Indian J Dairy Sci 78(4): 344-349
https://doi.org/10.33785/IJDS.2025.v78i04.007

RESEARCH ARTICLE

Development of a lightweight deep learning model for the identification and classification of Indigenous cattle breeds

Shruti Arya[1], Indu Devi[1] (✉), Divyanshu Singh Tomar[2], Man Singh[3], T. K. Mohanty[1], Radhika Warhade[1] and Anshul Gautam[1]

Received: 06 January 2025 / Accepted: 05 June 2025 / Published online: 23 August 2025

Abstract: This study aimed to develop a lightweight deep learning model for the identification and classification of Tharparkar and Hariana cattle breeds as they are phenotypically similar-looking and have subtle differences in visual appearance. Images were collected from 115 cows of each breed under natural conditions. A CNN-based semantic segmentation model was developed to accurately identify the cow as a Region of Interest in the given image. The IoU value of 84.15% and F1-Score of 87 % of the segmentation model for the cow region suggested that the model was capable in segmenting the cow pixels. The masked image as output from the segmentation model was used as input for the final breed classifier model. The recall value of 86 % and precision value of 88 % of the segmentation model for the cow region indicated that the model effectively identified cow regions with high accuracy, minimizing false positives. The model requires approximately 618 milliseconds and 3.27 million parameters to perform segmentation for one image. The accuracy of the classification model for the Tharparkar and Hariana class was found to be 72.5%. Precision, recall value, and F1-Score for the Hariana breed were 73.7%, 70.0%, and 71.8%, respectively. Whereas precision was 71.4%, recall value was 75.0%, and F1-Score was 73.2% for Tharparkar. This study attempted to differentiate white-coloured breeds using a deep learning method without the help of manual help and experts. Further research on robust datasets and fine-tuning of the model parameters may lead to better accuracy in breed classification.

[1] Livestock Production Management Division, ICAR-National Dairy Research Institute, Karnal-132001, Haryana, India.
[2] Assistant Professor, Livestock Production Management, Rani Lakshmi Bai Central Agricultural University, Jhansi- 284003, Uttar Pradesh, India
[3] Assistant Professor, Livestock Production Management, Lala Lajpat Rai University of Veterinary & Animal Sciences, Hisar – 125011

(✉)Indu Devi, Livestock Production Management Division, ICAR-National Dairy Research Institute, Karnal-132001, Haryana, India.
Email: indulathwal@gamil.com

Keywords: Breed identification, Deep learning model, Image processing, Similar-looking cattle breeds

Introduction

Indigenous animals play a crucial role in the national economy of our country in terms of milk, draught animal power, organic manure and urine (medicinal value). India has 53 registered breeds of Indigenous cattle (NBAGR, 2024). The population of pure Indigenous cattle breeds is declining due to indiscriminate breeding practices (Saini et al. 2021) and not-so-sound specific breed identification and importance among farmers in field conditions. Indigenous cattle breeds such as Tharparkar and Hariana are well adapted under Indian agro-climatic conditions and well known for their production potential and survivability on poor feed and fodder resources. Therefore, these breeds need to be identified for their unique traits by stakeholders accurately during breeding, purchasing and selection decisions. Tharparkar cattle have a white or greyish colour, broad, flat-convex forehead, medium-sized body, deep, robust build, prominent naval flap, straight limbs and fine feet, as well as an attentive and lively carriage. Hariana cattle have slim and long faces, a flat forehead, with a well-defined bony projection at the poll's centre, small horns and sheath, compact and proportionately built bodies, and coats are often white or light grey and cover the black skin. Short-horned, narrow-faced and grey cattle breeds such as Hariana and Tharparkar are phenotypically closer (NBAGR, 2023). Despite having similar white and greyish colours, these two breeds differ in a few important physical features, and experts take these phenotypic distinctions into account when determining which of the two breeds is a pure breed. Sometimes, experts are unavailable, and it is difficult to distinguish similar-looking native breeds like Hariana and Tharparkar, so purity of a breed may be misinterpreted. Further, it is time-consuming and human biasness can also be present. Genetic characterization for the identification of Indigenous breeds requires expensive laboratory facilities. Therefore, a suitable method with reasonable accuracy and economic viability is required for recognition of the breed. Computer vision and deep learning-based approaches can aid in developing an artificial intelligence-assisted system for breed identification (Abu Jwade et al. 2019). Convolutional Neural Network (CNN) predict automatically the type of object or animal breed present in a different set of images by learning the important discriminating features in the images of the training dataset (Binta et al. 2023). CNN-enabled deep learning models can perform image classification equal or better than human vision in many cases of human biomedical field such as deep learning method performed better or comparable to human (expert) technicians for scoring of leg movement (Carvelli et al. 2020), diagnosis of retinal diseases using image based deep learning method (Kermany et al. 2018). Lightweight deep learning models can be executed on small devices with limited resources and computational capabilities (Wang et al. 2022). The uniqueness of the lightweight model is that it is lighter than the state-of-the-art model's architecture with comparable/higher accuracy, and model training reproducibility. The present study was conducted to develop a 'lightweight deep learning model' for the identification and classification of visually similar looking cattle breeds (Hariana and Tharparkar).

