MATHEMATICAL SCIENCES

A STUDY ON POTATO DISEASE LEAF CLASSIFICATION BASED ON RESNET

Zhang Hongzhi,
National University of Uzbekistan named after Mirzo Ulugbek ,Faculty of Applied Mathematics and intellectual Technologies.
JiNing Normal University, School of Mathematics and statistic, Ulanqab, Inner Mongolia, P. R. China
Jining Normal University Scientific Research Project: Convolutional Neural Network-based Potato Leaf Disease Identification Method (jsky2024001)
Wang Hailiang,
Beijing Huaxia Chun Song Technology Co., Ltd.
Varlamova L.
National University of Uzbekistan named after Mirzo Ulugbek ,Faculty of Applied Mathematics and intellectual Technologies.
https://doi.org/10.5281/zenodo.16539247

Abstract

In recent years, convolutional neural networks (CNNs) have achieved significant research results in image classification and recognition.To achieve the identification of potato leaf diseases and enable timely control, this paper takes early blight and late blight of potatoes as the research objects and constructs a potato disease identification model based on a deep convolutional neural network—ResNet to obtain better identification performance.

Keywords: potato; diseased leaves; convolutional neural network; ResNet; dataset

In recent years, with the increasing area of potato planting, the incidence of early blight and late blight area and the degree of harm is also increasing, due to the early onset of early blight and late blight on the leaves, the two diseases are similar in symptoms, it is difficult to distinguish between the error of judgement is prone to delay the time of treatment. The traditional identification method relies on expert judgement, which requires a lot of professional knowledge, is affected by subjective factors, and is time-consuming, and cannot diagnose potato diseases in a timely and efficient manner. Therefore, the identification and accurate positioning of disease spots is an important means of potato disease control [1-2], so it is of great practical significance to study an efficient and accurate intelligent identification method for potato diseases [3]. To achieve the identification of potato leaf diseases and enable timely prevention and control, this paper takes early blight and late blight of potatoes as the research objects and constructs a potato disease identification model based on a deep convolutional neural network—ResNet. Through potato leaf identification, the model determines whether the leaf is healthy, infected with early blight, or infected with late blight.

healthy early blight late blight

Figure 1: Comparison of healthy and diseased potato leaves

1 Data set organisation and publication

In the field of artificial intelligence, data plays a crucial role. One of the key focuses and contributions of this study is the release of a dataset of potato disease leaves [4], hereinafter referred to as the 'Potato Disease Leaf Dataset v1'.

1.1 Dataset Overview

This dataset consists of two parts: a development dataset and a test dataset. Each part contains the same categories: early blight leaves, late blight leaves, and healthy leaves. The number of images in each category is shown in Table 1:

Table 1

Potato Leaf Disease Dataset v1 Classification and Quantity

Category/Dataset	Development dataset	Test dataset
Early blight leaves	3615	363
Late blight leaves	2424	67
Healthy leaves	1172	363
Total	7211	793

1.2 Dataset design

During the design of this dataset, the rationality of the distribution of the development and test datasets was fully considered. Dai Guowei et al. [5] used Yolo V5 [6] and the PlantVillage dataset [7] to train the model, and then tested it on another dataset. The accuracy rate of the test results was 49.79%, which was not ideal.

In many current studies on potato disease leaf recognition, the PlantVillage dataset is frequently used, with both training and testing data sourced from this dataset. However, this dataset was collected under single geographical and environmental conditions, while potato cultivation regions and varieties are diverse worldwide. This results in studies using the PlantVillage dataset [7] achieving high accuracy rates, yet performing poorly on other potato datasets.Therefore, in the process of creating this dataset, the following methods were considered and adopted to improve the rationality of the dataset and make it more meaningful for practical problems:

(1) The ratio of the test dataset to the development dataset should be around 1:10. This ratio accommodates the amount of data required to identify disease characteristics while also ensuring the diversity and adequacy of the test dataset.

(2) The test dataset and the development dataset should have as many obvious differences as possible. These differences are crucial for detecting the effectiveness of artificial intelligence algorithms, such as the following distinctions:

First, ensure that the development dataset and test dataset come from different growing regions.

