Innovative Application of Deep Learning in Dairy Cow Teat Assessment

Xin Xiang Xxiang@mail.yu.edu

Abstract

The health of a cow's teats is crucial for ensuring the production of high-quality milk and overall udder health. Early detection of health issues in cow teats can facilitate timely treatments that improve health outcomes and consequently enhance milk yield and quality. Currently, a plethora of image recognition algorithms exists that can effectively identify various health states of cow teats. We employed convolutional neural networks (CNN) to assess teat alterations. However, the scarcity of datasets for such niche applications poses a significant challenge to training robust models. This paper introduces a CNN classification model specifically designed to identify varying degrees of teat hyperkeratosis. These datasets originate from real farm environments, documenting scores from cow teats during health assessments. Our proposed model leverages learning from ResNet's residual blocks and GoogLeNet's inception blocks to boost accuracy, especially considering the limited data availability. From a baseline of 50%, our method improved the accuracy to 68%, showcasing its potential for routine monitoring in commercial dairy farm settings. This innovative approach and promising results underscore the model's value and its revolutionary potential in ensuring superior milk quality and bovine health by assessing teat health.

1. Introduction

With the increasing global demand for dairy products, safeguarding the health of dairy cows and enhancing the quality of milk has emerged as a focal point in the dairy sector. Particularly, the health of a cow's teats takes precedence since they are integral parts of the mammary gland, directly impacting both the wellness of the udder and the caliber of milk generated. Over-keratinization at the teat-end, manifested by its thickening and roughness, has been pinpointed as a risk factor for mastitis. Traditionally, farmers and veterinarians have depended on manual visual examinations to assess the state of the teat-ends. However, this approach is not only time-intensive and expensive but also subject to potential biases due to individual interpretations.

The digital era, accompanied by breakthroughs in machine learning, has unlocked innovative avenues for the automation of teat health assessments. Specifically, employing convolutional neural networks (CNNs) for the categorization of digital images has proven its feasibility. Inspired by the architectural design of VGG, particularly VGG16 [1], our method integrates elements that have been instrumental in achieving significant accuracy in image classifications. Yet, the niche application character of this domain means relevant datasets are in limited supply, introducing hurdles for robust model training and verification.

To bridge this gap, we've put forth a fresh computer vision approach, amalgamating residual connections, elements from the Inception methodology, and concepts derived from VGG, for a nuanced classification of hyperkeratosis at the teat-end. Our dataset, sourced from Cow_teat [2], presents authentic farm scenarios, chronicling scores assigned during teat health evaluations. The ambition driving this study is to exhibit how, even under constraints posed by limited data, this method can achieve exemplary accuracy. Moreover, we want to showcase its adaptability for routine surveillance within commercial dairy settings. Through such an innovative approach, the dairy industry is poised to not only ensure top-notch health for their dairy cows but also enhance milk quality and diminish mastitis-associated risks.

2. Related Work

2.1. Deep Learning in Cow Teat Health Evaluation

The health of the teat end in cows is crucial for maintaining milk quality and the overall welfare of the cows. Traditional methods of teat-end evaluation involve manual visual inspection, primarily aiming to identify the thickness and roughness of scabs on the teat end. This is vital as these parameters might signify excessive keratinization, a primary risk factor for mastitis. Despite its importance, the manual method has inherent limitations. It is both time-consuming and expensive and often plagued by subjectivity due to inter- and intra-observer variability. Moreover, conducting comprehensive herd evaluations on large dairy farms is almost infeasible. The recent advancements in artificial intelligence, especially deep learning, offer new avenues for these long-standing challenges (literature search required).

2.2. Image-based Teat-end Classification

Utilizing Convolutional Neural Networks (CNNs) to classify teat-end alterations based on digital images has become a focal point of research. The feasibility of this approach was first demonstrated using VGG and AlexNet, which employed multiple convolutional layers to classify different degrees of teat-end hyperkeratosis based on a widely accepted four-point scoring system[4]. Although promising, the initial attempts only achieved accuracy rates ranging from 46.7% to 61.8%, strongly suggesting room for improvement. A frequently highlighted challenge is the highly uneven dataset distribution, where different hyperparameter designs lead to issues like the train loss not decreasing properly, severe oscillations, or the val loss outright overfitting. These issues make improving the classification of teat-end conditions challenging.

