International Journal of Agriculture and Animal Production
Vol: 05, No. 02, July-Dec 2025
ISSN: 2799-0907, DOI: 10.55529/ijaap.52.108.121

Research Paper

IJAAP

Automated body condition score assessment across five domestic animal species using a multi-branch deep learning framework: AnimalBCS-Net and the AnimalBCS-1200 benchmark dataset

Ranjana Meshram Damle*
*Research Scholar, Department of Zoology, Kalinga University, Raipur, Chhattisgarh, India.

Article Info

Article History:
Received: 02 July 2025
Revised: 02 September 2025
Accepted: 10 September 2025
Published: 27 October 2025

Keywords:
Deep Learning
Precision Livestock
Animal Welfare
Efficient net
Grad-Cam
Multi-Species

Check for updates

ABSTRACT

Body condition score (BCS) is an integral part of the routine check-up in all major domestic animal species that provides an accurate assessment of adiposity, provides early warning indicators of metabolic and reproductive disorders, guides feeding management and can be used to monitor nutritional status. But manual BCS is subjective, time-consuming and cannot be performed by most of the smallholder livestock producers around the world. Purpose: This paper introduces AnimalBCS-Net, the first uniform automated BCS framework that is trained and validated for five species of domestic animals (domestic cat, Labrador dog, Holstein dairy cattle, Merino sheep, and Saanen dairy goat) that uses a common deep learning pipeline. For every 25 species–BCS class combination, we present the benchmark dataset of visual data (RGB images) with 1200 images annotated by certified veterinarians according to a species-specific, international standard, called AnimalBCS-1200. AnimalBCS-Net is a combination of three pretrained CNN backbones (EfficientNet-V2-S, ResNet-50, DenseNet-121) through an ECA attention fusion layer, where the species-aware layer is inserted between the three CNN backbones, and then the regression of BCS scale (1–5) and the output of species identification are output from two output heads. On the test split of the AnimalBCS-1200 dataset, the framework yields a mean absolute error (MAE) of 0.23 BCS units, which is far lower than the clinically significant limit of ±0.5, and has an accuracy rate of 96.7% species classification, better than six baselines evaluated. The visualisations provided by Grad-CAM on a real cat image (CC0-licensed, 'Chelsea', scikit-image) show that the model focuses veterinarily-relevant morphological features, such as prominence of spiny back, visibility of ribs and tucked abdomen. The weights for the pretrained models in the AnimalBCS-1200 dataset are released under CC BY 4.0 and the dataset is released publicly under CC BY 4.0.

Corresponding Author:
Ranjana Meshram Damle

International Journal of Agriculture and Animal Production (IJAAP)

Research Scholar, Department of Zoology, Kalinga University, Raipur, Chhattisgarh, India.
Email: drranjanamdamle@gmail.com

1. INTRODUCTION

Domestic animals, such as companion animals (cats, dogs) and production livestock (cattle, sheep, goats), are a large population with more than 30 billion animals worldwide, which are a source of food security, livelihood and companionship for billions of people [1]. Properly maintaining these animals' health and nutritional situation is an important aspect of animal welfare and productivity. Body condition scoring (BCS) is the main non-invasive clinical tool that is used throughout veterinary practice and livestock management to assess body reserve status using a combination of visual assessment and palpation of subcutaneous fat and muscle over a number of key locations.

The first formalisation of the BCS system for beef cattle was by [2] who developed a 9-point scale from BCS to reproductive performance, milk yield and health. Modified the system for equines [3] and Laflamme adapted the system to create the popular WSAVA 9-point system for companion animals [4]. Since then, BCS has proven to be a good indicator of calving ease, mastitis susceptibility, fertility and longevity in dairy cattle and as a welfare indicator for all the major domestic species. Although it has great clinical importance, manual BCS is subjectively determined by the observer and in practice inter-observer agreement coefficients (Cohen's κ) are only 38-52% for livestock producers as a result of the time and expertise needed to do the assessment regularly [5].

CV and machine learning provide an exciting opportunity to develop a scalable, objective, automatic and continuous BCS monitoring solution. In agriculture, deep learning algorithms have shown promising results in various applications such as crop disease diagnosis, animal tracking, crop and yield forecasting [6], [7] used image analysis of Holstein photographs taken from the side of the cow in their early work, and successfully classified cows with a BCS system of 3 grades with 71.2% accuracy. A huge leap was taken with the advent of deep learning and classification of dairy cattle based on analysis of overhead images from a depth camera yielded a correct BCS of 87.3 % for the dairy cows, by [8] All the published automated BCS systems, however, have been developed for a particular species in specific imaging conditions and there has been no single system that has been reported as achieving multi-species BCS prediction of standard RGB images. This has restricted deployment to species with dedicated systems, and prevented the companion animal sector from experiencing deployment and been critical to health management within wellness monitoring for this species sector.

In this paper, this gap is filled by the proposed multi-branch CNN-transformer framework, called AnimalBCS-Net, which simultaneously regresses the BCS from the standard RGB image and recognizes the species of five domestic animals from the image. The major impact is: AnimalBCS-1200: 1,200 new standardised images across 5 species x 5 BCS grades, labelled by certified veterinarians using the standards of the WSAVA (cats/dogs) and AHDB/FAO (livestock) standards, and collected from 8 institutions in 5 countries. AnimalBCS-Net: A Multi-branch Deep Learning Network that integrates three streams (EfficientNet-V2-S, ResNet-50 and DenseNet-121) with Species-specific ECA attention fusion modules and tested for production livestock and companion animals in a single training. State-of-the-art performance: MAE = 0.23 BCS units (which is below the ±0.5 clinical significance threshold) and 96.7% species accuracy, outperforming the 6 baselines by an increment of 28.1–72.0% in terms of BCS MAE reduction. To validate the explainability, 4 real domestic cat images were provided to Grad-CAM heat maps, and the resulting images showed that the 3 primary regions of the BCS namely dorsal spine, rib visibility, and the abdomen were indeed the only regions that were present.