Materials and Methods

Selection of animals and Image dataset construction

For this study, the images (2-D) of cows from both breeds i.e., Tharparkar and Haryana (115 cows each) were captured using a mobile phone camera with a resolution of 2296 × 4080 pixels. The study was conducted at ICAR-NDRI, Karnal from 2022-23. The images of Tharparkar cows were collected from Livestock Research Centre, ICAR-NDRI, Karnal and Hariana cows were collected from livestock farm, LUVAS, Hisar; during the daytime under the natural environment of animal sheds (as shown in Fig. 1). For easy computation through CNN, captured images were resized to 288 × 512 from 2296×4080 pixels. The images of 95 cows from each breed were used as a training dataset and images of the remaining 20 cows from each breed were kept as a testing dataset. A great number of images are needed to train the CNN model as a classifier in order to prevent overfitting. Three additional pictures were created from a single original image by applying three augmentation (geometric) procedures to the images: rotation (-10.0° to 10.0°), horizontal shift (60 pixels), and vertical shift (60 pixels). Table 1 provides information about the dataset of photos utilised. Semantic segmentation of cows from the surroundings was done in the current investigation to prevent background interference. A convolutional encoder-decoder type model was used for the semantic segmentation. The encoder part of the network was responsible for extracting features from the input image, while the decoder part samples the feature maps from the encoder to generate a segmentation map. Fig. 2 depicts the encoder-decoder type CNN model's architecture. 48 feature maps made up the first convolutional layer, which was then regularised using batch normalisation and applied non-linearity by means of a ReLU activation function. Sequentially, the CNN model comprised seven such layers with increased feature maps, each accompanied by batch normalization and ReLU activation. To manage computation and enhance feature extraction, max-pooling layers were inserted after the second, fourth, sixth, and seventh convolutional layers. After the convolutional layers, a series of deconvolutional layers (also known as transposed convolution or upsampling) were engaged to recover the spatial information lost during max-pooling. These layers were followed by a sequence of 2D convolution operations, batch normalization, and ReLU activation functions, for the refinement of features. Finally, the output layer employed a softmax activation function to generate probability score for each class. For the training of the CNN model for the semantic segmentation function, RGB images, as well as their corresponding ground truth (also called mask images), are required. In this study, a labelled dataset was created using Label-studio software. During the annotation process, each pixel in the images was labelled, with "0" representing the background and "1" representing the cow region.

Segmentation of cow from background using CNN model

To avoid the interference of background, semantic segmentation of cows from the background was carried out in the present study. A convolutional encoder-decoder type model was used for the semantic segmentation. The encoder part of the network

Table 1: Details of the training dataset used in the present study

Cattle breed	No. of images captured per cow	Total cows	Total images captured (original)	No. of images selected for training*	No. of images created through augmentation (3 images from 1 image)	Augmented images	Total number of images per breed for training dataset (augmented + original)
Hariana	4	120	480	95 × 4 = 380	380 × 3 = 1140	1140	1140 + 380 = 1520
Tharparkar	4	120	480	95 × 4 = 380	380 × 3 = 1140	1140	1140 + 380 = 1520
Total cow images used in the study as a training dataset							1520 + 1520 = 3040
Cow images used as testing dataset							200
Total cow images from both breeds							1520 + 1520 +200= 3240

