*ICIC Express Letters*
Volume **15**, Number **6**, June **2021**
 ISSN 1881-803X
pp. 531–543

# ENSEMBLE CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES FOR LAND USE CLASSIFICATION IN ECONOMIC CROPS AERIAL IMAGES

Sangdaow Noppitak[1] and Olarik Surinta[2,*]

[1]Department of Information Technology
Faculty of Informatics
Mahasarakham University
Maha Sarakham 44150, Thailand
61011261007@msu.ac.th

[2]Multi-Agent Intelligent Simulation Laboratory (MISL)
Department of Information Technology
Faculty of Informatics
Mahasarakham University
Maha Sarakham 44150, Thailand
*Corresponding author: olarik.s@msu.ac.th

Received October 2020; accepted January 2021

Abstract. *The analysis of land use and land cover is a task of remote sensing and geographic information systems. Nowadays, deep learning techniques can analyze land use and land cover with high performance. In this paper, we focus on the classification of land use for Thailand's economic crops based on the convolutional neural network (CNN) technique. We evaluated the ensemble CNN framework on Thailand's economic crops aerial image dataset called the EcoCropsAID dataset. Five economic crops categories, including rice, sugarcane, cassava, rubber, and longan, were collected using the Google Earth program. Economic crops aerial images obtained between 2014 and 2018 were considered. There were 5,400 images with approximately 1,000 images per class. Due to the ensemble CNN framework, we first proposed to use eight pre-trained CNN models consisting of InceptionResNetV2, MobileNetV2, DenseNet201, Xception, ResNet152V2, NASNetLarge, VGG16, and VGG19 to discover the best baseline CNN model. Second, three simplistic data augmentation techniques (rotation, width shift, and height shift) are applied to increasing the accuracy of the CNN models. As a result, we found that the three best models were VGG16, VGG19, and NASNetLarge architectures, respectively. Finally, we created an ensemble CNN framework that consisted of 3 CNNs based on the best CNN models. We also compared three ensemble methods, that were weighted average, unweighted average, and unweighted majority vote. From our experiments, the results show that the VGG16 outperforms other CNN models. Consequently, the classification performance on Thailand's economic crops aerial image dataset was significantly improved when the weighted average ensemble method was employed.*
**Keywords:** Land use classification, Economic crops aerial image, Deep learning, Ensemble convolutional neural networks, Ensemble method, Data augmentation technique

1. **Introduction.** Thailand is a country that mainly exports agricultural products that are economic crops, including rice, corn, cassava, sugar, rubber, palm oil, tapioca, and longan [1]. Hence, the government sector has to analyze the information and forecast the world economy, especially in agriculture, which requires the consideration of many factors outside the country, such as the world agricultural economy, and crude oil price. The domestic factors include the amount of water in reservoirs, rainfall, land use, etc. [2]. Without appropriate planning of land use, negative consequences might ensue, such as

DOI: 10.24507/icicel.15.06.531

selection of the wrong plant products and quantities inconsistent with export to international markets.

Land use and land cover can be interrogated by remote sensing (RS) and geographic information system (GIS) tasks that can be used to analyze, plan, and manage the quality of human life in respect of many issues. For example, the rapidly growing population causes environmental problems, air pollution, and temperature rise in urban areas [3]. Since the flooding in Ho Chi Minh City, Vietnam, Schaefer and Thinh [4] proposed using GIS and RS methodologies to evaluate the changes of land cover and agricultural protection sites. In the future, planners can use the proposed methods in decision support for managing land use. To address the problems caused by the rapid urbanization of human activities in the Halda river located in the south-eastern region of Bangladesh, Chowdhury et al. [5] used remote sensing and GIS to assess land use and land cover changes in the Halda watershed. Populations in the Nile Delta, Egypt, require basic security because of the unplanned urban growth [6] and propose a remote sensing technique to estimate the land use change according to the unplanned urban growth by monitoring the human activities that change in agricultural and urban areas. Consequently, remote sensing and GIS are used to classify land use in problematic regions due to the landscape (i.e., hills and lowlands) [7].

Nowadays, deep learning is a well-known technique proposed to address land use and land cover classification. The techniques are widely used in land use classification, for example, of buildings, paved roads, vegetation density, grassland, and water bodies [8]. Also, urban planning and management use deep learning models to classify land use in urban areas using high-resolution satellite imagery (VHR) [9]. It can be seen that deep learning can be employed to address the land use classification.

In this paper, we propose an ensemble convolutional neural network framework (CNN) for classification of land use in economic crops aerial images. The contributions of this paper can be summarized as follows.

1) We propose ensemble methods with deep convolutional neural networks, called the ensemble CNN method for land use classification. Using the ensemble method, we present the weighted average approach to find the optimal weight by applying weights to the output probabilities of each baseline CNN model. Based on the CNN architecture, we discover further efficient models by employing pre-trained CNN models from eight CNN architectures, including InceptionResNetV2, MobileNetV2, DenseNet201, Xception, ResNet152V2, NASNetLarge, VGG16, and VGG19. Our experimental results indicate that the three best CNN models on the economic crops aerial image dataset are VGG16, VGG19, and NASNetLarge, respectively. We also realize that the simplistic data augmentation techniques, such as rotation and height shift techniques, could improve the performance of the CNN method. We have demonstrated that the data augmentation technique significantly increases the performance of the CNN model.

2) Our paper aims to enable the use of Thailand's economic crops aerial image dataset, namely the EcoCropsAID dataset, for land use classification. In this paper, we collect the aerial image data between 2014 and 2018 by using the Google Earth program. The image quality is different depending on the different remote sensor types used by the Google Earth program; importantly, the aerial image quality is different. The EcoCropsAID dataset consists of 5,400 images that contain five categories: rice, sugarcane, cassava, rubber, and longan.

The remainder of this paper is organized as follows. Section 2 reviews the literature in the domain of land use classification using deep learning architectures. The proposed framework is introduced in Section 3. Section 4 describes the economic crops aerial image dataset. In Section 5, the results of the proposed method are presented and discussed. The last section concludes and advises feasible domains for future work.

