Enhancing the Classification Accuracy of Rice Varieties by Using Convolutional Neural Networks

Nga Tran-Thi-Kim[1,2,3]*, Tuan Pham-Viet[4], Insoo Koo[5], Vladimir Mariano[6], and Tuan Do-Hong[1]
[1] Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology (HCMUT), Ho Chi Minh City, Vietnam
[2] Vietnam National University Ho Chi Minh City, Ho Chi Minh City, Vietnam
[3] Department NongLam University, Ho Chi Minh City, Vietnam
[4] Department University of Education, Hue University, Vietnam
[5] Department University of Ulsan, Korea
[6] YSEALI Academy at Fulbright University Vietnam, Ho Chi Minh City, Vietnam
Email: pvtuan@hueuni.edu.vn; iskoo@ulsan.ac.kr; vladimir.mariano@fulbright.edu.vn; do-hong@hcmut.edu.vn

Abstract—**The aim of this study is to enhance the classification accuracy of rice varieties that are quite similar in external observation. In this study, 17 rice grain varieties popularly planted in Vietnam are classified with an Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) models. The two CNN models (modified VGG16 and modified ResNet50) are based on pre-trained VGG16 and Resnet50 models. Two datasets are used in the experiments: a feature dataset extracted using an extended improved local ternary pattern (extended ILTP) method, and an image dataset generated with a data augmentation technique. The feature dataset was fed into the ANN, while the image dataset was fed into the CNN models. The highest classification accuracies of ANN, modified VGG16, and modified ResNet50 models are 92.82%, 96.41%, and 97.88%, respectively. The results show that the modified VGG16 and ResNet50 models significantly improved classification accuracy of the 17 varieties of rice. In addition, the experiments show that the dimensions of the image dataset can affect the performance of the CNN models. This research can be developed for applications of rice varieties classification and identification.**

***Index Terms*—artificial neural network, convolutional neural network, local binary pattern, improved local ternary pattern, rice varieties**

I. Introduction

Rice cultivars are planted in many countries and are the main food for a large segment of the world's population. Rice belongs to the genus Oryza. According to Vaughan [1], there are about 22 rice species, but only Oryza sativa and Oryza glaberrima are cultivated. To deal with the changing climate and the impacts of insects and disease, many rice varieties have been bred to deal with the challenges.

Each rice variety is usually suitable for certain growing conditions like climate, soil, and water. So selecting of a suitable variety for the growing conditions is an important part of crop. In fact, identification of rice varieties is often based on the experience of experts and farmers. However, the external appearance of rice is often similar, so it is easy to confuse one variety with another. This lead to wrong selections of rice varieties for the growing conditions, so crop yields and the quality of the harvested rice will not be high. For these reasons, it is necessary to have an automatic, quick, and highly accurate classification model for classifying of rice varieties. Besides, enhancing the accuracy of classification for rice varieties helps to reduce confusion among varieties, and this leads to increases crop yields. Therefore, in this study, we classified 17 rice varieties with a similar external appearance that are popularly planted in Vietnam.

A. Related Work

For classification of rice varieties, image processing is applied to extract the features of each one and classify them with machine learning methods such as the artificial neural network (ANN), the support vector machine (SVM), random forest (RF), etc. Besides the traditional methods, applying convolutional neural network (CNN) models to rice images without extracting features still achieves high accuracy.

In traditional methods, features of rice grains (including color, morphological, and texture) are fed into machine learning methods for varieties classification. Features used include the mean, variance, and standard deviation of components in color spaces such as RGB, HSV, and YIQ [2-7], morphological features, such as length, width, diameter, surface area, extent, compactness, and solidity [3–5, 7, 8]. Statistical approaches to texture include basic descriptors and the gray-level co-occurrence matrix (GLCM) for rice varieties classification [5–7, 9, 10]. The authors in [5] extracted four texture features (mean, standard deviation, uniformity, and third moment) computed from the grayscale intensity. These texture features along with color and morphological features gave an accuracy of 90.54% when Random Forest (RF) was used for classification of six rice varieties.

Manuscript received November 18, 2022; revised January 17, 2023; accepted February 1, 2023.

*Corresponding author: Nga Tran-Thi-Kim (email: ttknga.sdh16@hcmut.edu.vn).

doi: 10.18178/ijeetc.12.2.150-160

In [11], texture features computed from the GLCM gave higher accuracy than color and morphological features for classification of 17 rice varieties. In [6] and [9], texture features from the GLCM were used for classification of six varieties and nine varieties, respectively, at 92% accuracy. However, in some research on rice varieties classification with more classes, the accuracy achieved was not high. In [7], texture features from the GLCM along with brush ratio, color, morphological features, and Fourier descriptors obtained 89.1% accuracy for classification of 30 varieties. In [10], the authors used texture features from the GLCM for an accuracy of 87.8% from classification of 15 rice varieties.

In applications for texture classification, local binary pattern (LBP) [12] and extended LBP [13–16] were the texture features used widely because of their simple implementation, high discrimination, and high classification accuracy. To increase the discriminant in a local pattern, local ternary pattern (LTP) [17] uses three-value codes instead of the two-value codes used in the LBP. However, this led to an increase in the dimensions of the output LTP histogram. Therefore, the authors in [18] applied uniformity measure U [19] proposing an improved local ternary pattern (ILTP) for bark classification. Based on the ILTP method, the extended ILTP in [20] was proposed to classify 17 rice varieties grown in Vietnam. This proposed descriptor considered not only grayscale intensity but also the color properties of the rice grains. In this method, the ILTP was applied in R, G, B, and gray components of rice grains, and the descriptors were fed into an SVM, achieving classification accuracy of 95.53% for the 17 rice varieties.

After extracting features of the rice grains, they have been used in classifiers such as the SVM, a neural network (NN), RF, and k-nearest neighbors (KNN) for classification of the varieties. In research on rice varieties classification, the neural network has been the machine learning method used most popularly [2–4, 6, 10]. The authors in [2] used color features with an ANN for classification of 15 rice varieties, obtaining average accuracy of 94.33%. The ANN model was constructed with 7 neurons in the input layer and 15 neurons in the output layer. The network applied a backpropagation algorithm for training, with the termination error at 0.01, the learning rate at 0.05, and the momentum coefficient at 0.6.

The authors in [4] classified five rice varieties planted in Iran. Thirty-nine features of the rice varieties were used with multilayer perceptron (MLP) and neuro-fuzzy neural networks. The MLP model included an input layer, an output layer, and two hidden layers. The topological structure of the neuro-fuzzy model contained an input layer and an output layer with 60 rules. The highest classification accuracies from the MLP and the neuro-fuzzy network were 98.40% and 99.73%, respectively. In [6], the authors combined MLP models with feature sets for classification of nine rice varieties. MLP models have four layers (an input layer, two hidden layers, and the output layer). The number of neurons in the hidden layers was computed based on the number of attributes in each feature set. The highest overall accuracy achieved was 92% from a model combining all feature sets. The authors in [10] used the ANN for rice varieties classification based on their edge images. An ANN model was constructed with an input layer, a hidden layer, and an output layer, as done in research by the authors [2]. The average accuracy from the ANN was 87.80% for classification of 15 rice varieties.

