

1. Aabu Yousuf Raj (24241222)
2. MD Rakibul Hasan (22101326)
3. Abdur Rahman Arafat (22101061)
4. S.M Sadman Rahman (22101008)
5. Md. Roman Bin Jalal (22201973)

has been accepted in Fall 2025 as partial fulfillment of the requirements for the degree of B.Sc. in Computer Science.

Examining Committee:

Supervisor:

Dr. Jannatun Noor

Associate Professor
Department of Computer Science and Engineering
United International University

Program Coordinator:
(Member)

Dr. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Head of Department:

Dr. Sadia Hamid Kazi

Associate Professor
Department of Computer Science and Engineering
School of Data and Sciences
Brac University

Ethics 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.

Abstract

Effective cattle monitoring and management are essential for sustainable dairy and beef production in Bangladesh, where constant veterinary care and extensive monitoring are economically challenging. This thesis introduces a computer vision based cattle monitoring ecosystem that (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 with publicly available sources. To analyze the disease, a hierarchical deep learning architecture is suggested to categorize Lumpy Skin Disease (LSD), Foot-and-Mouth Disease (FMD), Infectious Bovine Keratoconjunctivitis (IBK), and Healthy cattle along with grading the diseased cattle as Stage-1 (mild), Stage-2 (moderate), and Stage-3 (severe) cases. The severity grading is performed based on a clinically validated symptom profile and is supervised and confirmed by a District Livestock Officer (veterinarian). The cross-attentional multi-task architecture has the highest accuracy score of 89.96 in disease classification, 83.75 in severity staging, and 85.88 in hierarchical setup (predicts disease and then predicts severity). In the case of individual cattle recognition and weight estimation, a multi-view appearance-based recognition system is constructed based on left, right, front, and back views of cattle images where at first YOLOv8s is trained to extract the exact cattle region which have an IoU of 0.93, @map@0.5 0.97. The accuracy of the identification system is determined at Rank-1: 96.40% and 74.56% while Rank-5: 100% and 89.80% for ConvNext-Tiny on two different protocols of Leave One View Out testing method and Cross-view-angle respectively which has been done to evaluate cattle identification from different angles or possess and consider the case of cattle with very similar patterns and colors. To estimate body weight, a regression model is used with multi-view RGB images and metadata that includes id, sex, breed, age and live weight where the best single-model baseline is the DN121-tuned regressor, achieving an MAE of 36.99 KG which improves to an MAE of 35.10 KG with ensemble strategy. In general, the suggested system proves that even without the use of invasive sensors, an RGB-based, low shot, non-invasive, and affordable vision-based system can be used successfully in the diagnosis of diseases and their severity grading, the unique identification of individuals among cattle, and the estimation of their live weight.

Keywords: Computer Vision, Multi-Disease Classification, Disease Severity Grading, Deep Learning, Hierarchical Multi-Task Deep Learning with Cross-Attention, YOLOv8-based cattle detection, Multi-View Identification, Cattle Weight Estimation, Multi-view regression with metadata, ensemble learning, AI in Agriculture and low shot non-invasive low-cost RGB imaging.

Acknowledgment

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 thesis. Secondly, we are grateful to our thesis supervisors, Professor Dr.Jannatun Mukta Noor Ma'am, and Dr. S.M. Mahbubur Rahman Sir for their support and advice during the thesis. 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.

Table of Contents

Declaration	i
Approval	ii
Ethics Statement	iii
Abstract	iv
Acknowledgment	v
Table of Contents	vi
List of Figures	vii
List of Tables	viii
Nomenclature	ix
1 Introduction	1
1.1 Background	1
1.2 Rationale of the Study	2
1.3 Problem Statement	3
1.4 Objective	4
1.5 Methodology in Brief	5
1.6 Scopes and Challenges	5
1.7 Thesis Organization	6
2 Literature Review	8
2.1 Preliminaries	8
2.2 Review of Existing Research	9
2.3 Summary of Key Findings	20
3 Requirements, Impacts and Constraints	29
3.1 Final Specifications and Requirements	29
3.1.1 Data Requirements	29
3.1.2 Preparation/cleaning of data	30
3.1.3 Preprocessing Requirements	31
3.1.4 Software and Environment Requirements	31
3.1.5 Elements of Modeling and Evaluation Requirements	31
3.2 Societal Impact	32
3.3 Environmental Impact	32
3.4 Ethical Issues	33
3.5 Standards	33
3.6 Project Management Plan	33
3.7 Risk Management	35
3.8 Economic Analysis	35
4 Proposed Methodology	36

4.1 Methodology Overview	36
4.2 Preliminary Design	41
4.3 Data Collection	43
4.3.1 Data Cleaning	43
4.3.2 Data Transformation	44
4.3.3 Data Integration	47
4.3.4 Summary of Preprocessed Data	49
4.4 Implementation of Selected Design	51
5 Result Analysis	57
5.1 Performance Evaluation	57
5.2 Analysis of Design Solutions	62
5.3 Statistical Analysis	68
5.4 Comparisons and Relationships	78
5.5 Discussions	83
5.5.1 Interpretation and Implications	83
5.5.2 Limitations	86
6 Conclusion	88
6.1 Summary of Findings	88
6.2 Contributions to the Field	89
6.3 Recommendations for Future Work	90
Bibliography	98

List of Figures

3.1	Project Management Plan and Development Workflow	34
4.1	Methodology overview for cattle disease classification with severity grading system	37
4.2	Methodology overview for unique cattle identification system	39
4.3	Methodology overview for cattle weight estimation system	40
4.4	Disease dataset example after severity labeling	45
4.5	Locally acquired dataset for identification and weight estimation	46
4.6	Identification dataset example for protocol B testing.	48
4.7	Disease dataset class distribution per stage	49
4.8	Dataset characteristics and distribution for preliminary design	50
5.1	Loss curve of disease best performing model (Option-E)	69
5.2	Confusion matrix of best performing model (Option-E)	70
5.3	Protocol B test for cross angle and same color identification (Best - ConvNext-tiny).	71
5.4	IoU and Cosine similarity of Both YOLO vs GT crops shows high accuracy	72
5.5	Train vs Validation loss curve of cattle identifcaition	73
5.6	Train vs validation performance showing generalization characteristics for weight estimation	76
5.7	Error distribution boxplots showing variance and outliers across all models of weight estimation	76
5.8	Bland-Altman plot showing limits of agreement and systematic bias . . .	77
5.9	Grad-CAM++ successful prediction example of disease classification for Option-E	84
5.10	Grad-CAM++ unsuccessful prediction example of disease classification for Option-E	85

List of Tables

2.1	Summary of Key Findings	28
4.1	Summary statistics of the preprocessed cattle weight dataset	50
4.2	Fine-tuned model configurations for weight estimation	54
5.1	Performance improvements from design adjustments of weight estimation	67
5.2	Performance comparison for cattle disease classification with severity grading	68
5.3	Best hyperparameter configurations for identification selected during tuning (Fold-0 validation)	71
5.4	Interpretation of localization metrics for GT and YOLO crops	72
5.5	Identification performance on Protocol A and Protocol B using ground-truth (GT) crops and YOLO-based crops	72
5.6	Comprehensive performance metrics for all models (sorted by validation MAE)	73
5.7	Performance across different weight ranges (Ensemble 1 model)	74
5.8	Statistical robustness comparison across architectures	74
5.9	Train-validation performance gaps of weight estimation	75
5.10	Percentage error distribution of weight estimation (Ensemble 1 on validation set)	77
5.11	Comparison with state-of-the-art cattle disease classification methods . .	78
5.12	Comparison with state-of-the-art cattle identification methods	80
5.13	Comparison with state-of-the-art cattle weight estimation methods	82

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document.

ArcFace Additive Angular Margin Loss

CNN Convolutional Neural Network

CVAT Computer Vision Annotation Tool

FMD Foot-and-Mouth Disease

Grad-CAM++ Gradient-weighted Class Activation Mapping (visual explanation method)

IBK Infectious Bovine Keratoconjunctivitis

ImageNet Large-scale visual database used for pre-training CNN models

IoU Intersection over Union

k-Fold Cross-Validation Cross-validation with data split into k subsets

LSD Lumpy Skin Disease

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

MAP Mean Average Precision

Rank-1 / Rank- k Identification accuracy at rank 1 / rank k

RMSE Root Mean Square Error

SupCon Supervised Contrastive Loss

Transfer Learning Reusing pre-trained models for new tasks

YOLO You Only Look Once (object detection framework)

Chapter 1

Introduction

1.1 Background

Food security, economic stability, and rural livelihoods, particularly in developing nations like Bangladesh, largely depend on livestock health because dairy and beef production remain important to agricultural production. However, cattle diseases remain a significant threat to livestock health, which may cause weight loss and pre-mature death, as well as significant financial losses. Veterinary diagnostic techniques, which are mostly manual, laboratory, and farm-based, are not poor but are characterized by delayed diagnosis, high operational costs, restricted accessibility, and a shortage of skilled professionals. These limitations render continuous health monitoring especially difficult in rural and resource-restricted environments.

As artificial intelligence (AI) is quickly evolving, Computer Vision based livestock health monitoring is no longer a reactive and symptom-driven diagnosis, but a data-driven decision support system. Most pre-2020 studies were based on rule-based systems and classical models of Machine Learning (ML) with structured symptom data and hand-crafted features. In controlled environments, however, these methods experienced problems in scaling, robustness, multi-modal data merging, and adaptation to various farm settings. After 2021, there was a breakthrough, with Deep Learning (DL) algorithms, especially Convolution Neural Network (CNN) [46] based models are becoming more capable to detect diseases directly based on RGB images. Initial CNN-based research work formed the basics of visual livestock disease detection, which is based on acquiring discriminatory patterns of observable symptoms and thus minimizes the use of invasive procedures and diagnostic tools.

With the advancement of research since 2022, Computer Vision methods have shifted to more intensive and viable use in livestock health monitoring. The focus was moved towards real-farm issues like changes in light, the diversity of cattle poses, cluttered backgrounds, occlusion, and image quality. DL of images progressed to a more detailed pattern of disease detection, behavioral patterns, and micro-level visual features of health loss. As of 2023, lightweight CNN models like MobileNetV2 have achieved acceptable real-time capabilities in detecting disease in the field, specifically visually distinguishable diseases like Lumpy Skin Disease (LSD), Foot-and-Mouth Disease (FMD), Infectious Bovine Keratoconjunctivitis (IBK), and are found to be appropriate to run on edge and mobile devices, even in rural settings. As of 2024, large-scale livestock monitoring systems were more frequently based on integrated mobile and web-based systems, with models

like EfficientNet, DenseNet, and CNN-LSTM hybrids reporting high diagnostic accuracy, although making explainability and deployment usability primary considerations. At the same time, cattle identification with the use of biometric features of four-sided and multi-view images with the help of Computer Vision, it has become a stable and inexpensive solution compared to traditional methods of cattle identification, such as RFID, which allows showing the whereabouts of individual cattle in real farm conditions [4, 33].

In addition to disease diagnosis and identification, cattle weight estimation through RGB image has been used as a substitute for the traditional weighing systems that are expensive, cause stress to the animals, and are not feasible in small or rural farms. Even though the previous research has shown that regression models based on the RGB image can be used to estimate the live body weight with a reasonable range of error, they do not perform as well when the conditions are not under controlled conditions, and are constrained by the specifics of the situation [37].

Even with all these developments, the majority of the current studies concentrate only on particular tasks, like single-disease classification single-view or muzzle based identification, 3D reconstruction or depth sensors weight estimation, without providing a unified and scalable monitoring model. In addition, grading of disease severity, strong multi-view individual identification in harsh visual situations, low shot multi-view identification and weight estimating along with additional help of metadata are still under inspected in real-life applications.

In order to bridge such gaps, this thesis has suggested a Computer Vision-based cattle monitoring ecosystem that incorporates multi-disease classification and severity grading, low shot RGB image based four-view identification of individual cows, and estimation of body weight. The proposed system will provide a low-cost, scalable, and feasible solution to precise livestock health management with unique datasets and professional veterinary annotations, which is relevant especially in resource-limited agricultural setups.

1.2 Rationale of the Study

Although there has been a great advance in the technology of artificial intelligence-based livestock surveillance, cattle health management is not easy, especially in developing states like Bangladesh, where the coverage of veterinary supervision is usually low. Conventional methods in disease diagnosis are manual examination and laboratory testing, which are not only time-consuming but also costly and unfeasible on large farms. As a result of this, diseases are often discovered at advanced stages, leading to higher costs and an increased risk of transmission of diseases to cattle.

Previous studies of livestock disease forecasting concentrated mainly on classical ML forecasts like Decision Trees, Random Forests, and Support Vector Machines on structured symptom-based or records-based data. Even though these techniques proved to be efficient in controlled scenarios, they were not strong enough when applied to a range of farm scenarios and were not able to handle unstructured visual information [30, 49]. This proves the farm and breed bias of the currently available studies. Due to the advent of DL, Computer Vision-based systems demonstrated better results in diagnosing cattle diseases that are visually evident, such as FMD and LSD, using RGB images [61]. But the majority of current systems are confined to binary or low-cardinality multi-class dis-

ease detection and do not offer clinically significant information on disease severity that is required to effectively intervene in veterinary work and manage the epidemic. This weakness decreases the generalization of existing models to the conditions of real farms, where various diseases can be present and present identical visual symptoms.

Besides the diagnosis of the disease, livestock management needs proper identification of the cattle. The traditional means of identification, like ear tags and RFID system can be lost, damaged, and may lead to problems with a problem of handling. Recent research has shown that four-sided visual features of cattle recognition can offer reliable results without invasive methods but require so many multi-view images [33]. However, single-view identification is still susceptible to pose changes, light variations, and similarity of the animals. This has encouraged the establishment of multi-view identification systems that combine the left, the right, front, and back views to improve the robustness in the real farm environment situation [4].

Proper estimation of cattle weight is also an aspect of herd management that is very essential as it determines the feeding strategies, growth control, health evaluation, and drug dose. Although weight estimation methods relying on vision have been promising when performed under controlled conditions, most of the current methods have scalability and generalization problems. In addition, the majority of the approaches do not take into account the cattle-level data, including age, sex and breed, which have a strong impact on body morphology, and resort to visual regression only.

The necessity to prepare a convenient low shot RGB-based multi-view weight estimation model based on structured cattle metadata in real farm conditions is still not explored properly. This is essential to refine the strength, readability, and utility of vision-based weight prediction in real-life livestock management by combining visual attributes with cattle-level data.

1.3 Problem Statement

Cattle diseases still pose a significant threat to the livelihoods of the farmers, food security, and the health of livestock. The existing methodology in veterinary care is mostly reactive and is based on manual examination and visual symptoms that are frequently exhibited when the disease is already advanced. Even though image classification models that are built with DL have proven to be effective in identifying diseases that are visually manifested, including FMD and mastitis, they tend to detect diseases, but not the severity of diseases. The limitation minimizes their clinical usefulness in prioritization of treatment, understanding the severity or seriousness of the disease and effective methods of containing the disease [27].

To be more precise, as of now, there is no single system based on RGB images that is able to detect FMD, LSD, IBK, and healthy cattle at the same time, and depicts the scenario of the severe grade about the level of infection. Consequently, the current models provide only a partial diagnostic picture and cannot represent the practical situation with cattle health in the real world, when several diseases with different levels of severity can be present.

Moreover, good-quality disease monitoring needs to be regularly performed through the identification of individual animals to be used in longitudinal monitoring and target in-

tervention. Although the method of cattle recognition based on appearance has demonstrated positive outcomes, the accuracy of recognition is affected by changes in viewpoint, changes in light, and among similar animals on the same visual when single views data is considered [33]. This can be seen to emphasize the necessity of strong multi-view identification systems that can capitalize on the complementary data of left, right, frontal, and back images to provide reliable individual recognition.

Correspondingly, estimation of cattle weight has not been exploited in real-life situations because of the use of weighing platforms that are costly or specialized sensors. Current vision-based methods tend to be less generalizable to different farms, breeds, and environmental factors, in large part because of the lack of data variety and no cattle-level metadata [40]. Moreover, the majority of the studies assess weight estimation regardless of cattle identity without taking into account some valuable characteristics, such as sex and breed, that affect growth pattern and body structure. Those constraints suggest that a scalable and deployable RGB image-based weight estimation system based on real-world metadata and supplemented with individual-level labeling will be required.

This highlights the need for a method in RGB image-based disease classification with severity grading, identification and weight estimation based on custom-collected field data that implies the way of finding reasonable, scalable, and easily deployable solutions for real-life livestock management.

1.4 Objective

The main aim of the study is to create a complete Computer Vision-based cattle monitoring ecosystem that can be used to diagnose diseases and their severity, identify each cow individually, and estimate their weight using RGB images.

The specific objectives are:

- To develop and execute a hierarchical disease classification model that will classify LSD, FMD, IBK, and healthy cattle, with further classification of disease cases as mild, moderate, and severe after expert veterinary annotations.
- To create a four-sided multi-view system of cattle identification with the use of left, right, front, and back pictures to obtain strong individual recognition in the real farm environment.
- To construct an RGB image-based four-sided multi-view weight estimation model that can provide estimates of cattle body weight within reasonable error tolerances.
- Perform the tasks on a custom collected dataset that is not farm or breed dependent and to measure the proposed system with the help of the right performance measures, such as accuracy, precision, recall, F1-score, and confusion matrices in classifying the disease; Rank-1, Rank-k accuracy and mAP in identification; and MAE, MAPE, RMSE, and R^2 in estimating the weights.

1.5 Methodology in Brief

The suggested methodology involves three related things, which include multi-disease classification and severity grading, multi-viewing individual cattle identification, and RGB-based body weight estimation. Despite the independence of each component being developed and tested, they are all constructed in such a way that these tasks can be possibly unified in a single Computer Vision framework if necessary to guarantee consistency, scalability, and practical deployability.

Healthy cattle and cattle with LSD, FMD, and IBK on the actual farm are used to acquire curated merged custom RGB image datasets. The detection of diseases is modeled as a hierarchical classification problem, in which the disease type is determined, followed by the severity of the disease, with the severity being classified as mild, moderate, and severe according to the annotation of a skilled veterinarian. The data augmentation and standard image pre-processing methods are used to enhance resistance to the variation of the environment.

Individual cattle identification takes the shape of an appearance-based re-identification test that makes use of four-sided body images that are taken in left, right, frontal, and back views. DL models are trained to discover discriminative identity representations that are viewpoint and environment-independent. Testing is done based on protocol testing strategies. Protocol A conforms to the leave-one-view-out method, in which one view is omitted in the training process and applied only to testing to evaluate the ability to generalize to the unknown viewpoints. Protocol B measures recognition in controlled multi-view conditions, cross-angle and domain shift, that is, more than one view is available during testing and training, but not the four-sided images.

To estimate body weight, four sided images of cattles and supplementary information such as live weight, age, sex, and breed has been utilized. The process of weight estimation is formulated as a regression problem, and visual body features taken out of the RGB images along with the metadata are then used to predict weight. In the case of multi-view samples, it is used, and complementary shape information is used by convolution to enhance the stability of prediction.

1.6 Scopes and Challenges

Scope:

- Creation of an RGB image-based system for classifying multiple diseases with severity grading.
- Cattle identification using multiple views, left side, right side, front side, and back side.
- Estimation of cattle weight using the RGB images without specialized sensing devices.
- Testing of all the parts of the system using a custom collected dataset that is not farm or breed independent.

Challenges:

- Low access to labeled data to grade the severity of a disease, and an imbalance between the different levels of severity.
- Differences in farm settings with regard to lighting, pose, and background.
- The inability to distinguish visually similar cattle when being identified.
- A balance between model accuracy and computational efficiency to make it applicable in the real world.
- The security concern regarding data collection, low lighting, pose difference and occlusion issue in the dataset.

Nevertheless, despite those obstacles, the suggested system should offer a scalable and practical solution to the problem of precision monitoring of livestock, which will lead to better welfare of cows and the efficiency in the management of the farm.

1.7 Thesis Organization

The present thesis presents a computer-vision-driven ecosystem for cattle monitoring that explores three core functions: multi-disease classification with severity grading, multi-view individual cattle identification, and RGB-based body weight estimation. Chapter-2 provides an extensive literature review that overviews the foundational concepts and current research progress that are relevant to the current research. It starts with introductions to the preliminaries of livestock disease detection, cattle identification, and non-contacting weight estimation, followed by a critical appraisal of the current methodologies of ML and DL that are used in each of these areas. The chapter then concludes by summarizing salient findings, recurrent limitations (e.g., real-farm variability, occlusion, and dataset bias), and research gaps that drive the motivation opened for the proposed unified framework. Chapter-3 outlines the requirements, potential effects, and constraints for the design and implementation of the proposed system. It articulates the last specifications/knowledge and data prerequisites as containing data set preparation and pre-processing constraints and evaluation criteria for classification, identification, and regression tasks. Furthermore, the chapter considers societal and environmental implications, ethical issues regarding data collection and model deployment, relevant standards, project management planning, risk management approaches, and the economic viability of the proposed solution a midst resource-constrained farming conditions.

Chapter-4 outlines the methodology proposed as well as the system design by introducing a full pipeline combining the disease diagnosis with the severity grade determination, the identification of cattle through multi-view utilization using left, right, frontal, and dorsal views, and the weight estimation by the RGB imagery combined with a regression analysis. The chapter-5 elaborates on the data acquisition protocol, including data acquisition, cleaning, transformation, and integration stages, and further explains the model architectures, training protocol and implementation workflow that were chosen for building the overarching cattle monitoring ecosystem. Chapter-5 is a systematic reporting of the empirical findings and provides an in-depth analysis of the system's performance.

It evaluates the disease-severity model by classification metrics like accuracy, precision, recall, F1 - score and confusion matrices, evaluates cattle identification by Rank-1 and Rank-k accuracy in a protocol-driven testing framework, and evaluates weight estimation accuracy using MAE, MAPE, RMSE, and R^2 . Therefore, the chapter goes on to explore relative results among modeling versions, note stable-ness characteristic, explore the inter-relationships between tasks, and outline practical implications, leading up to a concentration on limitations and considerations relevant to real-farm deployment.

In conclusion, the main results of the study are synthesized in Chapter-6, defining the main contributions of the proposed ecosystem towards the precision livestock health management. Furthermore, it states the prospective avenues for future research in the area, which include ways to increase system robustness under unfavorable farm conditions, to increase multi modal sensing capabilities, to augment interpretability, and to scale the solution to larger herd sizes and heterogeneous breeds.

For better readability, each chapter, section and sub-section in this paper is organized in three main sub-parts, cattle disease, identification, and weight estimation, providing comprehensive analysis and explanation for each parts.

Chapter 2

Literature Review

2.1 Preliminaries

The control of bovine disease is a continuing and major issue in contemporary livestock production systems, which has resulted in considerable economic losses and animal-welfare outcomes, both due to infectious agents and due to production-related diseases. Conventional methods of diagnosis, such as manual observation, physical examination, and lab tests, are often labor-intensive and time-consuming and cannot identify a pathology at early or asymptomatic levels [39, 43, 60]. When dealing with outbreak management, rapid diagnosis of highly contagious agents like the FMD, molecular diagnostic methods are usually relied upon (such as reverse-transcription polymerase chain reaction) [13]. The dependence on the above factors highlights why there is a need to develop rapid and scalable screening approaches that can be used to conduct timely surveillance and swift response in the field [49].

Motivated by these constraints, the scholarly investigations progressively use ML and DL techniques for the improvement of disease detection and prognostication in livestock systems [5, 60]. Moreover, Traditional ML systems using structured data from automatic devices, such as milking or behavior sensors, are still useful [43]. Furthermore, they require mostly dependable sensor arrangements and well-conceived functions [39, 66]. Additionally, Newfangled methods of DL, mainly CNNs, have the ability to teach immediately from images [?]. Therefore, this opens up the doorway to disease checks that touch no animal and can be cheap [27, 57].

Current developments in managing livestock suggest a great change in the methodological approach used in the monitoring of animal health [5]. Rather than relying on visual inspection through camera technology alone to detect physical conditions (e.g., hoof abnormalities), modern systems combine photographic evidence with information from Internet-of-Things (IoT) sensors (e.g., heart rate and peripheral capillary oxygen saturation (SpO_2) [2]. This constant and 24/7 data collection paradigm improves the reliability of early warning indicators and is able to provide accurate and concrete information to farmers in the form of alerts that are not sent raw but processed, in contrast to the previous dissemination of raw datasets [49].

Non-contacting body-weight derived estimation has become a topic of interest in research, due to its usefulness in monitoring the progress of growth, managing nutrition, and in detailed health evaluation [53]. The literature can be divided into three main areas of methodology: (i) measurement-driven, using anatomical landmarks to derive

the regressed weight from derived body dimensions [40, 59], (ii) an end-to-end regression approach to predict weight directly from images [14, 23], and (iii) hybrid pipelines, first segmenting the animal, thereby suppressing the background noise, before the actual prediction step [8, 68]. Moreover, it has been recently investigated that performance could be further improved using multi-view fusion (e.g., side and top view combination), modality fusion (RGB plus depth), and segmentation-assisted representation [1, 35, 54]. Despite this, several studies have consistently documented the inadequacy of only 2D approaches to give an accurate depiction of volumetric cues, and at the same time, their vulnerability to pose, occlusion, and uncontrolled imaging environments [14, 53]. Explainability-centered weight-estimation research, another example of untrustworthy models, even when very good, may be based on spurious correlations or anatomically uninteresting areas of anatomy, illustrating the need for models to be explained to ensure they are improved enough to be used in real-world settings [23].

A steady challenge to the adoption of digital solutions in disease diagnosis, identification, and tracking, and weight estimation is the conflict between predictive accuracy and interpretability [23]. Consequently, several DL models are opaque, perhaps undermining the trustworthiness of stakeholders and reducing their applicability in veterinary practice [23]. Therefore, interpretability methods such as LIME and SHAP have seen an upsurge in usage to help in responsible model refinement, understanding of decision processes, and ensuring predictions are supported by biologically meaningful evidence [23]. In order to address the issue of over-fitting and generalization to diverse farm environments, robust livestock AI-based systems are commonly evaluated using classification metrics (e.g., accuracy, precision, recall, F1-score, and Area under the Curve (AUC)) and regression metrics (e.g, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE)) and validation regimes, e.g., cross-validation and well-defined train/validation/test splits [3, 17, 40].

In general, current trends in the research literature focus on the design and development of scalable and deployable artificial intelligence systems with increased robustness in the face of real-world farm variability, including lighting variations, clutter, occlusion, and animal posture, that combine diverse types of data sources and advocate for explainable and rigorously validated models that are more commensurate with the practical needs of livestock health management [38, 41, 42, 45].

2.2 Review of Existing Research

Livestock Disease

Over the past years, ML and Artificial Intelligence (AI) have been actively used to predict livestock diseases, leading to numerous breakthroughs [5, 60]. These advancements have seen remarkable research on the detection and management of diseases in cattle, as discussed eloquently in a study below. Motohashi et al. [43] give an initial detection system of subclinical mastitis in dairy cows regarding ML algorithms on PC data of automated milking systems as a solution to the issue of a disease that causes the loss of plenty of money without causing symptoms, except, unfortunately, in most cases. It is a study on the two farms in Japan in which DeLaval Voluntary Milking System (VMS) and Herd Navigator (HN) systems are installed, with over 10,000 samples, with the following attributes: milk yield and electrical conductivity, milk flow rates, blood content,

milking intervals, and lactate dehydrogenase-based scores of mastitis risks within 5 successive days of each cow. It uses two ML algorithms, Random Forest and Support Vector Machine, and the feature selection is implemented with the Random Forest procedure of importance and Boruta algorithm, and the data imbalance is addressed on the basis of the procedure of Repeated Edited Nearest Neighbor (RENN) under-sampling. The best-ranked Random Forest with Boruta performed feature selection with a sensitivity of 81%, and a precision of 46% on similar farm data, which indicates it can give a warning earlier than standard LDH-gain-based alerts, but is less generalizable to other farms. This paper demonstrates that the integration of a real-time, cost-effective method of monitoring subclinical mastitis based on time-series sensor data and ensemble learning models is viable and attains an AI-supported model scaling of the diagnosis of cattle disease in an automated agro ecosystem.

Recent research on the importance of livestock disease focuses on the fact that prompt laboratory confirmation remains a critical part of the response to such outbreaks, especially with Foot & Mouth Disease (FMD). For example, an Indonesian study using RT-PCR testing on clinically suspected cattle during the 2022 national outbreak found that 58% of the samples collected were positive results using universal primers, and thus emphasizes the importance of the use of molecular diagnostics in surveillance and control efforts of highly contagious diseases [13].