2. **Related Work.** This research focuses on the land use classification on aerial images using deep learning algorithms. Many studies in remote sensing and geoinformatics have mainly experimented on satellite images. In this paper, however, we experiment with aerial images collected in RGB color space. We survey aerial image datasets that have been used for land use and land cover classification tasks. The datasets, such as UC Merced land use [10], RESISC45 [11], and aerial image dataset (AID) [12] datasets, were created from the Google Earth program, except the EuroSAT dataset [13] that was collected from the Sentinel-2 satellite. We collected the aerial image data from the Google Earth program that considers only five economic crops consisting of rice, sugarcane, cassava, rubber, and longan. The information on our proposed aerial image dataset is presented in Section 4.

Many convolutional neural network (CNN) architectures, such as AlexNet, CaffeNet, GoogleNet, VGGNet, PlacesNet, Inception, and ResNet, were proposed to address the land use and land cover classification of the satellite images [14]. The CNN architectures are also proposed to extract the deep features from the aerial image. Xia et al. [12] proposed benchmark AID dataset, for aerial scene classification. The AID dataset contains 10,000 images and 30 categories. They then applied CaffeNet, VGG-VD-16, and GoogLeNet to extracting the deep features and classifying them using the Liblinear supervised classification. These methods achieved around 86% accuracy on the AID dataset. Pilipović and Risojević [15] evaluated three CNN architectures: GoogLeNet, ResNet, and SqueezeNet, on the high-resolution remote sensing dataset. The deep features were extracted using fine-tuned CNN models and presented to the support vector machine (SVM) method as a classifier. The results showed that the GoogLeNet fine-tuned features combined with the SVM outperformed the other methods on the UCM dataset. Additionally, the ResNet fine-tuned features showed the best result on the AID dataset.

Han et al. [16] proposed a framework that combines the deep feature and discriminative evaluation methods for scene classification and annotation, called the semi-supervised generative framework (SSGF). To evaluate the performance, when considering only the deep feature method, the ResNet model combined with supervised learning outperforms all CNN models on the aerial image datasets, including AID, UCM, NWPU-RESISC45, and WHU-RS19. The SSGF method that combined ResNet, VGG-S, and VGG-16 still achieved the best accuracy result on all datasets.

For aerial scene classification, Zheng et al. [17] proposed a deep scene representation approach. In this approach, deep features are extracted by the multi-scale max-pooling method and then given to the Fisher vector method to encode the multi-scale features into a global representation. Various pre-trained CNN models were evaluated, including AlexNet, CaffeNet, GoogLeNet, and VGGNet, on different aerial scene datasets. This approach achieved an accuracy above 93% on UCM, WHU-RS19, RSSCN7, and AID datasets.

It can be seen that the CNN architectures can be proposed to address several classification tasks, such as land use, land cover, and aerial scene. This research presents an ensemble CNN framework that combines the three beneficial baseline CNN models to classify aerial images. The ensemble CNN can decrease the generalization error and prediction variance. The proposed architecture is then explained as follows.

3. **Proposed Ensemble Convolutional Neural Network Architecture.** The ensemble method aims to enhance the accuracy results of the classification tasks. This method combines various classifiers instead of applying only an individual classifier and provides more robust results [18, 19]. Figure 1 illustrates the ensemble convolutional neural network (CNN) architecture. The proposed architectures consist of two schemes.

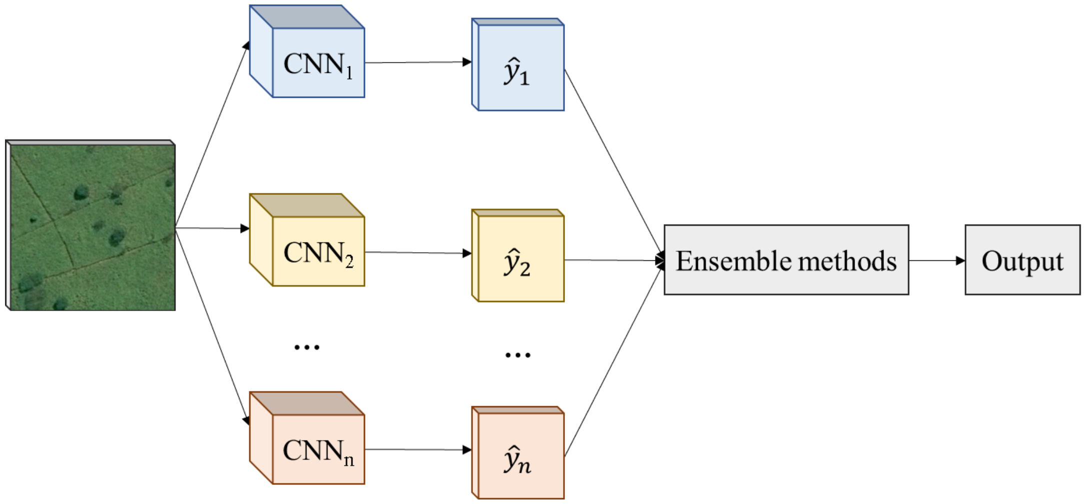



FIGURE 1. The framework of ensemble convolutional neural networks for land use classification in economic crops aerial images

*In the first scheme*, we first discover the best baseline CNN model from several state-of-the-art CNN architectures, including VGG16, VGG19, Xception, ResNet, InceptionResNet, MobileNet, DenseNet, and NASNetLarge. Second, the data augmentation techniques are employed to improve the performance of the CNN models. Finally, according to our experiments, we combine the three best CNN models. Then, the probability distribution, which is computed by the softmax function, is assigned to classify using the ensemble methods. The detail of the first scheme is explained in Sections 3.1 and 3.2.

*The second scheme* aimed to compare three ensemble methods: weighted average, unweighted average, and unweighted majority vote, regularly applied for neural networks. The ensemble methods are described in Section 3.3.

3.1. **Convolutional neural network (CNN) architectures.** In this study, we propose CNN architectures to address the land use classification problem on Thailand's economic crops aerial image dataset (EcoCropsAID dataset). To find the three most beneficial baseline CNN models, we compare the performance of eight CNN architectures, including Xception, VGG16, VGG19, ResNet152V2, InceptionResNetV2, MobileNetV2, DenseNet201, and NASNetLarge [20, 21, 22, 23, 24, 25]. We observe that the VGG16, VGG19, and NASNetLarge architectures achieved high performance on the EcoCropsAID dataset based on our experiments. In this study, we then created ensemble CNN using the best three CNN architectures to enhance land use classification performance. We further describe the baseline CNN architectures and the data augmentation techniques as follows.

