CISTER
Research Centre in
Real-Time & Embedded
Computing Systems

Journal Paper

Automated sheep facial expression classification using deep transfer learning

Alam Noor
Yaqin Zhao
Anis Koubâa*
Longwen Wu
Rahim Khan
Fakheraldin Y.O. Abdalla

*CISTER Research Centre

CISTER-TR-201014

2020/08

Automated sheep facial expression classification using deep transfer learning

Alam Noor, Yaqin Zhao, Anis Koubâa*, Longwen Wu, Rahim Khan, Fakheraldin Y.O. Abdalla

*CISTER Research Centre

Polytechnic Institute of Porto (ISEP P.Porto)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail: aska@isep.ipp.pt

https://www.cister-labs.pt

Abstract

Digital image recognition has been used in the different aspects of life, mostly in object classification and detections. Monitoring of animal life with image recognition in natural habitats is essential for animal health and production. Currently, Sheep Pain Facial Expression Scale (SPFES) has become the focus of monitoring sheep from facial expression. In contrast, pain level estimation from facial expression is an efficient and reliable mark of animal life. However, the manual assessment is lack of accuracy, time-consuming, and monotonous. Hence, the recent advancement of deep learning in computer vision helps to classify facial expression as fast and accurate. In this paper, we proposed a sheep face dataset and framework that uses transfer learning with fine-tuning for automating the classification of normal (no pain) and abnormal (pain) sheep face images. Current state-of-the-art convolutional neural networks (CNN) based architectures are used to train the sheep face dataset. The data augmentation, L2 regularization, and fine-tuning has been used to prepare the models. The experimental results related to the sheep facial expression dataset achieved 100% training, 99.69% validation, and 100% testing accuracy using the VGG16 model. While employing other pre-trained models, we gained 93.10% to 98.4% accuracy. Thus, it shows that our proposed model is optimal for high-precision classification of normal and abnormal sheep faces and can check on a comprehensive dataset. It can also be used to assist other animal life with high accuracy, save time and expenses.

Computers and Electronics in Agriculture 175 (2020) 105528

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Automated sheep facial expression classification using deep transfer learning

Alam Noor (MS) (AI Researcher)[a,b,*], Yaqin Zhao (PhD)[a,*], Anis Koubaa (PhD)[b,c], Longwen Wu (PhD)[a], Rahim Khan (PhD)[a], Fakheraldin Y.O. Abdalla (PhD)[a]

[a] Department of Information and Communication Engineering, Harbin Institute of Technology, Harbin 150001, China
[b] Robotics and Internet-of-Things Prince Sultan University, Saudi Arabia
[c] CISTER, INESC-TEC, ISEP, Polytechnic Institute of Porto, Portugal

ARTICLE INFO

Keywords:
CNN architectures
Fine-tuning
Sheep face dataset
Sheep face classification
Transfer learning

ABSTRACT

Digital image recognition has been used in the different aspects of life, mostly in object classification and detections. Monitoring of animal life with image recognition in natural habitats is essential for animal health and production. Currently, Sheep Pain Facial Expression Scale (SPFES) has become the focus of monitoring sheep from facial expression. In contrast, pain level estimation from facial expression is an efficient and reliable mark of animal life. However, the manual assessment is lack of accuracy, time-consuming, and monotonous. Hence, the recent advancement of deep learning in computer vision helps to classify facial expression as fast and accurate. In this paper, we proposed a sheep face dataset and framework that uses transfer learning with fine-tuning for automating the classification of normal (no pain) and abnormal (pain) sheep face images. Current state-of-the-art convolutional neural networks (CNN) based architectures are used to train the sheep face dataset. The data augmentation, L2 regularization, and fine-tuning has been used to prepare the models. The experimental results related to the sheep facial expression dataset achieved 100% training, 99.69% validation, and 100% testing accuracy using the VGG16 model. While employing other pre-trained models, we gained 93.10% to 98.4% accuracy. Thus, it shows that our proposed model is optimal for high-precision classification of normal and abnormal sheep faces and can check on a comprehensive dataset. It can also be used to assist other animal life with high accuracy, save time and expenses.

1. Introduction

Digital image processing mostly engages many techniques such as face and pose recognition, identification, and classification of humans and animals by computer. The inspection and assessment tool of farm animals through digital image recognition is useful to establish the welfare standard of animal products, protection, and monitoring their lives. The animal health and intervention strategies by manual observations occur low accuracy estimation (Main et al., 2003; Napolitano et al., 2009; Guesgen et al., 2014; Stubsjøen and Valle, 2011). Different manual techniques have been developed for animal species, yet few of them have been used to assess sheep life (Napolitano et al., 2009; Guesgen et al., 2014; Stubsjøen and Valle, 2011; Leach et al., 2012; Yang et al., 2015; McLennan et al., 2016). While observing pain level assessment of sheep living in their natural environment is a critical task (McLennan, 2018). Pain in sheep is caused by diseases such as footrot (Dolan et al., 2003) and mastitis (Dolan et al., 2000) commonly found in sheep. To evaluate and identify pain is essential to know the real causes that hinder the sheep growth (Flecknell, 2008). Well-organized pain assessment has the advantage of diagnosing the illness in a short time. Recently, various modern automated techniques have been used for facial expression of an animal pain level, such as post-vasectomy pain in mice (Leach et al., 2012). Similarly, the hands-on approach Localize Sparsely Distributed Facial Landmarks (LSDFL) is being used for sheep pose, but it can be affected by the large dataset. In LSDFL, the facial expression identification introduced for the pain of mastitis and footrot (Yang et al., 2015). In last, Sheep Pain Facial Expression Scale (SPFES) has been developed, which is a standard methodology to find pain in sheep as the manual assessment (McLennan et al., 2016). But the manual check is less successful because the sheep are not examined in natural surroundings. Also, maintaining large animal farms is time-consuming and costly and can lead to animal life in unhealthy states.

