Master Thesis

# A improved model for wheat diseases detection on small dataset

by

**Kairui Wang**

(2731737)

*First supervisor:* Hongyu Liu
*Daily supervisor:* Hongyu Liu
*Second reader:* Anna Bon

July 19, 2024

*Submitted in partial fulfillment of the requirements for the VU degree of Master of Science in Computer Sciences*

# A improved model for wheat diseases detection on small dataset

Kairui Wang
Vrije Universiteit Amsterdam
Amsterdam, The Netherlands
k4.wang@student.vu.nl

## ABSTRACT

*Context.* Identifying and treating these diseases has become a significant challenge due to their rapid spread and limited knowledge among farmers. Recent research suggests that state-of-the-art deep learning models with pre-trained weights on large datasets can achieve high accuracy (above 95%) in disease identification. However, Existing popular models with high accuracy are based on massive data. To address this, this paper proposes a Siamese network with the triplet loss function as a more efficient and practical approach, as it outperforms around 10% accuracy than existing models with a cross-entropy loss function on small datasets. This approach may help farmers develop a tailored identification model for their wheat and crops.

## 1 INTRODUCTION

In recent years, with the rapid development of information technology, machine learning, image processing, and other technologies to identify crop disease images, research hotspots in disease prevention and control have become new[1, 2]. Early methods of identifying crop diseases[3] based on image processing technology had disadvantages, such as manually extracted feature sets that may not be representative, incomplete, or redundant. In comparison, deep learning networks can automatically learn features and have achieved good application results in crop disease image classification.

However, existing data sets are still challenging in meeting the data volume requirements of wheat disease identification methods based on traditional deep learning, and one salient sample is Plant Village[4], which has 50000 crop images. Still, the number is small regarding the specific crops, and the variety of diseases is not comprehensive enough. While there are many types of crop diseases, single-type samples are occasional, and even the incidence of some diseases is strongly correlated with the region, which causes the versatility and generalization of the model to be guaranteed. From previous researches[5, 6], we found that part of their data is collected online, part of the data is collected in filed by their camera, while the photographed data is not open to the public, which leads it is tough to create a rich and complete data set.

Even so, there are still some studies that yielded promising results in the detection of wheat diseases. Lakshay Goyal et al.[7] proposed an improved architecture with a higher accuracy rate than VGG16 and RESNET-50. Genaev et al.[8] delve into EfficientNet-B0 and finetune its parameters, proving that EfficientNet-B0 can achieve better accuracy than most mainstream models. However, both of them are working on a large data set collected from various sources. In addition, they have no guarantee that the models they experimented with can be used in other real datasets or fields that do not have access to amounts of data.

Therefore, studying methods to solve crop disease(wheat in our paper) identification problems from the perspective of small sample detection has vital practical significance. Although some researchers have progressed in this area[9], they mainly focus on combining the image and text models, leveraging the correlation and complementarity between the two types of disease data for collaborative recognition. This paper aims to find a way that has declined when few samples have superior accuracy compared to directly porting over models training on large data sets. Thus, we propose an improved method based on small samples to overcome these limitations and compare the EfficientNetB0 model with triplet and cross-entropy loss. The traditional Cross-Entropyloss may face weak noise-robust and model overfitting during training, convergence, and other issues[10].

To solve these problems, we introduce triplet loss to optimize the feature representation of the model. Our experimental results show that the method based on triplet loss can improve the performance of the EfficientNetB0 model [11] in plant disease identification tasks in small sample cases. This improved method can not only alleviate the class imbalance and overfitting problems but also improve the generalization ability and robustness of the model. Therefore, it has the potential to become an important research direction in the field of plant disease identification in the future, providing farmers with more accurate and reliable disease identification tools and helping them take timely measures to protect crops and improve crop yield and quality.

We summarize our contributions as follows.

⋆ Propose a new method: a Siamese network with the triplet loss function, which can improve the wheat disease detection accuracy and efficiency compared with traditional deep learning models ( with cross-entropy loss function) on a small dataset with no supporting data.

⋆ Get a higher accuracy and performance by the Siamese network compared to the triplet loss function and traditional deep learning models (EfficientNet B0) on collected and limited data and analyze their classification and clustering ability, exploring the trend of the performance gap between the two models changing with the amount of data.

