Indonesian Journal of Electrical Engineering and Computer Science
Vol. 34, No. 3, June 2024, pp. 1582~1592
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v34.i3.pp1582-1592

Plant pathology identification using local-global feature level based on transformer

Manh-Hung Ha, Duc-Chinh Nguyen, Manh-Tuan Do, Dinh-Thai Kim, Xuan-Hai Le, Ngoc-Thanh Pham
Faculty of Applied Sciences, International School, Vietnam National University, Hanoi, Vietnam

Article Info

Article history:

Received Jan 7, 2024
Revised Feb 12, 2024
Accepted Feb 29, 2024

Keywords:

Attention
Convolutional neural network
Deep learning
Image classification
Local-global feature

ABSTRACT

Deep learning plays a crucial role in addressing the challenge of plant disease identification in the field of agriculture. Detecting diseases in plants requires extensive effort, along with a comprehensive understanding of various plant diseases and increased processing time. Balancing both speed and accuracy in predicting leaf diseases in plants can significantly improve crop production and reduce environmental damage. In this paper, we examined deseases on popular plants in agriculture. We proposed a novel model to predict crop pathology on a feature space of global-local based on transformer aggregation. Paticular, we use refined feature of different layer to correlate semantics from high-level feature and low-level feature. Besides, to capture the extended temporal scale across the entire image, we employ a transformer to discern long-range dependencies among frames. Subsequently, the enhanced features incorporating these dependencies are inputted into a classifier for preliminary crop pathology prediction. The plant village dataset and VietNam strawberry disease (VNStr) dataset were utilized for training and disease classification in the experiments. Extensive experiments show that the proposed method outperforms by 99.18% and 94.05% accuracy in plant village and VNStr, respectivly. The model after being judged was applied on Android devices and therefore is easy to use.

Corresponding Author:

Manh-Hung Ha
Faculty of Applied Sciences, International School, Vietnam National University
Hanoi, 100000, Vietnam
Email: hunghm@vnu.edu.vn

1. INTRODUCTION

In agriculture, pests and diseases have long posed a serious threat to crop growth and the storage of agricultural products. If only manual identification is based on the naked eye, it requires a long observation time, leading to no timely prevention and handling measures. That will greatly affect productivity and output. In recent years, with the continuous development of science and technology, especially computer software technologies, the application of high technology to agricultural production has developed rapidly. Through the use of cameras in agriculture, images of pests and diseases on agricultural crops can be identified more conveniently [1]. Traditionally, machine learning algorithms have been used which is time-consuming because it requires manual extraction of features from images and is fed as input to the algorithm for classification [2]-[4]. But for deep neural network (DNN) learning methods including multiple layers processing image elements and estimating features automatically for classification will save more time for end to end training [5]. Therefore, this project with the aim of implementing and selecting the most effective method to quickly and accurately identify plant diseases by classifying diseased leaves of plants, which helps farmers be more proactive in preventing and timely handling, avoiding affecting crop yields.

Identifying plant diseases involves recognizing symptoms that commonly show up in different parts of the plant, such as leaves, stems, and fruit pulp. However, accurate classification of diseases requires specialized knowledge. Reaching rural areas, particularly in less developed countries where smallholder farmers grow most crops, poses challenges. However, leveraging advanced technology enables expert-level diagnosis even in these challenging locations. Consider the prevalence of affordable smartphones and extensive internet coverage; these factors create an ideal foundation for a smartphone-based disease diagnosis service. Farmers can simply capture images of their crops, and the mobile disease detection system will efficiently identify and label any diseases present. This streamlined process can significantly reduce crop damage by eliminating the need for time-consuming steps, such as expert farm visits, in the conventional diagnostic procedure [6].The above studies have achieved good results and provided good research ideas and methodologies for crop disease detection. The real agricultural environment has many special characteristics compared to other areas of detection tasks, and the crop disease research described above also faces the following problems: (i) the different distances between the crop and the camera result in different resolutions and scales of disease spots featured on the image, which can cause significant differences in the features of the same disease on different images; (ii) due to differences in leaf pose and camera angle, the features of the same crop disease image may vary greatly on images from different shooting angles; and (iii) there are a wide variety of crop diseases, and the leaves, flowers and fruit parts of crops are susceptible to different diseases. The features of different disease images vary widely, and even the same disease can vary greatly in form, profile, color, size and other features on different parts of the crop. These reasons contribute to the fact that existing disease detection models cannot be well applied in real agricultural scenarios.

Convolutional neural networks (CNNs) have showcased cutting-edge performance in tasks such as image classification and various computer vision applications. This study investigated the effects of both local feature and global feature based on transformer (the proposed overal pipe-line of DNN shown in Figure 1). While earrlier the traditional approaches frequently have explored the impact of pathology in an independent feature space of high-level or low-level features, they have lack sufficient semantics for accurate classification, while high-level features may not offer detailed boundary information. To tackle this challenge, we introduce an innovative model for predicting crop pathology within a global-local feature space. Our approach involves utilizing refined feature pyramids from various layers to integrate semantics from both high-level and low-level features. Additionally, to capture the extended temporal context of the entire image, we employ a transformer to grasp long-range dependencies among frames. Local feature interpretability provides detailed information for model decisions, while global interpretability offers general and holistic insights into model learning. This dual approach global-local feature based on transformer achieve various objectives, including building confidence in the model, identifying and rectifying biases and errors as well as optimizing overall performance. Subsequently, the refined features incorporating long-range dependencies are input into a classifier for coarse crop pathology prediction for enhancement prediction accuracy even further. Our contribution of this paper are as follows:

- We propose the novel model association with local-global feature based on transformer.
- We peformance the effective local-global feature representation to the the proposed DNN on public dataset and real-life dataset from Ministry of VietNam Agriculture.
- We intergrate the apropriate model on popular smartphones which assists farmers to use easily.