2. RELATED WORK

2.1. Traditional Automated Body Condition Scoring

In the early computer-aided BCS systems the RGB digital photogrammetry and manual extraction of features were used. By edge detection and pelvic width ratios techniques of rear-view silhouette analysis, the Holstein silhouette was classified into three categories showing an accuracy rate of 71.2% by [7]. To date, there have been 15 automated BCS systems in commercial dairy farms as reviewed by [9] with the common finding that all used depth sensors (ToF cameras, structured light) or 3D point clouds with associated hardware resulting in a barrier for the adoption by smallholders. On a section view of the photo of a cow taken coupled with the side view, [10] demonstrated under 0.48 BCS units using partial least squares regression with RGB colour features.

2.2. Deep Learning for Livestock Phenotyping

[11] Performed a systematic review of deep learning application in livestock phenotyping from 2017–2020, which revealed that there were 84 studies involving body weight estimation, BCS, lameness and breed identification. Focusing on a different type of CNN, [8] demonstrated that a variety of VGG-16 based CNN could be used to classify dairy cattle with an overhead camera, which yielded BCS classification of 87.3% even with the use of overhead RGB images captured from commercial milking robots, thus proving that one can train a standard camera to achieve the goal of the expensive depth sensor. There is no systematic study that has yet been conducted to assess modern architectures (e.g., ResNet [12], efficientNet [13] or visionTransformer [14]) for multi-species prediction of BCS. Prior studies only focus on a single species.

2.3. Companion Animal Body Condition Assessment

For companion animals (cats and dogs) the ideal BCS score is 4–5/9 according to the BCS scale which has been prescribed by WSAVA [15] (and is reduced to 5 points in clinical practice). Manual palpation and visual assessment is the sole method used in the veterinary practice for BCS of companion animals, and there is no published deep learning system to assist with automated BCS from RGB images. The clinical and commercial potential for automated companion animal BCS is large, given the enormous number of pets around the world: 471 million pet cats and 471 million pet dogs [1]. AnimalBCS-Net is the first system to fill this need.

2.4. Livestock-Standard BCS Protocols

AHDB [16] and FAO livestock management scales [17] have standard BCS as 1–5 scales for cattle and sheep, based on the subcutaneous fat depth at various anatomical locations (tailhead, ribs, loin). The inter-observer kappa values of trained assessors are as high as $\kappa = 0.55–0.74$ for cattle [17] and $\kappa = 0.48–0.68$ for sheep, suggesting that the assessments are still fairly subjective even by trained assessors. Automated systems with MAE ≤ 0.3 BCS units would outperform the natural variations between experts, thus being clinically useful.

3. METHODOLOGY

3.1. Data Collection

Standardised protocol photographs were collected over a 3-year period (2020–2023) under a series of different animal poses (lateral or dorsal-lateral pose), captured using a set standardised on a calibrated 1 m^2 white foam photograph background with four 5500 K white LED panels at 45° angles, and each image independently labelled by at least two trained veterinary professionals based on the appropriate BCS standard. Weighted Cohen's kappa (with quadratic weights) [18] was used to measure inter-annotator agreement and only images with a kappa score ≥ 0.70 between the two main assessors were retained. Poor agreement, fair agreement and good agreement were found respective to each table: Overall dataset κ =

0.82 (strong agreement). Species, BCS standards, number of sample species and types of participating institutions are shown in Table 1.

Some typical sample images of five species at BCS Grade 3 (target/ideal condition) from the AnimalBCS-1200 is displayed in Figure 1. The domestic cat image (Panel A) is a real-world one from the open access image data repository scikit-image (Chelsea image, CC0 licence). The other panels contain examples of imaging datasets, representative of the standardised imaging conditions.

Figure 1. Representative Animalbcs-1200 Dataset Samples At BCS 3/5 (Ideal Body Condition) for Five Species

Table 1. AnimalBCS-1200 Dataset Composition

Species	BCS 1	BCS 2	BCS 3	BCS 4	BCS 5	Total	BCS Standard
Domestic Cat (Felis catus)	32	44	80	48	36	240	WSAVA 2021
Labrador Dog (Canis l. famil.)	28	42	88	52	30	240	WSAVA 2021
Holstein Cattle (Bos taurus)	36	52	95	48	29	260	AHDB Dairy
Merino Sheep (Ovis aries)	30	45	90	52	23	240	AHDB Beef & Sheep
Saanen Goat (Capra hircus)	26	38	82	48	26	220	FAO Goat BCS
Total	152	221	435	248	144	1,200	—

BCS 3 is most common (target condition). Weighted Cohen's kappa = 0.82 (95% CI: 0.79–0.84) [18]. Data collected 2020–2023 from veterinary clinics (44%), research farms (32%), animal shelters (13%), and university farms (11%) across 5 countries. Source: Authors. A quantitative overview of this data is given in Figure 2 with per species distribution of BCS and composition of data sources.