* Corresponding authors at: Department of Information and Communication Engineering, Harbin Institute of Technology, Harbin 150001, China.
E-mail addresses: pinkheart_gold@yahoo.com (A. Noor), yaqinzhao@hit.edu.cn (Y. Zhao), akoubaa@psu.edu.sa (A. Koubaa), wulongwen@hit.edu.cn (L. Wu), rahimkhan9001@yahoo.com (R. Khan), fakheraldin.abdalla@hit.edu.cn (F.Y.O. Abdalla).

https://doi.org/10.1016/j.compag.2020.105528
Received 15 April 2020; Received in revised form 21 May 2020; Accepted 23 May 2020
Available online 06 June 2020
0168-1699/

Similarly, other animals can be affected by diseased sheep if they don't investigate in time.

To address these problems, this research used the deep transfer of convolutional neural network state-of-the-art architectures, which is the most excellent tool of computer vision for small databases and low computational powers. The contribution of our research is that we developed a sheep face dataset using state-of-the-art power of Deep Learning (CNN architectures), which has been used for image classification with millions of images from different categories. In this paper, the transfer learning approach is used to classify normal (without pain), and abnormal (with pain) sheep faces as a facial expression in sheep. The proposed method improves accuracy, efficiency, and consistency in sheep face classification. The given approach can be used for sheep as a binary classifier. This technique is widely used in human facial rating for emotion, but here we develop it for sheep facial classification.

The remaining sections of the paper organized as follows. In Section 2, we explain related work. In this section, we briefly discuss the advantage of pre-trained CNN models and their applications in the area of image classification. We also present the related work for the Sheep facial expression and Sheep face dataset collection with pass through pre-processing methods. In Section 3, the proposed method has been evaluated with different parameters of deep transfer learning and architectures. We describe the experimental setup and results with the discussion in Section 4. Finally, we conclude and present future work in Section 5.

2. Related work

2.1. Sheep face dataset

We have developed a sheep face dataset of corpus sheep faces collected from different sites like ImageNet, NADIS, Pixabay, Flickr, and Gettyimages with high resolutions in compliance with the rule of SPFES standard. It consists of 1650 images for training, 350 for validation followed by testing. Among this sheep corpus dataset, 1400 plus were healthy, and 900 plus were abnormal images. The normal sheep data also consisted of Lamb's faces but not in anomalous sheep data. While testing, the dataset includes 126 abnormal and 224 normal images classified by humans as the normal and abnormal database. The data set is available at the Mendeley.[1] In the given dataset, the sheep's facial expression belongs to SPFES [11]. SPFES have defined as the abnormality of the face of the sheep face with essential features related to ears, nose, and eyes. For Ears, the SPFES introduced pain level with ear rotation, both frontal faces, and profile. Ear with pinna visible, less visible and not visible has 0, less, and high pain, respectively. While the nose pain level defined by nostril shape as shallow "U" shaped nose has no pain and shallow or extended "V" shape nose has low and high pain each. Similarly, eye pain levels defined in three terms. The fully opened eye represents that it has no pain, and partly closed eye has pain (Lu et al., 2017). We have divided the sheep face dataset into two parts, one for the normal face which has no such features and another abnormal look with all such pain levels. Few sample images of our primary dataset have been shown in Fig. 1.

2.2. Pre-trained models

In this research, different pre-trained architectures have been used as a transfer learning such as AlexNet (Krizhevsky et al., 2012), VGG16 (Simonyan, 2014), GoogleNet (Szegedy et al., 2015), DenceNet201 (Huang et al., 2017), Inceptionv3 (Szegedy et al., 2016), ResNet50 (He et al., 2016), and DarkNet (Redmon, 2013) to classify the sheep faces. Previously, these models have been trained on more than one million images with 1000 different categories and achieved state-of-the-art performance in ImageNet, CIFAR-10, and CIFAR-100 competitions. Pre-trained models have a vibrant feature, so we have not used a scratch model. These architectures are state-of-the-art CNN in computer vision and mostly used for deep learning objectives and applications. All of these architectures used classification problems except DarkNet that used as a backbone for different versions of YOLO for detection. AlexNet, VGG16, and GoogleNet are classic networks. GoogleNet has 60K parameters with a clear structure and sigmoid/tanh used after the convolution layer causing non-linearity after pooling. AlexNet made the computer vision community to apply deep learning, and is more similar to LeNet but has much more significant with 60 million parameters. The three essential tools, ReLU, multiple GPUs, and local response normalization, must be considered to make AlexNet state-of-the-art. The keys of VGG16 are used as a 3×3 filter with stride 1 and the same padding convolution layer and 2×2 window size with stride 2 of max pooling. It consists of approximately 138M parameters and from lower to higher layer's height as well as with a decreased width and increased channel size.

To use a robust classic neural network (AlexNet, VGG16, and GoogleNet) with a large number of layers for training, causes vanishing and explodes gradient problems for massive datasets with millions of images. In that case, residual learning by using skip connection became possible to train a network with more than 100 layers. Skip connection allows to feed activation from the previous layer to another layer even much more rooted in the neural network. It converges faster than other pre-trained models. ResNet50 has 50 layer's network and classify many animals and other objects.