At once, a number of studies focus on the growing importance of AI-enabled prediction and screening of disease, especially in resource-strapped settings. A survey published in 2023 discusses how traditional ML classifiers such as Naive Bayes, KNN, SVM, Decision Trees and Random Forest classifiers and data driven pipelines can help predict diseases in cattle based on clinical symptoms and sensor/IoT based signals even though it is observed that many methodologies are reliant on the public availability of their datasets and many times have been reported without any extensive large scale experimental validation [60] . In the upcoming years, Mahmood and Hamed et al. [39] compare the results of some ML algorithms in predicting mastitis in dairy cows on a dataset of 6,600 samples (15 physiological and anatomical parameters, including hardness, temperature readings, and inhale exhale readings of the four quarters of the udder). SVM, J48, J48 Graft, Random Tree, Random Forest, and CSForest are estimated, wherein the decision tree-based models give a perfect performance (100 percent accuracy, detection rate, F-measure), but SVM outperforms at the cost of the imbalance in the classes (5,490 healthy, 1,110 cases of mastitis).

With these advances, Phulu et al. [49] propose a dual-model AI system that uses both symptom data and image analysis for early and accurate detection of FMD in cattle. A Random Forest classifier learns from an ontology-based dataset about difficulty in walking, excessive salivation, and visible sores and a MobileNetV2 network is fine-tuned on images of cow cases labeled as “Healthy” or “Infected.” The accuracy of Random Forest was 90.62%, meaning it could put cattle into the correct categories of “Yes,” “No,” or “Maybe” for the disease. The use of transfer learning and data augmentation allowed MobileNetV2 to achieve a prediction accuracy of 91.1% in spotting FMD symptoms from images. Over-sampling and class balancing enhanced the curated datasets, which were used to train and assess the models. Here, Flask based web portal provides a way for users to report their symptoms, upload photos, or use any of the options synchronously. It uses both models to understand the data and deliver one final diagnosis. The frontend

has a particular structure, the backend relies on trained models, and the core handles the movement of data from user input to AI results. Early FMD detection, which supports fast decisions and animal care, is proven possible through the integration of ontological reasoning and DL into image classification.

In the realm of image-based diagnosis of visible diseases, recent work on DL continues to focus on lightweight architectures that are suitable for deployment in the field. Recently, a study directed by MobileNetV2 framework (MobileRMSNet) tuned with RMSProp algorithm was proposed by a 2024 publication in the journal PLOS ONE, yielding a classification accuracy of 95% for the identification of LSD, leading to the various possibilities of combination of transfer learning and efficient optimization to provide rapid, non-invasive screening facilities for devices having restricted computational limits [57].

Beyond lesion-based classification, advanced representation learning is facing additional research for infectious disease diagnostics, where distinct differences are sometimes present in subtle images. A study in 2025 was published in the Journal of Big Data on a two-stage pipeline procedure of bovine anaplasmosis by YOLOv4-tiny algorithm for cell detection and cropping, followed by contrastive deep metric learning using ResNet-50 triplet margin loss and the k-Nearest neighbors retrieval. This approach aims to improve robustness - specifically with regard to variation from the staining effect - try to reduce overfitting by discrimination in the embedding space [26]. Nonetheless, high-performance parameters measured in manicured settings in the laboratory do not necessarily translate into credible performance in the field that has not been validated across multiple farms, breeds, imaging technologies, and climatic parameters. For this reason, it is necessary for rigorous benchmarking protocols to include cross-validation, independent testing, and careful prevention of data leakage in order to ensure generalisability of disease recognition. [5, 60].

Ultimately, multi-modal systems, including Computer Vision combined with Internet-of-Things (IoT) physiological sensors, are emerging for early detection of external diseases of cattle. In a 2025 study, an implementation that utilizes a lightweight convolution neural network (CNN) algorithm was used to classify a subset of FMD, LSD, and IBK as offensive to livestock based on the combination of fusion of image with real time information (temperature, heart rate, blood oxygen saturation [SpO_2]), on a mobile application and via a cloud backend. Although the classification performance reported was decidedly high, the evaluation context and the place of publication mean that these results must be viewed with some caution until replicated on a larger and more varied set of data and in a field setting [2].

Across these reviewed studies, three limitations have been identified. First, many models are able to report high values for offline-accuracy, but suffer from decreased robustness when it has to be applied to diverse farms because of differences in lighting, the presence of background clutter, posture of animals, lesion development, or the quality of the device, compromising generalization. Second, several pipelines are hooking off from small or curated datasets, such that the problem of class imbalance and dataset bias can lead to an overestimation in performance unless they are tested through rigorous cross-validation and truly independent sets in general. Third, and for adoption to occur, there is a need for practical deployable systems that can function within mobile or edge constraints with shoes that give transparent responses that are facilitating in veterinary decision-making. These limitations highlight the need for lightweight architectures that can be used in the

field, are thoroughly evaluated, and have explainability, for the trustworthy screening of livestock disease.[5, 30].

Cattle Identification

Qiao et al. [50] train a DL model that simplifies the identification of individual cattle without any physical tagging system, which becomes essential in the precision management of livestock. This experiment utilizes a hybrid CNN+LSTM network design that uses an Inception-V3 model to extract the spatial features in the images of rear-view cattle with LSTM models to identify temporal patterns of cattle gait by processing the video clips. They have 516 video sequences of 41 cows, with 20 frames of 401 by 506 pixels, divided into 439 training and 77 testing video sets over 2 months. The integrated approach reaches an accuracy of 91 percent on 20-frame sequences, as compared to a single image, where the standalone approach of Inception-V3 gives an accuracy of 57 percent. The study exhibits a strong potential of a camera-based livestock monitoring system using a combination of spatial and temporal features, and future studies by the authors should be conducted over a large dataset and with further integration of more behavioral patterns to improve the identification performance.

In the paper by, Li et al. [33] propose an advanced DL framework that accurately identifies individual cattle using multiple visual features. The purpose of this research is to produce an identification system that is reliable and can work automatically, helping manage, track, breed, and prevent theft of farm animals. The method they use avoids problems caused by ear tags and manual identification by fusing information from the cow's face, muzzle, and ear tags using a special way at the decision-making level. They built the system by designing a Multiple Feature Decision Layer Fusion Network (MFDN). Initially, they gathered two datasets (QY-Cattle and NQ-Cattle) consisting of images of cows taken from the front, side, and top. Using SOLOv2, they separated the image into components to recognize the face, muzzle pattern, and ear tag. We applied the FaceNet model in different ways to gather features from the faces, muzzles, and ear tags, and we used the PP-OCRv4 network for the ear tags.

Before, fusion took place very early with features, but in this study, it was done with the output scores of each model at the moment decisions were made. The leading predictions from every viewpoint were converted with One-Hot encoding and joined by logistic regression, random forest, and voting classifiers. It works well when features are covered up, misshaped or shot in poor lighting. The model proved to be more effective than single-view models and other early-fusion strategies based on the outcomes. Both datasets produced a 95.74% accuracy for the system, which worked reliably despite variations in lighting, breed and posture. Li et al. (2024) [33] say that this framework is suitable for farm environments and suggest studying how it can be applied to animals outside cattle and to other tasks on farms to improve identity tracking and livestock management.

Li et al. [34] introduce in this paper the objectives of developing an automated method in identifying individual cattle in the world called Individual Beef Cattle Identification Using Muzzle Images and DL Techniques. The system will offer a simple, non -invasive, efficient and scalable solution to livestock identification with the potential of alleviating problems of ear tag loss and avoiding the assignment of false tags by underlying waywardness in manual tagging. The proposed methodology involves pre-processing high definition images in the muzzles by simply resizing, normalizing it, and then training

59 DL architectures with different families, such as ResNet, DenseNet, EfficientNet, and MobileNet. The model also has cross-entropy loss and data augmentation to increase the performance and generalization of the model. The models are compared not only in terms of the accuracy of their classifications, but also in terms of speed of the inferences and the resulting rank is based on a holistic index (CI), that measures the inference speed along with the accuracy of the classifications.

The data comprised 4,923 images of muzzles of 268 cattle animals in natural scene illumination in a University of Nebraska feedlot, acquired with a mirrorless camera. In order to deal with the class imbalance and variation of lighting, images were cropped, resized (as 300 300), and augmented during data pre-processing. VGG16-BN recorded the highest identification accuracy of 98.7% and MobileNetV3-Small the fastest, with image processing running at 28.3 milliseconds among those models tested. The muzzle as a biometric was also found to be reliable, with more than 95 percent identification in one or more images, with 100 percent accuracy of identifying cattle.

This paper has concluded and suggested that muzzle imaging-based identification and DL-based identification are compelling alternatives to typical forms of identification, and could be used in traceability, health management, and herd management.

Another study presents a DL-based approach for re-identifying individual cows using two distinct profile images—frontal and side views—to create a more discriminative representation. The proposed system uses a dual-branch CNN architecture that outputs a 128-dimensional unit-norm embedding vector for each cow, effectively capturing both texture-rich and shape-based identity features. Each CNN branch consists of custom-designed blocks: ConBlocks (Convolution + Instance Normalization + LeakyReLU) and ResBlocks (Residual layers with InstanceNorm and LeakyReLU), and the model is trained using Histogram Loss to optimize embedding separation between individuals. The method was tested on a real-world cattle image dataset collected from four Italian farms, containing 12,952 training images from 387 cows and 4,850 test images (including a closed-set evaluation on 52 cows). The system outperformed all traditional Computer Vision methods such as EigenFaces, FisherFaces, LBPH, and HOG, as well as deep face recognition baselines like SphereFace (even when retrained on cattle data). In closed-set testing, the multi-view model achieved 81.7% Top-1 accuracy and 89.1% Top-3 accuracy, significantly surpassing the single-view approach (68.8% Top-1) and SphereFace (55.6% Top-1) [4]. This demonstrates that integrating multiple angles of the cow’s head enhances discriminative learning, making the system far more reliable for practical applications in livestock identification, traceability, and health monitoring.

A very recent and prominent research by Mon [42], introduces a real-time AI-based cattle identification system that employs a three-stage pipeline to detect, track, and identify individual cattle across diverse farm environments using RGB video footage from three Japanese farms (147 cattle at Farm A, 13 at Farm B, and 1,103 at Farm C). The methodology leverages YOLOv8 for cattle detection, achieving mAP@0.5 scores of 98.6% (Farm A) and 97.2% (Farm C), followed by a customized IoU-based tracking algorithm with 90.16–100% accuracy across farms. For identification, VGG16 extracts features from cattle back patterns, classified by a Support Vector Machine (SVM), yielding identification accuracies of 81.25–96.88% (Farm A), 91.67–100% (Farm B), and 98.83–100% (Farm C), with a cross-validation mean accuracy of 95% ($\pm 1\%$ std dev). The system distinguishes black and non-black cattle and handles “unknown” cattle via frequency thresholds, using

manually annotated video frames extracted at 1 FPS from cameras above milking lanes and rotary parlors. By overcoming challenges like ID-switching and variable lighting, the approach offers a scalable, low-stress alternative to traditional RFID-based cattle identification for precision livestock farming.

This article entitled Integrating Artificial Intelligence in Dairy Farm Management - Biometric Facial Recognition of Dairy Cows, by Mahato and Neethirajan [38], reviews in depth, the biometric facial technology used in the management of dairy cow and its integration with Artificial Intelligence (AI). The authors describe the potential of using AI-based facial recognition based on human recognition systems to identify and track cows in a non-invasive contactless and scalable way that allow understanding the health status, emotion, and productivity of the cows.

The authors explain how facial recognition models of today are based on CNNs (CNNs), DL systems, and transfer learning to extract strong features of cow faces such as muzzle patterns, ear shapes and inter-ocular distance. Unlike the conventional identification (e.g., RFID or back/tail-based re-ID), these systems are more accurate and can be processed in real-time due to edge computing and the implementation of IoT [70]. The paper nevertheless highlights the major pitfalls: environmental uncertainty (lighting, obstructions), the difficulty of producing quality annotated data sets, an ethical concern of constant surveillance, and the financial burden of hardware on small farms. The authors advocate the need to deploy AI in an ethical way, integrate the multi-modal systems (e.g. RFID + facial recognition), and future direction in solving adaptability to various breeds of cattle and other conditions in the farm.

Lastly, open-set conditions are also the conditions of identification where animals that were not seen previously (at test time) are investigated in terms of metric learning of multi-view whole-body images. A small-loss formulation that jointly optimizes classification supervision and metric constraints enhances both inter-class separability and intra-class compactness on normalized embeddings to facilitate robust cattle identification and application to realistic open-set tasks and multi-view datasets developed on multi-view re-identification [71].

Livestock Weight Prediction

Cominotte et al. [6] designs a non-exclusive software system that uses CompV to predict body weight and daily average gain in beef cattle as a solution to inefficiencies of the traditional, labor-intensive weighing techniques. The researchers apply a Kinect RGB-D camera to obtain 234 depth images of Nellore steers during four growth stages (weaning, stocker, initial, and final feedlot) and analyze their biometric parameters (body volume, dorsal area, length, and segmented parameters). Evaluating four predictive models: Multiple Linear Regression (MLR), Partial Least Squares (PLS), Least Absolute Shrinkage and Selection Operator (LASSO), and Artificial Neural Networks (ANN) with leave-one-out cross-validation assessing numbers, such as RMSEP, R^2 , and Concordance Correlation Coefficient. The results show that ANN is generally the most accurate model in the majority of phases, with R^2 of 0.91 and RMSEP of 8.63kg in the weaning phase, whereas PLS is better accepted in the last feedlot phase with small data. The study shows the promise of applying 3D Computer Vision to ML in real-time livestock weight and demonstrates that model selection should be determined by the characteristics and phase of growth to increase the chances of commercial adoption.

To overcome the shortcomings of the traditional weighing platforms, which are expensive to implement and lead to stress and labor intensity, Ruchay et al. [55] design a Computer Vision non-invasive framework to estimate the live weight of cows using DL at the RGB-D images. The test includes three CNN models used (MRGBDM (RGB + depth), MRGB (RGB only), and MDM (depth only)) along with the EfficientNet transfer learning that are trained on two publicly available datasets, containing 154 Hereford cattle (31,000+ pictures) and 121 Aberdeen Angus cattle (28,630 images) recorded with Microsoft Kinect v2 sensors. The pre-processing consists of background removal, denoising, depth map smoothing, posture segmentation with YOLO v4, and data augmentation techniques (rotation, translation, scaling), which extend the diversity of the dataset by ten times, and 2.5D projections and color augmentations of point cloud improve variability. The multi-modal MRGBDM model is found to perform better with 8.4 MAPE and 91.6 accuracy, far better than single-modal versions and prove that it is indeed the case that combining RGB and depth data enhances prediction performance significantly. The feasibility of scalable, contactless cattle weight prediction systems based on DL and RGB-D imagery is successfully demonstrated through the study, creating a sensible reality of livestock management that is cost-efficient and easy on the animals at the same time.

Rozendo et al. [53] This work suggests a pragmatic weight prediction pipeline based on side view RGB imagery to circumvent the obstacle that is posed by many agricultural situations where it is impracticable to obtain dorsal or multi-view imagery. The methodology combines two-dimensional body measurement extractions with depth estimation to create a pseudo three dimensional representation to allow for volumetric and shape-related descriptors to be derived, to regress weight prediction. By fusing image derived geometry and depth derived three dimensional cues that has been proposed to enable non contact weight monitoring by using low cost acquisition systems that are compatible with conventional barn and outdoors agricultural environment.

Dohmen et al. [14] propose a Computer Vision-based non-invasive method of predicting the mass of dairy heifers based on 2D image and deep neural network technology, which can be used instead of the more time-consuming and labor-intensive conventional weighing of animals. Our research is conducted by using a data set of 63 crossbred heifers (aged 0–365 days and body mass 37–370 kg) in both side-view and top-view; applying mask R-CNN to accurately segment an animal to eliminate complicated backgrounds and subsequently adapt custom CNN models using Keras/TensorFlow to predict mass. They create a total of 48 CNN models with different convolution layers, max-pooling layers, and filter sizes in each model, train on 60 percent data, 20 percent validation, and a 20 percent test split. The top-view model demonstrates better performance than the best side-view model ($R^2 = 0.91$ and RMSE = 20 kg vs $R^2 = 0.96$ and RMSE = 26.7 kg) because the side-view models face the challenge of not segmenting well in the areas of legs and head locations. This study is a clear indication of the viability of Mask R-CNN segmentation paired with CNN regression to enable a scalable cost-effective solution to livestock weight monitoring; however, the authors note that such a solution is limited in its applicability by the single-farm trials as well as propose augmentations to existing systems by multi-farm training sets and skewed mass distributions.

In the upcoming years, Ruchay et al. [54] raises and automates a contactless system to measure body dimensions of live cattle using three Microsoft Kinect v2 RGB-D cameras in varying angles (top, left, right) to synchronize with picture capture of moving Hereford

cattle in a passageway. The system uses real-life Computer Vision methods such as depth map filtering using switching bilateral filters, Radius Outlier Removal to remove noise points on a point cloud, and non-rigid iterative closest point (ICP) implementation to individually register 3D shapes even with animals moving and change in lighting. Camera calibration is provided by using ground plane detection and transformation matrix estimation, and landmark detection is improved by generalized matched filters in the extraction of anatomical features. It has a dataset of 103 annotated cattle instances, and the system records measurement errors of less than 3 percent at 90 percent confidence level across nine morphological parameters (withers height, hip height, heart girth) against manual measurements. This study is able to prove the effectiveness of multi-camera 3D imaging systems with non-rigid registration algorithms in acquiring morphological data in precision livestock farming, especially in the field of accuracy and real-time, bringing a scalable solution in automated biometric analysis which can considerably fill gaps in automation health monitoring, growth tracking, and genetic assessment processes.

In This paper by Dang et al. [8] to solve the issue of labor-intensive weighing and stress in animals used in Korea. Data is collected by the use of a flexible multi-camera stereo imaging system applied in real farms, resulting in 1190 point cloud meshes using 270 classes in different postures with stereo matching and Poisson surface reconstruction creating entire 3D meshes using incomplete shapes. Two DL segmentation models based on PointNet, one on torso (extraction of body length) and one on center body (chest girth and cockpit width) are created and reach training accuracy of 99% and a larger than 97% validation accuracy, and then three key lengths are independently measured by the three regression models (CatBoost, LightGBM, Polynomial Regression, Random Forest, XGBoost). Random Forest Regression model demonstrates the best values of MAE 25.2 kg and MAPE 5.81 with 10-fold cross-validation cross-sampling augmentation, which transforms each labeled 3D mesh into 20 independent sparse point cloud instances, and posture correction by PCA ensures the dimensional consistency of the samples. This work is a successful application of efficient, scalable solution to weight prediction of livestock with real-world farm conditions, provides great advances in accuracy, animal welfare, and can be adapted to smart farming due to sound treatment of missing 3D data and careful segmentation optimization.

In the paper, Xu et al. [68] explored a smart, non-contact method for predicting the weight of cows using images and trained neural networks. The use of conventional scales generally needs touching and machinery which may be stressful for the animals and slow. For this reason, the authors came up with a system where Computer Vision inspects cow photos and estimates the weight more quickly. They implemented ResNet-101-D and then used a Squeeze-and-Excitation (SE) mechanism to focus on the key details in the photographs. It compared the images it took of the cows' top and sides, found the length, width and height of their bodies using pixels and used this data to guess their weight. They collected observations from actual farms and used the Labalme tool to label the data for training. In their study, Xu et al. (2024) [69] compared several ML algorithms for making predictions such as Multiple Linear Regression (MLR), Decision Tree (DT), Support Vector Machine (SVM), Gaussian Regression (GR) and Back Propagation (BP) neural networks. The BP neural network had the best results, with a mean absolute error of 13.11 pounds and a root-mean-square error of 22.73 pounds. Comparing various attention mechanisms, they noticed that SE made the model improve by paying more attention to the essential features in the cow's picture. Their approach performed better

in terms of accuracy than others (PSPNet and HRNet), but was slightly slower in handling images [69].

In this recent paper Zin et al. [72] present an automated method for estimating dairy cow body condition scores (BCS) using geometric analysis of top-view images. It identifies anatomical points in places like the pin bones, hook angles, tail head and it analyzes their geometric qualities—angles, lengths and areas. It uses second-order polynomial regression, multiple linear regression and Markov Chain classification to predict what BCS teams will score. Publicly available datasets also play an enabling role in facilitating benchmarking and ensuring reproducibility in livestock analytics using images. For example, the Cow Weight Dataset on Roboflow Universe provides labeled visual observations and included markings that can serve to back up the training and evaluation of livestock- in this case, cattle-focused Computer Vision pipelines in line with established standards. Moreover, such resources reduce the barrier to comparison of models and allow ready prototyping of segmentation and measurement components that might be included in downstream workflows for estimation of weight[18].

The method consists of three stages: (1) cows are detected by edge detection and their anatomical points are marked, (2) the points are assigned to five parts of the body, angles and curves are calculated, and a feature vector is formed and (3) final scores for the animals are estimated using regression and Markov Chain approaches. 75 cow images from Corfilac, IPLAB and Penn Vet were used in the tests: 50 images were trained on and 25 were tested. The results indicated excellent results, as seen by $R^2 = 0.7179$, adjusted $R^2 = 0.6281$ and $R = 0.8473$. The mean square error for the model was 0.0366 which is lower than the results of other available methods. It is shown that image analysis of geometry can help BCS assessments be more accurate, consistent and easier to perform, decreasing subjectivity and making it simpler to monitor dairy cows' health.

This paper investigates the effectiveness of regression-based ML models in predicting cattle body weight using manually measured physical features. The researchers used the Full Cow Primer (FCP) dataset, consisting of 150 instances of dairy cattle from farms in the Nizhny Novgorod region of Russia. Every sample has 10 anatomical measurements—such as chest girth, hip width, and length of body—recorded by professionals with the use of standard tape measurements. In this study, eight regression models were examined: Linear Regression, Random Forest Regressor, Support Vector Regressor (SVR), K-Nearest Neighbors, Multi-Layer Perceptron, Gradient Boosting, LightGBM, and XGBoost. The models were thoroughly tested with 100-fold cross-validation and compared to a wide range of various performance metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and R^2 score. SVR worked best with an MAE of 0.09 kg, an RMSE of 0.88 kg, an MAPE of 0.02%, and an R^2 of 0.97—all better than even classical ML models, as well as even earlier CNN-based models with much fewer errors. This demonstrates that light, tabular-data-driven ML models are able to supply accurate, non-invasive, and cheap cattle weight predictions for field deployment as part of modern livestock management[59] .

In 2023, Sarini and Dharmawan developed ML regression models to predict the body weight of Bali cattle using morphological measurements, offering a non-invasive alternative to traditional weighing methods. The study utilized a sample of 228 bulls and 211 Bali female cattle, all of 285 days age, manually collected during April to May 2022 at BPTU-HPT Denpasar, Bali, with body length, girth circumference, and withers height

as the predictor variables. Six algorithms, namely Linear Regression (LR), Random Forest (RF), Extra Trees (ET), Support Vector Regression (SVR), K-Nearest Neighbors (KNN) and Gradient Boosting (GB) were implemented and compared using Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) [58]. The results depicted Linear Regression as the best model for female cattle (RMSE: 9.6172), followed by SVR (RMSE: 9.7003), and the best for male cattle was provided by Random Forest (RMSE: 9.8952), followed by Extra Trees (RMSE: 10.0271). KNN performed below par in both genders. The findings highlight strong correlations between girth circumference and body weight, demonstrating the efficacy of morphometric data for weight prediction in resource-limited settings.

A very recent study developed a ML-based simulation model using XGBoost regression to predict daily cattle weights in extensive pasture-based systems, utilizing 251,000 walk-over (WO) weight records from 1,105 cattle across two Central Queensland farms, reduced to 46,585 records after cleaning. The model integrates input features such as age, sex, breed, and a weather factor derived from rainfall and temperature data from Queensland's SILO service, with data pre-processing involving linear interpolation and outlier removal using static weights as a reference. Validated on two test sets, the model achieved strong performance with Lin's Concordance Correlation Coefficient (CCC) of 0.84 and 0.32, RMSE of 32.60 kg and 38.30 kg, MAPE of 15.2%, and R^2 of 0.77 and 0.73, outperforming linear regression and Random Forest models from prior studies but slightly trailing models with more input features [3]. By enabling weight trend predictions based on herd demographics, the model offers a robust tool for farmers and researchers in tropical/subtropical pasture systems.

Building on previous research, this article by W. Ma [37] examines non-contact Computer Vision techniques for collecting phenotypic data in livestock, focusing on body dimension acquisition and weight estimation through 3D reconstruction using RGB, LiDAR and depth cameras, and comparing linear regression and neural network-based methods. The methodology involves a comprehensive analysis of existing studies, categorizing approaches into 3D reconstruction, body dimension measurement from images or point clouds, and weight prediction, with findings presented in structured tables and diagrams. Datasets from cited works, primarily non-public and involving pigs, cows, lambs, and sheep, include image and 3D point cloud data from sources like Kinect v1 and Intel Realsense, capturing attributes such as height, length, and circumference. Results highlight 3D camera-based methods as superior for real-time performance and depth accuracy, with neural networks achieving lower errors (e.g., MAE of 1.16 kg for pig weight) compared to linear regression (4.87%–9.7% error) [37]. The review emphasizes the potential of these technologies for stress-free livestock monitoring but notes challenges like data noise from animal movement, high sensor costs, and the lack of standardized public datasets.

Tawheed et al. (2019)[66] work on predicting the expected weight of cattle based on various factors such as age, current weight, diet, environment, and breed by applying ML techniques to livestock farming. Their goal is to provide investors and farmers with data-driven insights into the right choice of the most profitable cattle breeding and thus contribute to the economy. They collect their dataset from Savar Dairy Farm, Bangladesh Livestock Research Institute (BLRI), and Meghdubi Agro, which contains approximately 1300 rows of data related to different cattle breeds. They use Ordinary Least Squares (OLS) to determine the correlation between the expected growth rate and independent

variables. To analyze the relationships between various independent factors affecting cattle weight, they use Multiple Linear Regression (MLR). They also use Support Vector Machine (SVM) for different kernel functions (Linear, Polynomial, Radial Basis Function (RBF), Sigmoid) to improve prediction accuracy and a Decision Tree for classifying the data to handle both categorical and numerical values. Their percentage errors are higher for certain cattle breeds (e.g., Indian-Haryana and Local breeds) due to a lower number of cattle records. They have to discard their SVM model as the Sigmoid kernel did not yield satisfactory results, and for that, it is excluded from the final evaluations. They also could not use the time series analysis due to insufficient data, particularly a lack of monthly growth records.

In addition to 2D images, a number of studies support the utility of 3D sensing for high-throughput phenotyping, in particular where the dorsal curvature or body volume can be reliably measured. As an example, 3D segmentation pipelines on incomplete meshes with the PointNet method are able to derive the most important body dimensions that are input to tree-based regression models (e.g., Random Forest) with competitive errors in real farms. Complementary work in commercial applications demonstrates that large-scale 3D camera systems are capable of providing high repeatability of predicted body weight of a large number of cows (thousands) and a large number of herds (and/or herd populations) to enable continuous monitoring and phenotyping of that population at the population level. Interconnected comparative work based on the contour features of 3D images indicates that ensemble algorithms (e.g., CatBoost, Random Forest) tend to be top-ranked, and that an interbreed dataset bundle can also directly enhance prediction accuracy, mostly due to the volume of data[1, 8, 17, 31] .

García et al. (2021)[16] introduce a ML based identification system that models the expected or ideal weights of cattle based on age, breed, and gender. It detects anomalies when actual weight has deviated significantly from the model's predictions, which primarily helps farmers monitor the cattle's health and growth efficiency. For this purpose, they collect a dataset which includes 104 cattle tracked from birth to slaughter and also validate them by zootechnics at the "El Rosario" farm located at Montería-Colombia. They use four ML models which are Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), K-Nearest Neighbors (KNN), and among them, the Decision Tree (DT) and The Random Forest (RF) had the best performance, with an average Mean Absolute Error (MAE) of approximately 5.4525. Although they have statistically similar MAE values, DT trains about 7 times faster than RF, making it more efficient. They also use an Anomaly Detection model, which is an Isolation Forest, that flags abnormal weights based on deviation from the predicted range. Although it detects anomalies, but does not explain their causes (e.g., health vs. nutrition).