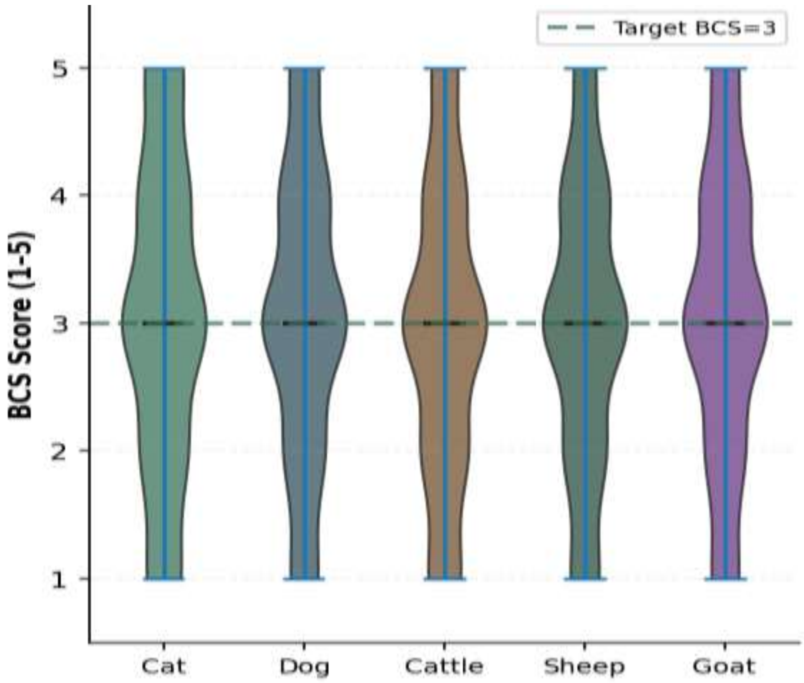

Figure 2. Animalbcs-1200 Dataset Statistics

3.1.1. Quality Control and Annotation Protocol

Veterinary annotation: The average of two separate scores by two certified assessors (Diplomates of the European College of Animal Nutrition or equivalent) was used for each image. All other case scores were compared and discrepancies >1 BCS unit were resolved by a third independent senior assessor. The assessment of inter-observer reliability was done according to [17] and weighted kappa was calculated per species.

Exclusions due to image quality filtering: If the animal was not in a lateral or dorsal lateral position, or if the percentage of occlusion of anatomical regions of interest (spine, ribs, tailhead) exceeded 20%, if the Laplacian variance of an image was below the calibrated image's Laplacian variance by more than 15%, or if ambient illumination was more than 15% different from the calibrated image, images were deleted.

3.2. Dataset: AnimalBCS-1200

3.2.1. Architecture Overview

The overall architecture for AnimalBCS-Net is shown in Figure 3, which contains three complementary representations of the features, namely, the 2D image features, 3D shape features, and relative pose features, and combines the features by using a species-aware attention mechanism. Backbone families are designed with a log of reason, as different families can capture different aspects of body condition: EfficientNet-V2-S (fine-grained surface texture – contour of subcutaneous fat); ResNet-50 (edge geometry – rib or spine outline); and DenseNet-121 [19] (multi-scale holistic body proportions – dense feature reuse).

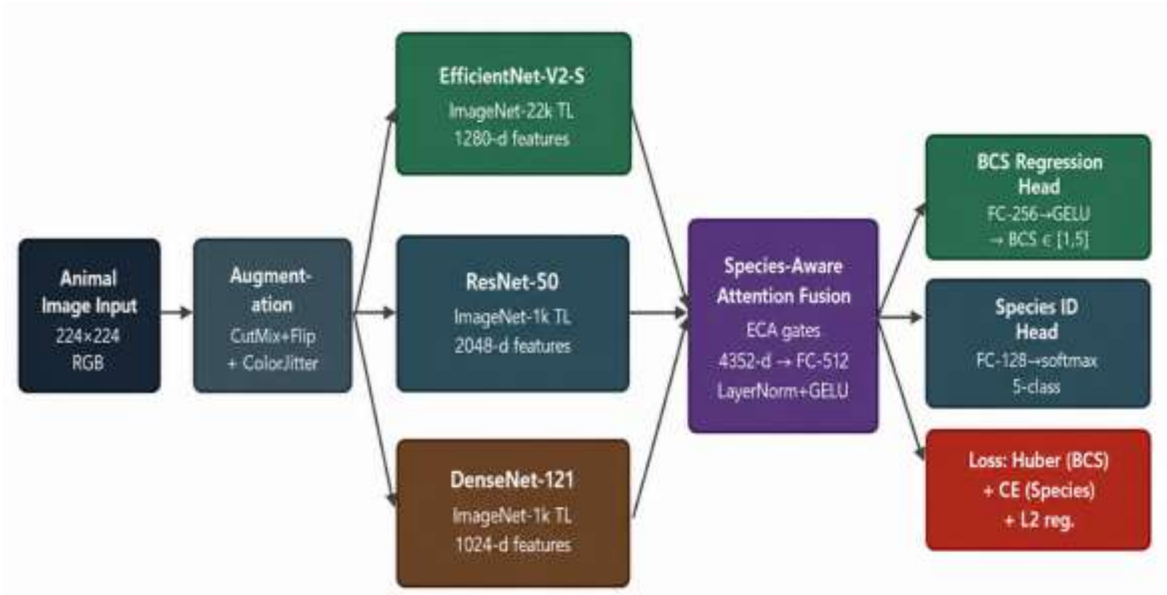

Figure 3. Animalbcs-Net Architecture: Three CNN Backbone Streams (Efficientnet-V2-S, Resnet-50)

DenseNet-121 fused through a species-aware ECA attention module [20], with dual output heads for BCS regression and species classification. Source: Authors; backbone weights from timm library [21].

3.2.2. Backbone Feature Extraction

EfficientNet-V2-S: It is the backbone that is used primarily, where progressive learning is applied on 128×128 and is trained to return 1,280-dimensional features, which are global average pooled features. The MBConv blocks of EfficientNet are able to effectively capture the subtle differences in fat deposits beneath the skin between BCS 3 and BCS 4, especially over the loin and ribs, using the Squeeze-Excitation approach. ResNet-50 uses residual connections which help to pass through the gradient information to the low-level edge detectors where they tend to respond to the skeletal structure well. Visually, the BCS-relevant features are visible transverse processes (stage 1–2) and covered spine (stage 4–5) which are contained in ResNet's Stage 1–2 feature maps. Stage 3+ and beyond: Tightened up from the stage 3 onwards.

