.

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Development of Korean Food Image Classification Model Using Public Food Image Dataset and Deep Learning Methods

MINKI CHUN[1], HYEONHAK JEONG[2], HYUNMIN LEE[3], TAEWON YOO[4], AND HYUNGGU JUNG[5]

[1]University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, Seoul 02504 Republic of Korea (e-mail:cna81136@gmail.com)
[2]University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, Seoul 02504 Republic of Korea (e-mail:drummycrane@gmail.com)
[3]University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, Seoul 02504 Republic of Korea (e-mail:yunoa64@gmail.com)
[4]Kyung Hee University, 1732 Deogyeong-daero, Giheung-gu, Yongin-si 17104 Republic of Korea (e-mail:twyoo@khu.ac.kr)
[5]University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, Seoul 02504 Republic of Korea (e-mail:hjung@uos.ac.kr)

Corresponding author: Hyunggu Jung (e-mail: hjung@uos.ac.kr).

This paragraph of the first footnote will contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

ABSTRACT Food image classification is useful in diet management apps for personal health management. Various methods for classifying food images in a particular country have been proposed in extant studies. However, knowledge of Korean food image classification is limited. The objective of this study was to train a classification model for Korean food images. To train the classification model, we collected Korean food images from the AI hub, a public food image dataset. The images were preprocessed and augmented for model training. The proposed model was evaluated in an experiment using a Korean food image dataset and the proposed model effectively classified Korean food images based on transfer learning. The findings of the study revealed that the model's performance in classifying food images depended on the type of food. The findings have implications for the classification model training process using CNN and the Korean public food image dataset. Future work is required to improve the performance of a classification model, especially as it pertains to its poor performance for some food image types.

INDEX TERMS computer vision, convolutional neural networks, image classification, machine learning, supervised learning

I. INTRODUCTION

FOOD image classification is useful in diet management apps for personal health management. For example, the mobile food image classification used in diet management apps enables users to determine their daily dietary requirements and calorie intake [1], [2]. However, food classification based on image processing and computer vision difficulties challenging because the visual differences between food classes are not significant [3]. Furthermore, Asian foods, including Korean foods, generally have more diverse recipes than Western foods [4]. It is also difficult to define the structure of food because food can be deformed, such as by placing or hiding some ingredients in entirely different positions, depending on the cooking method [5]. Therefore, it is not easy to specify a class of food through image classification because foods may be visually similar to foods of other classes and visually different from foods in the same class.

Prior researchers have attempted to solve the difficult task of food image classification using traditional machine learning (ML) or deep learning methods. The traditional ML methods that have been used include 1) color histogram and, Gabor texture method for feature extraction, 2) combining an individual's diet pattern with the results of image analysis, 3) using neural networks; 4) support vector machine (SVM)-based image segmentation, and 5) using food images of specific domains [6]–[10]. Deep learning methods are widely used to train food image classification models. The deep learning-based methods differ according to whether fine-tuning was performed [11] or not. Other researchers have trained a classification model using fine-tuning. However, some of them proposed a system using the model and others did not [12], [13].

Nevertheless, none of these researchers have proposed a method for Korean food image classification using the public Korean food image dataset. Thus, a food image classification model that classifies more than 100 types of Koran food was developed in this study using a public food image dataset. In addition, although Korean food belonging to the same food group have similar characteristics, the differences between food groups have been simplified or ignored in previous studies. In this study, we considered the differences between the 27 food groups defined by the Korea Culture and Information Service[1]. Analysis of the Korean food image dataset is still in its infancy. In this study, we examined the characteristics of Korean food image datasets that affect the classification accuracy.

We trained a model based on the convolutional neural network (CNN) that classifies 150 classes of Korean food images using a Korean food image dataset. First, we obtained 150 classes of Korean food image datasets from the AI hub [14] operated by the National Information Society Agency[2]. Second, we reduced model complexity and enhanced accuracy through image preprocessing. Subsequently, we augmented the preprocessed image set to prevent overfitting. Third, we extracted features using augmented image datasets and pre-trained deep neural networks (DNNs) (i.e., ResNet-50, ResNet-101, ResNet-152, MobileNetV2, InceptionResNetV2, and NasNetLarge) and fine-tuned our trained model. We also used the max-pooling, density, and dropout layers to train the classification model. We then compared the performance (e.g., accuracy and training time) of the DNNs. Hence, our study makes the following contributions to the target users of the model, which automatically classifies visually similar Korean food images, and the community of researchers who use the Korean food image dataset:

- We present a classification model that automatically classifies images of the visually similar classes of Korean food into 150 food classes.
- We present a preprocessed dataset that for training Korean food image classification models.
- We present the accuracy and learning time of several pre-trained DNNs in training a model to classify Korean food images into 150 food classes.