Recently, the CNN has been an effective model for the classification and recognition problem [21-24]. Different from the traditional methods, input images were fed directly into the CNN without implementing feature extraction [25]. Some CNN models are often applied for image classification such as LeNet-5, AlexNet, ZFNet, VGGNet, GoogLeNet, ResNet, DenseNet. For rice varieties, authors have applied CNNs to enhance classification accuracy of varieties of rice grains or rice kernels [26–29]. Fourteen rice varieties in Thailand were classified using five CNN models (VGG16, VGG19, Xception, InceptionV3, and Inception ResNetV2) [27]. To compare classification performance with statistical machine learning methods, LR, LDA, k-NN, and SVM were applied to features extracted from rice grains, including shape, color, and texture. Among the CNN algorithms, an Inception ResNetV2 model achieved the highest accuracy at 95.15%, while the highest accuracy from statistical machine learning methods was 90.61% by an SVM.

Deep learning algorithms such as VGG16, VGG19, and DenseNet121 and an SVM classifier combined with hand-crafted features and were used to classify six rice varieties in Vietnam [28]. The results showed that classification accuracy from a CNN model was enhanced significantly, compared to the SVM. The highest accuracy was 99.04% with DenseNet121, higher than other CNN models and the SVM. ANN, DNN, and VGG16 models were applied to classify five rice kernel varieties [30]. Accuracy was 99.87%, 99.95%, and 100% from the ANN, the DNN, and VGG16, respectively. Another approach to rice classification was reported in research that combined a CNN model with hyperspectral imaging. Hyperspectral images were used with a VGGNet architecture to classify four rice varieties [26]. The highest accuracy from VGGNet was 87%, higher than from an SVM at 84%. Research used hyperspectral imaging with a CNN to obtain 100% accuracy when classifying seven varieties of rice [29].

Among the CNN models, the VGG16 model is used commonly due to the simple network architecture with 13 convolutional layers. In addition, the kernel size (3×3) is small so that it is possible to reduce the parameters of the model but still achieve high effectively. With the ResNet50 model, using shortcuts from the front layer to the next layer to overcome the vanishing gradient problem during the training process [31]. Therefore, two modified models from VGG16 and ResNet50 are selected to apply in our study for rice varieties classification.

From the literature, research in classification of rice varieties achieved accuracy of 90.54% with six varieties grown in Vietnam [5], 92% with nine varieties grown in Sri Lanka [6], and 92% with six varieties grown in Odisha, India [9]. With more varieties to classify, accuracy achieved was not high: 87.8% with 15 varieties [10], and 89.1% with 30 varieties [7]. Aside from ANNs, CNN models such as VGG16, ResNet, and DenseNet121 can improve classification accuracy of rice varieties. However, the number of classes in that research was still too few.

B. Main Contributions

In our study, 17 rice varieties were classified with an ANN and modified VGG16 and ResNet50 models. An extended ILTP feature set was fed into the ANN, while an image dataset was fed into the modified VGG16 and ResNet50. The overall accuracy and the confusion matrixes of the models were computed to evaluate and compare the performance of the classifiers.

The main contributions are summarized as follows.

- We built large datasets of 17 varieties of rice grown in Vietnam, including a feature dataset and an image dataset. The feature dataset was generated from the extended ILTP method, while data augument technique was applied to improve the dimension of the image dataset.
- We proposed using an ANN model and two modified CNN models based on the pre-trained VGG16 and ResNet50 models for classification of the 17 rice varieties. In these two modified CNN models, a new FC layer with was used to replace the old FC layers of VGG16 and ResNet50.
- Experiments were conducted to evaluate and compare the performances from these three models. Two CNN models including modified VGG16 and modified ResNet5 showed results superior to those from the ANN. Therefore, these proposed modified models can enhance significantly the classification accuracy of 17 rice varieties.

The rest of this paper is organized as follows. Material and methods are presented in Section II. Results and discussion are in Section III, followed by conclusion in Section IV.

II. MATERIAL AND METHODS

A. Rice Varieties

For the objective was to classify of rice varieties that have been easy to confuse each other, we selected 17 varieties that are quite similar in external observation, especially in color as shown in Fig. 1. Seventeen rice varieties commonly planted in Vietnam were used in this study include DT8, HT1, IR4625, IR50404, IR6976, ML48, MO6162, OM108, OM3673, OM4218, OM429, OM4900, OM5451, OM6976, OM8108, OMCS2012, and RVT. In our study, the rice grain images were scanned at a resolution of 2400 dpi. There were five rice grains in each image. Each variety was represented by 200 images, from which 100 images were used for training, with the other 100 for testing. The images for training and testing were fixed for all experiments in this study.

B. Datasets

There were two datasets used for classification of the 17 rice varieties: a feature dataset and an image dataset.

Fig. 1. The 17 rice varieties commonly planted in Vietnam and used in this study.

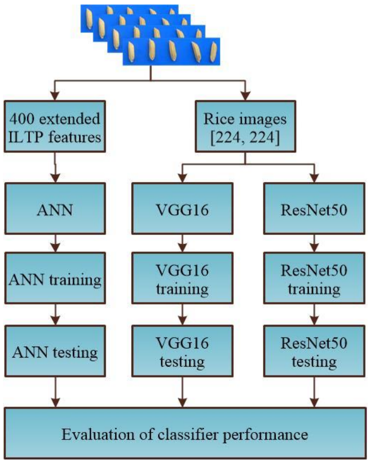


Fig. 2. Framework for classification of 17 rice varieties.

(a) (b) (c)

Fig. 3. Rice image pre-processing: (a) one rice image, (b) a rice image extracted from the blue background, and (c) a rice image after rotating each grain in the same direction while being kept straight.

The feature dataset was fed into the ANN model, whereas the image dataset was fed directly into the modified VGG16 and ResNet50. The framework of our work is presented in Fig. 2. Before extracting features and generating the image dataset, the original images were pre-processed. With each image, rice grains were rotated in the same direction and kept straight, as seen in Fig. 3.

1) Feature dataset

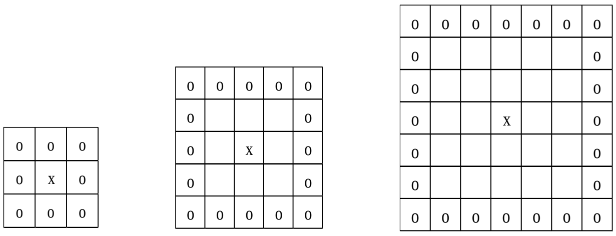
Extended ILTP feature [20] inherited the advantages of the LTP and ILTP, including being invariant to rotation, having high discriminative features, and a decreases in the dimension of the feature set. Therefore, the feature dataset was extracted via the extended ILTP method for the 17 varieties of rice used in this study. Extended ILTP features were extracted based on the ILTP [18] but included color properties. Therefore, each rice grain image was converted into grayscale, red (R), green (G), and blue (B) components. Then, ILTP features were extracted for each component.