Residual Neural Network is computationally expensive. So, the inception model (Inceptionv3) became useful for saving computation, keeping non-linearity, doing the non-trivial operation and keeping the height and width the same, while making some channel smaller by using the network in the network (1×1 filter) (Lin et al., 2014). Instead of applying directly 2×2 or 3×3 window size from one layer to another, one can use a bottleneck (1×1) layer to shrink down the volume and does not hurt performance but save a massive computation. Inception complicates the convolution neural network but works well. However, it has relationships problem between input and output.

Densely connected convolutional networks (DenseNet201) made relationships between input and output shorter with $L(L + 1)/2$. DenseNet has the effective performance for feature propagation strengthen and reuse, low vanishing gradient problems, and reduces the parameters. It has used on CIFAR-10, CIFAR-100, SVHN, and ImageNet as state-of-the-art with less computational power.

2.3. Transfer learning

In computer vision applications like bioinformatics and robotics, the way from scratch need very high computation power (GPUs) and large datasets (Tan et al., 2018). To make much faster progress in training, using transfer learning might be interested and significant. Transfer learning allows us to sort of transfer knowledge from extensive data to small data. Even one can use CPU and small dataset to train the model much faster rather than randomly initialized weight from the scratch network. Deep Learning is a connection-oriented with three different areas, supervised (training data with labeling), unsupervised (training data without labeling), and semi-supervised learning (training data with few labeling). The convolutional neural network is a supervised technique that needs input "x" and labels output "y". It consists of convolution layers with activation function followed by pooling layers and ends up with fully connected to softmax layers. CNN has simplicity, scalability, and domain transfer-ability. Here different architectures of CNN have been studied, which followed the rule of thumb but have used different approaches like Plain network, residual, inception, and dense techniques. Every architecture has three stages: the input layer, hidden layers (convolution base), and output layers. Instead of using these dense architecture layers, we have changed the last three layers of

[1] https://doi.org/10.17632/y5sm4smnfr.5.

Fig. 1. Normal and abnormal sheep face images. The first row is the normal and second row for abnormal sheep faces by using SPFES.

Table 1
Sheep dataset statistical approach.

Categories	Training set	Validation set	Testing set	Total
Number of images	1650	350	350	2350
Percentage	82.5%	17.5	17.5%	100%
Number of normal Sheep faces	959	224	224	1407
Number of abnormal sheep faces	491	226	226	943

every architecture with weight and bias learn rate factor as a fine-tuning for a binary classifier. The fully connected layer outputs are binary vector representation for normal and abnormal sheep Images. Additionally, the softmax layer activation function used to predict the output: each architecture parameter and dimensions presented in Table 1. Mostly ReLU and Leaky ReLU have used in state-of-the-art pre-trained models as an activation function for hidden layers. The idea of transfer learning has shown in Fig. 2 for all trained models. Every architecture with hidden and dense layers has different parameters. To improve the performance of pre-trained models during training, we will introduce the most effective and advanced techniques like Layers Freezing, data augmentation, regularization, fine-tuning, visualization, and custom read function method for non-standard datasets. These techniques briefly explained in Section 3.

3. Deep learning methodology

In this section, a full pipeline of advanced deep learning techniques is used for both normal and abnormal sheep faces.

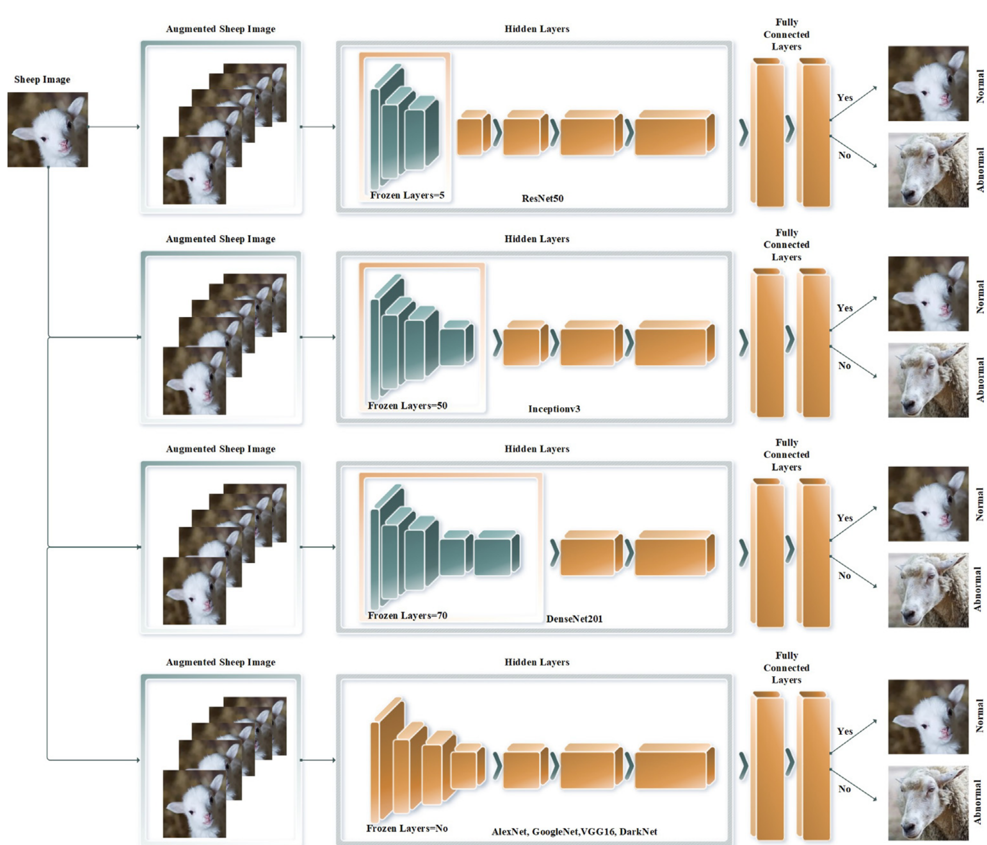

Fig. 2. Summary of all architecture's representation.

