


Performance Enhancement of Automatic Wood Classification of Korean Softwood by Ensembles of Convolutional Neural Networks¹

Ohkyung Kwon¹ ^{2,†} · Hyung Gu Lee² · Sang-Yun Yang^{3,4} · Hyunbin Kim³ · Se-Yeong Park^{3,5} · In-Gyu Choi^{3,4,6} · Hwanmyeong Yeo^{3,4}

ABSTRACT

In our previous study, the LeNet3 model successfully classified images from the transverse surfaces of five Korean softwood species (cedar, cypress, Korean pine, Korean red pine, and larch). However, a practical limitation exists in our system stemming from the nature of the training images obtained from the transverse plane of the wood species. In real-world applications, it is necessary to utilize images from the longitudinal surfaces of lumber. Thus, we improved our model by training it with images from the longitudinal and transverse surfaces of lumber. Because the longitudinal surface has complex but less distinguishable features than the transverse surface, the classification performance of the LeNet3 model decreases when we include images from the longitudinal surfaces of the five Korean softwood species. To remedy this situation, we adopt ensemble methods that can enhance the classification performance. Herein, we investigated the use of ensemble models from the LeNet and MiniVGGNet models to automatically classify the transverse and longitudinal surfaces of the five Korean softwoods. Experimentally, the best classification performance was achieved via an ensemble model comprising the **LeNet2**, **LeNet3**, and **MiniVGGNet4** models trained using input images of $128 \times 128 \times 3$ pixels via the averaging method. The ensemble model showed an F1 score greater than 0.98. The classification performance for the longitudinal surfaces of Korean pine and Korean red pine was significantly improved by the ensemble model compared to individual convolutional neural network models such as LeNet3.

Keywords: automatic wood species classification, convolutional neural networks, ensemble methods, LeNet, VGGNet

¹ Date Received February 18, 2019, Date Accepted April 24, 2019

² National Instrumentation Center for Environmental Management (NICEM), Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

³ Department of Forest Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

⁴ Research Institute of Agriculture and Life Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

⁵ Department of Forest Biomaterials Engineering, Kangwon National University, 1 Gangwondaehakgil, Chuncheon 24341, Republic of Korea

⁶ Institutes of Green Bio Science and Technology, Seoul National University, 1447 Pyeongchang-daero, Daehwa-myeon, Pyeongchang 25354, Republic of Korea

[†] Corresponding author: Ohkyung Kwon (e-mail: zoom@snu.ac.kr, ORCID: 0000-0002-6307-0060)

1. INTRODUCTION

Wood species identification is essential for many fields of science, engineering, and industry, but also wooden cultural heritage (Kim and Choi, 2016; Eom and Park, 2018; Lee *et al.*, 2018; Park *et al.*, 2018) in Korea. There are various ways of wood identification by utilizing morphological and spectroscopic features of woods. Yang *et al.* (2017), Park *et al.* (2017), and Yang (2019) proposed wood identification by spectroscopic and chemical characteristics of Korean softwood species. However, still, the most common methods for the wood species identification utilize visual and morphological features of the wood.

There has been a demand for automatic wood species identification by computer-aided machine vision identification systems based on visual and textural features (Koch, 2015). Most of the machine vision identification systems were designed to use in a laboratory environment (Tou *et al.*, 2007; Khalid *et al.*, 2008; Hermanson *et al.*, 2011). Hermanson *et al.* (2013) from USDA Forest Products Laboratory developed XyloTron system that is a field-deployable wood identification system. In more recent years, researchers adapted deep learning techniques for feature extraction and classification of images of wood at various scale. Hafemann *et al.* (2014) developed the convolutional neural network (CNN) models for macroscopic (41 classes) and microscopic (112 species) images of wood. Tang *et al.* (2017) proposed an automatic wood species identification of macroscopic images from 60 species of tropical timbers. Kwon *et al.* (2017) developed an automated wood species identification system for five Korean softwood species. These studies used macroscopic images from the cross-sectional plane of wood and taken by either a digital camera or smartphone camera. Ravindran *et al.* (2018) utilized transfer learning of CNN models to identify 10 neotropical species in the family Meliaceae.

Although the automatic wood identification software successfully classified the five species (cedar, cypress, Korean pine, Korean red pine, and larch) of Korean softwood (Kwon *et al.*, 2017), there are practical limitations of the system stemming from the training images, which are from the transverse plane of the wood species. The transverse surface is very rough and often characteristic pattern of growth rings was hidden by the rough surface. Sometimes the end surface is covered by paint preventing crack development along rays in the transverse surface. At mills, they process lumbers running longitudinal direction, and thus it is hard to catch the transverse surface of the lumber without stopping the process.

These practical limitations of wood species identification at the fields lead us to develop a new model suitable for the longitudinal surfaces from the three principal surfaces of wood. As we investigate patterns on the surfaces of longitudinal plane, the patterns are not as clearly distinguishable as those from the transverse surface. Also, there are considerable variations of patterns due to a mixture of earlywood, latewood, and rays, which are not in the exact orthogonal planes. The vast varieties of patterns cause drops of classification performance of the automatic wood species identification system for the longitudinal surfaces of lumbers.

