Learning Deep Transferability for Several Agricultural Classification Problems

Nghia Duong-Trung[1]
FPT University
Can Tho City, Viet Nam

Luyl-Da Quach[2]
Tay Do University
Can Tho City, Viet Nam

Chi-Ngon Nguyen[3]
Can Tho University
Can Tho City, Viet Nam

Abstract—This paper addresses several critical agricultural classification problems, e.g. grain discoloration and medicinal plants identification and classification, in Vietnam via combining the idea of knowledge transferability and state-of-the-art deep convolutional neural networks. Grain discoloration disease of rice is an emerging threat to rice harvest in Vietnam as well as all over the world and it acquires specific attention as it results in qualitative loss of harvested crop. Medicinal plants are an important element of indigenous medical systems. These resources are usually regarded as a part of culture's traditional knowledge. Accurate classification is preliminary to any kind of intervention and recommendation of services. Hence, leveraging technology in automatic classification of these problems has become essential. Unfortunately, building and training a machine learning model from scratch is next to impossible due to the lack of hardware infrastructure and finance support. It painfully restricts the requirements of rapid solutions to deal with the demand. For this purpose, the authors have exploited the idea of transfer learning which is the improvement of learning in a new prediction task through the transferability of knowledge from a related prediction task that has already been learned. By utilizing state-of-the-art deep networks re-trained upon our collected data, our extensive experiments show that the proposed combination performs perfectly and achieves the classification accuracy of 98.7% and 98.5% on our collected datasets within the acceptable training time on a normal laptop. A mobile application is also deployed to facilitate further integrated recommendation and services.

Keywords—*Medicinal Plant Classification; Grain Discoloration Classification; Transfer Learning; Deep Learning*

I. INTRODUCTION

Rice is not only a major food crop in Vietnam but also an important export product. The economy of different countries, as well as Vietnam, is highly dependent on the export of the particular commodity as Vietnam is the 5th leading exporter of rice in the world [1], [2]. Rice is a highly nutritive cereal and is consumed as essential food in most of Asian countries [3], [4]. In Vietnam, rice crop is subjected to various diseases which affect its quality and reduce the entire production. In the recent year, a new harvest reducing disease, grain discoloration, is becoming a serious problem to the reduction of rice crops [5], [6], [7], [8]. In this sense, the rice grain discoloration is considered as a potential risk to the rice-producing countries and various reports from number parts of the rice industry about this disease strongly requires a solution. Accurate classification of the grain discoloration is essential before proposing any practical control schemes. Nevertheless, either effective intervention or accurate prediction showing a complete solution to the disease is currently unexplored.

According to World Health Organization, about 70% of the world's population relies on plants for their primary health care and some 35,000 to 70,000 species have been used as medicament [9], a figure corresponding to $14 - 28\%$ of the 250,000 plants species estimated to occur around the world [10], [11], and equivalent to $35-70\%$ of all species used worldwide [10]. Medicinal plants are of crucial importance to the health of human beings. The medicinal value of these plants, both wild and planted, lies in some chemical substances that produce a physiological action on the human body. Many of these indigenous medicinal plants are used as food ingredients and medical purposes [13], [15], [16]. Furthermore, the special significance of medicinal plants in conservation peduncles from the major cultural, livelihood or economic roles that they play in many people's lives. Various sets of recommendations have been compiled relating to the conservation of medicinal plants[12]. Vietnam is home to an estimated 12,000 species of high-value plants, of which 10,500 have been identified, and approximately 3,780 species have medicinal properties. Vietnamese medicinal plant plants account for approximately 11% of the 35,000 species of medicinal plants known worldwide. The market size for Vietnamese herbal products and medicinal dietary supplement products at an estimated US $100 million[14]. With its abundant indigenous plant varieties, medicinal plants, and associated traditional knowledge, it is undoubtedly that Vietnam's biodiversity has a crucial role in contributing to sustainable livelihoods over many generations through the provision of food security and health care [20], especially for local people living in remote areas who are directly dependent on resources exploitation. People in many rural areas of Vietnam classify plants according to their medicinal values. Classification is considered an important activity in the preparation of herbal medicines [21]. Despite the importance of research on medicinal plants, there are a few works have been conducted in the literature. The most recently intensive work was done almost decades ago [17], [18]. It is necessary to make people realize the importance of medicinal plants before their extinction. The knowledge of herbal medicines should be maintained and passed along future generations. It is important for practitioners and botanists to know how to identify and classify the medicinal plants through computers and devices. Accurate classification of the medicinal plants is essential before developing any recommendation and services.

From the machine learning perspective, the mentioned problems could be addressable by the adoption of a new rapid solution that can bring experts, farmers, policymakers, and strategists into one choir. Traditionally, a major assumption in many machine learning algorithms is that the training and

future data must in the similar feature space. They address isolated tasks. Any differences may be eliminated before learning or they have no equivalent covariance during training a model. However, in many real-world applications, this assumption may not hold. The isolation insists on an entire learning procedure from dataset collection, model training, model evaluation and model tuning. Thus, there is obviously a demand for computing infrastructure and financial support. Transfer learning, however, attempts to change it by developing methods to transfer knowledge learned in one or more source tasks and use it to improve learning in a related target task [32], [33], [24]. Modern object classification models have millions of parameters and can take weeks to fully train. Transfer learning is a technique that shortcuts a lot of this work by taking a fully-trained model for a set of predefined categories like ImageNet [22], [23], and retrains from the existing weights for new implemented classes. The goal of transfer learning is to improve learning in the target task by leveraging knowledge transferability from the source task. Some work in transfer learning is in the context of inductive learning [47] and involves extending well-known classification and deduction algorithms such as Markov logic Networks, Bayesian networks, and neural networks.

Transfer learning methods tend to be extensively dependent on the machine learning algorithms being used to learn the prediction tasks, and can often merely be considered extensions of those algorithms [34]. Tremendous progress has been made in image classification and recognition, primarily thanks to the availability of large-scale annotated datasets. Since Krizhevsky et al. [25] won the ImageNet 2012 competition, there have been much interest and work toward the revival of deep convolutional neural networks [26], especially in the task of image classification [27], [28], [29]. However, in this research, we aim neither to maximize absolute performance nor to build a complete model from scratch, but rather to study transfer results of several well-known convolutional architectures. We use the reference implementation provided by Tensorflow [30], [31] so that our experiment results will be comparable, extensible and usable for a large number of upcoming research. TensorFlow is a modern machine learning framework that provides tremendous power and opportunity to developers and data scientists. One of those opportunities is to utilize the concept of transfer learning to reduce training time and complexity by repurposing pre-trained models.

