p-ISSN 2615-787X e-ISSN 2615-790X
Accredited by Directorate General of Higher Education, Research, and Technology, Republic of Indonesia, No. 225/E/KPT/2022

Tropical Animal Science Journal, September 2025, 48(5):402-411
DOI: https://doi.org/10.5398/tasj.2025.48.5.402
Available online at https://journal.ipb.ac.id/index.php/tasj

Comparative Performance Analysis of YOLOv10-Based Models with CBAM and SPPFCSPC for Body Condition Score Assessment in Beef Cattle

F. Ariadi[a], F. Utaminingrum[a,*], & B. A. Atmoko[b]
[a]Faculty of Computer Science, Brawijaya University
Jl. Veteran, Kampus UB 65145, Indonesia
[b]Research Center for Animal Husbandry, National Research and Innovation Agency (BRIN)
Jl. Raya Jakarta-Bogor, Bogor-16915, Indonesia
*Corresponding author: f3.utaminingrum@ub.ac.id
(Received 26-03-2025; Revised 09-07-2025; Accepted 10-07-2025)

ABSTRACT

Body condition score assessment serves as a critical metric for evaluating the health, nutritional status, and overall well-being of beef cattle, playing a pivotal role in herd management and productivity optimization. Traditional manual BCS assessment methods are inherently subjective, labor-intensive, and impractical for large-scale operations, thereby necessitating an automated and data-driven approach. This study investigates the performance of YOLOv10-based deep learning models, incorporating the convolutional block attention module (CBAM) and spatial pyramid pooling-fast cross-stage partial connections (SPPFCSPC) to enhance feature extraction, classification accuracy, and computational efficiency in BCS estimation. A total of 432 annotated images representing five BCS categories (1–5) were used for model training and evaluation. The models were assessed using precision, recall, and F1 Score, with expert-labeled ground truth ensuring robustness. Results show that the YOLOv10x variant achieved the highest classification accuracy of 88.2%, highlighting its superior detection capability. YOLOv10m exhibited a balanced trade-off between accuracy and computational efficiency, achieving an F1 Score of 79.2%. The integration of CBAM improved precision but slightly reduced recall, whereas SPPFCSPC enhanced recall at the expense of increased computational complexity. Notably, YOLOv10n achieved the fastest inference time of 1.0 ms but with a lower accuracy of 82.4%, underscoring the trade-off between model depth and real-time applicability. These findings validate the effectiveness of attention-based and multi-scale feature learning strategies for improving the automation of BCS classification in beef cattle.

Keywords: cattle assessment; BCS; CBAM; roboflow; SPPFCSPC; YOLO

INTRODUCTION

The beef cattle industry is a fundamental pillar of Indonesia's agricultural sector, playing a crucial role in national food security and economic stability. However, the sector faces persistent challenges, including a declining cattle population and inefficiencies in livestock management. Between 2021 and 2022, Indonesia's beef cattle population decreased by 374,676 heads, from 17,977,214 to 17,602,538, due to factors such as low reproductive efficiency, aging breeder stocks, and post-pandemic disruptions in livestock supply chains. This trend highlights an urgent need for innovative strategies to enhance productivity and ensure sustainable growth. In tropical livestock production systems, where heat stress and feed variability are common, continuous monitoring of BCS is essential for early intervention. An automated BCS assessment system enables timely nutritional and reproductive decisions, particularly in resource-limited smallholder contexts where trained personnel are often unavailable. Previous studies have demonstrated that digital image-based approaches can provide objective body condition score (BCS) estimates, with cow body shape modelling from 2D images successfully used to predict BCS in dairy cattle (Azzaro *et al*., 2011).

To address these challenges, smart and precision farming has emerged as a transformative approach, leveraging automation, artificial intelligence (AI), and data-driven methodologies to improve decision-making in livestock management. A critical component of this approach is the BCS assessment, which serves as a key indicator of cattle health, nutrition, and productivity. Monitoring BCS ensures that cattle remain in optimal condition throughout growth cycles and maintenance targets. It also enables rapid responses to emerging health or productivity issues, helping to correct deficiencies related to production, health, and reproduction. Proper BCS management is essential for preventing calving difficulties, as extreme weight loss during early lactation may lead to metabolic disorders (Spoliansky *et al*., 2016). Additionally, overweight cattle are prone to ketosis, mastitis, and infertility, which are closely linked to reduced fertility rates. BCS is an indicator of energy reserves, typically evaluated on a five-point scale, where 1 represents emaciated cattle and 5 indicates obesity. CS) serves as an indicator of energy reserves and has been shown to correlate with metabolic health and insulin sensitivity in dairy cattle (Ghaffari *et al*., 2023).

Despite its importance, conventional BCS assessment remains highly subjective, labor-intensive, and time-consuming, limiting its scalability for large-scale cattle operations. Traditionally, trained experts perform manual BCS evaluations using either palpation techniques (assessing fat deposits and bone structure) or visual inspection (observing body shape). However, manual BCS assessment requires experienced personnel, is time-consuming, and relies on subjective estimations. The development of AI- and IoT-based BCS assessment systems offers a promising solution, enabling accurate, objective, and fully automated BCS evaluation without human intervention (Long *et al*., 2022).

Advancements in computer vision and deep learning offer promising solutions to automate and enhance BCS evaluation. Studies utilizing depth sensors and AI-driven models have demonstrated strong correlations between projected body volume and actual body weight, achieving mean absolute error (MAE) below 20 kg in weight estimation (Xiong *et al*., 2023). However, most existing approaches still suffer from high computational demands and suboptimal real-time performance, necessitating the development of more efficient AI models for practical on-farm applications.

