



Udderly accurate: A deep learning based modelling for determination of dairyness of Sahiwal cow using computer vision

NARESH KUMAR DAHIYA^{1✉}, SHALINI BHASKAR BAJAJ¹, A P RUHIL² and VIVEK JAGLAN³

AMITY University, Gurugram, Haryana 122 413 India

Received: 1 September 2023; Accepted: 19 October 2023

ABSTRACT

Dairy farming is a crucial agricultural practice for food and nutritional security. Selection of the best milching animals has always been a challenging task for optimising production efficiency. The efficacy of milk yield production is contingent upon a number of factors including inherent linear traits. Linear traits refer to the quantifiable physical characteristics that are associated with the production and reproduction capabilities of dairy animals. This article presents an innovative approach for the classification of Sahiwal cows into high, medium and low yielder categories based on images featuring linear traits through emerging deep learning and computer vision techniques. A large dataset of 4110 images highlighting important linear traits such as udder size, shape, and texture of different categories of Sahiwal cows has been created for training, validation and testing of the model. Images were collected from the herd of Sahiwal cows maintained at the National Dairy Research Institute, Karnal. The dataset was pre-processed using image augmentation techniques to enhance the model's robustness. Different architectures of CNN models namely InceptionV3, ResNet50 and GoogleNet were trained and optimised. The Inception V3 model demonstrated the best result with 85.64% testing accuracy among all these models in classifying the cow. The developed model can be used under field conditions to determine the dairyness of a cow in real time mode in place of human experts. Additionally, the model's interpretability is evaluated through feature visualisation, showcasing the importance of different udder features in milk yield prediction.

Keywords: CNN, Computer vision, Cows, Deep learning, Dairyness, Dairy farming, Linear traits, Sahiwal, Udder features

Agriculture, especially dairy farming, faces the pressing challenge of enhancing income while using fewer resources efficiently. In dairy farming, careful selection of animals plays a pivotal role in optimising production and resource management. Farmers initially relied on assumptions about physical traits associated with an animal's dairy capabilities, referred to as morphological or linear-type attributes. These traits are identifiable features on the animal's body and are closely linked to milk production (Brotherstone 1994). The International Committee for Animal Recording has recognized 18 such traits for categorising cows. Scientific studies by Brotherstone (1994), Berry *et al.* (2004), Khan *et al.* (2016) have established the significant influence of linear traits on production and reproduction, making them valuable for animal classification. Among these traits, udder characteristics, such as rear udder height, udder length, width, and teat distances, are strongly correlated with milk yield (Balhara *et al.* 2022). In modern dairy farming, these traits are essential for breeding and selection decisions (Bewley J 2009, Banhazi *et al.* 2012).

Accurately measuring linear traits is challenging, labour-intensive, and stressful for animals. Furthermore, human errors are possible despite the utmost care. Traditionally, experts visually inspect an animal's dairyness based on linear traits. However, such expertise takes years to develop and may not be readily available to every dairy farmer. Therefore, the need for a computer vision-based system (Alzubaidi *et al.* 2021) that can act as a virtual expert in assessing dairyness through images became apparent.

Advancements in computer vision technology (Krizhevsky *et al.* 2012, Hossain *et al.* 2022) and deep learning algorithms (Guo *et al.* 2016), particularly convolutional neural networks (CNN), offer promising solutions to this complex challenge, enabling the classification of agricultural products through visual inspection.

This research proposes the development of CNN-based computer vision models to classify Sahiwal cows into low, medium, and high yielders based on images of morphological traits related to udder characteristics. The Sahiwal breed was chosen for its high milk productivity, low maintenance, disease resistance, and adaptability to India's hot and dry climates. This model will empower stakeholders to select suitable cows for their herds,

Present address: AMITY University, Gurugram, Haryana.
²ICAR-NDRI, Karnal, Haryana. ³DPGITM, Gurgaon, Haryana.
✉Corresponding author email: naresh.dahiya28@gmail.com

contributing to improved dairy farming efficiency.