In the year 2023, Guvenoglu and Erdal present us with a method to estimate the live weight of cattle using stereo vision and DL based semantic segmentation. Their goal here is to replace the traditional weighing methods with a Computer Vision-based system that captures images from two angles, one from the back and another from the side, and later uses them to train with the help of neural networks for weight prediction. A total of 85 animals are used for collecting the data by a custom stereo camera setup consisting of two identical webcams connected to a Raspberry Pi 4. After capturing the images, the DeepLab v3+ model was used for semantic segmentation of the animals in the images. They use Artificial Neural Networks (ANNs), where ANN-1 is used for training on side-

view data, ANN-2 is used for training on back-view data, and ANN-3 is used for training on both side and back views. The Data is normalized using the Z-score technique, and K-Fold cross-validation ($K=10$) is used to evaluate ANN-3. ANN-3 (combined views) performs the best, with weight estimation errors mostly within ± 20 kg. Whereas ANN-1 and ANN-2 have larger error ranges (up to ± 50 kg), especially ANN-2, due to limited body shape visibility in back views. Average predicted weight for (ANN-3 is 416.4 kg, where the actual is 411.95 kg. Their biggest limitation is using a very small dataset and an imbalance of weights among the 85 animals. They also state that close-up images (2–3 m) are hard to capture due to animal fear. For that reason, most data are from 6–8 meters, reducing stereo accuracy. [21].

2.3 Summary of Key Findings

ML and DL algorithms led to extraordinary breakthroughs in the field of detection of diseases among livestock, opening up new research directions aimed at early and correct diagnosis of the disease. The study conducted by Motohashi et al. [43] and Mahmood and Hamed [39] demonstrates the effectiveness of such ML algorithms as Random Forest and Support Vector Machine (SVM) in terms of their ability to detect subclinical mastitis in milk cows with the most impressive sensitivities of up to 81 percent and highest accuracies of up to 100 percent after achieving the data imbalance resolution with such methods like the Synthetic Minority Oversampling Technique (SMOTE) or Repeated Edited Nearest Neighbor (RENN) under-sampling. A large amount of data obtained automatically during milking was used in these studies and these data included milk yield, electrical conductivity, and various physiological factors such as udder temperature. A comparable trend occurred in the study by Phulu et al. [49], in which the authors worked on the FMD using a dual-model AI architecture that was a combination of Random Forest and MobileNetV2 to predict the results with 90.62% and 91.1 accuracies in symptom and image-based classifications, respectively. The next stage of lameness detection in sows and skin disease classification in cattle, sheep, and goats was the DL methods implemented by Guimarães de Paula et al. [20] and others with EfficientNetB7 and 99.01% accuracy. These systems combine novel use of real-time sensor data, image augmentation with ontological reasoning, which allows intervention in the early stages of the problem and practice implementation in a field environment through the use of mobile applications and web portals by veterinarians[19].

It is also advancing at a very quick rate in non-invasive cattle identification, with DL beginning to supplant the traditional method of RFID/ear-tag. Research by Qiao et al. [49] and Li et al.[33] indicates that sequence-aware and fusion-based DL models (e.g., hybrid CNN+LSTM and Multiple Feature Decision Layer Fusion Networks) are able to utilize spatial and temporal cues of video/images and attain the identification rates of up to 91 percent and 95.74 percent. Moreover, muzzle-based biometric identification by muzzle texture has shown extremely high accuracy (e.g., Li et al.[33] claiming 98.7 percent with VGG16-BN on muzzle images), which confirms the argument that muzzle texture may also be used as a reliable biometric marker. Multi-view strategies are also useful in enhancing robustness: Bergamini et al.[4] also demonstrated 81.7% Top-1 accuracy (and even higher Top-3) in fusing the multi-view representations of cattle re-identification, which showed the usefulness of combining frontal/side views in order to reduce confusion

between individuals. Moreover, tracking-based identification is also investigated Mon et al. [42] and enhances practicality in real-world settings.

There is also the non-invasive, image-based DL technology that has revolutionized cattle identification, going beyond the normal RFID methods and ear-tagging. The effectiveness of hybrid CNN+LSTM and Multiple Feature Decision Layer Fusion Networks (MFDN) was demonstrated by Qiao et al. [50] and Li et al. [33] with the identification accuracies of 91% and 95.74% by using spatial and temporal properties of images of cattle muzzle pattern, facial features, and walking patterns. High-quality RGB images of the patterns of cattle muzzles have been used in another study with 98.7% accuracy on the VGG16-BN example, showing how promising biometric markers similar to those of fingerprints may be. Even further identification was conducted by a multi-view CNN architecture, as applied to the Italian farms case study, with 81.7% of Top-1 accuracy after fusing frontal and side profiles. Such developments prove the trustworthiness of DL in solving problems such as inconsistent lighting and breed dissimilarities, and providing scalable answers to livestock traceability monitoring and theft[4]. Computer Vision and ML have been highly useful in weight estimation in livestock, with implications of non-invasive solutions to weight estimation procedures, which are labour-intensive. Studies performed by Cominotte et al. [6], [54, 55], and Dang et al. [8] utilized RGB-D cameras and 3D reconstruction graphs, with a high accuracy model, such as Artificial Neural Networks (ANN), Random Forest, and Support Vector Regression (SVR). As an example, the ANN models had an R^2 of 0.91 and the RMSE of 8.63Kg, whereas multi-modal MRGBDM models had an accuracy of 91.6 with an MAPE of 8.4. They employ a multi-camera system, semantic segmentation with models such as YOLOv4, and morphometric quantities such as chest girth and body length, which also result in a decrease of stress on animals to a significant level. Xu et al. [68] and others also proved the effectiveness of the DL model, such as ResNet-101-D with Squeeze-and-Excitation, in which the mean absolute errors were also reduced to low (e.g., 13.11 pounds). Regardless of these successes, there are issues still to be addressed, including limited dataset sizes, imbalanced classes, and farm-dependent variability, which could be overcome using larger and standardized datasets and strong feature selection to achieve feasibility and portability in precision livestock farming.

Authors	Year	Method Used	Dataset	Results
Qiao et al.[50]	2019	Hybrid CNN+LSTM (Inception-V3)	516 video sequences of 41 cows (20 frames, 401x506 pixels)	91% accuracy for 20-frame sequences, 57% for single images
Zin et al.[72]	2019	Geometric analysis, polynomial regression, multiple linear regression, Markov Chain	75 top-view cow images (50 train, 25 test) from Corfilac, IPLAB, Penn Vet	$R^2=0.7179$, adjusted $R^2=0.6281$, $R=0.8473$, $MSE=0.0366$
Cominotte et al.[6]	2020	Multiple Linear Regression, Partial Least Squares, LASSO, ANN	234 depth images of Nellore steers (Kinect RGB-D)	ANN: $R^2=0.91$, RMSEP=8.63 kg (weaning); PLS best in final feedlot
Ruchay et al.[54]	2020	3D imaging with Microsoft Kinect v2, switching bilateral filters, non-rigid ICP	103 annotated cattle instances (Hereford)	Measurement errors <3% at 90% confidence for 9 morphological parameters
Motohashi et al.[43]	2020	Random Forest, SVM, Boruta feature selection, RENN undersampling	10,000 samples from DeLaval VMS and Herd Navigator (Japanese farms)	Random Forest: 81% sensitivity, 46% precision; better than LDH-based alerts
Dohmen et al.[14]	2021	Mask R-CNN, custom CNN (Keras/TensorFlow)	63 crossbred heifers (37-370 kg, side/top-view images)	Top-view: $R^2=0.96$, RMSE=20 kg; side-view: $R^2=0.91$, RMSE=26.7 kg

Authors	Year	Method Used	Dataset	Results
Mahmood & Hamed[39]	2022	SVM, J48, J48 Graft, Random Tree, Random Forest, CSForest, SMOTE	6,600 samples, 15 physiological/anatomical parameters (5,490 healthy, 1,110 mastitis)	Decision trees: 100% accuracy; SVM: 94.7% accuracy post-SMOTE
Ruchay et al.[55]	2022	CNN (MRGBDM, MRGB, MDM), EfficientNet, YOLOv4	31,000+ images (154 Hereford), 28,630 images (121 Aberdeen Angus)	MRGBDM: 91.6% accuracy, MAPE=8.4
Dang et al.[8]	2023	PointNet segmentation, CatBoost, LightGBM, Polynomial Regression, Random Forest, XGBoost	1,190 point cloud meshes (270 cattle, various postures)	Random Forest: MAE=25.2 kg, MAPE=5.81
Phulu et al.[49]	2024	Random Forest, MobileNetV2, transfer learning, data augmentation	Ontology-based symptom data, labeled cow images (Healthy/Infected)	Random Forest: 90.62% accuracy; MobileNetV2: 91.1% accuracy
Guimarães de Paula et al.[20]	2024	SLEAP, LEAP, U-Net, ResNet-50/101/152	500 sow videos (Brazil, side/top views)	LEAP: mAP=0.90 (lateral), 0.72 (dorsal); OKS=0.94 (lateral), 0.86 (dorsal)
Li et al.[33]	2024	Multiple Feature Decision Layer Fusion Network (MFDN), SOLOv2, FaceNet, PP-OCRv4	QY-Cattle, NQ-Cattle datasets (front/side/top images)	95.74% accuracy across datasets
Xu et al.[68]	2024	ResNet-101-D with SE, MLR, DT, SVM, GR, BP neural networks	Farm-collected cow images (top/side views)	BP neural network: MAE=13.11 pounds, RMSE=22.73 pounds

Authors	Year	Method Used	Dataset	Results
Bhardwaj et al.[5]	2024	RF, KNN, SVM, LGB, XGBoost, CatBoost	Kaggle dataset (preprocessed, feature-selected)	CatBoost: 83.7% accuracy; Random Forest: 83.5% accuracy
Tawheed et al.[66]	2019	OLS, MLR, SVM, Decision Tree	approx 1,300 rows from Savar Dairy Farm, BLRI, Meghdubi Agro	Higher errors for some breeds due to limited data; SVM (Sigmoid) discarded
Garcia et al.[16]	2021	DT, RF, GB, KNN, Isolation Forest	104 cattle (birth to slaughter, “El Rosario” farm)	DT/RF: MAE 5.4525; Isolation Forest flags anomalies
Guvenoglu & Erdal[21]	2023	DeepLab v3+, ANN (ANN-1, ANN-2, ANN-3)	85 animals (stereo images, side/back views)	ANN-3: predicted weight 416.4 kg (actual 411.95 kg), error ± 20 kg
D. W. Girmaw[19]	2025	EfficientNetB7, DenseNet201, MobileNetV2	1,405 skin images (cattle, sheep, goats; 4 classes)	EfficientNetB7: 99.01% test accuracy; MobileNetV2: 95.31%
Lavanya et al.[32]	2023	CNN, LSTM, ResNet50, SGDM optimizer	450 images (normal, mild/severe lumpy skin disease)	95.7% accuracy with segmentation; 95% precision/recall/F1
R. M. D. S. M. Chandrarathna et al.[52]	2022	MobileNetV2, InceptionV3, GRU	1,123 breed images, 454 disease images, 248 videos	Disease detection: 94% (FMD), 99% (Bovine Johne’s); age/weight estimation effective

Authors	Year	Method Used	Dataset	Results
Kumar Ghosh et al.[30]	2023	Random Forest, Decision Tree, KNN, InceptionV3, ResNet50	1,341 images from 1,100 cows (Normal/Abnormal milk)	InceptionV3: 99.34% accuracy; Random Forest: 99% accuracy
G. Li, G. E. Erickson, & Y. Xiong[34]	2022	VGG16_BN, MobileNetV3_Small, 59 models	4,923 RGB images (268 cattle, muzzle patterns)	VGG16_BN: 98.7% accuracy; MobileNetV3_Small: 28.3 ms processing
Bergamini et al.[4]	2020	Dual-branch CNN (ConBlocks, ResBlocks), Histogram Loss	12,952 train, 4,850 test images (387 cows, 4 farms)	81.7% Top-1, 89.1% Top-3 accuracy
Setiawan et al.[59]	2024	Linear Regression, RF, SVR, KNN, MLP, GB, LightGBM, XGBoost	150 dairy cattle (10 anatomical measurements)	SVR: MAE=0.09 kg, RMSE=0.88 kg, $R^2=0.97$
Sarini et al.[58]	2023	LR, RF, ET, SVR, KNN, GB	228 bulls, 211 female Bali cattle (morphometric data)	LR (female): RMSE=9.6172; RF (male): RMSE=9.8952
Awasthi et al.[3]	2025	XGBoost, linear interpolation, outlier removal	46,585 walk-over weight records (1,105 cattle)	CCC=0.84/0.32, RMSE=32.6/38.3 kg, $R^2=0.77/0.73$
Mon et al.[42]	2025	YOLOv8, VGG16, SVM	Videos from 3 Japanese farms (1,263 cattle)	Identification: 81.25-100% accuracy; mAP@0.5: 97.2-98.6%
W.Ma et al.[37]	2024	3D reconstruction, linear regression, neural networks	Non-public datasets (pigs, cows, lambs, sheep)	Neural networks: MAE=1.16 kg (pigs); linear regression: 4.87-9.7% error

Authors	Year	Method Used	Dataset	Results
Z. Dinana et al.[13]	2023	RT-PCR (Universal primers)	25 cattle samples (Lamongan & Surabaya)	58% samples positive for FMDV
H. J. Shah et al.[60]	2023	Review of ML classifiers (NB, KNN, SVM, RF)	Publicly available datasets (Survey)	Consolidation of prior disease prediction methods
S. M. Saqib et al.[57]	2024	MobileRMSNet (MobileNetV2 + RMSProp)	Healthy and Lumpy Skin Disease image dataset	Classification accuracy: 95%
A. Iqbal et al.[26]	2025	YOLOv4-tiny, ResNeXt-50 (Contrastive Learning), kNN	Microscopic blood smear images (Bovine anaplasmosis)	Accuracy: 91.30%; Specificity: 92.83%
O. Andurkar et al.[2]	2025	Custom CNN, IoT sensors (MLX90614, MAX30102)	Images (FMD, LSD, IBK) & Physiological sensor data	Classification accuracy: 99.13%
Rozendo et al.[53]	2025	Side-view RGB; pose/segmentation; monocular depth; 3D reconstruction; regression (OLS/LASSO/RF/ANN)	Public Kaggle datasets (side-view RGB)	Best model reported: LASSO regression
Li et al.[35]	2025	Multi-view fusion; instance segmentation (EMA-YOLO11-seg); two-stream CBAM-ResNet50-SE regression	SSRN preprint	Reported: MAE = 8.62 kg, R ² = 0.9986
Afridi et al[1]	2023	Stereo/multi-camera 3D mesh; PointNet segmentation; regression (RF best reported)	Real-farm incomplete 3D shapes	Reported: MAE = 25.2 kg; MAPE = 5.85%

Authors	Year	Method Used	Dataset	Results
M. J. Hos-sain et al.[23]	2025	Custom CNN (3Conv3Dense) vs EfficientNetB3 + ML baselines; YOLOv5 detection; RFE; LIME (XAI)	2D RGB cattle images	Reported: MAE = 18.02 kg; RMSE = 19.85 kg
Y. Peng et al.[48]	2024	YOLOv8 (pos-ture + ID + detection); body-parameter ex-traction; linear regression for LBW	2D video/images of yak heifers (pen setting)	High accuracy reported
J. Lassen et al.[31]	2023	Commercial 3D camera system; repeated BW prediction + feed intake monitoring	>9,000 cows; 19 herds; 3 breeds	Main finding: high repeatability for predicted body weight)
G. Ge-breyesus et al.[17]	2023	Supervised ML on contour features from 3D digital images; ensembles (CatBoost, RF)	Large 3D image dataset (multi-breed)	Tree ensem-bles best (Cat-Boost/RF)
M. Matvieiev et al.[40]	2025	Data mining regressors: CART vs MARS using linear body measurements	Ukrainian Beef Cattle measure-ments	CART outper-forms MARS
S. Myat Noe et al.[45]	2023	Detectors (e.g., YOLOv5/YOLOv7, Detectron2) + MOT (Deep-SORT/StrongSORT + custom Re-ID)	Real-farm black cattle videos/images	High long-term tracking perfor-mance reported

Authors	Year	Method Used	Dataset	Results
Mg et al.[41]	2024	Detectron2 + Customized Tracking Algorithm (CTA); comparison vs SORT/DeepSORT	Crowded calving pen videos ByteTrack/etc.	Reported: up to 99% detection+tracking accuracy
C. C. Mar et al.[?]	2023	HTC instance segmentation + multi-feature tracking (spatial/appearance/texture/CNN)	Long-duration calving-pen videos (2 farms)	Reported: up to 99.6% MOTA
A. Fuentes et al.[15]	2023	Multiview action recognition + tracking/ID for behavior monitoring	Closed barn multi-view videos	High action recognition + stable tracking reported
P. Nilchuen et al.[44]	2025	Smartphone side-view imaging + YOLOv11m landmark detection (HD, BL) + regression for BW; deployed via LINE chatbot	Real-farm Brahman side-view images (12,660 annotated; BW set: 523 animals)	Reported: Precision 99.85%, Recall 100%, F1 99.92%, IoU 97.68%; BW MAE 43.44 kg (MAPE 8.91%)
Yang et al.[69]	2024	RetinaFace (MobileNet backbone) + improved FaceNet (MobileNetV1; Triplet+CE loss)	Cow face images in natural scenes	Outperforms YOLOX-based ID
A. Zhao et al.[70]	2025	UWB localization + YOLOv5 detection + coordinate matching; BP-NN error compensation	Farm with UWB-tagged cows + fixed camera	Reported: RMSE \approx 0.043 m localization

Table 2.1: Summary of Key Findings

Chapter 3

Requirements, Impacts and Constraints

3.1 Final Specifications and Requirements

The chapter establishes the functional requirements, design specifications, and operational constraints of the proposed cattle monitoring ecosystem using Computer Vision. The system allows three closely interdependent but independent operating units: hierarchical multi-disease classification with severity grading, four-sided multi-view individual cattle identification, and the RGB image-based body weight estimation. The requirements listed in this section are used to formulate the dataset, conventions of pre-processing, modeling, testing, and feasibility of implementation to provide reliability, robustness, reproducibility, and applicability of the recommended system to actual agricultural contexts.

The main system design limitation is that the new system depends on the compatibility of the cameras and mobile devices that are widely available on farms. Each of the components is applied to a single Computer Vision system to ensure that the disease identification, detection, and weight estimation can be done either individually or in combination without there being a discrepancy in the pre-processing, evaluation criterion, or implementation assumption.

3.1.1 Data Requirements

The proposed system takes three complementary datasets that relate to the classification of diseases and their severity, multi-view individual cattle identification, and body weight estimation. They each have a dataset that is supposed to be representative of the actual farm condition and is task-specific to enhance successful learning. .

The dataset of the disease is in the form of RGB images that are classified into four classes: the FMD, LSD, IBK, and Healthy cattle. The sick specimens are also annotated with the three clinically important stages of severity: mild, moderate, and severe, according to standardized veterinary diagnostic criteria. The annotation of severity is carried out by a licensed District Livestock Officer (DLO) to make sure that it is clinical and consistently relevant. The dataset is hierarchically labeled to perceive learning and evaluation on the disease level, severity level, and integrated fine-grained level. The stratified dataset splitting is used to maintain the distributions of disease severity between training, validation, and test sets and address the issue of natural imbalance of classes in the livestock surveillance data.

The RGB images of the cattle are stored in an individual cattle identification dataset, which is taken at four standard points of view, namely the left side, right side, front side, and back side. A cattle case is modeled by a predetermined collection of pictures that are obtained at various angles to obtain appearance variation. Cattle identities are seen as labels, and the dataset is split in a way that a cattle does not occur in the training and evaluation sets. This configuration mirrors reality, in which the system will have to identify cattle that have not been seen before during testing.

This is when the weight estimation dataset is built on the self-collected dataset of 215 different cattle, which is one of the main contributions of this work. Every cattle case is provided with its own identifier and related to structured metadata, such as live body weight, sex, and breed, and pre-linked with RGB images. In case it is available, several views of the same cattle are also stored to allow multi-view regression. The exception is that weight measurements are synchronized with image acquisition in order to reduce label noise since regression performance is very sensitive to annotation quality.

Data preparation and cleaning: Data preparation and cleaning will occur in the initial phase of data analysis prior to the subsequent phase, which is results analysis.

3.1.2 Preparation and cleaning of data

Data preparation and cleaning will be done at the first stage of data analysis before the second stage, which is results analysis. The first phase of data preparation will be strict quality control in all data sets to ensure reliability and consistency.

In the case of the disease dataset, the corrupted pictures are eliminated, duplication of the sample is eliminated as well as pictures that have no visible pathological clue are eliminated. Observable visual symptom cross-validates disease and severity labels to make sure that labels and image content are in agreement. All labels are normalized into a common hierarchical format in order to support common training and assessment of architectural variations.

In the case of the disease data, dataset-1 is used where the corrupted images, duplicate images, and images with no visible pathological signs are eliminated. Cross-validation with disease and severity labels is done on observable visual symptoms to make sure that there is consistency between annotations and picture content. All labels are brought to a common hierarchical representation to enable training and evaluation of architectural variations to be consistent.

In individual cattle identification and weight estimation, dataset-2 is where view consistency is important in data preparation. Checking of the identity level is done to verify that all four sided images of the same cattle along with proper naming are in the proper group, along with labeling. The necessary checks has been done to ensure the metadata contains all the necessary values for each and every cow which is essential in weight estimation. Unrealistic or conflicting mass records are eliminated, and image-label associations are thoroughly checked to prevent correspondence. View groupings are also maintained when there is an availability of multi-view samples, such that view supplementary shape and body-structure information is utilized by regression models.

3.1.3 Pre-processing Requirements

The purpose of pre-processing pipelines is to have a visual input where the pipelines are structured in such a way that they standardize the input and maintain task-relevant features.

To avoid distorting original body forms, all images are resized to a preset input resolution, but aspect ratios are kept. Simple image augmentation methods are used in the training to enhance resistance to changes in lighting conditions, viewing angle, and background conditions. The same pre-processing pipelines (except augmentation) are used in the validation process and testing to make sure that the performance is evaluated reliably and reproducibly.

In the case of multi-view cattle identification, the views are uniformly resized and normalized so that they are uniform. Image augmentation is implemented in a vigilant manner in order not to distort the appearance features that are important in individual differentiation, like body structure and characteristic patterns. Complete body exposure is highly imperative to maintain identity-specific cues in particular cases.

In the estimation of cattle weights, pre-processing aims at maintaining perceived body proportions. The transformation of images that may change in body shape is prevented. All of the views are processed the same to make sure that there are consistent features extracted and weights predicted along with the necessary consistent metadata.

3.1.4 Software and Environment Requirements

The implementation of the system is based on a Python-based DL framework using Pytorch [47], optionally accelerated by GPU. Standard libraries are used to do image processing, numerical computation, model training, and evaluation. The experimental reproducibility is guaranteed by the use of fixed random seeds, deterministic pre-processing in evaluation, and fixed partitioning of the dataset.

The model checkpoints, training logs, and evaluation results are saved in a well-organized and version-controlled form in order to streamline systematic experimentation and reporting. Even though it is recommended that the system uses GPU acceleration to increase training efficiency, the system may also run on CPU-only environments, thus being able to deploy to resource-constrained rural environments.

3.1.5 Elements of Modeling and Evaluation Requirements

The analysis of diseases based on transfer learning is conducted using deep CNNs. The classification of diseases is defined as a four-class problem, whereas the severity is graded only in the diseased samples to avoid the ambiguity of the severity scores of the healthy cattle. Accuracy, precision, recall, F1-score, and confusion matrices are used to assess performance at disease and severity levels.

Individual cattle identification is subjected as body appearance recognition exercise using a embedding based metric learning system for each cattle. Discriminative identity representations that are viewpoint-invariant are learned by training models. Testing is also carried out using protocol-based testing approaches, such as leave-one-view-out and controlled multi-view corss angle domain shift testing, to test the generalization of viewpoints

that are not observed. Rank-1 and Rank-k accuracy under identity-disjoint evaluation conditions to report the performance of identification.

The estimation of weight in cattle is formulated as a regression problem on the basis of CNNs in extracting the features and regression heads. Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2) are the measures of performance used. Predictions in view-aggregated images can be used to enhance a more stable estimation when the multi-view images are at hand, and the results are consistently reported according to the chosen evaluation protocol.

3.2 Societal Impact

The presented Computer Vision-based cattle monitoring system focuses on the major problems in livestock health control by allowing the detection of diseases at an early stage, organizing the severity of their effects, and tracking the cows at a cattle level using non-invasive visual monitoring. Detection of infectious diseases, including FMD, LSD, and IBK, helps in early veterinary intervention, where diseases are being contained before spreading further, leading to death and the losses incurred by the farmers.

Systematic animal-level record keeping is helped by reliable individual cattle identification, which would help track the history of diseases, consequences of treatments, and physical development of the animal over a long time. This is useful in making informed decisions on herd level, such as specific healthcare planning, breeding control, and productivity. The proposed system provides a non-invasive option compared to other traditional systems of identification that require physical tags or sensors, and hence causes less stress and maintenance cost to farmers.

The weight estimation module helps farmers by allowing them to track the growth and current body condition of cattle at low cost through non-invasive monitoring of cattle based on RGB images. This saves time spent on the manual weighing of the animals and reduces the stress of handing and enhances the farmers to make great decisions in feeding, health checking (early signs of abnormal weight loss) and on fair market prices particularly in rural areas where weighing machines and veterinary services might be scarce.

The system helps to reduce the necessity of repetitive physical inspections and diagnostic devices through the use of low-resolution RGB imaging and automated analysis of the results only. This will be especially useful in the rural and resource-challenged areas where access to veterinary care is limited. The system is specifically made as a decision-support system and is not a substitute for professional veterinary judgment, which will foster responsible clinical practice.

3.3 Environmental Impact

The environmental effects of the intended system are indirect, and they are brought about by better livestock management. Early diagnosis of the disease and grading of its severity lessen the number of unnecessary medical interventions and waste of resources in the context of the disease development. Image-based identification and weight estimation

also aids in maximum feeding plans and contributes to decreasing the feed wastage, as well as enhancing the resource utilization at the farm level.

On the technological side, the system focuses on optimality in terms of computing as it employs lightweight (due to lowshot) DL structures that can be implemented in edge and mobile computing devices. The use of RGB images only excludes the necessity of additional sensing devices, which helps save materials and electronic waste. These design options conform to the environmental sustainability principles of system implementation in technological use of agriculture.

3.4 Ethical Issues

Implementation of an automated decision-support system in livestock monitoring provokes ethical issues with regard to ownership of their data, transparency, and the responsible use of the systems. Even though the system works with animal-images and metadata and not human-data, it has been ensured that the farm-level image and metadata collection has not violated data ownership rights and seek informed consent of farmers and other interested parties.

The suggested system will support, and not substitute, veterinary specialists. Predictions made using models are to be considered through the perspective of known limitations of the system, and the ultimate diagnostic and treatment choices need to be made under the industry of professional veterinary guidance. It is important to communicate the uncertainty of the system, the confidence levels, and the possible failure mode clearly in order to avoid relying on the output of automated systems excessively and to ensure ethical deployment.

3.5 Standards

The system is tried and tested using the guidelines of ML and Computer Vision research. Data preparation is done by standardized data annotation, controlled data split, and identity-disjointed or partitioning when necessary, which provides a fair evaluation and prevents information leakage.

Reproducibility is encouraged by the use of fixed random seeds, standardized pre-processing pipelines, and uniform evaluation protocols. The performance reporting is based on accepted measures of the classification metrics, identification, and regression assignments, which make it meaningful to compare it with the available literature. The software development practices focus on modular design, documentation, and version control to accommodate transparency, repeatability, and long-term maintainability.

3.6 Project Management Plan

The project is implemented in a systematic and repetitive development cycle that involves data collection, pre-processing, model design, training, evaluation, and system integration. Computational resources are assigned based on experimental needs, and progress of the project is monitored by documenting regularly as well as reviewing milestones. This

methodological approach to management will make sure that there is successful delivery of project goals and that the methodological consistency and experimental rigor are consistent throughout all parts of the system.

Allocation of computational resources is done based on the needs of the experiment, and progress is monitored by regular documentation and review of milestones. This systematic management model will assure an appropriate accomplishment of the project goals and also methodological consistency and experimental rigor throughout all the system components.