was responsible for extracting features from the input image, while the decoder part samples the feature maps from the encoder to generate a segmentation map. The architecture of the encoder-decoder type CNN model is shown in Fig. 2. 48 feature maps comprising the first convolutional layer, which was then put through a ReLU activation function to introduce non-linearity and batch normalisation for regularisation. Sequentially, the CNN model comprised seven such layers with increased feature maps, each accompanied by batch normalization and ReLU activation. To manage computation and enhance feature extraction, max-pooling layers were inserted after the second, fourth, sixth, and seventh convolutional layers. After the convolutional layers, a series of deconvolutional layers (also known as transposed convolution or upsampling) were engaged to recover the spatial information lost during max-pooling. These layers were followed by a sequence of 2D convolution operations, batch normalization, and ReLU activation functions, for the refinement of features. Finally, the output layer employed a softmax activation function to generate probability score for each class. For the training of CNN model for semantic segmentation function, RGB images, as well as their corresponding ground truth (also called mask images), are required. In this study, a labelled dataset was created using Label-studio software. During annotation process, each pixel in

Fig. 1: Cow images captured (a) Front side of Tharparkar cow (b) Side view of Tharparkar cow (c) Front view of Hariana cow (d) Side view of Hariana cow

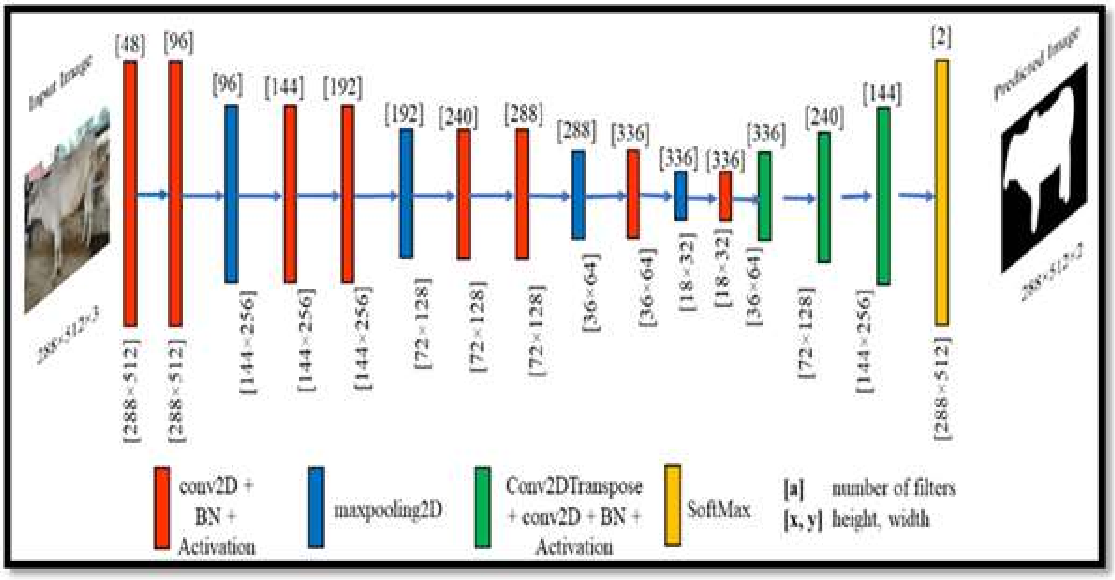

Fig. 2: The architecture of encoder-decoder type CNN model used for segmentation of cow

the images was labelled, with “0” representing the background and “1” representing the cow region.

Training and evaluation of the segmentation model

For this, 100 images of the front view and 100 images of the left side view were selected for each breed, resulting in a total of 400 images. Out of a total of 400 images from the collective data set of both breeds data set, approximately 70% of the annotated images (280 images) were used for training, the remaining 20% images were utilized for validation, and the remaining 10% were reserved for testing purposes of segmentation model. During the training process, the developed semantic segmentation model was trained using the training dataset. The model was trained for 100 epochs, with a batch size of ‘8’ images. The CNN model was trained using the Keras library (Chollet, 2015), with TensorFlow (TensorFlow Developers 2021) as the backend. The Adam optimizer (Kingma and Ba, 2017) was used to optimize the model parameters. Following the training phase, the model’s performance was evaluated using the validation dataset. Various evaluation metrics (Chen et al. 2021) such as intersection over union (IoU), precision, recall and F1 score; were calculated to assess the accuracy and effectiveness of the model in segmenting the cow region from the background.

