Indian J Dairy Sci 77(3): 243-251
https://doi.org/10.33785/IJDS.2024.v77i03.008

RESEARCH ARTICLE

Development of convolutional neural network models for evaluation of body condition scores of Holstein Friesian crossbred cows

Shriramulu[1], Heartwin A Pushpadass[1] (✉), Magdaline Eljeeva Emerald Franklin[1], Manimala Kanagaraj[2], Jeyakumar Sakthivel[3], Mukund A. Kataktalware[4], Sivaram Muniandy[5] and Ramesha P. Kerekoppa[6]

Received: 20 September 2023 / Accepted: 17 October 2023 /

Abstract: Body condition scoring (BCS) is an efficient tool to monitor the nutritional status of cows by subjective assessment of the amount of fat or stored energy in them. This method of scoring is cumbersome, laborious and inefficient because it takes more time to score due to increasing number of animals in modern dairy farms, besides involving high cost. Therefore, this study proposed a system based on convolutional neural network (CNN) models to automate BCS of cows by image analysis. The digitally-captured images were processed using GIMP software to subtract the background from the captured images of cows. The background-subtracted images were used to detect the edges and contours using fuzzy logic edge detection method in Matlab software. Finally, the images with body contours and edges were employed as input dataset for the development of CNN models. The image dataset was classified into two groups based on the incremental BCS system of 0.25 (CNN model 1) and 0.5 (CNN model 2). The classification accuracy of the first model for 0.25 and 0.50 error ranges was 61.41% and 80.31%, respectively. Similarly, the second model achieved classification accuracy of 81.45% and 93.54%, for the same error ranges. The CNN model 2 was relatively better as compared to the first model owing to the wider BCS range of classification.

Keywords: Convolutional neural network; Body condition scoring; Classification accuracy; Image processing; Deep learning

Introduction

Assessing the body condition score (BCS) of cows is important to measure the changes in subcutaneous body fat throughout the stages of lactation. This is because the anatomical characteristics of cows such as external shape and tissue cover with frame size can be related to their current nutritional status and productivity. The extremely fat or skinny cows give reduced milk yield, pose risk of metabolic diseases and have low reproductive performance (Albornoz et al. 2021). During early lactation, depletion of their fat reserves takes place, while in late lactation, accumulation of fat takes place. Therefore, it is crucial to manage the BCS of cows at various stages of lactation so as to improve their health condition and productive performance. The most common approach for evaluating the BCS is the subjective method that uses a standardized scale based on the amount of tissue coverage at the hindquarters of the cow.

Body condition scoring is visually done by experts by assessing the rear regions of cows such as the loin, tail and pelvis, and providing a score ranging from 1 to 5. This technique is vulnerable to bias in many situations and requires the animal to be confined during evaluation. It is also time-consuming, restricting the total number of animals that can be examined as well as the frequency with which they can be evaluated. Although the significance of BCS has been evidenced in numerous studies, more than half of the farm producers never record the BCS of cows in their herds regularly (Caraviello et al. 2006). Similarly, the survey conducted by German fresh cow management found that only 36% of farmers measured the BCS of their herd regularly (Heuwieser et al. 2010).

The lack of implementation of regular BCS measurement is due to practical difficulty in scoring and management of BCS scores by the experts and the farmers (Song et al. 2019). To incorporate high-quality manual scoring into farm management, farmers must employ experienced assessors on a regular basis or obtain training to score their own cows, both methods are time-consuming and costly. Hence, there is a need to develop modern hybrid and

---

[1]Dairy Engineering Section, ICAR-National Dairy Research Institute, Southern Regional Station, Adugodi, Bengaluru-560030, India

[2]Department of Information Science and Technology, Anna University, College of Engineering, Gunidy Campus, Chennai - 600 025, India

[3] Livestock Research Centre, ICAR-National Dairy Research Institute, Southern Regional Station, Adugodi, Bengaluru-560030, India

[4] Livestock Production and Management, ICAR-National Dairy Research Institute, Southern Regional Station, Adugodi, Bengaluru-560030, India

[5] Dairy Economics and Statistics Section, ICAR-National Dairy Research Institute, Southern Regional Station, Adugodi, Bengaluru-560030, India

[6] Dairy Production Section, ICAR-National Dairy Research Institute, Southern Regional Station, Adugodi, Bengaluru-560030, India

Heartwin A. Pushpadass (✉)
Email: heartwin1@gmail.com; P.Dhass@icar.gov.in; Mobile: +91 9448927370

machine learning computing tools using the assistance of image processing to determine the BCS in an automated way. Such a system will be a user-friendly technique to estimate the BCS of large herds, analyze the BCS data set, and facilitates in easy identification of the variation in BCS of cows so as to take timely and appropriate management decisions.

Image processing is the important step for training and testing the dataset as it extracts the features that will be utilized for estimation of BCS. Developing a predictive BCS model using conventional feature extraction using manual method is less effective and less precise owing to the small amount of samples and feature points, which result in a poor fit (Fischer et al. 2015). In general, 3D vision-based systems achieved more accurate results (Li et al. 2019; Song et al. 2019; Alvarez et al. 2018). This is because the concavity information of the cows' body surface, provided by the 3D image characteristics, is more closely related to the fat accumulation beneath the skin.