**VGGNet:** In 2015, Simonyan and Zisserman [21] invented VGG Networks, namely VGGNets. The VGGNets are designed according to the depth of the weight layers with 16-19 layers. These networks are divided into five convolutional blocks and the max-pooling layer follows each block. In each block, to create the feature maps, the small convolution filters with the size of $3 \times 3$ are computed. Then, in each block, the feature maps are reduced by half size of the previous block. In contrast, the feature maps are increased by double layers of the last block. Moreover, the FC layer is employed as the classifier. Additionally, the two fully connected (FC) layers with 4,096 and the one final FC layer are the outputs of the network. According to the EcoCropsAID dataset, in our framework, the final FC layer is designed as five. The architecture the VGGNets is shown in Figure 2.

**NASNet:** Zoph and Le [25] proposed a neural architecture search (NAS) that generates the CNN architecture using the recurrent neural network (RNN) with reinforcement learning. In 2018, Zoph et al. [26] developed the learned transferable architecture by

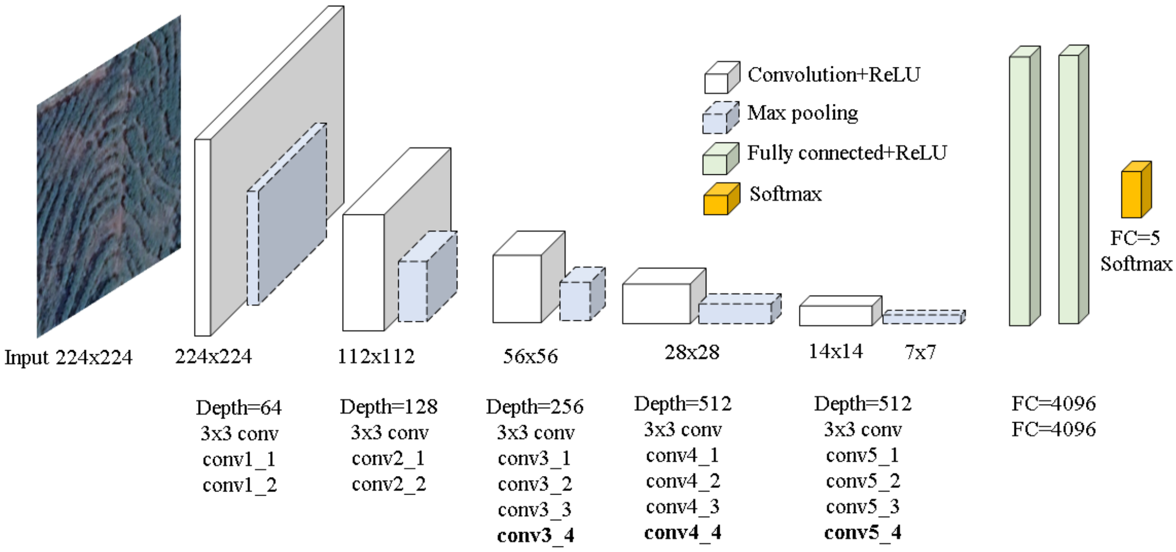



FIGURE 2. Network architectures of VGG16 and VGG19 (Bold)

extending the NASNet architecture. In this architecture, the RNN method is employed as the search method to explore the best CNN architecture. The NASNet architecture includes the normal and reduction cells searched by the RNN method. For the transfer architecture, the best CNN architecture is created based on learning from the small dataset (i.e., the CIFAR-10 dataset). Consequently, the CNN architecture is transferred to learn with a large dataset (i.e., imageNet). The NASNet architecture comprises NASNetLarge and NASNetMobile.

3.2. **Data augmentation techniques for aerial images.** The concept of data augmentation is to solve the problem of having an insufficient amount of data by increasing the number of training data [27]. Therefore, the new image is a synthesis from the original images using different augmentation methods, and new diversity images are also generated. Image data augmentation techniques are classified into two main groups: basic image manipulations (i.e., flip, color space, crop, rotation, translation, etc.) and deep learning approaches (i.e., adversarial training, neural style transfer, and meta-learning) [28].

Due to the aerial image data augmentation techniques, two categories are introduced, including instance-based augmentation (such as geometric, color, deformation, enhancement, and brightness) and fusion-based augmentation (i.e., the RGB channels from the satellite images are combined) [29].

Our experiments perform three data augmentation techniques: rotation, width shift, and height shift. The example images of the basic manipulation techniques are shown in Figure 3.

3.3. **Ensemble methods.** Due to an increment in the classification performance, the output probabilities of the CNN models are combined and classified using the ensemble methods. A brief explanation of the ensemble methods is as follows.

**Unweighted average method:** In this method, the CNN output probabilities of each model, which are computed by the softmax function, are averaged [30]. The highest probability is decided as a result. The output ($\widehat{y}_i$) is computed by Equation (1):

$$\widehat{y}_i = \frac{1}{n}\sum_{j=1}^{n} y_i \tag{1}$$

where $y_i$ is the output probabilities of the CNN model and $n$ is the number of the CNN models.