The paper is structured as follows. Section 2 deals about the previous related works. Section 3 introduces about proposed DNN. Section 4 depicts the results and discussion, followed by a conclusion.

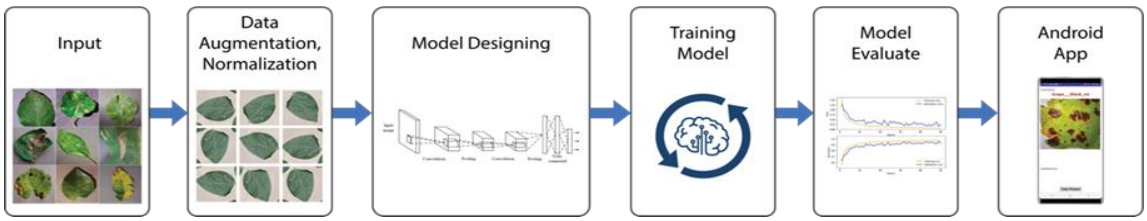

Figure 1. Classification pipe-line

2. RELATED WORK

In previous years, machine learning and advanced deep learning techniques have been applied to the plant disease detection problem. Several methods were implemented in the field of plant disease detection, a literature review was carried out to compare different methods proposed earlier. Many studies had proposed different solutions to detect diseases.

Jasim and Al-Tuwaijari [7], delved into the exploration of a deep learning neural network algorithm. Utilizing Tensorflow, they processed the data to make it suitable for training, subsequently constructing a model for plant disease detection. The implemented application underwent evaluation by specialists, achieving a training accuracy of 98.29% and testing accuracy of 98.029% across the entire dataset. Patel and Joshi [8], global researchers took a serious stance on addressing plant disease issues, focusing on prevention and early detection. Various techniques, including machine learning and deep learning, were explored. Kiran *et al.* [9] and Alagumariappan *et al.* [10] addressed the problem by employing step-by-step image processing techniques such as image acquisition, preprocessing, segmentation, feature extraction, and classification. Most of these methods relied on manual features and conventional machine learning techniques. Mahalakshmi *et al.* [11], concentrated on recognizing paddy plant diseases using image processing. The system took paddy leaf images as input, converting red, green and blue (RGB) images to grayscale, applying morphological opening operations to reduce noise, and concluding with image segmentation. Thorat *et al.* [12] presented an internet of things (IoT)-based system for automatic plant leaf disease detection. Sensor devices collected plant and leaf images, and image processing, k-means clustering, and artificial neural networks were applied. The device classified diseases based on monitoring temperature, humidity, and plant moisture, achieving an accuracy exceeding 90%. Ferentinos [13] implemented various deep learning model architectures, based on CNN architectures, to identify plant diseases using leaf images of both healthy and diseased plants. Beside [14], [15] demonstrates that a deep CNN model can be constructed and improved by adjusting multiple parameters to enhance identification accuracy performance.

The cutting-edge model DeepPestNet [16] with 11 hidden layers achieved 98.92% accuracy. Transfer learning models were widely used to obtain high efficiency more easily, in [17], the VGG19 architecture was compared in detecting diseases on apple trees and showed VGG19 functioned better with 87.7% accuracy performance. Sravan *et al.* [18] implemented to adjust hyperparameter of the ResNet-50 architecture to get 99.26 higher accuracy on the plant village dataset. VGG CNN architecture performed the best. It accomplished the accuracy of approximately 99.53% in classifying 17,548 plant leaf images. Moreover, deep learning structures were combined simultaneously with traditional handcrafted feature methods to extract maximum features from input images. The achieved efficiency was 99.79% for apple leaf datasets. Raut and Fulsunge [19], demonstrated proficient and accurate techniques for plant disease detection and identification employing image processing. They utilized the K-means clustering algorithm and various support vector machine (SVM) methods [20], systematically organized for the identification of both plant and fruit diseases. Image segmentation and feature extraction were employed to preprocess images for training. Currently, attention mechanisms have been applied as plugins to enhance the efficiency of the DNN model [21], [22].

3. PROPOSED DNN

The classification pipeline involves multiple steps, illustrated in Figure 2, starting with a manual analysis of the dataset to identify class imbalances. Class imbalance, where certain classes are underrepresented compared to others, can lead to biased model predictions. To address this issue, data augmentation techniques are applied to augment the underrepresented classes and mitigate the impact of class imbalances. After class balancing next crucial step is to normalize the dataset after augmented to be suitable for the input of the model. After that, we design a shallow CNN model and at the same time use, several state-of-the-art CNNs are available, which is network is trained on the ImageNet dataset, called transfer learning. In transfer learning method, we trained model by random initialization, fine-turning and feature extraction. In the case of random initialization, the network was trained from scratch with no frozen layers. Fine-tuning involved initializing the network with pre-trained weights, employing conventional transfer learning. For feature extraction, all layers of the network were frozen during the experiment, and the network served solely as a feature extractor. The architecture of CNN was designed in this experiment is shown in Table 1. When all models that we selected were trained, we will select the best model. In the last, we will apply the model to the android app by using TensorFlow Lite API to convert model to (.tflite) format. An overview of the classification pipe-line can be seen in Figure 1.