3.1. Layers freezing

Deep learning has fewest parameters but take time to train, and also sometimes the model does not end up with high accuracy. Freezing layers is a technique that speeds up the training process and increases efficiency and prevents the weights of initial layers from being modified. While initially, every model trained without layers freezing and then freeze the layers one by one. AlexNet, GoogleNet, VGG16, and DarkNet performed well with very high accuracy in a short time as compared to layers freezing. While ResNet50, Inceptionv3, and DenseNet201 achieved higher efficiency in a short time with freezing of 5, 50, and 70 layers, respectively.

3.2. Pre-processing

The first part of training a convolutional neural network is a dataset. Datasets are in two forms; standard (pre-processed by the organization and share among competitors as secondary data) and non-standard (used for the first time as a primary data). The standard datatype can be read by custom read function, but non-standard data should be passed through custom read function with pad-array. Without pad array, the shape of images changes by custom read function but did not able to classify with high performance during training. Custom read function made image shape constant during image processing for resizing and sharpening. We used different architectures which have different input size with a specific dimension image for training.

3.3. Data augmentation

Datasets are the key to deep learning and machine learning. Big datasets are very expensive, while small datasets caused a tendency of overfitting during training. Pre-trained models are susceptible to new unseen data, which memories the features of the training set, but model behavior does not generalize the validation set. Data augmentation is a way to avoid overfitting and produce new training and validation data from the existing smaller dataset. So, the data augmentation method used for creating new sheep images for the training. The simplest way of new data creation is practical data augmentation techniques, such as flipping, translating, rotation, and so on (Vasconcelos and Vasconcelos, 2017). We have used scaling, random rotation, translation, and reflection. The model generated different virtual images from each original image. The example of new images is shown in Fig. 3.

3.4. Regularization

Augmentation is not only the key to reduce overfitting. Instead of the model also tries to capture noise in the training dataset, and these noises are random data points which do not have exact properties of the data. We have used Ridge regression to make the model safe from under-fitting and over-fitting. Lambda is the tuning parameter of L2 regularization, and the selection of lambda during training was from 0.001 to 0.02.

Stochastic Gradient Descent with Momentum (SGDM) Stochastic gradient descent with momentum (Loizou and Richtárik, 2017) works better and faster than SGD due to the exponentially weighted average. It accelerated the gradient vector in optimal to actual derivative and more quickly converging toward the optimal solution. We used SGDM because classic SGD takes too much time, and Adam optimizer degraded the performance of training accuracy at smaller datasets.

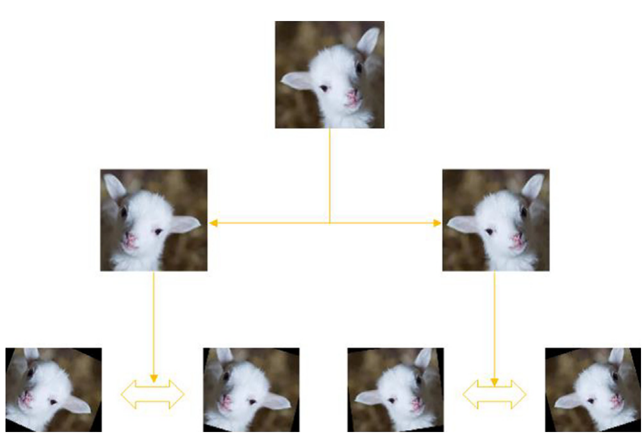

Fig. 3. Augmented images with flipping and rotation, the top image is original and bottom images augmented.

3.5. Fine-tuning

Hyper-parameters sittings are very important as a fine-tuning for a pre-trained model (Hermessi et al., 2019). We have not changed the learning rate for some architectures first layers by using layers freezing technique, while for others to adjusts the learning rate was necessary to train a model with slight changes in the weights of architectures first layers. When added new dense layers, the higher learning rate of the final layer is suitable for a faster change as compared to the first layers to update weights quickly. Having Mini-batch size and no freezing to must freezing layers techniques played an essential rule in getting very high accuracy of different architectures. Fig. 4 presents the workflow of the proposed method.

4. Experimental results

In the following section, we presented the training accuracy and loss of different models and their experimental results. Tools used for the given research have also been mentioned in this section.

4.1. Tools and setup

The experimental work has been performed with MATLAB 2018b (9.5.0) on Linux Server (4.20.13) with a single Nvidia GPU GeForce GTX 1070. We have used different CNN architectures on the sheep face dataset, which has been divided into training, validation, and test sets. The distribution of dataset is 70% for training and 15% each used for validation and test purpose, respectively.