MATERIALS AND METHODS

Data collection: Among all the linear traits, udder traits are the most influential in assessing milk productivity. Therefore, these traits were diligently considered during the image-capturing process for dataset creation. Images were collected from three distinct perspectives covering various udder traits: the back view, the side view of the udder, and the bottom view of the udder. The back view (Fig. 1A) of the cow captured rump width, rear udder width, rear leg view; the side view (Fig. 1B) covered udder balance, teat length, naval udder distance, udder height, udder depth, udder attachment; and the bottom view (Fig. 1C) captures teat placement, milk veins. This comprehensive approach ensured that the dataset encompassed essential aspects for developing a reliable model. Different types of commonly available user-friendly handheld mobile devices were used for image capturing for the robustness of the model. Images were captured just before milking to get clear representations of the udder traits. To ensure diversity within the dataset, images were clicked in an open area as well as inside the milking area while maintaining a consistent distance between the camera and the cow. Additionally, the height of the cameras was adjusted proportionally to the animal's height. A total of 4110 images were captured for this study. The images were collected from the herd of Sahiwal cows maintained at the cattle yard of the institute.

Platform for development: Convolutional Neural Network (CNN) architecture of the Deep learning framework of MatLab-2021 was used for the development



Fig. 1. Images from the data set prepared for the classification of Sahiwal cows based on linear traits. A. rear view; B. udder side view; C. bottom view.

of the classification model. A workstation equipped with a powerful NVIDIA GeForce GTX 1080 Ti GPU (Graphics Processing Unit) was employed considering the high requirement of computational intensity for training and testing of CNN models.

CNN model: Convolutional Neural Network (CNN), a type of deep learning model Hossain M E *et al.* (2022) specifically designed for processing and analysing images has been used to develop a classification model for determining the dairyness of Sahiwal cows. By stacking convolutional layers, activation functions, pooling layers, and fully connected layers, CNNs can learn hierarchical representations of visual data, extracting relevant features at different levels of abstraction. This makes CNNs powerful for tasks such as image classification object detection, segmentation, and more.

The CNN starts with one or more convolutional layers. Each layer applies a set of learnable filters (kernels) to the input image. The filters convolve over the image, performing element-wise multiplication and summation to produce feature maps. These feature maps capture different patterns present in the input image, such as edges, textures, or shapes. After the convolution operation, an activation function (e.g. ReLU) is applied element-wise to introduce non-linearity and help the model learn complex relationships in the data. Pooling layers follow the convolutional layers. They downsize the feature maps by selecting the maximum value (max pooling) or average value (average pooling) within a small window. Pooling reduces the spatial dimensions, retaining the most relevant information while minimizing computational complexity. After passing through several convolutional and pooling layers, the feature maps are flattened into a one-dimensional vector. This flattening step is necessary to connect the CNN's convolutional layers with fully connected layers for output. The basic architecture of CNN consisting of various layers is shown in Fig. 2.

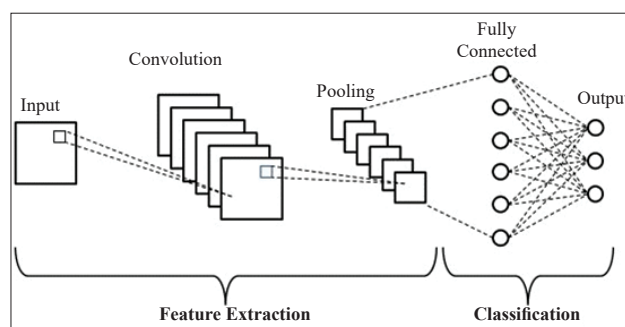


Fig. 2. Basic architecture of CNN.

Data pre-processing: The collected images (in RGB format) were pre-processed to ensure uniformity of images in the data set. Data augmentation techniques, such as random rotations, horizontal flips, cropping, translating and brightness adjustments were applied to augment the dataset and enhance its generalisation capability. Unwanted objects were removed from the images to reduce the size without losing any information at the initial stage. The images were

resized according to the requirements of the CNN model. Every image in the dataset was annotated with the category of animal vis-a-vis high, medium and low yielder.