In recent years, the area of deep learning has made great strides in the field of computer vision and image classification. Feature extraction is a key aspect of machine learning techniques. Deep learning algorithms fix the problem of feature extraction by automatically extracting relevant features from the raw input data rather than requiring pre-selected features. A deep learning model with various processing layers may learn complex input data attributes at various levels of abstraction (Cao et al. 2018). The convolutional neural network (CNN) is a deep learning method that is used to solve the complexity of developing, training and implementing machine learning models at any scale. It is a biologically-inspired concept of a deep network for feature detection, capable of learning purely complex features, and proves to be more effective in identification of objects (Zhang et al. 2017). In comparison to fully-connected networks with an equal number of hidden units, CNN architecture has the advantage of being simpler to train and has fewer parameters.

The major attraction of CNN is the concept of weight sharing to reduce the number of parameters that need to be trained to achieve greater generalization and only a few parameters are needed for training to avoid over-fitting (Arel et al. 2010). Secondly, the classification stage is combined with the feature extraction stage, both based on the learning process (LeCun et al. 1998). Thirdly, implementing large networks by use of general models in artificial neural networks (ANN) is more complex than implementation in CNN. In this study, a computer-aided CNN model was developed to evaluate the BCS of cows in order to minimize the error and bias that might arise during from interpretation of BCS by experts. The automatic BCS system was developed using various network architecture and general models used in deep-learning techniques, by selecting the body features of cows and extracting them from multiple viewpoints, so that the image-based BCS classification could be accurate and relevant.

Materials and Methods

Data collection

To develop and evaluate the CNN model, five dairy farms were visited and 503 images of Holstein Friesian cows with varying body muscle and fat in the hindquarters were acquired. A digital SLR camera (Nikon D5100, Nikon Corporation, Japan) was used to capture the images manually at ISO speed of 100, exposure time of 1/100 s and resolution of 4928×3264 pixels by ensuring the same distance and angles during image acquisition such that consistency could be maintained and error could be minimized. The images were acquired after cleaning and milking the animals. The vital areas of the cows were imaged at two suitable viewing angles for evaluation of BCS. The first dorsal view provided information on spinal parameters such as hook bones, pin bones and tail head of the cow. The second view was the side image, which gave the edge information on short and long ribs, and the area between pin and hook bones.

Dataset preparation

A scorecard containing images from the two viewing angles and detailed physiological information of the cows was developed. The images were printed and manually evaluated by three veterinary experts and their BCS were obtained on a 5-point scale. This scale is not breed specific, where score 1 represented thin cows and score 5 represented obese cows as explained by Gillund et al. (2001). The dataset of scores by experts was divided into 'training', 'testing' and 'validation' sets. Additionally, the image dataset was classified into two groups based on the incremental BCS system of 0.25 and 0.5 and each CNN model architecture was applied for prediction. This classification helped in evaluating the model performance for each increment of BCS. For development of CNN model, 70% of images were used, while 30% of them were used for testing and validation. Figs. 1 (a) and (b) show the percentage distribution of body condition scores over training and test set for (a) model 1 and (b) model 2.

- Model 1: Increment of 0.25 unit class of BCS
- Model 2: Increment of 0.5 unit class of BCS

System information and software tools

A 2.0 GHz CPU running on 64-bit Windows 10 was used to train the model. The model implementation program for BCS estimation was written in Python (v. 3.7.0, Python Software Foundation, Wilmington, Delaware, United States). The 'Keras' library in Python was used to build the CNN-based image classification. It is an advanced application programming interface (API) that operates on the basis of Tensorflow and offers advanced building blocks for creating deep learning models (Manaswi, 2018). Computer unified device architecture deep neural network

Fig. 1 Percentage distribution of body condition scores over training and test set for (a) model 1 and (b) model 2

