mathematics

MDPI

Article

Optimized Deep-Learning-Based Method for Cattle Udder Traits Classification

Hina Afridi [1,2], Mohib Ullah [1,*], Øyvind Nordbø [2,3], Faouzi Alaya Cheikh [1] and Anne Guro Larsgard [2]

[1] Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjøvik, Norway
[2] Geno SA, Storhamargata 44, 2317 Hamar, Norway
[3] Norsvin SA, Storhamargata 44, 2317 Hamar, Norway
* Correspondence: mohib.ullah@ntnu.no

Abstract: We propose optimized deep learning (DL) models for automatic analysis of udder conformation traits of cattle. One of the traits is represented by supernumerary teats that is in excess of the normal number of teats. Supernumerary teats are the most common congenital heritable in cattle. Therefore, the major advantage of our proposed method is its capability to automatically select the relevant images and thereafter perform supernumerary teat classification when limited data are available. For this purpose, we perform experimental analysis on the image dataset that we collected using a handheld device consisting of a combined depth and RGB camera. To disclose the underlying characteristics of our data, we consider the uniform manifold approximation and projection (UMAP) technique. Furthermore, for comprehensive evaluation, we explore the impact of different data augmentation techniques on the performances of DL models. We also explore the impact of only RGB data and the combination of RGB and depth data on the performances of the DL models. For this purpose, we integrate the three channels of RGB data with the depth channel to generate four channels of data. We present the results of all the models in terms of four performance metrics, namely accuracy, F-score, precision, and sensitivity. The experimental results reveal that a higher level of data augmentation techniques improves the performances of the DL models by approximately 10%. Our proposed method also outperforms the reference methods recently introduced in the literature.

Keywords: dairy industry; udder classification; supernumerary teats; udder conformation traits; milk production; deep learning models; neural networks; cattle images

MSC: 68T01

Citation: Afridi, H.; Ullah, M.; Nordbø, Ø.; Cheikh, F.A.; Larsgard, A.G. Optimized Deep-Learning-Based Method for Cattle Udder Traits Classification. *Mathematics* **2022**, *10*, 3097. https://doi.org/10.3390/math10173097

Academic Editor: Jakub Nalepa

Received: 31 July 2022
Accepted: 19 August 2022
Published: 29 August 2022

1. Introduction

The dairy industry is contributing substantially to meet the accelerating food demand of the world. For instance, dairy products are consumed by 80% of the population of the world [1]. To offer dairy products to a steady growing world population without increasing the land use connected to feed, the efficiency of dairy cows should be improved. Breeding organizations use genetic selection for increasing milk yield, decreasing the need for feed and improving the longevity of cows for enhancing resource efficiency. The cow's udder conformation plays an important role for the longevity of dairy cows. A robust, easy-to-milk long-lasting udder is preferred, and breeding organizations put a lot of emphasis on such traits in the breeding goals [2–4]. The udder of a cow characterized by good udder conformation is consistent in size and capacity relative to the cow's age and number of lactations. Good udder conformation is also represented by teats which are appropriate in number, size, length, and squarely placed under the quarter and perpendicular to the ground. For example, supernumerary teats are in excess of the normal number of teats. They are not capable of producing milk. In fact, they represent a typical congenital abnormality [5]. We depict both udders characterized by good conformation and supernumerary teats in Figure 1.

Figure 1. Udder conformation and supernumerary teats examples. The first column shows an udder characterized by good conformation. The second column shows an udder representing a supernumerary teat, and the last column shows an udder representing two supernumerary teats annotated with green circles.

In the dairy industry, the traits related to udder conformation are scored manually by a trained technician. Most traits are scored on a linear discrete scale ranging from 1 to 9, where 1 and 9 correspond to the most extreme cases [6]. However, early life measurements may not reflect the same quality later in life. Therefore, scoring udder conformation repeatedly on a cow might give a better and more objective reflection of the true udder conformation [7]. To study changes of udder conformation over long period of time, repeated observations are required, which is a costly and time consuming process. Developing a more automatic data collection and analysis method for such traits would directly increase the prediction, contribute to genetic gain, and reduce the associated cost. To meet these challenges, dairy industries can collect data related to udders using sensors and cameras. The data can be analyzed by deep neural networks (DNNs) to automate the decision-making processes. DNNs are trainable multi-layer models consisting of multiple feature-extraction phases, succeeded by a fully connected classification stage. Recently, DNNs have presented encouraging performance in many applications, including behavior recognition of pigs and cattle [8], analysis of animal migration patterns [9], and modeling animal biodiversity [10]. These approaches have been depicted to have promising generalization capabilities when properly trained on available data. This evolution has therefore led to a notion that DNNs could lead to better performance regarding the analysis of udder conformation and the related traits.

In this paper, we propose optimized DNNs for the classification of udder\non-udder images and supernumerary teats. Our proposed DNNs architectures could also be used to classify whether a cow has supernumerary teats or not. The classification based on cow imaging is a very challenging problem because, in addition to the udder characteristics, the analysis is sensitive to deformations due to image capture conditions. We propose two main approaches when considering DNNs. Firstly, we develop two customized DNN architectures characterized by a few layers to perform the classification. This approach provides high flexibility in terms of modeling. Secondly, we explore two existing DNN architectures, namely the VGG16 model [11] and the wide residual network (WideResNet) [12]. This approach reduces the time to develop new architectures. The VGG16 model and the WideResNet model are well-known networks used for many applications. To perform the classification using different DNNs, we collected both RGB and depth data. We used the data both separately and jointly. In fact, fusing depth information with RGB data for image-based udder classification can help to increase the performance. We investigate the optimal manner to perform this data fusion considering different DNNs architectures. We carried out comprehensive experiments to determine the impact of considering data separately and jointly. We also investigate the impact of image variations on the performances of DL models in terms of different data augmentation techniques. We experimentally highlight the behavior and characteristics of DL models by taking into account the data

augmentation techniques including blurring, flipping, rotation, brightness, and contrast. The main contributions of this paper are:

- To the best of our knowledge, we are the first to propose optimized DL models for udder\non-udder and supernumerary teat classification. We also investigate famous DL methods in conjunction with our proposed method.
- We collect and present our own dataset that can be used to analyze udder conformation and udder traits of cows. Our new dataset will open a new research direction since no such dataset is available publicly.
- We present comprehensive experiments to evaluate the impact of depth information and different data augmentation techniques on the performances of the DL models.
- We present detailed experiments from many aspects and comparisons with famous DL methods and other reference methods.

The rest of the paper is organized as follows. In Section 2, an overview of related work is presented. The details about our proposed and existing DL models are presented in Section 3. Experimental results on our collected dataset are shown in Section 4. We present the discussion in Section 5 and conclusions in Section 6.

2. Related Works

We divide the literature into two categories: classic methods and DL-based methods. This classification is motivated by the fact that classic methods are related to cattle traits in general. However, DL-based methods, recently published, are mostly related to cattle health.

In the first category, Kappes et al. [13] used multivariate analysis to investigate the functional traits including lameness score, udder cleanliness score, and udder depth. The purpose was to understand the connection between functional traits and milk production. However, they did not take into account the genetic parameters and the breeding values that represent an estimate of an animal's genetic merit for a particular trait. To address this issue, De et al. [14] used Bayesian inference to estimate breeding values considering two-trait analyses, and Carvalho et al. [15] estimated genetic parameters of different conformation and management traits. Furthermore, Shorten [16] discovered that the milk yield is strongly associated with udder traits. Therefore, the author modeled 3D images using a machine learning approach to determine individual cow milk yield and the lactation milk yield. The udder volume before milking, udder volume after milking, and the difference in udder volume before and after milking were estimated. It was found that the machine learning method could potentially be used for on-farm prediction of milk performance and udder traits from dairy cows. Understanding milk production yield and cattle longevity from pedigree is a complex process. Therefore, researchers also studied candidate genes and genetic basis in cattle. For example, Jiang et al. [17] and Colinet et al. [18] introduced a Bayesian approach to identify candidate genes for production, reproduction, and conformation traits. Their approach improved the understanding of the genetic basis of cattle complex traits. Stefani et al. [19] used a probabilistic approach to consider udder, feet, and leg traits for genetic gains. They estimated the heritabilities of the udder, feet, and leg traits. They found the possibility to promote early genetic gains in the longevity of animals by choosing slightly correlated traits to achieve a reasonable set of feet and legs, as well as well-positioned teats of medium length. Poppe et al. [7] investigated changes in udder conformation genetically and phenotypically within and between parities. They estimated genetic correlations between automatic milking systems (AMS)-based udder conformation traits and classifier-based udder conformation traits, longevity, and udder health. Their work concluded that udder conformation hardly transforms genetically between parities and is greatly repeatable within parities.