## 2 RELATED WORK

Early crop disease identification methods mainly relied on image processing technology. Among them, manually extracting feature sets had some problems[12], such as being labor-intensive, needing agricultural engineers and phytopathologists to be carried out properly, and the labeling made by humans could be incomplete or redundant. In contrast, deep learning networks can automatically learn features and perform well in crop disease image classification. To improve the effect of crop disease identification. Khan et al. [5] point out that deep learning models, mainly CNN, perform

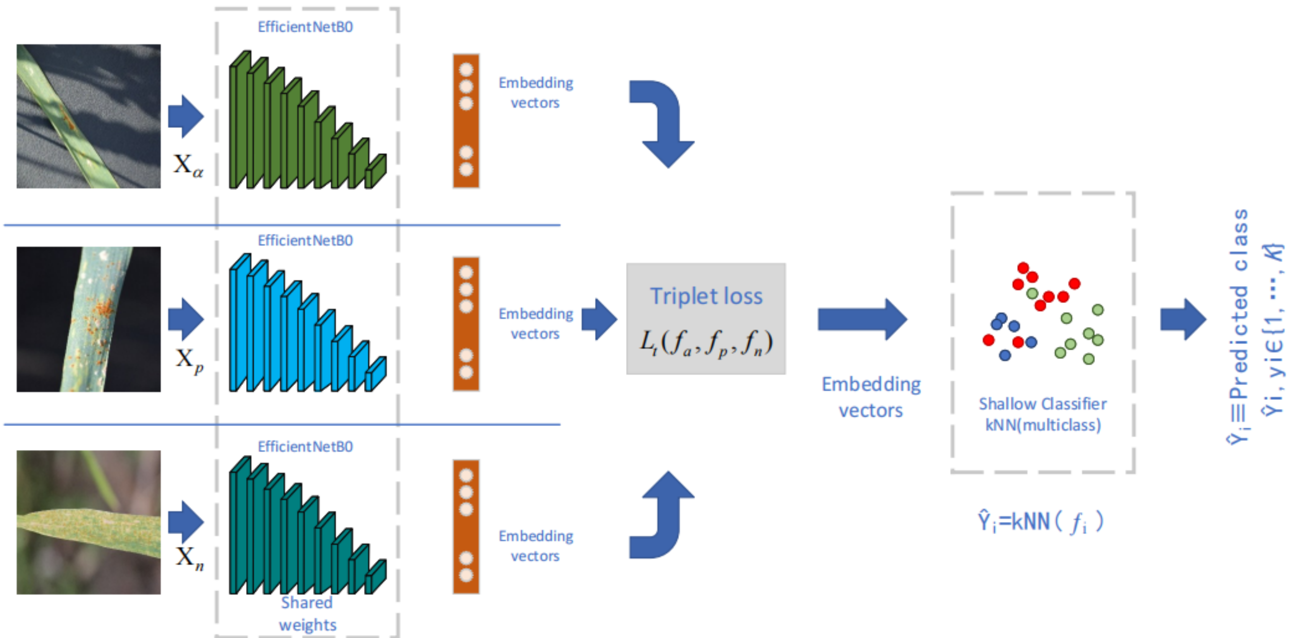



**Figure 1: Architecture of proposed model**

much better than traditional image processing(best accuracy reach 99.53%).

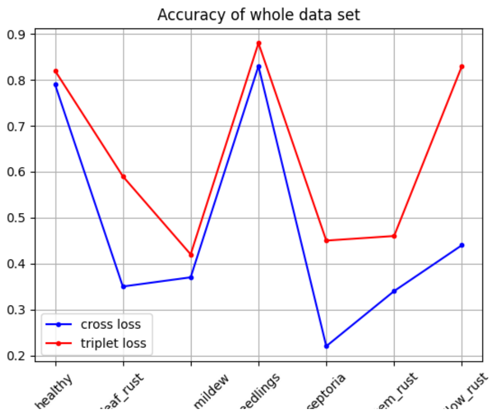
In recent years of research, the deep learning model yielded satisfactory results when supported by sufficient data. Jiang proposed an improved VGG16 model based on multi-task learning ideas to study three rice leaf diseases and two wheat leaf diseases[13]. Comparative experimental results show that this method is better than the single-task model, ResNet50 model, and DenseNet201 model. Gao proposed a dual-branch residual neural network (DECA ResNet) to solve the problem of low accuracy of existing crop disease identification methods. This model gets good results in disease identification on the Challenger2018 data set[14]. From the review of Hamed et al., we can see that most mainstream deep learning models can achieve accuracies beyond 95% on their data set.