Second, ensure that the images in the development dataset and test dataset have different lighting conditions.

Third, ensure that the images in the development dataset and test dataset have different backgrounds.

Fourth, ensure that the development dataset and test dataset come from different potato varieties. Validating the model based on these differences allows for a better focus on the model's generalisation ability.To support the above requirements, the development and test datasets in this dataset use different data sources as much as possible, and the self-built dataset was captured at different time periods and from different angles.

1.3 Data sources for the dataset

This dataset is compiled from multiple sub-datasets. These sub-datasets include both those created by the authors of this study, which were photographed and collected in actual potato growing areas, and those from open datasets on the Internet. These datasets are non-redundant and come from different growing areas.

(1) Self-built dataset

The authors of this study collected a total of 1,350 images of early blight leaves in Nan Village, Jining District, Ulanqab, Inner Mongolia Autonomous Region, People's Republic of China.

(2) PlantVillage-Dataset

The PlantVillage Dataset[7] contains images of leaves affected by various plant diseases, including 2,152 images of potato leaves.

(3) Potato Disease Leaf Dataset(PLD)

The Potato Disease Leaf Dataset (PLD) [8] is a potato leaf dataset that includes healthy leaves, early blight, and late blight, with a total of 4,072 images.

(4) Potato Leaf (Healthy and Late Blight)

Potato Leaf (Healthy and Late Blight) [9] is a potato leaf dataset that includes healthy leaves and late blight, with a total of 430 images.

Maintaining a diverse dataset of potato growing regions and potato varieties is critical for the application of artificial intelligence in agriculture. Therefore, the dataset released in this paper is a long-term, open dataset.

2 ResNet-based potato disease leaf classification model

2.1 Introduction to ResNet

ResNet (Residual Network) is a deep convolutional neural network architecture proposed by Kaiming He et al. [10] from the Microsoft team in 2015.At the time, it demonstrated exceptional performance in image classification, segmentation, and object detection tasks, marking a significant milestone in the field of deep learning.Its core idea is to introduce residual modules (ResidualModule) and residual connections (Residual Connection) to construct deeper network structures, thereby addressing issues such as gradient vanishing and network degradation that commonly arise during the training of DCNNs. ResNet-50 is the most commonly used structure, where the input image first passes through initial convolution layers and pooling layers to extract basic image features. The network then proceeds to a four-layer structure, with each layer containing 3, 4, 6, and 3 residual modules, respectively.

Each residual module consists of several residual units. Each residual unit consists of a main branch and a residual branch. The main branch comprises multiple convolutional layers, while the residual branch is either an identity mapping or includes additional convolutional layers. The outputs of the two branches are added together, passed through a ReLU activation function, and finally output through a global average pooling layer and a fully connected layer to produce the final classification result.

2.2 The innovation of ResNet

Before the advent of ResNet, the best-performing neural networks had relatively few layers. For example, in the field of computer vision, the number of layers in the best neural networks in some competitions is shown in Table 2:

Table 2

Brief introduction of best image classification models

Image Classification	Birth Year	Neural Network Layers	Description
LeNet	1998	5	MNIST state-of-the-art
AlexNet	2012	8	ILSVRC 2012 First place
VGG	2014	19	ILSVRC 2014 Second place
GoogLeNet	2014	22	ILSVRC 2014 First place

However, another hypothesis is that larger models can solve more complex problems because they can approximate more complex feature spaces through more parameters. Increasing the depth of a neural network results in a larger model, and the challenge lies in how to train large-scale neural networks because the parameter evaluation algorithm is based on gradient updates. The deeper the neural network, the more unstable the gradient becomes, leading to gradient explosion or gradient vanishing.The innovation of ResNet is precisely that it solves this problem.