In the second category of DL-based methods, the literature is mostly related to cattle health management. The use of DL-based methods in different domains is very wide. For example, Manzo and Pellino [20] explored pretrained DL models for cloud image description and classification. Shanthamalluet al. [21] presented different DL applications.

Regarding cattle, Porter et al. [22] explored a DL framework to monitor teat tissues which is a challenging problem for udder health management in dairy cows. They found that their method can be exploited to evaluate teat-end condition in a systematic and proper manner. However, they did not take into account the mastitis disease which is a significant economic and health problem in dairy farms. To address this challenge, Ebrahimi et al. [23] investigated a method for early sub-clinical mastitis detection. They exploited DL-based methods to identify the patterns of risk factors of mastitis. They evaluated the method on a large dataset of milk recordings to decode important models for the detection of sub-clinical mastitis. To further improve the detection speed and achieve automatic recognition of dairy cow mastitis, Xudong et al. [24] used a DL network based on the bilateral filtering enhancement of thermal images. To identify mastitis in the early stage, Fadul et al. [25] used different machine learning techniques to develop predictive and prescriptive decision support tools to identify cows positive for clinical mastitis during their first lactation. They found that two different methods working concomitantly, one for predicting the imminent risk and the other one for the overall risk during the first lactation, could help in the short-term, mid-term, and long-term decision-making process. Cattle health management can be further improved by considering heritability that measures the strength of the relationship between performance and breeding value. For this purpose, Nye et al. [26] used a composite DL-based method to extract phenotypic information for morphological features to evaluate conformation characteristics. They used pedigree and image information to estimate high heritabilities considering useful biological data.

Our literature review reveals that the previous research is either focused on different cattle traits using classic methods or cattle health considering DL-based methods. Moreover, the literature is very limited in the case of DL methods. Therefore, there is a research gap to understand and explore cattle udder conformation traits. Our work falls into the second category of DL-based methods. Therefore, our work also contributes toward the expansion of DL-based methods for conformation traits in cattle.

3. Materials and Methods

We note that limited research has been carried out to explore DL-based methods for predicting conformation traits in dairy cattle. Here, we propose two approaches to explore DL-based methods. In the first approach, we develop two custom DL networks that are flexible in term of modifying their depths and other hyperparameters. The first network, depicted in Figure 2, consists of three convolutional layers, three maxpooling layers, and two fully connected layers. The second network, depicted in Figure 3, consists of four convolutional layers, four maxpooling layers, and two fully connected layers. Considering the convolutional layers, our networks extract convolutional features from the input by exploiting square filters to produce feature maps. Each filter is represented by different weight parameters which are learned during the training stage. Maximum pooling is an operation to find the largest value in each patch of each feature map. This operation underlines the most present feature in the patch. In fact, features tend to capture the spatial presence of various patterns in the stack of the feature map. The input to the fully connected layer is the output from the final convolutional and maxpooling layers representing high-level features. To illustrate the operation of a convolutional layer using filters, it is formulated in Equation (1).

$$Conv(o^{[l-1]}, S^{[n]})_{x,y} = \Omega^{[l]}\Big(\sum_{i=1}^{\eta_H^{[l-1]}}\sum_{j=1}^{\eta_W^{[l-1]}}\sum_{s=1}^{\eta_C^{[l-1]}} S^n_{i,j,s} o^{[l-1]}_{x+i-1,y+j-1,s} + b_n^{[l]}\Big) dimen(Conv(o^{[l-1]}, S^{[n]})) = (\eta_H^{[l]}, \eta_W^{[l]}) \quad (1)$$

where $o^{[l-1]}$ represents the output from the previous convolutional layer $(l-1)$ which is input for the current layer l. $S^{[n]}$ shows the number of filters to be applied to the input. The dimension of the input in terms of height, width, and number of channels is represented by $\eta_H^{[l-1]}$, $\eta_W^{[l-1]}$, and $\eta_C^{[l-1]}$, respectively. The rectified linear activation function is shown

as $\Omega^{[l]}$. Therefore, the output $o^{[l]}$ from layer l and its dimension by applying filters $S^{[n]}$ is formulated as in Equation (2),

$$o^{[l]} = \Omega^{[l]}(Con(o^{[l-1]}, S^{(1)})), \Omega^{[l]}(Con(o^{[l-1]}, S^{(2)})), \Omega^{[l]}(Con(o^{[l-1]}, S^{n_C^{[l]}}))dimen(Conv(o^{[l]}) = (\eta_H^{[l]]}, \eta_W^{[l]}, \eta_C^{[l]}) \quad (2)$$

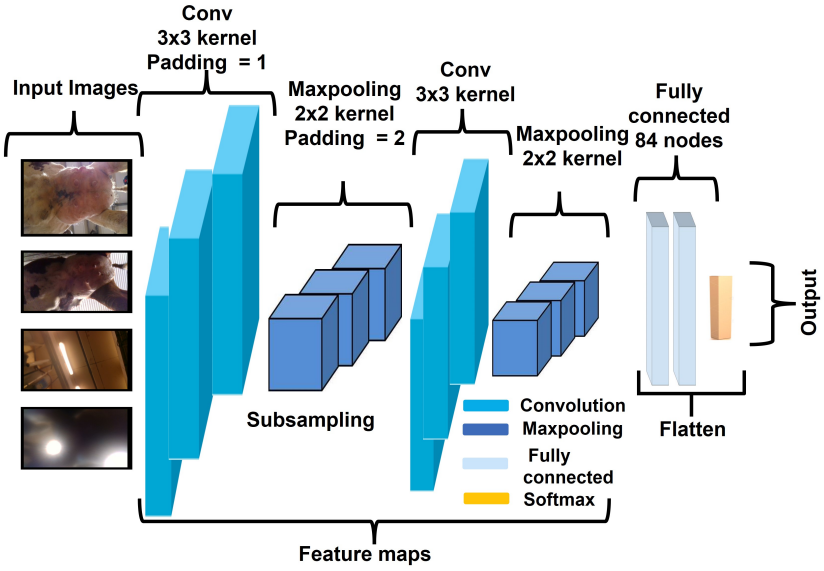

Figure 2. Our proposed network 1. We depict a convolutional neural network (CNN) in terms of input, convolutional, maxpooling, fully-connected, and softmax layers. The types of layers are the same both in our proposed and existing convolutional neural networks. We also provide information related to the kernel size, padding, and subsampling.

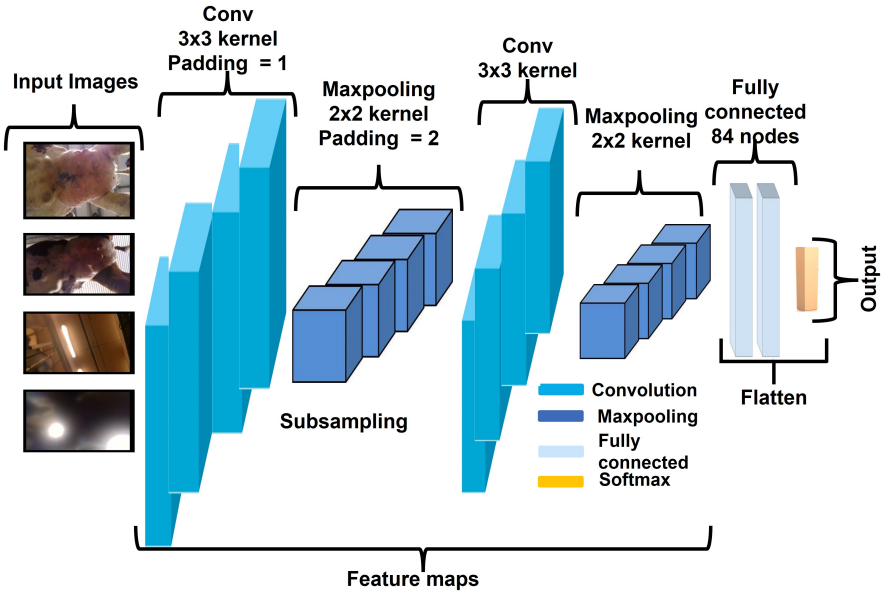

Figure 3. Our proposed network 2. We depict a convolutional neural network (CNN) in terms of input, convolutional, maxpooling, fully-connected, and softmax layers. The types of layers are the same both in our proposed and existing convolutional neural networks. We also provide information related to the kernel size, padding, and subsampling.