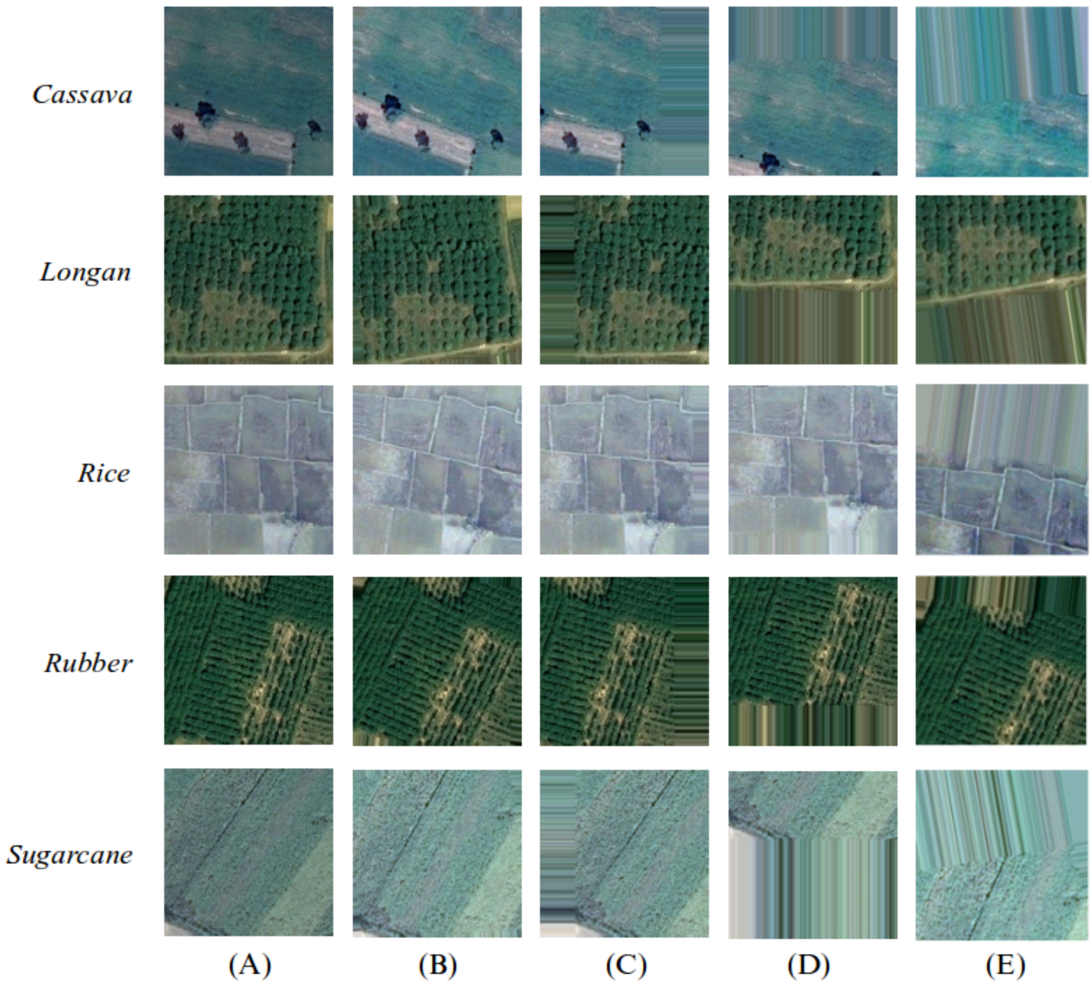



FIGURE 3. Example of data augmentation techniques: (A) Original image, (B) rotation, (C) width shift, (D) height shift, and (E) combination between rotation and width shift

**Weighted average method:** Due to the classification performance of the CNN model, the different weights are applied to the output probabilities. Hence, the higher weight is assigned to the CNN model that achieved a higher classification rate [31]. The weighted average method is given by Equation (2):

$$\widehat{y}_i = \frac{1}{n}\sum_{j=1}^{n} \alpha y_i \tag{2}$$

where $\alpha$ is a weight that multiplies with the output probabilities $y_i$ of the CNN models.

**Unweighted majority vote method:** The $\arg\max$ function is applied to the output probabilities of each CNN model and determined as the predicted labels. For each class, the number of votes is counted. Then, the most maximum votes are decided as the final decision [32]. The unweighted majority vote method is calculated by Equation (3):

$$\widehat{y}_i = \frac{1}{n}\sum_{j=1}^{n} \arg\max(y_i) \tag{3}$$

where $\arg\max$ returns the maximum value of the output probabilities $y_i$ of the CNN models.

4. **Thailand's Economic Crops Aerial Image Dataset.** We introduce the novel economic crops aerial image dataset, namely the EcoCropsAID dataset. This dataset was collected in Thailand from five economic crops that were cultivated in different provinces and regions between 2014 and 2018. The aerial images of economic crops were gathered based on Agri-Map Online provided by the Ministry of Agriculture and Cooperatives and the National Electronics and Computer Technology Center (NECTEC). The Agri-Map Online is an agriculture map that all departments under the Ministry of Agriculture and Cooperatives use as an agriculture management tool. Subsequent agricultural information is accurate and up-to-date [33]. Then, the Google Earth application was employed to capture aerial images after we selected the economic crops areas in which images were to be collected. It is quite a complex dataset because the Google Earth program used several remote imaging sensors [12] to record the aerial images.

The EcoCropsAID dataset includes five categories (rice, sugarcane, cassava, rubber, and longan) and contains 5,400 images. Each class has around 1,000 images. To prepare the aerial images of the economic crops, we recorded the image with $600 \times 600$ pixels and stored it in the RGB color format. Sample aerial images of this dataset are shown in Figure 4. As seen in Figure 4(A), the first row is the cassava field, that is at commencement of planting. However, the third row was a change in land use from rice field to cassava field because strange lines have appeared. Also, Figure 4(E) (first and second rows) shows a change in land use from rice field to sugarcane field. The pattern of the longan area is similar to the rubber area, as shown in Figure 4(D) in the first and second rows. As seen in Figure 4(C), the first row presents a wet area. The second row of Figure 4(C) illustrates the not yet planting area, while the third row is the beginning of planting. The challenges of classification on the EcoCropsAID dataset are 1) many different image resolutions and colors are contained in the EcoCropsAID dataset due to the various remote imaging sensors, 2) the similarity of patterns amongst each class, for example, longan and rubber, and 3) the difference of pattern inside the same class, for example, cassava and rice.

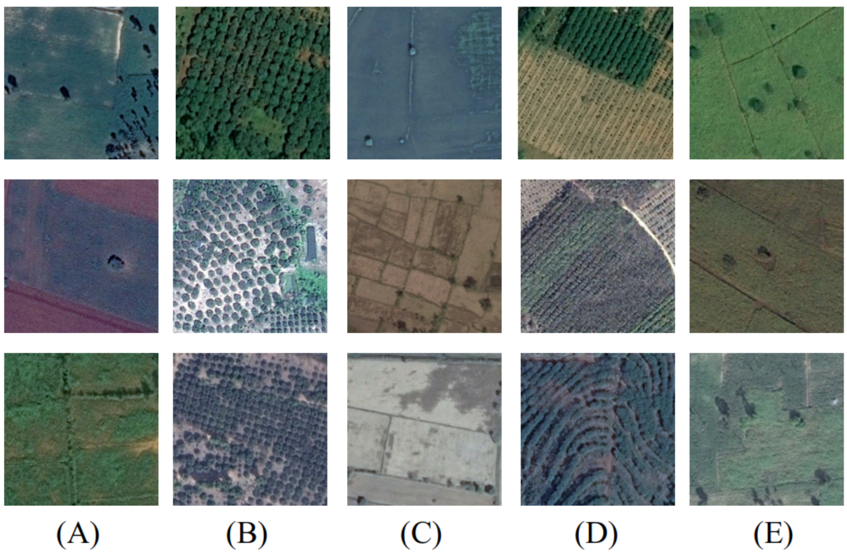



Figure 4. Example of economic crops aerial images: (A) Cassava, (B) longan, (C) rice, (D) rubber, and (E) sugarcane

5. **Experimental Results.** In this section, we present experiments on our economic crops aerial image dataset. All experiments are tested in the same environment. We used the TensorFlow as the deep learning framework that runs on Intel(R) Core-i9-9900K CPU @ 3.60GHz ×16, 32GB RAM, and GPU NVIDIA GeForce GTX 1080Ti. The experiment results are explained as follows.