Although breakthroughs have been made in locating and identifying crop diseases and pests using deep learning models, the models' performance still depends on the number of samples in the training data set. Therefore, learning techniques that can be learned from a small data set, also known as few-shot learning(FSL), are needed. A lot of work has been done in this area. They utilize techniques like transfer learning, data augmentation, meta-learning, etc. In the study by Afifi et al. [15], they take data from Plant Village[4] and coffee leaf dataset[16] as source domain and train a CNN on it, then use it as the feature extractor for small sample recognition model which boosts the small recognition model to a promising accuracy. Nuthalapati and Tunga proposed using Mahalanobis distance based on Transformer architecture[17], which can significantly improve the state-of-the-art accuracy of this dataset when the dataset is in cross-domain settings. When it comes to a specific domain (wheat disease detection in our case), this method can not ensure accuracy on the small data set. Moreover, both of these experiments have a premise: they need a vast data set as an intermediary, and their performances will decline when it comes to one specific crop, as we discussed.

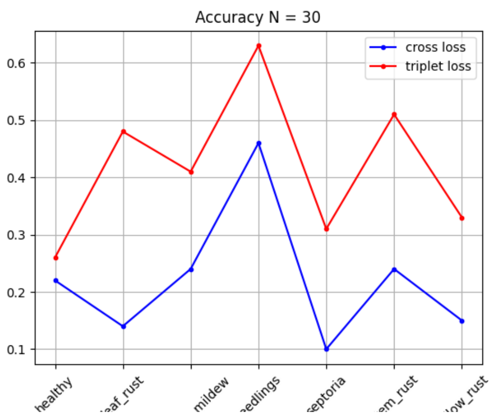
This paper proposes an improved method based on small samples without collecting large amounts of relevant data. This improved method introduces triplet loss on the EfficientNetB0 model to solve the problem of crop disease identification in small sample situations. Triplet loss is a classic metric learning method. By learning the relative distance between samples, the distance between samples of the same type is reduced as much as possible, and the distance between samples of different types is enlarged as much as possible, thereby improving the model's discriminative ability. In this method, triplet loss is constructed by selecting an appropriate sample triplet (anchor, positive, negative), and the feature representation ability of the model is optimized by minimizing triplet loss. This method can effectively utilize small sample data and improve the model's performance in crop disease identification tasks.

## 3 METHODOLOGY

In this section, we will explain the techniques used in the whole pipeline for wheat disease detection. The experiments were generally set to one way but with different components. Firstly, we trained the feature extractor with a small dataset(Baseline model with cross-entropy loss function vs Siamese network with triplet loss function). Then, we trained a k-nn as a shallow classifier with the same training dataset and used this trained k-nn classifier to quantify and visualize the classification results.




(a) Accuracy when using the whole dataset




(b) Accuracy when each class has 30 images

**Figure 2: Comparison of accuracy of full data vs. 30 images per class**

## 3.1 Baseline model and improved architecture

In our experiments, we used EfficinetNetB0 as the backbone, one of the state-of-the-art convolutional neural networks in crop disease detection. From the experiment results of Nigam et al. [18], EfficientNetB4 reached 99.35% accuracy on wheat disease detection when there is sufficient training data(over 10000 images in their case), which beyond all well-known and cutting-edge models (e.g., VGG19, RestNet152, InceptionV3, etc.), due to the computing resource constraints, we choose EfficientNetB0 (accuracy above average) that in the same framework as B4.

For our improved architecture(as shown in Fig 1), we conducted a Siamese Network and took EfficientNetB0 as the backbone of it; it is not the first time that Simases Network has been adopted in image classification. in He et al. work[19], they build a twofold Siamese network Named SA-Siam for real-time object tracking. As in Figure, our Siamese Network has three branches. We take three different images as input, each of which will be feature extracted by an independent backbone model. Then, the triplet loss function will receive these embedding vectors to adjust better strategy to do classification.

## 3.2 Defination of small data size

There is no specific data to distinguish between big and small data sets. Whether a large or small data set usually depends on the task and model. We can find similar experiments and literature for our task and refer to them according to their data volume. In the study of Afifi et al.[15], they compared the accuracy of different architectures in detecting leaf and coffee bean diseases on a small dataset. In their experiments, the data is divided into 32 categories, with a maximum of 50 images per category. This means 1600 images could be considered a small dataset in crop disease detection. Lin et al. did few-shot learning for leaf diseases[20], and their data was divided into ten categories of 200 images each, which means they take 2000 images as a small dataset. In contrast, our dataset has a total of 2414 images, and we only used 80% of them, or about 1900 images, to train the model, so our data can also be regarded as a small dataset.