We considered a local region with radius R and P neighbors (as seen in Fig. 4) for image component I^k in which k = 0, 1, 2, 3 (corresponding to grayscale, R, G, B image components). LTP was computed as in Eq. (1) and Eq. (2):

$$\mathrm{LTP}_{P,R} = \sum_{p=0}^{P-1} s_{\mathrm{LTP}}(g_p - g_c)3^p \tag{1}$$

$$s_{\mathrm{LTP}}(x) = \begin{cases} 1 & \text{if} \quad x \ge t \\ 0 & \text{if} \quad |x| < t \\ -1 & \text{if} \quad x \le -t \end{cases} \tag{2}$$

where g_c and g_p are the values of the central pixel and its neighbors, respectively; $s_{\mathrm{LTP}}(x)$ is the threshold function; and t is the threshold value selected by the user.



(a) b) (c)

Fig. 4. Example of local patterns with different values of P and R: (a) P = 8, R = 1, (b) P =16, R = 2, and (c) P = 24, R = 3.

For simplicity, the pattern will be categoried into the upper and lower pattern. Let $\mathrm{ULTP}_{P,R}$ and $\mathrm{LLTP}_{P,R}$ are upper binary code and lower binary code, respectively. Therefore, $\mathrm{ULTP}_{P,R} = (\mathrm{Ub}_{P,R}(0), \mathrm{Ub}_{P,R}(1), \cdots, \mathrm{Ub}_{P,R}(P\text{-}1))$, and $\mathrm{LLTP}_{P,R} = (\mathrm{Lb}_{P,R}(0), \mathrm{Lb}_{P,R}(1), \cdots, \mathrm{Lb}_{P,R}(P\text{-}1))$, where $\mathrm{Ub}_{P,R}(h)$ and $\mathrm{Lb}_{P,R}(h)$ denote the bit value at neighbor h of upper and lower binary code, respectively, with h = 0, 1, 2, ···, P–1. In this study, the binary code was started at row 2 column 1 of the pattern (h=0). $\mathrm{Ub}_{P,R}(h)$ and $\mathrm{Lb}_{P,R}(h)$ were calculated as in Eqs. (3), (4), (5), and (6):

$$\mathrm{Ub}_{P,R}(h) = s_U(g_h - g_c) \tag{3}$$

$$s_U(x) = \begin{cases} 1 & \text{if} \quad x \ge t \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$\mathrm{Lb}_{P,R}(h) = s_L(g_h - g_c) \tag{5}$$

$$s_L(x) = \begin{cases} 1 & \text{if} \quad x \le -t \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $s_L(x)$ and $s_U(x)$ are the threshold functions for lower and upper patterns, respectively.

Suppose that MIN and MAX are the minimum and maximum value of all pixels of the image. So if we choose (t >g_c–MIN) or (t >MAX–g_c) then lower binary code or upper binary code are all equal 00000000. Therefore, the limit for the value of t should be: ($t \le g_c$–MIN) and ($t \le$ MAX–g_c).

To decrease the computational complexity, uniformity measure (U) [19] denoted the times of changing from 1 to 0 (or vice versa) in the binary code, and was calculated with Eq. (7) and Eq. (8) for upper and lower binary codes, respectively:

$$U(\mathrm{ULTP}_{P,R}) = \left|\mathrm{Ub}_{P,R}(P-1) - \mathrm{Ub}_{P,R}(0)\right| + \sum_{h=1}^{P-1}\left|\mathrm{Ub}_{P,R}(h) - \mathrm{Ub}_{P,R}(h-1)\right| \tag{7}$$

$$U(\mathrm{LLTP}_{P,R}) = \left|\mathrm{Lb}_{P,R}(P-1) - \mathrm{Lb}_{P,R}(0)\right| + \sum_{h=1}^{P-1}\left|\mathrm{Lb}_{P,R}(h) - \mathrm{Lb}_{P,R}(h-1)\right| \tag{8}$$

Therefore, U was not changed when the binary code was calculated at different starting positions in a pattern. So this method was invariant to image rotation.

Relying on the U values, central pixel of upper pattern and lower patterns were assigned values of 0 to P+1. Let $U\_I^k_{P,R}(c)$ and $L\_I^k_{P,R}(c)$ respectively, denote central pixel of upper pattern and lower pattern for image component k. They were labelled as seen in Eq. (9) and Eq. (10), where U_{UT} and U_{LT} are threshold U values of the upper and lower patterns, respectively.

$$U\_I^k_{P,R}(c) = \begin{cases} \sum_{h=0}^{P-1} \mathrm{Ub}_{P,R}(h) & \text{if } U(\mathrm{ULTP}_{P,R}) \le U_{\mathrm{UT}} \\ P+1 & \text{otherwise} \end{cases} \tag{9}$$

$$L\_I^k_{P,R}(c) = \begin{cases} \sum_{h=0}^{P-1} \mathrm{Lb}_{P,R}(h) & \text{if } U(\mathrm{LLTP}_{P,R}) \le U_{\mathrm{LT}} \\ P+1 & \text{otherwise} \end{cases} \tag{10}$$

These operators were implemented for all local patterns of the component image, and then, the labels occurrence probabilities were constructed to form the component histogram. At the end of the process, all histograms were concatenated from the output histograms for gray, R, G, and B components. The output feature of extended ILTP method was the concatenated histogram. Due to the uniformity measure (U), the label of upper or lower patterns was obtained values in range of 0 and (P+1). So the number of labels (bins of historgram) assigned for each component image equal 2(P+2), and 8(P+2) for rice image (includes 4 component images). Therefore, the uniformity measure U helps decrease the number of bins in the output historgram (or output feature vector). With each rice sample of five grains, the extracted features were computed from the averages of the five grains.

The procedure for calculating the ILTP for each image component is presented in Fig. 5. In this example, $U(\mathrm{ULTP}_{8,1})$ equals 2 (=U_{UT}) and $U(\mathrm{LLTP}_{8,1})$ equals 4 (>U_{LT}). Therefore, the label of the upper and lower patterns were assigned values 1 and 9, respectively.

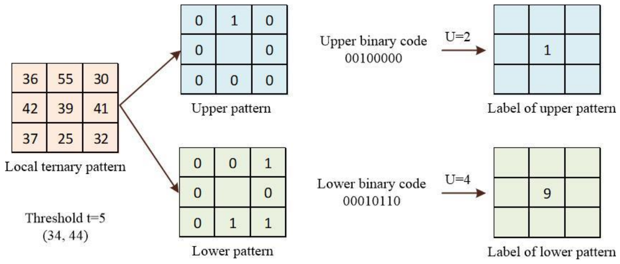


Fig. 5. An example calculation of the ILTP pattern (P=8, R=1) when $U_{\text{UT}} = U_{\text{LT}} = 2$.

Referring to [20], threshold values t equaled 9, and the radius of the neighborhood R was 6 (P = 48), which gave the highest effective classification for the 17 rice varieties. Therefore, a feature dataset containing 400 attributes ($8(P+2)$) was generated for this study.