Among AI methodologies, object detection models such as You Only Look Once (YOLO) have demonstrated exceptional capabilities in real-time applications. The latest iteration, YOLOv10, introduces NMS-free training and efficiency-accuracy-driven design, significantly reducing computational overhead and latency compared to previous versions (Wang *et al*., 2024). These characteristics make YOLOv10 a highly promising foundation for automated BCS detection in the Indonesian beef cattle industry. Recent work on YOLOv10 has focused on enhancing inference speed and model efficiency. For instance, YOLOv10-S has been shown to be 1.8 times faster than RT-DETR-R18 at similar accuracy levels on the COCO dataset, with 2.8 times fewer parameters and floating-point operations per second (FLOPs). Compared to YOLOv9-C, YOLOv10-B reduces latency by 46% while requiring 25% fewer parameters (Wang *et al*., 2024). While these advancements improve computational efficiency, further enhancements in feature extraction and spatial attention mechanisms are necessary to optimize BCS detection accuracy and robustness.

Therefore, addressing the accuracy efficiency trade-offs in existing AI-based BCS models remains critical for practical deployment, particularly in smallholder tropical cattle systems. This study proposes an enhanced YOLOv10-based BCS detection framework by integrating the convolutional block attention module (CBAM) and spatial pyramid pooling-fast cross-stage partial connections (SPPFCSPC) to improve feature extraction and computational performance. In summary, this research aims to develop a computationally efficient and accurate BCS detection model based on YOLOv10 and to evaluate its performance using mean average precision (mAP), precision, recall, and F1 Score.

MATERIALS AND METHODS

Research Implementation

This study was conducted in December 2024 at the Beef Cattle Pen, Livestock Development Center, Faculty of Animal Science, Universitas Gadjah Mada, Yogyakarta, Indonesia. The research aimed to develop an automated BCS assessment system by employing a YOLOv10-based deep learning architecture enhanced with CBAM and SPPFCSPC.

Imaging Equipment and IT Infrastructure

For image acquisition, a Hikvision 4MP AI CCTV camera was used, offering high-resolution video streaming with built-in artificial intelligence (AI) capabilities to enhance object detection. The camera was strategically mounted at a fixed height and angle to ensure consistent image capture for all cattle. Lighting conditions were standardized to minimize variations in image quality, following imaging protocols outlined by Xiong *et al.* (2023) and Spoliansky *et al.* (2016) to ensure consistent and replicable visual data for livestock monitoring.

Data processing and model training were conducted using a high-performance computing system, equipped with an Intel Core i9-12900K processor, 64GB RAM, and an NVIDIA RTX 3090 GPU (24GB VRAM). This setup facilitated efficient data handling and deep learning computations.

Livestock Dataset

The dataset comprised 225 high-resolution images of beef cattle, systematically collected through controlled sampling sessions at the Faculty of Animal Science, Universitas Gadjah Mada. As shown in Table 1, the data acquisition process was conducted across three independent sessions, with each session involving 15 cattle, and five images captured per individual, resulting in 75 images per session and a total of 225 images. To ensure consistency and minimize variability, images were captured from a standardized single-angle perspective, providing uniform visual representations for subsequent analysis.

The study involved a total of 15 female beef cattle, comprising 12 Peranakan Ongole (PO), 2 Simmental, and 1 Brahman Cross. These animals were purposefully selected to represent the genetic and physiological diversity commonly found in Indonesian beef cattle production systems. The exclusive use of females also aligns with the practical context of BCS monitoring, as body condition is particularly critical in managing reproduction and lactation performance in female cattle.

The sampled cattle exhibited heterogeneous physiological characteristics, encompassing variations

Table 1. Session-based image data collection from 15 female beef cattle (12 PO, 2 Simmental, 1 Brahman Cross)

Session	Number of cattle	Breed composition	Images cattle	Total images cattle
Session 1	15	12 PO, 2 Simmental, 1 Brahman Cross	5	75
Session 2	15	Same set of cattle (repeated measures)	5	75
Session 3	15	Same set of cattle (repeated measures)	5	75
Total	15	12 PO, 2 Simmental, 1 Brahman Cross		225

in sex (male and female), age groups (young and mature), and reproductive status (pregnant, lactating, or non-lactating). The structured multi-session data acquisition approach facilitated a diverse representation of body conditions while maintaining environmental consistency, thereby enhancing dataset robustness for deep learning-based BCS estimation (Patterson *et al*., 2021). BCS assessment can be conducted using visual indicators alone or a combination of visual and tactile estimations of key skeletal structures to evaluate fat cover. Key anatomical regions for BCS evaluation include the backbone, hips, tailhead, pins, rump, thigh, short ribs, and long ribs, as shown in Figure 1.

These regions are critical visual and tactile points for determining fat distribution and overall body condition. Visual assessment evaluates fat thickness over the ribs, tailhead, and backbone, providing essential cues about the cattle's nutritional and health status. Exclusion criteria involved images with motion blur, partial visibility, or obstructed views that could affect accurate annotation. Figure 2 labels each image based on BCS categories (1–5) by two independent expert evaluators from the Faculty of Animal Science, Universitas Gadjah Mada, with more than five years of experience in manual BCS scoring. Inter-annotator agreement was verified, and any discrepancies were resolved through discussion to ensure biologically valid ground truth. While the dataset consisted of over 400 annotated images, these were derived from 15 individual cattle. The limited number of unique animals is acknowledged as a constraint regarding biological variability and population representation.

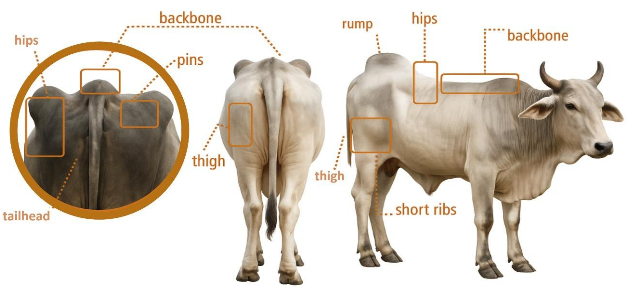

Figure 1. Key anatomical landmarks for body condition score (BCS) assessment in beef cattle. Primary visual zones depicted in this diagram include the tailhead, hips, thigh, rump, pins, short ribs, and backbone. This illustration reflects the typical morphology of beef breeds used in this study, adapted from Patterson *et al.* (2021).