From our experience of development of the CNN models for the transverse surface, we figured out that different CNN models show better accuracy for different species. In the previous study, we chose LeNet3 model considering overall accuracy and balance between accuracies for each species, not biased to select a model showing the highest accuracy for a species but not for others. It was a close call to choose between LeNet3 and MiniVGGNet3. LeNet3 showed the highest accuracy for cedar, but MiniVGGNet3 showed the best for cypress and larch. However, the accuracy variation of LeNet3 was less than that of MiniVGGNet3.

We expect a decrease of classification performance

of LeNet3 and MiniVGGNet3 for the images from the longitudinal surfaces of lumbers due to complex but less distinctive features for identification. A remedy for the decrease of classification performance is the utilization of a group of predictors called an ensemble (Rosebrock, 2017). Ensemble methods generally refer to training a large number of models and then combining their output predictions via voting or averaging to yield an increase in classification accuracy. By utilizing an ensemble model, it is possible to increase the classification performance by a combination of the strengths of different CNN models.

In this study, we developed an ensemble model for automatic wood species identification system utilizing smartphone images from transverse and longitudinal surfaces of lumbers of five Korean softwood species (cedar, cypress, Korean pine, Korean red pine, and larch). We proposed a method for the selection of an optimal ensemble model. Precision, recall, and F1 score were calculated and constructed a confusion matrix to describe the classification performance of the selected ensemble model.

2. MATERIALS and METHODS

2.1. Sample preparation

Five Korean softwood species [cedar (*Cryptomeria japonica*), cypress (*Chamaecyparis obtusa*), Korean pine (*Pinus koraiensis*), Korean red pine (*Pinus densiflora*), and larch (*Larix kaempferi*)] were under investigation by an automatic wood species identification utilizing ensembles of different CNN models. We purchased fifty lumbers of each species of $50 \times 100 \times 1200 \text{ mm}^3$ (thickness \times width \times length) from several mills participating in the National Forestry Cooperative Federation in Korea. The lumbers in each species were from different regions of Korea.

Images of the transverse surface were from wooden

blocks of $40 \times 50 \times 100 \text{ mm}^3$ (R \times T \times L) prepared from each lumber (50 wood samples per species). For images of the longitudinal surface, we prepared lumber of $40 \times 50 \times 600 \text{ mm}^3$.

2.2. Image acquisition and dataset preparation

We used smartphones (iPhone 7, Samsung Galaxy S3, Samsung Galaxy Tab4 Advanced) to obtain macroscopic pictures of the sawn surfaces of cross sections of the specimen. During the image acquisition process, the smartphones placed on a simple frame as stable support. The camera in an iPhone 7 model has a f/1.8 lens and phase detection autofocus function and produces an image of 12 megapixels. The camera produces a color picture of 3024×4032 pixels. Galaxy S3 model has a f/2.6 lens and autofocus function and produces a color image of 3264×1836 pixels. Galaxy Tab4 Advanced has a camera with 5 megapixel CMOS and autofocus function. It produces a color images of 1280×720 pixels.

We prepared 33,815 images of 512×512 pixels by utilizing a sliding window method; 25,361 images of all (75 %) for training and the other 8,454 images (25 %) for validation. Table 1 lists the number of images for each species corresponding to the wood species and surfaces as well as class name and index. We used the class indices in Fig. 2, 3, and 4 as tick labels in the x axis.

2.3. Ensemble models and methods

An ensemble of CNN models requires several operational CNN models such as LeNet (Lecun *et al.*, 1998) and VGGNet (Simonyan and Zisserman, 2014). We already developed and demonstrated the classification performance of LeNet3 and MiniVGGNet3 models in the previous study (Kwon, *et al.*, 2017). During the development, we investigated the perfor-

Table 1. Class name and its designated index for species and surface combination

Species - Surface	Class name	Class index	Number of images*
Cedar - Transverse	Cedar-C	0	3095
Cedar - Longitudinal	Cedar-L	1	2250
Cypress - Transverse	Cypress-C	2	3420
Cypress - Longitudinal	Cypress-L	3	3090
Korean Pine - Transverse	KoreanPine-C	4	3960
Korean Pine - Longitudinal	KoreanPine-L	5	3952
Korean Red Pine - Transverse	KoreanRedPine-C	6	3060
Korean Red Pine - Longitudinal	KoreanRedPine-L	7	3714
Larch - Transverse	Larch-C	8	3330
Larch - Longitudinal	Larch-L	9	3944

(* Number of images = total number of images for each class.)

Table 2. The architecture of LeNet models.