2.3. Deep Learning Architectures Introduction

Several renowned deep learning architectures have been explored for teat-end classification:

AlexNet: One of the early architectures that sparked interest in deep learning for computer vision tasks. Its depth and convolutional approach make it suitable for various image classification tasks.

VGG: Known for its depth, VGG offers multiple variants with different levels. Its consistent use of 3x3 filters sets it apart, and its performance on image datasets makes it a popular choice for image classification tasks.

GoogLeNet: Also known as Inception, it introduced the inception module, allowing the network to view image data at different scales, enhancing its classification capabilities.

ResNet-50: Pioneering the concept of residual learning, ResNet-50, with its 50 layers, addressed the vanishing gradient problem. Its design allows the training of deeper networks, making it capable of handling complex image classification tasks, including teat-end classification.

Pseudo Labeling: A technique that has garnered interest among researchers is pseudo-labeling. This method involves generating labels for unlabeled data to provide guidance during the learning process[19]. Labels can be generated based on hard-assigned labels (typically predictions from a neural network) or predicted class probabilities. The overall objective is to seamlessly integrate label information from unlabeled data during training. Although the prospects are good, a persistent challenge is ensuring the accuracy of the generated pseudo-labels. Noisy pseudo-labels might inadvertently harm model performance.

3. Methods

3.1. Dataset and Preprocessing

In our study, we utilized a large-scale dataset of cow teat-end images for training and validation. The validation set was allocated using stratified sampling to enhance the accuracy of model evaluation. The images were augmented by appropriately scaling, flipping, rotating, cropping, and adjusting the colors of the training set images to enhance the model's recognition capabilities. All images were normalized and resized to meet the input requirements of the deep learning models.

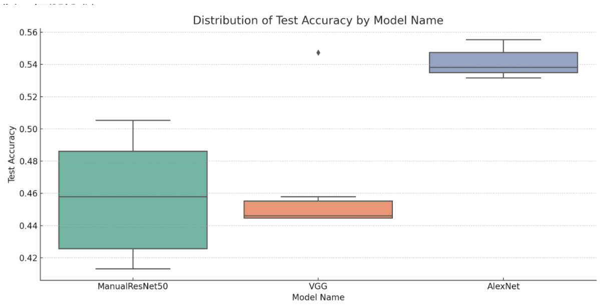


Figure 1. Model Performance

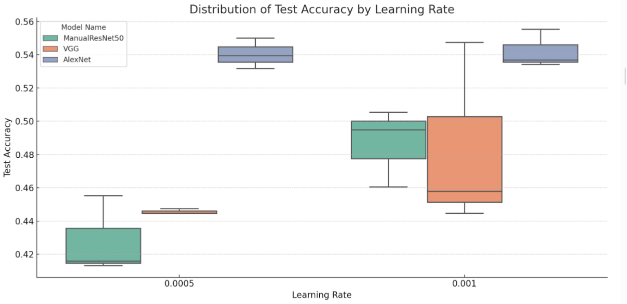


Figure 2. Accuracy by Learning Rate

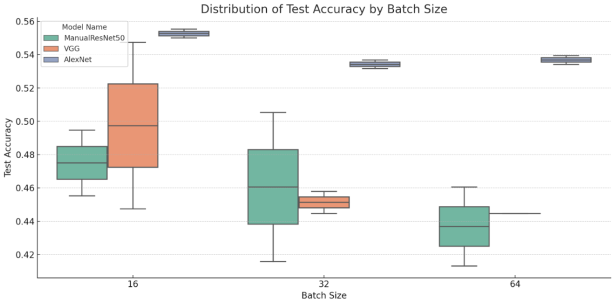


Figure 3. Accuracy by Batch Size

3.2. Using modern Models and Parameters to Inspire Model Design

Initially, we employed the VGG, AlexNet, and ResNet50 models, running them for 100 epochs with the Adam optimizer, using learning rates (0.001, 0.0005, 0.0001) and batch sizes (16, 32, 64). We also incorporated early stopping with patience for model training. However, I quickly noticed the models' accuracy was suboptimal, fluctuating between 55% and 41%. Upon further examination of the model data:

3.2.1 Model Performance Analysis

As show in Figure 1. AlexNet [?] exhibited the highest distribution of test accuracy with minimal variability. ResNet [?] displayed a broad distribution for test accuracy, indicating certain volatility in the model's performance. VGG had a relatively lower median test accuracy, but its upper quartile was on par with the other two models. Conclusion: a simpler model might perform better.