II. RELATED WORK

Previous researchers have attempted to classify food images using ML algorithms using traditional ML or DL. We also found studies that created and evaluated food image classification models in specific domains.

[1]https://www.kocis.go.kr/
[2]https://eng.nia.or.kr/

A. TRADITIONAL ML-BASED FOOD IMAGE CLASSIFICATION

ML-based food image classification models are characterized based on whether they consider domains or not. Researchers that did not consider domain classified food images using support vector machine (SVM) [15] and random forest (RF) method [16], instead of DL methods [1], [6]–[9], [17]–[21]. Various methods, such as color histograms and Gabor texture features, have been used to extract features from food images. Among them, in four studies, the researchers collected food images needed for the classification model training directly from participants or the web [6], [17]–[19]. Hoashi et al. [6] showed that the SVM classifies food images into several groups through multi-kernel learning. They developed an automatic food image recognition system for 85 food classes by converging various image features using a bag-of-features, color histogram, Gabor feature, or gradient histogram. Similarly, Beijbom et al. [18] extracted features from food images using color, histogram of oriented gradients, scale-invariant feature transforms, local binary patterns, and filter responses from the MR9 filter bank. They encoded the obtained features with locality-constrained linear encoding via k-means clustering and used SVM to classify the features from the images. Wang et al. [17] collected 1453 food images before meals and remnants from 45 people and classified food images into 56 classes. Dalakleidi et al. [19] created a food image classification model by collecting food images containing six classes from the web. They used speed-up robust features, color averages of images, and local binary pattern features to extract features from images and classify food images using an SVM. On the other hand, in another study, researchers trained a food image classification model using a PFID dataset consisting of fast-food images. Farinella et al. [20] extracted features using a texton descriptor and classified seven food classes using SVM classifiers.

Researcher has established that the accuracy of food image classification can be improved by combining personal meal patterns with image analysis results. In each study, the researchers trained a model to classify food images using the UNICT-FD889 dataset [1], [7]. Farinella et al. [7] used descriptors such as the bag of scale-invariant feature transform [22] and classified food and non-food images using the SVM. Likewise, McAllister et al. [1] compared the food image classification performance of the classical ML model, including RF and the neural-network-based classification model. RF entails using ensembles of several decision trees.

Researchers have reported that the classification model using neural networks showed high performance in food image classification. In one study, researchers collected food image data online. Minija and Emmanuel [8] calculated food calories from food images collected online using an SVM with a scatter search approach (SS-SVM). Other researchers collected food image data by asking participants to take pictures of food. Zhu et al. [21] developed an algorithm that automatically recognizes areas with specific food in images and identifies food based on its characteristics.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3227796

Another group researchers proposed a food image classification method using image segmentation results and SVM. Inunganbi et al. [9] developed a system for classifying food images tailored to the nutritional needs of patients with diabetes. They proposed food image segmentation using RF, extracted image features, and classified them using an SVM.

On the other hand, some researchers have focused on domain-based food image classification [10], [23]–[25]. Joutou and Yanai [23] collected 50 classes of food images from the web, including Japanese food. They proposed a classification model that extracts features using bag-of-features, color histograms, Gabor texture features, and gradient histograms and classifies features with SVM. Yang et al. [10] collected 18 classes of Chinese food images using a mobile phone. They created a local descriptor by extracting the keypoints of the image using the scale-invariant feature transform algorithm and generated RGB histograms. The RF method was also used to classify food images [24], [25]. Fahira et al. [24] created and evaluated traditional Indonesian food image classification models using several classical ML methods, including the RF. They found that the RF was the best at classifying traditional Indonesian food images in Java. Fahira et al. [25] reported that classical ML methods, such as RF methods can predict food images with similar accuracy, although less effectively than DL methods.