The idea behind ResNet is to stack input information again after weighted operations and activation functions are performed on neurons, thereby making the changes between the outputs and inputs of neurons more gradual, as shown in Figure 2:

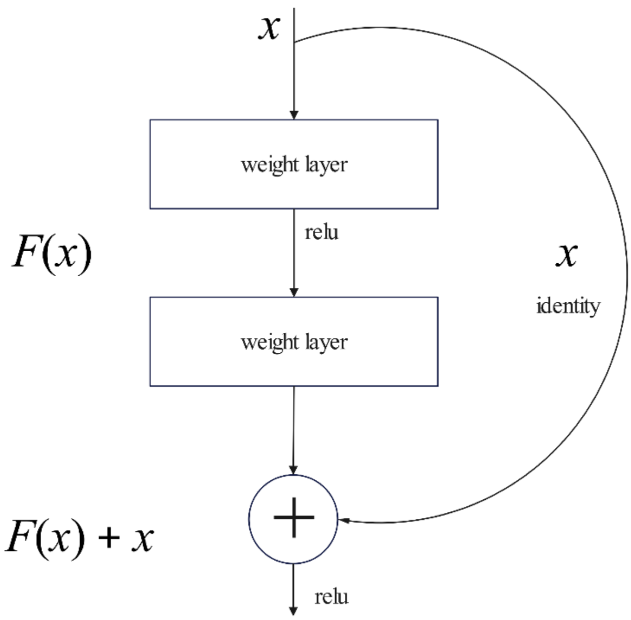

Figure 2: Residual learning: a building block

Add the input of this layer to the output of this layer. This method is called skip connection, and the block composed of this is called a residual block.

Assuming that the target function to be learned is $H(x)$, and that $F(x)$ learns $H(x)$ before using the residual, then after adding the residual, we obtain $H(x)=F(x)+x$, and thus $F(x)=H(x)-x$, i.e., the residual. This structure is called the residual structure.

The use of residual structures ensures that the stochastic gradient descent (SGD) algorithm can obtain stable gradients for parameter learning. Additionally, the authors of ResNet employed bottleneck structures to reduce the computational requirements for model training. The application of these techniques has made it feasible to train deep neural networks.

2.3 ResNet pre-trained model

Given the outstanding performance of the ResNet model in image classification tasks, the PyTorch[11] open-source machine learning framework provides multiple pre-trained models of ResNet v1 on ImageNet[12], namely resnet18, resnet34, resnet50, resnet101, and resnet152, which correspond to the architectural designs in the ResNet paper[13], as shown in Figure 3:

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		[3×3, 64 3×3, 64] ×2	[3×3, 64 3×3, 64] ×3	[1×1, 64 3×3, 64 1×1, 256] ×3	[1×1, 64 3×3, 64 1×1, 256] ×3	[1×1, 64 3×3, 64 1×1, 256] ×3
conv3_x	28×28	[3×3, 128 3×3, 128] ×2	[3×3, 128 3×3, 128] ×4	[1×1, 128 3×3, 128 1×1, 512] ×4	[1×1, 128 3×3, 128 1×1, 512] ×4	[1×1, 128 3×3, 128 1×1, 512] ×8
conv4_x	14×14	[3×3, 256 3×3, 256] ×2	[3×3, 256 3×3, 256] ×6	[1×1, 256 3×3, 256 1×1, 1024] ×6	[1×1, 256 3×3, 256 1×1, 1024] ×23	[1×1, 256 3×3, 256 1×1, 1024] ×36
conv5_x	7×7	[3×3, 512 3×3, 512] ×2	[3×3, 512 3×3, 512] ×3	[1×1, 512 3×3, 512 1×1, 2048] ×3	[1×1, 512 3×3, 512 1×1, 2048] ×3	[1×1, 512 3×3, 512 1×1, 2048] ×3
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figure 3: ResNet architecture for training ImageNet models

ResNet 50 is a 50-layer deep neural network based on the ResNet architecture. As the number of network layers increases, the number of parameters in the network also increases, leading to more stringent computational requirements and a corresponding increase in the computational power of the required hardware. By leveraging the pre-trained ResNet model, fine-tuning the potato disease leaf recognition model through transfer learning becomes simpler: this not only saves time but also requires fewer computational resources than before. The architecture of ResNet50 [14] is shown in Figure 4:

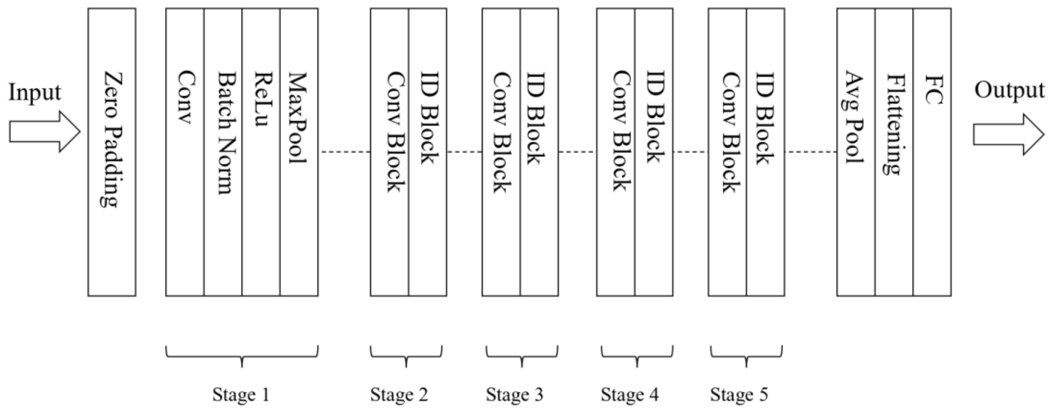

Figure 4 : ResNet50 Model Architecture

As shown in the figure above, the network outputs data from the left side, undergoes computations in the middle, and outputs data from the right side. The input layer of ResNet 50 uses zero padding. Then comes Stage 1, which performs convolution, batch normalisation, ReLU activation function, and max pooling. From stage two to stage five, four sets of convolution-ID modules are used. The convolution-ID module is also referred to as the BottleNeck module, as its shape resembles the neck of a bottled mineral water bottle. The design of the BottleNeck module in ResNet 50 is shown in Figure 5:

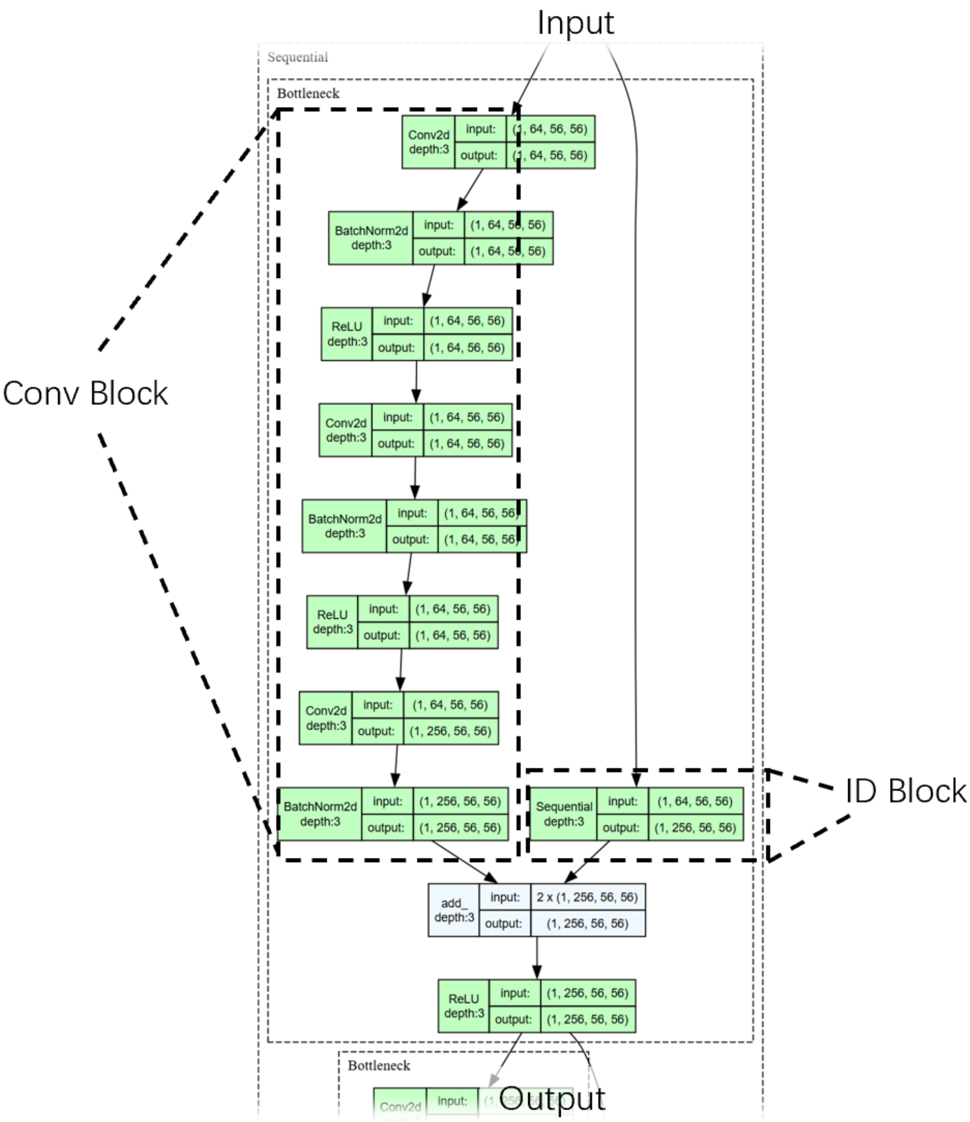

Figure 5: Bottleneck Module

2.4 Neural network architecture

This paper collects and labels three types of potato leaves: early blight, late blight, and healthy leaves. First, a three-classification problem is studied, using artificial intelligence algorithms for identification. The architecture used is outlined in Figure 6:

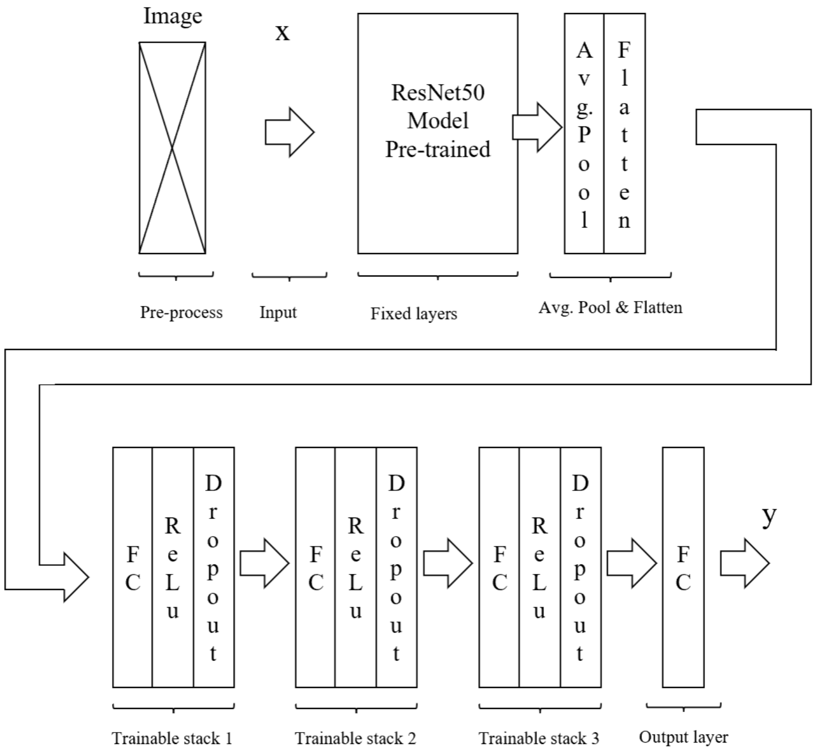

Figure 6: Transfer learning using a pre-trained ResNet model

The network architecture, from left to right, consists of image preprocessing, a pre-trained model, and an output layer. The parameters of the pre-trained model are fixed and do not update. Parameter updates for the trained network occur in Trainable Stack 1-3 and the output layer. The parameters in the network are described in Table 3:

Table 3

Parameters in model

Parameters	Number
Trainable params	2,647,747
Non-trainable params	23,508,032
Total	26,155,779

The Avg. Pool & Flatten layer in Figure 6 performs pooling and flattening. There are no parameters; it is simply a mapping operation. After flattening, each image becomes a 2048-dimensional vector. The detailed design of the neural network for this module is shown in Figure 7:

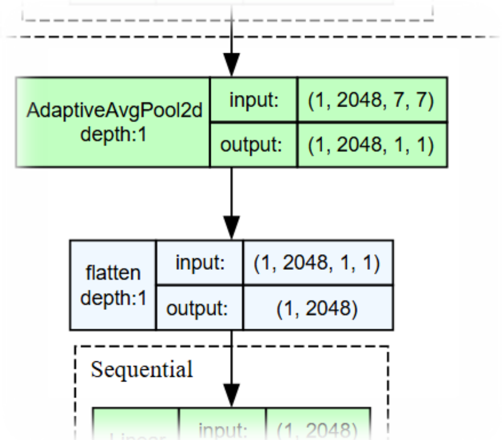

Figure 7: transform in Avg. Pool & Flatten Layer

The network design utilises transfer learning technology, with its core being the training of the ResNet encoder using ImageNet to better extract disease features from potato leaf images. Subsequently, the model is trained using a feedforward network, ReLU activation function, and random dropout technology. Through multiple fully connected layers, the classification dimension is gradually reduced, ultimately forming a 3-dimensional output layer. The classification process is shown in Figure 8:

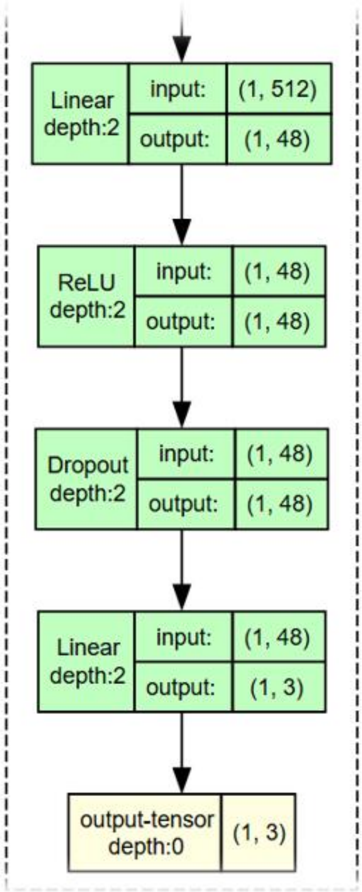

Figure 8: Trainable Stack 4 to output layer

2.5 Dataset splitting

In the Potato Leaf Disease Dataset v1, potato leaf images are divided into a development dataset and a test dataset. To validate the algorithm's effectiveness and ensure it can accurately identify the pathological features of the disease, the test dataset specifically includes images that differ significantly from those in the development dataset. These images are categorised based on potato variety, growing region, and different imaging devices. During the training process of this programme, the model is first trained using the development dataset, and then the test dataset is used for inference to evaluate the model's performance.

During the training process, it is necessary to divide the development dataset into a development training set and a development validation set. This process is called dataset splitting, and its purpose is to prevent model overfitting and adjust network hyperparameters. During dataset splitting, we randomly select 1/10 of the data from each category in the development dataset as the development validation set, and the remaining data as the development training set.

2.6 Image preprocessing

Image preprocessing refers to performing certain operations on images before training models or predicting image categories. In this study, the following operations were performed.

(1) Resize

The original images vary in size, but each image must be converted into a vector with the same dimensions to be used as input for the neural network. In the case of transfer learning, the front end of the network is connected to a pre-selected ResNet network, so the input image size must first be adjusted to the format accepted by the pre-trained ResNet network. In this case, the image size needs to be resized to a width of 232 pixels and a length of 232 pixels.

(2) Normalize

Before inputting the original image into the network, it needs to be normalized. This is because the PyTorch ResNet pre-trained model was trained using images from ImageNet, and the RGB values of the pixels in the images were statistically analysed to obtain the mean and standard deviation. The RGB image was then converted into a decimal value in the range (0, 1) before being input into the network's input layer. Therefore, during our training process, we first resize the images and then normalize them using the same mean and standard deviation as ImageNet for preprocessing.