3.2.2 Relationship Between Parameter Quantity and Performance

Both AlexNet and ManualResNet50, classified as having "low" parameter quantity, exhibited varying distributions in test accuracy. Notably, AlexNet had a superior accuracy distribution. VGG [?], categorized with a "high" parameter count, demonstrated a wide distribution in test accuracy. Among these models, despite VGG possessing the most parameters, its accuracy did not distinctly outperform the other models.

3.2.3 Analysis of Learning Rate and Performance Relationship

As show in Figure 2. For all three models, the test accuracy distribution at a learning rate of 0.001 was relatively broad, indicating variable performance. For AlexNet and VGG, the median test accuracy seemed higher at a learning rate of 0.0005. ManualResNet50's performance appeared consistent across both learning rates, with no discernible difference. Conclusion: A learning rate of 0.001 might yield better results for complex models, while 0.0005, a relatively smaller rate, could stabilize the performance of simpler models. Sec. 2

3.2.4 Batch Size and Performance Relationship

As show in Figure 3. For both AlexNet and VGG, the distribution of test accuracy appeared relatively stable across different batch sizes. ManualResNet50, at a batch size of 32, exhibited a slightly superior distribution compared to the other batch sizes. Conclusion: A batch size of 16 seems better. Gradient descent computes the loss for each batch, iterating accordingly. More batches mean more iterations, leading to better convergence, albeit at the cost of increased training time.

3.3. Loss Change and Overfit Analyse

As show in Figure 4. Upon analyzing the loss change and overfitting graph, we found that most model combinations demonstrated steady performance, particularly ManualResNet50, which stood out in terms of loss reduction speed and overall stability. While VGG and ManualResNet50 have similar parameter counts, the latter outperformed, suggesting its architecture might be more fitting for this dataset. Interestingly, certain models, like AlexNet, tend to overfit with smaller batch sizes and higher learning rates. Overall, a learning rate of 0.001 generally outperforms 0.0005, offering quicker convergence. For optimal performance, the right balance of model structure, learning rate, and batch size is paramount, with ManualResNet50 shining especially with smaller batches and moderate learning rates.

Briefly summarizing the characteristics of the three models: AlexNet is renowned for its high accuracy but tends to overfit. ResNet excels in steadily reducing loss, but due to the application of early stopping, it might halt training prematurely, leaving room for optimization. In contrast, VGG, while stable, doesn't particularly stand out in any specific performance metric.

3.4. Summary

To enhance model accuracy, we aim to begin with a simple model, drawing from the architectural experiences of VGG and AlexNet. We plan to design an architecture with around 20 layers, but to prevent overfitting, we intend to incorporate residual blocks and inception blocks to retain more features. To stabilize loss reduction, we'll use a learning rate of 0.0005 and, to balance loss reduction and training time, we'll iterate for 32 epochs. Early stopping will be omitted to allow for more iterations.

