Body Condition Score Classification of Dairy Cows Using Images with CNN Models

Sooram Kang
Dept. Mathematics and Statistics, Chonnam National University,
61186, S. Korea
slkang2001@gmail.com

Youngsik Park
Bigdatainsight, 61186, S. Korea
wpeak09@gmail.com

JaeBuhm Chun
Rural Development Administration, 54875, S. Korea
jbchun01@korea.kr

Myung Hwan Na
Dept. Statistics, Chonnam National University, 61186, S. Korea nmh@chonnam.ac.kr

ABSTRACT

Per capita milk consumption in Korea has been steadily declining over the past 20 years. As a result, the number of dairy farms and cows is also declining. Conversely, the number of cows per farm has increased significantly. Due to a decrease in consumption and an increase in the price of raw materials, it is necessary to improve management efficiency to increase the income of dairy farmers. For this purpose, it is important to check and predict information such as milk production capacity and health of cows. BCS is used as a scoring tool for dairy cows in farms based on the appearance of them. However, since it is difficult for individual cows to be continuously scored by experts, it is difficult to use them in real farms. This study proposes a model for discriminating BCS of individual cow using images. By acquiring images with a camera installed in a fixed position, the BCS of individual cow can be determined through an image analysis model. As the image model, Resnet, Inception, and VGG were used, and BCS-rated cattle and their images were trained. As a result, the VGG19 model performed the best with 90% accuracy.

KEYWORDS

BCS, Dairy Cow, Smartfarm, Image Analysis, CNN

1 INTRODUCTION

As per capita milk consumption has decreased over the past 20 years, the number of dairy farms and cows has also been continuously decreasing. According to Statistics Korea, the number of dairy farms in 2002 was 11,716 and the number of cows was 543,587. In the second quarter of 2022, the number of dairy farms halved to 6,056. And the number of cows decreased to 389,125. However, the number of cows per farm increased rapidly from 46 to 64. For this reason, dairy farming, unlike other livestock industries, has been largely automated.

Even thought, the continuous decrease in milk consumption and the increase in raw material prices make it difficult for dairy farms to achieve business performance, and it is difficult to avoid the decline and aging of the labor force in the livestock industry as well as agriculture. Therefore, there is a need for a method that can increase productivity per cow and do this automatically while predicting. BCS is being used as a tool to measure the health and productivity of cow. BCS is scored through visual and tactile assessment by experts, and its accuracy and usefulness have been confirmed by previous studies [1,2].

However, assessment by the expert is not only difficult to utilize BCS in actual farms, but also requires a continuous scoring for the growing period of individual cow. In this study, an analysis was performed to estimate BCS of cow using images. Image analysis models were trained with labeled images of cows, and BCS was scored between 1 to 9 for each cow through the images. Image models were compared with Resnet50, Inception V3, and VGG19. As the models pre-trained through many images, the task can be performed with a small number of images.

2 EXPERIMENTAL METHOD AND DATA SET

2.1 Convolutional Neural Network

In the existing deep learning method, the input of an artificial neural network composed of only fully connected layers is limited to a one-dimensional form. However, image data has a three-dimensional structure including color. When image data is flattened in one dimension to learn through a fully connected layer, the learning effect is very poor due to information loss. This is because spatial information is important for image data analysis. By flattening the image data, the properties connected in chunks disappear.

Convolutional Neural Network (CNN) is an artificial neural network model that can solve these shortcomings and utilize spatial information. The CNN model consists of a part that extracts features of an image and a part that classifies the class. Feature extraction is a form in which a convolution layer and a pooling layer are stacked in multiple layers. The convolution layer finds a pattern by repeatedly applying a filter of a certain size to the entire image. The pooling layer serves to reduce the dimension of features extracted through the convolution layer. Since the two layers are stacked in multiple layers, the model can learn from the low-dimensional features of the image to the high-dimensional features.

In order to classify the image class using the features extracted through this, a flattening process that makes multidimensional features into one dimension is performed. A one-dimensional feature is learned in a fully connected layer, and class classification is performed. Various image analysis models have appeared using CNN, and models that are lighter and have better performance are continuously appearing. Transfer learning is to perform different tasks using a model that has learned a lot of image data while having many layers and deep structures. Transfer learning has a relatively good learning effect even if only a small amount of image data is used.