Building upon these key insights, we propose design recommendations for classification of grain discoloration and medicinal plants. To the best of own knowledge, in this paper, the authors have made several contributions:

- Firstly, we have collected grain discoloration samples, and medicinal plants samples and that can be served as benchmark datasets.

- Secondly, we have proposed the combination of deep learning via convolutional neural networks, the idea of transfer learning and several real-world agricultural classification problems that is previously unstudied in the literature.

- Thirdly, we have proved that knowledge from very diverse source task can be very helpful to a target even if the source task may not be sufficiently related.

- And lastly, a mobile application providing the most affordable ways for millions of people to access information is also deployed to facilitate further integrated recommendation and services.

II. PROPOSED METHODOLOGY

In the recent years, the growth of classification datasets and the manifold directions of object classification research provide an unprecedented need and a great opportunity for a thorough evaluation of the current state of the field of categorical object detection [22], [48]. Taking ImageNet [23] dataset as an example, it is a dataset of over 15 million labelled high-resolution images with around 22,000 categories. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) uses a subset of ImageNet of around 1000 images in each of 1000 categories. In all, there are roughly 1.2 million training images, 50,000 validation images and 100,000 testing images.

Convolutional neural networks (CNNs) have recently gained outstanding image classification performance in the large-scale challenges [25], [35], [22], [36], [38]. The success of CNNs is achieved by their ability to learn rich level image representations as its hidden layers can be integrated theoretically unlimited. However, learning CNNs requires a very large number of annotated image samples and an estimation of millions of model parameters. This property obviously prevents the application of CNNs to problems with limited training data. There is a phenomenon in deep neural networks such that when trained the model on images, it tends to learn first-layer features that resemble either Gabor filters or color blobs [37]. Such first-layer features appear not to be specific to a particular dataset, but generally in the way that they are applicable to different tasks. The transition of knowledge is eventually transferred from the first layer to the last layer of the network. Expectedly, several large-scale datasets can be used to train the learning models, and then the learned models are applied to a particular target task where the parameters of the last layer are re-weighted based on its own dataset. The idea of transferring knowledge along deep neural networks have been explored by many previous researches [37], [39], [40], [46]. Going to that research direction, we explore the performance of several state-of-the-art convolutional neural networks upon our collected data.

A. Inception-Based Models

Google introduced the inception deep convolutional architecture as GoogleNet or Inception-v1 in [42]. Later the Inception architecture was refined in several ways, mainly by the introduction of batch normalization and the reduction of internal covariate shift [43]. The third iteration of the architecture [44] was improved by additional factorization ideas which will be referred to as Inception-v3 in our implementation. The authors are aware of the latest version, Inception-v4 [45], but do not include it into this work. We aim to reserve it for further investigation where we are going to conduct a thorough performance comparison of various models upon our further collected datasets. In this section, the authors are going to summary Inception-v3 served as the basement for their implementation.

Basically, prominent parts of an image can have utterly large size variation. Thus, choosing the right kernel size for the
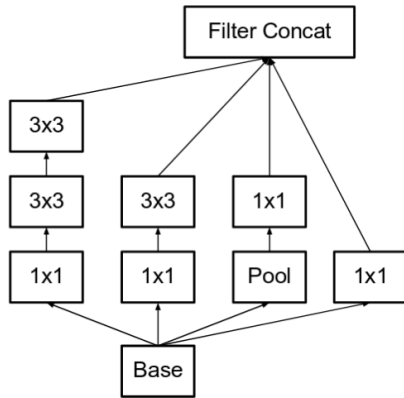
Fig. 1. Module A: Inception modules where each 5×5 convolution, the left-most convolution of the original Inception module, is replaced by two 3×3 convolution [44].
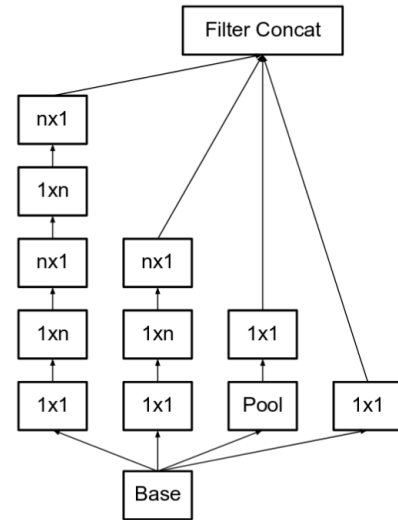

Fig. 2. Module B: Inception modules after the factorization of the $n \times n$ covolutions, where $n = 7$ [44].

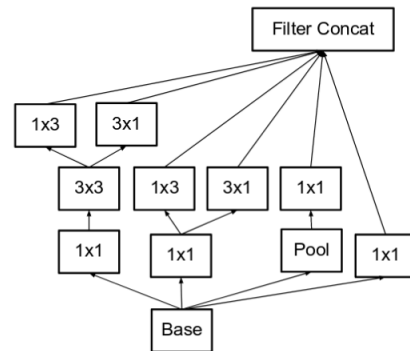Fig. 3. Module C: Inception modules with expanded the filter bank outputs [44].

convolution operation is very hard. A small kernel is favored for information that is locally distributed while a large kernel is favored for information that is globally distributed. The problem is addressed by the idea of inception [42] where filters with multiple sizes operate on the same level. As the result, the network might get wider but the computational expense is significantly reduced. Moreover, neural networks perform better when convolutions did not alter the dimensions of the input extremely. Reducing the dimensions may cause loss of useful information but improvement of computational efficiency. The balance point is known as representational bottleneck [44]. Another improvement of Inception-v3 over the original Inception may come from utilizing the idea of factorizing convolutions. The aim of factorizing convolutions is to reduce the number of connections and/or learning parameters without diminishing the network efficiency. By using appropriate factorization representation, convolutions can be made more efficient in terms of computational expensiveness and architecture complexity. For example, a 3×3 convolution is 2.78 times less computationally expensive than a 5×5 convolution, see Fig. 1. Thus, stacking two 3×3 convolutions is actually a boost in performance. Similarity, $n \times n$ convolutions are factorized into $1 \times n$ and $n \times 1$ convolutions, see Fig. 2. Last but not least, the third inception module is used for promoting high dimensional representations, see Fig. 3. To recapitulate, the outline of our adapted implementation architecture is described in Table V and Fig. 4.