Data Preprocessing and Augmentation

Captured images were preprocessed by resizing them to 640 × 640 pixels to match the input requirements of the YOLO model. From Table 2, techniques improved the model's robustness to lighting, orientation, and background conditions variations. The YOLOv10 model was trained using supervised learning with 432 labeled images. The model was optimized using the Adam optimizer, with a learning rate of 0.001 and a batch size of 16, trained over 100 epochs on an NVIDIA RTX 3090 GPU to accelerate computation. Adam was selected due to its adaptive learning rate and faster convergence properties compared to traditional optimizers such as SGD or RMSprop, making it particularly effective for training deep neural networks on relatively small and heterogeneous datasets. To increase data diversity and model robustness, the original 225 images were augmented using standard techniques (rotation, brightness, blur, flipping, etc.), resulting in 432 labeled images used for model training.

Architecture Model

An enhanced YOLOv10n architecture was developed by integrating CBAM and SPPFCSPC to improve feature extraction, detection accuracy, and computational efficiency for automated BCS assessment. As illustrated in Figure 3, the model consists of three primary components: Backbone, Neck, and Head, each

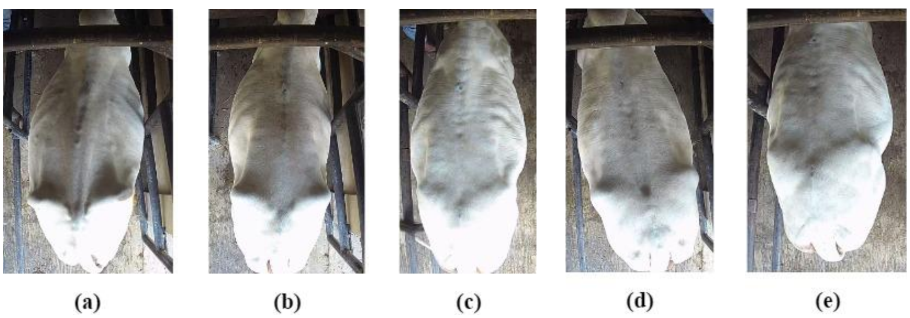

Figure 2. High-resolution images of beef cattle classified by body condition score (BCS) ranging from 1 to 5. (a) BCS 1, (b) BCS 2, (c) BCS 3, (d) BCS 4, and (e) BCS 5.

Table 2. Training hyperparameters for YOLOv10-based BCS classification

Hyperparameter	Value
Learning rate	0,01
Batch size	16
Number of epochs	100
Optimizer	Adam
Input image size	640x640 pixels
Data augmentation	Blur, Brightness, Saturation, Rotation, 90 Rotate, Flip

modified to optimize multi-scale feature representation and object detection performance.

The backbone is designed for feature extraction, incorporating convolutional layers, CSP blocks, SCDown, and depthwise separable convolutions to enhance computational efficiency. In the YOLOv10n + SPPFCSPC configuration (Figure 3c), SPPFCSPC is added to refine multi-scale feature aggregation further, improving context-aware encoding critical for accurate BCS classification. The neck is responsible for feature refinement and fusion, ensuring that spatial and contextual information is effectively utilized. In the YOLOv10n + CBAM model (Figure 3b), CBAM is integrated into the neck to enhance channel and spatial attention, improving detection precision by emphasizing fine-grained fat distribution differences.

The head performs final BCS classification and object detection using a dual-branch prediction mechanism. The One-to-Many Prediction Head evaluates multiple bounding boxes, increasing robustness in complex scenes, while the One-to-One Prediction Head ensures accurate and efficient classification with minimal redundancy. These modifications collectively enhance the accuracy, computational efficiency, and real-time applicability of BCS assessment, making the model well-suited for scalable, precision livestock management.

In this study, nine YOLOv10-based model configurations were developed and evaluated to explore the impact of attention mechanisms and pooling strategies across different architecture scales. These configurations consist of three baseline models: YOLOv10n, YOLOv10m, and YOLOv10x, representing small, medium, and large variants, respectively. Each baseline was further enhanced with either the CBAM or the SPPFCSPC, resulting in six additional variants. This structured approach allows for a comprehensive analysis of how model complexity and specific architectural enhancements influence detection accuracy, computational efficiency, and BCS classification performance.