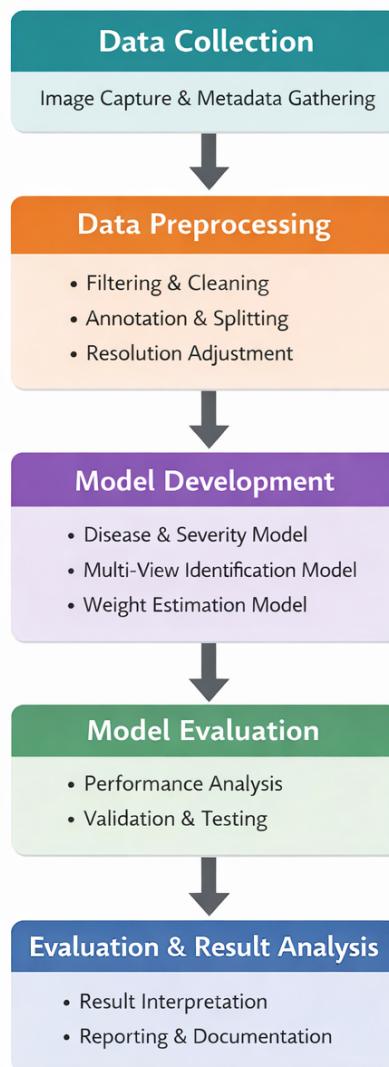


Figure 3.1: Project Management Plan and Development Workflow

3.7 Risk Management

Several risks are involved in the development and implementation of the proposed system. The risks associated with data are class imbalance, the inadequate presence of severe cases of disease, and the imbalanced image acquisition conditions. These dangers are addressed by stratified dataset division, data augmentation control, and hierarchical disease severity modeling.

The risks associated with models are overfitting, poor generalization in new farm setting and the visual similarity of cattle in the process of identification. The mitigation measures consist of identity-disjoint assessment protocols, multi-view learning, and more detailed class and identity-level performance analysis. Another risk that is caused by deployment is concerned with computational constraints and is mitigated by prioritizing efficient model architectures and lightweight pre-processing pipelines that can be deployed to edge and mobile platforms.

3.8 Economic Analysis

The key design feature of the proposed system is economic feasibility. The system can be used with only RGB images, and it does not need special sensors or weighing platforms, so its implementation and maintenance are much cheaper. Early detection of diseases and grading of severity helps lower the expenses of losses due to early treatment, the avoidance of loss of productivity, and the spread of diseases.

Individual cattle recognition and weight prediction on an image basis further grow the cost-effective cattle control through optimization of feeding plans, drug development, and surveillance over time. This is because the system does not need a lot of infrastructure to make it fit in the small farm holder and commercial farms, and allows the farms to engage in economically viable livestock management practices.

Chapter 4

Proposed Methodology

4.1 Methodology Overview

To classify diseases, our methodology develops the hierarchical cattle disease diagnosis system, which is a two-level task where the model predicts classes (Healthy, FMD, IBK, LSD), then provides severity to diseased cases: Stage-1 (mild), Stage-2 (moderate), and Stage-3 (severe). To guaranty a valid assessment with class imbalance, we had a fixed stratified 15 percent test sample and assessed the rest of the data with stratified 5-fold cross-validation to make the generalization estimates and equitable experimental comparisons. Five different modeling approaches have been experimented with, beginning with a basic single-pass baseline and onward to hierarchical separation to minimize inconsistency, followed by a shared-backbone learning using masking and ordinal staging. To be interpretable and analyze critical error cases, Grad-CAM++[47] was used to represent test predictions to identify the regions of images that determine decisions, which is useful to justify clinical relevance and explain typical failures.

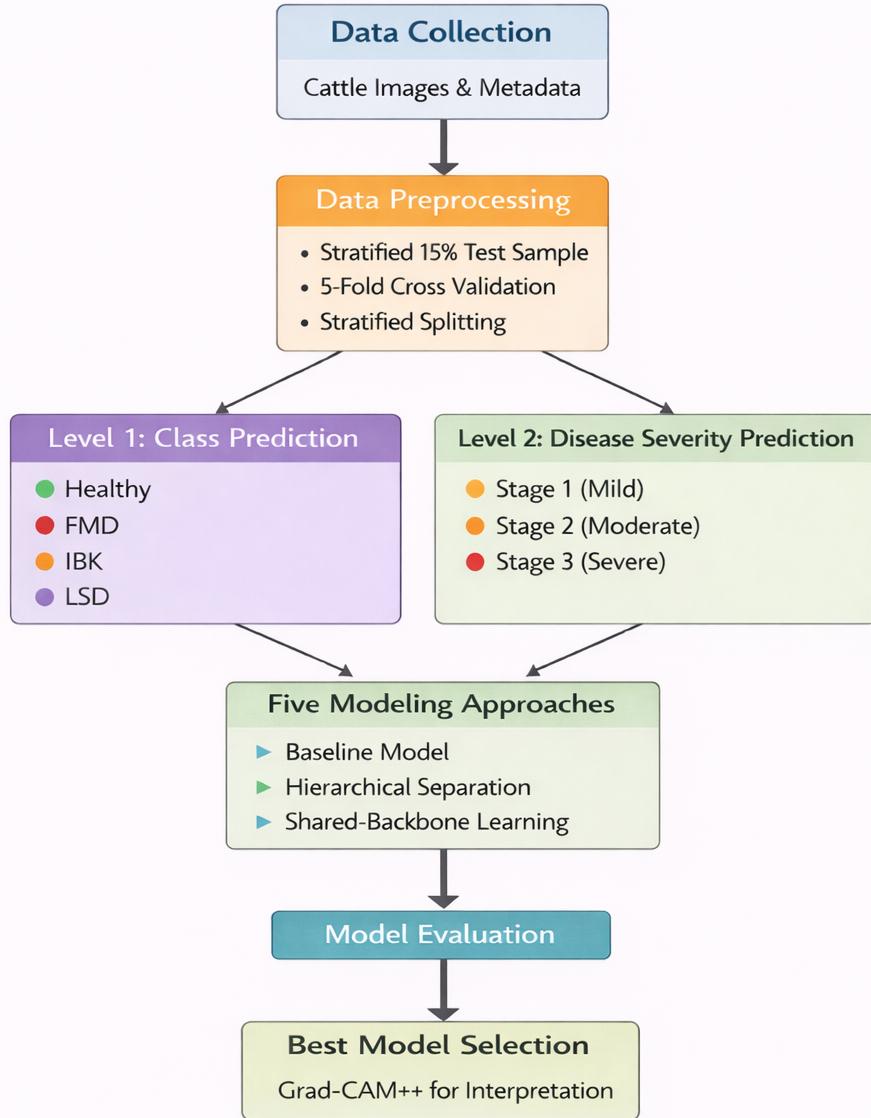


Figure 4.1: Methodology overview for cattle disease classification with severity grading system

For cattle identification, the methodology is built upon, supervised learning architecture that applies ArcFace [67] and SupCon [28] loss in a deep learning architecture to achieve high-quality feature embedding. The information is a multi-view image of a single cow at four different angles (front, back, left, right) and bounding box annotations in the

YOLOv8s [51] format identifying the object of interest. The first data processing stage consisted of parsing of these annotations to extract and visualize ground truth regions and then validation with YOLOv8s detector to measure the quality of these annotations as Intersection over Union (IoU) and cosine similarity between the ground truth and predicted crops. The core identification model is based on a convolutional neural network backbone which is pre-trained on ImageNet [10] and includes a projection head which learns to transform visual features into a high-dimensional embedding space with supervised contrastive learning. The training process takes into account ArcFace and SupCon combined loss, a method that injects an angular margin penalty to improve inter-class separability and intra-class compactness in the embedding space; thus, the embeddings will be more discriminative to identify each cow. Hyperparameter optimization targeted optimization of the parameter of the angular margin, the learning rate schedule, embedding dimension, temperature scaling of the contrastive loss, and composition of the batch to provide adequate positive and negative pairs per training iteration. The model has been end-to-end trained using data augmentation techniques such as random cropping, horizontal flipping, and jittering of colors to enhance stability in changing viewing angles and environmental variations. Performance measures were based on rank-k accuracy, mean average precision, and equal error rate in held-out test sets, where special emphasis was placed on cross-angle identification conditions where query and gallery images depict viewing different faces of the same person.

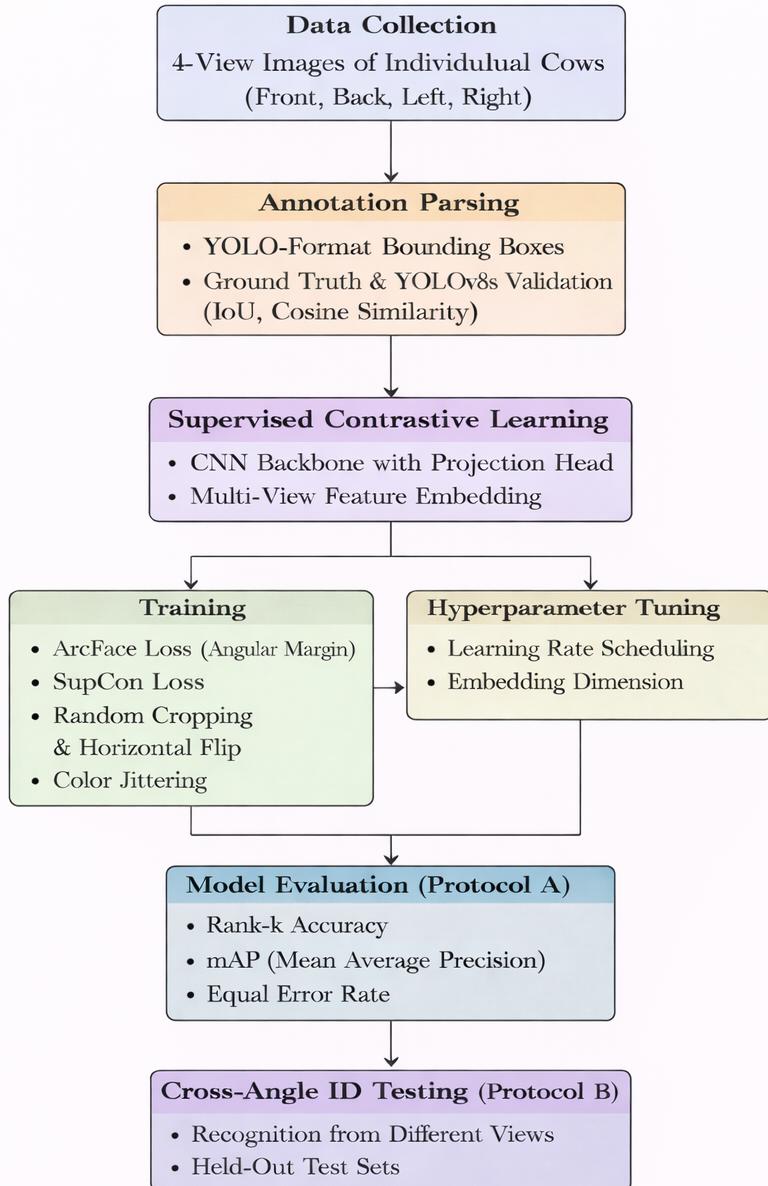


Figure 4.2: Methodology overview for unique cattle identification system

To estimate cattle weight, our research uses a full DL model of cattle weight estimation by multi-view image analysis. The transfer learning process is performed by the state-of-the-art CNN based architectures to extract certain visual features of the Cattle 4-sided images (front, back, left, and right view). In the design process, the collection and pre-processing of data, the model architecture and configuration, optimization of

training, and ensemble learning strategies were adopted to obtain strong and precise weight predictions. The general procedure: (1) multi-view image data collection and pre-processing, (2) transfer learning with pre-trained CNN backbones, (3) domain adaptation by fine-tuning approaches, (4) building up an ensemble model for better generalization, and (5) evaluation of comprehensive performances by a variety of metrics.

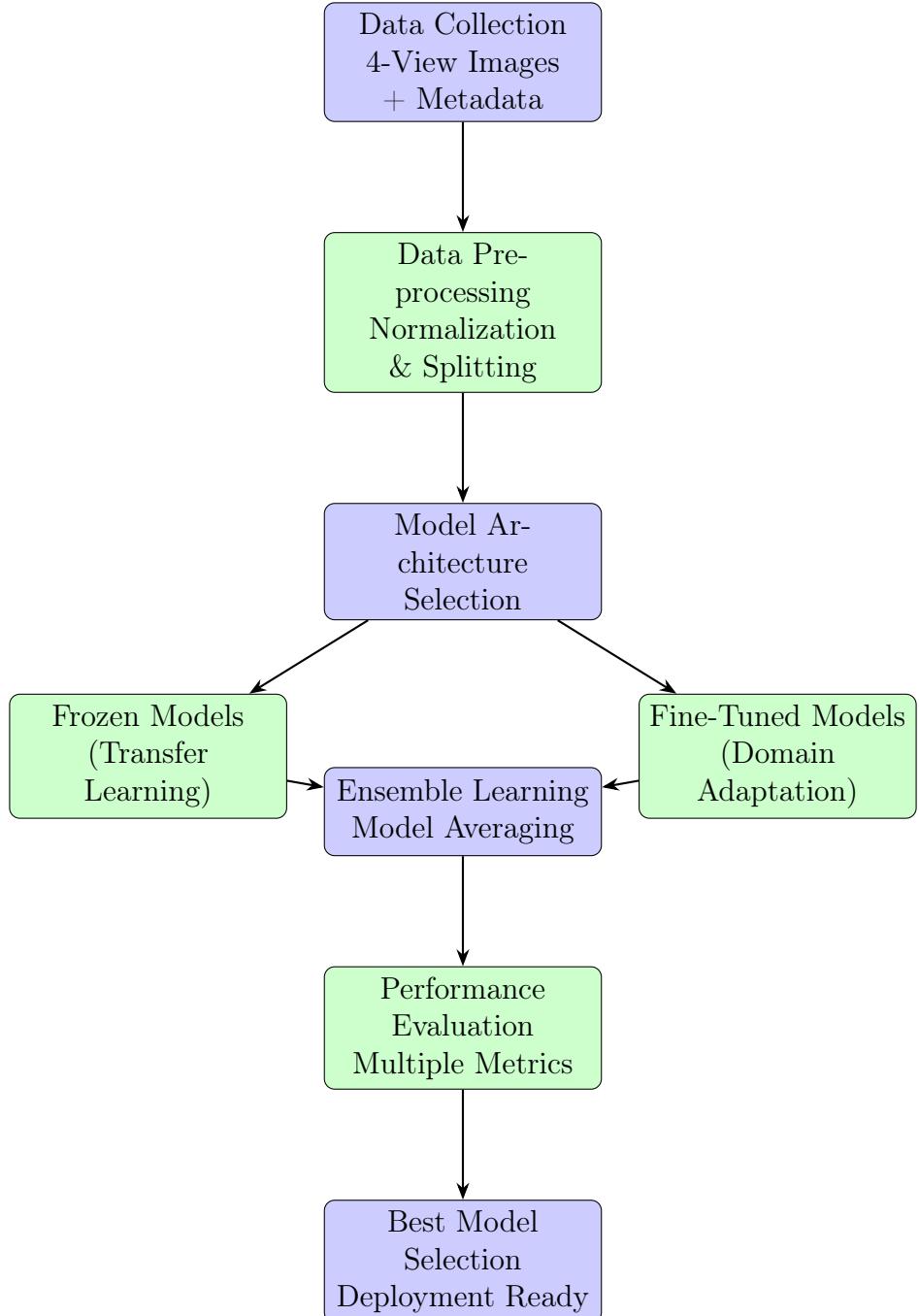


Figure 4.3: Methodology overview for cattle weight estimation system

4.2 Preliminary Design

The preliminary design for **disease classification** begin with defining the hierarchical label space which is necessary because severity is undefined for healthy cattle and staging cues are typically subtler than disease cues. Disease classification depends on disease-specific patterns such as vesicular lesions for FMD, ocular opacity and discharge for IBK, and nodular lesions for LSD. Severity staging depends more on extent, intensity, distribution, and systemic signs visible in imagery. To test how the hierarchy should be modeled, five options were designed with EfficientNet-B1[65] backbone and compared under identical data splits and a shared training protocol.

- **Option A - Flat 10-class:** The flat 10-class model represented the simplest baseline by treating the task as a single 10-way classification problem with the following labels:

- Healthy
- FMD_Stage-1, FMD_Stage-2, FMD_Stage-3
- IBK_Stage-1, IBK_Stage-2, IBK_Stage-3
- LSD_Stage-1, LSD_Stage-2, LSD_Stage-3

It was used to test whether the hierarchy can be learned implicitly without architectural guidance; however, it is expected to struggle because it treats all errors equally and does not enforce logical structure.

- **Cascaded two-stage:** The cascaded pipeline modeled the hierarchy explicitly by first predicting the disease class (Healthy/FMD/IBK/LSD) and then predicting severity only when the output is a disease class. This option was chosen because it aligns with clinical reasoning and isolates staging from the healthy class, but it was expected to be sensitive to error propagation since a wrong disease prediction makes the severity decision unreliable.
- **Shared-backbone multi-task:** Shared-backbone multi-task was a single feature extractor with two prediction heads, one designed to predict the disease and one to predict the severity of the disease, thus both tasks are trained on the same visual representations (lesions, inflammation, changes in texture). This design was aimed at being efficient at inference with a single forward pass and staging that is better by reusing disease-relevant features.
- **Masked multi-task:** The masked multi-task choice was an extension of shared multi-task learning with the computation of severity loss only in diseased test samples. This masking was required due to the fact that severity is not defined in healthy cattle and without masking the model may receive misleading gradients which decreases staging reliability and generates logically conflicting learning cues.
- **Ordinal-aware severity modeling:** The ordinal-conscious approach used severity stages as ordered and not discrete classes, since Stage-1, Stage-2 and Stage-3 are

progressive stages of disease presence. This was meant to minimize the adjacent-stage confusion by instilling the mild - moderate - severe structure into the learning objective that is more clinical.

The proposed **cattle identification** system employs a two-phase pipeline, the first step of which consists of a YOLOv8s detection module where input images are resized into 640x640 and the output of the detection module is the coordinates of detected cattle bounding box-coordinates (x_1, y_1, x_2, y_2) with the highest confidence score. The identified regions are then set to be cropped and normalized to 224 x 224 pixels in the second step and a Deep CNN-based re-identity module is used that uses metric learning to extract features and produce 512-dimensional L2-normalized embedding vectors. This module will be used to uniquely identify an individual cattle through learning high-dimensional visual embeddings, as opposed to simple classification. The formulated problem is defined as a deep metric learning problem, in which CNN backbones (ResNet-50 [22] and ConvNeXt-Tiny [36]) are used to project images of the same cow to clustered vectors in a 512-dimensional space. In order to guarantee strong discrimination, the hybrid loss is a combination of Supervised Contrastive SupCon loss and ArcFace loss are used by the module. These two-fold nature explicitly works to maximize the inter-identity distance using angular margins and, at the same time, optimize the intra-identity distance to the minimum to enable the model to generalize with just a little labeled data per individual.

The **weight estimation** module is also a regression-based DL system, which involves learning a non-linear and complex mapping between a set of multi-view images of cattle and the body weights of the cows. The preliminary design considerations are:

Architecture Selection: Three CNN state-of-the-art architectures were selected on the basis of their demonstrated performance in the computer vision task, and constraints in the realization of the trade-off between computational speed and accuracy:

- **DenseNet121:** Dense Convolutional Network with 121 layers [25], for dense connectivity patterns that promote feature reuse and gradient flow is chosen for its parameter efficiency (7.98M parameters) and strong feature extraction capabilities.
- **MobileNetV2:** Lightweight architecture (3.47M parameters) with depthwise separable convolutions and inverted residual blocks which is chosen for its computational efficiency and potential deployment on resource-limited devices [56].
- **InceptionV3:** Multi-scale feature extraction architecture (23.85M parameters) using inception modules with parallel convolutional paths that is selected for its ability to capture features at multiple spatial scales simultaneously [64].

Multi-View Integration Strategy: This preliminary design incorporates four views of each cattle in a shared weight feature extractor approach:

$$\mathbf{F}_{\text{view}} = \text{CNN}_\theta(\mathbf{I}_{\text{view}}) \quad \text{for view} \in \{\text{front, back, left, right}\} \quad (4.1)$$

where \mathbf{I}_{view} represents the input image for each view, and CNN_θ is the shared CNN backbone with parameters θ .

The extracted features from all views are then concatenated and processed through a regression head:

$$\mathbf{F}_{\text{combined}} = \text{Concat}(\mathbf{F}_{\text{front}}, \mathbf{F}_{\text{back}}, \mathbf{F}_{\text{left}}, \mathbf{F}_{\text{right}}) \quad (4.2)$$

$$\hat{w} = f_{\text{reg}}(\mathbf{F}_{\text{combined}}; \phi) \quad (4.3)$$

where f_{reg} is the regression head network with parameters ϕ , and \hat{w} is the predicted weight.

Transfer Learning vs Fine-Tuning: Two training paradigms were designed:

1. **Frozen Feature Extraction (v1 models):** Pre-trained ImageNet weights remain frozen, and only the regression head is trained. This approach leverages general visual features with minimizing computational cost.
2. **Fine-Tuning (v2 models):** Top layers of the pre-trained backbone are unfrozen and jointly trained with the regression head to enable domain-specific feature adaptation.

4.3 Data Collection

The **diseases dataset (Dataset-1)** was compiled by combining several public sources and a special field collection in order to cover all the target conditions. LSD, FMD and most of the Healthy cattle images were obtained by combining two Kaggle datasets [12, 62] and one Robo-flow dataset [63], whereas IBK images and few Healthy cattle images were obtained as a separate and locally collected dataset that are not represented by the public sources. The combined approach was required since none of the datasets had enough samples or classes for all the desired disease types to develop the model.

For **identification and weight estimation**, a detailed dataset (**Dataset-2**) with 215 samples of cattle was collected locally, consisting of different breeds of cattle, different ages, and different weight classes, so they can finalize the model with various cattle characteristics.[1, 35] The dataset was systematically collected from 15 different locations where each cattle was photographed from four different views (front, back, left, right) using mobile phone cameras under farm environmental conditions and metadata with cattle ID, age(years), sex, breed, and live weight (kg), which was measured in weight scale. The data collection process was a complex process as it needed the consent of the farm owners, an available manual scale for collecting live weight and most of the farm owners do not want to volunteer for this hassle because of security or privacy concerns.

4.3.1 Data Cleaning

For **disease classification, dataset-1**, data cleaning is necessary to make the merged dataset fit well to the same supervised learning: the inappropriate and duplicate samples were eliminated before the process of training. To prevent unwanted artifact generation in the pre-processing, all the files were checked as readable, corrupted images were eliminated, and standardized to a valid JPEG format. Then python library "imagededup" [9] was used to detect and remove duplicates and near-duplicates, along with a manual duplicate or near-duplicate check, which is important since duplicates in particular can

lead to train-test leakage and artificially inflate the performance based on memorization as opposed to actual generalization.

For identification and weight estimation, dataset-2, the raw dataset, consisting of 860 images across 215 unique individuals, was manually labeled using the CVAT [7] tool. This process isolates the cattle region of interest (ROI) from complex farm backgrounds, generating tight bounding box crops with no padding. This step effectively removes irrelevant background noise and ensures that the model focuses exclusively on the cattle’s discriminative visual features.

The data cleaning process involved a number of critical steps to ensure that the data is of high quality and consistent:

Missing Value Detection and Removal: Each record was validated to ensure it was complete (all required fields, which include ID, weight, and all four view images, were present). Records for cattle for which cattle ID and/or weight were unmeasured were removed from the dataset. During the loading phase, the system verified the existence of all four view images and then included a sample after all the checks.

Image Availability Validation: A flexible image loading mechanism was provided for the different naming conventions and .jpg, .JPG, .jpeg, .png file extensions. Verification done to ensure that all (front, back, left, right) were available for each cattle sample. In order to keep the data consistent, cattle that had incomplete view sets were excluded.

Data Integrity Checks: The consistency of metadata was checked by relating cattle ID with the image directory. Weight measurements were endorsed to ensure that they were within biologically possible ranges (59 kg to 654 kilos in the final data).

Cleaning Results: After these steps, 215 samples successfully passed the cleaning phase with complete four-view image sets and valid weight measurements.

4.3.2 Data Transformation

For disease classification, dataset-1, data transformation includes processed images and labels in a form suitable for transfer learning while preserving clinical meaning. All images were converted to RGB, resized to the backbone-specific input size (240×240 for EfficientNet-B1), and normalized using ImageNet mean and standard deviation to match pretrained feature statistics. In the original collected datasets, severity stage labels were not provided. Since severity detection is clinically crucial for treatment prioritization, each diseased image was therefore assigned a severity stage: Stage-1 (mild), Stage-2 (moderate), and Stage-3 (severe) using specific symptom-based criteria, and the annotations were validated by **Dr. Mahbubur Rahman (District Livestock Officer, Naogaon)** to ensure consistent staging across FMD, IBK, and LSD.

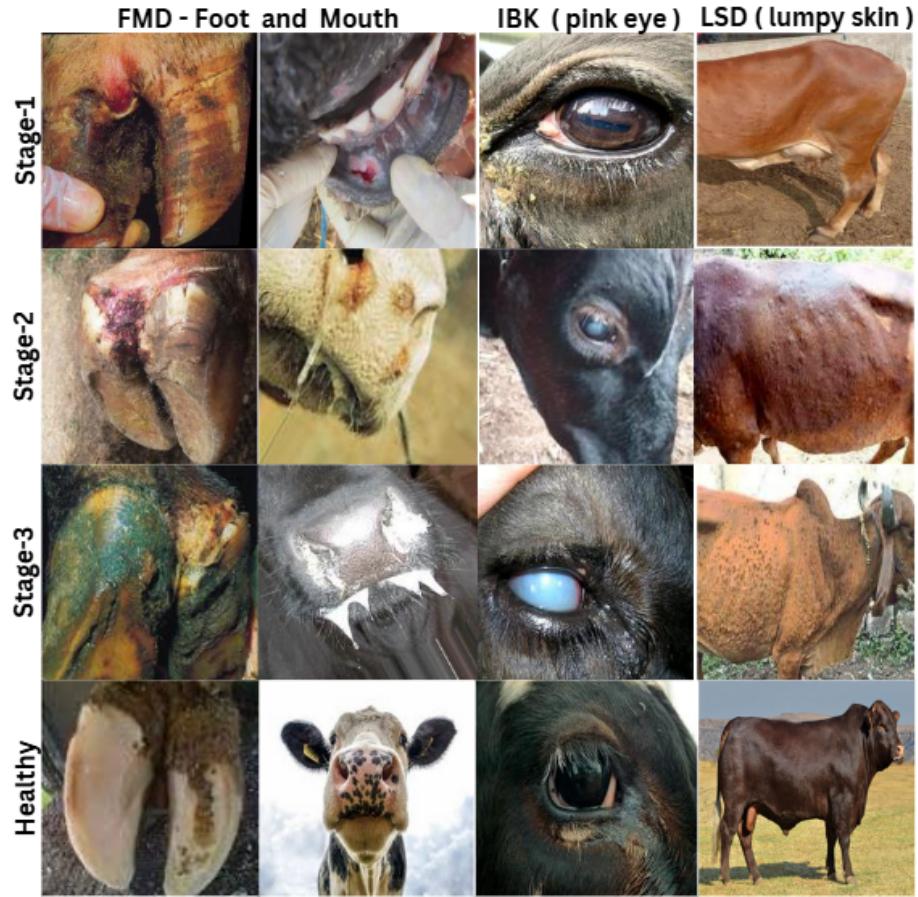


Figure 4.4: Disease dataset example after severity labeling

For identification and weight estimation, dataset-2, to satisfy the input requirements of the CNN backbones, all images are resized to a fixed resolution as necessary for specific architecture using bi-cubic interpolation. RGB values are normalized using standard ImageNet statistics to facilitate transfer learning. During training, aggressive augmentation is applied to prevent overfitting, which includes random horizontal flipping ($p = 0.5$), color jittering (brightness, contrast, saturation), and random erasing to simulate occlusion. Validation and testing transformations remain deterministic, as only resizing and normalization are applied there. The actual dataset images are showcased below.



Figure 4.5: Locally acquired dataset for identification and weight estimation

Multiple transformation steps were applied to prepare the raw data for DL model consumption:

Object Detection and Cropping:

Before using the standard image pre-processing pipeline, the object detection model of YOLOv8s (You Only Look Once version 8) was used to automatically detect and extract the bounding boxes of cattle in the images collected by us. This pre-processing step ensures that the model specifically focuses on the cattle and removes the irrelevant background information and improves the quality of the feature extraction.

- **YOLOv8s Detection:** Pre-trained YOLOv8s model used to detect cattle in each view
- **Bounding Box Extraction:** Detected regions cropped with no padding
- **Quality Assurance:** Manual verification of detection accuracy for dataset consistency

Image pre-processing Pipeline:

1. **Resizing:** All the cropped cattle images are normally resized to a uniform dimension of 224×224 pixels in order to match the input requirements of the pre-trained CNN architectures using the bi-cubic interpolation to preserve the quality of the image.