B. Depthwise Separable Convolution based Model

A standard convolutional layer takes as input a $D_F \times D_F \times M$ feature map $\mathbf{F}$ and produces a $D_G \times D_G \times N$ feature map $\mathbf{G}$ where D_F is the spatial width and height of a square input feature map[1], M is the number of input channels, N is the number of output channels, D_G is the spatial width and height of an output feature map. The layer is parameterized by a convolution kernel $\mathbf{K}$ of size $D_K \times D_K \times M \times N$ where D_K is the spatial dimension of the kernel, M is the number of input channels and N is the number of output channels. The mapping from $\mathbf{F}$ to $\mathbf{G}$ is done by applying a kernel of size

---

[1]We assume that the feature map has the same spatial dimensions but it generally works with arbitrary sizes and aspect ratios.

$D_K \times D_K \times M$ on the feature map $\mathbf{F}$. The kernel captures the details not only on one channel but also the correlation to every other channel. Generally speaking, it captures the relationship inside a channel and the relationship between channels. Fig. 5 visualizes a standard convolution.

MobileNets are a class of convolutional neural network designed by researches at Google [31]. They are coined "mobile-first" [50], [51] in that they are architected from the ground up to be resource-friendly and run quickly. It is designed to effectively maximize accuracy while being mindful of the restricted resources for an on-device or embedded application. The models are effectively small, low-latency, low-power parameterized to meet the resource constraints of a variety of use cases. They can be built upon for classification, detection, embeddings and segmentation similar to how other popular large-scale models. The main difference between the MobileNet architecture and a traditional CNNs is instead of a single 3×3 convolution layer followed by batch normalization [43] and ReLU [52], MobileNets split the convolution into a 3×3 depthwise convolution (see Fig. 6) and a 1×1
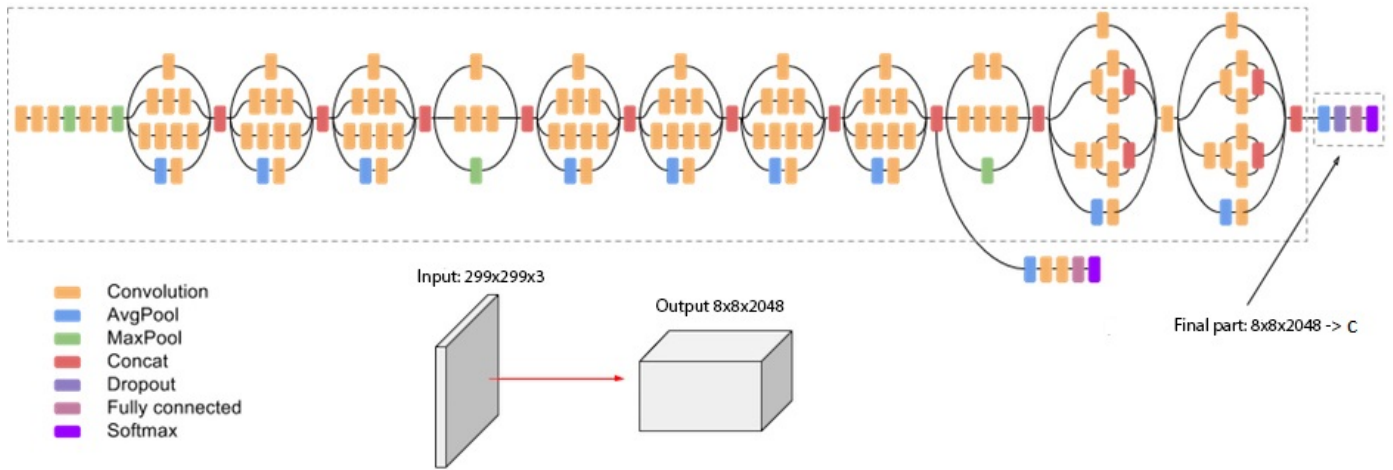
Fig. 4. The visualization of inception-based architecture used in our implementation.
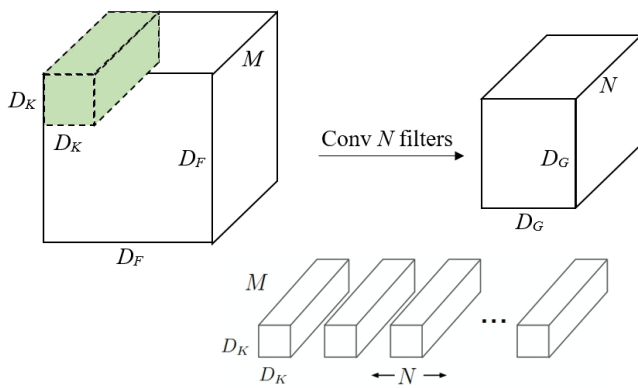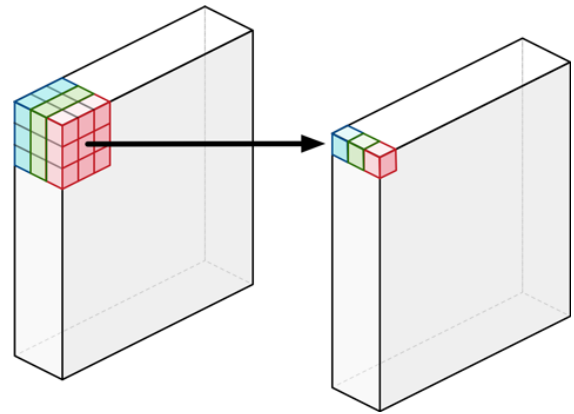

Fig. 5. A standard convolution.

Fig. 6. A depthwise convolution.

pointwise convolution (see Fig. 7)[2]. The depthwise convolution applies a single filter to each input channel while the pointwise convolution combines the outputs of the depthwise ones. In this section, we describe the model in more details.

The standard convolution operation has the effect of filtering features based on the convolutional kernels. However, in MobileNets, the filtering and combination processes are split into two separated stages by using the idea of depthwise separable convolution where the depthwise convolution and the pointwise convolution perform the filtering stage and the combination stage respectively. The depthwise convolution can be written as:

$$\hat{\mathbf{G}}_{k,l,m} = \sum_{i,j} \hat{\mathbf{K}}_{i,j,m} \cdot \mathbf{F}_{k+i-1,l+j-1,m} \qquad (1)$$

where $\hat{\mathbf{K}}$ is the depthwise convolutional kernel. The m_{th} filter in $\hat{\mathbf{K}}$ is applied to the m_{th} channel in $\mathbf{F}$ to produce the m_{th}
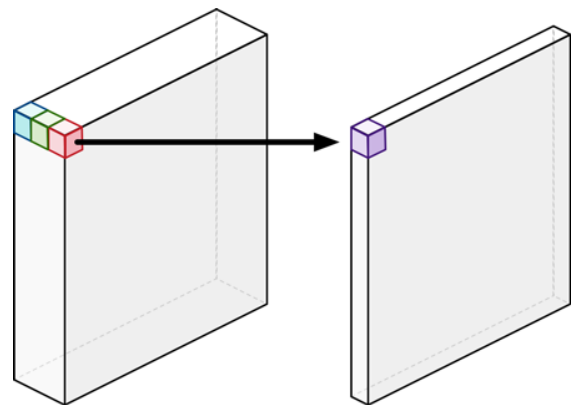