Model development: Various classification models for determining the dairyness of Sahiwal cows were developed based on the most three popular and well-established architectures of CNN namely InceptionV3, ResNet50, and GoogleNet. These architectures were used because they are already pre-trained on a large data set and have shown remarkable depth and exceptional performance in image recognition tasks. Transfer learning approach was used for the development of CNN classification models. This approach passes on the benefits of utilising the pre-trained layers knowledge and generalizes well to the limited data set. The last fully connected layer of a pre-trained model is removed and replaced with a new layer incorporating softmax activation for facilitating multi-class classification suiting to the classification problem in the present work. The rest of the model remains frozen with already predefined model parameters. The InceptionV3, ResNet50 and GoogleNet CNN models were fine-tuned on the dataset of Sahiwal cow images using deep learning framework. The output layer was designed with three classes namely high, medium and low yielders.

The classification models were trained and validated on the collected data set. The data set was stratified into training and test data subsets. The training set constituted 80% of the data set (3288 train images) and the validation set comprised 20% (822 images). The partitioning was executed randomly, ensuring equitable representation records in each subset of classification categories, i.e. high, medium and low yielder. The training set is used to train the model fine-tune hyperparameters and monitor the model's performance. The test set was used to evaluate the final model. During the training process, different combinations of training parameters were explored including the Stochastic Gradient Descent (SGD) optimizer and Adam optimizer, with varying learning rates as 0.001, 0.01, and 0.1, as well as varying batch sizes of 8, 16, and 32. The training process was iterated over 10 and 100 epochs to ensure extensive learning and model refinement.

Performance evaluation

In order to evaluate the performance of the developed models following performance metrics were used:

Confusion matrix: It is a cross table that records the number of occurrences with respect to actual verses prediction classification.

Accuracy: It measures the ability of a model to correctly predict the objects on the entire data set. It is directly computed from the confusion matrix as the proportion of the sum of numbers of correctly predicted outcomes (true positive) of each class out of total data (eq. 1):

$$\text{Accuracy} = \frac{\text{Sum of true positives predicted in each class}}{\text{Total number of objects classified}} \quad (1)$$

Precision: Precision measures the trustworthiness of the

model when it predicts an individual object of interest as positive, precision is computed class-wise as the proportion of a number of objects predicted correctly (true positive) in a particular class out of a total number of objects predicted by the model in that class (eq. 2):

$$\text{Precision}_k = \frac{TP_k}{\text{Total number of objects predicted by model for Class } k} \quad (2)$$

TP_k = Number of true positive cases predicted by the model for class k.

Recall: This metric measures the model's predictive accuracy for positive cases for a particular class. In the multiclassification model, recall is computed class-wise as the proportion of a number of true positive objects predicted by the model for a particular class out of a total number of actual objects in that particular class (eq. 3).

$$\text{Recall}_k = \frac{TP_k}{\text{Actual number of objects in Class } k} \quad (3)$$

RESULTS AND DISCUSSION

CNN models, viz. InceptionV3, ResNet50 and GoogleNet were trained using the training set (80%) and predictive accuracy and robustness were validated using the validation set (20%) on the data set of Sahiwal cow images. The output layer of these networks was set to three classes namely high, medium and low yielders. The training and test data set had almost equal number of observations in each class for proper training and testing. The models were trained by varying momentum and learning rate as 0.001, 0.01, and 0.1, as well as batch size as 8, 16, and 32 to get the optimised results. The optimiser stochastic gradient descent with momentum and a learning rate of 0.001 gave the best result. Results of these models, i.e. InceptionV3, ResNet50 and GoogleNet for test dataset were recorded in the form of a confusion matrix along with class-wise precision and recall matrices shown in Tables 1, 2, 3, respectively. The overall accuracy of models is tabulated in Table 4.

The efficacy of these models was evaluated on the performance metrics precision, recall and accuracy computed based on the confusion matrix. It can be observed from Table 4 that the accuracy of all these models is more than 60%. However, among all models, InceptionV3 has emerged as the best performer in terms of overall accuracy with 85.64% in identifying the cow images correctly across all classes. This may be because it is the latest model among all three models and has extracted more important features from the images due to its use of residual blocks, which enable the training of very deep neural networks. Also, the skip connections in ResNet50 allow the gradients to flow more directly through the network during training, addressing the vanishing gradient problem.