3.1. Data augmentation

Data augmentation is a widely used method to augment the dataset by applying various transformation techniques, thereby increasing the number of samples. This exposes the model to diverse aspects of the training data, effectively mitigating overfitting. The augmentation layer incorporates random flipping in both horizontal and vertical directions, as well as random rotation and random zoom of images. Additional data augmentation steps encompass random flipping, random rotation, and random zoom of images, illustrated in Figure 1.

Table 1. The architecture of CNN with shape of parameters

Layer (type)	Base bond output shape	Param (#)	Local-global	Transformer (mult-head attention)	FCs
input_1 (InputLayer)	[(None, 229, 229, 3)]	0			
Data augmentation layer (random horizontal and vertical flip, Random rotation, Random zoom)	[(None, 229, 229, 3)]	0			
conv2d (Conv2D) (3x3)	(None, 147, 147, 64)	#1792			
BatchNorm ()					
max_pooling2d (MaxPooling2D)	(None, 147, 147, 64)	0			
conv2d_1 (Conv2D)	(None, 128, 128, 32)	#18464			(None, 512) #1049088 (None, 128) #65664 (None, 38) #4902
BatchNorm ()					
max_pooling2d_1 (MaxPooling2D)	(None, 64, 64, 32)	0		(None, 2048)	
	Reshape **LF1** (None, 131072)	0	(None, 133120)		
conv2d_2 (Conv2D)	(None, 64, 64, 16)	#4624			
BatchNorm ()					
max_pooling2d_2 (MaxPooling2D)	(None, 32, 32, 16)	0			
	Reshape **LF2** (None, 16384)	0	(None, 18432)		
conv2d_3 (Conv2D)	(None, 32, 32, 8)	#1160			
BatchNorm ()					
max_pooling2d_3 (MaxPooling2D)	(None, 16, 16, 8)	0			
	Reshape **LF3** (None, 2048)	0	(None, 4096)		
Flatten (Flatten)	(None, 2048)	0			
dropout (Dropout)	**GF** (None, 2048)	0			

3.2. LGT architecture

The architecture is formulated using a multi-level CNN model. The initial convolutional layer takes an image input shape of 256×256×3, utilizing 64 filters with a kernel size of 3×3, employing "same" padding, and a strides setting of 1×1. The second convolutional layer maintains the shape of the first layer, incorporating max pool size 2×2 and strides of 2×2 for additional features. In the third convolutional layer, the image input shape becomes 128×128×64, utilizing 32 filters with a kernel size of 3×3, "same" padding, and a strides setting of 1×1. The fourth layer introduces max pool size 2×2 and strides of 2×2. Moving to the fifth convolutional layer, the image input shape is 64×64×32, with 16 filters, a kernel size of 3×3, "same" padding, and strides of 1×1. The sixth layer incorporates a max pool size of 2×2 and a stride of 2×2. In the seventh convolutional layer, the image input shape is 32×32×3, utilizing 8 filters with a kernel size of 3×3, "same" padding, and strides of 1×1. rectified linear unit (ReLU) activation functions are applied in both conv2d layers and dense layers.

The flatten layer uses 2048 units of the dense layer and among them, 50 percent is dropped by the ReLu activation function. In the dropout layer, 0.2 is used as rate to drop units the flatten layers. The last layer (dense_2), we used 38 units with a softmax activation function for the classifier. Transformer technique (multi-head attention) was applied to upgrade the efficiency compared to just using the CNN normal model. The input of transformer block are local-features which are combined with global-feature. These local-features are defined as output channels of the fourth layer, the sixth layer and the eighth layer. The global-feature is the output channel of the last dropout. Overal proposed local-global based multi-head attention (LGT) that inspide from [23] was showed on Figure 2.

The model of the proposed method is shown in Figure 2 with three key parts: extract features form input image by using a CNN pretrained model to obtain local features (LF) và global feature (GF), the combination of local features and global feature part and use a transformer to capture the long-range dependencies of feature levels. At first, A 10-layer CNN network is used to extract features of the input image. Like conventional CNN networks, layers will perform layer-by-layer convolution to extract features. But instead of just using the output data on the final layer, we took advantage of the data at the lower-levels where the rich details of the boundaries are preserved they are called local-features. These local-features are defined as output channels of the fourth layer, the sixth layer and the eighth layer. The output of the dropout layer is called the global-feature. Collecting multiple outputs of feature pyramids of different layers allows the model to obtain both rich details of the boundaries of low-levels and semantics for classification of high-levels. Global-feature and local-feature vectors are catenated to form higher dimensional fusion features. Fusion features explores data features for describe their rich insider information. The three output of features from fusion global and local will become the input data of the transformer block. The fusion features that are combined from the local features on each level of feature pyramids and transformer block captures the long-range dependencies of different feature levels.