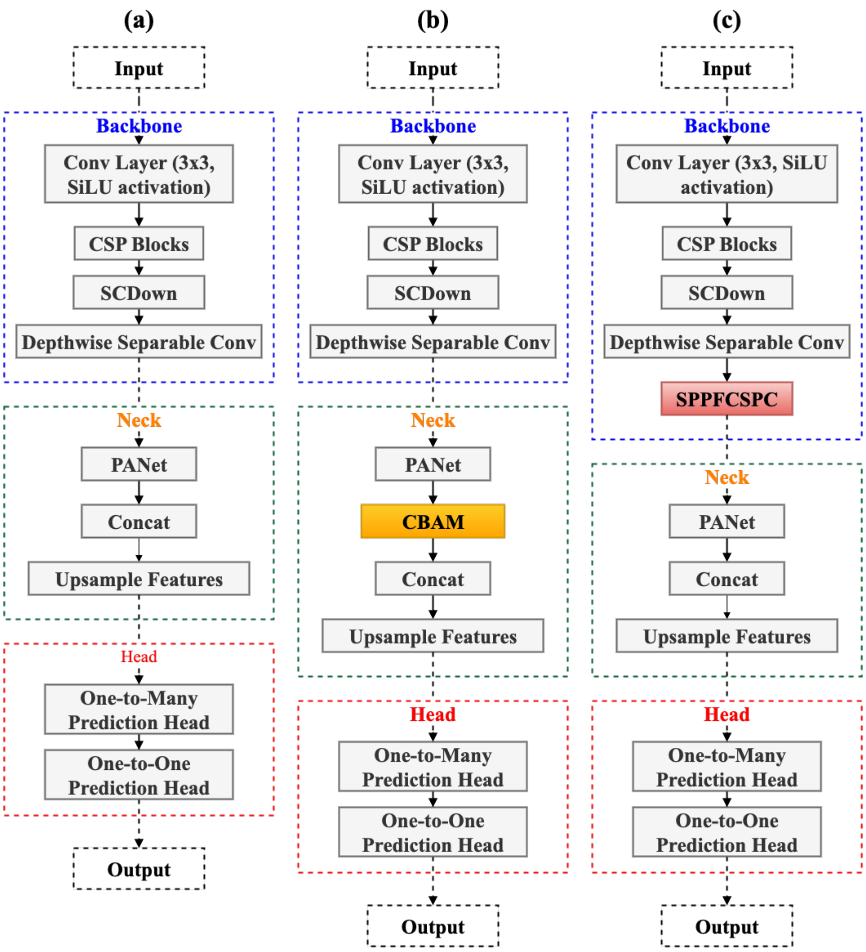

Figure 3. Architectural configurations of YOLOv10-based models evaluated in this study. (a) YOLOv10 (baseline), (b) YOLOv10 with convolutional block attention module (CBAM) integrated into the neck, (c) YOLOv10 with spatial pyramid pooling-fast cross stage partial connections (SPPFCSPC) embedded in the backbone.

Convolutional Block Attention Module

The CBAM is an attention mechanism designed to enhance CNN performance by emphasizing relevant features. CBAM is a lightweight module that integrates seamlessly with CNN architectures. Inspired by various attention mechanisms like squeeze-and-excitation networks (SENet) from 2017, Residual Attention Network from 2017 (Wang *et al*., 2024), and spatial transformer network (STN) from 2015, CBAM contains two sub-modules, the channel attention module and the spatial attention module, sequentially within a single attention block.

In Figure 4, designed to highlight critical features within each channel of a feature map. It takes the feature map as input and computes channel attention through two aggregation operations: average pooling and max pooling. These results are combined and then passed through fully connected layers and a sigmoid activation function to produce a channel attention map, which weighs the original feature map.

SPPFCSPC

The SPPFCSPC module (spatial pyramid pooling fast convolutional special pyramid pooling convolution) draws inspiration from previous modules and techniques aimed at enhancing object detection capabilities, particularly in detecting small targets and accelerating processing. Among these is SPPF, or spatial pyramid pooling fast, an optimized version of SPP designed for speed, introduced in YOLOv5 (Jooshin et al., 2024).

In Figure 5, pooling is performed sequentially across four different window sizes, with each pooling result passed to the subsequent pooling layer. SPPFCSPC incorporates the principles of SPPF, forwarding the pooled outputs to the next layer to retain spatial information and enhance processing speed. By integrating cross stage partial connections (CSPC), the module reduces computational redundancy and optimizes gradient flow, which is crucial for deep network architectures (Wang *et al*., 2023).

The combination of SPPF and CSPC allows for efficient multi-scale feature extraction, ensuring better object detection performance, particularly for small-scale features relevant to BCS assessment. This refinement is particularly beneficial in scenarios where precise localization of anatomical features is required for accurate classification (Nagy *et al*., 2023).

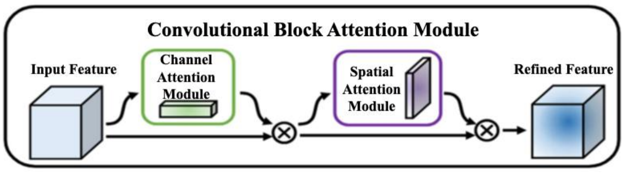

Figure 4. Structure of the convolutional block attention module (CBAM), showing sequential channel and spatial attention modules. Adapted from Woo *et al.* (2018).

Evaluation Metrics

The performance evaluation of each YOLOv10-based model was conducted using four main metrics: Precision, Recall, F1-Score, and Mean Average Precision (mAP). As defined in Equation (1), precision measures the proportion of true positive predictions out of all positive predictions made by the model, reflecting its accuracy in identifying relevant instances. Recall, shown in Equation (2), quantifies the proportion of actual positives that were correctly identified, indicating the model's sensitivity. These two metrics are combined into the F1-Score (Equation 3), which represents the harmonic mean of Precision and Recall and is particularly useful when evaluating imbalanced datasets. Moreover, the classification performance across all classes is summarized using mean average precision (mAP), as shown in Equation (4).

Precision= TP/(TP + FP) (1)

Recall= TP/(TP + FN) (2)

F1-score= 2 · (Precision · Recall)/(Precision + Recall) (3)

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (4)$$

Computational Efficiency Analysis

To evaluate the computational performance of the proposed enhanced YOLOv10 model, three key metrics were analyzed: inference time (ms), number of parameters (million), and floating-point operations per second (FLOPS, GigaFLOPS/G). A lower inference time is crucial for real-time BCS assessment, ensuring efficient large-scale cattle monitoring.

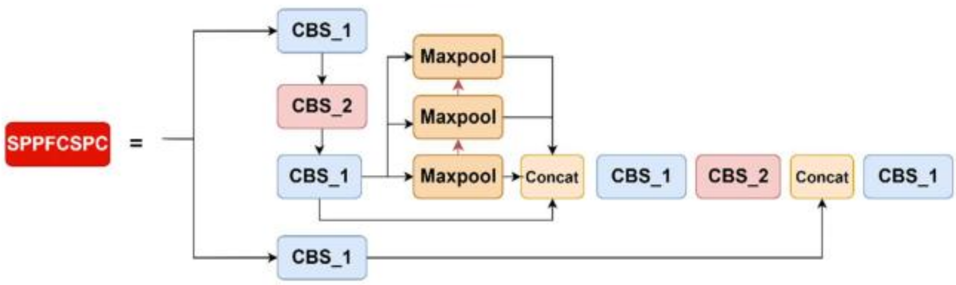

Figure 5. Illustration of the spatial pyramid pooling-fast cross-stage partial connections (SPPFCSPC) module used in the enhanced YOLOv10 architecture. Adapted from Jooshin *et al.* (2024).