A maxpooling layer downsizes the features of the input. Therefore, the output from the layers and its dimension are formulated as,

$$Maxpool(o^{[l-1]})_{x,y,z} = \Theta^{[l]}((o^{[l-1]}_{x+i-1,y+j-1,z})_{(i,j)\in[1,2,\ldots,f^{[l]}]})dimen(o^{[l]}) = (\eta_H^{[l]}, \eta_W^{[l]}, \eta_C^{[l]}) \quad (3)$$

where $\Theta^{[l]}$ represents the maxpooling operation, and $f^{[l]}$ denotes the size of the maxpooling filter. Before applying the operation of a fully connected layer, the output from the previous layer is transformed into a one-dimensional vector. The output from a fully connected layer considering j^{th} node of the l^{th} is formulated as,

$$z_j^{[i]} = \sum_{l=1}^{n_{i-1}} w_{j,l}^i o_l^{i-1} + b_j^i \tag{4}$$

where o_l^{i-1} represents the input from a previous convolutional or maxpooling layer, and $w_{j,l}^i$ and b_j^i depict the weight and bias parameters. Secondly, we consider the existing DL networks, the VGG16 model [11] and the Wide-ResNet-28 model [12], due to their success in many other applications. The VGG16 model is a convolution neural network (CNN). This model consists of convolution layers of 3×3 filter with a stride equal to 1. The first two convolutional layers have 64 channels. The model uses the same padding and maxpooling layer of 2×2 filter of stride equal to 2. The next two convolutional layers are represented by 256 filters. Subsequently, three fully connected layers follow the convolution and maxpooling layers. The last layer represents the softmax layer. An objective function is used to optimize the VGG16 model training process. The objective function is formulated as,

$$J(w, b) = \frac{1}{m} \sum_{g=1}^{m} \Delta(o_g^{w,b}, o_g) \tag{5}$$

where Δ represents the cost function, m represents the training samples, and $o_g^{w,b}$ and o_g show the output from the final layer and the actual label of a training sample, respectively. We use a cross-entropy loss function to optimize the classification models. The cross-entropy loss function for binary classification is formulated as,

$$\Delta = -[o_g log(o_g^{w,b})] + [(1 - o_g) log(1 - o_g^{w,b})] \tag{6}$$

We also explore wide residual network (WideResNet) [12] which is an improved version of the residual network. It is found that increasing the depth of the residual network does not improve the performance. The WideResNet network is motivated by the fact that the widening of ResNet blocks provides a more consolidated approach of enhancing the performance of deep residual architectures compared to increasing their depth. This approach presents substantial performance improvement over [27]. The WideResNet model can produce more representative features from input samples compared to traditional CNN methods. The WideResNet model is based on the idea of skip connection. To illustrate this idea, we consider the stacking of convolution layers, which generally develop CNN, allows a deep model to learn from the features at the lower level in a hierarchical setup. Nevertheless, a given layer is considered to only connect with its two adjacent layers. The information flow from previous layers might be lost during the process of backpropagation to update the weights. Therefore, this process is not very effective. The skip connections sustain a low number of parameters and hold the feature information across all layers. The input of a given layer might be an aggregation of previous layers. This process consolidates the gradient flow efficiently since it exploits a superhighway and the skip connections in the gradient backpropagation algorithm. It is also investigated in the literature [28] that the shallow networks can entail substantially more parameters than deeper networks. Therefore, the researchers of the residual networks transformed them into thinner networks to train deeper networks. As the gradient parameters propagate through the deeper architecture, there is nothing to push them to flow through the residual block. Hence, it can avoid learning anything during training. It is feasible that there are either only limited blocks that encode informative representations or several blocks provide shallow information with insignificant contribution to the final output. Based on these observations, the WideResNet network is built on top of the work in [27] to address the aforementioned limitations. We explore the WideResNet network, namely WRN-28-2, due to its widespread

adoption and availability. This is the network with depth 28 and width 2, including batch normalization [29] and leaky ReLU nonlinearities [30]. As can be seen in Figure 4, the depth equal to 28 represents the total number of layers in the network, and the width equal to 2 represents the widening factor (multiplies the number of features in convolutional layers).

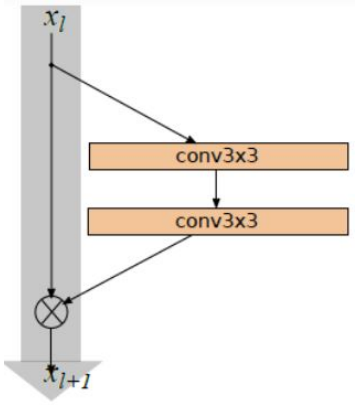

Figure 4. WideResNet has a depth and width. The depth represents the total number of layers in the network, and the width represents the widening factor.

4. Results

In this section, we provide the details of the dataset and its characteristics, experimental results we obtain on the dataset, and discussion about the analysis.

4.1. Dataset and Its Characteristics

We captured the cattle images using a handheld device consisting of an Intel RealSense D415 camera connected with a tablet and external batteries integrated in the handle of the stick holding the camera and the tablet (see Figure 5). We installed a customized software for capturing the cattle images. We placed the camera on the floor underneath the cow udder, taking the photo vertically upwards as shown in Figure 5. We collected the images in different housing systems, tied-stalls, loose housing systems, and the milking parlor. Before capturing an image, we specified the ID of the animal and number of hours since milking. We collected both RGB and depth images of 1400 dairy cattle. However, we dropped a large number of images due to poor quality in terms of severe blurring, partial visibility of the udder, and no visibility of udder due to cattle movement. We depict four sample images in Figure 6 representing udders of different cows. The first row shows the original RGB images, and the bottom row shows the corresponding depth images. We also depict four sample images in Figure 7 representing supernumerary teats. We show udders in the first two columns of the first row representing more than four teats. In the last two columns of the first row, we show udders representing four teats. The bottom row shows the corresponding depth images. We provide details of the dataset in Table 1 for both udder and supernumerary teats images. It is worth noting that the total number of images remains the same as shown in Table 1 when we integrate RGB channels with the depth channel. Therefore, we create four channel images representing both the depth and RGB information. For experimental analysis, we use 80% of the images for training and 20% for validation.

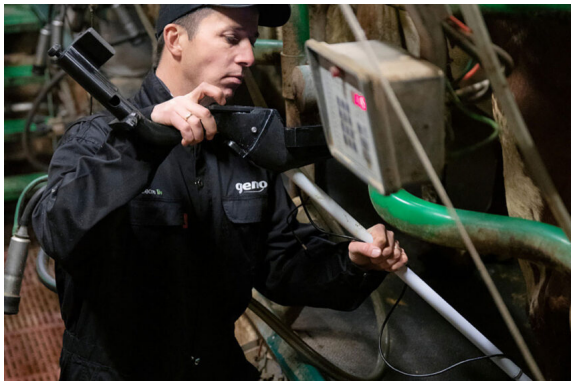

Figure 5. Data collection at cattle farm. Øyvind Nordbø (co-author) is collecting cattle images with a handheld device by placing it on the floor underneath the cow udder.

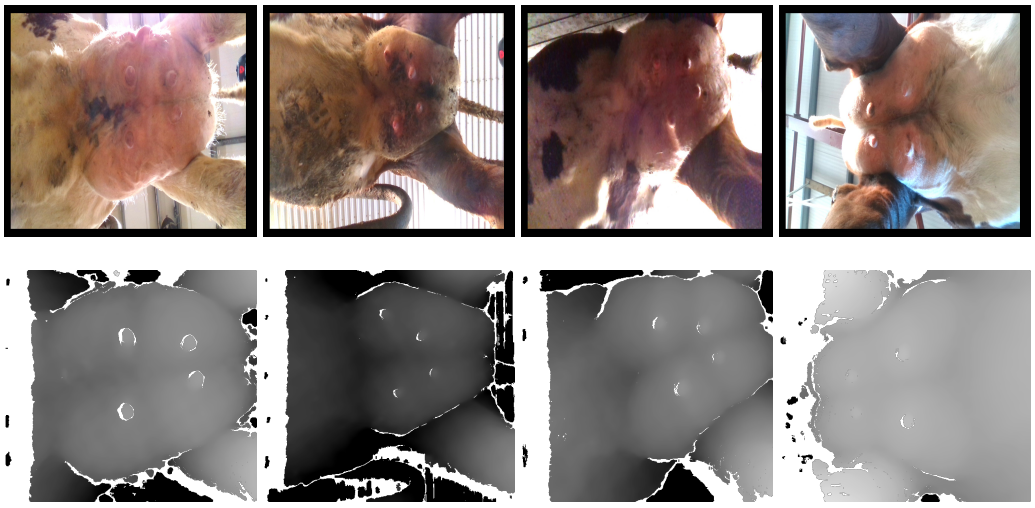

Figure 6. Udder images of cattles. The first row shows the RGB images. The bottom row shows the same images captured with the depth camera.