## 3.3 Cross-Entropy Loss

Categorical cross-entropy is a fundamental loss function for tackling multi-class classification challenges in machine learning. It assesses the disparity between two probability distributions—the actual distribution of class labels and the distribution predicted by the model. The loss is computed using the following formula:

$$L_c = -\sum_i y_i \cdot \log(\hat{y}_i) \tag{1}$$

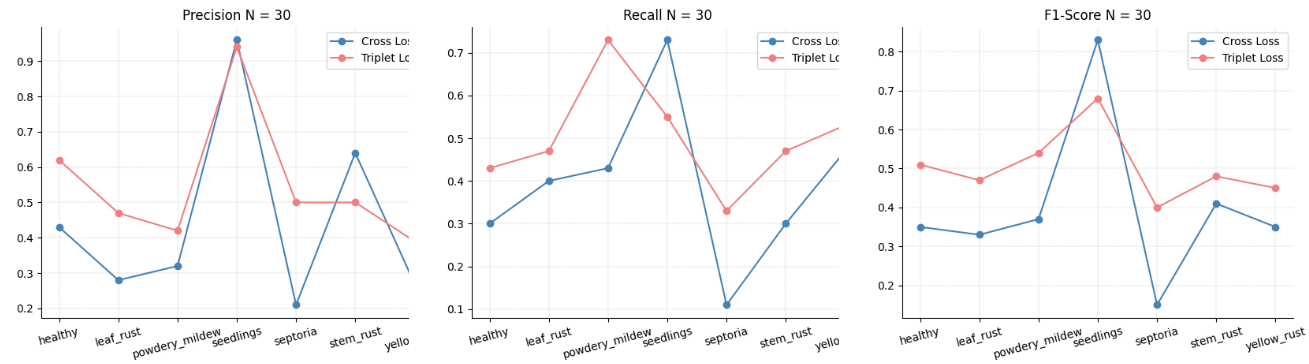
In the equation, $y_i$ represents the true probability of the i-th class, and $\hat{y}_i$ denotes the predicted probability of the i-th class as given by the model. The objective of the loss function is to minimize $L_c$ by adjusting the model parameters during training. This process encourages the predicted probability distribution$\hat{y}_i$ to align more closely with the true distribution$y_i$.

In the experiment, we trained EfficientNetB0 with cross-entropy loss as the baseline to represent a model with sufficient training data, which is exactly what Nigam et al. doing in their paper[18]

## 3.4 Triplet Loss

The triplet loss function enhances the distinguishability of embedding representations within an Euclidean vector space, particularly for classification tasks. It achieves this by minimizing the distance between embeddings of the same class while maximizing the distance between embeddings of different classes, thereby adjusting network parameters to optimize this objective. In practical implementation, a Siamese network architecture is often employed. This architecture comprises three subnetworks with shared weights interconnected by the triplet loss function. During training, three images are chosen: an anchor ($x_a$), a positive sample ($x_p$), and a negative sample ($x_n$). Each image is processed by one of the three subnetworks. Notably, the anchor and positive sample belong to the same class, whereas the negative sample represents a distinct class.

As a result, embedding vectors are produced to encapsulate the essential features corresponding to each image class. The triplet




**(a) Precision** **(b) Recall** **(c) F1-score**

**Figure 3: Comparison of Precision, Recall, and F1-score when N=30**

loss function then computes the distances between these three embedding vectors, typically utilizing the Euclidean distance function as expressed in Equation2.

Consequently, the network parameters undergo iterative adjustments to minimize the distance between anchor and positive sample embeddings while maximizing the distance between anchor and negative sample embeddings. This iterative optimization process enables the network to learn discriminative features essential for practical classification tasks.

$$\mathcal{L}_{\text{triplet}}(A, P, N) = \max\left(d(A, P) - d(A, N) + \alpha, 0\right) \quad (2)$$

In the equation, A, P, and N represent the anchor, positive, and negative samples. The *alpha* is a margin that specifies the minimum difference between the distance between the anchor-positive and the anchor-negative pair. It ensures that the loss is only computed if the negative example is sufficiently far from the anchor compared to the positive example, and the d represents the distance, the so-called Euclidean distance computed by equation3.

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (3)$$

### 3.5 Advantages of triplet loss on small datasets

In small datasets, triplet loss can be more data-efficient because it focuses on learning relative distances between examples. By leveraging triplets (anchor, positive, and negative), the model can learn more effectively from fewer examples by emphasizing relational information rather than absolute class probabilities. Cross-entropy loss tends to perform poorly on small datasets because it relies on a substantial amount of data to estimate class probabilities and learn robust decision boundaries accurately. Typically, small datasets are imbalanced, and the triplet loss function is less sensitive to class imbalance since it works with triplets. The selection of triplets can be designed to ensure a balanced representation of different classes, improving the model's robustness. In contrast, the cross-entropy loss function is more sensitive to class imbalance, which can be problematic in small datasets where some classes might be underrepresented.