2) Image dataset

The original image dataset for training contained 100 rice images. To increase the efficiency in the training process, a data augmentation technique was applied to generate more rice images for the training set. In addition, this technique helped to avoid the overfitting problem when learning samples of the dataset [31]. Because rice grains were straightened and rotated in the same direction, new samples in this study were generated by changing the order of the grains in the images, as presented in Fig. 6. After generating more samples, the image dataset contained seven image sets with different dimensions, as listed in Table I.

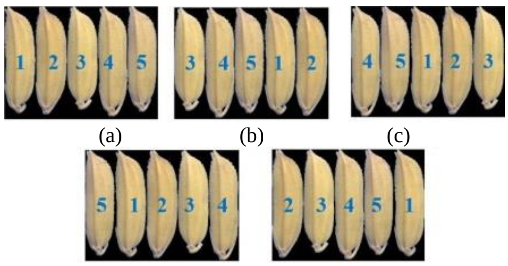


Fig. 6. More rice images were generated by changing the order of the grains in each image: (a) 1, 2, 3, 4, 5, (b) 3, 4, 5, 2, 1, (c) 4, 5, 1, 2, 3, (d) 5, 1, 2, 3, 4, and (e) 2, 3, 4, 5, 1.

TABLE I: THE NUMBER OF IMAGES PER CLASS IN THE IMAGE DATASET

Image Dataset	The number of images per class	
	Training	Testing
Set 1	750	100
Set 2	1000	100
Set 3	1250	100
Set 4	1500	100
Set 5	1750	100
Set 6	2000	100
Set 7	2250	100

C. Models for Classification of Rice Varieties

1) ANN

An artificial neural network consists of many units (also called perceptrons or neurons) that are organized in different ways to form the structure of the network. A unit receives a real value input, processes the input, and delivers a single output. The input can be new data or the output from other neurons. Similarity, the output of units can be final output or can be the input for other neurons [32]. The structure of perceptron j with input vector $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ is illustrated in Fig. 7.

In Fig. 7, $\mathbf{w} = (w_{j1}, w_{j2}, \cdots, w_{jn})$ denotes weight vector on connections from the input to neuron j, θ_j is the bias of neuron j, $\sum$ is a combination function, $f(a_j)$ is an activation function, and o_j is the output of the perceptron. In addition, a_j is computed with Eq. (11).

$$a_j = \sum_{i=1}^{n} w_{ji} x_i + \theta_j \tag{11}$$

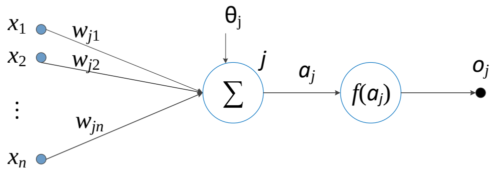


Fig. 7. The structure of perceptron j.

which can be rewritten as follows

$$a_j = \mathbf{wx} + \theta_j \tag{12}$$

Then, a_j is transferred to the activation function to form the output of the perceptron. Therefore, a unit is a transfer function of input $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ to generate o_j, as seen in Eq. (13).

$$o_j = f(\mathbf{wx} + \theta_j) \tag{13}$$

Some activation functions are used such as the linear function, the sigmoid function, the binary step function, the hyperbolic tangent function (Tanh), and the rectified linear unit (ReLU) activation function. In this study, the ReLU is used in the ANN model.

An ANN is formed from a collection of many neurons that are grouped into layers: an input layer, hidden layers, and an output layer. In this proposed model, there are 400 neurons corresponding to the number of features in the input layer, a hidden layer, and an output layer with 17 neurons corresponding to the 17 rice varieties. Referring to [33], the number of neurons in the hidden layer was computed as seen in Eq. (14):

$$N = \frac{I + O}{2} + \sqrt{Y} \tag{14}$$

where I is the neurons of the input layer, O is the neurons of the output layer, and Y is the number samples in the training set.

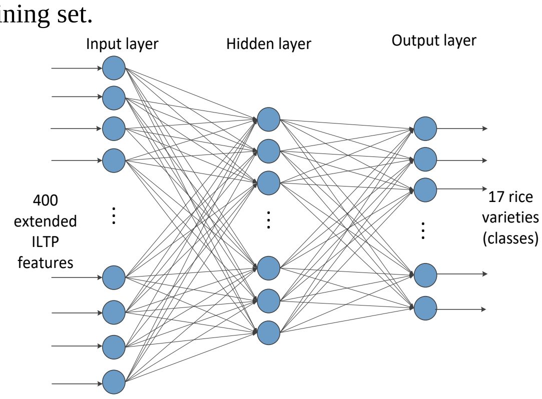


Fig. 8. The architecture of the ANN model in this study.

Therefore, the number of neurons in the hidden layer is 249. The structure of the ANN in our study is presented in Fig. 8.

2) CNN

A CNN is built with a deep learning algorithm that is effective in a lot of applications, especially those related to image classification, image semantic segmentation, object detection in images, etc. [34]. Instead of using extracted features of objects, as done in other traditional classifiers, the CNN is fed directly by input images. Then extracting features, learning and classifying are implemented throughout the layers within it. A CNN is structured with multiple basic layers: the convolutional layer, the activation layer (or non-linearity layer), the pooling layer, and the fully connected layer [35].

The convolutional layer is the most important part of a convolutional network. It consists of several convolutional kernels (or filters), which extract features of the input to create an output feature map. The feature map is generated by convolving the input and the kernels, and a non-linear activation function is applied to this convolved result [36]. The depth of a convolutional layer is determined by the number of filters used [37].

The purpose of the pooling layer is to reduce the dimension of the feature maps but still keep important content [23]. In this pooling layer, each non-overlapping region of the feature map is sampled to form the output. Some pooling techniques can be applied, such as max pooling, min pooling, average pooling, etc., among which max pooling is the most common. Therefore, the number of calculated parameters is reduced significantly [36].

After an image is processed with several convolutional layers and pooling layers, one or more Fully Connected (FC) layers are added as the last part of the CNN. For image classification, the fully connected layer plays the role of a classification layer. The fully connected layer takes input from the last pooling layer or convolutional layer in the form of feature maps. These feature maps are flattened to form a vector, and are then fed into the FC layer to get the final output from the CNN [30, 36].

In the CNN model, the activation layer is applied after each learnable layer, such as the convolutional and FC layers. The activation function is used to map the input to output within a certain range. Some activation functions are commonly used, such as sigmoid, tanh, ReLU, Leaky ReLU, etc. [30, 36].

For rice varieties classification, VGGNet and ResNet models gave highly accurate results in other research [26, 27, 30, 38]. Therefore, two CNN models applied in our study for classification of the 17 rice varieties were based on the VGG16 and ResNet50 models. In our proposed models, a transfer learning technique is applied to inherit the pre-trained VGG16 and ResNet50 models. Transfer learning technique was applied on pre-trained VGG16 and pre-trained ResNet50 models of ImageNet dataset. Therefore, after features extracted from convolutional part of pre-trained models, these extraction features become input of new CNN. So weights of modified models obtained from training with imageNet dataset were updated with new CNN.