5.1. **Experiments with CNN architectures and data augmentation techniques.** For the CNN experiments, we will present the best setting parameters of the CNN models on the EcoCropsAID dataset. However, to find the best setting parameters of the CNN models, we randomly select 25% (1,350 images) for the training data and test data. Eight CNN models are selected: Xception, VGG16, VGG19, ResNetV2, InceptionResNetV2, MobileNetV2, DenseNet, and NASNetLarge. We use the transfer learning technique to train eight pre-trained CNN models. We focus on performing two optimization algorithms, including stochastic gradient descent (SGD) and Adam optimizers. The batch size experiments used sizes of 4, 8, 16, 32, and 64. The learning rate and the number of training epochs are determined as 0.001 and 100.

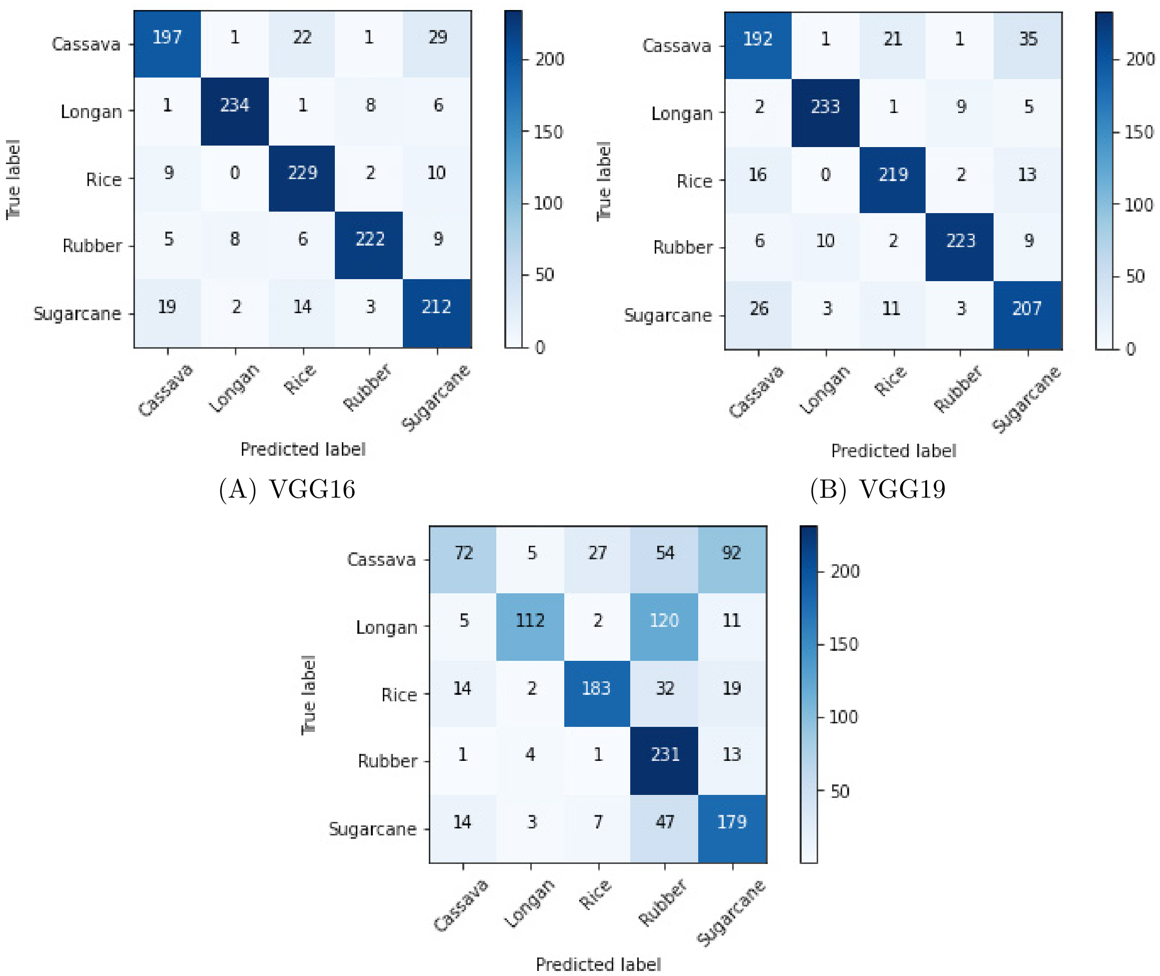
Table 1 shows the best CNN parameters and classification performances achieved from eight CNN models. The experiment results show that the VGGNet performs much better than other CNN models. The VGG16 significantly outperforms the VGG19. Also, VGGNets require less computation time (it took around 12 minutes). However, based on our experiments, the worst performance with approximately 48% accuracy is the InceptionResNetV2 and MobileNetV2 models. Subsequently, the NASNetLarge model requires more computation time and spends around 1 hour and 20 minutes. When comparing the accuracy between different optimizers and different batch sizes, we found that SGD obtained better results than Adam, except for Xception and NASNetLarge. The Adam optimizer performed better when using a small batch size, while the SGD optimizer gave better experiments when using a large batch size. Considering the results, as shown in Table 1, we selected and performed other experiments based on three CNN models: VGG16, VGG19, and NASNetLarge.