### 3.6 K-NN and t-SNE

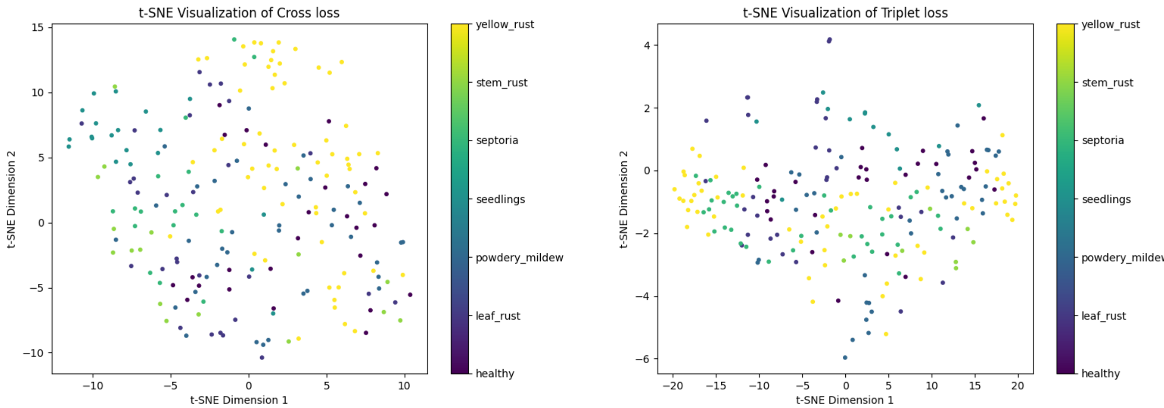
To derive different metrics from measuring the performance of the baseline model and improved architectures, the k-nn classifier was adapted as a shallow classifier after the model extracted image features in all experiments. The k-nn classifier takes feature embedding vectors as input and results in a length of seven vectors, which indicates the predicted disease type. The k-nn classifier was trained and fitted with the same training data as the feature extractor in different experiments. This ensures the model's performance and all metrics are based on limited data. In addition, the neighbor value (K) is the critical value of the k-nn classifier, so we did extra experiments on the validation set for different K values ranging from 1 to 80 and manually set the K value to 10, which led to a good performance on the baseline model and improved model.

To gain a deeper understanding of these two models' clustering capabilities, we used the t-SNE (t-Distributed Stochastic Neighbor Embedding) algorithm to visualize the feature embedding vectors extracted by models. t-SNE is a powerful technique commonly used for visualizing high-dimensional data in a lower-dimensional space, typically 2D or 3D. It works by first constructing a probability distribution over pairs of high-dimensional data points. This distribution reflects the similarities between data points: similar points are assigned higher probabilities of being picked together. Then, it constructs a similar probability distribution in the lower-dimensional space (usually 2D or 3D). The goal is to minimize the difference between these two distributions using gradient descent. In this way, it visualizes high-dimensional data (feature embedding vectors in our case) in lower-dimensional spaces, offering insights into how these two models perform on feature extraction and clustering.

## 4 RESULTS

### 4.1 Dataset and preprocessing

The dataset we were using is an open-source contributed by Genaev et al. [8], which has 2414 images and is categorized into seven diseases and one healthy type. Before using these images, there are some preprocessing steps. Every image is resized to 244x244 pixels and turned into an RGB image. In addition, some images have multiple labels; given that we only consider cases where images




**(a) Cross loss** **(b) Triplet loss**

**Figure 4: clustering ability when N=30**

contain one disease, we randomly choose one of the labels in our experiments.

## 4.2 Experiments and training settings

We generally have two sets of experiments based on different amounts of training data. First is the baseline; we adapted the EfficientNetB0 with pre-trained weights of ImageNet with cross-entropy loss function. Second is the Siamese Network with the triplet loss function. These two networks are trained as feature extractors. To quantify the classification performance, we used a k-nn classifier to receive the embedding vectors output by feature extractors to produce evaluation metrics and visualization results. In all experiments, data augmentation was adapted with RandomHorizontalFlip, RandomPerspective, and Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]). Adam optimizer with learning rate $\alpha = 10^{-4}$ was conducted over 50 epochs of all experiments.

