# Dairy Cows Teat-End Condition Classification using Separable Transductive Learning

Youshan Zhang, Parminder S. Basran, Ian R. Porter, Matthias Wieland
Cornell University
Ithaca, New York, USA

## Introduction

Mastitis remains one of the most frequently occurring diseases in dairy cows, often arising from intra-mammary infections by way of the teat canal. Machine milkings can affect teat canal integrity and lead to increased teat-end callosity- this can increase the risk of bacterial infections. Frequent monitoring of teat-end callosity and hyperkeratosis is critical for a mastitis prevention program. However, cow-side manual assessments of teat-health, which is the current best-practice, is time-consuming and suffers from inter- and intra-rater variability [1]. Another challenge is the inability to assess the entire herd in large-dairy farms. To address some of these challenges, a deep learning (DL) has been proposed [2]. The overall accuracy of this approach shows promise (46.7-61.8%) but also highlighted the need for improvement. In this paper, we describe modifications to this approach which yield a substantial improvement in accuracy while retaining the flexibility and accessibility of commonly used image classifiers such as GoogLeNet.

## Methods and Materials

The goal of the image classification problem is to improve the overall accuracy of the test images after the convolutional neural network (CNN) has been trained. The parameters of the CNN may be fit by optimizing a cross-entropy loss function (CE) which contrasts the extent of randomness between labels and predictions. If CE is high, differentiating information is also high. Low accuracy in our earlier work suggests the subtle differences between the different images, and between the classes, are not well quantified with CE alone. Therefore, two additional approaches are explored to identify and express differences in teat-end scores when training the CNN: separation loss (SL), and transductive learning (TL). Separation loss (SL) improves the inter-class dispersion in the training data (compacting differences in triangles/hatches in Fig. 1) so that the boundaries between various classes can be separable, and an image in the same category becomes more associated with each other. To determine how different images within a teat-end score are, we calculate the covariance matrix of each categories' samples, and then minimize the structural similarity between every two categories' covariance matrix [3].

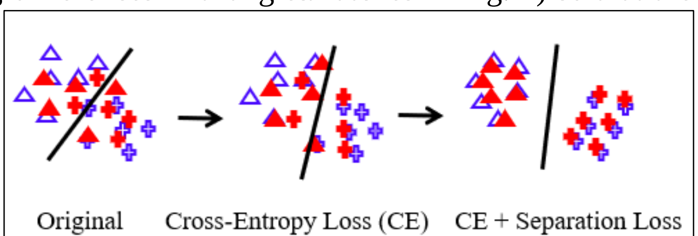
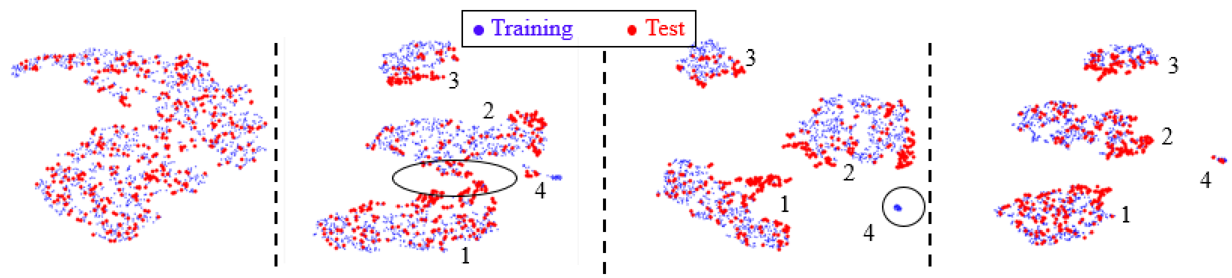



Fig. 1. An illustration of separation loss for training (blue) and test (red) data for 2 Classes.

To further improve the performance of the test dataset, we leverage transductive learning (TL) to mitigate the difference between the training and test data [4]. This approach intelligently separates the training and test teat-end images to optimize the deep learning. We employ the Maximum Mean Discrepancy loss to reduce the divergence between the training and test data [5]. We re-utilized the same dataset and the same network (GoogLeNet) as our earlier work but used a different set of training and test images for deep learning based on the TL approach [2]. The total number of teat-end images was 1529 in four categories (Score 1 to 4). Data was partitioned with 75% for training and 25% as test, and accuracies from the test data are reported. Training parameters were 100 epochs, 16 batch size, 3e-5 for learning rate, and ‘adam’ optimizer.

## Results and Discussion

Final accuracies from test data are: Original = 61.8%, CE = 67.1%, CE with SL = 70.8%, CE with TL = 69.2% and an aggregate of all 3 methods = 72.1% (10.3% higher than [2]). Note SL appears more important than TL, and both are higher CE. The proposed SL and the TL paradigm are effective in improving the performance of the test dataset. T-SNE was used to visualize these differences in 2 dimensions: this method demonstrates how the different algorithms cluster the teat-end scores [6]. The scores are not distinguishable with CE alone (Fig. 2a showing all 4 scores indistinguishable from one another) but become more discriminative after introducing SL and TL into the loss function. Comparing Fig. 2b and Fig. 2c with Fig. 2d, the scores cannot be correctly classified if the network is trained with a single loss function. There is contamination between Score 1 and 2 in for CE and CE+SL. CE+SL clusters all scores but does not carefully separate the training / test data for Score 4 (Fig 2c). Fig. 2d illustrates inter-class dispersion and intra-class compactness of the test datasets.




(a) CE (67.1%) (b) CE + SL (70.8%) (c) CE + TL (69.2%) (d) CE + TL + SL (72.1%)

Fig. 2. t-SNE views, which show how the system clusters the different Scores (and accuracies) for training and test dataset with different loss functions and different Scores.

To conclude, by carefully separating training and test data, and refining the loss function, we can improve the accuracy of deep learning classification of teat-end condition by as much as 10.3%.

## References


[1] Basran, P.S., Porter, I.R., Wieland M (2020). Technical Note: A digital technique and platform for assessing dairy cow teat-end condition. J. Dairy Sci. 103: 10703-10708.

[2] I. R. Porter, M. Wieland, P. S. Basran. (2021). Feasibility of the use of deep learning classification of teat-end condition in Holstein cattle. J. Dairy Sci., 104(4), 4529-4536.

[3] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, 2004.

[4] Y. Zhang, S. Xie, and B. D. Davison. Transductive learning via improved geodesic sampling. In Proceedings of the 30th British Machine Vision Conference, 2019.

[5] M. Long, H. Zhu, J. Wang, and M. I. Jordan. Unsupervised domain adaptation with residual transfer networks. In Advances in Neural Information Processing Systems, pages 136–144, 2016.

[6] L. Van der Maaten and G. Hinton, “Visualizing data using tSNE,” Journal of Machine Learning Research, vol. 9, no. 11, 2008.