While the incorporation of CBAM, and SPPFCSPC slightly increases computational overhead, optimizations such as depthwise separable convolutions and CSPC mitigate excessive latency, maintaining a balance between accuracy and efficiency. Research has shown that cross stage partial networks (CSPNet) effectively reduce redundancy while preserving high feature representation quality, making them suitable for lightweight yet accurate detection models (Wang *et al*., 2023).

The number of parameters and FLOPS indicate model complexity and computational cost, impacting real-time deployment feasibility. Although the proposed model introduces additional layers for enhanced feature extraction, leveraging attention-based mechanisms and efficient pooling strategies prevents excessive growth in computational demand. Studies on transformer-based architectures highlight the importance of optimizing attention mechanisms to maintain high accuracy with minimal processing overhead, ensuring real-time applicability.

RESULTS

Annotated Image Dataset from Roboflow

The dataset utilized in this study was annotated using Roboflow, where each image was manually labeled to identify key regions of interest (ROI), ensuring precise detection and segmentation of cattle bodies. These annotations served as the ground truth for model training and evaluation.

Figure 6 presents examples of annotated images, demonstrating the labeling consistency across different BCS categories (1–5). The high-quality annotations contribute to the model's ability to generalize well across diverse cattle conditions.

Confusion Matrix Results

Figure 7 illustrates the normalized confusion matrices for nine different YOLOv10-based model configurations, each reflecting the model's performance in classifying cattle across BCS categories 1 to 5.

Across the matrices, models such as YOLOv10x (Figure 7g) and YOLOv10n (Figure 7a) display stronger diagonals, suggesting more consistent and accurate predictions. In contrast, configurations involving single-module integrations (e.g., YOLOv10n + CBAM in Figure 7b or YOLOv10m + SPPFCSPC in Figure 7f) demonstrate relatively dispersed predictions, indicating potential confusion between adjacent BCS classes. The visual comparison underscores that deeper models or those with balanced attention and pooling mechanisms tend to yield improved category-specific accuracy, particularly in distinguishing subtle differences in intermediate BCS categories (BCS 2, 3, and 4).

Performance Metrics

A comparative analysis of different YOLOv10-based models for BCS assessment, focusing on accuracy, precision, recall, and F1 score, is presented in Table 3. The results indicate that YOLOv10x outperforms all other models, achieving the highest accuracy (88.2%), precision (88.2%), recall (83.6%), and F1 score (85.8%), demonstrating its superior capability in both classification accuracy and robustness. Among the YOLOv10n variants, the base YOLOv10n model (82.4% Accuracy, 80.9% F1 Score) performs better than YOLOv10n + CBAM (83.8% Accuracy, 69.5% F1 Score) and YOLOv10n + SPPFCSPC (71.2% Accuracy, 72.0% F1 Score).

For YOLOv10m models, YOLOv10m + CBAM (83.7% Accuracy, 78.5% F1 Score) slightly outperforms the base model (79.4% Accuracy, 79.2% F1 Score), while YOLOv10m + SPPFCSPC (75.4% Accuracy, 72.8% F1 Score) exhibits a decline in overall performance. This suggests that SPPFCSPC does not provide substantial performance gains in mid-sized YOLO architectures for BCS classification.

Additionally, YOLOv10x + CBAM (86.5% Accuracy, 79.7% F1 Score) and YOLOv10x + SPPFCSPC (84.0% Accuracy, 82.8% F1 Score) show that SPPFCSPC improves recall (81.7%) compared to CBAM (73.9%), but at the cost of lower overall accuracy. These findings emphasize the trade-off between feature extraction complexity and classification efficiency, where deeper models (YOLOv10x) perform better at multi-scale feature learning than their compact counterparts (YOLOv10n and YOLOv10m).

Computational Performance

The YOLOv10x model demonstrated the highest performance in BCS classification, achieving an overall accuracy of 88.2% and an F1 Score of 85.8%.

Despite this superior accuracy, the model incurred a substantially higher computational cost, which

(a) (b) (c) (d) (e)

Figure 6. Annotation results generated using roboflow for body condition score (BCS) categories. Images (a–e) represent annotated datasets for BCS categories: (a) Score 1, (b) Score 2, (c) Score 3, (d) Score 4, and (e) Score 5. The bounding boxes highlight the key regions of interest identified during the annotation process, providing a structured dataset for training and evaluating the YOLOv10 model.