Figure 7. Supernumerary teat images. The first row shows the original images where the first two images belong to the first class (udder representing more than four teats), and the second two images belong to the second class (udder representing four teats). The bottom row shows the same images captured with the depth camera.

Table 1. Dataset details. We provide the details of our dataset for both udder images and supernumerary images. The number of images are represented in terms of RGB images. We consider 80% of the images for training and 20% for validation. The number of images remains the same when we fuse the depth channel with the RGB channels.

Device	Location	Data	Udder and Non-Udder Images	Supernumerary Images
Intel RealSense D415 camera	Tied-stalls Milking parlor	Training	232	380
		Validation	62	96

It is worth exploring the underlying characteristics of our dataset. For this purpose, we explore uniform manifold approximation and projection (UMAP) [31] which is a novel non-linear dimensionality reduction technique for data visualisation. We are considering UMAP instead of PCA (principal component analysis) because PCA is significantly impacted by the outliers available in the data. Moreover, PCA is a linear projection, and it cannot capture non-linear dependencies. The UMAP technique preserves the global structure of the data for better representation by locating important patterns in the high-dimensional space and transforming them in the lower-dimensional space. The UMAP technique considers the notion that the distances among data samples vary across the manifold. That is, the space itself is warping: stretching or shrinking. We present the algorithm for the UMAP technique in Figure 8. We explore the UMAP technique to visualize the underlying structures of both the udder\non-udder images and the supernumerary teats images. The UMAP projection of the udder dataset is presented in Figure 9. As can be seen, the udder\non-udder images are easily distinguishable. This is also evident from the physical layout of the images (in Figure 6). We depict the UMAP projection of the supernumerary teats images in Figure 10. The underlying patterns in the images from both classes of the supernumerary dataset are very similar; therefore, the separability is very challenging. These analyses show that the classification of the supernumerary dataset would be difficult in comparison to the classification of the udder dataset.

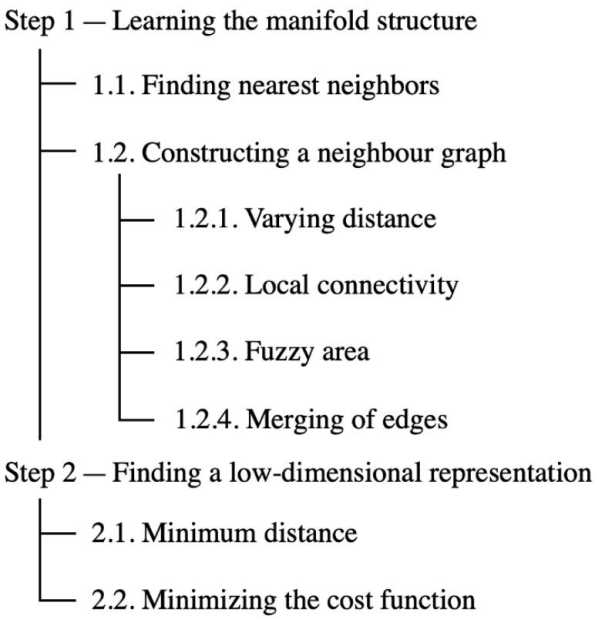

Figure 8. The UMAP algorithm [32]. The algorithm shows two major steps, namely learning the manifold structure and finding a low-dimensional representation.

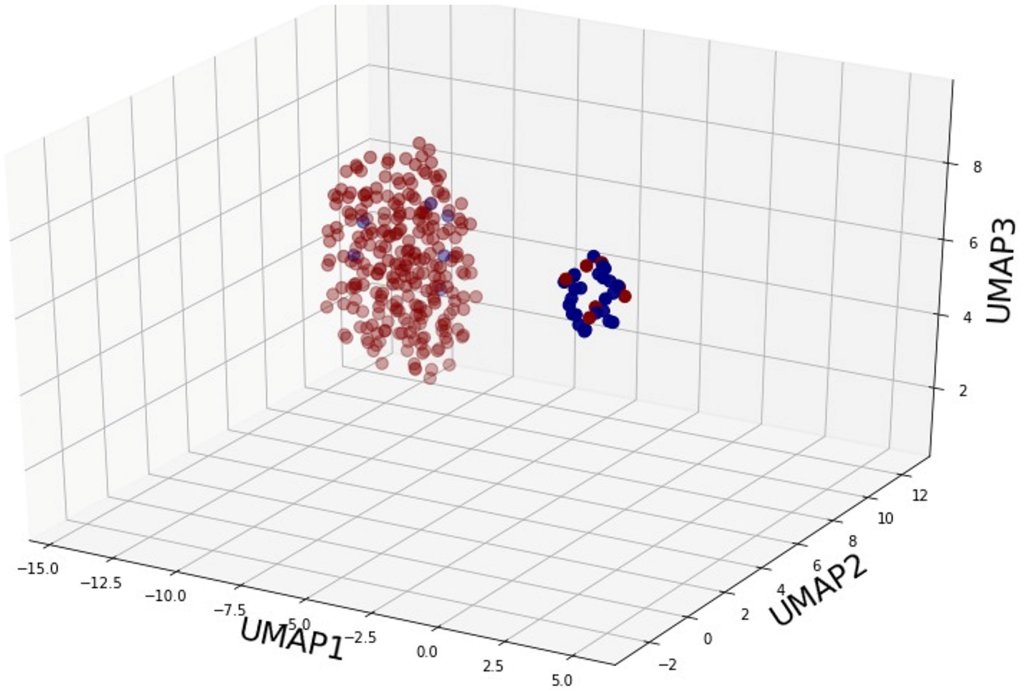

Figure 9. The Umap projection of the udder dataset. The three axes in the diagram represent the transformation of data from higher dimensions to lower UMAP dimensions. The underlying structures of udder\non-udder images show that they are easily separable.

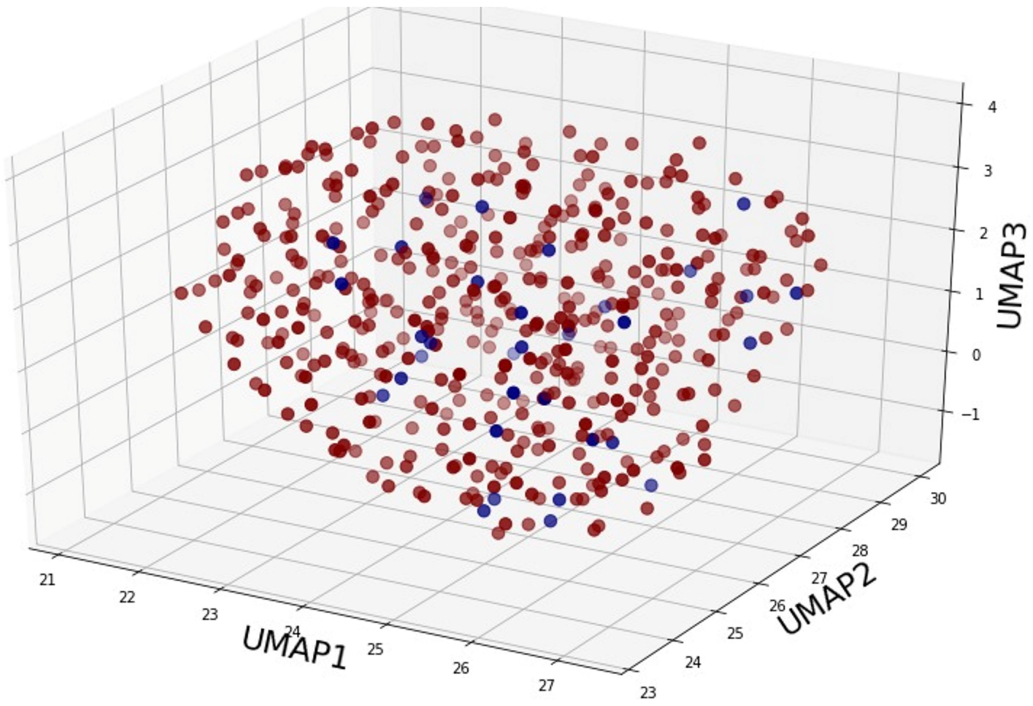

Figure 10. The Umap projection of the supernumerary teats dataset. The three axes in the diagram represent the transformation of data from higher dimensions to lower UMAP dimensions. The underlying structures of the supernumerary teats images show that images from both classes represent very similar patterns.