To achieve the balance of the dataset, in addition to all the data, we conducted experiments on the Baseline model and Siamese network with triplet loss function for each class with 30 images, a step size of 10, and up to each class 80 images. The reason we start the experiment with 30 images per class is that if the training and test data are too small, the results will be very chance, which is not conducive to our conclusions, and in the whole dataset, some classes have only about 100 images, if the training data is more than 80, then the test data set will be too small, resulting in unreliable test results.

## 4.3 Classification results

To show more prominently the disparity in experimental results due to how much data is available, In the discussion that follows, we mainly show the comparison between the set of experiments with the least amount of data (30 images per class) and the set with the most amount of data(Whole data set). These two experiments were conducted to validate the results, with two models utilized as controls to assess performance in few-shot learning scenarios.

The first experiment (Top of Fig 2)is trained on the entire dataset, with 60% of the data allocated for training and 20% for validation and testing. Notably, most classes (representing various wheat diseases) within the training data contain nearly 300 images, while the smallest class comprises 111 images. The second experiment(Bottom of Fig 2) adopts the baseline model to create a Siamese Network with a Triplet loss function. The training data is configured with 30 images per class (N=30). Figure 2 illustrates the prediction accuracy of each class for both experiments. The experimental results reveal that the triplet loss performs better than the categorical cross-entropy loss across all classes in both experiments. Moreover, the triplet loss function's advantage becomes more pronounced when dealing with limited data.

While accuracy provides valuable insights, a comprehensive evaluation of model performance necessitates the adoption of additional metrics. Figure 3 illustrates the precision, recall, and f1-score of the categorical cross-entropy loss and triplet loss when N=30. Notably, even within a specific class (e.g., "seeding"), the cross-loss function might outperform the triplet loss function, possibly due to the ease of capturing features associated with this disease, leading to a higher accuracy rate. However, on the whole, the triplet loss function consistently outperforms the cross-loss function for smaller sample sizes. These metrics substantiate our hypothesis, demonstrating that the triplet loss function yields greater accuracy than the Cross-Entropy loss function when dealing with limited data, specifically in the context of few-shot learning scenarios.

## 4.4 Clustering Results by t-SNE

The experiments employed EfficientNetB0 and a Siamese Network as feature extractors, coupled with a K-Nearest Neighbor (KNN) classifier. The KNN utilizes the embedding vector features extracted by the models as input and predicts results for the seven classes representing seven distinct diseases. To assess the clustering ability of the two models, we also applied T-distributed Stochastic Neighbor Embedding (t-SNE) to the embedding vectors, which visualizes high-dimensional data in a 2D plot.

Figure 4 displays the t-SNE plots of the two models during training, with each color representing a different disease type. The left graph depicts the baseline model with a cross-entropy loss function. However, some similar data points are grouped, and a significant portion of data is scattered throughout the plot, resulting in a generally loose distribution. Conversely, the right graph represents the Siamese network with the triplet loss function, exhibiting superior performance. Although data points of the same type are not entirely clustered, most are closely positioned, resulting in a more compact plot. Additionally, we computed the Davies–Bouldin index (DB index) for both models. This metric assesses clustering by comparing the compactness within clusters to the separation between clusters, thereby indicating the clustering ability of the models. From the DB index values, we conclude that the model trained with the triplet loss function demonstrates better clustering ability compared to the cross-entropy loss function ($DBindex_{\text{triplet}} = 4.34$ vs $DBindex_{\text{cross}} = 9.79$).

### 4.5 Exploring of the trend

In addition, in Figure 5, we show the accuracy of these two loss functions in different ranges of training data. The graph shows that the triplet loss function has been consistently more accurate than the cross-entropy loss function in small sample situations. Still, the gap narrows with increasing data volume, from around 10% gap when there are 30 images per class to around 6% when there are 80 images per class, proving a trend in the accuracy gap when changing the data volume. However, Because of our dataset's size limitations, we could not find a turning point for this scenario. More training data is needed to determine the limitation of the triplet loss function.