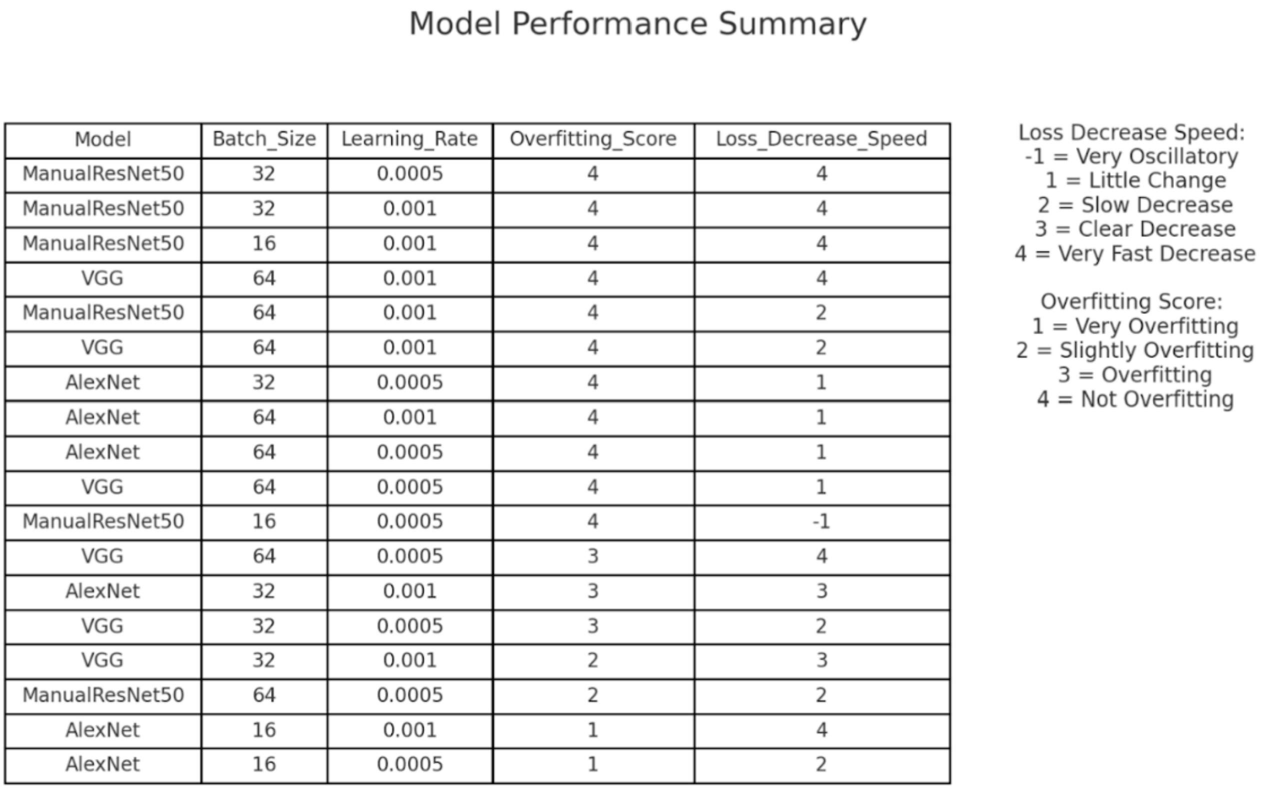
3.5. Model Architecture Design:

In this research, we designed a deep learning model named CowTeatResnet, which combines Residual Blocks and Inception modules. As show in Figure 5. By amalgamating these two modules, we aim to harness their strengths for enhanced classification accuracy. Specifically, the model first extracts features from the input image through a convolutional layer. These features are then passed through a combination of two sets of residual and inception blocks, using the residual blocks to prevent gradient explosion and remember more of the previous images. The Inception module is then used for more advanced feature combination and extraction. Finally, the model's predictions are output through three fully connected layers, corresponding to the number of target classes.

4. Results

4.1. Datasets

As show in Figure 6.Our dataset originates from multiple commercial dairy farms, and after continuous collection and integration, it encompasses thousands of digital images



Model Performance Summary

Model	Batch_Size	Learning_Rate	Overfitting_Score	Loss_Decrease_Speed
ManualResNet50	32	0.0005	4	4
ManualResNet50	32	0.001	4	4
ManualResNet50	16	0.001	4	4
VGG	64	0.001	4	4
ManualResNet50	64	0.001	4	2
VGG	64	0.001	4	2
AlexNet	32	0.0005	4	1
AlexNet	64	0.001	4	1
AlexNet	64	0.0005	4	1
VGG	64	0.0005	4	1
ManualResNet50	16	0.0005	4	-1
VGG	64	0.0005	3	4
AlexNet	32	0.001	3	3
VGG	32	0.0005	3	2
VGG	32	0.001	2	3
ManualResNet50	64	0.0005	2	2
AlexNet	16	0.001	1	4
AlexNet	16	0.0005	1	2

Loss Decrease Speed:
-1 = Very Oscillatory
1 = Little Change
2 = Slow Decrease
3 = Clear Decrease
4 = Very Fast Decrease

Overfitting Score:
1 = Very Overfitting
2 = Slightly Overfitting
3 = Overfitting
4 = Not Overfitting