Layer set	Layer	Parameter
Layer set 1 all LeNet models	Convolution 2D	20, 5 x 5, same, input
	Activation	RELU
	Max pooling	2 x 2, stride = 2 x 2
Layer set 2 all LeNet models	Convolution 2D	50, 5 x 5, same
	Activation	RELU
	Max pooling	2 x 2, stride = 2 x 2
Layer set 3 LeNet2, LeNet3, and LeNet4	Convolution 2D	50, 5 x 5, same
	Activation	RELU
	Max pooling	2 x 2, stride = 2 x 2
Layer set 4 LeNet3 and LeNet4	Convolution 2D	50, 5 x 5, same
	Activation	RELU
	Max pooling	2 x 2, stride = 2 x 2
Layer set 5 LeNet4	Convolution 2D	50, 5 x 5, same
	Activation	RELU
	Max pooling	2 x 2, stride = 2 x 2
	Dense	500
	Activation	RELU
	Dense	10
	Activation	Softmax

mance of variants of LeNet and MiniVGGNet architectures (Table 2 and 3) such as LeNet, LeNet2, LeNet4, MiniVGGNet, MiniVGGNet2, and MiniVGGNet4. Each model showed different performance on the classification of transverse images from five Korean softwood species.

We utilized these eight CNN models to construct ensemble models for classification of two types of lumber surfaces: (1) transverse and (2) longitudinal

surfaces. Each CNN model is trained for the images from the two types of surfaces of the five Korean softwood species. Then we examined combinations of two or three CNN models among the eight CNN models: (1) Sets of two CNN models: 28 combinations (e.g. LeNet-LeNet2, LeNet2-LeNet3, LeNet3-MiniVGGNet2, and so on) and (2) Sets of three CNN models: 56 combinations (e.g. LeNet-LeNet3-MiniVGGNet2, LeNet3-MiniVGGNet-MiniVGGNet3, and so on). The reason

Table 3. The architecture of MiniVGGNet models.

Layer set	Layer	Parameter	Layer set	Layer	Parameter
Layer set 1 all MiniVGGNet models	Convolution 2D	32, 3 x 3, same, input	Layer set 4 MiniVGGNet3 and MiniVGGNet4	Convolution 2D	128, 3 x 3, same
	Activation	RELU		Activation	RELU
	Batch normalization			Batch normalization	
	Convolution 2D	32, 3 x 3, same		Convolution 2D	128, 3 x 3, same
	Activation	RELU		Activation	RELU
	Batch normalization			Batch normalization	
	Max pooling	2 x 2		Max pooling	2 x 2
Dropout	0.25	Dropout	0.25		
Layer set 2 all MiniVGGNet models	Convolution 2D	64, 3 x 3, same	Layer set 5 MiniVGGNet4	Convolution 2D	128, 3 x 3, same
	Activation	RELU		Activation	RELU
	Batch normalization			Batch normalization	
	Convolution 2D	64, 3 x 3, same		Convolution 2D	128, 3 x 3, same
	Activation	RELU		Activation	RELU
	Batch normalization			Batch normalization	
	Max pooling	2 x 2		Max pooling	2 x 2
Dropout	0.25	Dropout	0.25		
Layer set 3	Convolution 2D	128, 3 x 3, same		Dense	512
	Activation	RELU		Activation	RELU
	Batch normalization			Batch normalization	
MiniVGGNet2, MiniVGGNet3, and MiniVGGNet4	Convolution 2D	128, 3 x 3, same		Dropout	0.5
	Activation	RELU		Dense	10
	Batch normalization			Activation	Softmax
	Max pooling	2 x 2			
	Dropout	0.25			

we investigated the performance of various combinations of the CNN models is that ensemble methods are computationally expensive. If fewer CNN models result in sufficient classification performance, it is not necessary to use an excessive number of models.

We applied two methods for prediction results from the ensemble models: (1) averaging and (2) max voting. From the performance measurement, we determined a better method for automatic wood species classification from the transverse and longitudinal surfaces of lumbers.

2.4. Model training and measures of performance

For training the ensemble models, we utilized a workstation with XEON CPU (14 cores) with 64 GB of memory as well as GPU with 24 GB (NVIDIA Quadro M6000). The operating system was Ubuntu 16.04 LTS with CUDA 8.0, Python 3.5, Tensorflow 1.2 and Keras 2.0.

Image patches covering at least several growth rings

from the transverse surface are necessary to utilize macroscopic features of different wood species. We determined the patch size according to the given condition for macroscopic features of all wood species. In the case of the smartphone camera without a zoom factor, the field of view (FOV) in 512×512 pixels was turned out to be a proper size. For other surfaces of lumber, we kept the same FOV to maintain the same feature ratio as in the transverse surface.

The original images were reduced to 64×64 and 128×128 pixels as input images for training purpose. Pixel values of the input images were normalized by 255. Image augmentation was performed as with the following parameters: rotation range = 30, shift range in width and height = 0.1, shear range = 0.2, zoom range = 0.2, and horizontal flip. We utilized the Stochastic Gradient Descent (SGD) algorithm as an optimizer with learning rate = 0.01 and Nesterov momentum = 0.9. The number of epochs was 250.