Fig. 7. A pointwise convolution.

channel in $\hat{\mathbf{G}}$. Hence, the computation cost of the depthwise convolution is the following:

$$D_F^2 \cdot D_K^2 \cdot M \qquad (2)$$

---

[2]Image courtesy to Matthijs Hollemans[54]

Because the depthwise convolution only filters input channel, we need to combine them to create new features by computing a linear combination of the output of depthwise convolution via 1×1 pointwise convolution. As the result, the computation cost of the depthwise separable convolution is the following:

$$D_F^2 \cdot D_K^2 \cdot M + D_F^2 \cdot M \cdot N \qquad (3)$$

which is the sum of the depthwise and the 1×1 pointwise convolutions. This factorization significantly reduces the computational cost [53]. More precisely, the reduction in computation by expressing convolution is the following:

$$\frac{D_F^2 \cdot D_K^2 \cdot M + D_F^2 \cdot M \cdot N}{D_F^2 \cdot D_K^2 \cdot M \cdot N} = \frac{1}{N} + \frac{1}{D_K^2} \qquad (4)$$

Although the base MobileNets architecture is already light and low latency, on-device applications may require the model to be lighter and faster. In order to build such less computationally expensive architecture, the model surfaces two hyperparameters, e.g. width multiplier and resolution multiplier, that we can tune to fit the resource and/or accuracy trade-off of our implemented model. The width multiplier allows us to thin the network, while the resolution multiplier changes the input dimensions of the image, reducing the internal representation at every layer. Given α and ρ be the width multiplier and resolution multiplier respectively. While the role of the width multiplier α is to thin a network uniformly at each layer, the role of the resolution multiplier is to reduce the resolution of the input image as well as the internal representation of every layer. The computational expense of a depthwise separable convolution with α is the following:

$$D_F^2 \cdot D_K^2 \cdot \alpha M + D_F^2 \cdot \alpha M \cdot \alpha N \qquad (5)$$

where $\alpha \in (0,1]$ with the typical settings of $\{1, 0.75, 0.5, 0.25\}$. In real world implementation, $\alpha = 1$ is the baseline MobileNets and $\alpha < 1$ is reduced MobileNets. Similarly, the computational expense for the core layers of the network as depthwise separable convolutions with width multiplier α and resolution multiplier ρ is calculated as:

$$\rho^2 D_F^2 \cdot D_K^2 \cdot \alpha M + \rho^2 D_F^2 \cdot \alpha M \cdot \alpha N \qquad (6)$$

where $\rho \in (0,1]$ with the implicit settings so that the input resolution of the network is $\{224, 192, 160, 128\}$. In real world implementation, $\rho = 1$ is the baseline MobileNets and $\rho < 1$ is reduced computation MobileNets.

Thanks to the idea of depthwise separable convolution, the network architecture is lighter, and consequently, the computation expense is significantly reduced. The reduction of computational cost and the number of parameters is quadratically by roughly α^2 and ρ^2. Given the value of $\alpha \in \{1, 0.75, 0.5, 0.25\}$ and the image resolution $\in \{224, 192, 160, 128\}$, Table I shows the comparison between a full convolution model, an inception-based model and 16 combinations of α and ρ in terms of the number of fused multiplication and addition operations, and the number of learned parameters. Readers should refer to the original papers for greater details.

TABLE I.   THE COMPARISON BETWEEN A FULL CONVOLUTION, INCEPTION-BASED MODEL AND 16 VERSIONS OF MOBILENETS

Models	Million Mult-Adds	Million Parameters
Full convolution	4866	29.3
Inception-based	5000	23.6
MobileNets 1.0-244	569	4.24
MobileNets 1.0-192	418	4.24
MobileNets 1.0-160	291	4.24
MobileNets 1.0-128	186	4.24
MobileNets 0.75-244	317	2.59
MobileNets 0.75-192	233	2.59
MobileNets 0.75-160	162	2.59
MobileNets 0.75-128	104	2.59
MobileNets 0.50-244	150	1.34
MobileNets 0.50-192	110	1.34
MobileNets 0.50-160	77	1.34
MobileNets 0.50-128	49	1.34
MobileNets 0.25-244	41	0.47
MobileNets 0.25-192	34	0.47
MobileNets 0.25-160	21	0.47
MobileNets 0.25-128	14	0.47

III.   EXPERIMENTAL SETUP

A. Dataset Collection

In this work, we have evaluated the proposed models on different data collections. The first dataset is samples of grain discoloration which is one of the most common diseases on rice in the Mekong Delta. The second dataset is a collection of medicinal plants which is essential for indigenous medical systems in Vietnam. We describe them in more details within this section.

1) Grain discoloration: Ministry of Agriculture and Rural Development of Vietnam has promulgated the resolution QCVN 01-166:2014/BNNPTNT regulating national technical regulation on surveillance method of Rice pests [41]. The resolution has described how to label grain discoloration. We randomly select a grain plant in a rice-growing field where grain discoloration disease is observed. By counting the number of affected grain of rice, called m, and the total number grain of rice, called n, on a grain plant, then the percentage, called p, of grain discoloration is calculated by the following equation: $p = \frac{m}{n}$.

Then, based on the value of p, rice experts classify the grain plant into four intensive levels of contamination of grain discoloration [41]. Table II shows the equivalence range of p and the assigned level. Level 1 is the less intensively contaminated whereas level 4 is the most intensively contaminated.