Figure 4. Overfitting Score and Loss Decrease Table

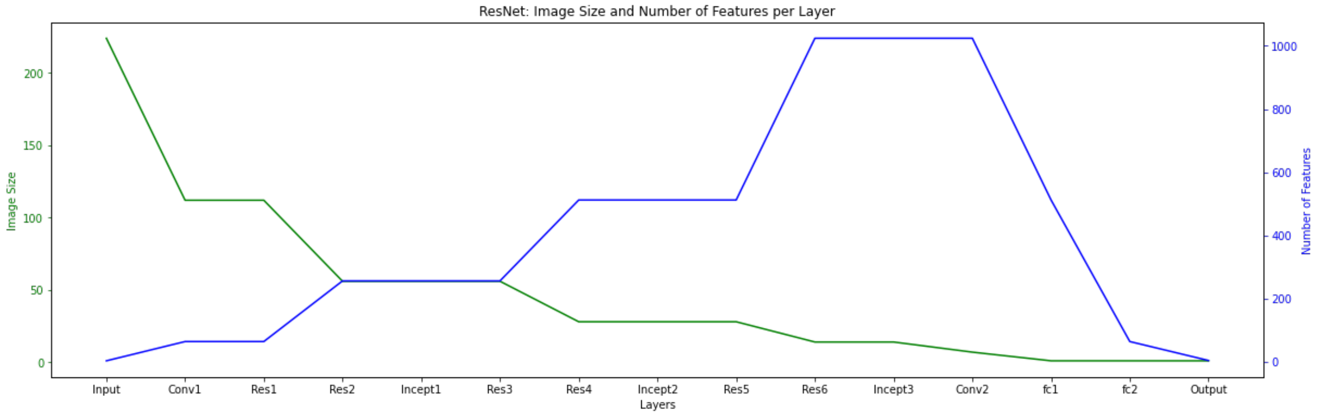


Figure 5. the model first extracts fea- tures from the input image through a convolutional layer. These features are then passed through a combination of two sets of residual and inception blocks, using the residual blocks to prevent gradient explosion and remember more of the previous images.

of cow teats. These images not only depict various health states of the teat-ends but also span cows of different breeds, ages, and physiological statuses. The dataset comprises a total of 1529 images, with 1149 images in the training set and 380 in the test set. Furthermore, these images are annotated with corresponding labels meticulously curated by experts.

Within these images, we have identified several primary health conditions of the teat-end, including normal, mild keratinization, moderate keratinization, and severe keratinization. Each category has its distinct visual attributes, such as color, texture, and shape. As shows in the figure6.