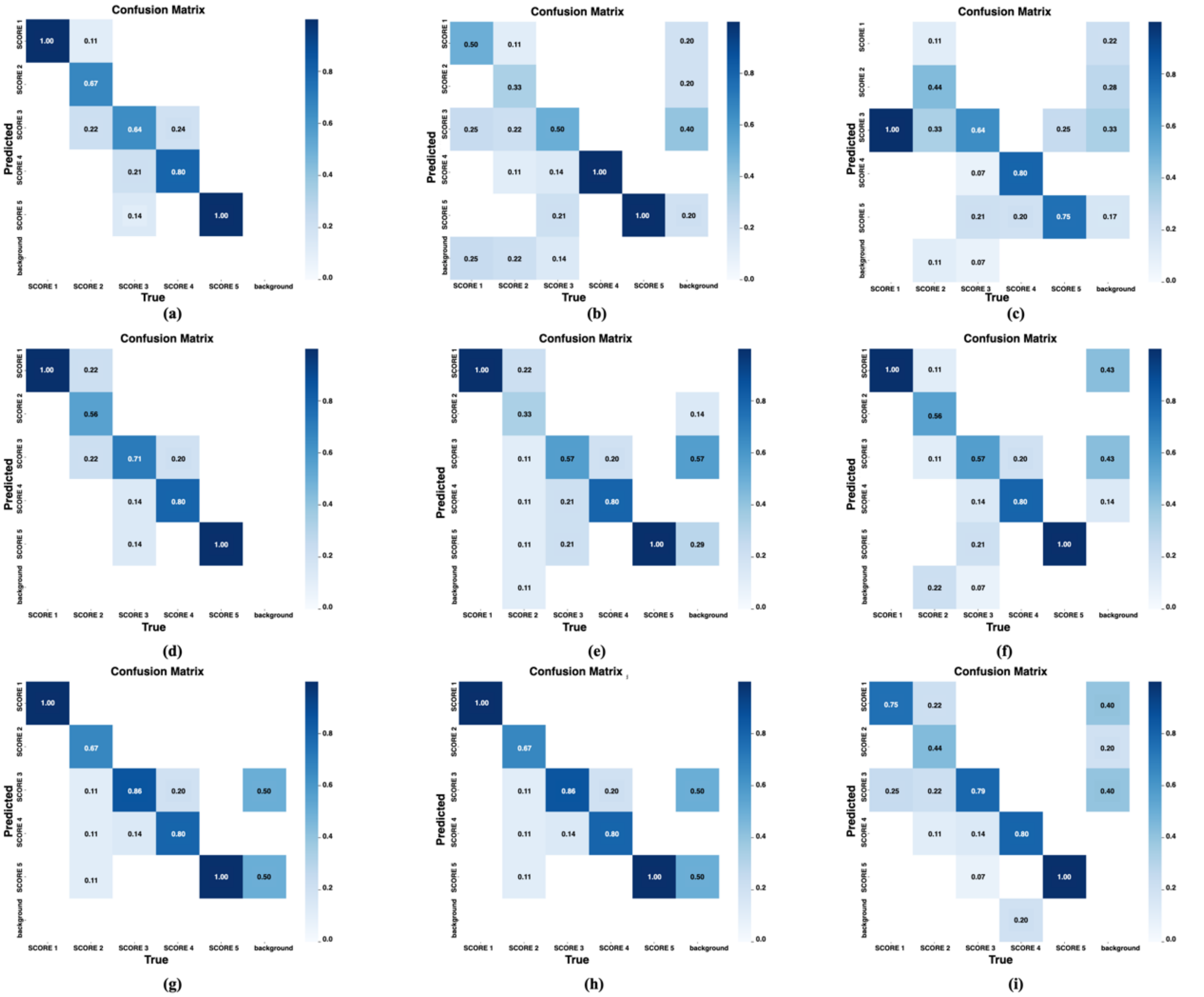

Figure 7. Normalized confusion matrices for the nine YOLOv10-based model configurations used for body condition score (BCS) classification. Each matrix represents classification performance across BCS categories (1–5), with predicted labels on the X-axis and actual labels on the Y-axis. Numerical values denote the normalized proportion of predictions for each class. Configurations shown include: (a) YOLOv10n, (b) YOLOv10n + CBAM, (c) YOLOv10n + SPPFCSPC, (d) YOLOv10m, (e) YOLOv10m + CBAM, (f) YOLOv10m + SPPFCSPC, (g) YOLOv10x, (h) YOLOv10x + CBAM, and (i) YOLOv10x + SPPFCSPC. CBAM= convolutional block attention module; SPPFCSPC= spatial pyramid pooling-fast cross stage partial connections.

Table 3. Performance comparison of YOLOv10-based models for body condition score (BCS) classification using mAP, precision, recall, and F1 score

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
YOLOv10n	82.4	82.4	79.4	80.9
YOLOv10n + CBAM	83.8	83.8	59.5	69.5
YOLOv10n + SPPFCSPC	71.2	71.2	72.7	72.0
YOLOv10m	79.4	79.4	78.9	79.2
YOLOv10m + CBAM	83.7	83.7	73.9	78.5
YOLOv10m + SPPFCSPC	75.4	75.4	70.4	72.8
YOLOv10x	88.2	88.2	83.6	85.8
YOLOv10x + CBAM	86.5	86.5	73.9	79.7
YOLOv10x + SPPFCSPC	84.0	84.0	81.7	82.8

Note: highlights the performance metrics of different models based on Accuracy, Precision, Recall, and F1 Score. mAP= mean average precision; CBAM= convolutional block attention module; SPPFCSPC= spatial pyramid pooling-fast cross stage partial connections.

may limit its applicability in resource-constrained environments. With an inference time of 5.7 ms, 31.59 million parameters, and 169.8 GFLOPS, YOLOv10x is computationally intensive and better suited for high-resource environments such as cloud-based systems or offline batch processing (Table 4).

In contrast, the YOLOv10n + CBAM model delivered a competitive accuracy (83.8%) with an

Table 4. Computational efficiency comparison of YOLOv10-based models for BCS classification based on inference time, parameters, and FLOPS

Model	Computational time (ms)	Number of parameters (Million)	FLOPS (G)
YOLOv10n	1.0	2.70	8.2
YOLOv10n + CBAM	0.9	2.70	8.2
YOLOv10n + SPPFCSPC	1.0	2.94	9.1
YOLOv10m	2.6	16.46	63.4
YOLOv10m + CBAM	3.0	19.81	68.6
YOLOv10m + SPPFCSPC	3.7	23.57	71.6
YOLOv10x	5.7	31.59	169.8
YOLOv10x + CBAM	6.3	31.59	169.8
YOLOv10x + SPPFCSPC	6.7	39.07	175.8

Note: compares the computational efficiency of each model based on Computational Time, Number of Parameters, and FLOPS. CBAM= convolutional block attention module; SPPFCSPC= spatial pyramid pooling-fast cross stage partial connections.

F1 Score of 69.5%, while significantly reducing the computational burden. It maintained a minimal inference time of 0.9 ms, a lightweight architecture with 2.70 million parameters, and only 8.2 GFLOPS—the same as the baseline YOLOv10n. The integration of CBAM refined attention to spatial and channel-wise features, improving precision without sacrificing efficiency.