We evaluated the classification performance of each CNN model as well as various combinations of the

		Predicted class		Measures
		Positive	Negative	
True class	Positive	True Positive TP	False Negative FN	Sensitivity TP/(TP + FN)
	Negative	False Positive FP	True Negative TN	Specificity TN/(FP + TN)
Measures		Positive predictive value (PPV) TP/(TP + FP)	Negative predictive value (NPV) TN/(FN + TN)	Accuracy (TP + TN)/(TP+FP+FN+TN)

Fig. 1. A confusion matrix and their meaning.

CNN models by constructing a confusion matrix. A confusion matrix (or an error matrix) is a table containing summaries of the classification performance of a classifier with respect to some test data. It visually presents the performance as a two-dimensional matrix with the true and the predicted class (Fig. 1). We prepared a confusion matrix to visually display other types of error arisen from different combinations of true and predicted conditions.

From the confusion matrix, we can calculate precision, recall, and F1 score that describe how a model performs. Precision is a positive predictive value, but recall is a true positive rate, sensitivity, or probability of detection. The F1 score is a good indicator of representing the performance of models according to a balance between precision and recall of a model, and we can calculate the following equation. We used F1 scores for comparison purpose for the classification performance of the models and ensemble sets used in this study. The F1 score can be calculated as

$$F1\ score = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

2.5. Selection of an optimal ensemble model

We need to investigate the classification performance

of eight types of ensemble models: (1) size of input image: $64 \times 64 \times 3$ or $128 \times 128 \times 3$, (2) ensemble method: averaging or max voting, and (3) number of CNN models used in an ensemble model: 28 combinations from set of two or 56 combinations from set of three. Due to a large number of ensemble models under investigation, we devised a method to choose an optimal ensemble model with the best performance. A selection of an optimal ensemble model is as followed:

- 1) For each ensemble model, make a confusion matrix with normalized values
- 2) Extract diagonal values from the confusion matrix
- 3) Calculate the mean and standard deviation of the extracted values
- 4) Calculate an SNR-like value as a measure for classification performance: SNR = Signal-to-Noise Ratio
- 5) Select an ensemble model resulting in the highest SNR-like value
- 6) Tabulate the SNR-like values corresponding to the cases to be tested
- 7) Determine an optimal ensemble model with an optimal method and the size of an input image

The SNR-like value can be calculated by an equation: $s = \text{mean}/\text{standard deviation}$. The SNR-like measure describes the degree of misclassification for each class, which treats some misclassified cases as noise - the higher the SNR-like value, the better the classification performance of the ensemble model.

3. RESULTS and DISCUSSION

3.1. Performance of LeNet-type, and MiniVGGNet-type models

All individual models (LeNet1~4 and MiniVGGNet1~4) showed lower classification performance for the ten classes (Table 1) from the transverse and longitudinal surfaces of the five Korean

Performance Enhancement of Automatic Wood Classification of Korean Softwood by Ensembles of Convolutional Neural Networks

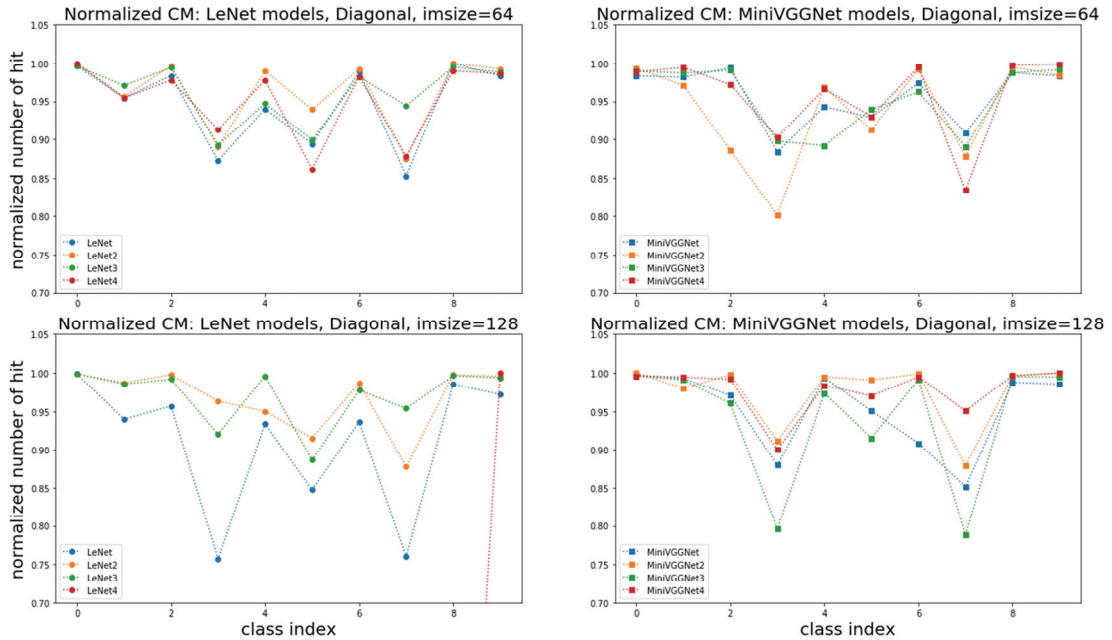


Fig. 2. Comparison of diagonal elements of confusion matrix from LeNet-type and MiniVGGNet-type models for different sizes of input images.