TABLE II.   DIFFERENT LEVELS OF CONTAMINATION OF GRAIN DISCOLORATION

level 1	level 2	level 3	level 4
$0.01 \leq p < 0.05$	$0.05 \leq p < 0.25$	$0.25 \leq p < 0.5$	$p \geq 0.5$

More than 1000 samples were collected from different rice growing areas of South Vietnam thanks to the help of rice experts from Cuu Long Delta Rice Research Institute as well as farm owners. Several photos taken by rice experts are presented in Fig. 8. We put a white board under the sample during taking pictures in other to isolate the background. All the collected samples were assigned to three different rice experts separately to conduct the label annotation. We keep the samples that have the same three annotated labels from three rice experts. At the final round of the collection procedure, 566 samples are
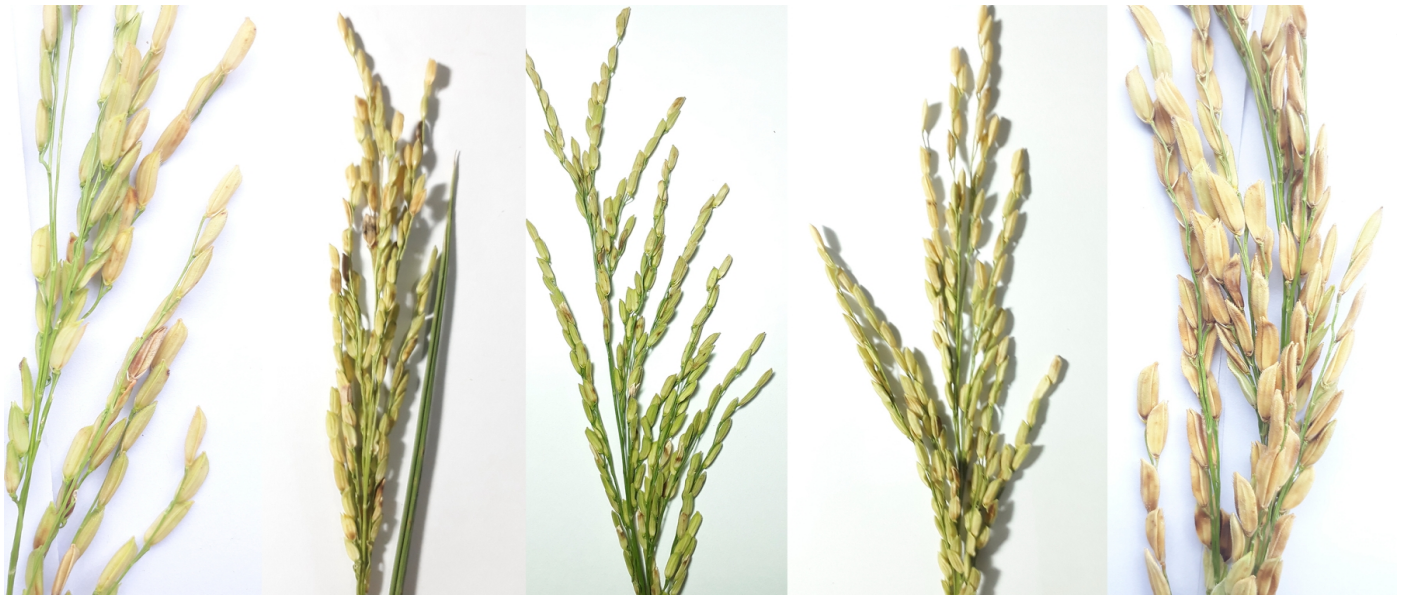
Fig. 8. Rice grain discoloration samples from major rice growing areas of South Vietnam. From left to right, the labels are level 2, level 3, level 2, level 1, and level 4 respectively. Photo by the authors (CC BY 4.0).

retained. We show their distribution into four intensive levels in Table III.

TABLE III. DATA OF GRAIN DISCOLORATION USED IN THE EXPERIMENTAL EVALUATION

	level 1	level 2	level 3	level 4	Total
Number of samples	159	156	113	138	566
Size (GB)	0.22	0.25	0.21	0.23	0.91

2) Medicinal plants: Approximately 5800 samples were collected from different growing areas of South Vietnam thanks to the help of botanic experts from botanic garden of Tay Do University as well as garden owners. Several photos taken by the authors and garden owners are presented in Fig. 9. Similar to the grain discoloration dataset, we put a white board under the sample during taking pictures in other to isolate the background. At the end of the collection procedure, 5816 samples from 20 different classes are retained. We show their distribution into classes in Table IV.

B. Implementation and Results

In our experiments, we set the required model hyper-parameters as follows. The learning rate is $\{0.01\}$. The number of epoch is $\{2000\}$. We have attempted varieties of learning rate, e.g. $\{0.1, 0.001\}$, and epoch, e.g. $\{1500, 2500, 3000\}$; however, the results are just slightly different. The model converges at around 1000th and 1500th epoch for the training and the test sets respectively. We re-train the last layer of the models by using our collected dataset. We randomly split the dataset into a training set, a validation set and a test set by the $80/10/10$ splitting schema.

The input size of the depthwise separable convolution based model is $n \times n \times 3$ for height, width and channel respectively. $n \in \{224, 192, 160, 128\}$. The input size of the inception-based model is $299 \times 299 \times 3$ by default. We also try different sizes of $\{244, 192\}$; however, the results are similar. These resolutions are common settings of running

TABLE IV. DATA OF MEDICINAL PLANTS USED IN THE EXPERIMENTAL EVALUATION

#	Label	Number of samples	Size (GB)
1	Cleistocalyx Operculatus	340	0.98
2	Polyscias Fruticosa	484	1.46
3	Premna Serratifolia L.	305	0.75
4	Gymnanthemum Amygdalinum	301	0.90
5	Cassia Alata L.	318	0.84
6	Laurales	453	0.97
7	Michelia Champaca L.	256	1.29
8	Ruellia Tuberosa L.	250	0.69
9	Durian	274	0.78
10	Ficus Racemosa	294	0.83
11	Laurus Nobilis	328	1.60
12	Schefflera Octophylla	295	1.15
13	Verbena Officinalis	300	1.90
14	Mangosteen	293	1.37
15	Phyllanthus Urinaria	315	2.08
16	Turnera Ulmifolia	355	1.78
17	Morus Alba	222	1.04
18	Pluchea Indica	248	0.16
19	Pouteria Lucuma	129	0.37
20	Zingiber Officinale	156	0.23
	Total	5816	21.17

deep convolutional networks. The classification decision is made at the softmax layer where its input is the probability distribution of investigated labels. In each combination, we cautiously re-run the model several times but the accuracy scores are unchanged. The architecture is described in Table V and Fig. 4 for the inception-based model. Whereas, the depthwise separable convolution based model is described in Table VI and Fig. 12.

Our experiments were conducted on a normal laptop Core i7-6500U with 2.5GHz clock speed, 16GB of RAM. An upper bound of RAM required for our models is 1.8GB and 1.4GB for the grain discoloration and the medicinal plants datasets respectively. The training time takes around 10 minutes and 45 minutes to complete 2000 epochs with the learning rate of 0.01 for the grain discoloration and the medicinal plants datasets respectively. The low-end GPU NVIDIA GeForce 940MX with

Fig. 9. Medicinal plant samples. From left to right and top to bottom, the labels are Cleistocalyx Operculatus, Polyscias Fruticosa, Premna Serratifolia L., Gymnanthemum Amygdalinum, Cassia Alata L., Laurales, Michelia Champaca L, Ruellia Tuberosa L., Durian, and Ficus Racemosa respectively. Photo by the authors (CC BY 4.0).