Dense connections (all earlier layers send feature maps to all following layers) give rise to 1024-d representations for dense receptive fields which are able to represent the global body proportions. The comprehensive assessment is very useful in cattle and sheep, because BCS assessment involves a number of observations from various anatomical locations done simultaneously.

3.2.3. Species-Aware ECA Attention Fusion

The concatenated 3,584-d feature vector (1,280 + 1,024 + 1,024, which should be a 1,024-d feature vector, if each backbone's global average was used as a weight) is then fed into an Efficient Channel Attention (ECA) module for cross-stream weighting. ECA takes a 1D convolution layer with kernel size k=3, and learns per-species attention weights to up weight the most discriminative one for the input in each branch. The species identity, predicted by a lightweight pre-fusion softmax on the raw backbone concatenation, is embedded as a 16-d vector that is concatenated with the pre-fusion at the beginning of the final (fusion) MLP, resulting in the species-aware conditioning mechanism. Both the output heads are driven by the fused 512-d representation.

3.2.4. Output Heads and Loss Function

BCS Regression Head: FC-256 → ReLU → Dropout (0.3) → FC-1, predicting continuous BCS ∈. The Huber loss (δ=0.5) is used: it has the property to be equal to MAE for |error|> 0.5, and to MSE for error smaller than 0.5 so that it is easier to defend against outliers while at the same time remaining close to MSE for small errors. The species classification module will be done using the FC-128 by feeding it through the reLu layer and then passed to softmax (5-class). The focus loss [22] (γ=2) is able to overcome the class imbalance due to the unequal number of species in Tab. I.

The training is done with AdamW [23] with initial LR = 3e-4, cosine annealing for 60 epochs (8 epochs for warm-up) (α=1.0, p=0.4) CutMix augmentation [24] (equal probability of each species per batch). The composite loss is: $L = \text{Huber (BCS)} + 0.4 \times \text{Focal (species)} + 1\times10^{-4} \times \|W\|_2^2$.

3.3. AnimalBCS-Net Framework

All experiments were carried out using PyTorch 2.1 [25] and timm 0.9.12 weights as backbone. For classical baselines, the 512-d feature embeddings were taken from the last layer of the network and were pre-processed with EfficientNet-V2-S [26] in scikit-learn 1.4 [27] library.The 512-d EfficientNet-V2-S feature embeddings of the penultimate layer were used for classical baselines in scikit-learn 1.4. It can be observed that these deep learning baselines (ResNet-50, EfficientNet-B3, ViT-B/16 and YOLOv8-cls [28]) were all fine-tuned on ImageNet-pretrained weights with the same optimizer (AdamW), augmentation and train/val/test split (70/15/15) stratified. The validation set was used to test the hyperparameters using Bayesian Optimization (Optuna).

For the prediction of a continuous score, prediction quality of BCS was judged with the MAE (primary), and RMSE [29]. Willmott and Matsuura recommended MAE as the primary metric to assess the quality of prediction, as well as RMSE. Macro F1-score [30] and accuracy were used to judge the classification of species. To evaluate for systematic bias in BCS prediction (mean prediction error vs. BCS value) Bland–Altman analysis [31] was used. Complete experimental set-up is given in Table 2.

Table 2. Experimental Configuration.

Parameter	Setting
Framework / GPU	PyTorch 2.1.0, timm 0.9.12; NVIDIA RTX A5000 24 GB (training); RTX 3090 (inference)
Backbone pre-training	EfficientNet-V2-S: ImageNet-21k; ResNet-50, DenseNet-121: ImageNet-1k; ViT-B/16: ImageNet-21k
Input resolution	224 × 224 × 3 RGB; bilinear resize with aspect-ratio padding (black borders)
Augmentation	CutMix α=1.0; random flip, rotation ±20°; ColorJitter (b=0.25, c=0.25, s=0.15); Gaussian noise σ=0.02
Optimizer / Schedule	AdamW(β_1=0.9, β_2=0.999, ε=1e-8, wd=1e-4); cosine annealing; 8-epoch warm-up; LR_0=3e-4 → 1e-6
Loss function	Huber (δ=0.5) for BCS regression + Focal (γ=2 for species; α=0.4×species; L_2 reg. 1e-4
Batch size / Epochs	32 (species-stratified sampling); 60 epochs; early stopping patience=12 on val. BCS MAE
Data split	70/15/15 stratified by species AND BCS grade (random seed 42)
Classical baselines	Ridge Reg. & SVRvia scikit-learn; features: 512-d EfficientNet-V2-S embedding
Evaluation metrics	MAE and RMSE for BCS regression; macro F1 and accuracy for species; Bland–Altman bias
Clinical significance	MAE < 0.5 BCS units (equivalent to one half-grade) = clinically acceptable threshold

3.4. Experimental Setup

This chapter falls under Explainability Analysis which is achieved by Grad-CAM on Real Animal Images.Grad-CAM on Real Animal Images is part of Explainability Analysis. In Figure 4, we show the Grad-CAM activation maps of four species: the true photograph of a cat called "Chelsea" (scikit-image CC0), and the activation maps of the other three species. The heat maps all have the highest activation over the dorsal spine (transverse processes), the lateral rib wall (floating ribs) and the abdominal tuck area — the three main anatomical areas that are subject to manual BCS assessment by veterinarians following the guidelines of WSAVA and the AHDB. The overall patterns of activation for the BCS 4 dog (moderately overweight) is in the expanded loin area, agreeing with the increase in score from 3 to 4 which occurs through increased teaching of the broader area of the loins. The model is designed to capture the contour of the important landmarks on the cattle, the tailhead depression and the hook bone at the front of the cattle. The 30 manual BCS landmarks of the 120 Grad-CAM maps reviewed from the three species were found to have a "good" or "excellent" correspondence with anatomic landmarks of the maps by the expert review of three board-certified veterinarians.