4.2. Results

The results of state-of-the-art CNN architectures are obtained using data augmentation, L2 regularization, and fine-tuning techniques by applying deep learning techniques on the sheep face dataset. The performance of every pre-trained model has been presented in Fig. 5. *By* using these pre-trained models, we achieved very high accuracy and low loss during the experiment. The best performance stands at 100% and 98.17% for training with 99.69% and 97.8% for validation (with VGG16 and ResNet-50 architecture), which show a small difference of 0.31% and 0.37%. DenseNet-201 comes next at 98.03% with difference 1.02% between training and validation, which is marginally lower. Also, GoogleNet, DarkNet, Inceptionv3, and AlexNet showed excellent results with this sheep face classification task. Initially, all of the pre-trained models have been trained without freezing any layer in which VGG16, GoogleNet, AlexNet, and DarkNet performance was satisfied, but ResNet-50, Inceptionv3, and DenseNet-201 faced with high variance on the given dataset. After all, we freeze 1 to 10 layers of classic architectures, but accuracy did not improve, and unfortunately, the model caused overfitting problems. While on the other hand, we got high accuracy by freezing layers of ResNet-50, Inceptionv3, and DenseNet-201. Initially, we had to freeze a few layers but did not achieve bias level performance instead of minimal changes. Finally, it performed very high accuracy with low variance by freezing 5, 50, 70

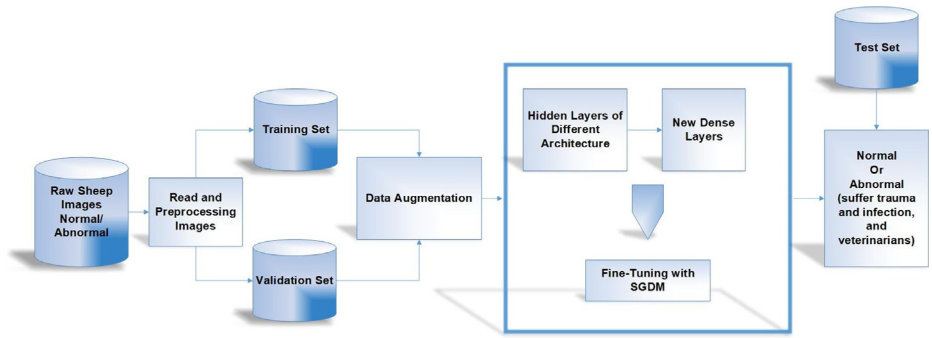

Fig. 4. Workflow of the proposed method for sheep face classification.

layers of ResNet-50, Inceptionv3, and DenseNet-201, respectively. To freeze more hidden layers, we degraded the training accuracy of the network but had very less overfitting problem. The final stage of the proposed method was to employ fine-tuning with freezing and unfreezing of the base convolutional layers. All the pre-trained models trained for 20 epochs. We changed the learning rates for the last layers and not modify the first hidden layers of learning rates because it had vibrant features obtained by trained on million images with different categories. Their weights updated during training, which worked for some state-of-the-art models, while others should freeze first layers by sitting the learning rate to zero. All models have learned features with 20 base and weight learning rates for the last layers instead of VGG16 and DarkNet, as given in Table 1. Fig. 6 shows the training and validation loss during the training of pre-trained models at sheep dataset.

Bayes error is a proxy for training error and beyond of its learning algorithm face with overfitting as we have used regularization to reduce the variance. Meanwhile, using the avoidable bias technique by applying different CNN models to minimize the difference between Bayes optimal error and training error. Therefore, VGG16 and ResNet50 performed with deficient error as compared to other CNN models, and their error closed to theoretical error. In Fig. 6, we saw that the validation loss of the pre-trained models is lower than the training loss in the first ten epochs but gradually changed by running the training set longer.

The classification performance of all CNN models using the confusion matrix for the test data shown in Fig. 7. We have predicted two possible classes normal and abnormal (diseased) Sheep faces. The classifier tested at 350 images, in which 224 has no pain and 126 with different pain levels. The diagonals show the true positives (TP) and true negatives (TN) for the classified test data. Outside, the diagonals predicted false positives (FP) and false negatives (FN), which showed the misclassification rate of the pre-trained models during test data. VGG16 leads at the top with zero misclassified images, while ResNet50 achieved an excellent result with only five false positive and for other pre-trained CNN models, as shown in Fig. 7.

The correctness and miss-classification rate of a test dataset should be checked and defined as given in Eqs. (1) and (2).

$$Accuracy = \frac{TP + TN}{Total} \tag{1}$$

$$ErrorRate = \frac{FP + FN}{Total} \tag{2}$$

In Eq. (3), sensitivity has also been defined as a TP rate. While Eq. (4) shows the ratio between TN values and actual negative values.