softwood species than that for the five classes from the transverse. Especially, classification performance for the longitudinal surfaces was significantly lower than for the transverse surfaces. Average and standard deviation of the recall values were LeNet models with $64 \times 64 \times 3$ images: (0.956, 0.046), $128 \times 128 \times 3$ images: (0.736, 0.401), MiniVGGNet models with $64 \times 64 \times 3$ images: (0.951, 0.050), $128 \times 128 \times 3$ images: (0.961, 0.055), (average, standard deviation) respectively. MiniVGGNet models trained with $128 \times 128 \times 3$ images showed better performance than LeNet models trained with $64 \times 64 \times 3$ images (Fig. 2).

LeNet models did not show significant differences for $64 \times 64 \times 3$ images but showed significant variations in performance trained with $128 \times 128 \times 3$ images (Fig. 2, left-top and left-bottom graphs). On the other hand, MiniVGGNet models showed large variations of performance trained with $128 \times 128 \times 3$ images. However, the best classification performance was

from a MiniVGGNet model trained with $128 \times 128 \times 3$ images (Fig. 2, right bottom graph).

From the normalized diagonal elements of the confusion matrices for the CNN models, we expected that combinations of LeNet2, LeNet3, MiniVGGNet2, and MiniVGGNet4 models would result in the excellent classification performance as ensemble models trained with input images of $128 \times 128 \times 3$. However, for the input images of $64 \times 64 \times 3$, it was difficult to predict an optimal combination of the individual CNN models would result in the best performance.

3.2. Performance of the ensemble models

Between the ensemble methods, there was no significant difference, but the averaging method showed marginally better performance than the voting method did (Fig. 3 and 4). Thus, the averaging method as an optimal ensemble method.

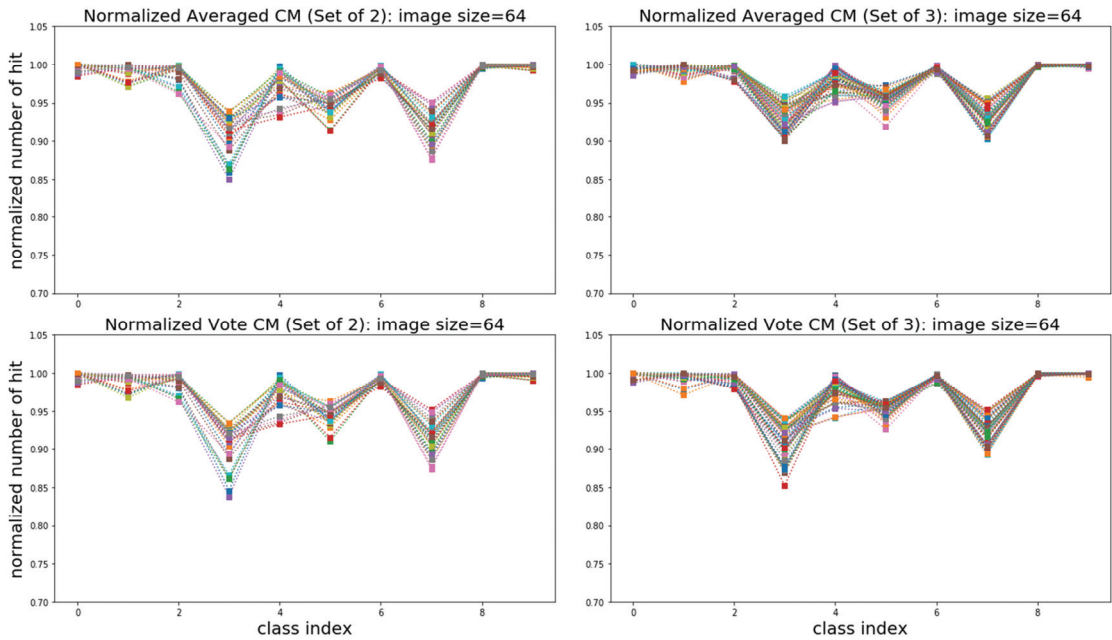


Fig. 3. Normalized diagonal values from confusion matrices by ensemble models for input images of $64 \times 64 \times 3$.