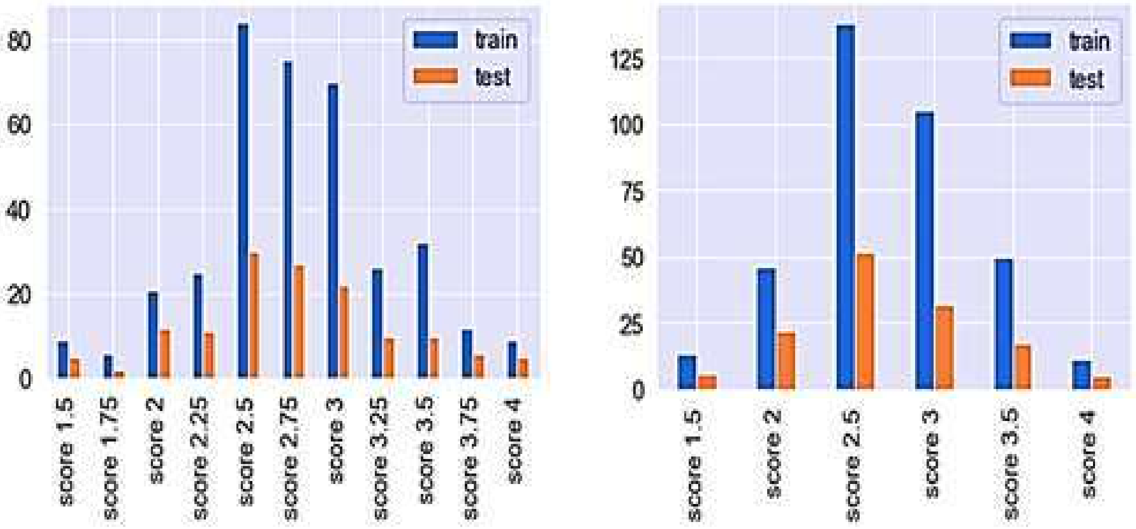
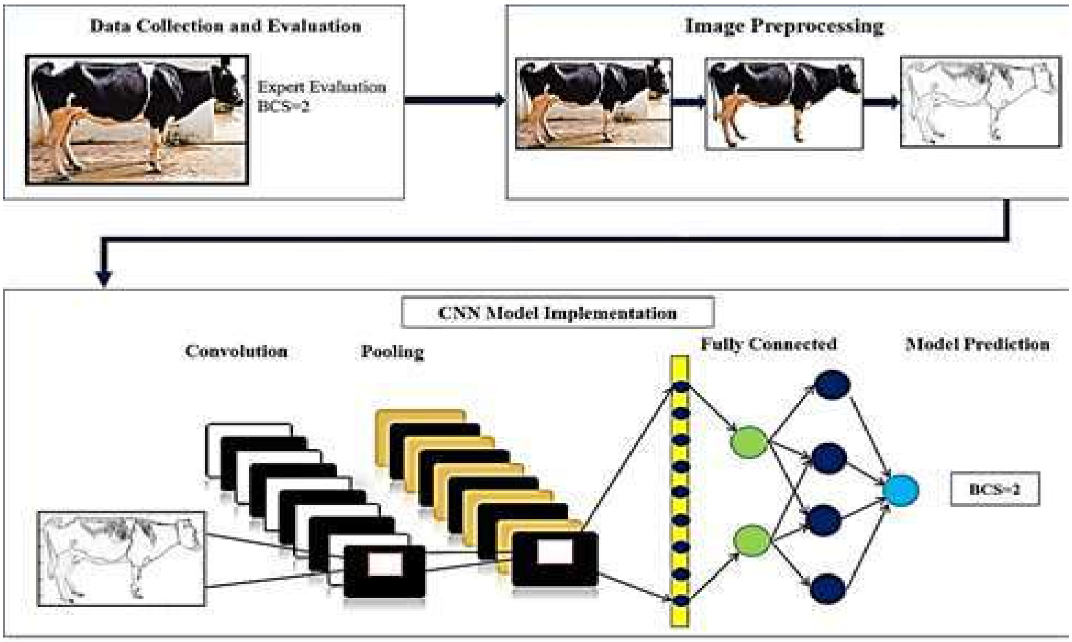


Fig. 2 Overview of the developed convolutional neural network system for estimation of body condition scores



(CuDNN) was used to run the Keras models for superior performance GPU acceleration. CuDNN is a library that offers fine-tuned implementations like forward and backward convolution, normalization, pooling and activation layers (Chetlur et al. 2014).

Background subtraction and fuzzy logic edge detection of images

The segmentation between background and digitally-captured images was subtracted using the GNU image manipulation program (v. 2.10.22, GIMP Development Team, Charlotte, North Carolina, USA). It is an open-source software for performing image-feature tasks such as modifying, organizing and programming images. The background-segmented images thus obtained were converted to edge images using the image processing toolbox in Matlab (v. R2022a, MathWorks, Natick, Massachusetts, USA). A Matlab code was implemented to locate the sharp changes in intensity and to identify the boundaries or contours of the captured images using fuzzy logic edge detection method. In this method, the cow images were converted to grayscale, and then a gradient image was produced using gradient filters. A fuzzy inference system was used to obtain edge images. It used fuzzy set theory, IF-THEN rules, and a fuzzy reasoning process to determine the output corresponding to crisp inputs. The image processing steps for digitally-captured images are shown in Fig. 2. The edge of a digital image is a collection of pixels with change in gray value as well as the area where the brightness of the local area of the image changes significantly.

Edge detection was used as a pre-processing step to obtain low-level boundary features that were passed on to subsequent processing steps like object detection and recognition. The edge-detected images thus obtained were used as input dataset for the development of CNN models.