4GB of RAM is activated by default.

Our experiment scenario consists of running 16 different combinations of MobileNets plus the default version of the inception-based model as described in Table I. During the training procedure, the parameters of the last model's layer were re-weighed. Hyper-parameters space is described previously. Only the overall best performance is reported. The implementation has achieved the best classification accuracy of **98.7%** on the grain discoloration dataset and **98.5%** on the medicinal plants dataset. The least accuracy scores of models is also a good result. The complete performance on 17 implementation in the experiment is presented in Tables VII and VIII.

TABLE V. THE OUTLINE OF INCEPTION-BASED ARCHITECTURE. AT THE SOFTMAX LAYER, c IS THE NUMBER OF PREDICTED LABELS

Type	Patch size/stride	Input size
Conv	$3 \times 3/2$	$299 \times 299 \times 3$
Conv	$3 \times 3/1$	$149 \times 149 \times 32$
Conv padded	$3 \times 3/1$	$147 \times 147 \times 32$
Pool	$3 \times 3/2$	$147 \times 147 \times 64$
Conv	$3 \times 3/1$	$73 \times 73 \times 64$
Conv	$3 \times 3/2$	$71 \times 71 \times 80$
Conv	$3 \times 3/1$	$35 \times 35 \times 192$
$3\times$ Inception	Module A (see Fig. 1)	$35 \times 35 \times 288$
$4\times$ Inception	Module B (see Fig. 2)	$17 \times 17 \times 768$
$2\times$ Inception	Module C (see Fig. 3)	$8 \times 8 \times 1280$
Pool	8×8	$8 \times 8 \times 2048$
Linear	logits	$1 \times 1 \times 2048$
Softmax	classifier	$1 \times 1 \times c$

TABLE VI. THE OUTLINE OF DEPTHWISE SEPARABLE CONVOLUTION BASED ARCHITECTURE. AT THE SOFTMAX LAYER, c IS THE NUMBER OF PREDICTED LABELS

Type / Stride	Filter shape	Input size
Conv / s2	$3 \times 3 \times 3 \times 32$	$n \times n \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
$5\times$ Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times c$

C. Mobile App Deployment

Mobile communication technology has quickly become the world's most common way of sharing information and widespreading services. A mobile application providing the most affordable ways for millions of people to access information should be developed to facilitate further integrated recommendation and services. In that sense, we have deployed an

TABLE VII.   THE CLASSIFICATION ACCURACY OF EVALUATED MODELS ON GRAIN DISCOLORATION DATASET

Models	Classification accuracy
Inception-based	98.7%
MobileNets **1.0-244**	97.4%
MobileNets **1.0-192**	87.2%
MobileNets **1.0-160**	94.9%
MobileNets **1.0-128**	92.3%
MobileNets **0.75-244**	89.7%
MobileNets **0.75-192**	84.6%
MobileNets **0.75-160**	94.9%
MobileNets **0.75-128**	84.6%
MobileNets **0.50-244**	87.2%
MobileNets **0.50-192**	94.9%
MobileNets **0.50-160**	87.2%
MobileNets **0.50-128**	94.9%
MobileNets **0.25-244**	87.2%
MobileNets **0.25-192**	92.3%
MobileNets **0.25-160**	94.9%
MobileNets **0.25-128**	84.6%

TABLE VIII.   THE CLASSIFICATION ACCURACY OF EVALUATED MODELS ON MEDICINAL PLANTS DATASET

Models	Classification accuracy
Inception-based	96.2%
MobileNets **1.0-244**	98.5%
MobileNets **1.0-192**	97.7%
MobileNets **1.0-160**	98.5%
MobileNets **1.0-128**	97.4%
MobileNets **0.75-244**	97.7%
MobileNets **0.75-192**	97.9%
MobileNets **0.75-160**	97.5%
MobileNets **0.75-128**	96.7%
MobileNets **0.50-244**	97.2%
MobileNets **0.50-192**	98.0%
MobileNets **0.50-160**	97.0%
MobileNets **0.50-128**	94.9%
MobileNets **0.25-244**	97.5%
MobileNets **0.25-192**	96.0%
MobileNets **0.25-160**	96.4%
MobileNets **0.25-128**	93.2%

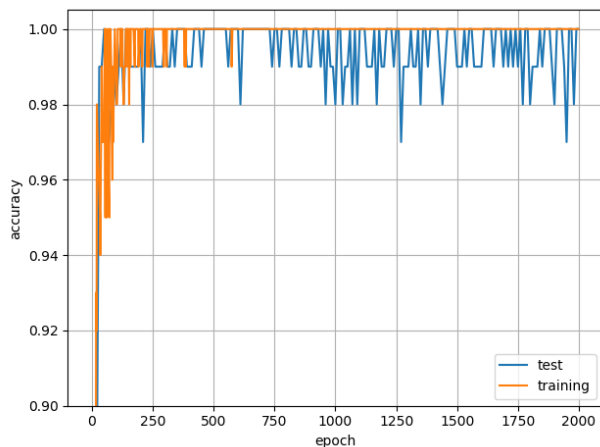Fig. 11.   The demonstration of our medicinal plant recognizer app.

Fig. 10.   The classification accuracy of the depthwise separable convolution based model in case of $\alpha = 1.0$, the input resolution 224×224, and the medicinal plant dataset.

Android application, called medicinal plant recognizer, of our experimented models especially for the task of classification of medicinal plants. After training our models, we integrate them into the mobile app that is used in two different scenarios, e.g. real-time and offline prediction. In the first scenario, a particular Android-based smartphone points at an arbitrary

medicinal plant and the prediction is made instantly. In the second scenario, a saved photo is added to the application and the prediction is made afterwards. The output is basically the probability distribution of the plant over labels. Fig. 11 shows the demonstration of our mobile application.

IV.   REMARKS AND DISCUSSION

One of the biggest advantages of the combination between the idea of transfer learning and the usage of the state-of-the-art deep convolutional neural networks is that it significantly reduces the heavy demand for the hardware infrastructure and the total training time. It helps developing countries, like Vietnam, come up with solutions in a timely and affordable manner. Instead of training a model using many high-end GPUs in week [42], [49] or even months [55], the pre-trained model is reused to straightforwardly re-weight the parameters in the last layer within 45 minutes. It reveals that the classification accuracy is very accurate. Admittedly, the models were picked in a somewhat ad hoc manner with the main constraint being that the computational complexity and rapid deployment can be made within a limitation of resources.