4.2. Experimental Results

We implemented our proposed method in a Python (version 3.8) programming environment using the Pytorch framework and the Numpy library. We performed coding and experiments on a personal computer consisting of an Intel(R) Core(TM) i7-10750H CPU @ 2.60 GHz 2.59 GHz and equipped with NVIDIA GeForce RTX 2080 Super with Max-Q design and with the configuration of Cuda toolkit 11 on a Windows 10 Pro operating system.

We do not have large data to train different DL networks. Therefore, in the first part of the experimental analysis, we consider the impact of different data augmentation techniques on the performances of our proposed network 1, our proposed custom network 2, the VGG16 model, and the WideResNet model for all of our classification tasks. We consider and explore different data augmentation techniques to produce modified versions of images in our dataset to increase its size. These techniques help to cope with data scarcity and insufficient data diversity. Bringing variations in the dataset improve the training process of the deep neural networks. We consider four different levels of augmentation in our experimental analysis, namely brightness and contrast, Gaussian blurring, horizontal–vertical flipping and rotation, and combining all these augmentations. We call the combination of these augmentations a full augmentation. The contrast isolates intensity levels between the dark and bright regions of an image. Brightness represents the specified magnitude of intensity level in an image. Both contrast and brightness are affected by each other. Therefore, we consider both in the first level. In the second level, we consider Gaussian blur which randomly blurs the image using a Gaussian distribution. In the third level, we consider horizontal–vertical flipping and rotation. In the fourth level, we use all the augmentation techniques namely brightness and contrast, Gaussian blurring, horizontal–vertical flipping, and rotation.

We present the results in terms of different performance metrics: accuracy, f-score, precision, and sensitivity as shown in the equations below. Accuracy is the fraction of predictions a model got right. Accuracy represents the ratio of the number of correct predictions to the total number of predictions. F-score represents harmonic mean for measuring a model's performance considering the analysis of binary classification. These metrics help determine the robustness of a machine learning model.

$$Accuracy = \frac{Numberofcorrectpredictions}{Totalnumberofpredictions} = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

$$F - score = \frac{2TP}{2TP + FP + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{10}$$

For udder\non-udder classification, we present the results in terms of accuracy for all the four networks in Figure 11. As can be seen, our proposed network 1, our proposed network 2, the VGG16 model, and the WideResNet model present higher accuracies in cases of horizontal–vertical flipping and rotation, and full augmentation techniques (brightness and contrast, Gaussian blurring, horizontal–vertical flipping, and rotation). The proposed network 1 and the proposed network 2 present 97% accuracy in the case of both horizontal–vertical flipping and rotation, and full augmentation. The VGG16 model presents 100% accuracy in the case of horizontal–vertical flipping and rotation and 96% accuracy in case of full augmentation. If we consider both horizontal–vertical flipping and rotation and full augmentation, our proposed method presents competitive results. We can also infer that the consolidated or intense augmentations present better results.

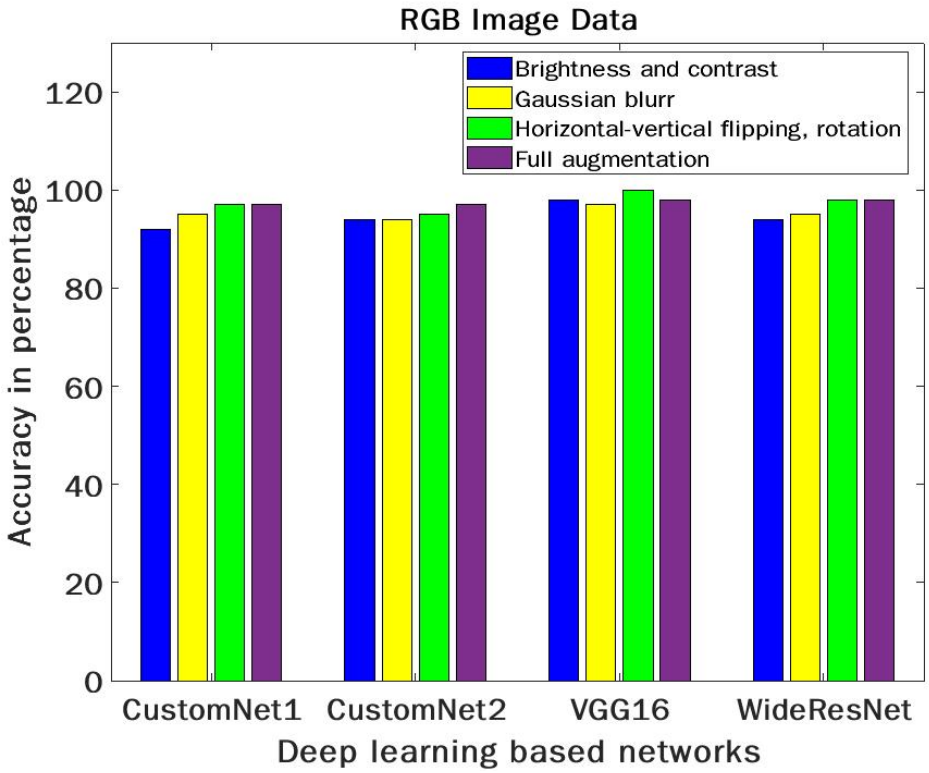

Figure 11. Udder classification using RGB image data. Considering different augmentation techniques, we present the results obtained from our proposed custom network 1, our proposed custom network 2, the VGG16 model [11], and the WideResNet model [12].

For udder\non-udder classification, we also consider both RGB and depth images. We present the results in terms of accuracy for all the four networks in Figure 12. In this case, we combine the depth channel with the three RGB channels to produce four channels of data. As can be seen, also in this case, our proposed network 1 and our proposed network 2 present competitive results in comparison with the VGG16 model and the WideResNet model. The accuracies are higher in cases of horizontal–vertical flipping and rotation and full augmentation techniques. Our proposed network 1, our proposed network 2, the VGG16 model, and the WideResNet present 94%, 97%, 97%, and 97% accuracies in the case of full augmentation. If we consider both horizontal–vertical flipping and rotation and full augmentation, our proposed method presents competitive accuracies. Moreover, combining both RGB and depth data, we can infer that the consolidated or intense augmentations present better results.

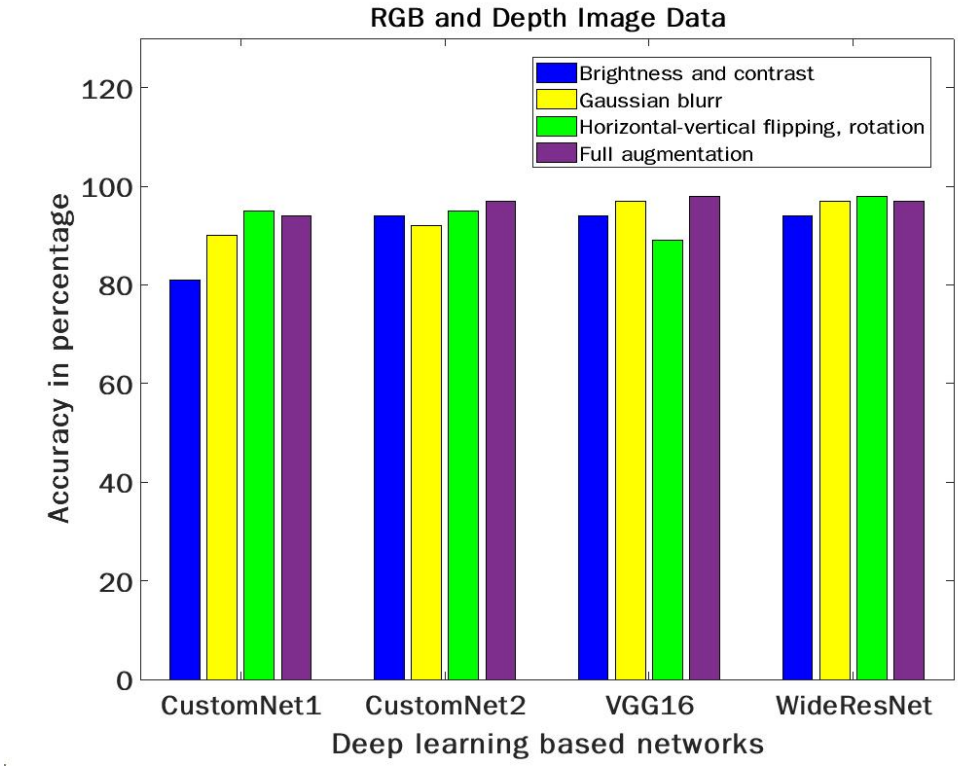

Figure 12. Udder classification using RGB and depth image data. Considering different augmentation techniques, we present the results obtained from our proposed network 1, our proposed network 2, the VGG16 model [11], and the WideResNet model [12].