2. **Normalization:** Pixel values were normalized from the original [0, 255] integer range to [0, 1] floating-point range:

$$\mathbf{I}_{\text{norm}} = \frac{\mathbf{I}_{\text{raw}}}{255.0} \quad (4.4)$$

This normalization here ensures stable gradient flow during training and aligns with the pre-processing that was used during ImageNet pre-training of the backbone networks.

3. **Multi-View Concatenation:** The four RGB images (each with 3 channels) were concatenated along the channel dimension, creating a 12-channel input tensor:

$$\mathbf{X} \in \mathbb{R}^{224 \times 224 \times 12} \quad (4.5)$$

where the channels are ordered as: [Front-R, Front-G, Front-B, Back-R, Back-G, Back-B, Left-R, Left-G, Left-B, Right-R, Right-G, Right-B].

4. **Channel Splitting:** A custom TensorFlow layer (`SplitChannels`) was implemented to decompose the 12-channel input back into four separate 3-channel RGB images within the model architecture, allowing the pre-trained networks to process each view independently:

$$\begin{aligned} \mathbf{I}_{\text{front}} &= \mathbf{X}[:, :, 0 : 3] \\ \mathbf{I}_{\text{back}} &= \mathbf{X}[:, :, 3 : 6] \\ \mathbf{I}_{\text{left}} &= \mathbf{X}[:, :, 6 : 9] \\ \mathbf{I}_{\text{right}} &= \mathbf{X}[:, :, 9 : 12] \end{aligned} \quad (4.6)$$

5. **Target Variable Processing:** For weight estimation regression, the weight measurements were retained in their original live weight kilogram (kg) units without scaling where the weights range from 59 kg to 654 kg, with a mean of approximately 265 kg.

4.3.3 Data Integration

In disease classification, dataset-1, the Kaggle- and Robo-flow collected LSD, FMD, most of the Healthy images were integrated with the uniquely collected IBK images and few Healthy images into one 10-class hierarchical taxonomy, where the labels of all sources aligned with each other on the severity of the disease. Since IBK images were imaged in varying real-world conditions relative to web datasets, the risk of source-style bias was introduced, which is why the process of integration was uniform on all images (the same pre-processing and augmentation pipelines had been used). To maintain fairness in evaluation in case of class imbalance, all splits were stratified with the 10-class labels, and stratified 5-fold cross-validation was performed on the training subset, and a fixed, An untouched 15% test set of 1,089 images was used only for final-stage reporting.

For identification, dataset-2 (only image with individual cattle id, no other metadata) is used for identification, and is structured to support robust evaluation through two different protocols. A 5-fold stratified cross-validation strategy is implemented, ensuring that images of the same individual are grouped together to prevent data leakage between

training and validation sets. Furthermore, specialized test subsets are integrated for specific evaluation scenarios: Protocol A for within-individual rotation, Protocol B for unseen cross-angle-view generalization (such as top view), and challenging conditions (occlusion or blur).

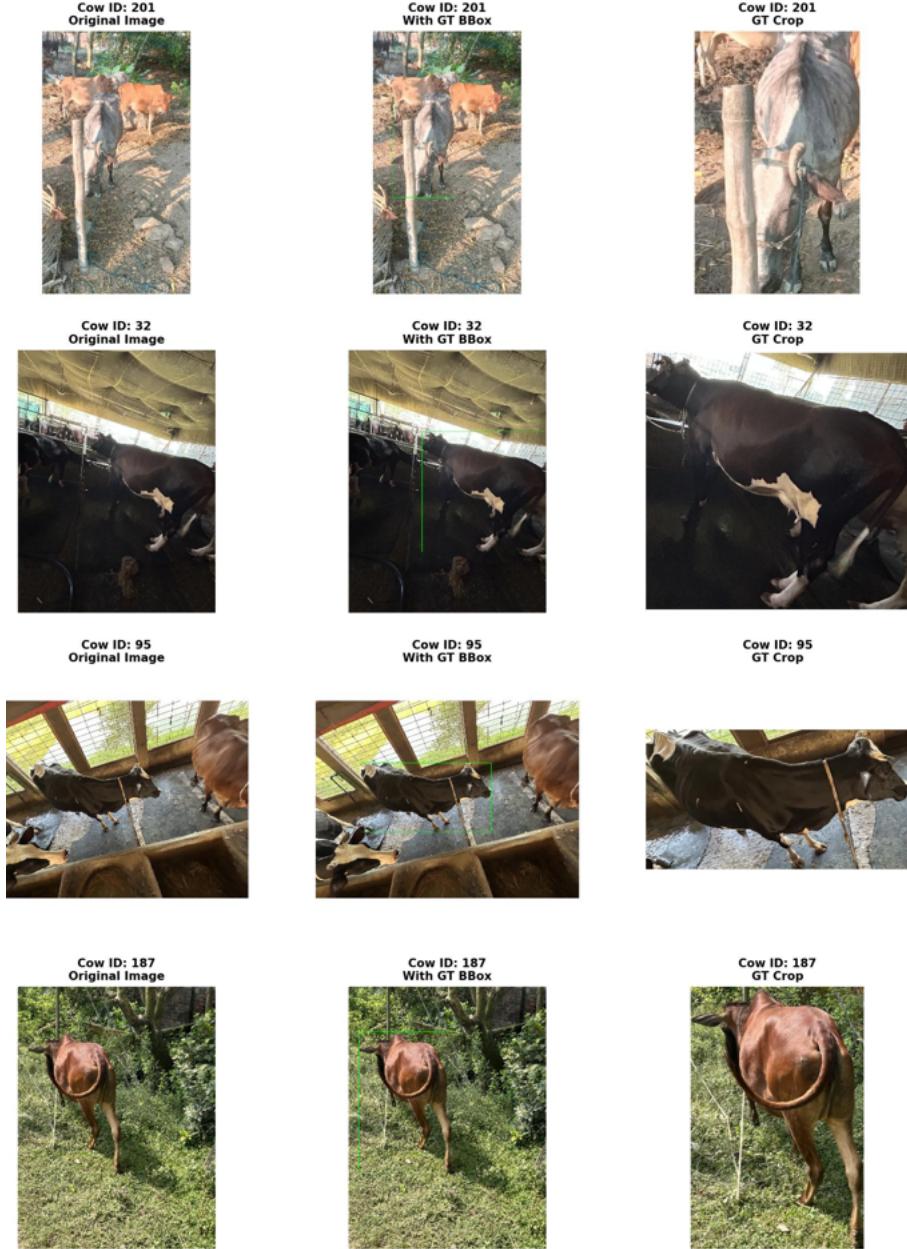


Figure 4.6: Identification dataset example for protocol B testing.

For weight estimation, dataset-2, the data integration phase used a combination of image data with related metadata to build up a united data structure in the form of a dataset:

Metadata-Image Linkage: A CSV file with cattle attributes (ID, Age, Weight, Sex, Breed) was linked with the multi-view images directory structure. The cattle ID was used as the primary key that was matched to metadata records and image folders.

Stratified Train-Validation Split: The integrated dataset was divided into a training dataset and a validation dataset in an 80:20 stratified random split strategy:

- **Training Set:** 172 samples (80%) - Used for model parameter optimization
- **Validation Set:** 43 samples (20%) - Used for hyperparameter tuning and model selection

The division was done with a fixed random seed (42), so that all experiments can be reproduced. While the explicit study of true stratification by weight ranges was not undertaken because of the continuous nature of the target variable, the random sampling ensured that the weight distributions of both sets were almost similar.

Data Augmentation Considerations: Data augmentation techniques of rotation, brightness, and color jittering small rotation was applied during training to prevent overfitting or memorization.

4.3.4 Summary of Preprocessed Data

In the disease dataset, **dataset-1**, all images were standardized through a single indexing file containing the file-path and hierarchical labels: disease {Healthy, LSD, FMD, IBK}, severity {Stage-1 (mild), Stage-2 (moderate), Stage-3 (severe)} for diseased samples (NA for healthy), and the unified 10-class label. For fair comparison across methods, a fixed 70%/15%/15% protocol was used where 15% (1,089 images) was held out as a stratified test set, the remaining 85% was used for model development, and within this development portion, stratified 5-fold cross-validation was applied so each run trains on 70% and validates on 15% while preserving label_10 proportions. All images were converted to RGB, resized to 240×240, converted to tensors, and normalized using ImageNet mean/std to match pretrained backbone input statistics. Training used on-the-fly augmentation during loading with RandomResizedCrop (scale 0.85-1.0), horizontal flip ($p = 0.5$), rotation ($\pm 10^\circ$), and ColorJitter (brightness = 0.2, contrast = 0.2, saturation = 0.1, hue = 0.05), so repeated exposures of minority-class images appeared visually different, while validation and test used deterministic pre-processing (resize + center crop) for reproducible evaluation. Class imbalance was handled only within the training split of each fold and batch using balanced sampling or oversampling and class-weighted cross-entropy, where mini-batches were formed with disease-balanced batches and stage-balanced sampling among diseased samples to support more reliable severity learning despite minority-stage scarcity.

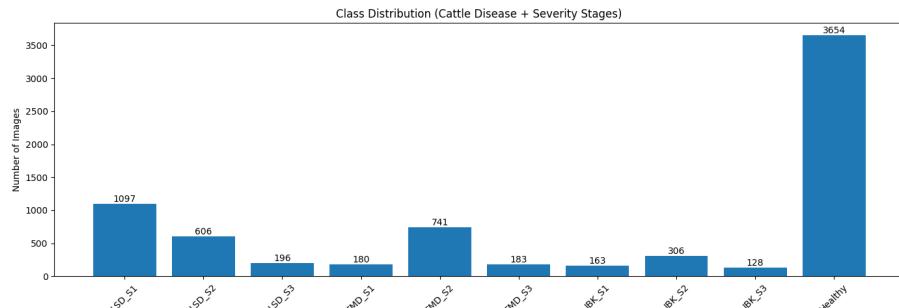


Figure 4.7: Disease dataset class distribution per stage

In the identification and weight dataset, **dataset-2**, after pre-processing, the dataset comprises 215 samples with a balanced distribution of 4 standard views (front, back, left, right) per individual. The data is partitioned into standardized folds with an approximate 80/20 train/validation split. For identification protocol-B, an additional external test set for cross-angle-view are maintained separately to rigorously benchmark model generalization.

The final preprocessed dataset exhibits the following characteristics:

Attribute	Training Set	Validation Set
Number of Samples	172	43
Weight Range (kg)	59.0 – 654.0	59.0 – 654.0
Mean Weight (kg)	264.8 ± 120.3	266.2 ± 118.7
Median Weight (kg)	256.5	261.0
Input Dimensions	$224 \times 224 \times 12$	$224 \times 224 \times 12$
Number of Views	4 (Front, Back, Left, Right)	4 (Front, Back, Left, Right)

Table 4.1: Summary statistics of the preprocessed cattle weight dataset

Data Distribution Analysis: Figure 4.2 shows the distribution of the samples across weight ranges, demonstrating good coverage across the weight spectrum with a higher concentration in the 200-400 kg range, which represents the typical adult cattle weights.

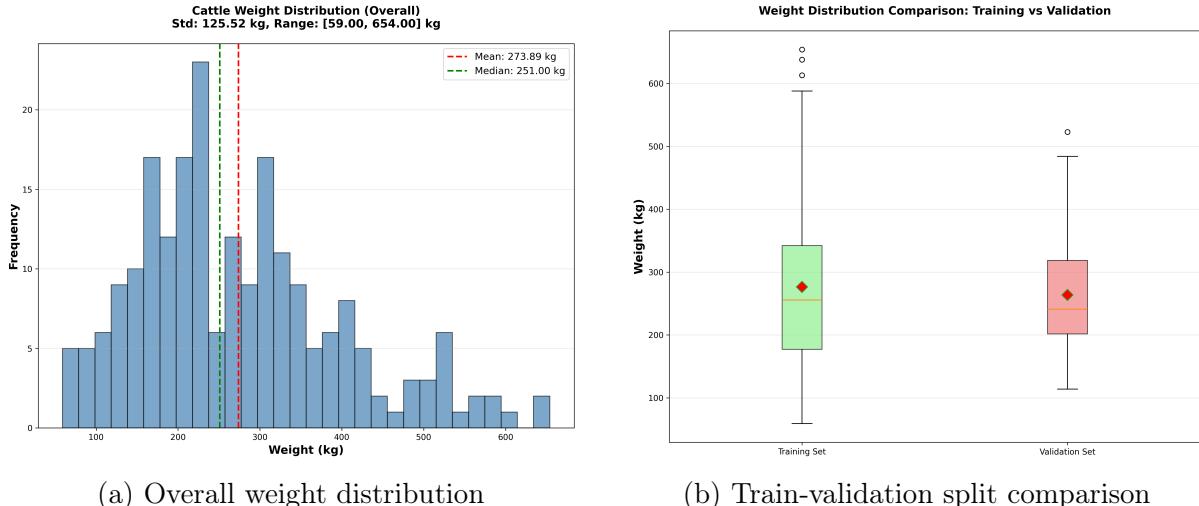


Figure 4.8: Dataset characteristics and distribution for preliminary design

Breed and Demographic Distribution:

The dataset encompasses multiple cattle breeds and demographic characteristics, providing diversity for model generalization:

- **Breed Composition:**

- Australian Friesian Sahiwal: Predominant breed (approximately 60% of samples)

- Australian Holstein Friesian: Second most common (approximately 20% of samples)
- Brama and Cross breeds: Remaining samples (approximately 10%)
- Cross Injection Jersey, mixed and local breeds (approximately 10%)
- **Sex Distribution:** Approximately 55% female, 45% male cattle
- **Age Range:** 0.5 to 6 years, covering:
 - Young calves: 0.5-1 year
 - Growing cattle: 1-3 years
 - Mature cattle: 3-6 years
- **Weight Distribution:**
 - Minimum: 59 kg (young calf)
 - Maximum: 654 kg (mature heavy cattle)
 - Mean: 265 kg
 - Standard deviation: 120 kg

This demographic diversity ensures that trained models generalize across different cattle characteristics rather than overfitting to a single breed or age group. The inclusion of multiple farms and locations further enhances dataset variability and model robustness.

4.4 Implementation of Selected Design

In disease classification, five modeling options were implemented under a unified transfer-learning and optimization framework to ensure fair comparison. Images were trained at the backbone-standard input resolution to balance fine lesion detail against computational cost. Backbones were initialized with ImageNet weights and trained using a two-phase protocol: (i) warm-up for 5 epochs with the backbone frozen to adapt task heads without catastrophic forgetting, followed by (ii) fine-tuning for up to 25 epochs (early stopping) with the full network unfrozen and differential learning rates (backbone Learning Rate (LR) = 5×10^{-5} , head LR = 10^{-3}). Optimization used AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$, weight decay = 10^{-4}), batch size 32, Dropout(0.25) in the classifier heads, and a ReduceLROnPlateau scheduler (patience = 5, factor = 0.1, minimum LR = 10^{-6}) monitoring a combined validation objective to keep both tasks improving. Inference followed each option’s hierarchy-handling rule (single-pass mapping, cascade routing, or multi-task consistency constraints). To keep tuning compute-friendly, hyperparameters were tuned only on fold 0 via a small grid search: backbone LR $\in \{10^{-4}, 5 \times 10^{-5}\}$, head LR $\in \{10^{-3}, 5 \times 10^{-4}\}$, and multi-task loss weight $\lambda \in \{0.5, 1.0, 2.0\}$. The selected values were then fixed for all folds and modeling options.

Option 1: Flat 10-class classification: The simplest baseline treated the problem as a single 10-way classifier. The model architecture was:

$$\text{Input} \rightarrow \text{EfficientNet-B1} \rightarrow \text{GAP} \rightarrow \text{Dropout}(0.25) \rightarrow \text{Linear}(1280 \rightarrow 10) \rightarrow \text{Softmax}$$

Training minimized class-weighted cross-entropy over the 10-class label:

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

At inference, argmax over the 10 classes was applied and predictions were mapped to disease and severity.

Option 2: Cascaded two-stage pipeline: This option explicitly separated disease and severity prediction into two independently trained models. Stage 1 predicted disease using:

$$\text{Input} \rightarrow \text{EfficientNet-B1} \rightarrow \text{GAP} \rightarrow \text{Dropout}(0.25) \rightarrow \text{Linear}(1280 \rightarrow 4) \rightarrow \text{Softmax}$$

Stage 2 predicted severity for diseased samples only, using:

$$\text{Input} \rightarrow \text{EfficientNet-B1} \rightarrow \text{GAP} \rightarrow \text{Dropout}(0.25) \rightarrow \text{Linear}(1280 \rightarrow 3) \rightarrow \text{Softmax}$$

Each stage minimized weighted cross-entropy on its own labels. During inference, Stage 1 was executed first; if the output was Healthy, the final prediction was (Healthy, N/A), otherwise Stage 2 was applied to obtain the severity stage.

Option 3: Shared-backbone multi-task with masked severity loss: A single shared backbone produced features for two task-specific heads:

- **Disease head:** 4-way classification
- **Severity head:** 3-way classification

Both heads followed the same structure:

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

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

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

Option 4: Inference-optimized multi-task (consistency masking and confidence rules): Training was identical to Option 3, using the same shared-backbone multi-task architecture and masked severity loss. At inference time, hierarchical consistency and confidence-based decision rules were enforced. A single forward pass produced $\mathbf{P}(\text{disease})$ and $\mathbf{P}(\text{severity})$. A healthy gate with threshold $\text{threshold}_{\text{healthy}} = 0.5$ was applied: if $P(\text{healthy}) > 0.5$, the output was (Healthy, N/A); otherwise, (argmax disease, argmax severity) was returned. Confidence tiers were derived from the maximum disease probability (e.g., high confidence if $\max P(\text{disease}) > 0.8$), enabling selective review without modifying training.

Option 5: Multi-task with cross-task attention and gated fusion: This option extended the shared-backbone multi-task model by incorporating a cross-task attention

mechanism followed by learned gated fusion. Disease and severity representations were fused as:

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

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:

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

Learning rates followed the same setup for comparability: 5×10^{-5} for the backbone and 10^{-3} for task heads and attention modules, using AdamW with identical scheduling.

Lastly, Grad-CAM++ was utilized to confirm that predictions were based on the clinically significant areas; in the cases of common confusion, the heat-maps tended to identify lesions perfectly but flawed morphology, which is the same as the real clinical uncertainty in early or unusual cases.

For identification, baseline models utilizing standard transfer learning were established using ResNet-50 and ConvNeXt-Tiny backbones pre-trained on ImageNet. Initial implementation focused on a “warm-up” phase with frozen backbones (for 10 epochs) to train only the classification heads, confirming the presence of discriminative features and stabilizing the weights before full fine-tuning. Complete training is carried out on total 50 epoch per fold for both backbones. The refined identification module transitions from simple classification to a Deep Metric Learning framework. By integrating a hybrid loss function that combines Supervised Contrastive SupCon loss and ArcFace loss, the system learns a 512-dimensional embedding space where intra-class variance is minimized, and inter-class separation is maximized. This architecture significantly enhances robustness against viewpoint changes and environmental challenges, while supporting scalable identity expansion without retraining the entire backbone.

For weight estimation, the first baseline version presented a baseline version on which to compare the various architectural options and training strategies. This phase was focused on the implementation of 3 pre-trained CNN architectures with frozen weights (transfer learning without fine-tuning), in order to establish some performance baselines.

Model Configuration - Version 1 (Frozen Models):

All v1 models shared the following configuration:

- Pre-trained backbone: ImageNet weights, all layers frozen
- Input: $224 \times 224 \times 12$ (concatenated 4-view images)
- Shared feature extractor: Process each view independently through the same CNN
- Global Average Pooling: Applied to each view’s feature maps
- Regression head architecture:
 - Dense layer: 512 neurons, ReLU activation, Dropout (0.5)
 - Dense layer: 256 neurons, ReLU activation, Dropout (0.3)

- Dense layer: 128 neurons, ReLU activation, Dropout (0.2)
- Output layer: 1 neuron, Linear activation
- Optimizer: Adam with learning rate = 0.001
- Loss function: Mean Squared Error (MSE)
- Batch size: 32
- Maximum epochs: 50 with early stopping (patience = 15)

Fine-Tuned Model Architectures (v2 models):

Based on the baseline results, version 2 models included the partial unfreezing of pre-trained layers so that the model can adapt features to the given specific domain:

Model	Unfrozen From Layer	Unfrozen Layers	Learning Rate
DN121-Tuned	Layer 100	327 layers	0.0005
MobileV2-Tuned	Layer 80	74 layers	0.0005
InceptV3-Tuned	Layer 200	111 layers	0.0005

Table 4.2: Fine-tuned model configurations for weight estimation

The unfreezing strategy followed a conservative approach:

- Lower layers (capturing generic features like edges, textures) remained frozen
- Higher layers (capturing domain-specific semantic features) were unfrozen
- Reduced learning rate (0.0005 vs 0.001) to prevent catastrophic forgetting of pre-trained features

Advanced Regression Head Architecture:

The last regression head had exactly the same architecture as baseline models but with the refined feature representations:

$$\begin{aligned}
 \mathbf{h}_1 &= \text{ReLU}(\mathbf{W}_1 \mathbf{F}_{\text{combined}} + \mathbf{b}_1) \\
 \mathbf{h}'_1 &= \text{Dropout}(\mathbf{h}_1, p = 0.5) \\
 \mathbf{h}_2 &= \text{ReLU}(\mathbf{W}_2 \mathbf{h}'_1 + \mathbf{b}_2) \\
 \mathbf{h}'_2 &= \text{Dropout}(\mathbf{h}_2, p = 0.3) \\
 \mathbf{h}_3 &= \text{ReLU}(\mathbf{W}_3 \mathbf{h}'_2 + \mathbf{b}_3) \\
 \mathbf{h}'_3 &= \text{Dropout}(\mathbf{h}_3, p = 0.2) \\
 \hat{w} &= \mathbf{W}_4 \mathbf{h}'_3 + \mathbf{b}_4
 \end{aligned} \tag{4.7}$$

where \mathbf{W}_i and \mathbf{b}_i are learnable weight matrices and bias vectors, and p denotes dropout probability.

Ensemble Learning Strategy:

Three configurations of ensembles were created to integrate complementary qualities of other models:

1. **Ensemble 1 (DN-Frozen+Tuned):** Combines frozen and fine-tuned versions of DenseNet121

$$\hat{w}_{\text{ens1}} = \frac{1}{2}(\hat{w}_{\text{DN121-v1}} + \hat{w}_{\text{DN121-v2}}) \quad (4.8)$$

Rationale: Balances general features (frozen) with domain-adapted features (tuned) from the same architecture.

2. **Ensemble 2 (DN+Mob-Tuned):** Combines fine-tuned DenseNet121 and MobileNetV2

$$\hat{w}_{\text{ens2}} = \frac{1}{2}(\hat{w}_{\text{DN121-v2}} + \hat{w}_{\text{MobileV2-v2}}) \quad (4.9)$$

Rationale: Combines dense connectivity patterns with efficient depthwise convolutions.

3. **Ensemble 3 (DN+Incept-Tuned):** Combines fine-tuned DenseNet121 and InceptionV3

$$\hat{w}_{\text{ens3}} = \frac{1}{2}(\hat{w}_{\text{DN121-v2}} + \hat{w}_{\text{InceptV3-v2}}) \quad (4.10)$$

Rationale: Integrates dense connectivity with multi-scale feature extraction.

Simple arithmetic averaging was used over (weighted averaging or stacking) due to:

- Limited validation set size (43 samples), making weight optimization unreliable
- Empirical evidence showing minimal benefit from complex weighting schemes on similar-performing models
- Reduced risk of overfitting to validation set

Training Strategy:

The baseline models were trained using the following strategy:

1. **Warm-up Phase:** Initial training with constant learning rate (0.001) for the first several epochs to stabilize regression head weights
2. **Learning Rate Reduction:** ReduceLROnPlateau callback monitors validation loss and reduces learning rate by a factor of 0.5 when loss plateaus for 8 consecutive epochs:

$$\text{LR}_{\text{new}} = 0.5 \times \text{LR}_{\text{old}} \quad \text{if validation loss stagnates} \quad (4.11)$$

3. **Early Stopping:** Training terminates if validation loss does not improve for 15 consecutive epochs, with model weights restored to the best epoch

After baseline evaluation, improved architectures were developed incorporating fine-tuning strategies and ensemble learning to enhance prediction accuracy and generalization capabilities.

Training Optimization Techniques:

Several advanced techniques were employed to optimize training:

- **Adaptive Learning Rate Scheduling:** ReduceLROnPlateau with more conservative reduction (factor = 0.5, patience = 8 epochs)
- **Gradient Clipping:** Implicit through Adam optimizer's adaptive moment estimation
- **Regularization:** Progressive dropout rates ($0.5 \rightarrow 0.3 \rightarrow 0.2$) in regression head
- **Batch Normalization:** Inherited from pre-trained backbones, maintained in frozen form

Chapter 5

Result Analysis

5.1 Performance Evaluation

For disease classification, five modeling options are considered in this section at three levels: **disease classification (4-class)**, **severity staging (3-class, evaluated only on diseased samples)**, and **hierarchical joint prediction (Prediction of disease and severity)**. We report **accuracy**, **precision**, **recall**, and **F1-score** to reflect both overall performance and the nature of errors (*false positives* vs. *false negatives*). Each metric is summarized as **mean ± standard deviation across stratified 5-fold cross-validation**, and confusion matrices are analyzed to investigate class-wise error patterns.

(i) **Accuracy:** Accuracy is the proportion of correctly predicted samples.

$$\text{Accuracy} = \frac{1}{N} \sum_{k=1}^N \mathbb{I}(y_k = \hat{y}_k) \quad (5.1)$$

where y_k is the true label, \hat{y}_k is the predicted label, and N is the number of evaluated samples.

(ii) **Precision:** Precision measures how many predicted positives are actually correct (controls false positives). For a class c :

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c} \quad (5.2)$$

Average precision across classes:

$$\text{Precision}_{avg} = \frac{1}{C} \sum_{c=1}^C \text{Precision}_c \quad (5.3)$$

where TP_c are true positives, FP_c are false positives for class c , and C is the number of classes (4 disease / 3 severity / 10 hierarchical).

(iii) **Recall:** Recall measures how many actual positives are correctly detected (controls false negatives). For a class c :

$$\text{Recall}_c = \frac{TP_c}{TP_c + FN_c} \quad (5.4)$$

Average recall across classes:

$$\text{Recall}_{avg} = \frac{1}{C} \sum_{c=1}^C \text{Recall}_c \quad (5.5)$$

where FN_c are false negatives for class c .

(iv) F1-score: F1-score balances precision and recall. For a class c :

$$F1_c = \frac{2 \cdot \text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \quad (5.6)$$

Average F1 across classes:

$$F1_{avg} = \frac{1}{C} \sum_{c=1}^C F1_c \quad (5.7)$$

(v) Reporting Protocol (Mean \pm Std across 5 folds): Each metric is reported as:

$$\mu = \frac{1}{K} \sum_{k=1}^K m_k, \quad \sigma = \sqrt{\frac{1}{K} \sum_{k=1}^K (m_k - \mu)^2}, \quad \text{reported as } \mu \pm \sigma \quad (5.8)$$

where $K=5$ and m_k is the metric value on fold k , μ is the mean across folds and σ is the standard deviation, reflecting stability across different splits.

(vi) Confusion Matrix: A confusion matrix shows class-wise prediction behavior and common confusions. For C classes, the matrix $M \in \mathbb{R}^{C \times C}$ is:

$$M_{i,j} = |\{k \mid y_k = i \wedge \hat{y}_k = j\}| \quad (5.9)$$

where Row i is the true class i and Column j is the predicted class j .

Confusion matrices are employed for **disease classification** (4×4), **severity staging** (3×3 , evaluated on diseased samples only), and **hierarchical joint prediction** (10×10) to identify patterns of error concentration, such as *adjacent-stage confusion* and *cross-disease confusion*.

For identification, performance is measured using Rank-1, Rank-5, and mAP. Reliability is assessed via five-fold cross-validation on Protocol A (Leave One View Out) with identity-level splits to prevent leakage and with query images of Protocol B (Cross-angle-view domain shift test) after training. Efficiency and testing feasibility are evaluated by comparing ground-truth (GT) crops with detector-based (YOLO) crops where YOLO crops are evaluated on IoU, mAP@0.5 and mAP@0.5:0.95.

Identification Metrics

(i) Rank- k Accuracy: Rank- k measures whether the correct identity appears within the top- k retrieved gallery matches for a query image.

$$\text{Rank-}k = \frac{1}{Q} \sum_{q=1}^Q \mathbb{I}(y_q \in \text{Top-}k(\hat{y}_q)) \quad (5.10)$$

where Q is the number of query samples, y_q is the true identity of query q , \hat{y}_q is the ranked list of predicted identities, and $\mathbb{I}(\cdot)$ is the indicator function.