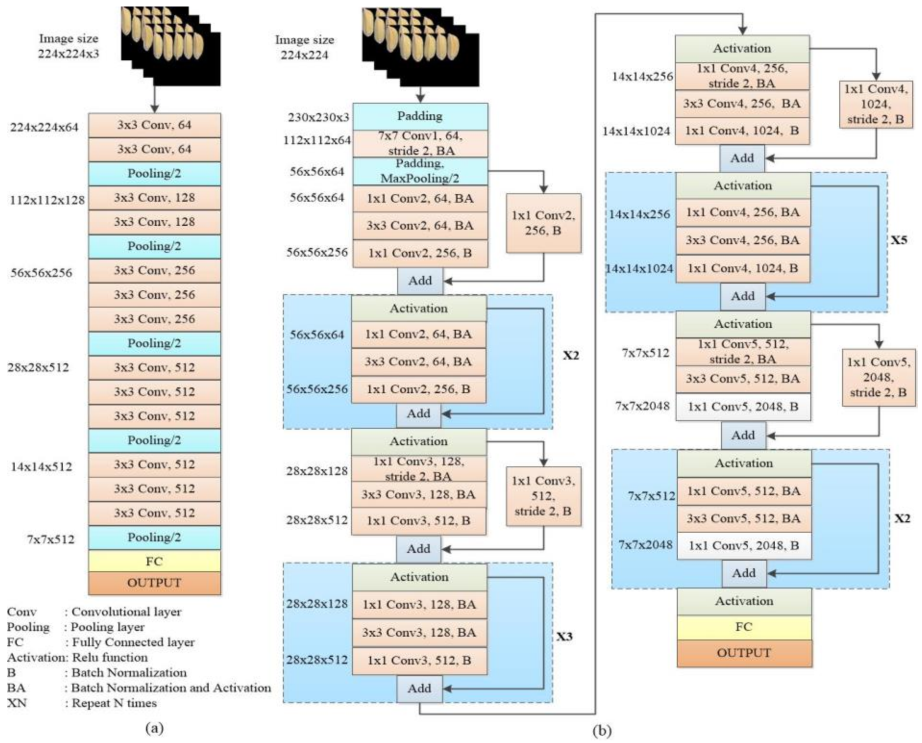


Fig. 9. Architecture of the CNN models: (a) the modified VGG16, and (b) the modified ResNet50

In these modified models, we removed old FC layers from traditional VGG16 and ResNet50, and then added a newly FC layer. Therefore, instead of using FC layers in the VGG16 and ResNet50 models, the output from the pre-trained models is flattened to feed a newly added FC layer. The new FC layer has 256 neurons with a ReLU activation function. At the end of the process, an output layer was added with 17 neurons and a Softmax activation function to label the 17 varieties of rice. The architectures of the modified VGG16 and ResNet50 models are presented in Fig. 9. In the proposed models, the numbers 64, 128, 256, 512, 1024, and 2048 represent the kernels used in the convolutional layers. In the ResNet50 model, using a shortcut connection from previous layers into the following layers is the importance difference compared to other CNN models. This can overcome the vanishing gradient problem in the training of deep networks [31].

TABLE II: HARDWARE SPECIFICATIONS

Hardware	Specifications
CPU	Intel Core i7-6700HQ 2.6 GHz
RAM	16 GB
Operating System	Windows 10
Programming language for image processing	Visual Studio C++ combined with Open Source Computer Vision Library
Framework	Keras
Programming language	Python 3.6

D. Hardware Specifications

In this study, experiments were implemented with the hardware specifications in Table II.

E. Evaluation of Classification Performance

1) Classification accuracy

Classification accuracy (ACC) was measured on a test set to evaluate the performance of the classification models. It is computed with (15):

$$ACC = \frac{N_d}{N_t} 100\% \qquad (15)$$

where N_d is the number of test samples that are labeled correctly, and N_t is the total number of samples belonging to the test set.

2) Confusion Matrix

The confusion matrix is applied to describe the number of samples labeled correctly in a class, and the number of samples mislabeled as other classes. An example of two-class confusion matrix was applied to evaluate classification performance, as seen in Table III, where L_{ij} is the number of samples in class C_i that is predicted as class C_j.

TABLE III: CONFUSION MATRIX

		Predicted class	
		C_1	C_2
Actual class	C_1	L_{11}	L_{12}
	C_2	L_{21}	L_{22}

III. RESULTS AND DISCUSSION

A. Classification with ANN

The extended ILTP feature set was fed into the ANN for classification of the 17 rice varieties. The feature dataset containing 3400 samples was divided into 50% for training and 50% for testing. The batch size was 4 and learning rate was 0.001, chosen by trial-and-error. The experiment was conducted for 1000 epochs. Classification accuracy when changing the number of epochs is presented in Fig. 10. As in Fig. 10, the classification accuracy was significant decreased at 600 epochs. It can be seen that the model has not been stable. Therefore, the number of epochs should be increase for the better performance of the model. From this result, the highest accuracy achieved was 92.82% when training for 1000 epochs. The resulting confusion matrix is presented in Fig. 11. From the confusion matrix, all testing samples of the IR4625, ML48, OM108, OM429, and RVT varieties were labeled correctly. However, the number of OM8108 samples with correct labels was quite low (54 out of 100 testing samples).

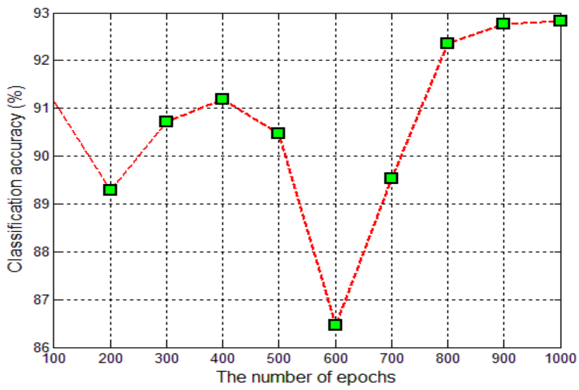


Fig. 10. Classification accuracy of the 17 rice varieties when changing the number of epochs.

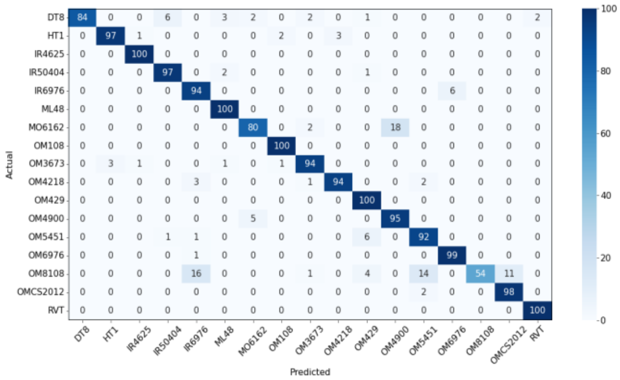


Fig. 11. The confusion matrix for the ANN model after classification of the 17 rice varieties.

B. Classification with CNN

In the image dataset, there were seven image sets with different dimensions, as seen in Table I. Each image set was fed into the modified VGG16 and modified ResNet50 models. The batch size was 64 and learning rate was 0.001, were chosen by trial-and-error method. Classification accuracies from these two models for the 17 rice varieties are presented in Table IV when the number of epochs was 50 for both models.