Model implementation

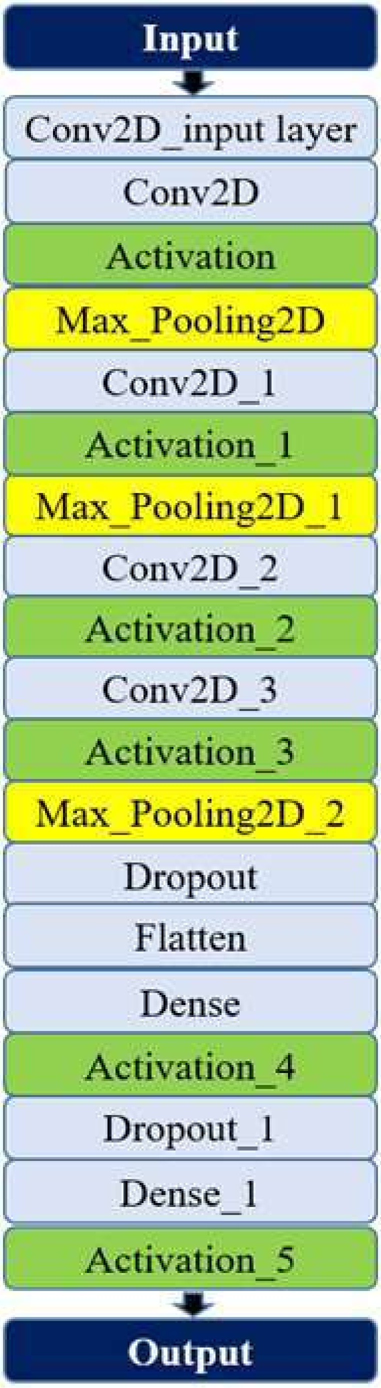
The CNN architecture was designed to evaluate the BCS of the image dataset. Fig. 2 shows the overview of the developed CNN-based BCS estimation system. There were three types of layers in the CNN architecture namely, convolutional, pooling and fully-connected. The convolutional layer computes the convolution operation of the input images using kernel filters to extract various features using spatial filters such as edges, lines and corners. After the convolutional layer, the pooling layer was placed. The pooling layers are effective to minimize the spatial resolution of input volume for the next convolutional layer, making the features robust against noise and distortion (Hasan et al. 2022). The pooling layers do not affect the dimension of volume. The pooling layer applied non-linear down sampling on the activation maps. As reduction in size results in loss of information, the process done by this layer is often referred to as sub-sampling.

Several convolutional and pooling layers are stacked on top of each other to extract more abstract features in the neural network through fully-connected layers. Fully-connected layers contain a variety of classification techniques depending on the structure and type of network. As their name suggests, the neurons in a fully-connected layer have complete connections to all activation functions in the previous layer (Hasan et al. 2022; Hiary et al. 2018). The function of fully-connected layer was to converge the attributes obtained from convolutional and pooling layers to the number of classes the dataset was intended to be classified.

Proposed architecture

The designed CNN architecture used the edge-detected images as input with size of 256×256 px (pixels). The first layer in the CNN applied 32 filters on the input images with size of 3×3 px, producing 32 feature maps of 254×254 px size. The convolved layer then passed through a max-pooling layer of 2×2 px size and produced image size of 127×127 px. The second layer applied 32 filters, each of 3×3 px size, producing 32 feature maps of 125×125 px size and passed through a max-pooling layer of 2×2 px size, producing an image size of 62×62 px. The third layer applied 64 filters, each of 3×3 px size, producing 64 feature maps of 60×60 px size. Subsequently, it passed through the max-pooling layer of size 2×2 px, producing image size of 30×30 px. In this architecture, rectified linear unit (ReLU) function was used as the activation function for both convolutional and fully-connected layers. Dropout was used to prevent overfitting by reducing the correlation between neurons. The architecture of the CNN model for estimation of BCS is shown in Fig. 3.

Fig. 3 Convolutional neural network model for estimation of body condition score



Statistical analysis

The classification performance of each CNN model was assessed using a set of metrics, which measured the accuracy of prediction besides comparing the predictive performance of different models. The confusion matrix, commonly used in classification problems (Çevik, 2020; Tripathi, 2021), is a tool to recognize different classes and show the details of correct and incorrect classification for each class. In a confusion matrix, the rows depict true observations and the columns depict predicted data. The diagonal of the confusion matrix represents the number of correctly predicted objects (Tatbul et al. 2018). In fact, for each class, four possible values can be identified. True-positive (TP) indicated that the true example was positive and the predicted example was positive. True-negative (TN) indicated that the true example was negative and the predicted example was negative. Meanwhile, a false-positive (FP) indicated that the true example was negative and the predicted example was positive. False-negative (FN) indicated that the true example was positive and the predicted example was negative (Tatbul, 2018; Chen et al. 2020). The proposed system was evaluated by the following parameters (Eqs. 1 to 4) that were based on the above factors, and were calculated using the information from confusion matrix.