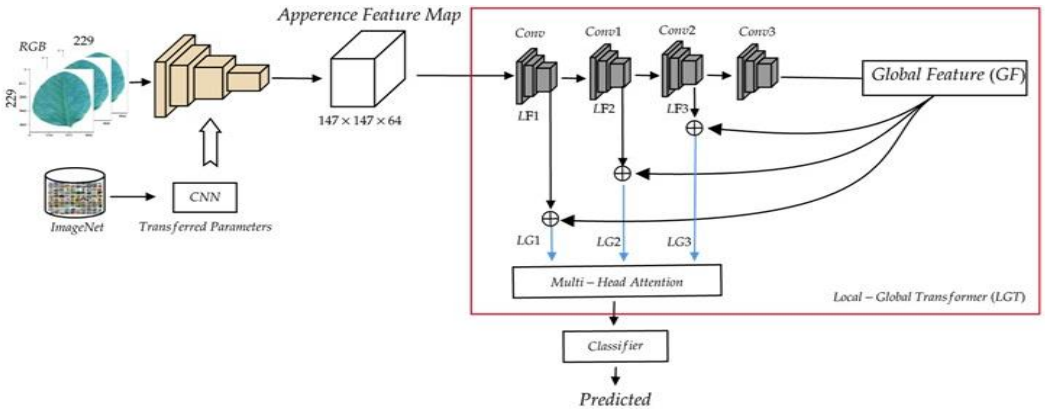

Figure 2. Overalls proposed local-global based transformer (LGT)

3.3. Hyperparameter

The deep learning model's parameters are acquired by minimizing the loss function via a stochastic process, such as stochastic gradient descent (SGD), Adam, BFGS, and root mean square propagation (RMSprop). In the presented model, a learning rate of 0.001 is employed in the ADAM optimization [24]. The categorical crossentropy serving is chosen loss function when training.

4. RESULTS AND DISCUSSION

4.1. Dataset

For this study, we have used a public dataset for plant leaf disease detection called PlantVillage [10]. The dataset consists of 54,303 RGB images with a size of 256×256×3 for each image, containing 38 crop diseases, i.e 26 disease classes, and 12 healthy plant classes. Some sample images of dataset are shown in Figure 3. The images contain a detailed description of the leaves before and after diseases affect them. In Kaggle, the plant village, a dataset open to the public, has now collected 54,309 images of lava plant diseases, contains 14 types of vegetables and fruits, such as apples, grapes, soybeans, apples, tomato, and corn contains total 26 diseases out of that 17 types of fungal diseases, 4 types of viral diseases and 1 type of diseases caused.

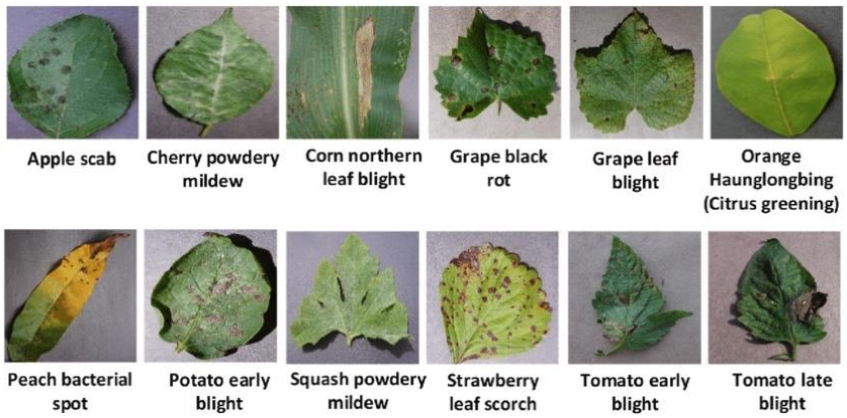

Figure 3. Some sample images from PlantVillage dataset

Dataset collection for strawberry disease identification from VietNam strawberry disease (VNStr): strawberries are highly valued in agriculture in each country. However, the strawberry diseases are very diverse and spread very quickly in a short time. This reduces the number of strawberries and causes financial losses to the farmers. Therefore, we created a system to detect strawberry disease with data including real-life images of healthy and diseased strawberry.

The complexity of the real world, we create models trained on taking images of strawberry disease in the farm setting from the Vietnam Ministry of Agriculture. We collected 1,000 images by using the scientific and common names of 5 classes (Figure 4) mentioned in our dataset include normal strawberry, powdery mildew disease of strawberries, black spot disease of strawberries, gray mold disease of strawberries, rubber disease of strawberries as shown in Figures 4(a) to (e) respectively. One of the most important factors for the classification (arrangement into groups) were the color, area, and density of the diseased part and the shape of the species. To enhance accuracy, we excluded erroneous images (e.g., non-strawberry, lab-controlled, and out-of-scope images). Each image underwent scrutiny by two individuals following guidelines to minimize labeling errors.

(a) (b) (c) (d) (e)

Figure 4. Example of VNStr datase; (a) normal, (b) powdery mildew disease, (c) black spot disease, (d) gray mold disease, and (e) rubber disease

4.2. Experimental results

4.2.1. Effect of our proposed DNN

In this performance, using pretraied model Resnet 101 followed six type have designed to evaluate the effective the diffirence level in structure as shown in Figure 2. Type 1 is only local feature 2(LF2) and classifier, type 2 is only local feature 3 (LF3) and classifier, type 3 is global feature (GF) and classifier, type 4 is LF1 and GF, type 5 is LF2 and GF, type 6 is proposed LF3 and GF. The proposed model with fusing LFs and the GF solves above problems, resulting in a model that converges faster, with a smoother curve and higher accuracy. The graph for training and validation accuracy and loss of LGT model in VNStr dataset is shown in Figure 5. Fusing the outputs of low feature levels with a GF level improved the loss and accuracy curves with faster convergence. Figure 6 list the six architecture performanced in PlantVillage data. The curve in Figure 6(a) is convergence slowly compared to Figure 6(b) because of learning deeply. When LF2 was fused with GF as shown in Figure 6(e), the fusion features still had the features of the low level, making the curve on the validation set not smooth and highly variable. When we incorporate higher layers in the LGT architecture (Figure 6(f)), the variability improves, but still falls short of expectations.