2.2 CNN Models

2.2.1 Resnet. CNN models before Resnet tried to create a deep structure, but simply increasing the number of layers did not improve the performance. This is because of the vanishing radiant problem, in which the influence of gradient decreases when performing backpropagation as the structure of the model deepens. The basic principle of Resnet is to use a structure in which an input x enters a layer and is sent out together with an output value, and it is called a residual block. Through this, learning becomes easier and converges better than learning the whole by learning only the output value, which is the residual effect. It has a deeper structure than previous models, but with improved complexity and performance.

2.2.2 Inception. Inception is a CNN model announced by Google, and if the existing CNN models have a structure in which convolution filters of the same size are continuously added to the input image one by one, Inception uses filters of several sizes at the same time. While reducing overall intra-network connections, detailed matrix operations are processed to be as dense as possible. As a result, the connectivity was reduced, but the total amount of computation increased due to the matrix operation. To reduce the amount of computation, Inception uses a 1x1 convolution layer.

2.2.3 VGG. The VGG model focused on the depth of the network and the impact on model performance. If the kernel size is large, the image size is reduced rapidly, making it difficult to create a deep layer. In addition, the number of parameters and the amount of computation are large. For this reason, VGG fixed the kernel size to 3x3. Using the 5x5 kernel once and using the 3x3 kernel multiple times output the same size feature map but using the 3x3 kernel several times showed better performance. This is because by using a small kernel, more ReLU functions can be used and more nonlinearities can be secured. VGG has a feature that requires a lot of computation with a deep structure.

2.3 Data set

An image data set created by Corfilac, IPLAB and the Penn Veterinary Medicine joint team was used for model training [3]. The data set consists of 207 images of 29 cows. The camera was installed at the exit of the milking robot, and all 207 images were taken at a fixed position. The resolution of the images is 704x576. The color, size and age of the cattle were not considered. The data set also has BCS scores for each image. BCS scores of each image was assigned by two experts. The position of the camera is fixed, but the position of each cow coming out of the exit is slightly different. Because the moving cow was photographed, the angle of the cow and the degree of coverage of the cow are slightly different. Also, there are many images in which the image resolution is not high, and the cow's edge is blurred. In general, it is known that image resolution and quality affect learning. Fig. 1 shows examples of data set

Figure 1: Examples of data set.

3 EXPERIMENTAL RESULT

3.1 Pre-processing

In order to use the data set for model training, two parts of the preprocessing process were performed. First, in the image part, the cow area is limited compared to the whole image, so to remove irrelevant information, each image was cropped at the same size. The image crop ratio is 4:3. The cropped area was slightly different depending on the position of the cow, but there was no significant difference in the background shown in each image. In the crop process, images with too large object sizes were excluded, and as a result, 204 images were used.

The BCS score for each image was given by two experts. Even though it is not a big difference, the BCS score of each image is slightly different depending on the expert. Therefore, the average of the two BCS scores was used. In general, there are two systems of BCS scores. One is a scoring system from 1 to 5 and the other is a system from 1 to 9. The data set used in this study consisted of a score between 1 and 5. A score of 1 indicates very skinny, and a score of 5 indicates obesity. However, there were only values between 2.5 and 4.5 in the data set, and no scores of 1 and 5 existed. Because of the small number of data and low resolution, the regression task was converted into a classification task to improve predictive power. Also, since the BCS scores are mostly concentrated at 3, scores from 1 to 5 were converted into scores from 1 to 9 in order to widen the distribution of scores and increase the number of classes. As a result, the BCS scores used to train the model widened from 5 to 9.

3.2 Model learning and test result

The models used for the analysis were Resnet50, Inception V3, and VGG19. The number of parameters of each model is 25.6M, 23.9M, and 143.7M, in the order of Inception V3, Resnet50, and VGG19. The input dimension of the three models is (224, 224, 3), and the cropped image is resized to 224x224 for training. All three models used weights trained in imagenet in advance. In each model, the last layer was replaced and two fully connected layers with 1000 and 500 nodes added, respectively. Relu function was used as the activation function and binary cross entropy was used as the loss function. The train/test ratio was 8 to 2, the epoch was 30, and the learning rate was 0.001. When dividing the train data and the test data, it was divided according to the data distribution for each class so as not to have an extreme ratio. The learning speed was fast in the order of Resnet50, Inception V3, and VGG19. Fig. 2 shows accuracy and loss graph by epoch.