For supernumerary teat classification, we present the results in term of accuracy for all four networks in Figure 13. As can be seen, our proposed method (custom network 1) presents higher accuracies when we use the augmentation techniques, i.e., horizontal–vertical flipping and rotation and full augmentation techniques (brightness and contrast, Gaussian blurring, horizontal–vertical flipping and rotation). The performance of our proposed method (custom network 1) is comparatively higher than other networks in the case of supernumerary teat classification since the data is limited. Therefore, the smaller network performs better. It is also worth noting that supernumerary teat classification is a more difficult task compared to udder\non-udder classification. Therefore, we have comparatively lower accuracies for all the networks. These results are aligned with analysis we performed with the UMAP technique, where we found that the separability of the images in the supernumerary dataset is more difficult. For supernumerary teat classification, we also consider both RGB and depth images. We present the results in term of accuracy for all four networks in Figure 14. As can be seen, also in this case, our proposed method (custom network 1 and custom network 2) presents higher accuracies in the case of full augmentation techniques. By taking into account all these results, we infer that more augmentations present better results for both RGB and depth data.

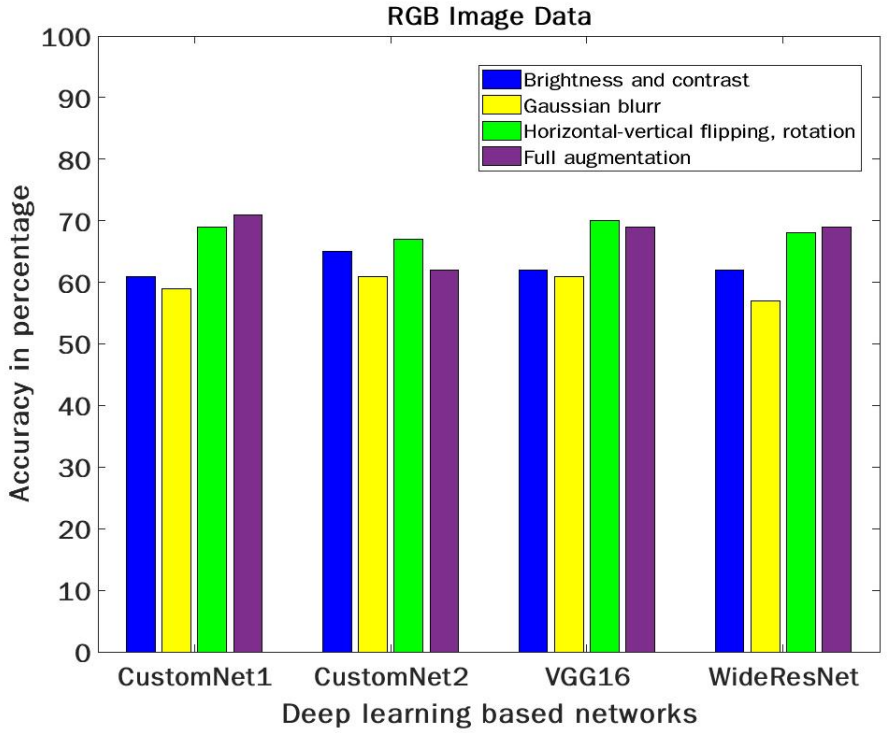

Figure 13. Supernumerary teat classification using RGB data. The results are presented in term of accuracy for our proposed custom network 1, our proposed custom network 2, the VGG16 model [11], and the WideResNet model [12].

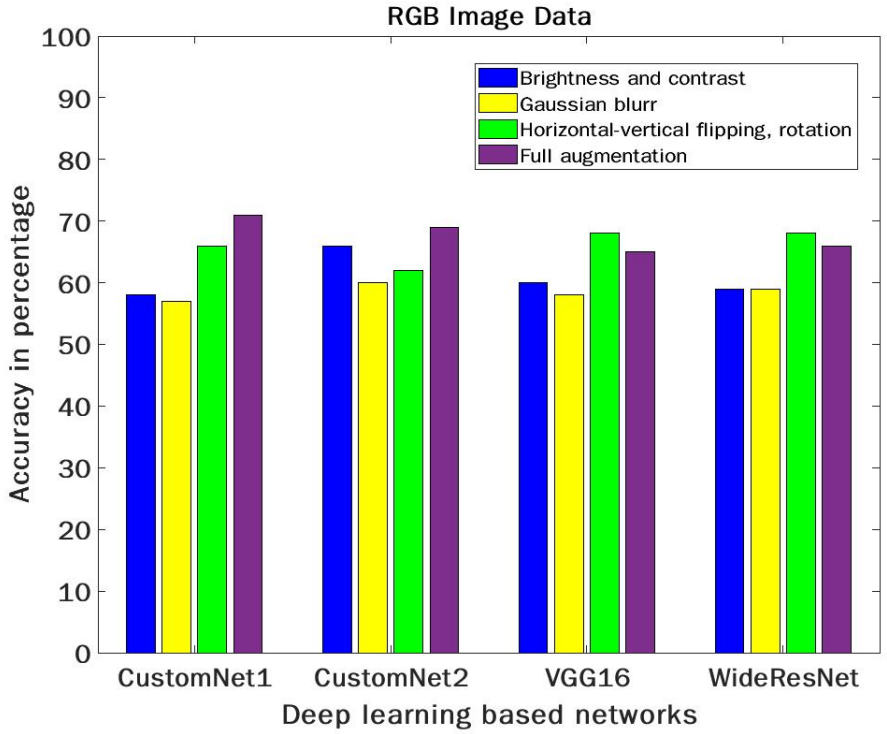

Figure 14. Supernumerary classification using RGB and depth image data. The results are presented in term of accuracy for our proposed network 1, our proposed network 2, the VGG16 model [11], and the WideResNet model [12].

In the second part of the experimental analysis, we consider the impact of RGB data and RGB data combined with the depth data on the performances of our proposed method (custom network 1 and custom network 2), the VGG16 model, and the WideResNet model for udder\non-udder image classification. We report the results in terms of F-score in Table 2. We also report the results in terms of precision and sensitivity in Table 3. As can be seen, our proposed method (custom network 1 and custom network 2) presents competitive performances in the case of RGB data. The performance in the case of combining the depth information with the RGB data is also competitive.

Table 2. F-scores for udder\non-udder classification. The results are presented for our proposed network 1, our proposed network 2, the VGG16 model [11], and the WideResNet model [12].

Models	RGB Images				RGB and Depth Images			
	Bright-Contrast	**Gauss. Blur**	**Hor-Ver Rot.**	**Full Aug.**	**Bright-Contrast**	**Gauss. Blur**	**Hor-Ver Rot.**	**Full Aug.**
Proposed CNet1	0.92	0.95	0.97	0.97	0.80	0.90	0.95	0.94
Proposed CNet2	0.94	0.93	0.95	0.97	0.93	0.92	0.95	0.97
VGG16 [11]	0.98	0.97	1.00	0.96	0.93	0.97	0.89	0.97
WideResNet [12]	0.94	0.95	0.95	0.96	0.93	0.97	0.95	0.97

Table 3. Precision and sensitivity for udder\non-udder classification. The results are presented for our proposed network 1, our proposed network 2, the VGG16 model [11], and the WideResNet model [12].

Models	RGB Images				RGB and Depth Images			
	Bright-Contrast	**Gauss. Blur**	**Hor-Ver Rot.**	**Full Aug.**	**Bright-Contrast**	**Gauss. Blur**	**Hor-Ver Rot.**	**Full Aug.**
Proposed CNet1	0.92 0.91	0.95 0.94	0.96 0.97	0.96 0.97	0.80 0.80	0.90 0.90	0.95 0.95	0.94 0.93
Proposed CNet2	0.93 0.93	0.93 0.93	0.95 0.95	0.96 0.96	0.93 0.94	0.93 0.91	0.95 0.94	0.97 0.96
VGG16 [11]	0.98 0.98	0.97 0.96	1.00 1.00	0.96 0.96	0.93 0.93	0.97 0.96	0.91 0.87	0.96 0.96
WideResNet [12]	0.93 0.93	0.95 0.95	0.95 0.95	0.96 0.96	0.93 0.94	0.96 0.97	0.96 0.94	0.96 0.96

To better understand the usage of both the RGB data and depth data combined with the RGB data, we also present the results in terms of F-score, precision and sensitivity for supernumerary teat classification. We present the F-scores for the same DL-based networks in Table 4. We present the precision and sensitivity in Table 5. We analyze the results by focusing on our proposed method that represents the smaller networks, namely network 1 and network 2. We consider smaller networks since we have seen that they perform better due to limited availability of data for supernumerary teat classification. In this case, we also analyze the results for only horizontal–vertical flipping and rotation and for full augmentation. Higher F-scores are highlighted in bold. Our proposed method (customer network 1) presents the highest F-score equal to 70% in the case of full augmentation. As can be seen in Table 5, the combination of RGB data with depth information also presents better performances mostly in the case of supernumerary teat classification which is a difficult task compared to udder\non-udder classification. In fact, the consolidation of input information by combining depth data with the RGB data is an important aspect, especially when we have a challenging problem such as supernumerary teat classification. Therefore, we present a consolidation framework characterized by hybrid input from both

the RGB and the depth camera. Depth data improves the input by adding significant variations to data to improve the training process.