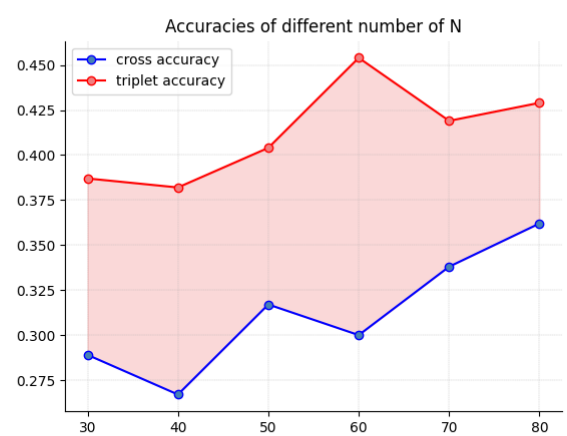



**Figure 5: Accuracy comparison of different number training data**

## 5 DISCUSSION

Deep learning technology for crop disease detection has gained increasing attention in recent years because of the pathologist-oriented, subjective, costly, time-consuming, and labor-intensive traditional detection methods[21]. Much work has been done in crop disease detection with a large amount of data, and promising accuracy has been achieved. Chen et al.[22] suggested an improved VGG model on a merged dataset of maize and rice datasets, with an accuracy of 92%. Gokulnath et al. proposed a resilient LF-CNN with an accuracy of 98.93% on the PlantVlillage dataset. An extensive dataset guarantees both the stability and accuracy of the model. While it is difficult to get a vast amount of training data for crop disease so far, like those articles we discussed in this paper, except for those experiments on Plant, most of the rest are non-public data obtained by some authors from different sources, and when it comes to specific crop varieties the data of PlantVilliage are also insufficient. Therefore, building a model that can learn from few amounts of data is critical. Argüeso et al.[23] proposed an architecture that can learn from a tiny amount of training data, 15 images per class. However, their model was pre-trained with the source dataset, meaning it still needs a supporting dataset.

This paper proposes using the Siamese network with a triplet loss function to learn from the limited dataset instead of the traditional CNN model. We have done a series of experiments around these two networks, from 30 images per class to 80 images per class (the maximum value of our dataset), to compare the performance of the proposed network and the traditional network (EfficientNetB0). The result shows that within our dataset's scope, the Siamese network's accuracy (around 40%) is about 10% more accurate than the baseline model(around 30%). To get more insights, we also looked into the different metrics of each type of disease and found out that, except for "seeding," Simases network performance is better for all the rest of the diseases. We think "seeding" is unique because the features of this disease are easy to recognize, so it does not need vast amounts of data to learn. In addition, both from the comparison of 30 images per class with the complete data and from the successive comparisons of the accuracy of different numbers of data, we can see that the gap between them is narrowing, which means the Siamese network with the triplet loss function only superior to baseline with cross-entropy loss function only with limited data.

The result and evaluation prove our proposal can achieve higher accuracy than traditional CNN models. However, it is worth noting that wheat diseases are geographically specific, and the same disease may manifest itself differently in different regions. Our data does not include all areas and all diseases, so we cannot guarantee whether they will perform the same in various datasets. In addition, Although our accuracy is higher than that of traditional CNNs, 40% accuracy is not enough to meet the requirements of production conditions for the time being. Moreover, due to the limited dataset, we haven't found the turning point in the amount of data where the traditional network overtakes the Siamese network. So, improving the diversity of datasets, increasing the volume of data, and finding the turning point will be the focus of future work.

## 6 CONCLUSION

Few-shot learning is a challenging subfield of machine learning and deep learning, making developing machine learning models in real-world settings feasible. For this reason, This thesis aims to propose a developed model: a siamese network with the triplet loss function and compare it with a traditional CNN model, EfficientNetB0, with the cross-entropy loss function on a limited wheat diseases dataset.

Our experiments are built on a tiny but real farmland dataset without any outside supporting dataset. A series of experiments have been done on two networks to observe the performance gap and how it changes with the amount of data and data volume from 30 images per class to 80 per class(nearly the maximum of the dataset), with ten images added at a time. From the results of the experiments, the triplet loss network demonstrates its superiority over cross-entropy loss on various metrics. Notably, in our in-depth analysis of the metrics for each disease, there was one exception: most metrics on "seeding" present that cross-entropy loss network perform better; we believe that it is because the features of this disease are too simple to need large amounts of data. Also, we analyze the generated latent space descriptors of two networks; the triplet network improves the DB-index parameter by 125% when N=30. Furthermore, a trend can be drawn from the results: As the amount of data increases, the accuracy gap between the two networks decreases, Which means there is a tipping point in the amount of training data beyond which the triplet loss network will no longer outperform the cross-entropy loss network.

Based on the results and findings of our experiments, this paper proposes an improved network with the triplet loss function that can perform better than the traditional model with the cross-entropy loss function with limited training data and no supporting data, which is closer to real scenarios and real needs in actual farmland.