The in-depth analysis of class-wise precision of each model (Tables 1, 2 and 3) reveals that the InceptionV3 model is the most trustworthy among all models and has shown consistency with precision varying class-wise from 82% to 90%. This means that the Inception model predicts

Table 1. Confusion matrix for Inception V3

		Inception V3 (predicted class)			Total	recall
		1	2	3		
Actual class	1	230	19	25	274	0.8394
	2	50	224	0	274	0.8175
	3	0	24	250	274	0.9124
Total		280	267	275	822	
Precision		0.8214	0.8389	0.9091		

(Class 1: High; 2: Medium; and 3: Low Yielder)

Table 2. Confusion matrix for ResNet50

		ResNet50 (predicted class)			Total	Recall
		1	2	3		
Actual class	1	210	40	24	274	0.7664
	2	60	155	59	274	0.5656
	3	0	74	200	274	0.7299
Total		270	269	283	822	
Precision		0.7777	0.5762	0.7067		

(Class 1: High; 2: Medium; and 3: Low Yielder)

Table 3. Confusion matrix for GoogleNet

		GoogleNet (predicted class)			Total	Recall
		1	2	3		
Actual class	1	180	40	54	274	0.6569
	2	50	150	74	274	0.5474
	3	54	50	170	274	0.6204
Total		284	240	298	822	
Precision		0.6338	0.625	0.5704		

(Class 1: High; 2: Medium; and 3: Low yielder)

Table 4. Overall accuracy of CNN models

Model	Accuracy
InceptionV3	85.64%
ResNet50	68.73%
GoogleNet	60.82%

the maximum true positive cases in each class predicted by the model. The maximum precision in this model is 90.9% for class 3, i.e. for low yielder class. It also indicates that the model is identifying the low yielder animal images very well with high accuracy. For ResNet50, precision fluctuates between 57% to 77%, and for GoogleNet, it ranges from 57% to 63%. Which is significantly low performance in accurately predicting individual images as true positives. The occurrence of class-wise false positive cases is higher, resulting in a precision range of 55% to 77% for both models. It implied that the performance of these two models is not reliable in predicting individual images as true positive, i.e. the class-wise false positive cases are higher therefore, these models cannot be trusted in the class-wise prediction of cow images. The class-wise analysis of the recall metric for each model divulges that Inception V3 models are highly capable of finding the true positive cases in the actual class data set with a maximum recall value of 91.24% for class 3. This means that the Inception model

can identify the images of low-yielder cows within the actual class dataset with 91% accuracy. The recall values in other models are much lower in comparison to the InceptionV3 model. This shows that the Inception model works very well in classifying the low yielder animals with more than 90% accuracy within the actual (recall) as well as predicted (precision) class. This will make farmers more confident about the selection of animals and accordingly can invest wisely. This work is much more familiar to the human approach of classification, whereas Shorten (2021) used deep learning and try to find out the volume of udder before and after milking in order to find out the milk yield of the dairy cow and Afridi Mohib *et al.* (2022) used the VGG-16 to find out whether udder is in the image captured or not. This study aligns closely with traditional human approaches to the classification of Sahiwal cows. Shorten (2021) employed deep learning techniques to quantify udder volume before and after milking, aiming to determine milk yield in dairy cows. On the other hand, Afridi Mohib *et al.* (2022) utilised the VGG-16 model to ascertain the presence of the udder in captured images. Consequently, the author's work stands out as a distinct and closely analogous approach to the human methodology.

The authors used Grad-CAM (Gradient-weighted Class Activation Mapping) to gain insight into the model during training. Grad-CAM visualises the regions of the input image that are "important" for the model's predictions. The results of Grad-CAM are shown in Fig. 3. Left hand side images in the figure are the input images given to the model whereas the right-hand side images are the corresponding input images with some red and blue colours, showing a significant part of the image contributing most in the model's prediction. From this figure, it may be observed that the model is only considering important parts of the