- Classification accuracy (CA), which is the ratio of the number of true observations to the total number of observations, was calculated using Eq. 1.

$$CA=\frac{(True\ Positive+True\ Negative)}{(True\ Positive+False\ Positive+True\ Negative+False\ Negative)} \quad (1)$$

- Precision (P), which is the ability of the classifier to make positive predictions, was estimated using Eq. 2.

$$P=\frac{TP}{(TP+FP)} \quad (2)$$

- Similarly, recall (R), which is used to detect all positive instances, was determined by dividing number of true positives by the total number of actual positives (Eq. 3).

$$R=\frac{TP}{(TP+FN)} \quad (3)$$

- The last parameter namely, weighted average or harmonic mean of precision and recall (F1-score), was calculated using Eq. 4.

$$F_1=2\times\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (4)$$

Results and Discussion

Background subtraction and fuzzy logic edge detection of images

Manual BCS scoring is the traditional method to determine the score of each cow by visual assessment. The images of cows from two viewing angles and their physiological information were compiled. The veterinary experts evaluated the images of these cows and classed them into different BCS, varying from 1 to 5 with increment of 0.25. The digitally-captured images were processed using GIMP software to subtract the background from the images of cows. It was accomplished by using the foreground select tool from the GIMP toolbox. The outline of the cow image was created by choosing each point on the images. Subsequently, the background subtraction methodology was used to obtain processed images without background. The background-subtracted images were used to detect the edges and contours of the cow's body using fuzzy logic edge detection method. A Matlab programme was written to identify the edges and contours of the cow images, and the images with body contours and edges were employed as the training and testing dataset for development of the CNN models.

Convolutional neural network models

Figs. 4 (a) and (b) depict the training and validation accuracy after 10 epochs of training for CNN models 1 and 2, respectively. The training and validation accuracy improved as the number of epochs increased. Similarly, Figs. 4 (c) and (d) depict the training and validation loss after 10 epochs of training for CNN models 1 and 2, respectively. It was observed from Figs. 4 (a) and (b) that the training accuracy of the models was greater than its validation accuracy. This was due to the fact that a model's training accuracy was its performance on examples used in its construction, but its validation accuracy was its performance on unknown data. The rapid increase in accuracy for initial epochs suggested that the model learned quickly. Subsequently, the accuracy curve flattened, indicating that not many epochs were further needed to train the model (Fig. 4 a). Junayed et al. (2021) developed and validated the CNN model, and the findings revealed that training and validation accuracy improved during early epochs and the curve flattened thereafter. In case of model 2 (Fig. 4 b), the training and validation accuracies were improved with number of epochs, suggesting that the model was well-fitted.

Fig. 4 (c) shows that the training loss had a flat line, and it had no chance of learning the training dataset due to insufficient data for extreme BCS classes. In Fig. 4 (d), the validation dataset did not give enough information to assess the ability of model to generalize. This was due to less number of images in the validation dataset as compared to the training dataset. Therefore, the curve for validation loss showed the noisy movements, while the curve for training loss showed linear reduction in loss with increasing number of epochs during training.

Precision of fit of convolutional neural network models

Tables 1 and 2 show the confusion matrix of the test dataset for each model. To consider various error ranges, the confusion matrix was represented by different colour scales. The yellow cells represented accurate predictions, which were consistent with the assessment of experts. The upper triangular portion of the

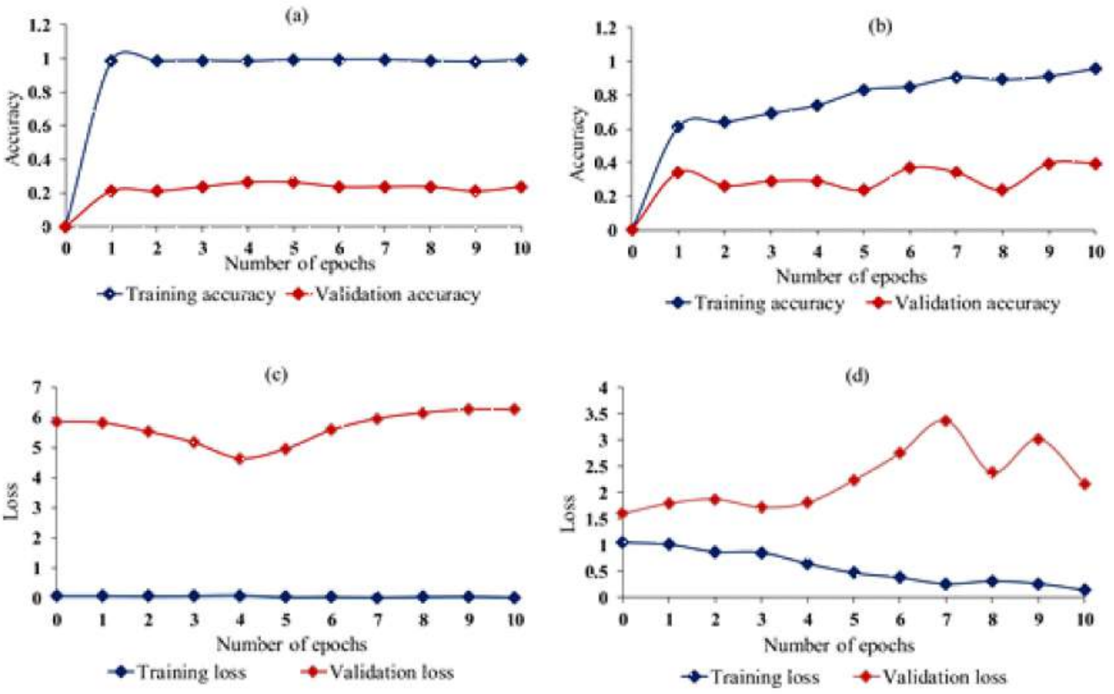


Fig. 4 Training and validation accuracy for (a) model 1 and (b) model 2 & Training and validation loss for (c) model 1 and (d) model 2