Whether to use the mean and standard deviation from ImageNet for preprocessing images depends on the following principles:

Use — the images to be classified are ordinary photos of natural scenes, such as people, buildings, animals, different lighting conditions, angles, backgrounds, etc.;

Do not use — the photos are special images such as medical images, satellite maps, or hand-drawn images;

The former is similar to the ImageNet dataset, while the latter requires recalculating the mean and standard deviation based on the dataset.

2.7 Training environment

The procedures used in this study employed the following environment.

(1) hardware resources

Computer configuration: Intel Core i5 processor, 16 cores; 32 GB memory; NVIDIA GeForce RTX 4060 Ti GPU graphics card.

(2) Software Framework and Version

The main software and versions used in this program: Python v3.11; PyTorch 2.4; CUDA 12.4.

(3) hyperparameters

The code for this program has been published on the GitHub hosting platform. During execution, the following hyperparameters were used: initial learning rate 0.0001; number of training epochs 1000; patience value for early termination 5.

(4) execution time

Based on the above environment, the program ran for 7.2 hours.

2.8 Training process data

(1) Loss function performance

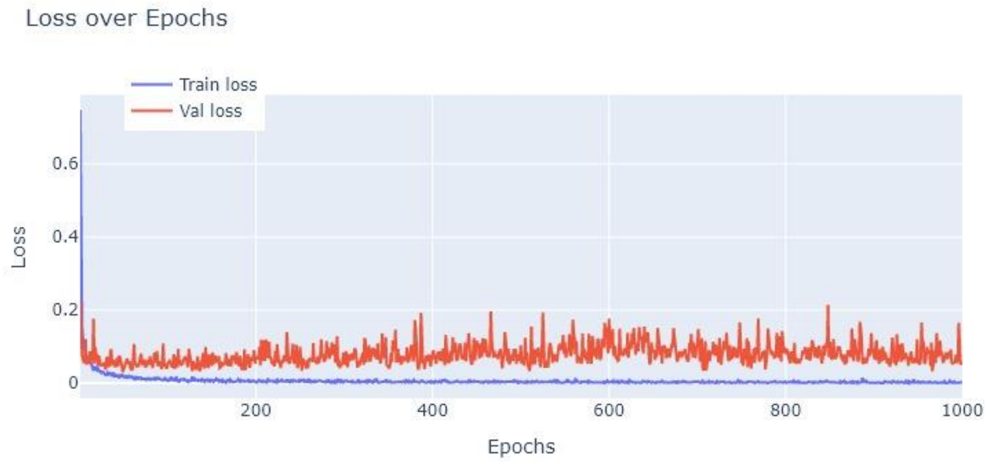

Figure 9: Changes in the loss function of the ResNet50 transfer learning model

(2) Accuracy performance

From the above training data, it can be seen that after entering the 20th round, the accuracy rate on the development training set has reached 99%, and has remained at around 99% with minor fluctuations thereafter. At the same time, the accuracy rate on the development validation set has fluctuated slightly around 98%. The loss function also converges rapidly on both the development training set and the development validation set, with minor fluctuations after the 20th training round. The performance of the transfer learning algorithm on the development dataset indicates that the data classification task can achieve an excellent accuracy of over 96%.