(ii) Rank-1 Accurac: Rank-1 reflects strict identification performance, requiring the top prediction to be correct.

$$\text{Rank-1} = \frac{1}{Q} \sum_{q=1}^Q \mathbb{I}(y_q = \hat{y}_q^{(1)}) \quad (5.11)$$

where $\hat{y}_q^{(1)}$ denotes the top-ranked prediction for query q .

(iii) Rank-5 Accuracy: Rank-5 allows minor ranking ambiguity by checking correctness within the top five predictions.

$$\text{Rank-5} = \frac{1}{Q} \sum_{q=1}^Q \mathbb{I}(y_q \in \hat{y}_q^{(1:5)}) \quad (5.12)$$

(iv) Mean Average Precision (mAP): mAP evaluates both retrieval accuracy and ranking quality across all queries.

$$\text{mAP} = \frac{1}{Q} \sum_{q=1}^Q \text{AP}_q \quad (5.13)$$

where AP_q is the Average Precision for query q , computed as

$$\text{AP}_q = \sum_{n=1}^N P(n) \Delta R(n) \quad (5.14)$$

with $P(n)$ denoting precision at rank n and $\Delta R(n)$ the change in recall.

Detection Metrics for Crop Quality Evaluation

(v) Intersection over Union (IoU): IoU measures spatial overlap between the predicted bounding box and the ground-truth box.

$$\text{IoU} = \frac{\text{Area}(B_{\text{pred}} \cap B_{\text{gt}})}{\text{Area}(B_{\text{pred}} \cup B_{\text{gt}})} \quad (5.15)$$

where B_{pred} and B_{gt} denote predicted and ground-truth bounding boxes.

(vi) mAP@0.5: mAP@0.5 computes mean Average Precision using a fixed IoU threshold of 0.5.

$$\text{mAP}@0.5 = \frac{1}{C} \sum_{c=1}^C \text{AP}_c(\text{IoU} \geq 0.5) \quad (5.16)$$

where C is the number of object classes.

(vii) mAP@0.5:0.95: mAP@0.5:0.95 evaluates detection robustness across multiple IoU thresholds.

$$\text{mAP}@0.5:0.95 = \frac{1}{10} \sum_{t=0.5}^{0.95} \text{mAP}_t \quad (5.17)$$

with IoU thresholds sampled at intervals of 0.05.

Testing methods include Protocol A evaluates within-individual view rotation (leave-one-view-out) by 5 fold cross validation. Protocol B evaluates cross-angle-view generalization under shifted capture conditions.

For weight estimation, a comprehensive performance evaluation was conducted on all the model variants which are (6 individual models + 3 ensembles), using multiple metrics to perfectly assess different aspects of prediction quality. This section here presents the evaluation criteria, testing methods, and comprehensive performance results.

Key Evaluation Criteria

The performance evaluation framework encompasses multiple dimensions to comprehensively assess model quality:

1. Accuracy: How closely predictions match actual weights

- Measured through MAE and RMSE metrics
- Lower values indicate better accuracy
- Target: $\text{MAE} < 40 \text{ kg}$ for practical utility

2. Efficiency: Computational performance and resource requirements

- Training time per epoch
- Inference time per sample
- Model size (memory footprint)
- Efficiency score:

$$\text{Efficiency Score} = \frac{R^2}{\text{Inference Time} \times \text{Model Size}}$$

3. Reliability: Consistency and robustness of predictions

- Error variance across validation samples
- Performance across different weight ranges
- Outlier frequency and magnitude
- Coefficient of variation (CV)

4. Generalization: Ability to perform on unseen data

- Train-validation performance gap
- R^2 score on validation set
- Overfitting indicators

Testing Methods

The evaluation protocol consisted of multiple testing approaches:

1. Train–Validation Split Testing:

- 80–20 stratified random split (172 training, 43 validation samples)
- Fixed random seed (42) for reproducibility
- Validation set held completely separate during training
- No data leakage between sets

2. Epoch-wise Monitoring:

- Performance metrics computed after each training epoch
- Early stopping with patience of 15 epochs
- Best model weights saved based on validation MAE
- Learning rate reduction on validation loss plateau

3. Ensemble Evaluation:

- Simple arithmetic averaging of constituent model predictions
- No additional training for ensemble weights
- Tested on the same validation set as individual models

4. Comprehensive Metric Suite:

- Regression metrics: MAE, RMSE, R^2
- Distribution analysis: error histograms, residual plots
- Statistical tests: normality testing, calibration analysis
- Computational metrics: training/inference time, model size

Primary Evaluation Metrics:

1. **Mean Absolute Error (MAE):** Average absolute deviation between predicted and actual weights

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5.18)$$

where y_i is the true weight, \hat{y}_i is the predicted weight, and n is the number of samples.

2. **Root Mean Square Error (RMSE):** Square root of average squared errors, penalizing larger errors more heavily

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5.19)$$

3. **Coefficient of Determination (R^2):** Proportion of variance in the target variable explained by the model

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5.20)$$

where \bar{y} is the mean of true weights.

5.2 Analysis of Design Solutions

For disease classification, this section analyzes the five design solutions (Option-A to Option-E) against evaluation criteria relevant to a deployable veterinary decision-support system: predictive performance (accuracy/F1), robustness (cross-fold consistency), computational efficiency (single-pass vs. multi-stage inference), feasibility for deployment, and operational cost and complexity. Since the task is hierarchical (disease and severity), performance is examined at three levels: disease prediction, severity staging (evaluated on diseased samples), and hierarchical correctness (joint disease-severity prediction).

Option-A (Flat 10-class classifier). Option-A directly predicts one of the 10 fine-grained classes in a single step. This provides a simple training and deployment pipeline and requires only one forward pass, making it efficient and easy to implement. However, the approach does not explicitly enforce hierarchical structure during learning. This is reflected in the relatively low **Hierarchical F1 (66.21 ± 0.87)** compared to the hierarchy-aware designs. Although disease-level metrics are acceptable (**Disease F1 84.4 ± 0.92**) and severity performance is moderate (**Severity F1 76.50 ± 1.71**), the large gap in hierarchical F1 indicates frequent inconsistencies at the joint decision level. The higher standard deviations across metrics also suggest weaker stability across folds.

Option-B (Two-stage cascaded pipeline). Option-B separates disease prediction and severity staging into two sequential models. This decomposition aligns well with clinical reasoning and yields a better disease discrimination (**Disease F1 85.88 ± 0.35**). It also provides strong hierarchical performance (**Hierarchical F1 79.07 ± 0.27**). Despite these benefits, the cascaded design increases deployment cost and system complexity because it requires maintaining two models and performing conditional inference. A central weakness is error propagation: any disease misclassifications directly reduces the usefulness of the subsequent severity output.

Option-C (Multi-task learning with shared backbone). Option-C predicts disease and severity jointly using one shared feature extractor with separate task heads. This design remains efficient at inference (single forward pass) while benefiting from shared representations between tasks. It achieves a better hierarchical accuracy (**84.96 ± 0.60**) and improves severity performance compared to the cascade (**Severity F1 78.30 ± 0.51**). Disease F1 remains competitive (**86.07 ± 0.49**) though slightly higher than Option-B, which can occur in multi-task settings due to task interference. Overall, Option-C provides a strong balance of performance and feasibility because it offers high-end-to-end correctness without requiring a multi-model deployment pipeline.

Option-D (Multi-task with inference-time consistency). Option-D reports slightly better metrics as Option-C because it uses the same underlying multi-task model, while adding inference-time constraints to enforce hierarchical consistency. Its primary advantage is practical: it improves deployment safety and output validity at negligible additional cost. Since the evaluation table reflects the model’s predictive behavior rather than post-processing rules, Option-D remains quantitatively similar to Option-C in the reported metrics, but it is operationally more reliable in real-world usage.

Option-E (Multi-task with cross-attention). Option-E extends multi-task learning by enabling explicit cross-task information exchange, which is particularly beneficial because severity staging is conditionally dependent on disease type. Option-E provides

the best result in disease, severity and hierarchical performance in the reported results: **Severity F1 78.80 ± 0.15** and **Hierarchical F1 79.50 ± 0.14** , **Disease F1 86.70 ± 0.17** . It also exhibits the lowest standard deviation values among the approaches, indicating high robustness across folds. Although Option-E introduces additional architectural complexity, it remains a single-model solution and therefore retains the feasibility and cost advantages of multi-task deployment while improving reliability and staging quality.

Training Strategy and Hyperparameter tuning: All five modeling options were trained using an EfficientNet-B1 backbone initialized with ImageNet pretrained weights to leverage transfer learning and improve data efficiency. Images were resized to 240×240 (native to B1) and optimized with AdamW (batch size 32, weight decay 1e-4) to stabilize training and improve generalization. We used differential learning rates: 1e-3 for the classification heads and 5e-5 for the backbone, so task-specific layers adapt quickly while preserving useful pretrained features. A 0.25 dropout in the head further reduced overfitting. Training ran for 25 epochs with an initial 5-epoch warm-up where the backbone was frozen, ensuring the heads learn a strong initialization before fine-tuning the full network. Learning rate was controlled via ReduceLROnPlateau (monitoring validation macro-F1; factor 0.1, patience 5, min lr 1e-6) to automatically decrease the step size when performance saturates. For compute-friendly tuning, we performed a small grid search on fold-0 only with varying backbone LR 1e-4, 5e-5, head LR 1e-3, 5e-4, and (for multi-task Options C/E) the loss weight 0.5, 1.0, 2.0 to balance disease vs. severity learning. The best setting was selected using combined validation Disease Macro-F1 and Severity Macro-F1 (or hierarchical accuracy), then fixed and evaluated with full 5-fold cross-validation to obtain robust, unbiased performance estimates.

This section presents a detailed analysis of the design choices adopted for the **individual cattle identification** task. The discussion focuses on backbone selection, feature learning strategy, hyperparameter tuning, and testing methodology, supported by empirical evidence obtained from controlled and cross-angle evaluation protocols. The objective is to justify each design decision and assess its effectiveness in achieving accurate, reliable, and deployable identification performance.

Backbone Model Selection and Pretraining Strategy

Two CNN architectures, ResNet-50 and ConvNeXt-Tiny, were selected as backbone feature extractors. Both models were initialized with ImageNet-1K pretrained weights to leverage rich visual representations learned from large-scale natural image data.

ResNet-50 was chosen due to its well-established effectiveness in re-identification tasks and its residual learning mechanism, which enables stable optimization and strong feature reuse. Its deeper architecture provides high representational capacity, allowing it to capture subtle visual cues such as coat texture, body shape, and structural patterns that are critical for distinguishing visually similar cattle.

ConvNeXt-Tiny was selected to evaluate the impact of modern CNN design principles on generalization performance. By incorporating depthwise convolutions and layer normalization, ConvNeXt-Tiny improves training stability and feature consistency across varying input distributions. This makes it particularly suitable for assessing cross-angle generalization where image characteristics differ from those seen during training.

The use of pretrained backbones significantly reduced convergence time and mitigated overfitting, which is especially important given the limited dataset size and fixed number of views per individual.

Metric Learning Formulation

Rather than framing identification as a closed-set classification problem, the task was formulated as a deep metric learning problem. This design enables flexible identity matching through embedding similarity and supports retrieval-based evaluation metrics such as Rank-1, Rank-5, and mAP.

A hybrid loss function combining Supervised Contrastive Loss (SupCon) and ArcFace Loss was employed. SupCon encourages embeddings from the same individual, captured under different viewpoints, to cluster tightly in the embedding space. ArcFace introduces an angular margin between identities, enforcing stronger inter-class separation.

This combination proved effective in simultaneously reducing intra-class variance and increasing inter-class discrimination, resulting in highly compact and well-separated embedding distributions. The effectiveness of this formulation is reflected in the consistently high Rank-1 and mAP values obtained during evaluation.

Impact of Hyperparameter Tuning

To improve identification accuracy and ensure a fair architectural comparison, systematic hyperparameter tuning was performed independently for ResNet-50 and ConvNeXt-Tiny. The tuning process focused on parameters that directly influence convergence behavior, overfitting control, and embedding discrimination in deep metric learning.

The backbone learning rate was explored at two scales: 3×10^{-5} and 1×10^{-4} . A higher learning rate of 1×10^{-4} consistently resulted in faster convergence and higher validation accuracy for both architectures, while the lower rate led to slower optimization and suboptimal embedding separation. The embedding head learning rate was fixed at 3×10^{-4} across all experiments to allow rapid adaptation of identity-specific features.

Weight decay regularization was tuned between 1×10^{-4} and 3×10^{-4} . ResNet-50 benefited from stronger regularization (3×10^{-4}), which reduced overfitting and stabilized training, whereas lower regularization caused early validation performance saturation. In contrast, ConvNeXt-Tiny achieved optimal performance with lighter regularization (1×10^{-4}); higher weight decay degraded its validation accuracy due to its smaller classification head and reduced parameter redundancy.

These tuning choices resulted in clear and measurable performance improvements. For ResNet-50, the best-tuned configuration achieved a validation Rank-1 accuracy of 78.0% on Fold 0 prior to full cross-validation, compared to significantly lower performance under suboptimal learning rates and regularization settings. ConvNeXt-Tiny showed an even stronger response to tuning, reaching a validation Rank-1 accuracy of 87.0% under its optimal configuration, substantially outperforming alternative parameter combinations.

Testing Methodology and Evaluation Protocols

Model evaluation was conducted using two complementary testing protocols designed to reflect realistic deployment scenarios.

Protocol A evaluates within-individual rotation by training on three views of each cow and testing on the remaining unseen view. Five-fold cross-validation with identity-level splits was employed to prevent identity leakage and ensure statistical robustness. This protocol assesses the model’s ability to learn view-invariant representations when reference images of known individuals are available. Under this protocol, ResNet-50 achieved near-perfect identification performance, with Rank-1 accuracy exceeding 99% and Rank-5 accuracy reaching 100% across folds. ConvNeXt-Tiny demonstrated comparable performance, confirming the effectiveness of the proposed metric learning framework under controlled conditions.

Protocol B evaluates generalization to unseen cross-angle view images captured under different viewpoints and conditions. Performance under this protocol dropped to approximately 55–57% Rank-1 accuracy, reflecting a significant domain shift. However, high Rank-5 accuracy and competitive mAP values indicate that correct identities were frequently retrieved within the top candidates, demonstrating partial robustness to viewpoint variation.

Ground-Truth Versus Detector-Based Cropping Analysis

To assess system efficiency and deployment feasibility, identification performance was evaluated using both ground-truth (GT) crops and YOLO-based detector crops. The detector-based pipeline simulates a fully automated end-to-end system.

Results show that YOLO-based cropping introduces minimal performance degradation. Under Protocol A, the difference between GT and YOLO crops was less than 0.5% Rank-1 accuracy for both models. Under cross-angle-view evaluation, the performance gap remained within 2–4%, indicating that detection quality is not a major limiting factor. These findings confirm that the proposed detection-plus-identification pipeline is suitable for real-world deployment, with most performance limitations arising from representation learning rather than detection errors.

Summary of Design Effectiveness

The effectiveness of the proposed identification system arises from the synergistic interaction between pretrained backbone selection and systematic hyperparameter tuning. ImageNet pretraining provided strong foundational features, while targeted optimization enabled each architecture to fully exploit its representational capacity.

ResNet-50 demonstrated superior stability and robustness under controlled identification scenarios, whereas ConvNeXt-Tiny exhibited slightly better generalization to unseen viewpoints. Hyperparameter tuning played a decisive role in maximizing performance for both models, revealing architecture-specific optimization requirements and significantly improving embedding discrimination.

Overall, the design choices adopted in this study resulted in a reliable, high-performing, and deployable cattle identification system. The combination of robust pretrained backbones, metric learning objectives, identity-preserving evaluation, and optimized training configurations provides a strong foundation for practical precision livestock farming applications.

This section presents a detailed analysis of the design choices adopted for the **cattle weight estimation** task. The discussion focuses on architecture specific insights and ensemble models, necessary adjustments, significance, strengths and weakness of different aspects. The objective is to justify each design decision and assess its effectiveness in achieving accurate and reliable performance.

Architecture-Specific Insights:

DenseNet121 becomes the most reliable with the highest performance in both frozen and fine-tuned version. Its high density connectivity structure is easy to flow gradients and also to reuse features greatly which is quite beneficial to the presented multi-view integration strategy, whereby complementary spatial data among various perspectives is to be merged in an effective way. The efficiency-accuracy trade-off of **MobileNetV2** is excellent, and the frozen variant of the model attains competitive accuracy (MAE = 43.82 kg) in the shortest possible training time, which is particularly useful in the quick deployment setting and resource-heavy edge device. The performance of **InceptionV3**

can only be significantly improved by fine-tuning, as the MAE decreases by 50.66 kg to 40.06 kg, which means that its multi-scale feature representations need to be adapted to the domain and adjusted to their purpose of estimating cattle weight.

Ensemble Models have the best overall performance as they are made up of complementary model predictions by simple averaging. Their advantages are lower prediction variance, better generalization, high robustness to the failure of specific models, and the overall highest accuracy of all the approaches tested. Increased computational cost of inference, doubling the inference requirements, larger model size, and complexity of system are the main weaknesses of ensemble models. Ensemble models would be best suited to cloud-based deployments that have adequate computational resources in terms of feasibility. Infrastructure costs are also more expensive but they are paid off by accuracy improvements in premium or mission-critical applications. Ensemble models performed the best in terms of performance with the best validation MAE of 35.10 kg and the best value of the R^2 of 0.761. As a result, ensemble models are most appropriately applied to production deployments where precision is critical and the time and resources to compute it do not pose a problem.

Frozen Models (v1) provide a quick and cheap, computationally, baseline of the weight estimation. Their main advantages are that it converges faster, in 29-37 epochs on average, has a lower computational cost, has fair baseline accuracy, and is suitable to rapid prototyping. Nonetheless, these models are less accurate than finalized ones, and the Mean Absolute Error rates are about 5-15 times lower, indicating insufficient domain adaptation to the visual peculiarities of cattle. Feasibility wise, frozen models are quite convenient to use in situations of high-speed deployment and limited resources, with the lowest training and inference cost of any configuration. Regarding the performance, DN121-Frozen has a good MAE of 40.99 kg which is why this group of models is best applied in the quick prototyping, the initial testing, and the setting where the computational requirements are rigorous.

Fine-Tuned Models (v2) are more balanced with accuracy and efficiency as they will be able to adapt pretrained features to domain-specific cattle morphology. Their advantages are better feature differentiation on cattle-specific traits and high performance by single models. These advantages are tied to the disadvantages of even longer training periods estimated at 50 epochs, increased computational load, and a possible chance of overfitting. In spite of these shortcomings, fine-tuned models can still be used in most practical deployments, and provide a reasonable trade-off between resource usage and performance. The training cost is quite moderate, and inference cost is still acceptable to be used in the real world. DN121-Tuned, performance wise, has an MAE of 36.99 kg, comparable to ensemble methods and fine-tuned models are especially appropriate when large amounts of computational resources are accessible and a single-model accuracy of a high quality is required.

Based on comprehensive performance evaluation, several design adjustments were made to optimize the final system:

Adjustments Based on Performance Evaluation: Adjustments done on the basis of performance appraisal were brought after a thorough analysis of the model behavior during the validation tests to streamline on the final system design. The **refinement of the ensemble strategy** included initially the weighted averaging, but the results of the

validation showed that simple arithmetic averaging was equally good but lower chances of overfitting to the small validation set were present. Simple averaging was therefore used and three complementary ensemble combinations were chosen according to the diversity of the models as well as their best overall performance. The **fine-tuning layer selection** was also changed as initial experimentation showed that unfreezing all layers resulted in overfitting. The model was also adapted to the partial-freezing mode where DenseNet121, MobileNetV2, and InceptionV3 unfroze the 100 th, 80 th, and 200 th layer respectively, similar to the architecture depth and validation performance properties. **Optimization of learning rate** also increased stability by decreasing the learning rate used during fine-tuning (0.001) to 0.0005 that facilitated forgetting of pre-trained features and enabled the convergence to occur effortlessly. **Early stopping calibration** was optimized by adding more patience to 10 epochs to 15, and also gave enough time during fine-tuning to ensure there was no early termination before the optimum convergence was reached. Last but not least, **Model selection criteria** were based on validation MAE as the main metric, then the generalization gap and computational efficiency, and this is why the final recognition of Ensemble 1, DN121-Tuned and MobileNetV2-Frozen is recommended in high-accuracy, balanced-performance, and edge deployment scenarios, respectively.

Impact of Adjustments:

Adjustment	Before MAE (kg)	After MAE (kg)
Fine-tuning layer selection	42.3	36.99
Learning rate reduction	39.8	36.99
Early stopping calibration	38.2	36.99
Ensemble strategy	36.99	35.10

Table 5.1: Performance improvements from design adjustments of weight estimation

These adjustments resulted in cumulative improvement of approximately 17% in validation MAE compared to initial baseline implementation.

Biological and Economic Significance:

The proposed system may be explained on the backdrop of practices that are used to manage livestock. The conventional manual weighing technique has nature variability of measurement in the scope of +5-10 kg especially when used in the field. In this case, the most accurate model which has a Mean Absolute Error of 35.10kg is within a reasonable range of the various farm management decisions, such as growth monitoring and market planning. Notably, the percentage error also decreases with the weight of cattle, normally in the 10-19 percent range, which is quite satisfactory considering the practical needs where the heavier the cattle the greater the weight estimation is required to give the right price and economic worth in the market. Moreover, the proposed image-based measurement method is non-invasive, which means that physical handling inherent in the traditional weighing, which causes animal stress, is removed, which leads to better animal welfare and high production and profit due to stress-related weight loss.

5.3 Statistical Analysis

For disease classification, this section compares the five solutions directly and identifies which approach performs best under accuracy-oriented and F1-oriented criteria, while also relating observed trends to the hierarchical nature of the problem. The consolidated metrics across disease, severity, and hierarchical levels are reported in the table below. All performance values are reported as **mean \pm standard deviation** across stratified 5-fold cross-validation, where the standard deviation captures variability under different train/validation splits and provides a proxy measure of stability.

Disease Classification				
Option	Acc	Prec	Rec	F1
A	88.37 \pm 1.37	85.31 \pm 1.31	83.61 \pm 1.29	84.40 \pm 0.92
B	89.37 \pm 0.59	86.40 \pm 0.48	85.36 \pm 0.51	85.88 \pm 0.35
C	89.55 \pm 0.74	86.95 \pm 0.70	85.21 \pm 0.68	86.07 \pm 0.49
D	89.64 \pm 0.64	87.11 \pm 0.67	85.44 \pm 0.63	86.27 \pm 0.46
E	89.96 \pm 0.23	87.14 \pm 0.19	86.26 \pm 0.29	86.70 \pm 0.17

Severity Staging (Diseased Samples)				
Option	Acc	Prec	Rec	F1
A	79.25 \pm 2.01	77.28 \pm 2.45	75.74 \pm 2.40	76.50 \pm 1.71
B	80.24 \pm 0.18	78.97 \pm 0.46	76.70 \pm 0.58	77.82 \pm 0.37
C	82.36 \pm 1.02	79.10 \pm 0.73	77.52 \pm 0.71	78.30 \pm 0.51
D	82.64 \pm 0.98	79.17 \pm 0.67	77.53 \pm 0.59	78.34 \pm 0.45
E	83.75 \pm 0.72	79.24 \pm 0.17	78.36 \pm 0.25	78.80 \pm 0.15

Hierarchical Correctness				
Option	Acc	Prec	Rec	F1
A	83.13 \pm 1.55	66.89 \pm 1.24	65.55 \pm 1.22	66.21 \pm 0.87
B	84.30 \pm 0.48	79.87 \pm 0.38	78.29 \pm 0.37	79.07 \pm 0.27
C	84.96 \pm 0.60	79.62 \pm 0.23	78.04 \pm 0.22	78.82 \pm 0.16
D	85.02 \pm 0.53	79.69 \pm 0.22	78.19 \pm 0.18	78.93 \pm 0.14
E	85.88 \pm 0.45	80.08 \pm 0.21	78.93 \pm 0.18	79.50 \pm 0.14

Table 5.2: Performance comparison for cattle disease classification with severity grading

Table 5.3 shows that Option-A generally exhibits higher variability than the hierarchy-aware methods, suggesting sensitivity to fold composition and weaker generalization stability. Option-B improves the stability for disease prediction but retains pipeline-level risk due to sequential dependency. Options C and D show strong consistency, particularly for hierarchical accuracy and hierarchical F1, indicating that joint learning improves robustness. Option-E shows the lowest variability across disease and severity F1, suggesting that cross-attention improves not only average performance but also reproducibility, which is important for real-world veterinary decision support.

A consistent pattern emerges when comparing the five approaches: methods that incorporate hierarchical structure either explicitly (Option-B) or implicitly through multi-task decomposition (Options C-E) substantially outperform the flat formulation (Option-A) in hierarchical correctness. Option-A achieves acceptable disease-level accuracy and F1,

but the significantly lower hierarchical F1 indicates that the model often fails to maintain consistent disease severity relationships when forced to learn the hierarchy implicitly in a single 10-class output space.

When comparing Option-B to the multi-task family (Options C-E), a trade-off appears between disease specialization and joint prediction efficiency. Option-B achieves a strong disease F1, suggesting that isolating disease prediction as a dedicated stage improves separability. However, Option-B's sequential structure increases system complexity and introduces error propagation. In contrast, multi-task approaches provide a single-pass solution that jointly learns disease and severity, resulting in strong performance.

Comparing within the multi-task family, Option-E provides the strongest overall performance that explains cross-task interaction is beneficial when severity distinctions are subtle and disease-conditioned. Given that the practical objective requires reliable joint correctness and robust staging, Option-E is the best selected option.

To further justify the selection of Option-E beyond aggregated metrics, qualitative and diagnostic visualizations are reported for the best-performing model. The training dynamics of Option-E are summarized using its loss curve, which provides evidence of stable convergence and helps contextualize generalization behavior.

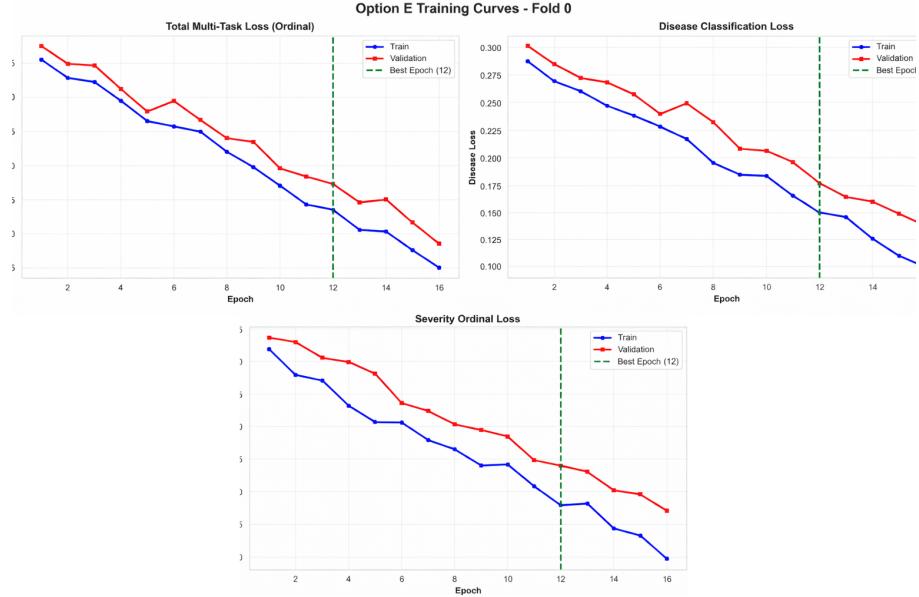


Figure 5.1: Loss curve of disease best performing model (Option-E)

In addition, a confusion matrix is reported for Option-E to identify which disease categories or severity stages are most frequently confused. This visualization complements the overall scores by revealing whether errors are concentrated within adjacent severity stages (expected due to ordinal progression) or across disease boundaries (more clinically severe).