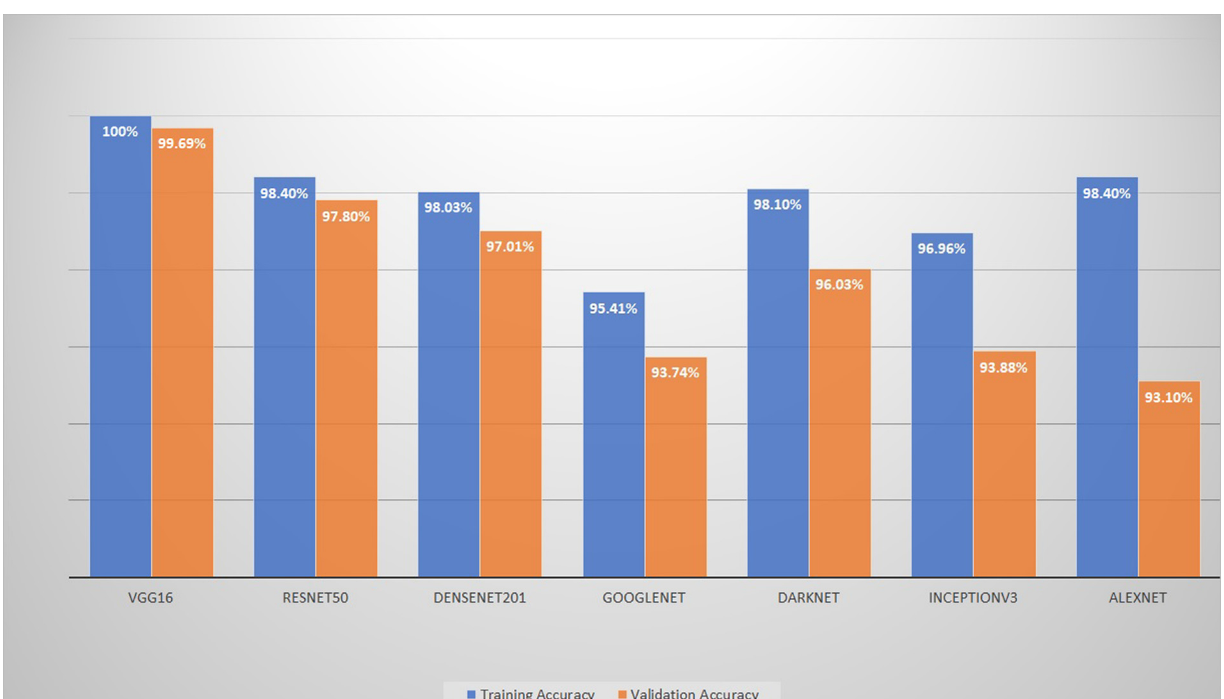

Fig. 5. Normal and abnormal sheep face classification on the different CNN state-of-the-art architectures.

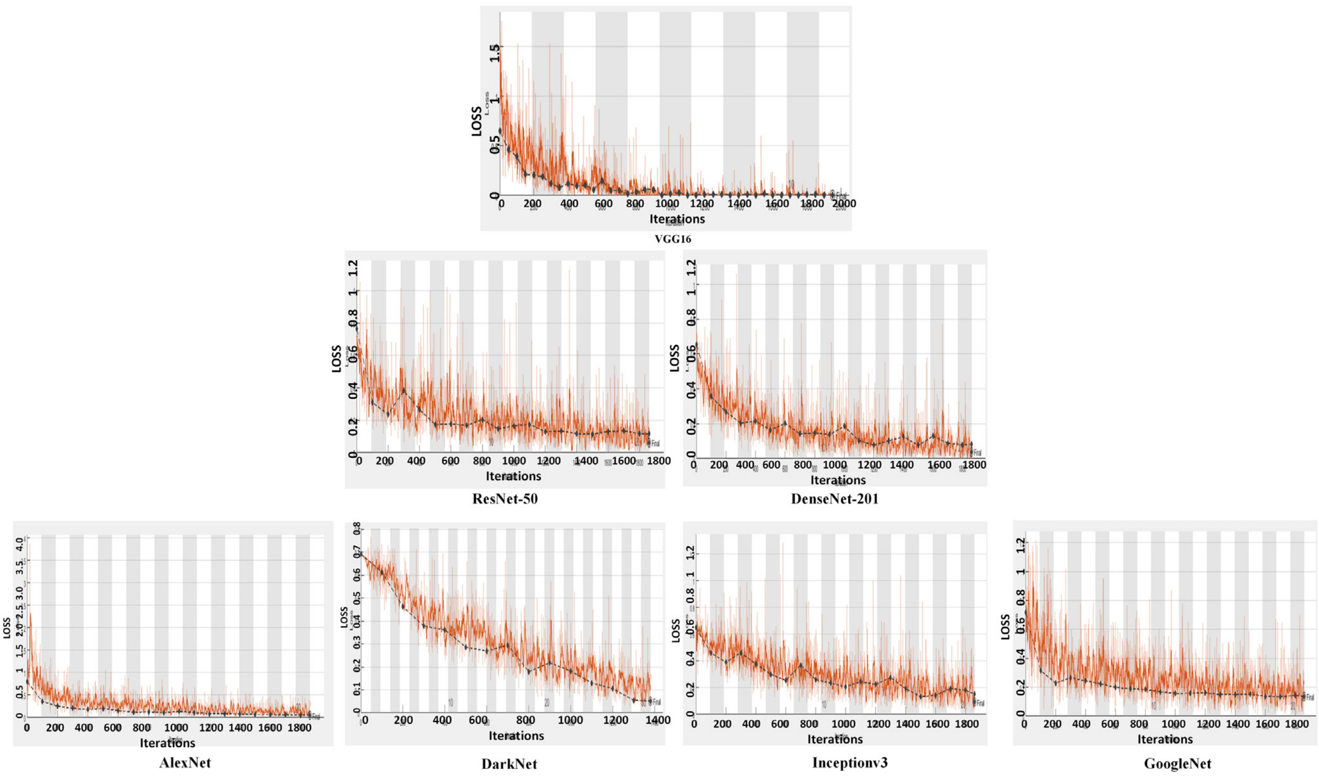

Fig. 6. Training and validation loss of all pre-trained models.

Fig. 7. Testing confusion matrix for all architectures with sheep face dataset.

Table 2
Models and layers with parameters.