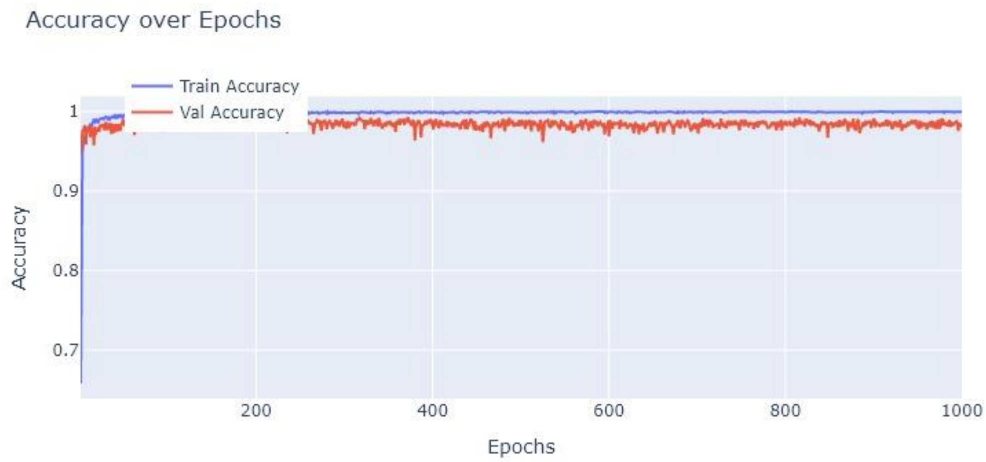

Figure 10: Changes in the accuracy of the ResNet50 transfer learning model

2.9 Prediction test set

The prediction test set is used to evaluate the trained model using actual potato leaves from production and daily life. The test data set in the Potato Disease Leaf Dataset v1 was designed for this purpose. We explore artificial intelligence models in the hope of predicting infinite, real data spaces with limited data.

The purpose of researching algorithms is to enable models to learn patterns rather than merely memorise limited training data. In the experiments conducted in this programme, the evaluation metrics achieved an accuracy rate of 66.32% (accurately predicting 516 out of 778 test images), a recall score of 0.739, and an F1 score of 0.594.

3 Conclusion

From the above research, it can be demonstrated that a neural network composed of transfer learning, fully connected neural networks, Dropout technology, and ReLu activation functions can perform well on real potato disease datasets. This paper constructed an open dataset of potato disease leaves and achieved an accuracy rate of 66.32% using the currently leading algorithm. This serves as baseline data, providing a data and comparison foundation for evaluating more advanced algorithms in the future.

References:

1. Zhao Lixin, Hou Fadong, Lt Zhengchao, et al. Image recognition of cotton leaf diseases and pests based on transfer learning[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020 , 36(7):184 -191.
2. Tian Haitao,Zhao Jun,Pu Fupeng(A method for recognizing potato's bud eye) [J]. Acta Agriculturee Zhejiangensis,2016,28(11) 1947-1953.
3. AHMAD LOTI N N,MOHD NOOR M R,CHANG S W,Integrated analysis of machine learning and deep learning in chili pest and disease identification[J].J Sci Food Agricul,2021,101(9):3582-3594.
4. Hailiang W, Hongzhi Z. Potato diseases dataset[EB/OL]. 2025[2025-03-24]. https://github.com/jnai-team/potato-diseases-dataset.
5. Dai Guowei, Fan Jingchao. Method for identifying early-late blight leaf diseases of potatoes:China,CN114677561A[P/OL].2022. https://patents.google.com/patent/CN114677561A/.
6. Khanam R, Hussain M. What is yolov5: A deep look into the internal features of the popular object detector[M/OL]. arXiv, 2024. DOI: 10.48550/ARXIV.2407.20892.
7. Mohanty S P, Hughes D P, Salathé M. Using deep learning for image-based plant disease detection[J/OL]. Frontiers in Plant Science, 2016, 7. DOI: 10.3389/fpls.2016.01419.
8. Rashid J, Khan I, Ali G, et al. Multi-level deep learning model for potato leaf disease recognition [J/OL]. Electronics, 2021, 10(17): 2064. DOI: 10.3390/electronics10172064.
9. Sankalana N. Potato leaf (healthy and late blight)[EB/OL]. 2024[2024-03-24].https://www.kaggle.com/datasets/nirmalsankalana/potato-leaf-healthy-and-late-blight/.
10. He K, Zhang X,Ren S,et al. Deep residual learning for image recognition[C]//Proceedings of CVPR 2016.Las Vegas:IEEE,2016: 770-778
11. Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library[M/OL].arXiv,2019.DOI: 10.48550/ARXIV.1912.01703.
12. Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C/OL]//2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009. DOI: 10.1 109/cvpr.2009.5206848.
13. He K, Zhang X, Ren S, et al.Deep residual learning for image recognition[M/OL]. arXiv, 2015. DOI: 10.48550/ARXIV.1512.03385.
14. Bendjillali R I, Beladgham M, Merit K, et al. Illumination-robust face recognition based on deep convolutional neural networks architectures[J/OL]. Indonesian Journal of Electrical Engi- neering and Computer Science, 2020, 18(2): 1015. DOI: 10.11591/ijeecs.v18.i2.pp1015-1027.