Masking of segmented image over original cow image

The masking process acted as a pre-processing step that significantly improved the overall performance of the classification CNN model. After successfully performing semantic segmentation of the cow using our custom-designed lightweight CNN model, masking was used as a significant step towards enhancing the accuracy of the classification CNN model. By combining the segmented image with the original image, we effectively masked out all the background pixels, leaving them with a value of ‘0’. Meanwhile, the pixels corresponding to the cow in the original image remained unchanged. This approach was strategically adopted to enable the classification CNN model to exclusively focus on the essential cow features while disregarding the background information.

Deep learning model architecture used for breed classification

The CNN model for cow classification processes four images that have different dimensions using a multi-input architecture. Two images have a size of 288 x 512 with three colour channels (RGB), while the other two images are grayscale versions of the former. Each input image was passed through a separate parallel layer that consists of a convolutional block, followed by a max-pooling layer. The convolutional block included a series of operations: a convolutional layer, a batch normalization layer for improving model training stability, and an activation layer to introduce non-linearity. These operations help in capturing important features from the input images. After processing each input image, the output feature maps from all parallel layers were concatenated using a concatenate layer, enabling the model to merge information from multiple views and colours. Following the concatenation, the concatenated features undergo another series of convolutional blocks and max-pooling layers.

The first convolutional block processes the concatenated features with two convolutional layers, followed by a max-pooling layer. This results in feature maps of size 36x64 with 32 feature maps. The subsequent convolutional blocks continue this process, reducing the size of feature maps while increasing the number of feature maps as described below:

1.After convolving the concatenated features with two consecutive convolutional blocks, a maxpooling layer was added resulting in feature maps sizes of 36, 64 from the size of 72, and 128 with 32 number of feature maps

2.Output feature maps from the previous maxpooling layer were convolved with two next consecutive convolutional blocks, maxpooling layer was added resulting in feature maps size of 18, 32 from size of 36, and 64 with 64 number of feature maps

3.Output feature maps from the previous maxpooling layer were convolved with next four consecutive convolutional blocks, maxpooling layer was added resulting in feature maps size of 9,16 from size of 18, 32 with 128 number of feature maps

4.Output feature maps from previous maxpooling layer were convolved with next four consecutive convolutional blocks, maxpooling layer was added resulting in feature maps size of 4,8 from size of 9, 16 with 216 number of feature maps

5.Output feature maps from the previous maxpooling layer were convolved with next four consecutive convolutional blocks, maxpooling layer was added resulting in feature maps size of 2,4 from size of 4, 8 with 306 number of feature maps

By the end of the model, the feature maps were further reduced in size and enhanced in complexity. A flatten layer was added to convert the 2D feature maps into a 1D array with 2448 units. The flattened features were then fed into fully connected dense layers. The first dense layer has 700 units with ReLU activation, followed by the second dense layer with 300 units and ReLU activation. These dense layers enable the model to learn intricate patterns and relationships among the features. Finally, the output layer, consisting of two units corresponding to the two classes (Hariana and Tharparkar breed), was activated using the softmax function. This activation converts the final layer’s raw scores into class probabilities, indicating the likelihood of the input belonging to Hariana or Tharparkar breed. For evaluation of classification model. Various performance metrics (accuracy, precision, recall, and F1 score) were employed to evaluate the classification model’s effectiveness in identifying cow breeds.

Software and Hardware used for processing

The study was conducted using an Intel® Core™ i5 personal computer. This ran at 2.2 GHz and contained 25 GB of RAM, a GDDR5 graphics card. Python programming language was used along with jupyter notebook for training and validation of the CNN model.

Results and Discussion

The quantitative results of the segmentation model for the cow dataset in terms of four metrics IoU, F1-score, precision, and recall are presented in Table 2.

The performance of a breed classification model is generally evaluated by computing the number of correctly predicted classes out of all predictions made. Table 3 shows a quantitative evaluation of the classification performance of the proposed model for distinguishing the Hariana and Tharparkar cattle breeds.