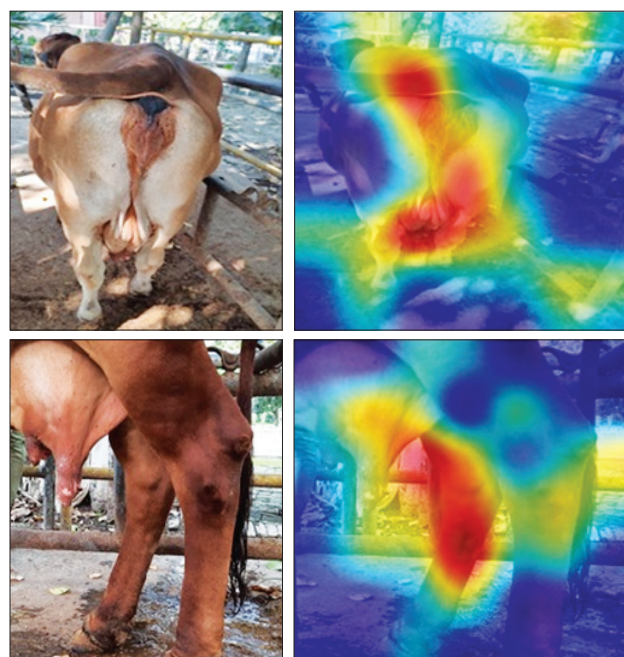


Fig. 3. Grad-CAM technique highlighting the important features of the image.

images for its prediction, and not the surrounding area or other objects. This technique has shown that the network is looking at the correct patterns in the image for classification which has built the confidence of authors in the developed model.

In conclusion, the performance of the InceptionV3 model with 85% accuracy was found best among all models. The present work has established a benchmark towards further studies in the determination of dairyness of Sahiwal cows through images. The results can be further improved by augmenting more images in the dataset. The developed model can be used in the absence of experts in identifying the best animals in the field conditions. Similar work can be extended to other breeds of cattle and buffalo.

REFERENCES

- Afridi M, Ullah Øy Nordbø F and Alaya C. 2022. Deep learning based udder classification for cattle traits analysis. *Electronic Imaging* **34**: 390–1.
- Alzubaidi L, Zhang J and Humaidi A J. 2021. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data* **8**(53).
- Balhara S, Balhara A K, Dahiya N, Singh R P, Ruhil A P and Himanshu. 2022. An algorithm-based approach for identification of most relevant linear traits for selecting high producing Murrah buffaloes. *Indian Journal of Animal Sciences* **92**(5): 618–23.
- Banhazi T M, Lehr H, Black J L, Crabtree H, Schofield P, Tschärke M and Berckmans D. 2012. Precision livestock farming: An international review of scientific and commercial aspects. *International Journal of Agriculture and Biological Engineering* **5**(3): 1–9.
- Berry D P, Buckley F, Dillon P, Evans R D and Veerkamp R F. 2004. Genetic relationships among linear type traits, milk yield, body weight, fertility and somatic cell count in primiparous dairy cows. *Irish Journal of Agricultural and Food Research* **43**:161–76.
- Bewley J. 2009. *Precision Dairy Farming: Opportunities, Challenges and Solutions*. The Dairy Practices Council, 40th annual conference held at Latham, NY.
- Brotherstone S. 1994. Genetic and phenotypic correlations between linear type traits and production traits in Holstein-Friesian dairy cattle. *Animal Production* **59**(2): 183–87.
- Guo Y, Liu Y, Oerlemans A, Lao S, Wu S and Lew. 2016. Deep learning for visual understanding: A review. *Neuro computing* **187**: 27–48.
- Hossain M E, Kabir M A, Zheng L, Swain D L, McGrath S and Medway J A. 2022. Systematic review of machine learning techniques for cattle identification: Datasets, methods and future directions. *Artificial Intelligence in Agriculture* **6**: 138–55.
- Khan M A and Khan M S. 2016. Genetic and phenotypic correlations between linear type traits and milk yield in Sahiwal cows. *Pakistan Journal of Agricultural Sciences* **53**(2): 483–89.
- Krizhevsky A, Ilya S and Geoffrey E H. 2012. ImageNet classification with Deep Convolutional Neural Networks. NIPS Proceedings of the 25th International Conference on Neural Information Processing Systems **1**: 1097–105.
- Shorten P R. 2021. Computer vision and weigh scale-based prediction of milk yield and udder traits for individual cows. *Computers and Electronics in Agriculture* **188**.