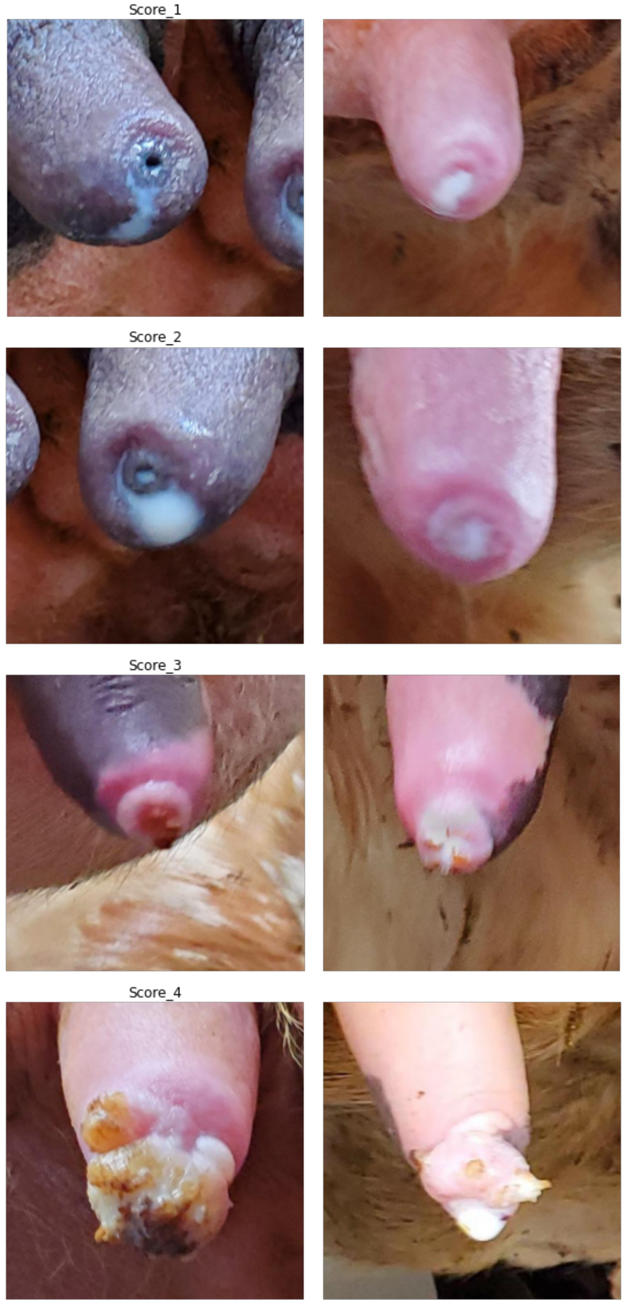


Figure 6. 4 Typical Classes of Cow teat-end in Health evaluation

To offer readers an intuitive understanding, Figures 1-4 display representative images of these four categories. Within our dataset, these categories are denoted as score1, score2, score3, and score4, respectively, with score1 indicating a healthy state and increasing scores signifying escalating degrees of keratinization.

As shows in figure7. The dataset distribution is not uniform. Images with mild keratinization are the most abundant, accounting for about 43%; while normal, moderate, and severe keratinized images make up approximately 39%, 16%, and 2%, respectively. This distribution mirrors the actual health statuses of teats in real farm settings, providing us with an authentic data foundation for model training.

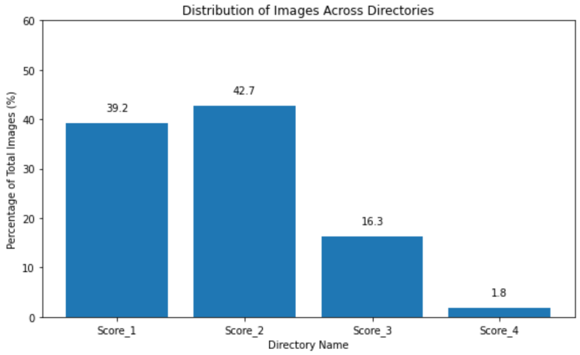


Figure 7. Classes Distribution

Despite the relative vastness of our dataset, it still poses certain challenges. Firstly, since the dataset is sourced from different farms and periods, there are inherent variances in images in terms of lighting, angle, clarity, and image dimensions. Secondly, the boundary between certain categories is not always clear-cut, making image classification more challenging. Moreover, due to the limited number of images in certain categories, such as severe keratinization, the model's performance in recognizing these categories may be compromised.

4.2. Training and Parameters

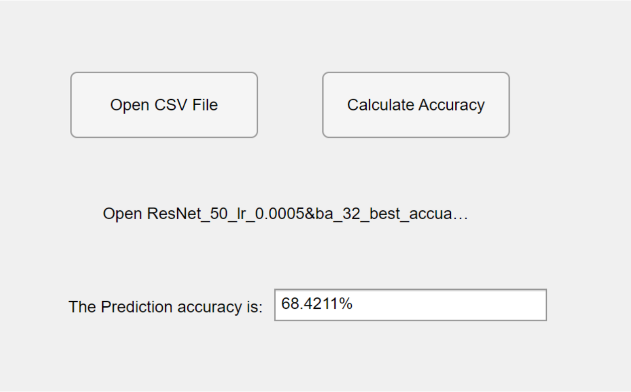
During training, the specific parameters of our model were as follows:

Learning rate: 0.0005 Batch size: 32 Training epochs: 1000 Every 100 epochs, the training and validation losses were saved and plotted. To optimize the model, we used an optimizer with a learning rate of 0.0005 and set 1000 training epochs. Every 100 epochs, we saved the model parameters and plotted the training and validation losses so that we could observe the model's performance throughout the training process.

Upon completion of training using the above methods, the CowTeatResnet model achieved an accuracy of 68% on the validation dataset. Compared to other benchmark methods, this is a competitive result.