One of the interesting phenomena to note in Fig. 10 is that the model might be overfitting. More precisely, the model obtains **100%** accuracy on the training set early but seems to fluctuate on the test set. We have attempted several values of the learning rate and the number of epoch but facing similar behavior. The observation strongly indicates that there is a lot of room for adding a more sizable volume of data.

The experiment results in this paper have pointed out many further research directions. Firstly, we have collected several benchmark datasets of grain discoloration and medicinal plants
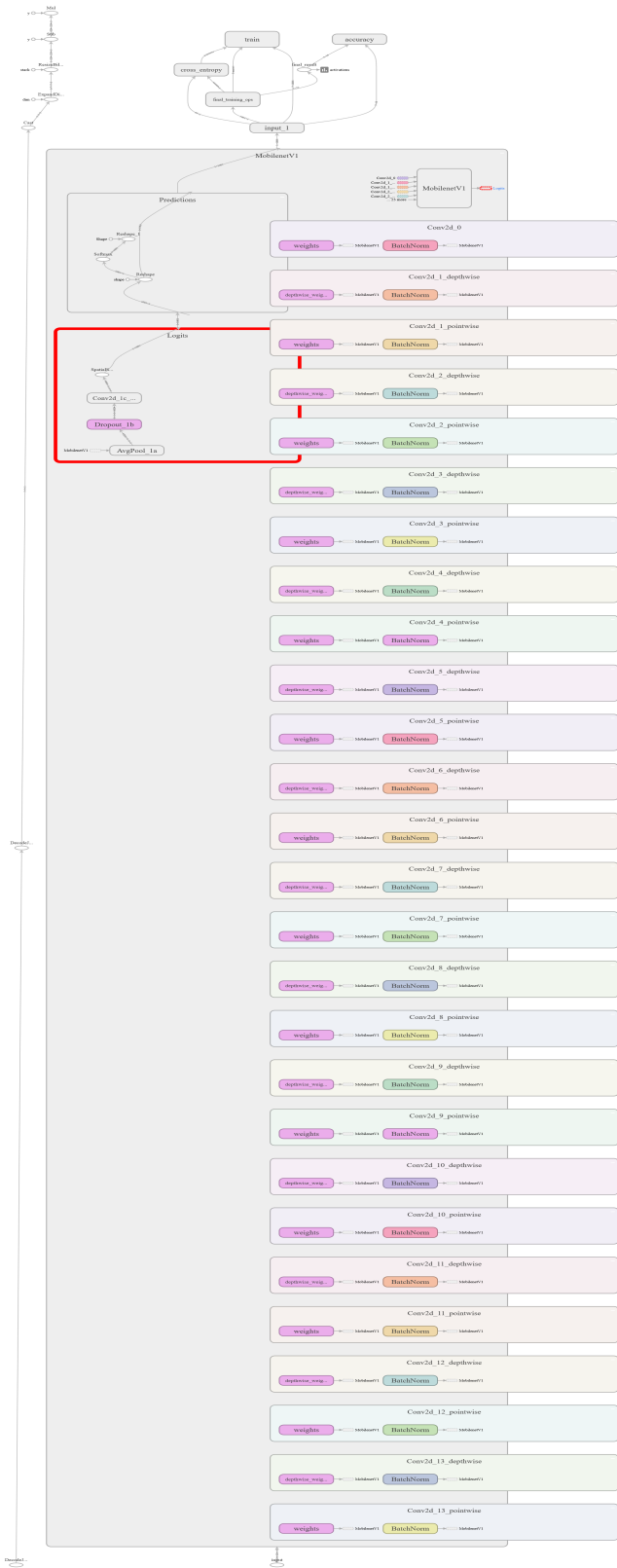
Fig. 12. The visualization of the depthwise separable convolution based architecture used in our implementation.

that serves as a preliminary preparation for accumulating

development. We aim to collect the most 70 used herbal plants described in the Decision No. 4664/QD-BYT [19] by Vietnam's Ministry of Health. Obviously, any recommendation systems should be developed upon the accurate classification results. Secondly, we have proved that the combination of deep learning via deep convolutional neural networks and the idea of knowledge transferability achieves notable results. Thirdly, we have addressed several agricultural classification problems that had been unstudied in the literature. And last but not least, we deploy the mobile version of the model to reach further users and development practitioners.

V. CONCLUSION

In this paper, we have proposed using an adapted deep learning architecture and investigated the idea of transfer learning upon a real-world classification problem. Although the experimented categories are not originally included in the ImageNet dataset, the combination works that well proving that this direction is worth investigating. The proposed transfer methodology has performed well on the unseen images of grain discoloration and medicinal plants samples. A mobile app of the best version of the depthwise separable convolution based model is also deployed. These works assist human beings in real-world classification and identification problems and are considered an essential task in agricultural research.

REFERENCES

[1] Trang, Tran Hoai Thao, and O. Napasintuwong. "Farmers' willingness-to-change and adoption of aromatic rice in Vietnam." Journal of ISSAAS (International Society for Southeast Asian Agricultural Sciences) 22.2 (2016): 50-65.

[2] World's Top Exports. "Rice Exports by Country". Available online: www.worldstopexports.com/rice-exports-country/. 2018.

[3] Muthayya, Sumithra, et al. "An overview of global rice production, supply, trade, and consumption." Annals of the new york Academy of Sciences 1324.1 (2014): 7-14.

[4] Barker, Randolph, Robert W. Herdt, and Beth Rose. The rice economy of Asia. Routledge, 2014.

[5] Luong, Van Vien. "Xac dinh nam gay benh lem lep hat tren lua tai huyen Cho Moi, Chau Thanh, Tri Ton, tinh An Giang vu thu dong 2008 (Identified fungal pathogens on wheat grain in Cho Moi District, Tri Ton, An Giang Province in 2008 autumn and winter crops)", 2014.

[6] Islam, Waqar, and Manzoor Ahmed. "Identification of different fungi associated with grain discoloration complex disease of rice and management of infested grains through fungicides." Int. J. Sci. Res. Agric. Sci 4.2 (2017): 30-35.

[7] Ashfaq, M., et al. "Grain discoloration: an emerging threat to rice crop in Pakistan." JAPS: Journal of Animal & Plant Sciences 27.3 (2017).

[8] Kanjanamaneesathian, M., and P. Meetum. "Efficacy of a simple liquid culture of Bacillus megaterium in suppressing grain discoloration disease of rice (Oryza sativa)." New Zealand Plant Protection 70 (2017): 196-202.

[9] Mamedov, N. "Medicinal plants studies: history, challenges and prospective." Med Aromat Plants 1.8 (2012): e133.

[10] Padulosi, S., D. Leaman, and P. Quek. "Challenges and opportunities in enhancing the conservation and use of medicinal and aromatic plants." Journal of Herbs, Spices & Medicinal Plants 9.4 (2002): 243-267.