Table 4. F-scores for supernumerary teat classification. The results are presented for our proposed network 1, our proposed network 2, the VGG16 model [11], and the WideResNet model [12] considering both RGB data and depth data.

Models	RGB Images				RGB and Depth Images			
	Bright-Contrast	Gauss. Blur	Hor-Ver Rot.	Full Aug.	Bright-Contrast	Gauss. Blur	Hor-Ver Rot.	Full Aug.
Proposed CNet1	0.65	0.57	0.68	**0.70**	0.57	0.55	0.64	**0.71**
Proposed CNet2	0.65	0.61	0.66	0.61	0.66	0.58	0.61	**0.68**
VGG16 [11]	0.61	0.61	0.69	0.69	0.59	0.56	0.67	0.64
WideResNet [12]	0.62	0.54	0.67	0.67	0.57	0.58	0.68	0.65

Table 5. Precision and sensitivity for supernumerary teat classification. The results are presented for our proposed network 1, our proposed network 2, the VGG16 model [11], and the WideResNet model [12].

Models	RGB Images				RGB and Depth Images			
	Bright-Contrast	Gauss. Blur	Hor-Ver Rot.	Full Aug.	Bright-Contrast	Gauss. Blur	Hor-Ver Rot.	Full Aug.
Proposed CNet1	0.79 0.55	0.60 0.54	0.68 0.67	0.70 0.69	0.62 0.52	0.55 0.54	0.65 0.63	0.72 0.68
Proposed CNet2	0.72 0.60	0.63 0.57	0.67 0.64	0.63 0.58	0.71 0.61	0.59 0.57	0.61 0.60	0.68 0.67
VGG16 [11]	0.61 0.60	0.61 0.61	0.72 0.67	0.69 0.68	0.59 0.59	0.57 0.55	0.67 0.66	0.63 0.63
WideResNet [12]	0.62 0.62	0.55 0.53	0.67 0.67	0.68 0.67	0.58 0.55	0.58 0.58	0.68 0.68	0.67 0.62

For udder\non-udder and supernumerary teat classification, we present additional experimentation results in comparison with the state-of-the-art methods, namely: DeepHolistic [33], DeepPropagate [34], MatchAnchoring [35], DeepLabel [36], and TempLearn [37]. Method [33] leverages unlabeled data to mitigate the reliance on the samples. They combine elements of the famous models to introduce a new technique. They guess the low-entropy labels for data-augmented unlabeled samples. Method [34] investigates the manifold assumption that similar samples should obtain the same prediction. They used a transductive label propagation technique to produce predictions on the data and use these predictions to produce pseudo-labels. Method [35] considers distribution alignment and augmentation anchoring. The distribution alignment supports the marginal distribution of predictions to be near to the marginal distribution of ground-truth samples. The methods provide informative versions of an input into the model and support each output to be close to the prediction for a weak version of the same input. The method [36] learns data characteristics from a limited number of samples by exploring the underlying properties. Method [37] exploits the temporal dynamics and inherent multimodal attributes in the data samples. They use the temporal gradient as an additional modality for more attentive feature extraction. They extract the fine-grained representations from the temporal gradient and impose consistency across different modalities. The results in comparison with these methods are presented in Table 6 considering the performance metrics: F-score, precision, and sensitivity. In Table 6, the third and the fifth columns show the F-scores for udder\non-udder and supernumerary teat classification. The fourth and the last columns show precision and sensitivity results. Our proposed method outperformed all other methods for both udder\non-udder and supernumerary teat classification. For udder\non-udder classification, higher results are obtained. However, the results are lower for the supernumerary teat

classification since it is a difficult task. It is also worth noting that the results of the methods are not stable since the labeling lacks the informative and unique samples of the data.

Table 6. For udder\non-udder and supernumerary teat classification, we present additional results in comparison with the VGG16 model [11], the WideResNet model [12], and the state-of-the-art methods ,namely: DeepHolistic [33],DeepPropagate [34], MatchAnchoring [35], DeepLabel [36], and TempLearn [37]. The third and the fifth columns show the F-scores for udder\non-udder and supernumerary teat classification. The fourth and the last columns show precision and sensitivity results.

Year	Models	Udder\ Non-Udder F-Scores	Udder\ Non-Udder Prec-Sensi	Supernum. Teat F-Scores	Supernum. Teat Prec-Sensi
Our method	Proposed CNet1	0.97	0.96–0.97	0.71	0.71–0.69
Our method	Proposed CNet2	0.97	0.96–0.97	0.61	0.63–0.58
2014	VGG16 [11]	0.96	0.96–0.96	0.69	0.71–0.65
2018	WideResNet [12]	0.96	0.95–0.97	0.67	0.68–0.66
2019	DeepHolistic [33]	0.85	0.84–0.86	0.56	0.52–0.60
2019	DeepPropagate [34]	0.93	0.95–91	0.59	0.60–0.58
2019	MatchAnchoring [35]	0.91	0.92–91	0.63	0.63–0.63
2021	DeepLabel [36]	0.93	0.95–0.91	0.59	0.62–0.57
2022	TempLearn [37]	0.87	0.86–0.89	0.55	0.54–0.55

5. Discussion

We have seen that the performance of the DL models improves with data augmentation techniques. In fact, the data augmentation techniques integrate data with more variations which help the models to better learn during the training stage. We have also seen that the integration of depth information with the RGB information did not improve the results significantly. The reasons are many. Firstly, the results for udder\non-udder classification are significantly high without considering the depth information. Therefore, there is not enough space for the depth information to further improve the results. Secondly, the depth information are more susceptible to distance from the camera. In the current settings, the distance information is not very relevant for udder\non-udder image classification and supernumerary teats classification. Furthermore, the supernumerary teats are small and mostly on the back side of the udder. Therefore, when we were capturing the depth images, the visibility of the supernumerary teats were poor due to a curved udder where the reflecting light from the udder surface is not homogeneous. The supernumerary teats were also dirty due to mud in some cases where the depth images cannot contribute to performance. The depth information consolidate the DL models since overfitting is still a challenge. Considering the limitations, our proposed method does not detect and classify other cattle traits. Our method should be tuned to generalize well for other cattle traits.

6. Conclusions

In this paper, we presented two important classification tasks, namely udder\non-udder classification and supernumerary teat classification. For this purpose, we proposed two DL-based networks: custom network 1 and custom network 2. We also explore the VGG16 model and the WideResNet model. We evaluated the performance of these networks using the dataset we collected. We analyzed the variations in the performances of the deep networks by considering different data augmentation techniques. We discovered that a higher level of augmentations generally improves the results of the networks. We also analyzed the variations in their performances by taking into account both the RGB data and depth data. We explored that the addition of the depth information could improve

the results for a challenging classification task, especially when limited data is available to train the models. We also performed a comparison with the reference methods.

In our future work, we would extend our work to classify other traits of cows. Most of the other traits are measured on a scale from 1–9. Therefore, we would consider them as regression problems. We would also explore the impact of augmentation techniques and depth information for other traits. For other traits, for example, teat length and udder depth, the depth images may contribute significantly to the performance of DL-based models.

Author Contributions: H.A. conceptualization, methodology and formal analysis; M.U. discussion, conceptualization, and editing; Ø.N., F.A.C. and A.G.L. discussion, editing, formulation. All authors have read and agreed to the published version of the manuscript.

Funding: We would like to thank the Research Council of Norway for funding this study, within the BIONÆR program, project number 282252 and the Industrial PhD program, project number 310239. We would also like to thank the Norwegian University of Science and Technology for supporting the APC through the open-access journal publication fund.

Data Availability Statement: The data presented in this study are available on request from the co-author Mohib Ullah (mohib.ullah@ntnu.no). The data are not publicly available due to its proprietary nature and ethical concerns.