DISCUSSION

The comparative evaluation of YOLOv10-based models for BCS assessment highlights the impact of CBAM and SPPFCSPC integration on classification performance and computational efficiency. The YOLOv10x model achieved the highest performance, with 88.2% accuracy, 88.2% precision, 83.6% recall, and an F1 Score of 85.8%, confirming its robustness in feature extraction and classification tasks. In contrast, YOLOv10n and YOLOv10m demonstrated moderate accuracy levels (82.4% and 79.4%, respectively), emphasizing the importance of model depth and complexity in improving detection performance. The integration of CBAM improved precision by focusing on relevant spatial and channel features. However, it slightly reduced recall, indicating a trade-off between sensitivity and specificity. Conversely, SPPFCSPC increased recall by enhancing multi-scale feature aggregation, but this came at the cost of reduced overall accuracy in some model configurations, particularly in mid-sized architectures such as YOLOv10m.

From a computational perspective, YOLOv10x exhibited the highest resource demand, with an inference time of 5.7 ms, 31.59 million parameters, 169.8 GFLOPS. This computational complexity renders it less suitable for real-time applications in low-resource environments, such as smallholder farms. Meanwhile, YOLOv10n maintained the lowest computational footprint 1.0 ms inference time, 2.69 million parameters, and 8.2 GFLOPS, making it a viable candidate for edge-based deployment where speed and efficiency are prioritized over absolute accuracy.

The implementation of automated BCS classification models provides practical support for on-farm decision-making. In cow-calf systems, maintaining cows at BCS 3 prior to breeding improves conception success and calf weaning weight. In feedlot operations, BCS monitoring allows for strategic identification of optimal fattening windows, transforming lean, healthy animals (BCS 2) into market-ready cattle (BCS 4 or 5) while minimizing overfeeding. This targeted approach improves feed efficiency, animal welfare, and profitability. Recent work by Li *et al.* (2025) further supports this by demonstrating that an improved YOLOv5-based system can perform real-time video-based BCS estimation across multiple cows simultaneously, balancing accuracy and efficiency for herd-level monitoring in practical farm conditions. Further analysis of the confusion matrices (Figure 7) revealed that the most frequent misclassifications occurred between intermediate BCS categories, particularly BCS 3 and 4. This pattern of misclassification is consistent across multiple model configurations, particularly those with higher sensitivity but reduced specificity. The difficulty in distinguishing BCS 3 from BCS 4 can be attributed to the subtle visual differences in fat distribution, especially around the tailhead and short ribs, which are often challenging to detect even by trained experts. In several models, such as YOLOv10n + CBAM and YOLOv10m + SPPFCSPC, BCS 3 was frequently overpredicted as BCS 4. This suggests that while attention and pooling modules enhance feature representation, further improvement in annotation consistency or region-focused learning may be necessary to reduce confusion between adjacent BCS categories. Although the model achieved robust classification performance, the limited number of individual animals (n= 15) may constrain its generalizability. Future research should include more diverse cattle populations to validate the model across different breeds, ages, and physiological conditions. In the present investigation, the YOLOv10x model integrated with the CBAM achieved the highest classification accuracy of 88.2% and an F1 Score of 85.8% (Table 3). Similarly, Jooshin *et al.* (2024) showed that variants of spatial pyramid pooling, including SPPF and SPPFCSPC, significantly enhanced the aggregation of multi-scale features in livestock imaging tasks. He *et al.* (2023) further validated the role of attention modules in their YOLOX-based pig BCS detection model, achieving a mAP of 80.06%. Likewise, He *et al*. (2023) proposed an improved YOLOX-based architecture incorporating attention mechanisms for automated body condition scoring in pigs, achieving a mean average precision (mAP) of 80.06%. Their findings underscore the efficacy of architectural enhancements in improving the reliability and precision of animal health assessment models. Moreover, the incorporation of SPPFCSPC in our study led to notable recall improvement, as shown in the YOLOv10x + SPPFCSPC configuration (81.7% recall, 84.0% accuracy). These results are consistent with the observations of Jooshin *et al*. (2024), who confirmed that spatial pyramid pooling variants, such as SPPF and SPPFCSPC, enhance the capacity for multi-scale feature aggregation, an essential function for capturing subtle anatomical variations in BCS-related visual tasks. Additionally, Utaminingrum *et al*. (2022) demonstrated that attention-based visual recognition models, when combined with statistical texture descriptors such as the gray-level co-occurrence matrix (GLCM), substantially improve classification performance in dynamic and unstructured environments. Although their work focused on assistive mobility systems, the underlying principles support the broader applicability of attention-enhanced detection models across diverse operational domains. Qiao *et al.* (2021) provide a comprehensive review emphasizing that integrated, non-contact AI vision systems are essential for scalable cattle monitoring tasks, including identification, BCS evaluation, and weight estimation. Their analysis shows that such systems, leveraging deep learning and computer vision, have matured to support real-world livestock operations at scale. The proposed 2D vision-based model offers a cost-effective and scalable alternative to 3D imaging systems, such as those used by Xiong *et al.* (2023), which, although accurate, are more complex and expensive for on-farm use. By aligning with established patterns in precision livestock farming, this model achieves a practical balance between accuracy, simplicity, and real-time feasibility, making it particularly well-suited to the constraints of tropical livestock systems.

The importance of feature selection and computational optimization in deep learning-based detection models has been demonstrated in various domains. Genetic algorithms (GA) and extreme learning machines (ELM) to optimize feature selection for road damage detection, achieving high classification accuracy while reducing computation time (Utaminingrum *et al*, 2023). By leveraging attention mechanisms such as CBAM and multi-scale pooling techniques like SPPFCSPC, the proposed YOLOv10 models effectively improve BCS classification while addressing computational constraints, ensuring their applicability in large-scale precision livestock management systems. By enabling objective and real-time monitoring of body reserves, the proposed AI-based BCS system can improve feed management, reduce metabolic disorders, and support reproductive planning in tropical beef cattle systems. The development of AI-based BCS systems represents a collaborative effort between animal scientists and computer engineers. From the animal science perspective, domain experts provide insight into anatomical features and validation of visual scoring standards.