In Table IV, the performance of modified VGG16 was quite effective with all image sets. For modified ResNet50, its performance was affected significantly when changing of dimensions of image sets. The classification accuracy of modified ResNet50 was low when the dimensions of training image sets were small or very large. In specially, modified ResNet50 combined with Set 1 was obtained the accuracy at 46.94%, lower than modified VGG16. As seen in Table IV, classification accuracy from the modified VGG16 was the highest at 96.41% for Set 3, and this rate decreased with sets 4 to 7. From the modified ResNet50, classification accuracy was higher than from the modified VGG16 for sets 4 to 6. The highest accuracy from the modified ResNet50 model was 97.88% with Set 4. From this result, Set 3 with 1250 training images per class was chosen for the modified VGG16 model.

TABLE IV: CLASSIFICATION ACCURACY OF CNN MODELS WHEN COMBINED WITH IMAGE SETS

Image dataset	ACC (%)	
	Modified VGG16	Modified ResNet50
Set 1	95.71	46.94
Set 2	95.71	69.76
Set 3	**96.41**	95.94
Set 4	96.24	**97.88**
Set 5	96.12	97.76
Set 6	93.41	97.76
Set 7	94.65	77.94

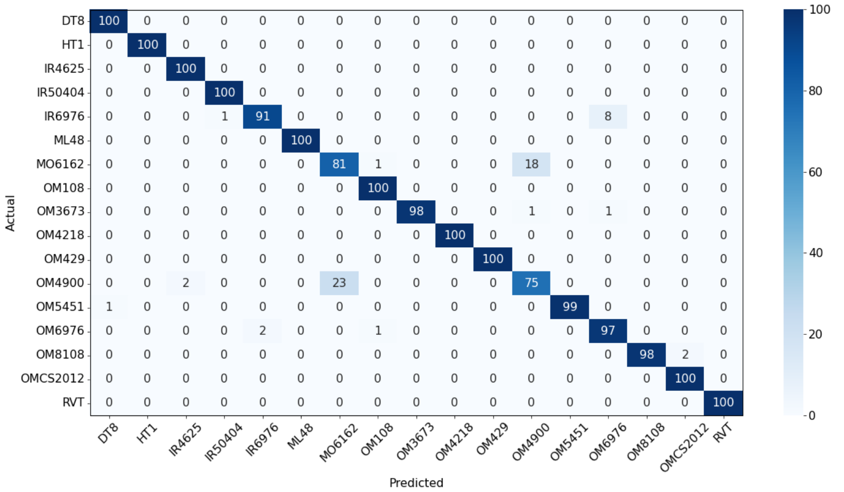


Fig. 12. The confusion matrix from the modified VGG16 model after classification of the 17 rice varieties.

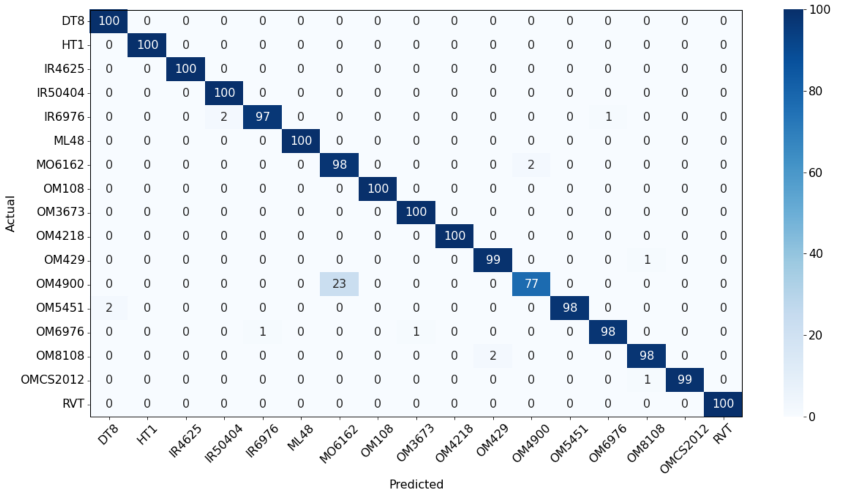


Fig. 13. The confusion matrix from the modified ResNet50 model after classification of the 17 rice varieties.

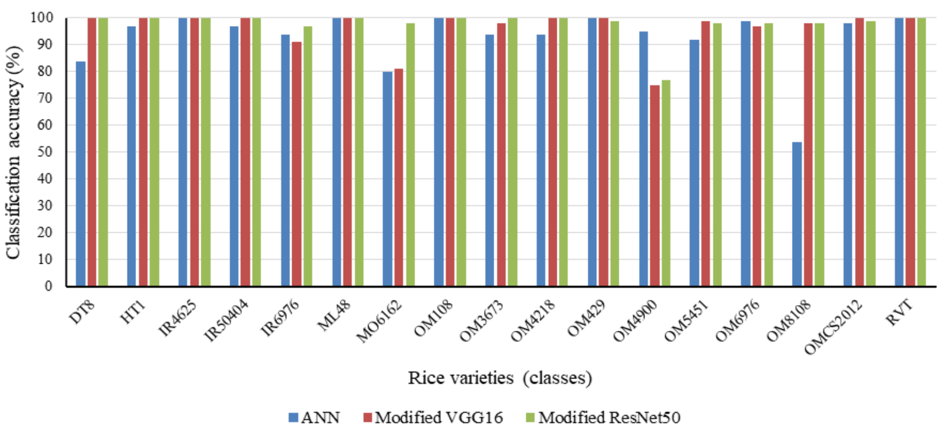


Fig. 14. Comparison of classification accuracies of the classes from the ANN, the modified VGG16 and ResNet50 models.

The confusion matrix of the modified VGG16 for classification of the 17 rice varieties with Set 3 is presented in Fig. 12. For the modified ResNet50, Set 4 with 1500 training images per class was chosen to classify the 17 rice varieties, and the confusion matrix is in Fig. 13.

In the confusion matrix from the modified VGG16 model, we can see that some varieties (including DT8, HT1, IR6425, IR50404, ML48, OM108, OM4218, OM429, OMCS2012, and RVT) were predicted correctly for all images in the testing set. A lot of MO6162 and OM4900 testing images were mislabeled as other classes (19 and 25 images, respectively). In the confusion matrix from the modified ResNet50 model, all test images from some classes (including DT8, HT1, IR4625, IR50404, ML48, OM108, OM3673, OM4218, and RVT) were labeled correctly. There was only one class (OM4900) that had a low correct prediction rate (77 images).

C. Classification Performance Comparison

The best classification accuracy by the ANN, modified VGG16, and modified ResNet50 models in the experiments from this study were 92.82%, 96.41%, and 97.88%, respectively, as presented in Table V.