[11] Farnsworth, Norman R., and D. D. Soejarto. "Global importance of medicinal plants." The conservation of medicinal plants 26 (1991): 25-51.

[12] Hamilton, Alan C. "Medicinal plants, conservation and livelihoods." Biodiversity & Conservation 13.8 (2004): 1477-1517.

[13] Okwu, D. E. "Flavouring properties of spices on cassava fufu. Afri. J. Root Tuber Crops 3: 18-20. 11. Okwu DE (2004). Phytochemicals and vitamin content of indigenous spices of South Eastern Nigeria." J. Sustain. Agric. Environ 6 (1999): 30-34.

[14] 3,780 Vietnamese medicinal plant species. Available online: http://www.intracen.org/blog/3780-Vietnamese-medicinal-plant-species/. 2016.

[15] Okwu, D. E. "Evaluation of chemical composition of indeginous species and flavouring agents." Global Journal of Pure and Applied Sciences 7.3 (2001): 455-460.

[16] Vernin, Gaston, and Cyril Parkanyi. "Chemistry of ginger." Ginger. CRC Press, 2016. 107-200.

[17] Van Duong, Nguyen. "Medicinal plants of Vietnam, Cambodia and Laos." Sl: Nguyen Van Duong 528p.-. ISBN 963730312 (1993).

[18] Do Tat Loi. "Nhung cay thuoc va vi thuoc Viet Nam (Vietnamese medicinal plants and herbs)". Available online: http://sachthucvat.blogspot.com/2017/02/2004-gsts-o-tat-loi-nhung-cay-thuoc-va.html. 2004.

[19] Ministry of Health, Viet Nam. "Quyet dinh ve viec ban hanh Bo tranh cay thuoc mau su dung trong co so kham benh, chua benh bang y hoc co truyen (Decision on the issuance of a set of model medicinal plants used in medical examination and treatment establishments by traditional medicine)". Available on-line: moh.gov.vn/LegalDoc/Lists/OperatingDocument/Attachments/389/Q\%C4\%90\%204664Q\%C4\%90-BYT.pdf. 2014.

[20] Ogle, Britta M., et al. "Food, feed or medicine: the multiple functions of edible wild plants in Vietnam." Economic Botany 57.1 (2003): 103-117.

[21] Wahlberg, Ayo. "Bio-politics and the promotion of traditional herbal medicine in Vietnam." Health: 10.2 (2006): 123-147.

[22] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. Ieee, 2009.

[23] Russakovsky, Olga, et al. "Imagenet large scale visual recognition challenge." International Journal of Computer Vision 115.3 (2015): 211-252.

[24] Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." IEEE Transactions on knowledge and data engineering 22.10 (2010): 1345-1359.

[25] Competition, ImageNet Large Scale Visual Recognition. Available online: http://www. image-net. org/challenges. 2012.

[26] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.

[27] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[28] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[29] Maggiori, Emmanuel, et al. "Convolutional neural networks for large-scale remote-sensing image classification." IEEE Transactions on Geoscience and Remote Sensing 55.2 (2017): 645-657.

[30] Abadi, Martín, et al. "Tensorflow: a system for large-scale machine learning." OSDI. Vol. 16. 2016.

[31] Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861 (2017).

[32] Torrey, Lisa, and Jude Shavlik. "Transfer learning." Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. IGI Global, 2010. 242-264.

[33] Torrey, Lisa, et al. "Transfer learning via advice taking." Advances in Machine Learning I. Springer, Berlin, Heidelberg, 2010. 147-170.

[34] Taylor, Matthew E., and Peter Stone. "Transfer learning for reinforcement learning domains: A survey." Journal of Machine Learning Research 10.Jul (2009): 1633-1685.

[35] Everingham, Mark, et al. "The pascal visual object classes (voc) challenge." International journal of computer vision 88.2 (2010): 303-338.

[36] Russakovsky, Olga, et al. "Imagenet large scale visual recognition challenge." International Journal of Computer Vision 115.3 (2015): 211-252.

[37] Yosinski, Jason, et al. "How transferable are features in deep neural networks?." Advances in neural information processing systems. 2014.

[38] Griffin, Gregory, Alex Holub, and Pietro Perona. "Caltech-256 object category dataset." (2007).

[39] Russakovsky, Olga, et al. "Detecting avocados to zucchinis: what have we done, and where are we going?." Proceedings of the IEEE International Conference on Computer Vision. 2013.

[40] Oquab, Maxime, et al. "Learning and transferring mid-level image representations using convolutional neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[41] National technical regulation on surveillance method of Rice pests. Available online: http://www.ppd.gov.vn/uploads/news/2014_06/QC% 20dich%20hai%20lua%20166.pdf. 2014.

[42] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[43] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." arXiv preprint arXiv:1502.03167 (2015).

[44] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[45] Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." AAAI. Vol. 4. 2017.

[46] Hoo-Chang, Shin, et al. "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning." IEEE transactions on medical imaging 35.5 (2016): 1285.

[47] López-Sánchez, Daniel, Angélica González Arrieta, and Juan M. Corchado. "Deep neural networks and transfer learning applied to multimedia web mining." International Symposium on Distributed Computing and Artificial Intelligence. Springer, Cham, 2017.

[48] Torralba, Antonio, Rob Fergus, and William T. Freeman. "80 million tiny images: A large data set for nonparametric object and scene recognition." IEEE transactions on pattern analysis and machine intelligence 30.11 (2008): 1958-1970.

[49] Zoph, Barret, et al. "Learning transferable architectures for scalable image recognition." arXiv preprint arXiv:1707.07012 2.6 (2017).

[50] Search Engine Land. "Google's shift to mobile-first: mobile moments that matter". Available online: https://searchengineland.com/googles-shift-mobile-first-mobile-moments-matter-263971. 2016.

[51] Forbes. "How To Prepare For The Mobile-First World Of 2018". Available online: https://www.forbes.com/sites/forbesagencycouncil/2018/01/26/how-to-prepare-for-the-mobile-first-world-of-2018/#3c7ebc753108. 2018.

[52] Krizhevsky, Alex, and Geoff Hinton. "Convolutional deep belief networks on cifar-10." Unpublished manuscript 40.7 (2010).

[53] Sifre, Laurent, and Stéphane Mallat. Rigid-motion scattering for image classification. Diss. PhD thesis, Ph. D. thesis, 2014.

[54] Matthijs Hollemans's blogs. Available online: http://machinethink.net/. 2018.

[55] Chollet, François. "Xception: Deep learning with depthwise separable convolutions." arXiv preprint (2017): 1610-02357.