B. DEEP LEARNING-BASED FOOD IMAGES CLASSIFICATION

Prior researchers have created and evaluated deep learning-based food image classification models [11]–[13], [26]–[38]. Deep learning-based food image classification methods are widely used because they enable various systems to extract and classify image features. Previous studies using deep learning methods to classify food images are categorized based on whether fine-tuning was performed or not, as well as whether a system was proposed or not. The first category of researchers did not use fine-tuning in the process of training a food image classification model through deep learning, and did not propose a system using the trained model [11], [26], [27]. In two studies, the researchers created new food image datasets for training image classification models [11], [26]. Samraj et al. [11] developed a visual food classification system to prevent errors in handling large amounts of food. They created an image classification model that classifies 170 food image classes. Meanwhile, Christodoulidis et al. [26] proposed a method of extracting patches from food images and then feeding them to a CNN. The method classified food images using the CNN's output and a voting scheme. Other researchers trained a food image classification model using the UEC FOOD 100 dataset consisting of 100 types of food images. Memiş et al. [27] compared the performance of several deep learning image classification models that classify 100 categories of food images using the UEC FOOD 100 dataset, and reported that ResNet-50 showed the best classification performance. [39].

The second category trained a deep learning-based classification model using fine-tuning; however, they did not propose a system using this model [12], [28]–[32]. In four studies in which a deep learning model was trained for image classification, some food image datasets (e.g., Food-11, Food-475) were used for training the image classification model [12], [28]–[30]. Islam et al. [12] developed a CNN model to classify food images using the food-11 dataset. Researchers have reported that the performance of the pre-trained Inception V3 model was relatively high when comparing the performance of their proposed model and other models using the Food-11 dataset [3], [40]. Similarly, Ciocca et al [28] developed a CNN-based food image classification model with residual network using Food-475. Researchers used Food-475 with a combination of food images and duplicates in the Food-524 database through syntactic analysis and reported that the more the data in the food domain, the more accurate the recognition of features obtained from the CNN trained on the database. Pan et al. [29] proposed DeepFood, a new framework that uses deep learning to extract rich and effective features from a dataset of food ingredient images and improves the average accuracy of multiclass classification using ML. They extracted image features using deep learning from datasets with 41 image classes, following which they selected features using principal component analysis, correlation-based feature selection, and integrated gradient methods. They then trained classification models using sequential minimal optimization (SMO). They reported that DeepFood effectively integrated an optimized classifier called deep learning and SMO. Additionally, Özsert et al. [30] proposed pre-trained deep-CNN (DCNN) structures in INISTA 2017 and compared their performance with pre-trained structures such as Alexnet and Capenet [41], [42]. They reported that learning optimization improved the performance of all comparative models. On the other hand, the researchers collected food images in a restaurant or constructed a dataset with food images taken by users using an app that allows collecting food images [31], [32]. Zhou and Lin [31] proposed a new food image classficaition based on the abundant relationships in the boost graph library algorithm. They proposed a model that modified the last fully connected layer to the bipartite graph label layer in the CNN model, and reported that the model proposed in a new food image benchmark dataset consisting of 975 menus showed high performance. Likewise, Kagaya et al. [32] applied CNN for food detection and recognition. They showed that CNN-based food image detection is significantly more accurate than traditional SVM-based methods.

The third category of researchers trained a food image classfication model through deep learning with fine-tuning, and proposed a system using this model [13], [33]. Liu et al. [13] proposed a network architecture called an inception module for food recognition. They developed a system that divides images captured using this module into blurry and clear images and segments the background. Similarly, Meyers et al. [33] presented a system that recognizes multiple foods in an image and predicts the nutritional content of a meal. They used the GoogLeNet CNN model, a multi-label classifier, a DeepLab model for food content segmentation, and CNN architecture from previous research by Eigen and Fergus [43] for volume estimation. Meyers implemented a mobile app that recognizes food from an image using a multi label classifier.