Table 1. Performances of the convolutional neural network architectures on the EcoCropsAID dataset

| Model | Optimizer | Batch size | Accuracy (%) | Training time | No. of parameters |
|---|---|---|---|---|---|
| InceptionResNetV2 | SGD | 16 | 48.00 | 29 mins | 54,828,261 |
| MobileNetV2 | SGD | 32 | 48.40 | 5 mins | 2,571,589 |
| DenseNet201 | SGD | 64 | 50.75 | 15 mins | 18,792,389 |
| Xception | Adam | 16 | 52.99 | 21 mins | 21,885,485 |
| ResNet152V2 | SGD | 64 | 59.87 | 19 mins | 58,833,413 |
| NASNetLarge | Adam | 8 | 62.29 | 1 h 21 mins | 87,356,183 |
| VGG19 | SGD | 64 | 85.92 | 12 mins | 20,149,829 |
| VGG16 | SGD | 16 | **87.57** | 11 mins | 14,840,133 |

Figure 5 shows the confusion matrices of VGG16, VGG19, and NASNetLarge models. The confusion matrix shows that the cassava class was misclassified as belonging to the sugarcane class. The VGG16, VGG19, and NASNetLarge misclassified 29, 35, and 92 images, respectively and this is because the patterns of the cassava and sugarcane classes are quite similar.

For data augmentation experiments, the EcoCropsAID dataset is divided into a training set 80% (4,320 images) and a test set 20% (1,080 images). As seen from the results in Table 2, we experimented with three CNN models, VGG16, VGG19, and NASNetLarge, as pre-trained models. We also considered three data augmentation techniques: rotation, width shift, and height shift to increase the performance of the CNN models. These data augmentation techniques do not destroy the aerial image spectral information [34]. We explore the optimal values by configuring the parameters of the data augmentation techniques as follows. The rotation technique parameter is 10, 20, and 30 degrees; width and height shift parameters are 0.1, 0.2, and 0.3 ratios. As a result, the optimal parameter




(A) VGG16 (B) VGG19

(C) NASNetLarge

Figure 5. The confusion matrices of (A) VGG16, (B) VGG19, and (C) NASNetLarge models

Table 2. Test accuracy (%) of the CNN architectures and data augmentation techniques on the EcoCropsAID dataset

| Model (Data augmentation) | Optimizer | Batch size | Accuracy | Training time | No. of parameters |
|---|---|---|---|---|---|
| NASNetLarge (Hshift) | Adam | 8 | 59.40 | 1 d 1 h 12 mins | 87,356,183 |
| NASNetLarge (Rotation + Hshift) | Adam | 8 | 79.60 | 1 d 1 h 29 mins | 87,356,183 |
| VGG19 (Rotation + Hshift) | SGD | 64 | 86.50 | 6 h 23 mins | 20,149,829 |
| VGG19 (Hshift) | SGD | 64 | 88.30 | 6 h 20 mins | 20,149,829 |
| VGG16 (Rotation + Hshift) | SGD | 16 | **91.50** | 6 h 39 mins | 14,840,133 |
| VGG16 (Hshift) | SGD | 16 | **91.50** | 6 h 39 mins | 14,840,133 |

of the rotation is 10 degree, width shift (Wshift) is 0.3 ratio, and height shift (Hshift) is 0.3 ratio. It should be noted that we examine data augmentation techniques by combining two more data augmentation techniques [27]: rotation + Wshift, rotation + Hshift, Wshift + Hshift, and Wshift + Hshift.

Table 2 shows the experimental results of the CNN models and data augmentation techniques. The experiments show that the VGG16 (Rotation + Hshift) and VGG16 (Hshift) are the best models with an accuracy of 91.50%. These VGGNets spend around 6 hours and 30 minutes when training. However, The NASNetLarge (Rotation + Hshift) obtains 79.60% accuracy, which is 11.90% lower than the VGGNet models. When training the model, the NASNetLarge requires more than 24 hours. The results also showed that the VGG16 model performed better than VGG19 and NASNetLarge models. It is concluded that the deep layers and number of parameters did not affect the land use classification accuracy.

5.2. **Experiments with ensemble methods.** In these experiments, we have evaluated the ensemble methods consisting of unweighted average, weighted average, unweighted majority vote methods. For the weighted method, we optimize the weighted parameters using the grid-search method. We perform four ensemble CNN methods as follows: 1) NASNetLarge (Rotation + Hshift) + VGG19 (Hshift) + VGG16 (Hshift), called E1 model; 2) VGG16 (Rotation + Hshift) + VGG16 (Rotation + Wshift) + VGG16 (Rotation + Wshift + Hshift), called E2 model; 3) VGG16 (Rotation + Wshift) + VGG16 (Rotation + Wshift + Hshift) + VGG16 (Wshift + Hshift), called E3 model; 4) VGG16 (Rotation + Hshift) + VGG16 (Rotation + Wshift) + VGG16 (Wshift + Hshift), called E4 model.

Table 3 provides the accuracy results of three ensemble methods on the EcoCropsAID dataset. We observed that the weighted average ensemble method insignificantly outperforms the average ensemble method on models E2-E4, but with model E1 it was equally accurate.

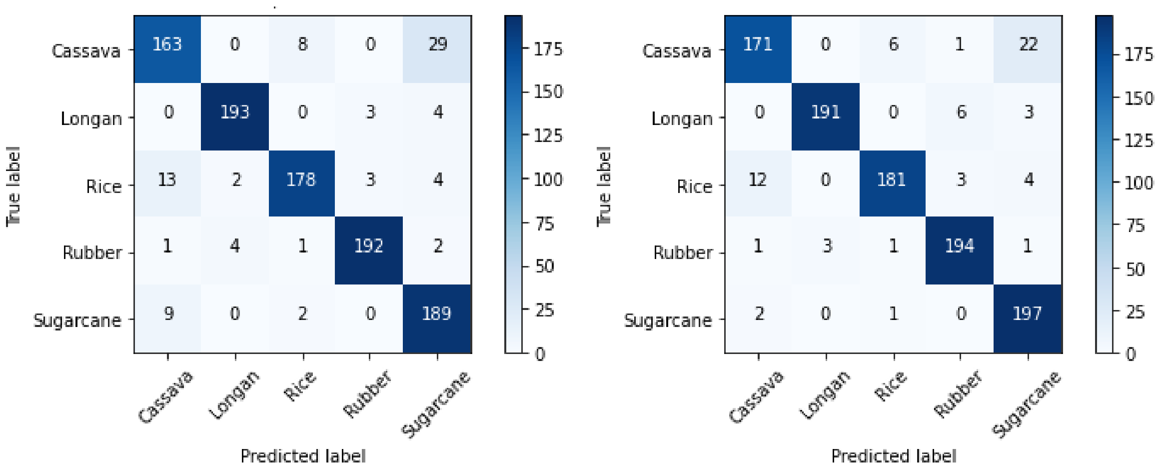
Table 3. Performances of the ensemble CNN methods on the EcoCropsAID dataset

| Model | Ensemble method | | |
|---|---|---|---|
| | Unweighted majority vote | Unweighted average | Weighted average |
| E1 | 92.00 | 92.60 | 92.60 |
| E2 | 91.90 | 92.40 | 92.70 |
| E3 | 92.30 | 92.30 | 92.70 |
| E4 | 92.50 | 92.60 | **92.80** |