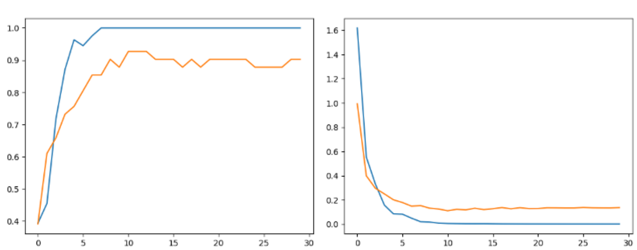

Figure 2: Accuracy graph by epoch on left, loss graph by epoch on right of VGG19 (blue is train and orange is test).

As a result of the experiment, the accuracy of the three models was high in the order of VGG19, Inception V3, and Resnet50. The accuracies of each model were 90% for VGG19, 80% for Inception V3, 63% for Resnet50. VGG19 with the best performance showed stable convergence from about 20 epochs, and the change in accuracy according to epochs was not significant. VGG19, which showed 90% accuracy, gave the correct answer for both 6 and 7 points out of a score of 5 to 9. On the other hand, the accuracy was 71% at 5 points and 75% at 7 and 9 points. In the case of 5 points, 6 and 7 points were answered in 14% of cases, respectively, and in the case of 8 points, 9 points were answered in 25% of cases. Conversely, in the case of 9 points, 25% of the answers were wrong with 7 points. The model showed an error of 10% but did not choose a different answer at all.

Fig. 4 is images in which VGG19 failed to classify correctly. Except for the first image, these images are not sharp. Also, the last picture is the one with the most tilted angle of the cow. If images with better resolution are collected through direct data collection, the predictive power will be improved.

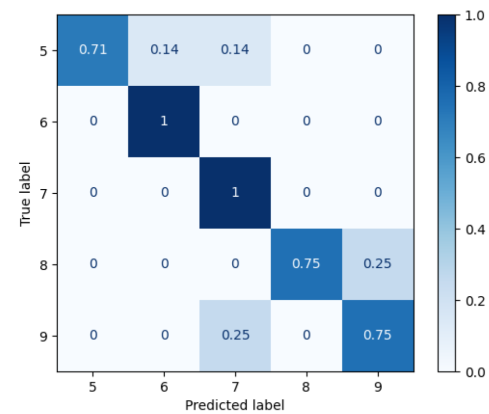

Figure 3: Confusion matrix for test data of VGG19.

Figure 4: Images for which VGG19 failed to correctly classify.

4 CONCLUSIONS

This study performed an analysis of estimating BCS, a tool for diagnosing the health and productivity of cows, using images. The data set used for the analysis has problems with low resolution and moving cattle, but has a BCS score judged by experts. To remove the background, the cattle area was cropped and the 5-point-based BCS score convert to 9-point-based as pre-process. The performance of three pre-trained CNN models was compared, and among them, VGG19 showed 90% accuracy. It is significant in that relatively high accuracy was obtained using 204 small image data. It is expected to enable quick and easy BCS scoring in an uncomplicated manner using images with high resolution and fixed positions of cattle through direct collection in the future.

ACKNOWLEDGMENTS

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry(IPET) through Smart Plant Farming Industry Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA)(421017-04).

This research was supported by the BK21 FOUR (Fostering Outstanding Universities for Research, NO.5120200913674) funded by the Ministry of Education(MOE, Korea) and National Research Foundation of Korea(NRF)

REFERENCES

[1] I.A. Wright, A.J.F. Russel. 1984. Partition of fat, body composition and body condition score in mature cows. *Animal Production.* 38 (1984), 23-32.

[2] F. Buckley, K. O'Sullivan, J.F. Mee, R.D. Evans, P. Dillon. 2003. Relationships among milk yield, body condition, cow weight, and reproduction in spring-calved Holstein-Friesians. *J Dairy Sci.* 86(7), (2003), 2308-19.

[3] G Azzaro 1, M Caccamo, J D Ferguson, S Battiato, G M Farinella, G C Guarnera, G Puglisi, R Petriglieri, G Licitra. 2011. *Journal of dairy science.* 94, (2011), 2126-37.

[4] Huang X, Hu Z, Wang X, Yang X, Zhang J, Shi D. 2019. An Improved Single Shot Multibox Detector Method Applied in Body Condition Score for Dairy Cows. *Animals.* (2019) 23;9(7):470.

[5] Dong Liu, Dongjian He, Tomas Norton. 2020. *Biosystems Engineering*. 194, (2020),.