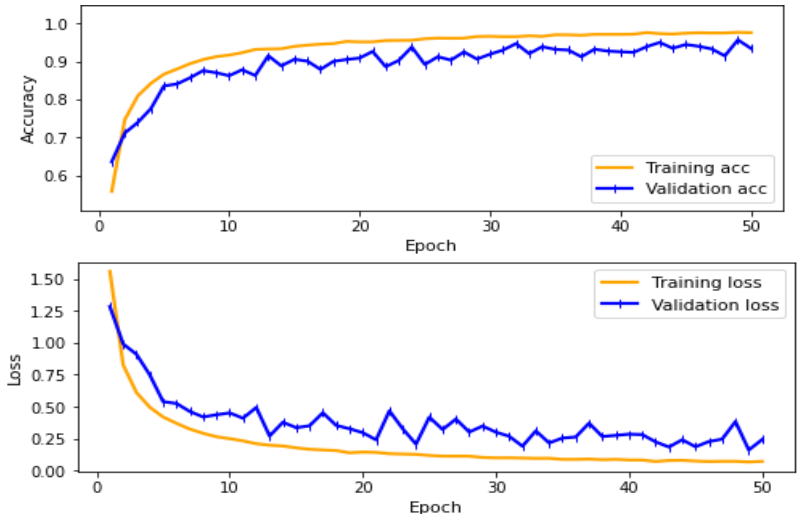

Figure 5. Training and validation accuracy and loss curve on VNStr dataset

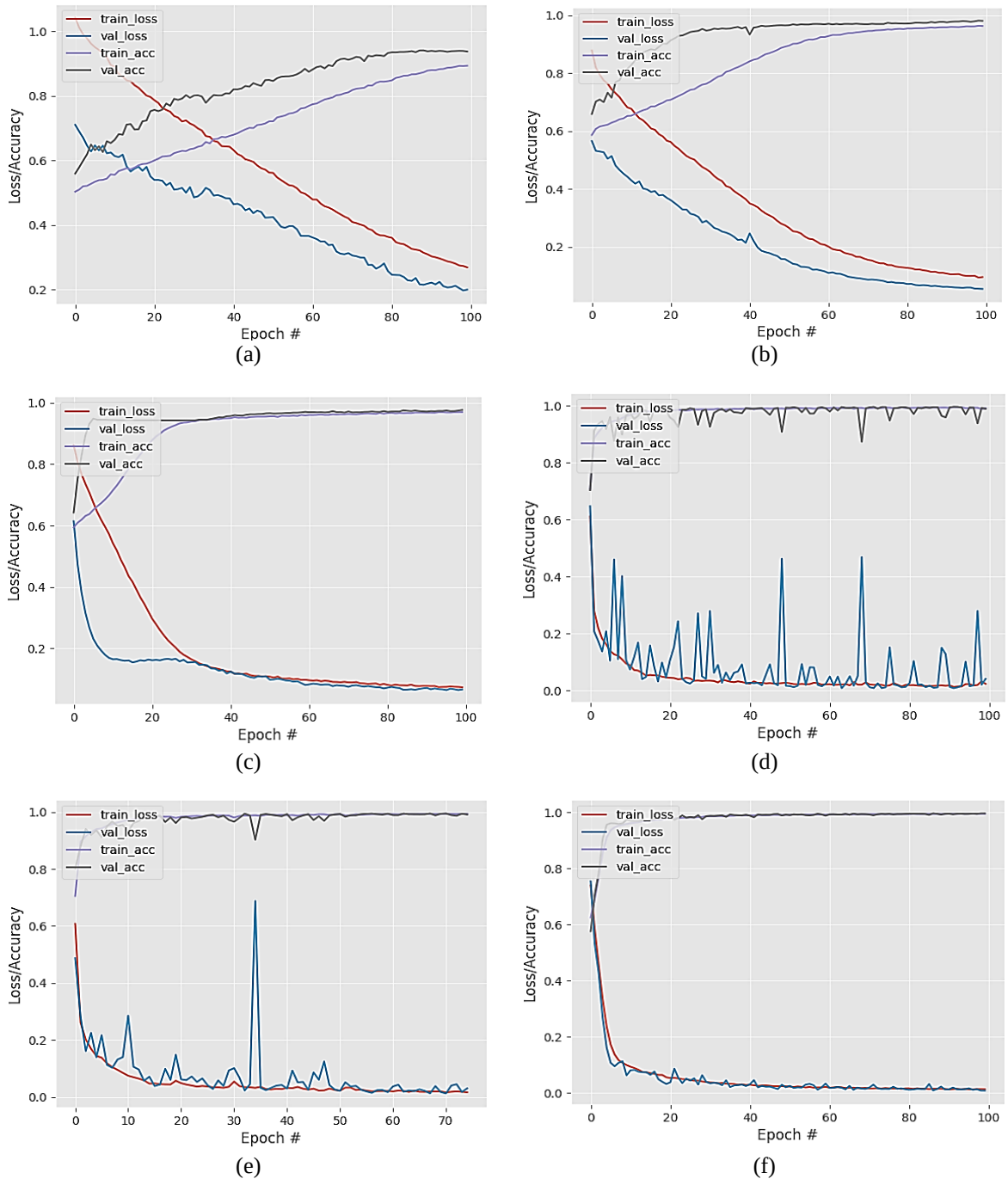

Figure 6. Illustration 6 type of PlantVillage data for training and validation: (a) LF2, (b) LF3, (c) GF, (d) LG2, (e) LG3, and (f) LGT

Effectiveness of our proposed on PlantVillage dataset: perform accuracy tests on PlantVillage dataset with 6 models, as shown in Table 2. With the first model, we only used the architecture of a CNN network with the first six layers (local feature 2 (LF2)) as input for the classifier block, obtaining an accuracy of 84.1% on the validation set and 85.8% on the test set. Accuracy increased as we replaced LF2 with higher-level features such as LF3 and GF. The accuracy values of the two models above were 88.3% on the validation set and 87.7% on the test set for LF3 combined with the classifier block, and 89.2% on the validation set and 88.9% on the test set for GF combined with the classifier block.