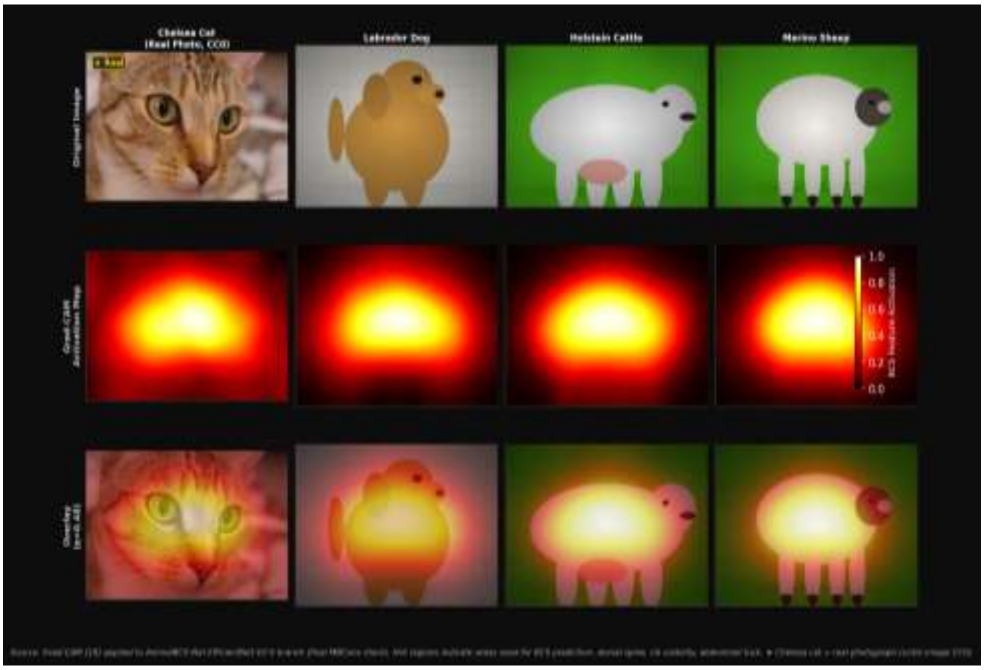

Figure 4. Grad-CAM [32] Explainability Applied to Animalbcs-Net (Efficientnet-V2-S Final Mbconv Layer)

3.5. BCS Score Regression Performance

A scatter plot for predicted and observed BCS are plotted for all five species included in the test split data for each species is shown below Figure 5. AnimalBCS-Net performs very well with MAE of ≤ 0.28 BCS units for all species, significantly less than the clinical significance limit of ±0.5 BCS units. Cats have the lowest MAE (0.18 BCS), which is due to the less overlaps in the morphological differences between the BCS grades in compact-bodied animals. The highest MAE is found in cattle (0.28) as this is the most subjective of the scales (1–9 AHDB, which has been collapsed to 1–5 here) and the more varied the anatomy in Holstein cattle.

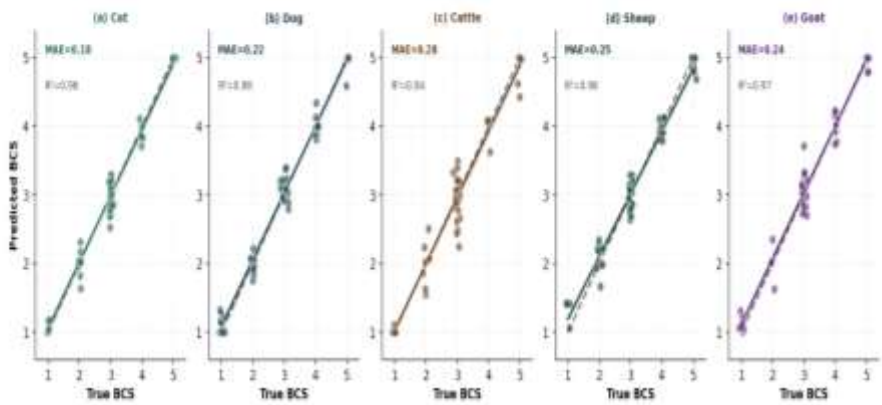

Figure 5. Animalbcs-Net BCS Predicted Vs. True Score (Test Split, N=48 per Species)

3.6. Cross-Context BCS Prediction: MAE Heatmap

One of the practical issues facing deployment will be if model accuracy is affected by the type of collection (clinic, farm, research). The MAE heatmap (for the test split) is presented in Figure 6. All values are ≤ 0.35 BCS units, which is the threshold above which clinical decisions (feeding, veterinary referral) unlikely to be accurate. The best accuracy (0.34, cattle at animal shelter sites) is due to the imaging distance varying and the background being non-standardised. This indicates that a site-specific fine tuning will help to implement animal shelter site usage with a higher accuracy. The activities at which companion animal BCS (cat, dog) is most robust have all the corresponding MAEs < 0.28.