For the Hariana class, the model correctly classified 14 samples as Hariana (TP) and misclassified 6 Hariana samples as Tharparkar (FN). Additionally, it classified 5 Tharparkar samples as Hariana (FP) and correctly classified 15 Tharparkar samples as Tharparkar (TN). The accuracy of the classification model for the Hariana class was found to be 72.5%, indicating that 72.5% of the samples were correctly classified. Similarly, for the Tharparkar class, the model correctly classified 15 samples as Tharparkar (TP) and misclassified 5 Tharparkar samples as Hariana (FN). It also classified 6 Hariana samples as Tharparkar (FP) and correctly classified 14 Hariana samples as Hariana (TN). The accuracy of the classification model for the Tharparkar class is also calculated to be 72.5%. The precision denotes the proportion of truly positive predictions done by the model out of all positive predictions. Precision for Hariana class and Tharparkar class is 73.7% and 71.4% respectively. The recall (true positive rate), denotes the proportion of truly positive predictions out of all real or actual positive samples. For the Hariana class, the recall is 70.0%, and for Tharparkar class, the recall is 75.0%. The F1-Score offers a balance between precision and recall by taking the harmonic mean of the two metrics. The F1-Score is 73.2% for the Tharparkar class and 71.8% for the Hariana class.

The quantitative metrics for segmentation model provides a comprehensive assessment of the model's accuracy in segmenting cow regions and its ability to minimize false positives and false negatives. Intersection over Union (IoU) of 89.47% for the background region indicated that the model precisely identified and segmented the background pixels in test images. This high IoU value signifies a strong alignment between the predicted background mask and the ground truth mask, leading to reliable background segmentation. The IoU value of 84.15% for the cow region suggested that the model effectively segmented the cow pixels, though it was slightly lower than the background IoU. It may be due to variations in the pose and texture of cow regions, which can make the segmentation task more challenging. The F1-Score of 90.30% for the background region showed a balance between precision and recall for background segmentation. This high F1-Score indicated that the model has minimized false positives and false negatives, resulting in accurate cow segmentation. Similarly, the F1-Score of 87.00% for the cow region revealed a well-balanced performance of the model in identifying cow pixels with good precision and recall. The recall value of 91.00% for the background region indicated that the model effectively detected and captured a high proportion of actual background pixels present in the ground truth mask. This high recall value indicates that the model misses a lesser number of background pixels during segmentation. The recall value of 86.00% for the cow region indicated that the model performed well in detecting cow regions in the test images. However, it slightly misses a small percentage of cow pixels present in the ground truth mask, which contributes to the overall false negatives. The model successfully recognised cow pixels during segmentation, with minimal false positives, according to the accuracy value of 88.00% for the cow region. The anticipated cow regions are certain to be a part of the cow in the pictures owing to this high level of accuracy. In situations where precise cow region segmentation is critical, this is crucial. The segmentation outcomes of our investigation were not correlated with any animal studies that we could uncover. So, we correlated with other similar agriculture studies to compare results of

Table 2: Evaluation of segmentation results obtained using models for background and cow

Evaluation metrics	Background (%)	Cow (%)
IoU	89.47	84.15
F1-Score	90.30	87.00
Recall	91.00	86.00
Precision	89.50	88.00

Table 3: Quantitative evaluation of classification results obtained using the proposed convolutional neural network

Class	TP	FN	FP	TN	Accuracy, %	Precision, %	Recall, %	F1-Score, %
Hariana	14	6	5	15	72.5	73.7	70.0	71.8
Tharparkar	15	5	6	14	72.5	71.4	75.0	73.2

segmentation of 'region of interest' from the input image. Li et al. (2017) utilized a CNN model for the detection of cotton bolls bearing in mind interference of the sky and got an IoU of 59.1% for images of sky interference. Singh et al. (2022) studied the CNN to segment and differentiate the targeted area as cotton bolls pixels from the background and attained IoU of 81.01% for the VGG16 model, 84.50% for the inception model and 83.65% for the ResNet model. We also obtained comparable results for segmentation of the region of interest (cow) from the background using CNN CNN-based segmentation model. The developed segmentation model had approx. 3.27 million trainable parameters and required 618 ms to process each input image. A model with a larger number of trainable parameters may have a more complex architecture, which could lead to increased computation during inference. Models with fewer trainable parameters may be simpler and more lightweight, leading to faster inference times.