TABLE V: COMPARISON OF CLASSIFICATION ACCURACIES FOR THE ANN AND THE MODIFIED VGG16 AND MODIFIED RESNET50 MODELS

Model	ACC (%)
ANN	92.82
Modified VGG16	96.41
Modified ResNet50	97.88

From the experiments, we can see that classification accuracy of the 17 rice varieties by the modified ResNet50 model was higher than the ANN and the modified VGG16 model. A comparison of classification accuracies for each class in these three models is illustrated in Fig. 14. From results in Fig. 14, classification accuracy of the classes by the CNN models (especially modified ResNet50) were higher than from the ANN. However, there was only one class (OM4900) for which this rate was not improved, classification accuracies were 95%, 75%, and 77% with the ANN and the modified VGG16 and modified ResNet50 models, respectively.

TABLE VI: RICE VARIETY CLASSIFICATION IN THE LITERATURE

Research	Year	Models	Number of classes	ACC (%)
Silva and Sonnadara [6]	2013	Neural network	9	92
Anami et al. [2]	2015	Neural network	15	94.33
Hong et al. [5]	2015	Random forest	6	90.54
Kuo et al. [7]	2016	Sparse-representation-based classification	30	89.1
Wu et al. [29]	2018	Deep convolutional neural network	7	100
Duc et al. [28]	2020	DenseNet121	6	99.04
Singh and Chaudhury [39]	2020	Neural network	8	99.63
Kiratiratanapruk et al. [27]	2020	InceptionResNetV2	14	95.15
Nga et al. [11]	2021	SVM combined with birnary particles swarm optimization	17	93.94
Nga et al. [20]	2022	SVM	17	95.53
Nga *et al*. (this study)		**Modified VGG16**	**17**	**96.41**
Nga *et al*. (this study)		**Modified ResNet50**	**17**	**97.88**

D. Research in Rice Variety Classification in the Literature

Recently, rice variety classification has concerned many scientists, so there has been a lot of research in this field. Some research and results are presented in Table VI. Because the rice variety datasets in the research were different, this table is presented to provide more information about the results of rice varieties classification.

IV. CONCLUSION

In this study, three models (an ANN along with modified VGG16 and modified ResNet50 models) were applied to classify 17 rice varieties commonly planted in Vietnam. Features from images of the 17 rice varieties were extracted via the extended ILTP method to form a feature dataset. The data augmentation technique was applied to generate seven image sets with different dimensions for image dataset. The classification accuracy achieved by the ANN was 92.82% when applied to the feature dataset. For the image dataset, the modified VGG16 obtained the highest accuracy at 96.41% with Set 3, while the modified ResNet50 achieved its highest accuracy at 97.88% with Set 4. From the results, we can see that the modified VGG16 and modified ResNet50 models can improve significantly the classification accuracy of the 17 rice varieties. In the future, more CNN models will be applied in this type of work to find the best classification performance.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Nga Tran-Thi-Kim and Tuan Pham-Viet conceived the idea. Nga Tran-Thi-Kim conducted the research, and wrote the paper. Tuan Do-Hong, Tuan Pham-Viet, Vladimir Y. Mariano, and Insoo Koo supervised the research, and critically reviewed and revised the paper.

ACKNOWLEDGMENT

We acknowledge Ho Chi Minh City University of Technology (HCMUT), VNU-HCM, for supporting this study. The authors also wish to thank Dr. Dang Minh

Tam from the Cuu Long Delta Rice Research Institute in Can Tho, Vietnam, for providing samples of the 17 rice varieties.

REFERENCES

[1] D. A. Vaughan, *The Wild Relatives of Rice: A Genetic Resources Handbook*, International Rice Research Institute, Manila, Philippines, 1994, pp. 3-5.

[2] B. S. Anami, N. M. Naveen, and N. G. Hanamaratti, "A colour features based methodology for variety recognition from bulk paddy images," *International Journal of Advanced Intelligence Paradigms*, vol. 7, no. 2, pp. 187-205, July 2015.

[3] I. Cinar, and M. Koklu, "Classification of rice varieties using artificial intelligence methods," *International Journal of Intelligent Systems and Applications in Engineering*, Ismail Saritas, Turkey, vol. 7, no. 3, pp. 188-194, 2019.

[4] A. R. Pazoki, F. Farokhi, and Z. Pazoki, "Classification of rice grain varieties using two artificial neural networks (MLP and Neuro-fuzzy)," *The Journal of Animal and Plant Sciences*, vol. 24, no.1, pp. 336-343, 2014.

[5] P. T. Thu Hong, T. T. Thanh Hai, L. T. Lan, V. T. Hoang, V. Hai, and T. T. Nguyen, "Comparative study on vision based rice seed varieties identification," in *Proc. of IEEE Seventh International Conference on Knowledge and Systems Engineering (KSE)*, Ho Chi Minh City, Vietnam, October 2015, pp. 377-382.

[6] C. S. Silva and U. Sonnadara, "Classification of rice grains using neural networks," in *Proc. of the Technical Sessions*, vol. 29, Kelaniya, Sri Lanka, March 2013, pp. 9-14.

[7] T. Y. Kuo, C. L. Chung, S. Y. Chen, H. A. Lin, and Y. F. Kuo, "Identifying rice grains using image analysis and sparse-representation-based classification," *Computers and Electronics in Agriculture*, vol. 127, pp. 716-725, Sep. 2016.

[8] S. D. Fabiyi, H. Vu, C. Tachtatzis, P. Murray, D. Harle, T. K. Dao, I. Andonovic, J. Jen, and S. Marshall, "Varietal classification of rice seeds using RGB and hyperspectral images," *IEEE Access*, vol. *8*, pp. 22493-22505, Jan. 2020.

[9] P. K. Sethy and A. Chatterjee, "Rice variety identification of western Odisha based on geometrical and texture feature," *International Journal of Applied Engineering Research (IJAER)*, vol. 13, no. 4, pp. 35-39, 2018.

[10] B. S. Anami, N. N. Malvade, and N. G. Hanamaratti, "An edge texture features based methodology for bulk paddy variety recognition," *Agricultural Engineering International: CIGR Journal*, vol. 18, no. 1, pp. 399-410, Mar. 2016.

[11] T. T. K. Nga, P. V. Tuan, D. M. Tam, I. Koo, V. Y. Mariano, and D. H. Tuan, "Combining binary particle swarm optimization with support vector machine for enhancing rice varieties classification accuracy," *IEEE Access*, vol. 9, pp. 66062-66078, May. 2021.

[12] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51-59, Jan. 1996.

[13] N. Alpaslan and K. Hanbay, "Multi-resolution intrinsic texture geometry-based local binary pattern for texture classification," *IEEE Access*, vol. 8, pp. 54415-54430, Mar. 2020.

[14] Z. Pan, X. Wu, and Z. Li, "Scale-adaptive local binary pattern for texture classification," *Multimedia Tools and Applications*, vol. *79*, no. 9, pp. 5477-5500, Mar. 2020.

[15] W. Wang, Q. Kou, S. Zhou, K. Luo, and L. Zhang, "Geometry-based completed local binary pattern for texture image classification," in *Proc. 2020 IEEE 3rd International Conference on Information Communication and Signal Processing (ICICSP)*, IEEE, Shanghai, China, Sep. 2020, pp. 274-278.