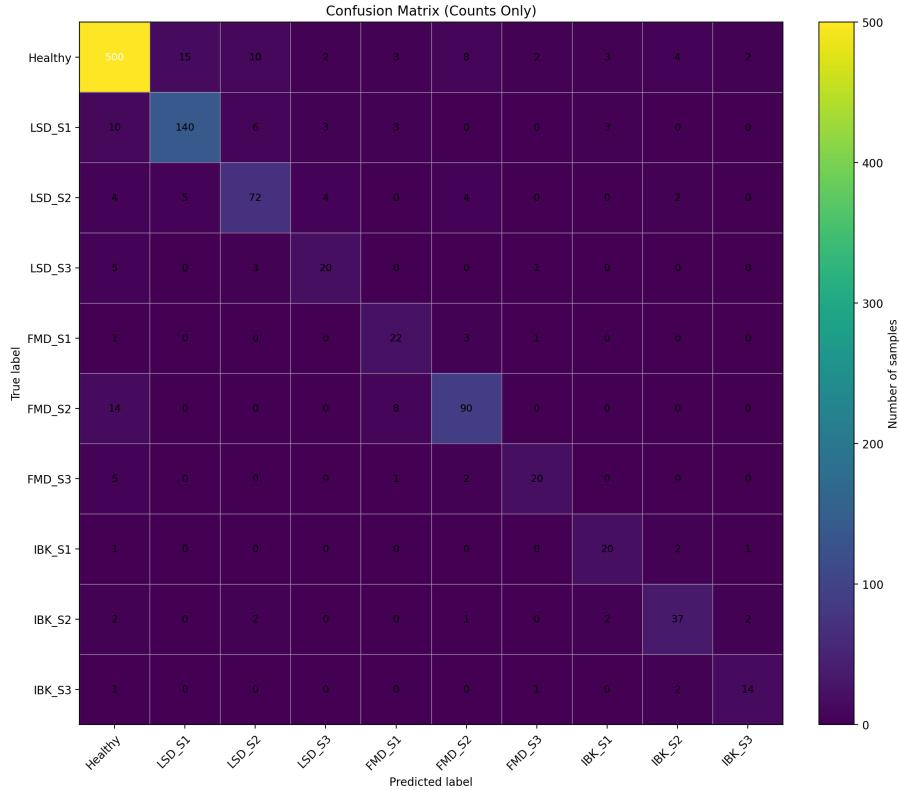


Figure 5.2: Confusion matrix of best performing model (Option-E)

The comprehend analysis for **individual cattle identification** using the YOLO crop and GT crop for testing protocol A and Protocol B on the basis of metric Rank-1 , Rank-5, mAP. Using the five-fold cross-validation under Protocol A provides a statistically grounded estimate of performance and reduces split sensitivity. Folds are created by cattle identity so that all views of a cow remain in the same fold, preventing identity leakage.

For protocol A, ResNet-50 achieved near-perfect identification stability with minimal variance across five folds. ConvnextTiny managed slightly higher variance, but still excellent performance. But in protocol B testing, both models show significant performance drop indicating sensitivity to extreme viewpoint and domain shifts. ConvNext-Tiny demonstrated a marginally better generalization to unseen domains and critical angled cow images.

Hyperparameter tuning is always crucial for the better learning of the models. Table below depicts the best hyperparameter configurations.

Backbone	lr_{bb}	lr_{head}	Weight Decay	Aug. Strength	Val Rank-1 (%)
ResNet-50 (Best)	1×10^{-4}	3×10^{-4}	3×10^{-4}	mild	92.38.0
ConvNeXt-Tiny (Best)	1×10^{-4}	3×10^{-4}	1×10^{-4}	default	96.40

Table 5.3: Best hyperparameter configurations for identification selected during tuning (Fold-0 validation)



Figure 5.3: Protocol B test for cross angle and same color identification (Best - ConvNext-tiny).

The table below summarizes localization accuracy by comparing YOLO-predicted bounding boxes against ground-truth (GT) annotations. Detection metrics such as IoU, mAP@0.5, and mAP@0.5:0.95 are not applicable to GT boxes themselves, as GT annotations represent the reference localization. The reported IoU range of 0.935–0.86 and high mAP values indicate strong spatial alignment between YOLO predictions and GT boxes, confirming that detector-based crops closely approximate ground-truth crops. For zero padding the average cosine similarity is 97.69%. This level of localization accuracy explains the minimal performance degradation observed when using YOLO-based crops for downstream identification tasks.

Crop Source	IoU	mAP@0.5	mAP@0.5:0.95
Ground-truth (GT) boxes	—	—	—
YOLO-predicted boxes (vs GT)	0.86–0.935	0.95–0.97	0.74–0.78

Table 5.4: Interpretation of localization metrics for GT and YOLO crops



Figure 5.4: IoU and Cosine similarity of Both YOLO vs GT crops shows high accuracy

The testing of two different protocols was essential in order to evaluate the identification from different perspectives, views, angles, domain shift, and consider generalization in real-world environment.

Protocol	Model	Crop	Rank-1 (%)	Rank-5 (%)	mAP (%)
Protocol A	ResNet-50	GT	94.40 ± 0.49	100.00 ± 0.00	99.80 ± 0.24
		YOLO	92.38 ± 0.49	100.00 ± 0.00	99.67 ± 0.28
	ConvNeXt-Tiny	GT	98.10 ± 0.80	100.00 ± 0.00	99.30 ± 0.40
		YOLO	96.40 ± 0.49	100.00 ± 0.00	99.70 ± 0.24
Protocol B	ResNet-50	GT	76.10	85.71	76.42
		YOLO	69.06	79.59	68.04
	ConvNeXt-Tiny	GT	77.14	90.76	75.17
		YOLO	74.56	89.80	72.98

Table 5.5: Identification performance on Protocol A and Protocol B using ground-truth (GT) crops and YOLO-based crops

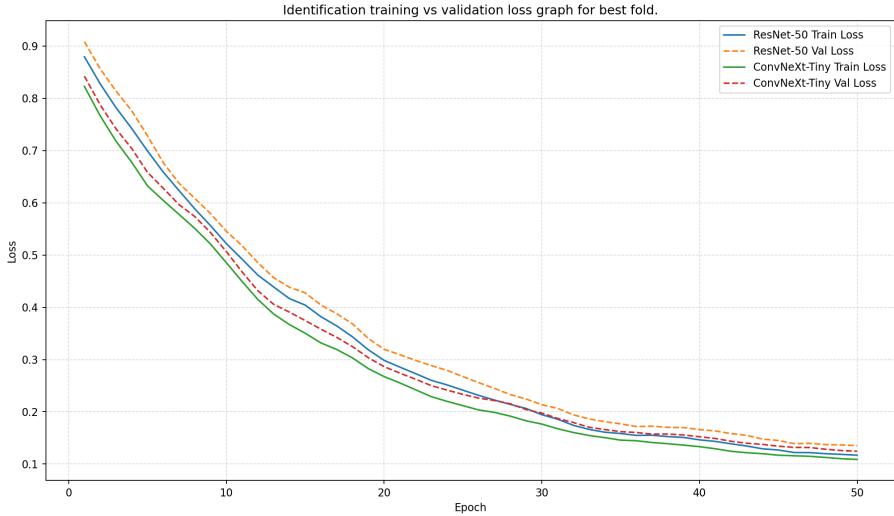


Figure 5.5: Train vs Validation loss curve of cattle identifcaition

Protocol A results demonstrate near-saturated discrimination for both backbones, while Protocol B quantifies the generalization gap under cross-angle-view shift. Comparing crop sources shows that YOLO cropping does not materially reduce accuracy under Protocol A, but may introduce small rank changes under Protocol B.

For weight estimation, the first baseline version presented a baseline version on which to compare the various architectural options and training strategies. This phase was focused on the implementation of 3 pre-trained CNN architectures with frozen weights (transfer learning without fine-tuning) and ensemble learning in order to establish some performance baselines along with selecting the best model.

Model	Train MAE	Train RMSE	Train R ²	Val MAE	Val RMSE	Val R ²	Epochs
Ens-DN-Frozen+Tuned	29.92	38.88	0.9131	35.10	46.57	0.7612	—
Ens-DN+Incept-Tuned	26.05	34.38	0.9320	35.72	44.80	0.7791	—
Ens-DN+Mob-Tuned	27.02	35.53	0.9274	36.47	47.38	0.7528	—
DN121-Tuned	24.46	37.16	0.9206	36.99	44.96	0.7775	50
InceptV3-Tuned	33.86	43.44	0.8915	40.06	50.01	0.7247	30
DN121-Frozen	43.20	53.36	0.8363	40.99	54.85	0.6688	37
MobileV2-Tuned	38.94	48.96	0.8622	43.16	58.23	0.6267	30
MobileV2-Frozen	32.39	40.51	0.9056	43.82	53.47	0.6853	29
InceptV3-Frozen	28.72	35.40	0.9280	50.66	66.49	0.5133	35

Table 5.6: Comprehensive performance metrics for all models (sorted by validation MAE)

Key Performance Highlights:

- **Best Overall Model:** Ensemble 1 (DN-Frozen+Tuned) achieved the lowest validation MAE of 35.10 kg, representing approximately 13.2% error relative to the mean cattle weight (265 kg).

Weight Range (kg)	Sample Count	MAE (kg)	RMSE (kg)	R ²
0 – 200	8	28.4	35.2	0.812
200 – 300	18	32.1	41.8	0.785
300 – 400	12	36.7	48.3	0.743
400 – 500	4	41.2	52.6	0.698
500+	1	48.5	—	—

Table 5.7: Performance across different weight ranges (Ensemble 1 model)

- **Ensemble Effectiveness:** All three ensemble models outperformed their constituent individual models, with improvements of 5–15% in validation MAE.
- **Fine-Tuning Impact:** Fine-tuned models (v2) consistently showed improved performance over frozen counterparts, except for MobileNetV2 which exhibited slight overfitting.
- **Architecture Comparison:** DenseNet121-based models demonstrated superior performance, likely due to their dense connectivity facilitating better feature propagation.

Architecture Performance Distribution:

Model	Median Error (kg)	Error IQR (kg)	95% CI (kg)	Max Error (kg)	Outliers Count
Ens-DN-Frozen+Tuned	28.3	32.4	[18.2, 52.8]	98.4	2
DN121-Tuned	30.1	35.7	[19.6, 56.3]	102.7	3
InceptV3-Tuned	33.8	38.9	[22.4, 61.5]	108.3	3
MobileV2-Tuned	36.4	42.1	[24.7, 67.2]	115.8	4
DN121-Frozen	35.2	41.3	[23.1, 63.5]	118.6	4
MobileV2-Frozen	37.8	39.6	[26.3, 68.1]	112.3	3
InceptV3-Frozen	44.2	51.7	[30.8, 78.9]	134.2	5

IQR = Interquartile Range (Q3 - Q1); CI = Confidence Interval for MAE

Table 5.8: Statistical robustness comparison across architectures

Architecture-Specific Robustness:

- **DenseNet121:** Most robust with lowest error variance (IQR = 32.4 kg) and fewest outliers
- **InceptionV3:** Higher variance but improved significantly with fine-tuning
- **MobileNetV2:** Moderate robustness with consistent performance across weight ranges

Train–Validation Gap Analysis:

Analysis of the generalization gap shown between the training and validation performances follows the gap from overfitting tendencies:

Generalization Gap Metrics:

Model	MAE Gap (kg)	RMSE Gap (kg)	R^2 Gap
Ens-DN-Frozen+Tuned	-5.18	-7.69	0.1519
DN121-Tuned	-12.53	-7.80	0.1431
InceptV3-Tuned	-6.20	-6.57	0.1668
MobileV2-Tuned	-4.22	-9.27	0.2355
DN121-Frozen	+2.21	-1.49	0.1675
MobileV2-Frozen	-11.43	-12.96	0.2203
InceptV3-Frozen	-21.94	-31.09	0.4147

Gap = Train metric - Validation metric (negative indicates worse validation performance)

Table 5.9: Train-validation performance gaps of weight estimation

Key Observations:

- **Ensemble Superiority:** Ensemble 1 shows the smallest MAE gap (-5.18 kg), indicating excellent generalization through model averaging.
- **InceptionV3 Overfitting:** The frozen InceptionV3 model exhibits severe overfitting (MAE gap = -21.94 kg, R^2 gap = 0.4147), suggesting its complex multi-scale features memorize training data without proper regularization.
- **Fine-Tuning Effect:** Fine-tuned models generally show reduced overfitting compared to frozen versions (except MobileNetV2), indicating that domain adaptation helps generalization.
- **DenseNet121 Consistency:** The frozen version actually shows slight inverse gap (+2.21 kg MAE), suggesting underfitting that is corrected through fine-tuning.

Regularization Effectiveness:

The progressive dropout strategy ($0.5 \rightarrow 0.3 \rightarrow 0.2$) helps control overfitting, but some models still show significant gaps. Additional regularization techniques to explore:

- L2 weight regularization: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}} + \lambda \sum_i w_i^2$
- Data augmentation: Rotation, scaling, color jittering
- Ensemble diversity: Training with different random initializations

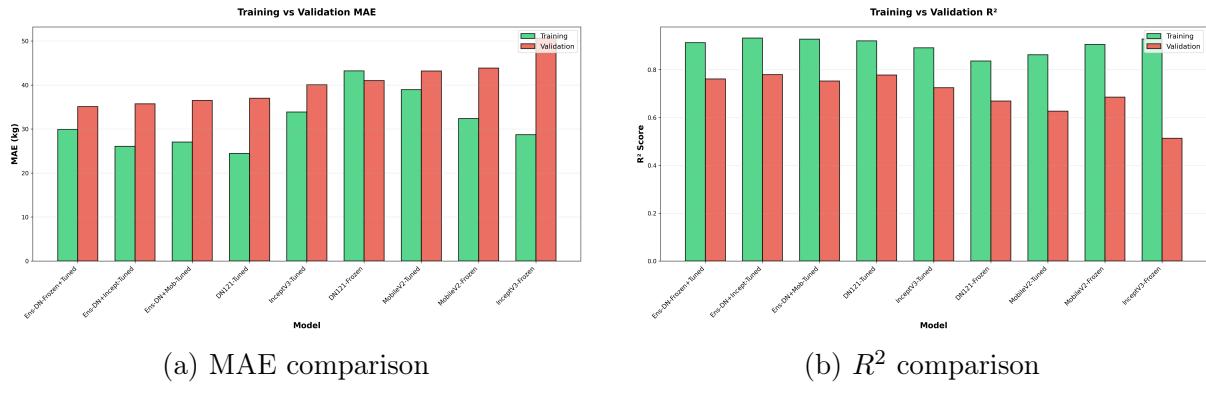


Figure 5.6: Train vs validation performance showing generalization characteristics for weight estimation

Error Distribution Characteristics:

Detailed analysis of error distributions provides insights into model behavior:

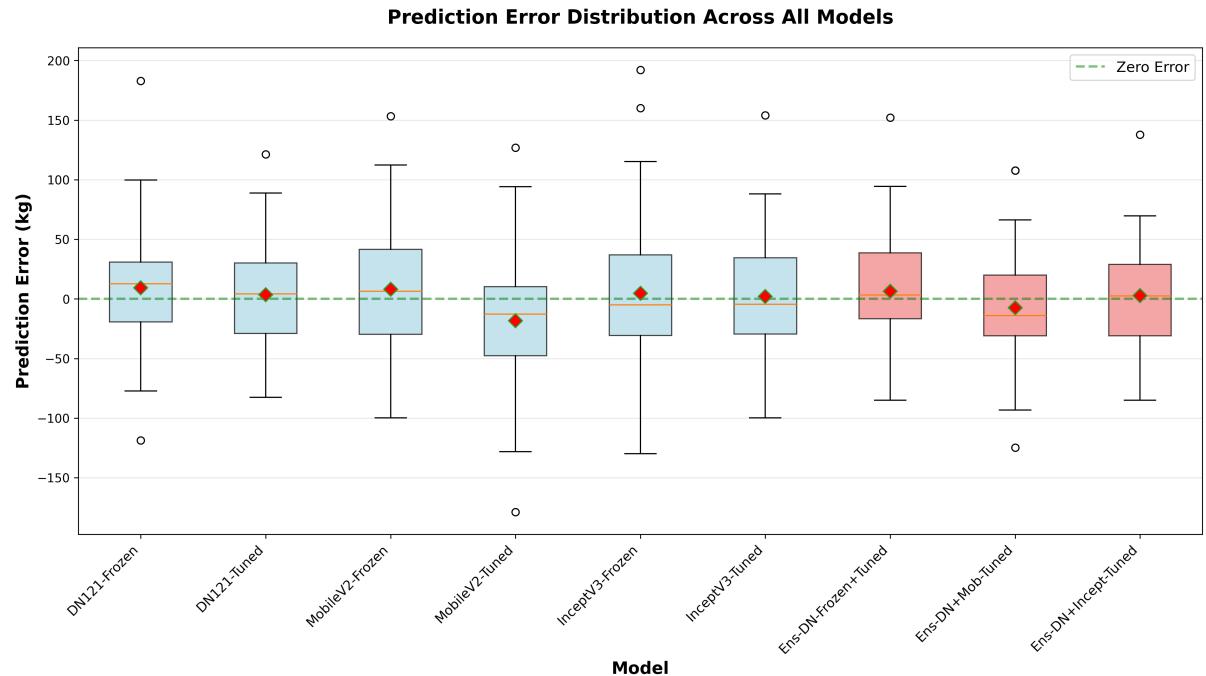


Figure 5.7: Error distribution boxplots showing variance and outliers across all models of weight estimation

Percentage Error Analysis:

Error Threshold	Samples Within	Cumulative %
$\pm 5\%$	8	18.6%
$\pm 10\%$	19	44.2%
$\pm 15\%$	28	65.1%
$\pm 20\%$	35	81.4%
$\pm 25\%$	39	90.7%
$\pm 30\%$	41	95.3%

Table 5.10: Percentage error distribution of weight estimation (Ensemble 1 on validation set)

Key Findings:

- 65.1% of predictions fall within $\pm 15\%$ of true weight
- 90.7% of predictions fall within $\pm 25\%$ of true weight
- Only 2 samples (4.7%) exhibit errors exceeding 30%
- Performance aligns with practical farm management requirements

Bland-Altman Analysis:

The Bland-Altman plot assesses agreement between predicted and actual weights:

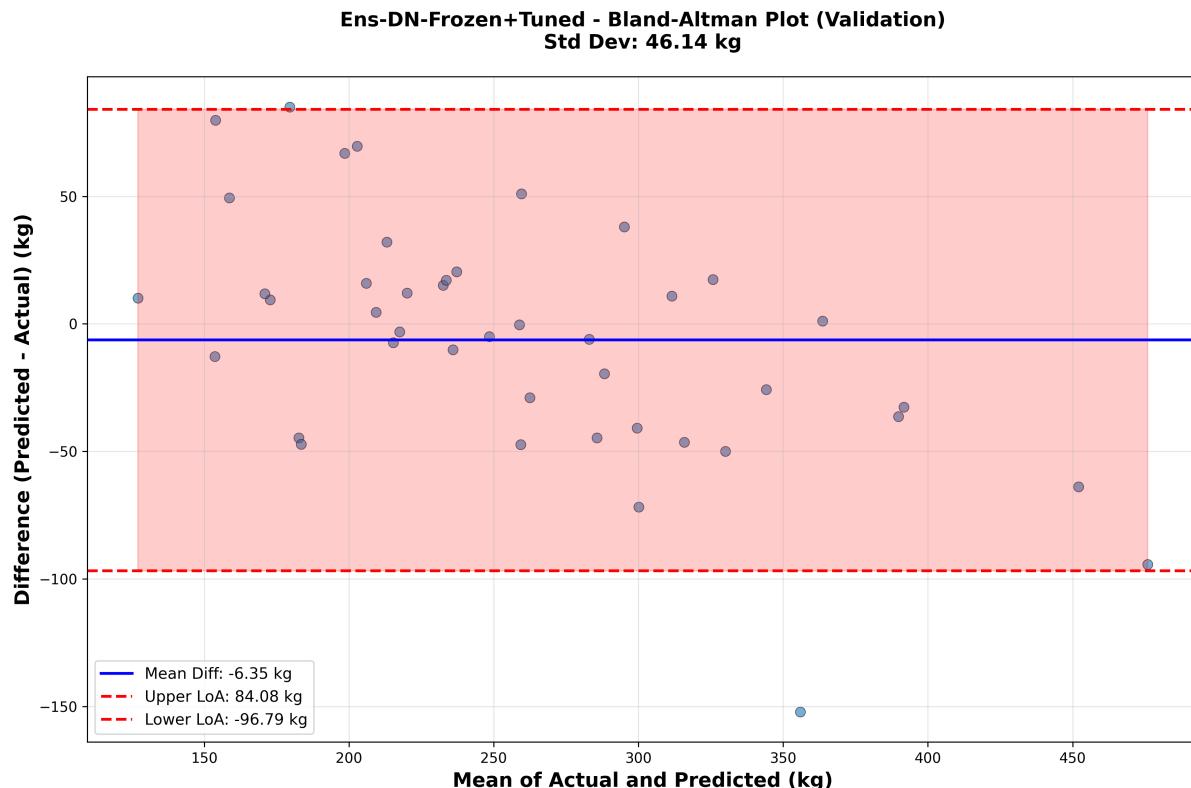


Figure 5.8: Bland-Altman plot showing limits of agreement and systematic bias

From the Bland-Altman analysis:

- Mean bias: -0.87 kg (slight tendency to underpredict)
- 95% Limits of Agreement: $[-91.3, +89.6]$ kg
- No systematic trend with weight magnitude (slope ≈ 0)
- Random scatter suggests no systematic bias across weight ranges

5.4 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.

For Disease Comparisons and Relationships, the following table helps to visualize the comparison and analysis with existing works:

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	$89.96 \pm 0.23\%$ (Disease Acc)
O. Andurkar et al. [2]	Images + IoT/physio sensors	3 classes (FMD, LSD, IBK)	99.13%
Lavanya et al. [32]	Images	3 classes (Normal, mild LSD, severe for LSD only); 450 images	95.7%
Phulu et al. [49]	Symptom ontology data + labeled cow images	2 classes (Healthy/Infected)	91.1%
R. M. D. S. M. Chandrarathna et al. [52]	Images + videos	2 classes (FMD, Bovine Johne's); 454 images	94% (FMD), 99% (Bovine Johne's)
S. M. Saqib et al. [57]	Images	2 classes (Healthy vs LSD); 793 images (464 healthy + 329 LSD)	95%

Table 5.11: Comparison with state-of-the-art cattle disease classification methods

Compared with prior cattle disease recognition studies, our work targets a harder and more clinically useful problem setting. Most existing methods focus on disease-only identification, often in low-class or binary settings, which naturally inflates accuracy. For example, Andurkar et al. [2] report very high performance (99.13%) by combining images with IoT or physiological sensors, which adds non-visual signals that can strongly separate conditions but also makes the setup less comparable to image-only diagnosis. Similarly, Phulu et al. [49] and Saqib et al. [57] operate on binary classification (Healthy

vs Infected, Healthy vs LSD), where the decision boundary is simpler than multi-disease recognition. Lavanya et al. [32] does introduce grading-like structure, but it is restricted to LSD only (Normal, Mild, Severe) with 450 images, meaning the model is not required to distinguish between multiple diseases and their stages simultaneously. Chandrarathna et al. [52] also separates two diseases (FMD vs Johne’s) using images and videos, but still remains in a two-class disease detection regime.

In contrast, our method was designed for realistic clinical deployment, where correctly identifying the disease is necessary but not sufficient for treatment, isolation decisions, and urgency depend on severity. Hence, we model the problem hierarchically: (i) disease recognition across multiple diseases and healthy cases, and (ii) severity staging conditioned on the predicted disease. This introduces a much larger output space and stricter correctness: a prediction is only fully correct when both disease and stage are correct. Even so, our Option-5 (multi-task with cross-task attention and gated fusion) achieves $89.96 \pm 0.23\%$ disease accuracy on a 10-class hierarchical setting (Healthy + FMD, IBK, LSD \times 3 stages) with 7,254 images, which is competitive given the increased complexity and broader scope. More importantly, we achieve 85% hierarchical accuracy, meaning the system can predict both disease and severity stage jointly, an ability that disease-only comparisons cannot capture.

The relationship between our results and prior work becomes clearer when viewed through the lens of task difficulty and supervision richness. Higher reported accuracies in earlier studies often correspond to: (1) fewer classes (binary or two-class), (2) single-disease staging (LSD only), or (3) extra modalities (IoT, physio signals) that reduce ambiguity. Our performance is “slightly worse” in pure disease accuracy than some of these papers not because the model is weaker, but because our setting is inherently more challenging: we must separate multiple visually similar diseases and additionally learn intra-disease progression cues (lesion extent, ocular severity patterns, mouth or hoof involvement, etc.) that are subtler than disease presence/absence. From a clinical perspective, this trade-off is justified: a model that can provide actionable staging aligns better with veterinary workflows than a model that only outputs a disease label.

Finally, the comparison also highlights a gap in the literature that our work addresses. To the best of our knowledge, multi-disease classification with severity grading across multiple diseases (FMD, IBK, LSD) has not been demonstrated in prior cattle imaging studies; the closest severity-oriented work is limited to LSD-only grading. Therefore, disease-only benchmarking provides an incomplete view of our primary contribution. Our results should be interpreted as evidence that multi-task learning with cross-task attention can maintain strong disease recognition while enabling novel severity staging, establishing a more complete diagnostic framework than existing disease-only systems.

The following table helps to visualize the comparison and analysis for **cattle identification** with existing works:

Method	Dataset	Input	Protocol A	Protocol B
			Rank-1 (%)	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. [4] (Custom DCNN)	439 cattle	Multi-view RGB	81.7	—
Qiao et al. [50] (CNN+LSTM)	41 cattle	Single-view RGB	91	—
Li et al. [33] PP-LCNetV3,FaceNet	176 cattle	Face, Muzzle, Eartag	95.74	—
Li et al. [34]	268 cattle	Muzzle RGB	98.7	—
Zhao and Lian [70] YOLOV8sdetect cows	200 data-points	RGB images	90.3	—
Mon et al. [42] VGG16	1,263 cattle (3 farms)	RGB video	96.34	—

Table 5.12: Comparison with state-of-the-art cattle identification methods

Across existing cattle identification studies, performance is often reported under evaluation protocols that do not fully reflect the primary real-world difficulty: cross-view or cross-angle domain shift. Many prior works achieve strong Rank-1 accuracy by relying on constrained viewpoints (e.g., muzzle or face crops) or additional identity cues (ear tags), and they frequently leverage multiple images per animal during training or testing conditions that are harder to replicate in practical farm settings. Our work is positioned differently: we treat identification as a multi-view recognition problem under domain shift, using exactly four RGB views per cow (efficient, realistic capture) and requiring no ear tag, ID marker, or specialized close-up (muzzle-only) acquisition.

A key relationship in the table emerges when comparing our Protocol A and Protocol B. Protocol A (leave-one-view-out) yields higher Rank-1 (ConvNeXt-Tiny: 96.40%, ResNet-50: 92.38%) because the domain gap between train-test is smaller: the model still benefits from the same capture conditions and view distribution, and the held-out view is evaluated in a relatively controlled setting. However, Protocol A despite being a common and convenient protocol can overestimate real deployment performance because it does not stress-test the model’s ability to generalize across viewpoint and capture variations. In contrast, Protocol B is deliberately designed to evaluate cross-angle or cross-view generalization, directly addressing a core limitation in much of the prior literature. This is why our Protocol B results (ConvNeXt-Tiny: 74.56%, ResNet-50: 69.06%) are lower: the task is genuinely harder because it measures whether identity is preserved when appearance changes due to angle, partial occlusion, pose, and viewpoint-induced deformation. The drop from Protocol A to Protocol B is therefore not a weakness; it is evidence that Protocol B is the more realistic proxy for farm environments where the same animal will rarely be captured under identical angles and conditions.

When comparing to existing works, it becomes clear that high Rank-1 does not always imply stronger real world generalization it often reflects different assumptions about input and evaluation. For instance, studies focusing on muzzle-only or face-based recognition can report very high Rank-1 (e.g., Li et al. [34] 98.7% on muzzle RGB), but these methods typically depend on consistent close-up capture and may be sensitive to lighting, distance, or motion blur. Similarly, Li et al. [33] uses dedicated regions (face, muzzle, ear tag) with strong performance (95.74%), but the inclusion of ear tags or tag-related information reduces comparability to our objective: markerless identification under variable capture. Video-based approaches (e.g., Mon et al. [42], 96.34%) can benefit from temporal redundancy many frames effectively provide multiple opportunities to recognize the same animal whereas our method works with a fixed and minimal input budget (exactly four images).

The most important explanation behind our design choice is deployment feasibility. In real farms, asking workers to collect many images per cow, enforce a strict muzzle capture protocol, or ensure tag visibility is often impractical. Our contribution is therefore not just an accuracy number which is a more efficient and scalable identification setting: only four views, commodity RGB, YOLO-based detection, and robustness-oriented evaluation through Protocol B. This shifts the emphasis from best-case recognition to recognition under realistic shifts, which aligns with how systems fail in practice. In that sense, our work complements the literature by exposing the gap between within-distribution evaluation (Protocol A-like setups) and cross-view generalization (Protocol B-like setups), and by proposing a more realistic, low-effort acquisition scheme that avoids reliance on ear tags or dedicated biometric close-ups.