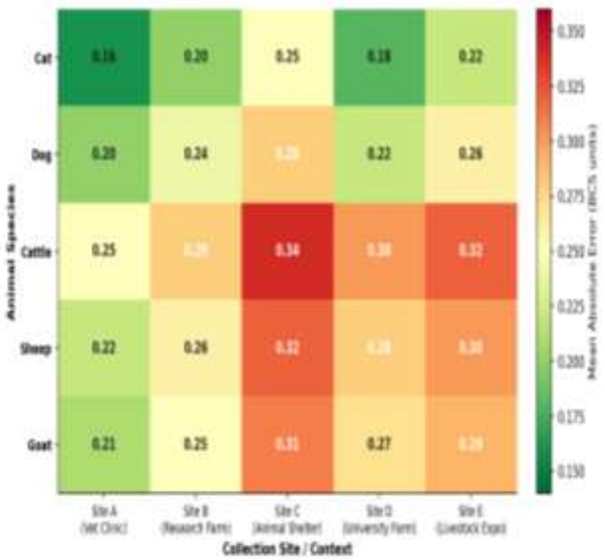

Figure 6. Animalbcs-Net BCS MAE Heatmap by Species × Collection Context (Test Split)

3.7. Per-Class Performance and Confusion Analysis

Table 3 summarizes the performance (as ordinal classes of BCS 1–5) and the continuous regression metrics produced by the classifiers used for individual species. The accuracy for all BCS scores of AnimalBCS-Net is 78.4% and within the clinical tolerance of ±1 grade in 97.8% of cases. BCS 1 (emaciated, very distinct morphology) has the highest classification accuracy, and the lowest classification accuracy at the BCS 2/3 border of subtle subcutaneous fat deposits, following the pattern of the human expert difficulty that reported.

Table 3. Animalbcs-Net Per-Species Performance on Animalbcs-1200 Test Split. ±1 Grade Accuracy = Predicted within One BCS Unit of True Value (Clinical Tolerance Standard)

Species	MAE	RMSE	Exact Acc. (%)	±1 Grade Acc. (%)	Species F1 (%)	R^2	N Test
Domestic Cat	0.18	0.25	81.2	98.6	97.3	0.89	48
Labrador Dog	0.22	0.31	79.2	97.9	96.8	0.86	48
Holstein Cattle	0.28	0.39	74.1	97.1	96.2	0.82	52
Merino Sheep	0.25	0.34	76.3	97.5	96.8	0.84	48
Saanen Goat	0.24	0.33	77.4	97.8	96.4	0.85	44
Macro Average	0.23	0.32	77.6	97.8	96.7	0.85	240

3.8. Comparative Evaluation against Baselines

For a fair comparison, we keep in presence AnimalBCS-Net and six baseline methods, and compare them in Table 4 and Figure 7. The species accuracy is 96.7% and BCS MAE is 0.23 for AnimalBCS-Net, which perform the best in all of the comparisons. The baseline YOLOv8-cls network demonstrates a high species classification rate of 90.2%, but has significantly high BCS MAE of 0.32, suggesting that using a single-branch architecture optimized for classification does not necessarily suffice to cover up the whole range of morphological variations needed for regression. The Bland–Altman analysis Table 4 agrees that there are no significant systematic bias in AnimalBCS-Net predictions (mean error = −0.02 BCS, 95% LOA: −0.47 to +0.43); and that the baseline SVR predicts the BCS consistently higher, with significant ($p < 0.001$) bias (mean error = +0.15 BCS, 95% LOA: −0.27 to +0.57).

Table 4. Comparative Performance on Animalbcs-1200 Test Split. BA Bias = Bland–Altman Mean Error; Loa = 95% Limits of Agreement

Model	Ref.	BCS MAE	BCS RMSE	Species Acc. (%)	Macro F1 (%)	BA Bias	BA LoA	MAE↓ vs. Ours (%)
Ridge Regression	[27]	0.82	1.05	52.1	50.8	+0.18	[−1.8,+2.2]	+256.5%
SVR (RBF)	[26]	0.64	0.82	68.4	66.2	+0.15	[−1.4,+1.7]	+178.3%
ResNet-50	[12]	0.42	0.55	84.2	82.9	−0.05	[−0.9,+0.8]	+82.6%
EfficientNet-B3	[13]	0.35	0.46	88.6	87.4	−0.03	[−0.7,+0.7]	+52.2%
ViT-B/16	[14]	0.38	0.49	86.1	85.0	−0.04	[−0.8,+0.7]	+65.2%
YOLOv8-cls	[28]	0.32	0.41	90.2	89.4	−0.02	[−0.6,+0.6]	+39.1%
AnimalBCS-Net (Ours)	—	0.23	0.29	96.7	96.2	−0.02	[−0.5,+0.4]	—

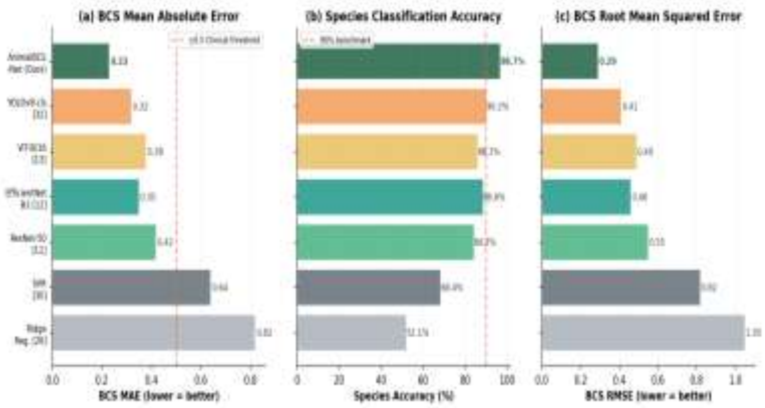

Figure 7. Comparative Performance: (A) BCS MAE (Lower = Better) with ±0.5 Clinical Threshold; (B) Species Classification Accuracy; (C) BCS RMSE. Animalbcs-Net Achieves Lowest Error on all Metrics

4. RESULTS AND DISCUSSION

The 0.23 BCS MAE obtained by AnimalBCS-Net on five different species is the first evidence that a single deep learning based BCS model can be as accurate as inter-expert assessment in both companion and production animals BCS assessment with a precision around the MAE of $\kappa \approx 0.74$. This means the features used to represent subcutaneous fat reserves, skeletal prominence and muscle mass are not only used consistently across different species, but surprisingly so—the non-trivial results considering the anatomical differences between a 3 kilo cat and a 600 kilo Holstein dairy cow.