When we fuse global and local features, the accuracy improves significantly. High-level features help achieve better results because they provide better semantics for classification, while lower levels (LF)

provide rich details of the boundaries. Especially, with the architecture GF+LF3+classifier, the accuracy on both val and test sets reached 98.1%.

For the proposed model, we used three local features fusion with a global feature, processing through the transformer and implementing the classifier block which gained 99.05% on the val set and 99.18% on the test set. For VNStr dataset, the oucome performence reach 93.11 on val set and 94.05 on test set. This result is much better than the conventional CNN models tested above.

Table 2. Performance comparison of individual modules with pretrained (ResNet101) on two datasets

Architectures	PlantVillage dataset (%)		Own data VNStr (%)	
	val	test	val	test
LF2+classifier	84.14	85.65	73.00	73.13
LF3+classifier	88.37	87.77	80.30	80.24
GF+classifier	90.21	91.43	83.90	84.16
LG2+classifier	95.86	97.98	87.81	88.07
LG3+classifier	98.10	98.10	90.90	90.96
Our proposed (LGT)	**99.05**	**99.18**	**93.11**	**94.05**

Effectiveness of own dataset VNStr: structures of LGT at variant simplified configurations based on transformer are evaluated by accuracy. In Table 3, the architecture of Resnet101 pretrained+LGT achieves the best performance with increasing test accuracy range of 19.92 % on VNStr dataset. The main reason is that the architecture of LGT can provide significant local-global information as well as correlation context for strawberry disease identification.

Table 3. Performance comparison of difference pretrained model CNN

Pretrained CNN	Architectures	PlantVillage dataset (%)		Own data VNStr (%)	
		val	test	val	test
MobileNetV2	LGT+classifier	83.27	85.69	74.00	74.13
MobileNetV3Small		87.74	87.57	81.70	81.44
MobileNetV3Lage		89.43	88.44	82.50	83.15
InceptionV3		98.22	98.53	92.70	92.46
Resnet34		99.87	98.88	92.91	92.05
Resnet50		99.00	99.00	93.01	93.98
Resnet101		99.05	99.18	93.11	94.05

4.2.2. Comparison state of the art on PlantVillage dataset

Table 4 presents a performance comparison between the proposed and previous DNNs for distinguishing the PlantVillage dataset. The proposed method may benefit from the cardiovascular benefits of leveraging Resnet101 pretrained CNN, thus, the outcomes of the experiments showcase that the proposed DNN, exhibits superior performance compared to traditional DNNs, achieving accuracy improvements ranging from 0.98% to 8.78%. Consequently, the proposed DNN emerges as a potent tool for robust disease recognition across diverse content-aware applications. Despite achieving promising results by utilizing state-of-the-art pretrained models in combination with training the proposed model, our work stops at trial and error, experimenting with various architectures based on experience as well as subsequently evaluating and comparing them to draw conclusions. In the next steps of our work, we aim to employ optimization methods to identify optimal hyperparameters and a comprehensive architecture that meets the general requirements for a broader range of data types.

Table 4. Comparition to conventinal work

References	Method	Year	Feature	Accuracy (%)
Wang *et al.* [25]	VGG16 model trained with transfer learning	2017	Deep	90.40
Khan *et al.* [3]	M-SVM	2019	Color, histogram, LBP	97.20
Pradhan *et al.* [6]	transfer learning-Xception	2022	Deep	98.75
Song and Gao [26]	Swin transformer	2022	Deep	98.70
Paymode and Malode [27]	Multi-crops leaf disease+VGG16	2022	Deep	98.40
Sadhasivam *et al.* [1]	DNN	2024	Deep	98.20
Proposed	Pretrained (Resnet101)+LGT		deep	**99.18**

4.3. Android application interface

Making the above trained model run on mobile devices will make theme easier to use for real life cases. However, due to limited physical capabilities and power consumption, we are forced to optimize the model to fit mobile devices. Tensorflow Lite is an open-source framework designed for on-device inference. TensorFlow Lite interpreter is the tool that runs models in the mobile devices. In Figure 7, our workflow is shown for the TensorFlow Lite use case on device mobile. The select the model are MobileNetV3Lage because of small enough paramester. The model has (.h5) format, we convert to (.tflite) format by using the TensorFlow Lite API. Then, we add the model in the format (.tflite) to Android Studio. In Android Studio, we designed a theme for the android app, which is available on activity_main.xml. A theme is shown on Figure 7.