Table 1 Confusion matrix of convolutional neural network model 1

		Confusion matrix										
	1.50	0	0	1	1	0	1	0	0	0	0	0
	1.75	0	1	0	0	0	1	0	0	0	0	0
	2.0	0	0	4	0	6	1	1	0	0	0	0
	2.25	0	0	0	4	3	1	2	0	0	0	1
	2.50	0	0	1	0	19	5	4	1	0	0	0
True class	2.75	0	0	0	0	4	15	6	1	1	0	0
	3.0	0	0	0	0	7	2	12	0	1	0	0
	3.25	0	0	1	0	0	3	2	3	0	1	0
	3.50	0	0	0	0	4	1	2	0	3	0	0
	3.75	0	0	0	0	1	0	0	2	1	1	0
	4.0	0	0	0	0	1	1	1	0	0	0	1
		1.50	1.75	2.0	2.25	2.50	2.75	3.0	3.25	3.50	3.75	4.0
		Predicted class										

Table 2 Confusion matrix of convolutional neural network model 2

		Confusion matrix					
	1.5	0	0	1	1	2	0
	2.0	0	9	11	2	1	0
True class	2.5	0	2	39	8	3	1
	3.0	0	0	13	17	4	0
	3.5	0	0	4	3	11	0
	4.0	0	0	2	2	0	0
		1.5	2.0	2.5	3.0	3.5	4.0
		Predicted class					

confusion matrix were indicative of overestimation by the models for the test dataset, while the lower triangular portion signified underestimation by the models. The green cells represented predictions within 0.25 unit of error, while the light orange cells represented predictions within 0.50 unit of error. This approach allowed simplifying the calculation of remaining parameters by using the values of confusion matrix with different error ranges. Majority of data in the whole dataset were represented along the diagonal of confusion matrix, while the remaining elements were near zero. It could be stated that the optimum condition was not entirely achieved. It was also evident that the BCS between 2 and 3 were more accurately predicted than the extreme BCS values owing to the less number of such extreme BCS values (or images) in the dataset, which made the model difficult to recognize the pattern of such extreme BCS during training (Rodríguez et al. 2018).

Classification accuracy and comparison of convolutional neural network models

Table 3 shows the classification accuracy of each model, which was calculated using the values of confusion matrix within various error ranges. The accuracy of the first CNN model for 0, 0.25 and 0.50 error ranges was 46.32%, 63.23% and 85.29%, respectively. Similarly, the second model achieved classification accuracy of 55.88%, 86.02% and 94.85% for the corresponding error ranges. The accuracy of the second model was relatively better as compared to the first model owing to the bigger BCS range of classification (0.5 instead of 0.25) of the dataset. As the BCS increment of the second model was higher, it enabled the model to differentiate the BCS values easily without much misinterpretation. If the increment between the classes of BCS was lower, it would be tough for the model to recognize and classify the dataset for each range. However, the overall accuracy of the second CNN model was better (94.85%) as compared to manual assessment (78%) in evaluating the BCS. It was observed that the models of this study achieved accuracy levels similar to earlier works for different error ranges. Bewley and Schutz (2008) assessed the BCS by manually marking 23 anatomical features from the rear contour of cows, and achieved an accuracy of 92.79% within 0.25 units of error. Similarly, Spoliansky et al. (2016) did BCS study using a low-cost camera, and achieved an overall accuracy of 71.35% within 0.25 error range. A recent study employing image analysis model for CNN achieved an accuracy of 78% within 0.25 error range (Alvarez et al. 2018). Yukun et al. (2019) developed a CNN-based BCS system to evaluate the back images of 686 cows, and the model obtained average accuracy of 45%, 77% and 98% within 0, 0.25 and 0.50 error ranges, respectively. From these studies, it could be stated that the method of data collection and feature extraction techniques used in image processing influence the prediction accuracy of BCS.

Table 3 Classification accuracy of Convolutional Neural Network models

Classification accuracy (%)		
Error range	Model 1	Model 2
0 (exact)	46.32	55.88
0.25	63.23	86.02
0.50	85.29	94.85

Tables 4 and 5 show the detailed evaluation of metrics such as precision, recall and F1-score for the two models with different error ranges. The last two rows of each table represent the combined results of the weighted and unweighted average metrics of the models for each class of BCS by considering the distribution in the test dataset. The weighted average was included because of the imbalance in the image dataset within each BCS class. In Tables 4 and 5, the zero values signified that the model could not predict positive samples for the BCS class of 1.5 and 4.0 regardless of the error range. It