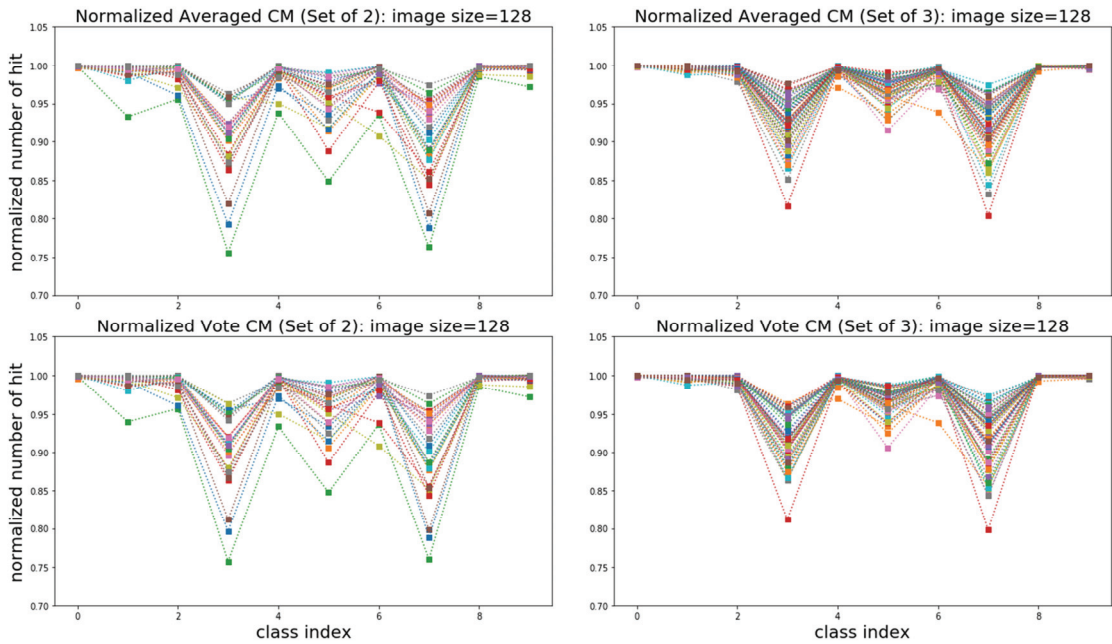


Fig. 4. Normalized diagonal values from confusion matrices by ensemble models for input images of $128 \times 128 \times 3$.

Table 4. The best results by SNR-like measure from ensemble sets for input image of $64 \times 64 \times 3$.

Ensemble model	Averaged	Voting
LeNet3 and MiniVGGNet3 (averaged, voting)	36.55	33.87
LeNet3, LeNet4, and MiniVGGNet3 (averaged) LeNet2, LeNet3, and MiniVGGNet3 (voting)	51.95	38.63

Table 5. The best results by SNR-like measure from ensemble sets for input image of $128 \times 128 \times 3$.

Ensemble model	Averaged	Voting
LeNet2 and MiniVGGNet4 (averaged, voting)	59.74	55.89
LeNet2, LeNet3, and MiniVGGNet4 (averaged, voting)	69.79	61.29

In general, the effect of input image size on the performance of the ensemble models was significant. For the ensemble models with 3 CNN models, ensemble models with input images of $64 \times 64 \times 3$ showed less variable than those with $128 \times 128 \times 3$ images did in performance of the ensemble models (Fig. 3 and 4). However, the highest performance was found from the results with $128 \times 128 \times 3$ images. Since input images of $64 \times 64 \times 3$ did not show significant differences in classification performance corresponding to the ensemble methods and the number of CNN models in an ensemble model, we determined the size of input images as $128 \times 128 \times 3$.

The ensemble models showed a similar pattern to the individual LeNet and MiniVGGNet models, but the increase of overall performance (Fig. 2, 3 and 4). Regardless of the sizes of input images and the ensemble methods (averaging and voting), ensemble models with 3 CNN models showed better performance than those with 2 CNN models (Fig. 3 and 4).

We determined the SNR-like measure for input images of $64 \times 64 \times 3$ (Table 4) and $128 \times 128 \times 3$ (Table 5). The ensemble models with 3 CNN models by the averaging method showed the best performance. The ensemble models with 3 CNN models by the

averaging method showed the best performance. Among the ensemble models with 2 and 3 CNN models, the best classification performance resulted by the combinations of [LeNet3-MiniVGGNet3] and [LeNet3-LeNet4-MiniVGGNet3] models trained with $64 \times 64 \times 3$ images, but those by [LeNet2-MiniVGGNet4] and [LeNet2-LeNet3-MiniVGGNet4] models trained with $128 \times 128 \times 3$ images.

Difference between [LeNet3-MiniVGGNet3] and [LeNet3-**LeNet4**-MiniVGGNet3] models was **LeNet4** model, which is an additional CNN model to the ensemble of 2 CNN models. Similarly, difference between [LeNet2-MiniVGGNet4] and [LeNet2-**LeNet3**-MiniVGGNet4] was **LeNet3** model. We thought that LeNet4 and LeNet3 models increased the classification performance of the ensemble models with 2 CNN models.