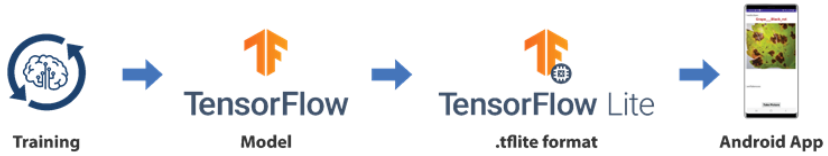

Figure 7. Workflow for TensorFlow lite use case on-device mobile

5. CONCLUSION

In this paper, different diseases on popular plants in agriculture were examined to release efficient models in classification. Instead of just using feature space of a high-level (global level), the recommended model utilized feature space from local levels and combined with transformer aggregation technique to elevate the predictions. The PlantVillage dataset and the VNStr dataset were employed for training and evaluating to identify the optimal model suitable for the given problem data. The proposed local-global feature-based transformer model achieved accuracies of 99.18% and 94.05% on the test set for the PlantVillage dataset and VNStr dataset, respectively. An agricultural model helped the training happened easily and did not waste many dealing materials on smartphone devices. This assisted the application on reality easily and effectively.

ACKNOWLEDGEMENT

This research is funded by International School, Vietnam National University, Hanoi (VNU-IS) under project number CS.2023-01.

REFERENCES

[1] M. Sadhasivam, M. K. Geetha, and J. G. M. Britto, "Efficient deep learning architecture for the classification of diseased plant leaves," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 1, pp. 198–206, Jan. 2024, doi: 10.11591/ijeecs.v33.i1.pp198-206.

[2] O. T. C. Chen, C. H. Tsai, H. H. Manh, and W. C. Lai, "Activity recognition using a panoramic camera for homecare," Aug. 2017, doi: 10.1109/AVSS.2017.8078546.

[3] M. A. Khan *et al.*, "An optimized method for segmentation and classification of apple diseases based on strong correlation and genetic algorithm based feature selection," *IEEE Access*, vol. 7, pp. 46261–46277, 2019, doi: 10.1109/ACCESS.2019.2908040.

[4] K. W. V. Geollegue, E. R. Arboleda, and A. A. Dizon, "Seed of rice plant classification using coarse tree classifier," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 2, pp. 727–735, Jun. 2022, doi: 10.11591/ijai.v11.i2.pp727-735.

[5] M. H. Ha and O. T. C. Chen, "Deep neural networks using residual fast-slow refined highway and global atomic spatial attention for action recognition and detection," *IEEE Access*, vol. 9, pp. 164887–164902, 2021, doi: 10.1109/ACCESS.2021.3134694.

[6] P. Pradhan, B. Kumar, and S. Mohan, "Comparison of various deep convolutional neural network models to discriminate apple leaf diseases using transfer learning," *Journal of Plant Diseases and Protection*, vol. 129, no. 6, pp. 1461–1473, Aug. 2022, doi: 10.1007/s41348-022-00660-1.

[7] M. A. Jasim and J. M. Al-Tuwaijari, "Plant leaf diseases detection and classification using image processing and deep learning techniques," in *Proceedings of the 2020 International Conference on Computer Science and Software Engineering, CSASE 2020*, Apr. 2020, pp. 259–265, doi: 10.1109/CSASE48920.2020.9142097.

[8] A. Patel and M. B. Joshi, "A survey on the plant leaf disease detection techniques," *Ijarcce*, pp. 229–231, Mar. 2017, doi: 10.17148/ijarcce.2017.6143.

[9] A. Kiran, S. K. Lokesh Naik, M. S. Raj, and S. K. Palvadi, "Plant disease detection using image processing with machine learning," in *2023 4th International Conference on Electronics and Sustainable Communication Systems, ICESC 2023 - Proceedings*, Jul. 2023, pp. 1590–1595, doi: 10.1109/ICESC57686.2023.10192986.

[10] P. Alagumariappan, N. J. Dewan, G. N. Muthukrishnan, B. K. B. Raju, R. A. A. Bilal, and V. Sankaran, "Intelligent plant disease identification system using machine learning †," in *Engineering Proceedings*, Nov. 2020, vol. 2, no. 1, doi: 10.3390/ecsa-7-08160.

[11] J. Mahalakshmi and G. Shanthakumari, "Automated crop inspection and pest control using image processing," *International Journal of Engineering Research and development*, vol. 13, no. 3, pp. 25–35, 2017.