CONCLUSION

The experimental results demonstrated that the YOLOv10x variant achieved the highest BCS classification accuracy (88.2%), with superior precision (88.2%), recall (83.6%), and F1 Score (85.8%), confirming its effectiveness for precise BCS estimation. In contrast, the YOLOv10n model achieved the lowest computational cost, with an inference time of 1.0 ms and 8.2 GFLOPS, indicating its suitability for real-time applications. These findings validate that integrating attention mechanisms and feature aggregation modules such as CBAM and SPPFCSPC effectively balances accuracy and computational efficiency. In tropical livestock systems, where manual BCS assessment remains labor-intensive and subjective, the proposed AI-based approach offers a scalable, non-invasive, and objective alternative.

CONFLICT OF INTEREST

The authors declare no conflict of interest with any organization or third party regarding the material discussed in this research.

ACKNOWLEDGEMENT

This research was financially supported by the Student Academic Research Scheme (UB STARS) of Universitas Brawijaya, Malang, Indonesia, under contract number 02388/UN10.A0101/B/RT.01.00/2024. The authors would also like to express their sincere appreciation to the Faculty of Computer Science, Universitas Brawijaya, and the Research Organization for Agriculture and Food, National Research and Innovation Agency (BRIN), for providing additional support and collaboration throughout the research process.

REFERENCES

Azzaro, G., Caccamo, M., Ferguson, J. D., Battiato, S., Farinella, G. M., Guarnera, G. C., Puglisi, G., Petriglieri, R., & Licitra, G. (2011). Objective estimation of body condition score by modeling cow body shape from digital images. Journal of Dairy Science, 94(5), 2126–2137. https://doi.org/10.3168/jds.2010-3467

Ghaffari, M. H., Sadri, H., & Sauerwein, H. (2023). Assessment of body condition score and body fat reserves in relation to insulin sensitivity and metabolic phenotyping in dairy cows. Journal of Dairy Science, 106(2), 807–821. https://doi.org/10.3168/jds.2022-22549

He, H., Chen, C., Zhang, W., Wang, Z., & Zhang, X. (2023). Body condition scoring network based on improved YOLOX. Pattern Analysis and Applications, 26(3), 1071–1087. https://doi.org/10.1007/s10044-023-01171-x

Jooshin, H. K., Nangir, M., & Seyedarabi, H. (2024). Inception-YOLO: Computational cost and accuracy improvement of the YOLOv5 model based on employing modified CSP, SPPF, and inception modules. IET Image Processing, 18(8), 1985-1999. https://doi.org/10.1049/ipr2.13077

Li, J., Zeng, P., Yue, S., Zheng, Z., Qin, L., & Song, H. (2025). Automatic body condition scoring system for dairy cows in group state based on improved YOLOv5 and video analysis. Artificial Intelligence in Agriculture, 15(2), 350–362. https://doi.org/10.1016/j.aiia.2025.01.010

Long, S. T., ToanN. C., GioiP. V., & HangP. T. (2022). Factors associated with the odds of pregnancy for dairy cattle after treatment of ovarian disorders in Northern Vietnam. Tropical Animal Science Journal, 45(3), 277-283. https://doi.org/10.5398/tasj.2022.45.3.277

Nagy, S. Á., Kilim, O., Csabai, I., Gábor, G., & Solymosi, N. (2023). Impact evaluation of score classes and annotation regions in deep learning-based dairy cow body condition prediction. Animals, 13, 194. https://doi.org/10.3390/ani13020194

Patterson, D. J., Perry, R. C., and Kiracofe, G. H. (2021). Body condition scoring of beef cattle. https://extension.missouri.edu/publications/g2230.

Qiao, Y., Kong, H., Clark, C., Lomax, S., Su, D., Eiffert, S., & Sukkarieh, S. (2021). Intelligent perception for cattle monitoring: A review for cattle identification, body condition score evaluation, and weight estimation. Computers and Electronics in Agriculture, 185, 106143. https://doi.org/10.1016/j.compag.2021.106143

Spoliansky, R., Edan, Y., Parmet, Y., & Halachmi, I. (2016). Development of automatic body condition scoring using a low-cost 3-dimensional Kinect camera. Journal of dairy science, 99(9),7714-7725. https://doi.org/10.3168/jds.2015-10607

Utaminingrum, F., Johan, A. W. S. B., Somawirata, I. K., Risnandar, & Septiarini, A. (2022). Descending stairs and floors classification as control reference in autonomous smart wheelchair. Journal of King Saud University - Computer and Information Sciences, 34(8), Part B, 6040–6047. https://doi.org/10.1016/j.jksuci.2021.07.025

Utaminingrum, F., Alqadri, A. M., Somawirata, I. K., Karim, C., Septiarini, A., Lin, C. Y., & Shih, T. K. (2023). Feature selection of gray-level co-occurrence matrix using genetic algorithm with extreme learning machine classification for early detection of pole roads. Results in Engineering, 20, 101437. https://doi.org/10.1016/j.rineng.2023.101437

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. arXiv preprint arXiv:2405.14458. https://doi.org/10.48550/arXiv.2405.14458

Wang, Y., Mücher, S., Wang, W., Guo, L., & Kooistra, L. (2023). A review of three-dimensional computer vision used in precision livestock farming for cattle growth management. Computers and Electronics in Agriculture, 206, 107687. https://doi.org/10.1016/j.compag.2023.107687

Xiong, Y., Condotta, I. C., Musgrave, J. A., Brown-Brandl, T. M., & Mulliniks, J. T. (2023). Estimating body weight and body condition score of mature beef cows using depth images. Translational Animal Science, 7(1), txad085. https://doi.org/10.1093/tas/txad085