Acknowledgments: We would like to thank Hans Snerting and Jan Atle Bakkenget {hans.snerting@tine.no, jan.atle.bakkenget@tine.no} (both advisors in Tine SA), for the image acquisition.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wethal, K.B. Genetic Analyses of New Milkability, Temperament, and Udder Health Traits for Norwegian Red Cows Based on Data from Automatic Milking Systems. Ph.D. Thesis, Norwegian University of Life Sciences, Ås, Norway, 2020.
2. Avlsmålet for NRF. Available online: https://www.geno.no/fagstoff-og-hjelpemidler/avlsprogram-for-norsk-rodt-fe/avlsmalet-for-nrf/ (accessed on 25 January 2022).
3. Gocheva-Ilieva, S.; Yordanova, A.; Kulina, H. Predicting the 305-Day Milk Yield of Holstein-Friesian Cows Depending on the Conformation Traits and Farm Using Simplified Selective Ensembles. *Mathematics* **2022**, *10*, 1254. [CrossRef]
4. Wang, L.; Li, M.; Pei, X.; Zhang, J. Optimal Breeding Strategy for Livestock with a Dynamic Price. *Mathematics* **2022**, *10*, 1732. [CrossRef]
5. Wen, H.; Luo, H.; Yang, M.; Augustino, S.; Wang, D.; Mi, S.; Guo, Y.; Zhang, Y.; Xiao, W.; Wang, Y.; et al. Genetic parameters and weighted single-step genome-wide association study for supernumerary teats in Holstein cattle. *J. Dairy Sci.* **2021**, *104*, 11867–11877. [CrossRef] [PubMed]
6. Section 5—ICAR Guidelines for Conformation Recording of Dairy Cattle, Beef Cattle, Dual Purpose Cattle and Dairy Goats. Available online: https://www.icar.org/Guidelines/05-Conformation-Recording.pdf (accessed on 26 January 2022).
7. Poppe, M.; Mulder, H.; Ducro, B.; De Jong, G. Genetic analysis of udder conformation traits derived from automatic milking system recording in dairy cows. *J. Dairy Sci.* **2019**, *102*, 1386–1396. [CrossRef] [PubMed]
8. Chen, C.; Zhu, W.; Norton, T. Behaviour recognition of pigs and cattle: Journey from computer vision to deep learning. *Comput. Electron. Agric.* **2021**, *187*, 106255. [CrossRef]
9. Cui, K.; Hu, C.; Wang, R.; Sui, Y.; Mao, H.; Li, H. Deep-learning-based extraction of the animal migration patterns from weather radar images. *Sci. China Inf. Sci.* **2020**, *63*, 1–10. [CrossRef]
10. Chalmers, C.; Fergus, P.; Wich, S.; Longmore, S. Modelling Animal Biodiversity Using Acoustic Monitoring and Deep Learning. *arXiv* **2021**, arXiv:2103.07276.
11. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
12. Oliver, A.; Odena, A.; Raffel, C.; Cubuk, E.D.; Goodfellow, I.J. Realistic evaluation of deep semi-supervised learning algorithms. *arXiv* **2018**, arXiv:1804.09170.
13. Kappes, R.; Knob, D.A.; Thaler, A.; Alessio, D.R.M.; Rodrigues, W.B.; Scholz, A.M.; Bonotto, R. Cow's functional traits and physiological status and their relation with milk yield and milk quality in a compost bedded pack barn system. *Rev. Bras. Zootec.* **2020**, *49*. [CrossRef]
14. de Oliveira, M.H.V.; de Vasconcelos Silva, J.A.I.; da Silva Faria, R.A.; de Paiva, J.T.; Malheiros, J.M.; dos Santos Correia, L.E.C.; Albuquerque, L.G.; de Genova Gaya, L. Genetic evaluation of weaning weight and udder score in Nellore cattle. *Livest. Sci.* **2021**, *244*, 104400. [CrossRef]
15. Carvalho, N.; Daltro, D.; Machado, J.; Camargo, E.; Panetto, J.d.C.; Cobuci, J. Genetic parameters and genetic trends of conformation and management traits in Dairy Gir cattle. *Arq. Bras. Med. VeterináRia Zootec.* **2021**, *73*, 938–948. [CrossRef]

16. Shorten, P. Computer vision and weigh scale-based prediction of milk yield and udder traits for individual cows. *Comput. Electron. Agric.* **2021**, *188*, 106364. [CrossRef]
17. Jiang, J.; Cole, J.B.; Freebern, E.; Da, Y.; VanRaden, P.M.; Ma, L. Functional annotation and Bayesian fine-mapping reveals candidate genes for important agronomic traits in Holstein bulls. *Commun. Biol.* **2019**, *2*, 1–12. [CrossRef]
18. Colinet, F.; Vandenplas, J.; Vanderick, S.; Hammami, H.; Mota, R.; Gillon, A.; Hubin, X.; Bertozzi, C.; Gengler, N. Bayesian single-step genomic evaluations combining local and foreign information in Walloon Holsteins. *Animal* **2018**, *12*, 898–905. [CrossRef]
19. Stefani, G.; El Faro, L.; Júnior, M.L.S.; Tonhati, H. Association of longevity with type traits, milk yield and udder health in Holstein cows. *Livest. Sci.* **2018**, *218*, 1–7. [CrossRef]
20. Manzo, M.; Pellino, S. Voting in transfer learning system for ground-based cloud classification. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 542–553. [CrossRef]
21. Shanthamallu, U.S.; Spanias, A. Machine and Deep Learning Applications. In *Machine and Deep Learning Algorithms and Applications*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 59–72.
22. Porter, I.; Wieland, M.; Basran, P. Feasibility of the use of deep learning classification of teat-end condition in Holstein cattle. *J. Dairy Sci.* **2021**, *104*, 4529–4536. [CrossRef]
23. Ebrahimi, M.; Mohammadi-Dehcheshmeh, M.; Ebrahimie, E.; Petrovski, K.R. Comprehensive analysis of machine learning models for prediction of sub-clinical mastitis: Deep Learning and Gradient-Boosted Trees outperform other models. *Comput. Biol. Med.* **2019**, *114*, 103456. [CrossRef]
24. Xudong, Z.; Xi, K.; Ningning, F.; Gang, L. Automatic recognition of dairy cow mastitis from thermal images by a deep learning detector. *Comput. Electron. Agric.* **2020**, *178*, 105754. [CrossRef]
25. Fadul-Pacheco, L.; Delgado, H.; Cabrera, V.E. Exploring machine learning algorithms for early prediction of clinical mastitis. *Int. Dairy J.* **2021**, *119*, 105051. [CrossRef]
26. Nye, J.; Zingaretti, L.M.; Pérez-Enciso, M. Estimating conformational traits in dairy cattle with deepaps: A two-step deep learning automated phenotyping and segmentation approach. *Front. Genet.* **2020**, *11*, 513. [CrossRef] [PubMed]
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. *Eur. Conf. Comput. Vis* **2016**, *9908*, 630–645.
28. Larochelle, H.; Erhan, D.; Courville, A.; Bergstra, J.; Bengio, Y. An empirical evaluation of deep architectures on problems with many factors of variation. In Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007; pp. 473–480.
29. Ioffe, S.; Normalization, C.S.B. Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.
30. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the ICML Workshop on Deep Learning for Audio, Speech and Language Processing, Atlanta, GA, USA, 16–21 June 2013; Volume 30, p. 3.
31. McInnes, L.; Healy, J.; Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, arXiv:1802.03426.
32. UMAP Algorithm. Available online: https://towardsdatascience.com/umap-dimensionality-reduction-an-incredibly-robust-machine-learning-algorithmb5acb01de568 (accessed on 12 February 2022)
33. Berthelot, D.; Carlini, N.; Goodfellow, I.; Papernot, N.; Oliver, A.; Raffel, C.A. Mixmatch: A holistic approach to semi-supervised learning. *arXiv* **2019**, arXiv:1905.02249.
34. Iscen, A.; Tolias, G.; Avrithis, Y.; Chum, O. Label propagation for deep semi-supervised learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 5070–5079.
35. Berthelot, D.; Carlini, N.; Cubuk, E.D.; Kurakin, A.; Sohn, K.; Zhang, H.; Raffel, C. Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. *arXiv* **2019**, arXiv:1911.09785.
36. Khaleghian, S.; Ullah, H.; Kræmer, T.; Eltoft, T.; Marinoni, A. Deep Semisupervised Teacher–Student Model Based on Label Propagation for Sea Ice Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 10761–10772. [CrossRef]
37. Xiao, J.; Jing, L.; Zhang, L.; He, J.; She, Q.; Zhou, Z.; Yuille, A.; Li, Y. Learning from temporal gradient for semi-supervised action recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 3252–3262.