Models and parameters	AlexNet	LeNet	VGG16	DenseNet-201	Inception-v3	ResNet-50	DarkNet
Input shape	227,227,3	224,224,3	224,224,3	224,244,3	299,299,3	224,224,3	448,448,3
Total Parameters	61M	60K	138M	20M	23.9	25.5M	51M
Dense layer	4096	1024	4096	1886	2048	2048	4096
Weight Learning Rate	20	20	10	20	20	20	10
Base Learning Rate	20	20	10	20	20	20	10
Binary Softmax	2	2	2	2	2	2	2

Table 3
The testing accuracy, Cohen's Kappa, F1 Score, Sensitivity and Precision using the Sheep dataset.

Models	Classes	Accuracy (%)	Error rate (%)	Se, Sp (%)	Precision (%)	CK (%)	F1 Score (%)
GoogleNet	Abnormal, Normal	96.29	3.71	90.86, 99.55	99.13, 94.89	90	94.61, 97.17
Inceptionv3	Abnormal, Normal	93.71	6.29	82.54, 100	100, 91.05	85.8	90.43, 95.32
DarkNet	Abnormal, Normal	95.71	4.285	98.41,94.20	90.51, 99.00	86.6	94.29, 96.57
AlexNet	Abnormal, Normal	96.29	3.71	90.50, 99.60	99.1, 94.9	91.77	94.61, 97.17
DenseNet201	Abnormal, Normal	97.43	2.57	94.40, 99.10	98.3, 96.94	94.24	96.36, 98.01
ResNet50	Abnormal, Normal	98.57	1.43	96.03, 100	100, 97.8	96.82	97.98, 98.90
VGG16	Abnormal, Normal	100	0.00	100, 100	100, 100	100	100.0, 100.0

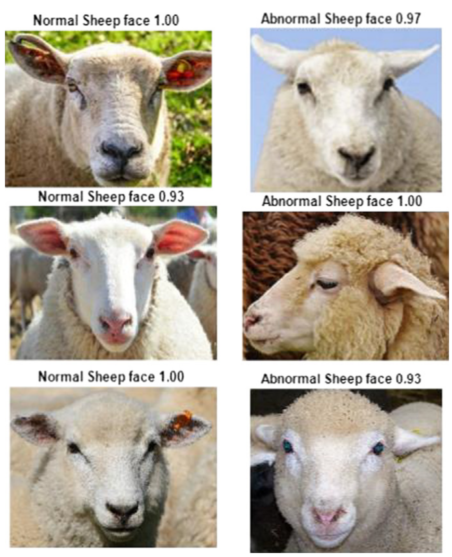

Fig. 8. The predicted outputs of normal and abnormal Sheep faces using pre-trained models. The first column represents the normal sheep faces, and the second column presents the abnormal sheep faces.

$$sensitivity = \frac{TP}{Actual\ Yes} \tag{3}$$

$$Specificity = \frac{TN}{Actual\ No} \tag{4}$$

Eq. (5) defines how often the model is correct in terms of the ratio of actual and predicted yes. And the prevalence determines how usually the yes condition occurs in our trained model, which is given in Eq. (6).

$$Precision = \frac{TP}{Predicted\ Yes} \tag{5}$$

$$Prevalence = \frac{Actual\ Yes}{Total} \tag{6}$$

Similarly, the null error rate (Eq. (7)) gives us a metric and a useful base when evaluating the model of how much is incorrect if the majority class is always expected.

$$Null\ Error\ Rate = \frac{Actual\ No}{Total} \tag{7}$$

While Cohen's Kappa is essentially a measure of how well the assessment was done relative to how well it actually by chance would have worked. In other words, if the model has a significant difference between the accuracy and the null error rate, a model will get a high Kappa score. As shown in Eq. (8) is;

$$Cohen's\ Kappa = \frac{P_o - P_e}{1 - P_e} \tag{8}$$

Where, $P_o = accuracy$ is an observed proportionate agreement and $P_e = P_{yes} + P_{no}$ is the probability of random agreement. The predicted likelihood of both showing yes to random is thus $P_{yes} = \frac{TP+FP}{Total}\cdot\frac{TP+FN}{Total}$. Similarly, both would predict no at random is $P_{no} = \frac{FN+TN}{Total}\cdot\frac{FP+TN}{Total}$.

F1 is a weighted average of the precision and recall, as shown in Eq. (9) and often referred to as either the F score or the F measure. F1 is generally more useful than accuracy, mainly if the class distribution is uneven. Accuracy works better if the model's FN and FP have the same costs. When the costs of FP and FN are significantly different, both precision and recall should be considered.

$$F1\ Score = \frac{2 * (Sensitivity * Precision)}{Sensitivity + Precision} \tag{9}$$

The accuracy (Ac), Error Rate (ER), Sensitivity (Se), Specificity (Sp), Precision (Pr), Prevalence (Pre), Null Error Rate (NER), Cohen's Kappa (CK) and F1 Score have been given in Table 2.

The Error Rate is very high except for VGG16 and ResNet50, which is 0 and 1.43, respectively. As shown in Table II, Cohen's Kappa of other pre-trained models is very low as compared to VGG16 and ResNet50 and DenseNet201. Therefore, VGG16 and ResNet have a significant impact on small datasets, especially animal datasets. (Table 3).

After the training of different convolutional neural network pre-trained models, some of the correct predictions of models for normal and abnormal sheep faces are given in Fig. 8.