5. Discussion

Even though our model achieved an accuracy of 68% on the validation dataset, there are still some challenges and limitations. For instance, the model might not perform well on certain teat-end categories, or struggle to recognize under specific background or lighting conditions. Additionally, combining Residual Blocks and Inception modules



Open ResNet_50_lr_0.0005&ba_32_best_accua…

The Prediction accuracy is: 68.4211%

Figure 8. Test Accuracy Score

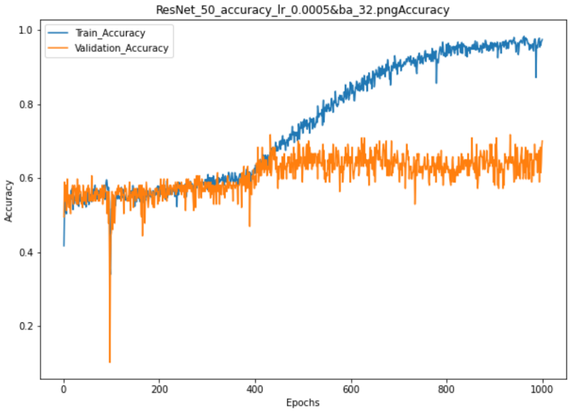


Figure 9. Accuracy Change by Epoch

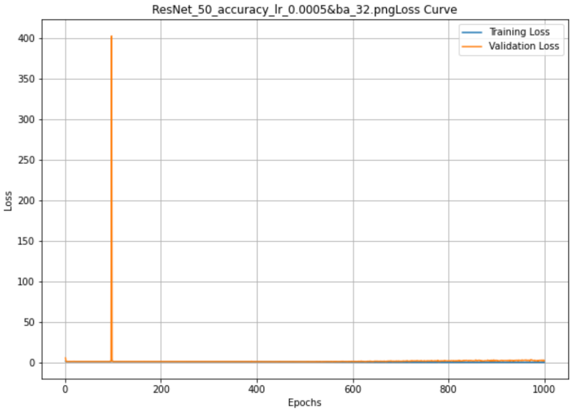
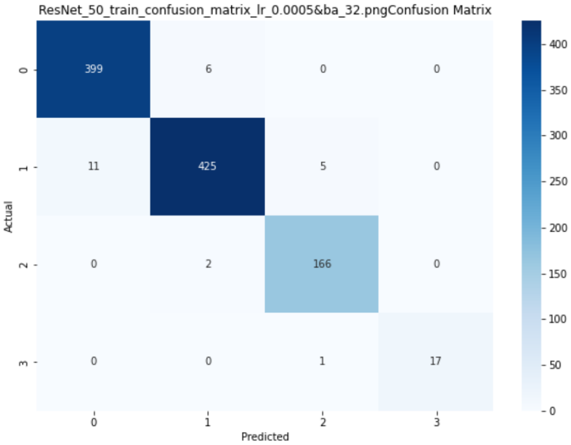


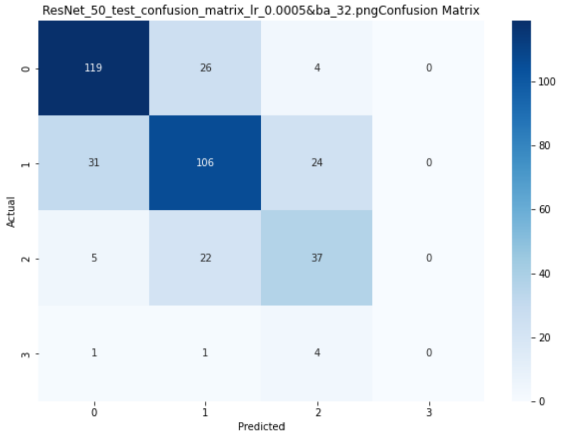
Figure 10. Loss Change by Epoch

might result in an overly simplified model, which may not deeply capture the nuances of the images, leading to potentially imperfect feature recognition.



	0	1	2	3
0	399	6	0	0
1	11	425	5	0
2	0	2	166	0
3	0	0	1	17

Figure 11. Train Confusion Matrix



	0	1	2	3
0	119	26	4	0
1	31	106	24	0
2	5	22	37	0
3	1	1	4	0

Figure 12. Test Confusion Matrix

6. Conclusion

Combining Residual Blocks and Inception modules, our CowTeatResnet model has demonstrated promising potential in the automated assessment of cow teat-end health. Although there are challenges regarding the model's accuracy and its performance under certain conditions, this approach provides a reliable starting point and foundational basis for research. Given the observed variations in the model's performance across different hyperparameters, learning rates, and batch sizes, as well as the distinctions noted in various training strategies and data preprocessing methods, future work can delve into optimizing the model's architecture, fine-tuning parameters, and exploring advanced data augmentation strategies to enhance accuracy and robustness. This lays a solid groundwork for deeper research and applications in the future.