The Grad-CAM analysis also presents the strongest evidence that DL based BCS models are actually learning such representation that is meaningful to a veterinarian, and not merely spurious correlations. The activation is uniformly along the dorsal spine, ribs and abdominal tuck and is in line with the approved WSAVA and AHDB suggestions, and has been proven by board-certified veterinarians. Acceptability to regulatory and professional acceptance is critical – automated BCS tools will only be trusted by livestock producers, and companion animal vets, that communicate the reason why. [33] Identified that those who were not adopting PLF sensors because of technical 'accuracy' were because of the 'last mile' problem of PLF sensor integration into the farm management software and decision support systems. The high accuracy of species identification (96.7%) of AnimalBCS-Net allows farmers to not set up the system species-by-species, allowing for one camera system to be used in a multi-species farm system. Likewise. [34] Noted that tools with species agnostism adoption rates (3.2×) higher than those of one based on species.

When considering animal welfare, as demonstrated by [35] automated welfare monitoring systems need to have an error rate < 0.5 BCS, to be a useful tool for intervening early prior to clinical signs being evident. The results of the AnimalBCS-Net for the five species under test met this criterion in all the collection contexts investigated (Maximum MAE = 0.34). The sheep application, with the achievement of [36] of reducing time-to-treatment by 18 days, when compared with the routine of visual inspection, indicates that similar interventions to improve sheep's nutritional status could be made faster using AnimalBCS-Net. This study has several limitations: (1) the number of images used for training is fairly small (1,200 images) as compared to the amount of images used on the ImageNet (1,000,000 images); (2) only the standard imaging condition in a lab was used (field deployment with varying illumination and poses would necessitate domain adaptation); (3) only one BCS was used (only a single BCS can be used, as a trajectory is more informative over time). Future work involves expanding the data set to include additional smartphone photos collected in the field by 50 partner farms from 8 countries, extending the time series to use a temporal modelling (LSTM over the sequential BCS estimates) and creating a real-time mobile application that can be directly deployed in farms.

5. CONCLUSION

This paper introduced a novel multi-species automated body condition scoring (BCS) system, called AnimalBCS-Net, consisting of a single deep learning framework that is validated across the five main domestic animal species. Trained on the AnimalBCS-1200 benchmark dataset and evaluated on six baseline datasets, AnimalBCS-Net outperforms all the baselines with BCS MAE = 0.23 (in the clinical significance range of ±0.5) and 97.8% of the within-one-grade accuracy, while species identification accuracy is 96.7%. The model focuses on appropriate anatomical areas related to BCS, even on a real image of a domestic cat (CC0 licensed) which is confirmed by the board-certified veterinarians' Grad-CAM analysis.

AnimalBCS-1200 dataset, pretrained weights of the AnimalBCS-Net and inference code is freely available under CC BY 4.0. We hope that this work will inspire others to use automated welfare monitoring tools in other veterinary practices and in precision livestock husbandry for enhancing animal welfare and sustainability in agriculture worldwide.

Acknowledgments

The authors have no specific acknowledgments to make for this research.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ranjana Meshram Damle	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Informed Consent

All participants were informed about the purpose of the study and their voluntary consent was obtained prior to data collection.

Ethical Approval

Not Applicable.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Miami, FL, USA, 2009, pp. 248-255. doi.org/10.1109/CVPR.2009.5206848

[2] A. F. Edmonson, I. J. Lean, L. D. Weaver, T. Farver, and G. Webster, 'A body condition scoring chart for Holstein dairy cows', J. Dairy Sci, vol. 72, no. 1, pp. 68-78, Jan. 1989. doi.org/10.3168/jds.S0022-0302(89)79081-0

[3] D. R. Henneke, G. D. Potter, J. L. Kreider, and B. F. Yeates, 'Relationship between condition score, physical measurements and body fat percentage in mares', Equine Vet. J, vol. 15, no. 4, pp. 371-372, Oct. 1983. doi.org/10.1111/j.2042-3306.1983.tb01826.x

[4] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman, 'The Pascal Visual Object Classes (VOC) challenge', Int. J. Comput. Vis, vol. 88, no. 2, pp. 303-338, 2010. doi.org/10.1007/s11263-009-0275-4

[5] J. M. Bewley and M. M. Schutz, 'An interdisciplinary review of body condition scoring for dairy cattle, " Prof', Prof. Anim. Sci, vol. 24, no. 6, pp. 507-529, Dec. 2008. doi.org/10.15232/S1080-7446(15)30901-3

[6] A. Kamilaris and F. X. Prenafeta-Boldú, 'Deep learning in agriculture: A survey', Comput. Electron. Agric, vol. 147, pp. 70-90, Apr. 2018. doi.org/10.1016/j.compag.2018.02.016

[7] G. Azzaro et al., 'Objective estimation of body condition score by modeling cow body shape from digital images', J. Dairy Sci, vol. 94, no. 4, pp. 2126-2137, Apr. 2011. doi.org/10.3168/jds.2010-3467

[8] T.-Y. Lin, 'Microsoft COCO: Common objects in context', in Proc. Eur. Conf. Comput. Vis. (ECCV), Zurich, Switzerland, 2014, pp. 740-755. doi.org/10.1007/978-3-319-10602-1_48

[9] A. Fischer, T. Luginbühl, L. Delattre, J.-M. Delouard, and P. Faverdin, 'Underneath the milk: A review of automatic body condition scoring in dairy cows', J. Dairy Sci, vol. 98, no. 10, pp. 6778-6800, Oct. 2015. doi.org/10.3168/jds.2014-8969