Previous studies also reported comparable results for livestock breed classification. Agh Atabay (2018) proposed a deep learning model for the identification of horse breeds and found average classification accuracy of 90.69% for VGG16, 90.05% for VGG19, 88.79% for InceptionV3, 95.90% for ResNet50, 93% for Xception model. In this study they have used an already pre-trained model for horse breed identification which might be the possible reason for higher accuracy than our study; so, with the help of transfer learning, we can increase our accuracy. Weber et al. (2020) recognised the Pantaneira cattle breed with 99% accuracy across all three networks (DenseNet-201, Resnet50, and Inception-Resnet-V2). This study had high accuracy in comparison to our study, which might be due to use of more number of images, use of already trained models and dark coloured animals. Pan et al. (2022) proposed a computer vision-based recognition system to recognize and classify the Nili-Ravi buffalo breed and achieved accuracy results as 78.67 - 85.33 %, and F1 score of 78.78 – 85.57% for machine learning-based classifiers for identification of buffalo breed. The results of this study are comparable with our results in terms of accuracy, precision, recall and F1 score and also provided a concept proof that the newly designed CNN model needs more comprehensive training from scratch for a new dataset because the model is not pre-trained on this type of dataset. Overall, the results of the present study indicated that the proposed CNN model shows reasonably good performance in classifying Hariana and Tharparkar cow breeds. However, there are some misclassifications, as indicated by the false positives and false negatives, which may provide valuable insights for further model improvement and optimization. Further research and fine-tuning of the model parameters may lead to even better results and higher accuracy in breed classification.

Conclusion

A lightweight deep learning model was developed for identification and classification of Tharparkar and Hariana Cattle Breeds with an accuracy of 72.5%. The accuracy of the developed deep learning model can be enhanced by using a large, diverse image dataset, better fine-tuning of model parameters and use of uniform/dark background.

Acknowledgements

The authors sincerely acknowledge the support and facilities provided by the Director, ICAR-National Dairy Research Institute, Karnal, Haryana, and the Vice-Chancellor, LUVAS, Hisar, Haryana, India to carry out this work successfully.

References

Abu Jwade S, Guzzomi A, Mian A (2019) On farm automatic sheep breed classification using deep learning. Comput Electron Agric 167: 1-13

Agh Atabay H (2018) Deep Learning for Horse Breed Recognition. CSI J Comput Sci Eng 15: 45-51

Binta Islam S, Valles D, Hibbitts T J, Ryberg WA, Walkup DK, Forstner MRJ (2023). Animal Species Recognition with Deep Convolutional Neural Networks from Ecological Camera Trap Images. Animals 13: 1526

Carvelli L, Olesen AN, Brink-Kjær A, Leary EB, Peppar PE, Mignot E, Sørensen HB, Jennum P. (2020). Design of a deep learning model for automatic scoring of periodic and non-periodic leg movements during sleep validated against multiple human experts, Sleep Med 69: 109–119

Chen Z, Ting D, Newbury R, Chen C (2021) Semantic segmentation for partially occluded apple trees based on deep learning. Comput Electron Agric 181: 105952.

Chollet F (2015) Keras. Available at: https://github.com/fchollet/keras.

Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell. 172:1122–31, e9.

Kingma DP, Ba J (2017) Adam: A Method for Stochastic Optimization. 3rd International Conference on Learning Representations, San Diego, USA

Li Y, Cao Z-G, Xiao Y, Cremers A (2017) DeepCotton: In-field cotton segmentation using deep fully convolutional network. J Electron Imaging 26: 1

NBAGR (2024) Database, Information System of Animal Genetics Resources of India. Available at: https://nbagr.icar.gov.in/

Pan Y, Jin H, Gao J, Rauf HT (2022) Identification of Buffalo Breeds Using Self-Activated-Based Improved Convolutional Neural Networks. Agriculture 12:1386.

Saini G, Yadav V, Kumar S, Pandey AK (2021) Importance of indigenous cattle and peculiarity of their reproductive cycle: A review. Pharma Innov J 10: 156–159

Singh N, Tewari VK, Biswas PK, Dhruw LK, Pareek CM, Singh DH (2022) Semantic segmentation of in-field cotton bolls from the sky using deep convolutional neural networks. Smart Agr Technol 2:100045

Wang CH, Huang KY, Yao Y, Chen JC, Shuai HH, Cheng WH (2022) Lightweight Deep Learning: An Overview. . IEEE Consum Electr M:1–12

Weber F de L, Weber VA de M, Menezes GV, Oliveira Junior AS, Alves DA, de Oliveira MVM, Matsubara ET, Pistori H, Abreu UGP (2020) Recognition of Pantaneira cattle breed using computer vision and convolutional neural networks. Comput Electron Agric 175:105548