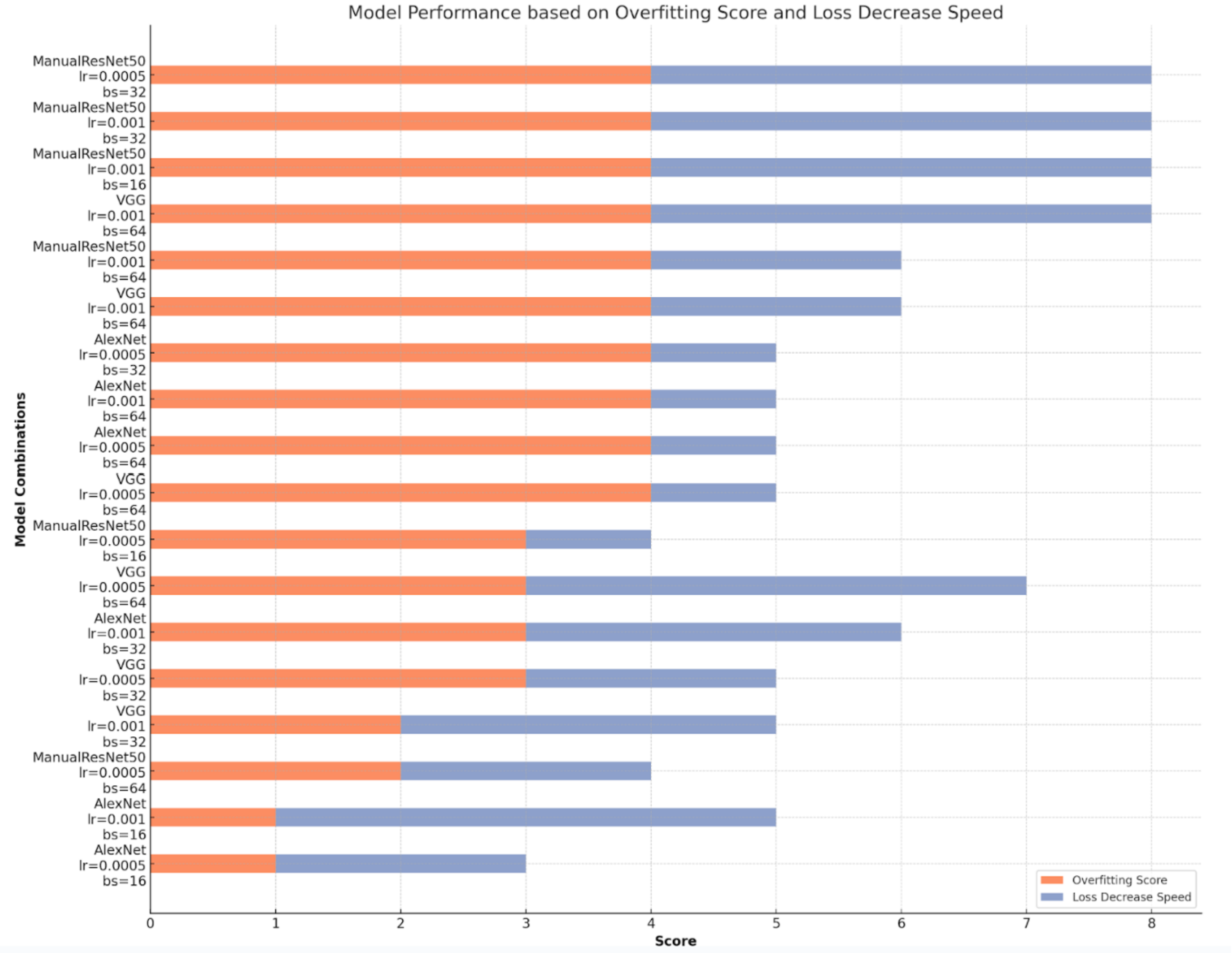


Figure 13. Overfitting Score and Loss Decrease Speed

References

[1] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*, 2015. 1

[2] Youshan Zhang, Parminder S Basran, Ian R Porter, and Matthias Wieland. Dairy cows teat-end condition classification using separable transductive learning. In *61st National Mastitis Council (NMC) Annual Meeting*, 2022. 1