[16] L. Liu, S. Lao, P. W. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Median robust extended local binary pattern for texture classification," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1368-1381, Jan. 2016.

[17] X. Tan and W. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635-1650, Feb. 2010.

[18] S. Fekri-Ershad, "Bark texture classification using improved local ternary patterns and multilayer neural network," *Expert Systems with Applications*, vol. 158, 113509, Nov. 2020.

[19] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, pp. 971-987, July 2002.

[20] T. T. K. Nga, P. V. Tuan, D. M. Tam, I. Koo, V. Y. Mariano, and D. H. Tuan, "Extending color properties for texture descriptor based on local ternary patterns to classify rice varieties," *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, pp.1-13, Mar. 2022.

[21] T. Lin, X. Chen, X. Tang, L. He, S. He, and Q. Hu, "Deep learning based classification of radar spectral maps," *International Journal of Electrical and Electronic Engineering and Telecommunications*, vol. 10, no. 2, pp. 99-104, Mar. 2021.

[22] W. Ayadi, W. Elhamzi, I. Charfi, and M. Atri, "Deep CNN for brain tumor classification," *Neural Processing Letters*, vol. 53, pp. 671-700, Jan. 2021.

[23] I. V. Pustokhina, D. A. Pustokhin, J. J. P. C. Rodrigues, D. Gupta, A. Khanna, K. Shankar, C. Ceo, and G. P. Joshi, "Automatic vehicle license plate recognition using optimal K-means with convolutional neural network for intelligent transportation systems," *IEEE Access*, vol. 8, pp. 92907-92917, May 2020.

[24] S. Ahlawat, A. Choudhary, A. Nayyar, S. Singh, and B. Yoon, "Improved handwritten digit recognition using convolutional neural networks (CNN)," *Sensors*, vol. 20, 3344, June 2020.

[25] P. Lin, X. L. Li, Y. M. Chen, and Y. He, "A deep convolutional neural network architecture for boosting image discrimination accuracy of rice species," *Food and Bioprocess Technology*, vol. *11*, pp. 765-773, Jan. 2018.

[26] Z. Qiu, J. Chen, Y. Zhao, S. Zhu, Y. He, and C. Zhang, "Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network," *Applied Sciences*, vol. 8, no. 2, p. 212, Jan. 2018.

[27] K. Kiratiratanapruk, P. Temniranrat, W. Sinthupinyo, P. Prempree, K. Chaitavon, S. Porntheeraphat, and A. Prasertsak, "Development of paddy rice seed classification process using machine learning techniques for automatic grading machine," *Journal of Sensors*, vol. 2020, Article ID 7041310, July 2020.

[28] P. V. H. Duc, T. Surinwarangkoon, V. T. Hoang, H. T. Duong, and K. Meethongjan, "A comparative study of rice variety classification based on deep learning and hand-crafted features," *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, vol. *14, no.* 1, pp.1-10, Jan. 2020.

[29] N. Wu, C. Zhang, X. Bai, X. Du, and Y. He, "Discrimination of chrysanthemum varieties using hyperspectral imaging combined with a deep convolutional neural network," *Molecules*, vol. 23, no. 11, 2831, Oct. 2018.

[30] M. Koklu, I. Cinar, and Y. S. Taspinar, "Classification of rice varieties with deep learning methods," *Computers and Electronics in Agriculture*, vol. 187, p. 106285, Aug. 2021.

[31] I. Z. Mukti and D. Biswas, "Transfer Learning Based Plant Diseases Detection Using ResNet50," in *Proc. 4th International Conference on Electrical Information and Communication Technology (EICT)*, 2019, Khulna, Bangladesh, pp. 1-6.

[32] T. M. Mitchell, *Machine Learning*, McGraw-Hill Science, 1997, pp. 81-126.

[33] A. Kannur, A. Kannur, and V. S. Rajpurohit, "Classification and grading of bulk seeds using artificial neural network," *International Journal of Machine Intelligence*, vol.3, no. 2, pp. 62-73, Sep. 2011.

[34] J. Wu, "Introduction to convolutional neural networks," *National Key Lab for Novel Software Technology*, Nanjing University, China, vol. 5, no. 23, 495, May 2017.

[35] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. 2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1-6.

[36] A. Ghosh, A. Sufian, F. Sultana, A. Chakrabarti, and D. De, "Fundamental concepts of convolutional neural network," in *Proc. Recent Trends and Advances in Artificial Intelligence and Internet of Things*, Springer, Cham, 2020, pp. 519-567.

[37] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 24-49, Mar. 2021.

[38] V. A. Patel and M. V. Joshi, "Convolutional neural network with transfer learning for rice type classification," in *Proc. of Tenth Int. Conf. on Machine Vision*, Vienna, 2018, pp. 282-289.

[39] K. R. Singh and S. Chaudhury, "Comparative analysis of texture feature extraction techniques for rice grain classification," *IET Image Processing*, vol. 14, no. 11, pp. 2532-2540, Aug. 2020.

Nga Tran-Thi-Kim received the B.E. degree in electronics and telecommunications engineering from the Da Nang University of Science and Technology, Danang, Vietnam, in 2002, and the M.E. degree in electronics engineering from The University of Danang, Danang, Vietnam, in 2009. She is currently pursuing the Ph.D. degree in electronics engineering with the Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology, Vietnam. Since 2009, she has been with the Faculty of Engineering and Technology, Nong Lam University Ho Chi Minh City, Ho Chi Minh City, Vietnam. Her research interests include computer vision, digital image processing, and machine learning.

Tuan Pham-Viet received the B.E. and M.E. degrees in electronics and telecommunications engineering from the Ho Chi Minh City University of Technology, Vietnam, in 2005 and 2011, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Ulsan, South Korea, in 2018. He was as a Postdoctoral Researcher with the Multimedia Communications System Laboratory, University of Ulsan. His research interests include optimizations, MIMO communications, and machine learning.

Insoo Koo received the B.E. degree from KonKuk University, Seoul, South Korea, in 1996, and the M.S. and Ph.D. degrees from the Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea, in 1998 and 2002, respectively. From 2002 to 2004, he was with the Ultrafast Fiber-Optic Networks Research Center, GIST, as a Research Professor. In 2003, he was a Visiting Scholar with the Royal Institute of Science and Technology, Stockholm, Sweden. In 2005, he joined the University of Ulsan, South Korea, where he is currently a Full Professor. His current research interests include next generation wireless communication systems and wireless sensor networks.

Vladimir Y. Mariano received the B.S. degree in statistics and the M.S. degree in computer science from the University of the Philippines Los Banos, and the Ph.D. degree in computer science and engineering from The Pennsylvania State University. His research interests include computer vision, digital image processing, and machine learning.

Tuan Do-Hong (Member, IEEE) received the B.S. and M.Eng. degrees in electrical engineering from the Ho Chi Minh City University of Technology, Vietnam, in 1994 and 1997, respectively, and the M.Sc. and Ph.D. degrees in communication engineering from the Munich University of Technology, Germany, in 2000 and 2004, respectively. Since 1994, he has been with the Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology. He is currently the Dean of the Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology. His research interests include signal processing and applications for digital communications, and image and video processing.