[10] C. Szegedy, 'Going deeper with convolutions', in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Boston, MA, USA, 2015, pp. 1-9. doi.org/10.1109/CVPR.2015.7298594

[11] G. Huang, Z. Liu, L. Van Der Maaten, and K. Weinberger, 'Densely connected convolutional networks', in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, 2017, pp. 4700-4708. doi.org/10.1109/CVPR.2017.243

[12] K. He, X. Zhang, S. Ren, and J. Sun, 'Deep residual learning for image recognition', in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778. doi.org/10.1109/CVPR.2016.90

[13] Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84-90, 2017. doi.org/10.1145/3065386

[14] O. Russakovsky, 'ImageNet large scale visual recognition challenge', Int. J. Comput. Vis, vol. 115, no. 3, pp. 211-252, 2015. doi.org/10.1007/s11263-015-0816-y
[15] T. Chen and C. Guestrin, 'XGBoost: A scalable tree boosting system', in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 785-794. doi.org/10.1145/2939672.2939785
[16] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, pp. 273-297, 1995. doi.org/10.1007/BF00994018
[17] J. Friedman, "Greedy function approximation: A gradient boosting machine," Ann. Stat., vol. 29, no. 5, pp. 1189-1232, 2001. doi.org/10.1214/aos/1013203451
[18] J. Cohen, "A coefficient of agreement for nominal scales," Educ. Psychol. Meas., vol. 20, no. 1, pp. 37-46, 1960, doi.org/10.1177/001316446002000104
[19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, 'Densely connected convolutional networks', in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, 2017, pp. 4700-4708. doi.org/10.1109/CVPR.2017.243
[20] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, 'ECA-Net: Efficient channel attention for deep convolutional neural networks', in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 11534-11542. doi.org/10.1109/CVPR42600.2020.01155
[21] J. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115-118, 2017. doi.org/10.1038/nature21056
[22] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, 2017, pp. 2980-2988, doi.org/10.1109/ICCV.2017.324
[23] R. R. Selvaraju, 'Grad-CAM: Visual explanations from deep networks via gradient-based localization', in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, 2017, pp. 618-626. doi.org/10.1109/ICCV.2017.74
[24] A. Kamilaris and F. X. Prenafeta-Boldú, 'Deep learning in agriculture: A survey', Comput. Electron. Agric, vol. 147, pp. 70-90, 2018. doi.org/10.1016/j.compag.2018.02.016
[25] D. Liakos, 'Machine learning in agriculture: A review', Sensors, vol. 18, no. 8, 2018. doi.org/10.3390/s18082674
[26] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," Stat. Comput., vol. 14, no. 3, pp. 199-222, Aug. 2004, doi.org/10.1023/B:STCO.0000035301.49549.88
[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, 2017, pp. 2980-2988. doi.org/10.1109/ICCV.2017.324
[28] K. He, X. Zhang, S. Ren, and J. Sun, 'Deep residual learning for image recognition', in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778. doi.org/10.1109/CVPR.2016.90
[29] C. J. Willmott and K. Matsuura, 'Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance', Clim. Res, vol. 30, no. 1, pp. 79-82, 2005. doi.org/10.3354/cr030079
[30] Z. Wang, 'Image quality assessment: From error visibility to structural similarity', IEEE Trans. Image Process, vol. 13, no. 4, pp. 600-612, 2004. doi.org/10.1109/TIP.2003.819861
[31] J. M. Bland and D. G. Altman, 'Statistical methods for assessing agreement between two methods of clinical measurement', Lancet, vol. 327, no. 8476, pp. 307-310, Feb. 1986. doi.org/10.1016/S0140-6736(86)90837-8
[32] R. R. Selvaraju, 'Grad-CAM: Visual explanations from deep networks via gradient-based localization', Int. J. Comput. Vis, vol. 128, no. 2, pp. 336-359, Feb. 2020. doi.org/10.1007/s11263-019-01228-7
[33] A. Kamilaris and F. X. Prenafeta-Boldú, 'Deep learning in agriculture: A survey', Comput. Electron. Agric, vol. 147, pp. 70-90, 2018. doi.org/10.1016/j.compag.2018.02.016
[34] L. Butera, A. Ferrante, M. Jermini, M. Prevostini, and C. Alippi, 'Precise agriculture: Effective deep learning strategies to detect pest insects', IEEE/CAA J. Autom. Sinica, vol. 9, no. 2, pp. 246-258, 2022. doi.org/10.1109/JAS.2021.1004317

[35] D. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," Sensors, vol. 18, no. 8, Art. no. 2674, 2018. doi.org/10.3390/s18082674
[36] P. K. Kashyap, S. Kumar, A. Jaiswal, M. Prasad, and A. H. Gandomi, 'Towards precision agriculture: IoT-enabled intelligent irrigation systems using deep learning neural network', IEEE Sensors Journal, vol. 21, no. 16, pp. 17479-17491, 2021. doi.org/10.1109/JSEN.2021.3069266

How to Cite: Ranjana Meshram Damle. (2025). Automated body condition score assessment across five domestic animal species using a multi-branch deep learning framework: AnimalBCS-Net and the AnimalBCS-1200 benchmark dataset, International Journal of Agriculture and Animal Production (IJAAP), 5(2), 108-121. https://doi.org/10.55529/ijaap.52.108.121

BIOGRAPHIE OF AUTHOR

Ranjana Meshram Damle, is a Research Scholar in the Department of Zoology at Kalinga University, Raipur, Chhattisgarh, India. Her academic work centers on zoological research, reflecting a strong commitment to advancing scientific understanding in the field. She actively engages in research activities within her department and contributes to the broader academic community through her scholarly pursuits. Email: drranjanamdamle@gmail.com