[12] M. N. R. Thorat, P. Pimpri, and S. Nikam, "Early disease detection and monitoring large field of crop by using IoT," *International Journal of Computer Science and Information Security*, vol. 15, no. 10, 2017.
[13] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, Feb. 2018, doi: 10.1016/j.compag.2018.01.009.
[14] I. A. M. Zin, Z. Ibrahim, D. Isa, S. Aliman, N. Sabri, and N. N. A. Mangshor, "Herbal plant recognition using deep convolutional neural network," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 5, pp. 2198–2205, Oct. 2020, doi: 10.11591/eei.v9i5.2250.
[15] T. Hidayat and A. R. Nilawati, "Identification of plant types by leaf textures based on the backpropagation neural network," *International Journal of Electrical and Computer Engineering*, vol. 8, no. 6, pp. 5389–5398, Dec2018, doi: 10.11591/ijece.v8i6.pp5389-5398.
[16] X. Fan, P. Luo, Y. Mu, R. Zhou, T. Tjahjadi, and Y. Ren, "Leaf image based plant disease identification using transfer learning and feature fusion," *Computers and Electronics in Agriculture*, vol. 196, p. 106892, May 2022, doi: 10.1016/j.compag.2022.106892.
[17] S. T., R. Khilar, and M. Subaja Christo, "WITHDRAWN: a comparative analysis on plant pathology classification using deep learning architecture – Resnet and VGG19," *Materials Today: Proceedings*, Jan. 2021, doi: 10.1016/j.matpr.2020.11.993.
[18] V. Sravan, K. Swaraj, K. Meenakshi, and P. Kora, "WITHDRAWN: a deep learning based crop disease classification using transfer learning," *Materials Today: Proceedings*, Feb. 2021, doi: 10.1016/j.matpr.2020.10.846.
[19] S. Raut and A. Fulsunge, "Plant disease detection in image processing using MATLAB," *International journal of innovative Research in science, Engineering and Technology*, vol. 6, no. 6, 2017.
[20] Z. Ibrahim, N. Sabri, and N. N. A. Mangshor, "Leaf recognition using texture features for herbal plant identification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 9, no. 1, pp. 152–156, Jan. 2018, doi: 10.11591/ijeecs.v9.i1.pp152-156.
[21] O. T. C. Chen, M. H. Ha, and Y. L. Lee, "Computation-affordable recognition system for activity identification using a smart phone at home," in *Proceedings - IEEE International Symposium on Circuits and Systems*, Oct. 2020, vol. 2020-October, doi: 10.1109/iscas45731.2020.9180826.
[22] M. H. Ha, T. A. Pham, D. T. Thanh, and V. L. Tran, "Attention correlated appearance and motion feature followed temporal learning for activity recognition," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 2, pp. 1510–1521, Apr. 2023, doi: 10.11591/ijece.v13i2.pp1510-1521.
[23] M. H. Ha and O. T. C. Chen, "Non-local spatiotemporal correlation attention for action recognition," Jul. 2022, doi: 10.1109/ICMEW56448.2022.9859314.
[24] M. H. Ha and O. T. C. Chen, "Deep neural networks using capsule networks and skeleton-based attentions for action recognition," *IEEE Access*, vol. 9, pp. 6164–6178, 2021, doi: 10.1109/ACCESS.2020.3048741.
[25] G. Wang, Y. Sun, and J. Wang, "Automatic image-based plant disease severity estimation using deep learning," *Computational Intelligence and Neuroscience*, vol. 2017, pp. 1–8, 2017, doi: 10.1155/2017/2917536.
[26] H. Song and Y. Gao, "Plant diseases recognition on digital images using swin transformer," in *ACM International Conference Proceeding Series*, Nov. 2022, pp. 219–223, doi: 10.1145/3581807.3581839.
[27] A. S. Paymode and V. B. Malode, "Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG," *Artificial Intelligence in Agriculture*, vol. 6, pp. 23–33, 2022, doi: 10.1016/j.aiia.2021.12.002.

BIOGRAPHIES OF AUTHORS

Manh-Hung Ha received the M.S. degrees in Information Communication Technology from University of Paris 13, France, and the Ph.D. degree with the Department of Electrical Engineering, National Chung Cheng University, Taiwan in 2014 and 2021, respectively. He was lecturer with the Faculty of Electrical Engineering, Phenikaa University, Hanoi, Viet Nam, from September 2021 to July 2022. Since July 2022, he has been a lecturer with the Faculty of Applied Science, International School, Vietnam National University, Hanoi, Vietnam. His major research interests include multimedia/image/video analytics, computer vision, speech signal processing, and machine learning. He can be contacted at email: hunghm@vnu.edu.vn.

Duc-Chinh Nguyen received the Degree of Engineer from the School of Information and Communications Technology at Hanoi University of Science and Technology, Vietnam in 2019. Since graduation, he has worked as a Web Development Engineer at Temona Inc., Tokyo, Japan. Currently, he is pursuing a Master's degree in Master of Informatics and Computer Engineering (MICE) at the International School, Vietnam National University, Hanoi. His research interests focus on computer vision, machine learning, and deep learning, particularly in graph-related architectures. He can be contacted at email: 22075057@vnu.edu.vn.

Manh-Tuan Do holds a Bachelor's degree in Electronics from the University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam, awarded in 2023. Following his graduation, he assumed the role of an Assistant Lecturer at the Faculty of Applied Science, International School, Vietnam National University, Hanoi, Vietnam. His research interests primarily focus on deep learning, with a specific emphasis on one-stage object detection models such as YOLO. He can be contacted at email: 19021129@vnu.edu.vn.

Dinh-Thai Kim was born in 1984. He received the M.S. degree in control engineering and automation from the Thai Nguyen University of Technology, Vietnam, in 2013, and the Ph.D. degree in electrical and communications engineering from Feng Chia University, Taichung, Taiwan, in 2020. He is currently a Lecturer with the International School, Vietnam National University, Hanoi. His research interests include computer vision, intelligent control, and robotics. He can be contacted at email: thaikd@vnu.edu.vn.

Xuan-Hai Le was born in 1982. He received the M.Sc. and Ph.D. degrees in control and automation from the Hanoi University of Science and Technology, in 2011 and 2019, respectively. He is currently a Lecturer with the International School, Vietnam National University, Hanoi. His research interests include deep learning and applying intelligent control in under-actuated systems for robots, overhead cranes, ships, UAVs, and self-balancing cars. He can be contacted at email: hailx@vnu.edu.vn.

Ngoc-Thanh Pham received the B.Eng., degree from the Hanoi University of Science and Technology, Hanoi, Vietnam, in 2009, and the M.Eng., and Ph.D. degrees from Deakin University, Geelong, Victoria, Australia, in 2013 and 2017, respectively, where he is currently working with the School of Engineering, Deakin University until 2021. Currently, he is a Lecturer with the International School, Vietnam National University, Hanoi. His research activities include advanced control and reinforcement learning, cyber security and detection, smart grid technology, and intelligent systems, machine learning. He can be contacted at email: pnthanh@vnu.edu.vn.