Based on the new measure (SNR-like value), we determined the LeNet2-LeNet3-MiniVGGNet4 model as the best ensemble model, but we need to investigate other aspects of the classification performance of other ensemble models by confusion matrix. Comparison of the F1 scores from the selected ensemble models according to the SNR-like measures showed no significant difference among the models (Fig. 5). However,

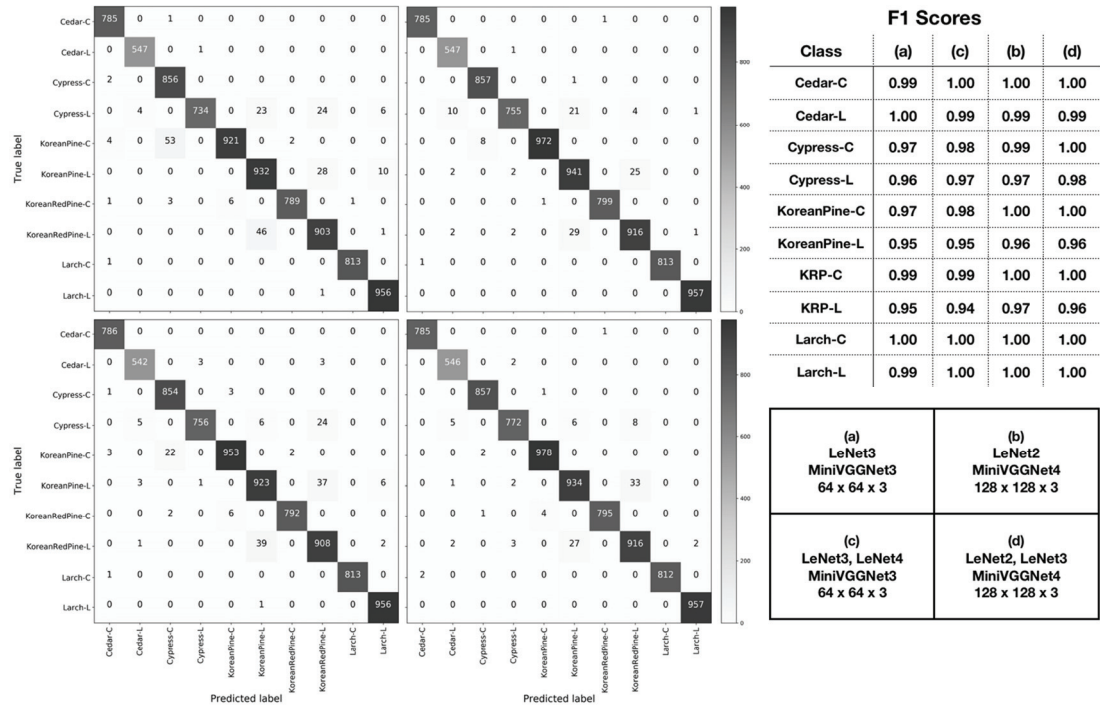


Fig. 5. Confusion matrices and F1 scores of several ensemble models by the averaging method.

Table 6. Performance measures of LeNet2-LeNet3-MiniVGGNet4 ensemble model by the averaging method.

Class name	Precision	Recall	F1 score	Support
Cedar-C	1.00	1.00	1.00	786
Cedar-L	0.99	1.00	0.99	548
Cypress-C	1.00	1.00	1.00	858
Cypress-L	0.99	0.98	0.98	791
KoreanPine-C	0.99	1.00	1.00	980
KoreanPine-L	0.97	0.96	0.96	970
KoreanRedPine-C	1.00	0.99	1.00	800
KoreanRedPine-L	0.96	0.96	0.96	950
Larch-C	1.00	1.00	1.00	814
Larch-L	1.00	1.00	1.00	957

the confusion matrices revealed the distribution of false negative (FP) cases, and it was possible to determine the best ensemble model from their FP distributions. In Fig. 5, the ensemble models with two CNN models (a and b) showed more misclassification cases than those with three CNN models (c and d). Between the

two ensemble models with three CNN models (c and d), the ensemble model (d) (trained with input images of $128 \times 128 \times 3$) showed less misclassifications than (c) (trained with input images of $64 \times 64 \times 3$). Therefore, we confirmed that the best ensemble model among the ensemble models under investigation in this study was

Table 7. Performance measures (F1 scores) of the best ensemble model (LeNet2, LeNet3, and MiniVGGNet4) by the averaging method and individual CNN models (LeNet3 and MiniVGGNet3).

Class name	LeNet3	MiniVGGNet3	LeNet2 + LeNet3 + MiniVGGNet4
Cedar-C	0.99	0.98	1.00
Cedar-L	0.98	0.94	0.99
Cypress-C	0.99	0.97	1.00
Cypress-L	0.94	0.98	0.98
KoreanPine-C	0.99	0.97	1.00
KoreanPine-L	0.92	0.84	0.96
KoreanRedPine-C	0.99	1.00	1.00
KoreanRedPine-L	0.90	0.85	0.96
Larch-C	1.00	1.00	1.00
Larch-L	0.99	0.97	1.00

the [LeNet2-LeNet3-MiniVGGNet4] model.

The measures of classification performance of the [LeNet2-LeNet3-MiniVGGNet4] model was listed in Table 6. Overall performance was excellent, but classification performance for the longitudinal surface of Korean pine and Korean red pine was lower than that for other surface images. However, the performance was significantly improved by the ensemble model compared to individual CNN models such as LeNet3 and MiniVGGNet3 (Table 7) for the longitudinal surfaces of Korean pine and Korean red pine.