Table 4 Classification measures of Convolutional Neural Network model 1

	Error range								
BCS score	0 (Exact)			0.25			0.50		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
1.50	0	0	0	0	0	0	1.00	0.33	0.50
1.75	1.00	0.50	0.66	1.00	0.50	0.67	1.00	0.50	0.66
2.0	0.57	0.33	0.41	0.57	0.33	0.42	0.90	0.83	0.86
2.25	0.80	0.36	0.49	0.88	0.64	0.74	0.89	0.73	0.80
2.50	0.42	0.63	0.50	0.56	0.80	0.66	0.83	0.97	0.90
2.75	0.48	0.55	0.51	0.74	0.78	0.76	0.84	0.96	0.87
3.0	0.30	0.54	0.38	0.58	0.64	0.61	0.85	1.00	0.92
3.25	0.42	0.30	0.35	0.55	0.50	0.52	0.90	0.90	0.90
3.50	0.50	0.30	0.37	0.60	0.30	0.40	0.83	0.50	0.62
3.75	0.50	0.20	0.28	0.66	0.40	0.50	1.00	0.80	0.89
4.0	0.50	0.25	0.33	0.50	0.25	0.33	0.50	0.25	0.33
Weighted average per class	0.47	0.46	0.43	0.66	0.60	0.60	0.85	0.85	0.84
Unweighted average per class	0.50	0.36	0.39	0.60	0.47	0.51	0.86	0.71	0.75

Table 5 Classification measures of Convolutional Neural Network model 2

	Error range								
BCS score	0 (Exact)			0.25			0.50		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
1.5	0	0	0	0	0	0	1.00	0.25	0.40
2.0	0.82	0.39	0.52	1.00	0.87	0.97	1.00	0.96	0.98
2.5	0.56	0.73	0.62	0.88	0.92	0.90	0.96	0.98	0.97
3.0	0.52	0.50	0.50	0.87	1.00	0.93	0.97	1.00	0.98
3.5	0.52	0.61	0.56	0.70	0.78	0.74	0.86	1.00	0.92
4.0	0	0	0	0	0	0	0.67	0.50	0.57
Weighted average per class	0.56	0.56	0.53	0.82	0.86	0.84	0.95	0.95	0.94
Unweighted average per class	0.47	0.45	0.37	0.58	0.60	0.59	0.91	0.78	0.80

occurred because in the image dataset these classes had very less data for adequate training and testing of the models.

In F1-score (the harmonic mean of precision and recall), the classes that included more dataset of images in training and are in the middle of the BCS scale (BCSe”2 and BCSd”3), achieved the best outcome. The mean F1-score was 0.50, 0.78 and 0.93 within the error ranges of 0, 0.25 and 0.50, respectively. Conversely, in case of extreme BCS scores (BCS 1 and 4), the F1 score was 0.26, 0.33 and 0.64 within the error ranges of 0, 0.25 and 0.50, respectively. It was also observed from the results of the confusion matrix and classification measures that the models showed difficulty in classifying the dataset of extreme BCS scores. These issues were attributed due to the lack of adequate distribution of such extreme BCS in the image dataset, and the limited ability of the models to extract features related to extreme BCS scores, which resulted in their average performance for that category. However, it is not a major concern as it is not common to have cows in such poor or obese physical condition in a typical dairy farm. Ferguson et al. (2006) and Rodríguez et al. (2018) reported similar difficulties in differentiating extreme BCS of less than 2 and greater than 3. The classes with more training dataset in the middle of the BCS scale showed good predictive results. The predictive performance of the models for extreme range of BCS could be improved by collecting a larger dataset consisting of extreme BCS. About 66% of images in the training dataset had BCS between 2 and 3. This implied that the majority classes (BCSe”2 and BCSd”3) were important in determining the overall accuracy. As a result, the overall accuracy dropped if the majority of classes were incorrectly categorized. On the contrary, if the minority classes (BCS 1 and 4) were mis-classified, the loss of overall accuracy would not increase considerably. Based on the results, it could be stated that the developed models achieved good results for the middle range of BCS scores and provided strength to models, despite the lack of extreme BCS scores in the dataset. Hence, the developed models would perform effectively for commercial dairy farms, which do not commonly have cows with poor or high BCS as they would not be very productive.

Conclusion

Body condition score (BCS) is an efficient tool to monitor the nutritional status of cows. The quality of manual BCS is decided by the assessor's experience and the standard of scoring technique. As a result, incorporating regular high-quality scoring of BCS of individual cows in a commercial farm as a routine procedure of management is challenging. In this study, two CNN models with two different error ranges were used to estimate the BCS of cows using edge detected images for different increments and error ranges of BCS. The classification accuracy of the models ranged between 85 and 95%, depending on the error range selected. The CNN models performed adequately for the middle range of BCS, wherein the data of most cows lied, but did not predict the extreme BCS classes very well. It is not a major concern as farms do not keep cows with such extreme BCS scores. However, there is further scope to make the system robust for all BCS classes by enlarging the image dataset to train and validate the CNN based system.