In the grid search experiments, we defined the range of the weighted parameters of 0-0.995. We explored the weighted parameters on the training set. The total amount of the weighted parameters is equal to one. It took about 18 minutes to search. The best-weighted parameters for each ensemble CNN model (E1-E4) were as follows: E1 = $[0.08, 0.27, 0.65]$, E2 = $[0.07, 0.86, 0.07]$, E3 = $[0.86, 0.09, 0.05]$, and E4 = $[0.07, 0.86, 0.07]$.

Additionally, as can be seen from Figure 6, the confusion matrix explains that the misclassified from cassava class to sugarcane class is reduced from 29 to 22 images and misclassified from sugarcane class to cassava class is reduced from 9 to 2 images. Consequently, it can be seen that the ensemble CNN method achieves higher efficiency than training with only the individual CNN architecture.

The confusion matrix of the VGG architecture is illustrated in Figure 6(A). The results show that 29 cassava images are misclassified as belonging to the sugarcane class. Also, 8 cassava images are misclassified as rice class. Subsequently, 13 rice and 9 sugarcane images are misclassified as cassava. The confusion matrix of the ensemble CNN method is shown in 6(B). When capturing the aerial images at the beginning of the cultivation period,




(A) VGG16 (B) Ensemble CNN

FIGURE 6. Illustration of confusion matrix of (A) the VGG16 architecture when using the data augmentation technique with the height shift technique and (B) the ensemble CNN with the weighted average method on the EcoCropsAID dataset

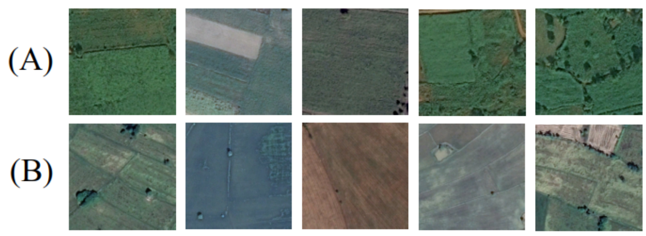



FIGURE 7. The results indicate that (A) several samples of ($\sim$ 29) cassava class are misclassified as sugarcane class and (B) ($\sim$ 13) rice class are misclassified as cassava class.

the cassava and sugarcane patterns always appear to be similar and it is challenging to distinguish between these two classes. The misclassification between the cassava and rice classes is shown in Figure 7.

6. **Conclusions.** In this paper, we have compared several convolutional neural network (CNN) architectures to discover the baseline CNN models for land use classification. For land use classification, we propose a novel dataset of Thailand's economic crops aerial images called the EcoCropsAID. This dataset includes five classes (rice, sugarcane, cassava, rubber, and longan) and contains 5,400 economic crops aerial images collected from the Google Earth program. We chose eight CNN models comprising InceptionResNetV2, MobileNetV2, DenseNet201, Xception, ResNet152V2, NasNetLarge, VGG16, and VGG19, in order to discover the best CNN model. From the experimental results, we conclude that the performance of the VGG16 and VGG19 models is unequivocally better than other CNN models. The experiments show that the VGG architecture significantly outperforms NASNetLarge with an accuracy of approximately 20% and InceptionResNetV2 with 37% accuracy. We have also demonstrated the impact of the data augmentation technique. Surprisingly, even the uncomplicated data augmentation techniques, such as rotation and height shift, improve the performance of the CNN architectures. The accuracy increased

by 17.31% when training the NASNetLarge architecture using data augmentation techniques.

We propose the ensemble CNN framework to achieve higher performance of land use classification. According to the ensemble method, three ensemble methods are then compared, including unweighted average, weighted average, and unweighted majority vote. The experiments showed that the three baseline CNN architectures combined with the weighted average ensemble method outperform when combined with the other ensemble methods. In order to evaluate the proposed framework, our ensemble CNN method achieved the recognition accuracy of 92.80% on the EcoCropsAID dataset. Although the ensemble method is an outstanding way to enhance the recognition performance, it does not improve the accuracy results if there is a high variance in the accuracy between each model.

In future work, to enhance the performance of land use classification, we will experiment on the various ensemble methods, such as the snapshot ensemble and stacked ensemble [35]. We will also consider self-supervised feature learning to extract the spatial feature [17] from the economic crops aerial image and classify the feature vector with other deep learning such as long short-term memory (LSTM) network. We are interested in using generative adversarial networks (GANs) [36] as the data augmentation technique.

## REFERENCES

[1] Office of Agricultural Economics, *Agricultural Economic Report: 3rd Quarter 2019 and Outlook for 2019*, Tech. Rep., Ministry of Agriculture and Cooperatives, Bangkok, 2019.

[2] Office of Agricultural Economics, *Agricultural Economic Outlook*, Tech. Rep., Ministry of Agriculture and Cooperatives, Bangkok, 2019.

[3] P. K. Mishra, A. Rai and S. C. Rai, Land use and land cover change detection using geospatial techniques in the Sikkim Himalaya, India, *Egyptian Journal of Remote Sensing and Space Science*, pp.1-11, 2019.

[4] M. Schaefer and N. X. Thinh, Evaluation of land cover change and agricultural protection sites: A GIS and remote sensing approach for Ho Chi Minh City, Vietnam, *Heliyon*, vol.5, 2019.

[5] M. Chowdhury, M. E. Hasan and M. M. Abdullah-Al-Mamun, Land use/land cover change assessment of Halda watershed using remote sensing and GIS, *The Egyptian Journal of Remote Sensing and Space Science*, vol.23, no.1, pp.63-75, 2020.