Different type of machine learning techniques has developed for image classification like SVM, shallow classifier, discrete wavelet packet transforms, and so on. However, Deep learning has the advantage of transfer learning with powerful features end to end learning. Sheep face classification is very expensive and takes a very long time by manual assessment. The proposed technique is the first study in our knowledge to use transfer learning for classification of Sheep dataset as deep learning needs considerable power and big datasets. While using fine-tuning and transfer learning, we eliminated the problems of the large dataset and very high power. The disadvantage of small sheep faces dataset size reduced by using data augmentation. We have used different deep learning state-of-the-art architectures with the same epoch in which VGG16 and ResNet50 carried out on the top of all architectures. Another disadvantage is to use a small number of abnormal images compared to normal images. For this problem, we are collecting more images for a pain rating scale of Sheep faces.

5. Conclusion

Efficient and reliable pain estimation in sheep is essential for management decisions. In this paper, using the sheep face dataset, we proposed transfer learning and fine-tuning with state-of-the-art pre-trained CNN architectures to classify normal and abnormal sheep images. Our models achieved 93.10% to 100% with best VGG16 and ResNet50 of accuracy for training, validation, and testing during trained models. By the experimental approach, we demonstrate the computer-based automated assessment of sheep face classification, which is useful in terms of time-saving, expenses, and accuracy. We employed the augmentation technique to increase the number of sheep images virtually. Our proposed method also can be used and generalized across different datasets. We are also collecting more images for future work to classify multi-class pain rating scales of sheep faces instead of binary classification.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This Paper is supported by the National Natural Science Foundation of China, China [Grand number: 61671185]. Also,this work is supported by the Robotics and Internet-of-Things Laboratory of Prince Sultan University, Saudi Arabia.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.compag.2020.105528.

References

Dolan, S., Field, L.C., Nolan, A.M., 2000. The role of nitric oxide and prostaglandin signaling pathways in spinal nociceptive processing in chronic inflammation. Pain 86, 311–320.

Dolan, S., Kelly, J.G., Monteiro, A.M., Nolan, A.M., 2003. Up-regulation of metabotropic glutamate receptor subtypes 3 and 5 in spinal cord in a clinical model of persistent inflammation and hyperalgesia. Pain 106, 501–512.

Flecknell, P., 2008. Analgesia from a veterinary perspective. BJA: British Journal of Anaesthesia 101, 121–124.

Beausoleil, Guesgen, Stewart, Minot, Stafford, 2014. The lorentz transformation and absolute time. Applied Animal Behaviour Science 159, 41–49. https://doi.org/10.1016/j.applanim.2014.07.008.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C., A survey on deep transfer learning, arXiv:1808.01974.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp. 770–778.

Hermessi, H., Mourali, O., Zagrouba, E., 2019. Deep feature learning for soft tissue sarcoma classification in mr images via transfer learning. Expert Systems with Applications 120, 116–127. https://doi.org/10.1016/j.eswa.2018.11.025.

Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: 30th IEEE Conference on Computer Vision and Pattern Recognition, pp. 2261–2269.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Communications of the ACM 60, 84–90.

Leach, M.C., Klaus, K., Miller, A.L., Scotto di Perrotolo, M., Sotocinal, S.G., Flecknell, P.A., 2012. The assessment of post-vasectomy pain in mice using behaviour and the mouse grimace scale. PLos One 7 e35656. doi: 10.1371/journal.pone.0035656.

Loizou, N., Richtárik, P., 2017. Momentum and Stochastic Momentum for Stochastic Gradient, Newton, Proximal Point and Subspace Descent Methods. ArXiv, abs/1712.09677.

Lin, M., Chen, Q., Yan, S., 2014. Network In Network. CoRR, abs/1312.4400.

Lu, Y., Mahmoud, M., Robinson, P., 2017. Estimating sheep pain level using facial action unit detection. In: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 394–399.

Main, D.C.J., Kent, J.P., Wemelsfelder, F., Ofner, E., Tuyttens, F.A.M., 2003. Applications for methods of on-farm welfare assessment. Animal Welfare 12 (4), 523–528.

McLennan, K.M., 2018. Why pain is still a welfare issue for farm animals, and how facial expression could be the answer. Agriculture 8, 127.

McLennan, K.M., et al., 2016. Development of a facial expression scale using footrot and mastitis as models of pain in sheep. Applied Animal Behaviour Science 179, 105–107. doi: 10.1016/j.applanim.2016.01.007.

Napolitano, F., De Rosa, G., Ferrante, V., Grasso, F., Braghieri, A., 2009. Monitoring the welfare of sheep in organic and conventional farms using an ani 35 l derived method. Small Ruminant 83, 49–57.

Simonyan, Zisserman, 2014. Very deep convolutional networks for large-scale image recognition, preprint arXiv, arXiv:1409.1556.

Hektoen, Janczak, Valle, Zanella, 2011. Assessment of sheep welfare using on-farm registrations and performance data. Animal Welfare 20, 239–251.

Szegedy, C., et al, 2015. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

Redmon, J. Darknet: Open source neural networks in c, http://pjreddie.com/darknet/.

Szegedy, C., Vanhoucke, V., Ioffe, J.S.S., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV 2818–2826.

Vasconcelos, C.N., & Vasconcelos, B.N. (2017). Increasing Deep Learning Melanoma Classification by Classical And Expert Knowledge Based Image Transforms. ArXiv, abs/1702.07025.

Yang, H., Zhang, R., Robinson, P., 2015. Human and sheep facial landmarks localisation by triplet interpolated features. In: IEEE Winter Conference on Applications of Computer Vision (WACV). doi: 10.1109/WACV.2016.7477733.