References

Albornoz RI, Giri K, Hannah MC, Wales WJ (2021) An improved approach to automated measurement of body condition score in dairy cows using a three-dimensional camera system. Anim 12(1): 72

Alvarez JR, Arroqui M, Mangudo P, Toloza J, Jatip D, Rodríguez JM, Mateos C (2018) Body condition estimation on cows from depth images using convolutional neural networks. Comput Electron Agric 155: 12-22

Arel I, Rose DC, Karnowski TP (2010). Deep machine learning-a new frontier in artificial intelligence research. IEEE Comput Intell Mag 5(4): 13-18

Bewley JM, Schutz MM. (2008) An interdisciplinary review of body condition scoring for dairy cattle. Prof Anim Sci 24(6): 507-529

Cao C, Liu F, Tan H, Song D, Shu W, Li W, Xie Z (2018) Deep learning and its applications in biomedicine. Genom Proteon Bioinform (1): 17-32

Caraviello DZ, Weigel KA, Craven M, Gianola D, Cook NB, Nordlund KV, Wiltbank MC (2006) Analysis of reproductive performance of lactating cows on large dairy farms using machine learning algorithms. J Dairy Sci 89(12): 4703-4722

Çevik KK (2020) Deep learning based real-time body condition score classification system. IEEE Access 8: 213950-213957

Chen W, Li Y, Xue W, Shahabi H, Li S, Hong H, Wang X, Bian H, Zhang S, Pradhan B, Ahmad BB (2020) Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating decision tree, and random forest methods. Sci Total Environ 701: 134979

Chetlur S, Woolley C, Vandermersch P, Cohen JM, Tran J, Catanzaro B, Shelhamer E (2014) cuDNN: Efficient primitives for deep learning. *ArXiv, abs/1410.0759*

Ferguson JD, Azzaro G, Licitra G (2006) Body condition assessment using digital images. J Dairy Sci 89(10): 3833-3841

Fischer A, Luginbühl T, Delattre L, Delouard JM, Faverdin P (2015) Rear shape in 3 dimensions summarized by principal component analysis is a good predictor of body condition score in Holstein dairy cows. J Dairy Sci 98(7): 4465-4476

Gillund P, Reksen O, Gröhn YT, Karlberg K (2001) Body condition related to ketosis and reproductive performance in Norwegian dairy cows. J Dairy Sci 84(6): 1390-1396

Hasan N, Ibrahim S, Aqilah Azlan A (2022) Fish diseases detection using convolutional neural network (CNN). Int J Nonlinear Anal 13(1): 1977-1984

Heuwieser W, Iwersen M, Gossellin J, Drillich M (2010) Survey of fresh cow management practices of dairy cattle on small and large commercial farms. J Dairy Sci 93(3): 1065-1068

Hiary H, Saadeh H, Saadeh M, Yaqub M (2018) Flower classification using deep convolutional neural networks. IET Comput Vis 12(6): 855-862

Junayed MS, Islam MB, Jeny AA, Sadeghzadeh A, Biswas T, Shah AS (2021) ScarNet: development and validation of a novel deep CNN model for acne scar classification with a new dataset. IEEE Access 10: 1245-1258

LeCun Y, Bottou L, Bengio Y Hafner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11): 2278-2323

Li WY, Shen Y, Wang DJ, Yang ZK, Yang XT (2019) Automatic dairy cow body condition scoring using depth images and 3D surface fitting. In: 2019 *IEEE* Int Conf Unmanned Syst Artif Intell IEEE pp. 155-159. 22-24 **November** 2019. Xi'an, China

Manaswi NK (2018) Understanding and working with keras. In: Deep Learning With Applications Using Python. pp. 31-43, Apress, Berkeley, CA

Rodríguez AJ, Arroqui M, Mangudo P, Toloza J, Jatip D, Rodrígue JM, Teyseyre A, Sanz C, Zunino A, Machado C (2018) Body condition estimation on cows from depth images using Convolutional Neural Networks. Comput Electron Agric 155: 12-22

Song X, Bokkers EAM, Van Mourik S, Koerkamp PG, Van Der Tol PPJ (2019) Automated body condition scoring of dairy cows using 3-dimensional feature extraction from multiple body regions. J Dairy Sci 102(5): 4294-4308

Spoliansky R, Edan Y, Parmet Y, Halachmi I (2016) Development of automatic body condition scoring using a low-cost 3-dimensional Kinect camera. J Dairy Sci 99(9): 7714-7725

Tatbul N, Lee TJ, Zdonik S, Alam M, Gottschlich J (2018) Precision and recall for time series. In: 32nd IEEE Int Conf Neural Netw Inf Process Syst 31, pp. 1920-1930. 3-8 December 2018. Montréal, Canada

Tripathi M (2021) Analysis of convolutional neural network based image classification techniques. J Innov Image Process 3(2): 100-117

Yukun S, Pengju H, Yujie W, Ziqi C, Yang L, Baisheng D, Yonggen Z (2019) Automatic monitoring system for individual dairy cows based on a deep learning framework that provides identification via body parts and estimation of body condition score. J Dairy Sci 102(11): 10140-10151

Zhang K, Zuo W, Chen Y, Meng D, Zhang L (2017) Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Trans Image Process 26(7): 3142-3155