[6] M. H. Elagouz, S. M. Abou-Shleel, A. A. Belal and M. A. O. El-Mohandes, Detection of land use/cover change in Egyptian Nile Delta using remote sensing, *The Egyptian Journal of Remote Sensing and Space Science*, vol.23, no.1, pp.57-62, 2020.

[7] M.-R. Rujoiu-Mare and B.-A. Mihai, Mapping land cover using remote sensing data and GIS techniques: A case study of Prahova Subcarpathians, *Procedia Environmental Sciences*, vol.32, pp.244-255, 2016.

[8] H. A. Al-Najjar, B. Kalantar, B. Pradhan, V. Saeidi, A. A. Halin, N. Ueda and S. Mansor, Land cover classification from fused DSM and UAV images using convolutional neural networks, *Remote Sensing*, vol.11, no.12, 2019.

[9] P. Zhang, Y. Ke, Z. Zhang, M. Wang, P. Li and S. Zhang, Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery, *Sensors*, vol.18, 2018.

[10] Y. Yang and S. Newsam, Bag-of-visual-words and spatial extensions for land-use classification, *The 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, San Jose, CA, USA, pp.270-279, 2010.

[11] G. Cheng, J. Han and X. Lu, Remote sensing image scene classification: Benchmark and state of the art, *Proc. of the IEEE*, vol.105, pp.1865-1883, 2017.

[12] G. S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang and X. Lu, AID: A benchmark data set for performance evaluation of aerial scene classification, *IEEE Trans. Geoscience and Remote Sensing*, vol.55, pp.3965-3981, 2017.

[13] P. Helber, B. Bischke, A. Dengel and D. Borth, EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol.12, pp.2217-2226, 2019.

[14] M. P. Vaishnnave, K. S. Devi and P. Srinivasan, A study on deep learning models for satellite imagery, *International Journal of Applied Engineering Research*, vol.14, no.4, pp.881-887, 2019.

[15] R. Pilipović and V. Risojević, Evaluation of convnets for large-scale scene classification from high-resolution remote sensing images, *IEEE EUROCON – The 17th International Conference on Smart Technologies*, Ohrid, Macedonia, pp.932-937, 2017.

[16] W. Han, R. Feng, L. Wang and Y. Cheng, A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol.145, pp.23-43, 2018.

[17] X. Zheng, Y. Yuan and X. Lu, A deep scene representation for aerial scene classification, *IEEE Trans. Geoscience and Remote Sensing*, vol.57, pp.4799-4809, 2019.

[18] D. Kreuter, H. Takahashi, Y. Omae, T. Akiduki and Z. Zhang, Classification of human gait acceleration data using convolutional neural networks, *International Journal of Innovative Computing, Information and Control*, vol.16, no.2, pp.609-619, 2020.

[19] H. Dyoniputri and Afiahayati, A hybrid convolutional neural network and support vector machine for dysarthria speech classification, *International Journal of Innovative Computing, Information and Control*, vol.17, no.1, pp.111-123, 2021.

[20] F. Chollet, Xception: Deep learning with depthwise separable convolutions, *Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp.1251-1258, 2017.

[21] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *The 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, 2015.

[22] K. He, X. Zhang, S. Ren and J. Sun, Identity mappings in deep residual networks, *arXiv.org*, arXiv: 1603.05027, 2016.

[23] C. Szegedy, S. Ioffe, V. Vanhoucke and A. A. Alemi, Inception-V4, Inception-ResNet and the impact of residual connections on learning, *The 31st AAAI Conference on Artificial Intelligence (AAAI)*, San Francisco, CA, USA, pp.4278-4284, 2017.

[24] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.-C. Chen, MobileNetV2: Inverted residuals and linear bottlenecks, *Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, pp.4510-4520, 2018.

[25] B. Zoph and Q. V. Le, Neural architecture search with reinforcement learning, *The 5th International Conference on Learning Representations (ICLR)*, Toulon, France, pp.1-16, 2017.

[26] B. Zoph, V. Vasudevan, J. Shlens and Q. V. Le, Learning transferable architectures for scalable image recognition, *Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, Utah, USA, pp.8697-8710, 2018.

[27] P. Pawara, E. Okafor, L. Schomaker and M. Wiering, Data augmentation for plant classification, *Advanced Concepts for Intelligent Vision Systems (ACIVS)*, Antwerp, Belgium, pp.615-626, 2017.

[28] L. Perez and J. Wang, The effectiveness of data augmentation in image classification using deep learning, *arXiv.org*, arXiv: 1712.04621, 2017.

[29] M. A. A. Ghaffar, A. McKinstry, T. Maul and T. T. Vu, Data augmentation approaches for satellite image super-resolution, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol.IV-2/W7, no.9, pp.47-54, 2019.

[30] C. Ju, A. Bibaut and M. Van Der Laan, The relative performance of ensemble methods with deep convolutional neural networks for image classification, *Journal of Applied Statistics*, vol.45, no.15, pp.2800-2818, 2018.

[31] X. Frazão and L. A. Alexandre, Weighted convolutional neural network ensemble, in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2014. Lecture Notes in Computer Science*, E. Bayro-Corrochano and E. Hancock (eds.), Cham, Springer, 2014.

[32] O. Surinta, L. Schomaker and M. Wiering, A comparison of feature and pixel-based methods for recognizing handwritten Bangla digits, *The 12th International Conference on Document Analysis and Recognition (ICDAR)*, Washington, D.C., USA, pp.165-169, 2013.

[33] Office of the Permanent Secretary, *Agri-Map Online*, Tech. Rep., Office of the Permanent Secretary for Ministry of Agriculture and Cooperatives, Bangkok, 2019.

[34] X. Yu, X. Wu, C. Luo and P. Ren, Deep learning in remote sensing scene classification: A data augmentation enhanced convolutional neural network framework, *GIScience & Remote Sensing*, vol.54, pp.741-758, 2017.

[35] R. Pari, M. Sandhya and S. Sankar, A multitier stacked ensemble algorithm for improving classification accuracy, *Computing in Science and Engineering*, vol.22, no.4, pp.74-85, 2020.

[36] Q. H. Cap, H. Uga, S. Kagiwada and H. Iyatomi, LeafGAN: An effective data augmentation method for practical plant disease diagnosis, *arXiv.org*, arXiv: 2002.10100, 2020.