## REFERENCES


[1] Wang Linbai, Zhang Bo, Yao Jingfa, Yang Zhihui, Zhang Jun, and Fan Xiaofei. Potato leaf disease recognition and potato leaf disease spot detection based on convolutional neural network. *Journal of Chinese Agricultural Mechanization*, 42(11):122, 2021.

[2] Bao Wenxia, Huang Xuefeng, Hu Gensheng, and Liang Dong. Identification of maize leaf diseases using improved convolutional neural network. *Transactions of the Chinese Society of Agricultural Engineering*, 37(6), 2021.

[3] Jayme G. A. Barbedo. Deep learning applied to plant pathology: the problem of data representativeness. *Tropical Plant Pathology*, 47(1):85–94.

[4] David. P. Hughes and Marcel Salathe. An open access repository of images on plant health to enable the development of mobile disease diagnostics. 2016.

[5] Image-based detection of plant diseases: From classical machine learning to deep learning journey. 2021:1–13.

[6] Vimal Singh, Anuradha Chug, and Amit Prakash Singh. Classification of beans leaf diseases using fine tuned CNN model. *Procedia Computer Science*, 218:348–356.

[7] Lakshay Goyal, Chandra Mani Sharma, Anupam Singh, and Pradeep Kumar Singh. Leaf and spike wheat disease detection & classification using an improved deep convolutional architecture. *Informatics in Medicine Unlocked*, 25:100642.

[8] Mikhail A. Genaev, Ekaterina S. Skolotneva, Elena I. Gultyaeva, Elena A. Orlova, Nina P. Bechtold, and Dmitry A. Afonnikov. Image-based wheat fungi diseases identification by deep learning. *Plants*, 10(8):1500.

[9] Chunshan Wang, Ji Zhou, Chunjiang Zhao, Jiuxi Li, Guifa Teng, and Huarui Wu. Few-shot vegetable disease recognition model based on image text collaborative representation learning. *Computers and Electronics in Agriculture*, 184:106098.

[10] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in Neural Information Processing Systems*, 31.

[11] M Tan and EfficientNet Le Q V. rethinking model scaling for convolutional neural networks. 2019. *arXiv preprint arXiv:1905.11946*, 1905.

[12] J. G. A. Barbedo and C. V. Godoy. Automatic classification of soybean diseases based on digital images of leaf symptoms.

[13] Zhencun Jiang, Zhengxin Dong, Wenping Jiang, and Yuze Yang. Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning. *Computers and Electronics in Agriculture*, 186:106184, 2021.

[14] Ronghua Gao, Rong Wang, Lu Feng, Qifeng Li, and Huarui Wu. Dual-branch, efficient, channel attention-based crop disease identification. *Computers and Electronics in Agriculture*, 190:106410, 2021.

[15] Ahmed Afifi, Abdulaziz Alhumam, and Amira Abdelwahab. Convolutional neural network for automatic identification of plant diseases with limited data. *Plants*, 10(1):28.

[16] José G.M. Esgario, Renato A. Krohling, and José A. Ventura. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Computers and Electronics in Agriculture*, 169:105162.

[17] Sai Vidyaranya Nuthalapati and Anirudh Tunga. Multi-domain few-shot learning and dataset for agricultural applications. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1399–1408.

[18] Sapna Nigam, Rajni Jain, Sudeep Marwaha, Alka Arora, Md. Ashraful Haque, Akshay Dheeraj, and Vaibhav Kumar Singh. Deep transfer learning model for disease identification in wheat crop. *Ecological Informatics*, 75:102068.

[19] Anfeng He, Chong Luo, Xinmei Tian, and Wenjun Zeng. A twofold siamese network for real-time object tracking. pages 4834–4843.

[20] Hong Lin, Rita Tse, Su-Kit Tang, Zhen-ping Qiang, and Giovanni Pau. Few-shot learning approach with multi-scale feature fusion and attention for plant disease recognition. *Frontiers in Plant Science*, 13:907916. Publisher: Frontiers.

[21] Sharada P. Mohanty, David P. Hughes, and Marcel Salathé. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7. Publisher: Frontiers Media SA.

[22] Chen J, Zhang D, Nanehkaran Ya, and Li D. Detection of rice plant diseases based on deep transfer learning. *Journal of the science of food and agriculture*, 100(7). Publisher: J Sci Food Agric.

[23] David Argüeso, Artzai Picon, Unai Irusta, Alfonso Medela, Miguel G San-Emeterio, Arantza Bereciartua, and Aitor Alvarez-Gila. Few-shot learning approach for plant disease classification using images taken in the field. *Computers and Electronics in Agriculture*, 175:105542.