4. CONCLUSION

In this study, we investigated the use of ensembles from LeNet and MiniVGGNet models to automatically classify the transverse and longitudinal surfaces of five Korean softwoods (cedar, cypress, Korean pine, Korean red pine, and larch). Images from cameras in mobile devices such as smartphone and tablet were used to provide macroscopic images to the ensemble models.

The experimental results showed that the best model was the ensemble model by the averaging method consists of [LeNet2-LeNet3-MiniVGGNet4] models trained with input images of $128 \times 128 \times 3$. The ensemble model showed F1 score of > 0.96 . Classi-

fication performance for the longitudinal surface of Korean pine and Korean red pine was significantly improved by the ensemble model compared to individual CNN models such as LeNet3.

ACKNOWLEDGMENT

The Forest Science & Technology Projects (Project No. 2016009D10-1819-AB01) provided by the Korea Forest Service financially supported this study.

REFERENCES

- Eom, Y., Park, B. 2018. Wood Species Identification of Documentary Woodblocks of Songok Clan of the Milseong Park, Gyeongju, Korea. *Journal of the Korean Wood Science and Technology* 46(3): 270-277.
- Hafemann, L.G., Oliveira, L.S., Cavalin, P. 2014. Forest Species Recognition Using Deep Convolutional Neural Networks. *22nd International Conference on Pattern Recognition (ICPR)*, pp. 1103-1107.
- Hermanson, J.C., Wiedenhoef, A.C. 2011. A brief review of machine vision in the context of automated wood identification systems. *IAWA Journal* 32(2): 233-250.

- Hermanson, J., Wiedenhoef, A.C., Gardner, S. 2013. A machine vision system for automated field-level wood identification. Global Timber Tracking Network. Presentation at Regional Workshop for Asia, Pacific and Oceania on identification of timber species and origins, Beijing, China.
- Khalid, M., Lee, E.L.Y., Yusof, R., Nadaraj, M. 2008. Design of an intelligent wood species recognition system. *International Journal of Simulation System, Science and Technology* 9(3): 9-19.
- Kim, S.C., Choi, J. 2016. Study on Wood Species Identification for Daeungeon Hall of Jeonghyesa Temple, Suncheon. *Journal of the Korean Wood Science and Technology* 44(6): 897-902.
- Kwon, O., Lee, H.G., Lee, M.-R., Jang, S., Yang, S.-Y., Park, S.-Y., Choi, I.-G., Yeo, H. 2017. Automatic Wood Identification of Korean Softwood Based on Convolutional Neural Networks. *Journal of Korean Wood Science and Technology* 45(6): 797-808.
- Lecun, Y., Bottou, L., Bengio, Y., Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11): 2278-2324.
- Lee, K., Seo, J., Han, G. 2018. Dating Wooden Artifacts Excavated at Imdang-dong Site, Gyeongsan, Korea and Interpreting the Paleoenvironment according to the Wood Identification. *Journal of the Korean Wood Science and Technology* 46(3): 241-252.
- Park, J.H., Oh, J.E., Hwang, I.S., Jang, H.U., Choi, J.W., Kim, S.C. 2018. Study on Species Identification for Pungnammun Gate (Treasure 308) in Jeonju, Korea. *Journal of the Korean Wood Science and Technology* 46(3): 278-284.
- Park, S.Y., Kim, J.C., Kim, J.H., Yang, S.Y., Kwon, O., Yeo, H., Cho, K., Choi, I.G. 2017. Possibility of Wood Classification in Korean Softwood Species Using Near-infrared Spectroscopy Based on Their Chemical Compositions. *Journal of the Korean Wood Science and Technology* 45(2): 202-212.
- Ravindran, P., Costa, A., Soares, R., Wiedenhoef, A.C. 2018. Classification of CITES-listed and other neotropical Meliaceae wood images using convolutional neural networks. *Plant Methods* 14:25
- Rosebrock, A. 2017. Deep Learning for Computer Vision with Python: Practitioner Bundle. PyImageSearch, pp. 73-83.
- Simonyan, K., Zisserman, A. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv.org. cs.CV.*
- Tang, X.J., Tay, Y.H., Siam, N.A., Lim, S.C. 2017. Rapid and Robust Automated Macroscopic Wood Identification System using Smartphone with Macro-lens. *arXiv:1709.08154v1 [cs.CY]*
- Tou, J.Y., Lau, P.Y., Tay, Y.H. 2007. Computer vision-based wood recognition system. In *Proceedings of International workshop on advanced image technology.*
- Yang, S.Y., Park, Y., Chung, H., Kim, H., Park, S.Y., Choi, I.G., Kwon, O., Yeo, H. 2017. Soft Independent Modeling of Class Analogy for Classifying Lumber Species Using Their Near-infrared Spectra. *Journal of the Korean Wood Science and Technology* 47(1): 101-109.
- Yang, S.Y. 2019. Classification of Wood Species using Near-infrared Spectroscopy and Artificial Neural Networks, Doctoral dissertation, Seoul National University