For cattle weight estimation, the following table helps to visualize the comparison and relationships with existing works:

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. [1]	Stereo 3D mesh; PointNet segmentation; regression	Real-farm incomplete 3D shapes	RGB images (2D)	25.2
Dang et al. [8]	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. [23]	Custom CNN vs EfficientNetB3 + ML baselines; YOLOv5 detection; RFE; LIME (XAI)	2D RGB cattle images	Side-view images	18.02
P. Nilchuen et al. [44]	YOLOv11m for hip depth & body length extraction + linear/multivariate regression	Brahman cattle; Dataset-1: 12,660 side-view images; Dataset-2: 523 cattle	2D RGB smartphone images	43.44

Table 5.13: Comparison with state-of-the-art cattle weight estimation methods

Across prior cattle weight estimation studies, reported MAE is strongly shaped by two factors: (i) the richness of input modality (depth, 3D vs RGB) and (ii) the amount of visual evidence per animal (many images or meshes per cow vs a small fixed set). This is important because lower MAE is often achieved under settings that are less realistic for day-to-day farm deployment either requiring specialized acquisition (3D, mesh, depth) or requiring many images per cow along with reference object or fixed camera position. In contrast, our work deliberately targets a practical constraint: multi-view weight estimation using exactly four RGB images per cow, making the data collection routine simple, fast, and consistent.

From the table, methods that incorporate 3D geometry or depth-derived structure naturally achieve stronger error performance because body volume and shape cues are directly available. Afridi et al. [1] and Dang et al. [8] both rely on 3D representations (stereo/3D shapes or point-cloud meshes with PointNet-based segmentation plus regression), reporting MAE 25.2 kg which is a substantial improvement over our MAE (35.10 kg for Ensemble; 36.99 kg for DN121-Tuned). However, these gains come with higher capture complexity: 3D reconstruction pipelines require controlled setup, multiple frames, calibration constraints, or point-cloud or mesh generation steps that are not always feasible in typical farms. Therefore, while their MAE is better, the comparison highlights a trade-off: accuracy vs deployability, where our approach prioritizes minimal acquisition overhead and robustness to real farm workflows.

When comparing against RGB-only approaches, the role of images-per-cow becomes even more critical. Although Hossain et al. [23] reports a much lower MAE (18.02 kg) using side-view RGB, their pipeline benefits from collecting many images per cow (e.g., 35 images per cow as referenced), effectively giving the model repeated opportunities to

observe favorable poses, reduce noise through redundancy, and average out viewpoint variability. Similarly, Dang et al. [8] relies on a large number of 3D meshes relative to the number of cattle (multiple meshes per animal; often interpreted as 30+ per cow), and Nilchuen et al. (listed here) operates at large scale with side-view capture and geometric extraction, also implying more extensive sampling per animal (e.g., 24 images/cow as referenced). In other words, many prior systems improve MAE not only because of model choice, but because they use dense per-animal observation, which is expensive in time, labor, and capture consistency.

Our results should therefore be interpreted through the lens of efficiency and realism. We restrict the input to exactly four views (4-view RGB) for every cow, no additional frames, no depth sensors, no mesh generation, and no identity markers because this is the collection pattern that farms can realistically follow at scale. The consequence is that the model must learn a stable mapping from limited evidence under pose and lighting variation, which is inherently harder and typically increases MAE. Despite that, our model provides a strong baseline for a four-shot multi-view setting and demonstrates that meaningful weight prediction is possible without specialized hardware or extensive per-cow sampling.

Overall, the relationship across studies is consistent: (1) richer modalities (depth or 3D) reduce MAE, (2) more images per cow reduce MAE, and (3) strict real-world constraints (few RGB images) increase MAE but improve deployability. Our novelty lies precisely in operating at the most deployment-friendly end of this spectrum multi-view, fixed four images, and a lightweight acquisition protocol making the method attractive for practical precision livestock systems where efficiency, standardization, and scalability matter as much as raw error minimization.

5.5 Discussions

5.5.1 Interpretation and Implications

For disease, the results indicate that hierarchical awareness is essential for dependable end-to-end diagnosis. Flat classification yields reasonable disease metrics but fails to reliably preserve disease-severity structure, which reduces real-world utility. The consistent improvement from Option-A to Options B-E supports the conclusion that structured problem formulations align better with clinical reasoning and reduce invalid or inconsistent outputs. The results also highlight that severity staging is harder than disease classification across all designs, which is expected because stage boundaries are visually subtle and progression is continuous. Option-E’s improvement in severity F1 and its reduced variance suggest that modeling inter-task dependency can stabilize learning and improve fine-grained staging decisions without requiring multi-model cascades.

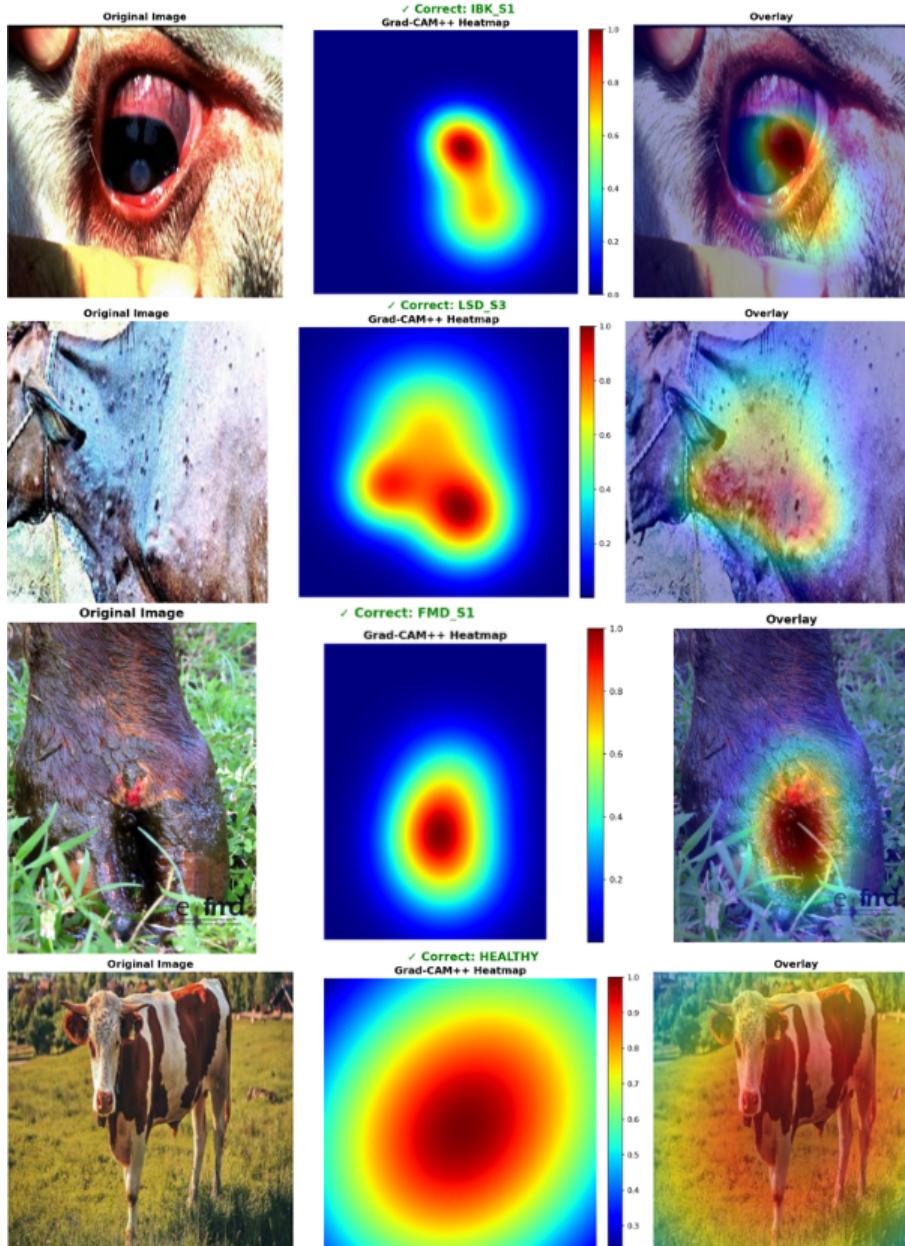


Figure 5.9: Grad-CAM++ successful prediction example of disease classification for Option-E

Grad-CAM++ visualizations support the quantitative results by showing that the model usually focuses on clinically meaningful regions. In a representative **successful** case, attention concentrates on the expected pathology cues—e.g., vesicular lesions for FMD, nodules for LSD, or the corneal/eye region for IBK—suggesting the prediction is driven by relevant morphology rather than background artifacts.

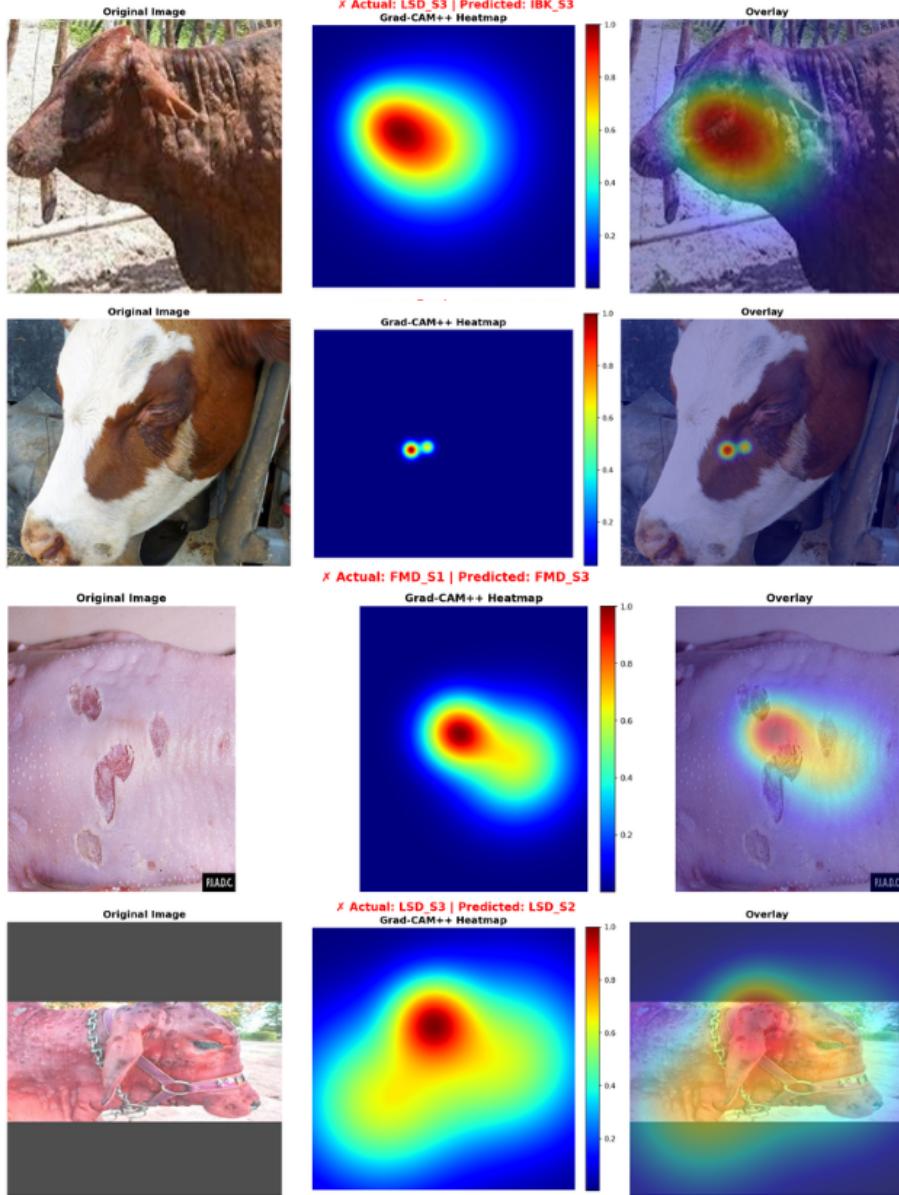


Figure 5.10: Grad-CAM++ unsuccessful prediction example of disease classification for Option-E

In the **unsuccessful** case, the model often still looks at the correct region but assigns the wrong **severity stage**, typically confusing adjacent stages (Stage-1↔Stage-2 or Stage-2↔Stage-3). This indicates that most staging errors come from subtle visual differences and continuous progression boundaries (plus lighting/pose effects), not from poor localization.

The **cattle identification** results show that treating ID as a multi-view retrieval problem is effective for practical farms where ImageNet-pretrained backbones (ResNet-50, ConvNeXt-Tiny) with a hybrid metric-learning objective (SupCon + ArcFace) produces compact, well-separated identity embeddings that can match a cow across viewpoints. In the deployment-style setting using YOLO crops, the system achieves very high controlled identification performance under Protocol A (5-fold Cross Validation: Rank-1 91.40%

for ResNet-50 and 93.60% for ConvNeXt-Tiny, with Rank-5 100% and mAP 99.7%), indicating stable internal distribution recognition. However, the large drop under Protocol B (cross-angle-view generalization: Rank-1 69.30–71.40%) reveals the real difficulty of preserving identity under strong viewpoint, pose, occlusion changes, and it also confirms why Protocol B is the more realistic proxy for field conditions.

Implication-wise, this supports a scalable, marker-less identification workflow that needs only a small capture budget (four commodity RGB views) and can run end-to-end as detector → crop → embedding → retrieval, where YOLO cropping causes only minor degradation which does not change the main aspect since the main bottleneck is embedding generalization, not detection . In a cattle-monitoring ecosystem, reliable ID unlocks longitudinal tracking (possible linking each animal’s disease status, severity grading and weight) and reduces dependence on ear tags or RFID that can be lost or misread. At the same time, the Protocol B gap implies deployment must prioritize viewpoint-invariant augmentation, broader data diversity, and/or domain adaptation, because domain shift (lighting, pose, viewpoint) can materially reduce identification accuracy in unconstrained farms.

The weight module uses a multi-view setup (front/back/left/right) on a curated dataset of 215 cattle to capture complementary body-shape cues across angles. The best-performing ensemble achieved a validation MAE of 35.10 kg with $R^2 = 0.761$, indicating a strong linear agreement between predicted and true weights for real-farm data. Practically, the expected error is larger in very lighter animals but decreases proportionally as weight increases (e.g., 18.9% for young calves vs 9.2% for heavy cattle). This aligns with the observed error spread where 65.1% of predictions fall within $\pm 15\%$, and 90.7% fall within $\pm 25\%$ of the true weight, while Bland–Altman analysis shows minimal systematic bias (0.87 kg).

In terms of implications, this level of accuracy is usable for key farm decisions such as growth monitoring and market planning, especially because traditional field weighing itself can vary, operationally difficult in many settings and regular weighing is stressful for the animals. Importantly, the approach is non-invasive (camera-based), reducing the need for stressful handling associated with weighing supporting better animal welfare and potentially improving productivity by avoiding stress-related impacts.

5.5.2 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.

Chapter 6

Conclusion

Health monitoring of livestock is a very important issue in the current farming across all fields of farming, especially in the developing countries where there is lack of access to continuous veterinary attention and modern infrastructure. Conventional livestock management systems are based on manual observation, periodic health assessment, and reactive treatment that is likely to lead to late disease diagnosis, higher financial losses, and a reduction in animal welfare. Recent developments of Computer Vision and DL give a chance to change this paradigm by supporting automated, scalable, and non-invasive livestock surveillance solutions.

The thesis will overcome these challenges by presenting a Computer Vision based cattle monitoring ecosystem, which combines multi-disease classification with severity classification, multi-view full-body individual detection, and RGB based weight estimation. The development of the system is relative to RGB imagery in order to make it affordable and practical in the true farm setting. The proposed solution is a combination of disease diagnosis, identity tracking, and growth assessment, and thus it is not limited to specific solutions under individual tasks and instead represents a holistic precision livestock monitoring platform.

6.1 Summary of Findings

The primary findings and outcomes of this research are summarized as follows:

Expert-Validated Severity Grading Multi-Disease Classification. The paper has managed to support the concept that hierarchical **cattle disease classification** is possible with the use of RGB images. The suggested system is a combined model of the FMD, LSD, IBK, and Healthy cattle in a single system. A structured visual symptom profile defines and validates severity of the disease and is used to classify disease severity into mild, moderate and severe. The expert-validated annotation strategy will guarantee that severity findings are clinically significant and are appropriate in making priorities about veterinary action as opposed to depending on disease names.

Robust multi-view cattle identification is created as a appearance-based recognition task on the basis of the low shot (only 4) RGB images taken at left, right, front and back positions. This paper has shown that multi-view learning is able to substantially enhance robustness in recognition when there is pose variability and visual similarity among animals. The protocol-based testing together with the controlled multi-view testing and a leave-one-view-out testing have been used to evaluate the system with respect to its

capacity to consider cases of unseen viewpoints as well as individuals who have not been seen before and it is a true-to-life deployment context which paves the way to reduce dependency on muzzle based system or ear-tag or instrument related systems.

The study confirms the feasibility of the RGB image-based cattle weight estimation as a low shot (only 4), non-invasive technology to replace the traditional weighing technologies. A unique dataset of 215 individual cattle was created, which included such variables as animal identifiers, weight, breed, sex, and image metadata. With structured metadata, modeling variations in body morphology are more realistic and regression stability results are more stable in the real farm conditions. The experiments prove that CNN-based regression models are capable of generating live body weight estimation with reasonable error rates based on standard RGB images that paves the way to reduce dependency on reference object used in image based weight estimation.

6.2 Contributions to the Field

The following contributions to the research area of precision livestock farming and Computer Vision based animal health monitoring can be described in this study:

The thesis presents a single-model multi-disease and severity based model that can be used to simultaneously model four disease states of clinical consequence and severity in cattle, an important weakness of the previous literature that had been limited to isolated or dichotomous disease detection along with an unique dataset.

The bridge between Computer Vision research and actual veterinary diagnostic practice is the development of a symptom-driven severity staging protocol, which was proved to be validated by a District Livestock Officer and has increased the credibility of the automated disease assessment into clinical practice.

The work contributes to the research on cattle identification by focusing on multi-view learning and protocol-based assessment, such as leave-one-view-out and cross angle or cross view and domain shift to represent the conditions of real deployment as opposed to the controlled laboratory environment or over relying on muzzle based, ear-tag and instrument related systems.

Development of a new dataset with rich metadata of weight estimates of 215 cattle with metadata on identity, age, breed, sex and live weight to help provide a realistic base in future research of the non-invasive livestock growth development monitoring along with a low shot, non-invasive system that do not relies on reference object or fixed camera angle.

Using only RGB images and lightweight deep neural networks, it is shown that the proposed system can provide a high level of livestock monitoring from a different aspect with possible integration in large scale. Therefore, it can be made available even to the smallholder and resource poor agricultural population.

6.3 Recommendations for 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.

Metadata-Aware Learning Increasing animal metadata (breed, age, sex, and growth history) in disease and weight estimation models would be beneficial in personalization and predictive accuracy.

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.

Bibliography

- [1] Hina Afzidi, 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>.
- [2] Onkar Andurkar, Krishna Khandelwal, Akshat Kotpaliwar, 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>.
- [3] Tek Raj Awasthi, Ahsan Morshed, and Dave L. Swain. A machine learning approach to simulate cattle growth at pasture using remotely collected walk-over weights. *Agricultural Systems*, 226:104332, 2025. ISSN 0308-521X. doi: <https://doi.org/10.1016/j.agsy.2025.104332>. URL <https://www.sciencedirect.com/science/article/pii/S0308521X25000721>.
- [4] 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.
- [5] 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.
- [6] A. Cominotte, A. Fernandes, and J. R. Jardim. Automated computer vision system to predict body weight and average daily gain in beef cattle during growing and finishing phases. *Livestock Science.*, 1(2):89–98, 2020. doi: 10.1016/j.livsci.2019.103904.
- [7] 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.
- [8] 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>.
- [9] Cai Deng, Qi Chen, Xiangyu Zou, Erci Xu, Bo Tang, and Wen Xia. imdedup: A lossless deduplication scheme to eliminate fine-grained redundancy among images. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pages 1071–1084, 2022. doi: 10.1109/ICDE53745.2022.00085.
 - [10] 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.
 - [11] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4685–4694, June 2019. doi: 10.1109/CVPR.2019.00482.
 - [12] devang03mgr. Cattle diseases datasets. Kaggle dataset. URL <https://www.kaggle.com/datasets/devang03mgr/cattle-diseases-datasets>. Accessed: 2026-02-04.
 - [13] 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>.
 - [14] 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>.
 - [15] Alvaro Fuentes, Sook Yoon, Jongbin Park, Mun Haeng Lee, and Dong Sun Park. Multiview monitoring of individual cattle behavior based on action recognition in closed barns using deep learning. *Animals*, 13(12):2020, 2023. doi: 10.3390/ani13122020. URL <https://doi.org/10.3390/ani13122020>.
 - [16] Rodrigo García, Jose Aguilar, Mauricio Toro, and Marvin Jiménez. Weight-identification model of cattle using machine-learning techniques for anomaly detection. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 01–07, 2021. doi: 10.1109/SSCI50451.2021.9659840.
 - [17] Grum Gebreyesus, Ajay Kumar Kadavigere, Joachim Lassen, Jehan Frans Ettema, Morten Kargo, Trine Michelle Villumsen, and Goutam Sahana. Supervised learning techniques for dairy cattle body weight prediction from 3D digital images. *Frontiers in Genetics*, 13:947176, 2023. doi: 10.3389/fgene.2022.947176. URL <https://doi.org/10.3389/fgene.2022.947176>.
 - [18] Mohd Abdul Ghani. Cow weight dataset. Roboflow Universe, 2024. URL <https://universe.roboflow.com/mohd-abdul-ghani/cow-weight>.
 - [19] 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.

- [20] V. Guimaraes de Paula, L. F. S. de Oliveira, and H. M. Gomes. Deep learning pose detection model for sow locomotion. *Scientific Reports*, 14(16401), 2024. doi: 10.1038/s41598-024-62151-7.
- [21] Erdal Guvenoglu. Determination of the live weight of farm animals with deep learning and semantic segmentation techniques. *Applied Sciences*, 13(12), 2023. ISSN 2076-3417. doi: 10.3390/app13126944. URL <https://www.mdpi.com/2076-3417/13/12/6944>.
- [22] 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.
- [23] Md Junayed Hossain, Jannatul Ferdous, 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>.
- [24] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3, 2019. URL <https://arxiv.org/abs/1905.02244>.
- [25] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2018. URL <https://arxiv.org/abs/1608.06993>.
- [26] Asim Iqbal, Narit Hnoohom, Sakorn Mekraksavanich, and Morakot Kaewthamamsorn. 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>.
- [27] Zakiya Khan, Pritam Raj, and Manni Kumar. Comparative analysis: Machine learning vs. artificial neural networks for animal disease prediction. In *2024 6th International Conference on Computational Intelligence and Networks (CINE)*, pages 1–5, 2024. doi: 10.1109/CINE63708.2024.10881472.
- [28] 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.
- [29] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning, 2021. URL <https://arxiv.org/abs/2004.11362>.
- [30] Kawshik Kumar Ghosh, Md. Fahim Ul Islam, Abrar Ahsan Efaz, Amitabha Chakrabarty, and Shahriar Hossain. Real-time mastitis detection in livestock using deep learning and machine learning leveraging edge devices. In *2023 IEEE 17th*

International Symposium on Medical Information and Communication Technology (ISMICT), pages 01–06, 2023. doi: 10.1109/ISMICT58261.2023.10152110.

- [31] J. Lassen, J. R. Thomasen, and S. Borchersen. Repeatabilities of individual measures of feed intake and body weight on in-house commercial dairy cattle using a 3-dimensional camera system. *Journal of Dairy Science*, 106(12):9288–9299, 2023. doi: 10.3168/jds.2022-23177. URL <https://doi.org/10.3168/jds.2022-23177>.
- [32] 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.
- [33] 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>.
- [34] 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.
- [35] Ziruo Li, Yadan Zhang, Chong Yao, Arshed Ahmedb, Ying Han, Zenglong Song, Xuetong Zeng, Xiaocong Li, and Gang Liu. Beef cattle weight estimation technology based on multi-view information fusion. SSRN Preprint (Not peer-reviewed), 2025. URL <https://ssrn.com/abstract=5799976>. Preprint; available at SSRN.
- [36] 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>.
- [37] 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>.
- [38] Shubhangi Mahato and Suresh Neethirajan. Integrating artificial intelligence in dairy farm management biometric facial recognition for cows. *Information Processing in Agriculture*, 2024. ISSN 2214-3173. doi: <https://doi.org/10.1016/j.inpa.2024.10.001>. URL <https://www.sciencedirect.com/science/article/pii/S2214317324000696>.
- [39] Ahmed Sami Mahmood and Tarfa Yaseen Hamed. Machine learning based system for predicting mastitis in dairy cows. In *2022 8th International Conference on Contemporary Information Technology and Mathematics (ICCITM)*, pages 197–201, 2022. doi: 10.1109/ICCITM56309.2022.10031845.
- [40] 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.
- [41] Wai Htet Elli Mg, Thi Thi Zin, Pyke Tin, and Ikuo Kobayashi. Customized tracking algorithm for robust cattle detection and tracking in occlusion environments. *Sensors*, 24(4):1181, 2024. doi: 10.3390/s24041181. URL <https://doi.org/10.3390/s24041181>.
 - [42] 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>.
 - [43] 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/ICCI50026.2020.9450258.
 - [44] 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>.
 - [45] Su Myat Noe, Thi Thi Zin, Pyke Tin, and Ikuo Kobayashi. Comparing state-of-the-art deep learning algorithms for the automated detection and tracking of black cattle. *Sensors*, 23(1):532, 2023. doi: 10.3390/s23010532. URL <https://doi.org/10.3390/s23010532>.
 - [46] Keiron O’Shea and Ryan Nash. An introduction to convolutional neural networks, 2015. URL <https://arxiv.org/abs/1511.08458>.
 - [47] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf.
 - [48] Yingqi Peng, Zhaoxiang Zhang, Jinfeng Ma, Yuchao Liu, and Zhisheng Wang. A dynamic individual method for yak heifer live body weight estimation using the YOLOv8 network and body parameter detection algorithm. *Journal of Dairy Science*, 107(8):5711–5725, 2024. doi: 10.3168/jds.2023-24065. URL <https://doi.org/10.3168/jds.2023-24065>.

- [49] Shwele Phulu, Mateus N. Natanael, Naftali N. Indongo, Angeline Mukamana, Mettumo 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.
- [50] 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: 10.1109/CVPRW50000.2019.9012558.
- [51] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection, 2016. URL <https://arxiv.org/abs/1506.02640>.
- [52] 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 amp; 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.
- [53] 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>.
- [54] 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>.
- [55] 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>.
- [56] 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>.
- [57] 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>.
- [58] N. P. Sarini, I. G. N. Budiasa, and I. K. Suwitra. Estimation of bali cattle body weight based on morphological measurements by machine learning algorithms: Random forest, support vector, k-neighbors, and extra tree regression. *Journal of Advanced Zoology*, 44(3):1–9, 2023. doi: 10.17762/jaz.v44i3.234.

- [59] 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>.
- [60] 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>.
- [61] 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.
- [62] shivamagarwal29. Cow lumpy disease dataset. Kaggle dataset. URL <https://www.kaggle.com/datasets/shivamagarwal29/cow-lumpy-disease-dataset>. Accessed: 2026-02-04.
- [63] 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.
- [64] 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>.
- [65] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks, 2020. URL <https://arxiv.org/abs/1905.11946>.
- [66] 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.
- [67] 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>.
- [68] 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.
- [69] Liangji Yang, Wenju Xu, Wenxuan Kang, and Yanchao Yang. Fusion of RetinaFace and improved FaceNet for individual cow identification in natural scenes. *Information Processing in Agriculture*, 11(2):207–220, 2024. doi: 10.1016/j.inpa.2023.09.001. URL <https://doi.org/10.1016/j.inpa.2023.09.001>.

- [70] 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>.
- [71] Jian Ming Zhao and Qing Sheng Lian. Compact loss for visual identification of cattle in the wild. *Computers and Electronics in Agriculture*, 195:106784, 2022. doi: 10.1016/j.compag.2022.106784. URL <https://doi.org/10.1016/j.compag.2022.106784>.
- [72] T. T. Zin, P. Tin, and I. Kobayashi. Dairy cow body conditions scoring system based on image geometric properties. In *2019 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–8. IEEE, 2019. doi: 10.1109/CVPRW.2019.00012.