Other researchers have proposed using deep learning methods for the food image classification of specific domains such as Brazil and Japan [34]–[38]. Among them, some have proposed food image classification models using food images of Australia, Brazil, Japan and America, respectively [34]–[36]. Islam et al. [34] created a food image classifier using a food image database based on Australian dietary guidelines. They classified food images into 22 classes through transfer learning using models that learned ImageNet and reported that the model using transfer learning and deep CNN (DCNN) classified Australian food images with high accuracy. Similarly, Freitas et al. [35] reconstructed the image datasets of the 50 most consumed food by Brazilians to create food image datasets with 10 classes. Researchers have created food image classification models using state-of-the-art deep learning methods and reported that, based on performance comparison, mask-RCNN yielded relatively good results. Likewise, Ege and Yanai [36] simultaneously learned food calories, classes, ingredients, and cooking directions using deep learning to train a classification model that estimates food calories with food images collected from Japanese and American food recipe sites. They reported that the multi-task CNN-based method effectively estimates food calories. In two studies, researchers proposed a deep learning model to classify Korean food images [37], [38]. They collected images of Korean food by taking pictures or searching for images on the Internet. Park et al. [37] collected more than 4,000 Korean food images consisting of 23 food groups, and trained a DCNN model, K-foodNet, to recognize food images. Yang et al. [38] collected 6462 Korean food images from 18 food classes. They employed a pre-trained DenseNet to create a double-layered hierarchical deep model that classifies 18 food classes into four groups in the first stage, and classifies each food class in the second stage.

C. LIMITATIONS OF PRIOR STUDIES

However, the review of previous studies reveals a few limitations. First, although previous researchers proposed image classification models trained using food from various countries, only few used the Korean food image dataset. Moreover, researchers who developed the Korean food image classifiers have pointed out that the dataset was insufficient [?], [38]. We trained our model on a larger dataset consisting of 27 food categories, including 150 food classes, which may lead to the discovery of additional challenges and opportunities in classifying Korean food images. Moreover, to our knowledge, this is the first attempt to train a model that classifies Korean food images using a public Korean food image dataset consisting of 150,000 images. Second, understanding of the classification performance for each Korean food category is limited. Although Korean foods in the same food group have similar characteristics, food groups have been simplified or ignored in previous studies. The dataset used in this study contained 27 food categories defined by the Korean Culture and Information Service [44]. Few models for classifying Korean food images classification have been proposed in a previous study. However, the analysis of the Korean food image dataset is insufficient. Therefore, we examined the characteristics of the Korean food image dataset that affect classification accuracy. Investigating food images is expected to provide researchers with opportunities to improve Korean food classification models.

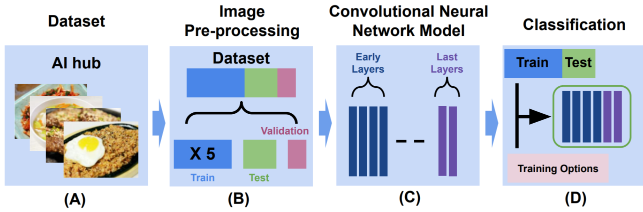


FIGURE 1. Training Process of Classification Model. (A) Collect Korean food images from the AI hub. (B) Preprocess image to reduce the complexity of the model network, increase the classification accuracy, and prevent overfitting. (C) Load pre-trained feature extraction network and add last classifier layers. (D) Initial training and fine-tuning the trained model.

III. METHODS

The objective of this study was to develop a classification model for Korean food image. We collected images from a Korean food image database to train a model forclassifying Korean food images. We obtained image data for model training through image preprocessing and augmentation. The classification model was then trained using the obtained data. Figure 1 shows the training process of the classification model. Each light blue box matches one to one with subsections in the Methods section.

A. DATASET

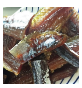
We used the Korean food image dataset provided by the AI hub [14]. The dataset was provided to develop health management applications that automate food image identification and daily nutrient intake calculations. The AI Hub provides 150 classes of Korean food images belonging to 27 food groups. Table 1 lists these food groups. The total number of food images we collected from the AI hub was 150,610. Every food class in the database consists of a minimum of 1,000 images and a maximum of 1,130 images.

B. IMAGE PRE-PROCESSING

We reduced the complexity of the model network and increase classification accuracy by preprocessing the food images before training the CNNs. We split the entire image dataset into 105,427 images for training, 30,122 images for testing, and 15,061 images for validation. All the images were resized to 331 pixels × 331 pixels to fit the input of the model.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3227796

TABLE 1. List of Food Groups in Korean Food Data.

#	Image	Group Name	Description	#	Image	Group Name	Description
1		Jeongol	A dish in which the ingredients are added to the broth and boiled; the ingredients are eaten first.	15		Myeon	Dough made from various flour is pressed to make it thin, and then boiled or fried in various ways.
2		Jjigae	A dish made with less broth than stew or soup and boiled with various ingredients such as meat, vegetables, and seafood.	16		Bokkeum	A dish in which food ingredients are seasoned and cooked while stirring in oil or soy sauce.
3		Tang	A thick broth made by boiling soup longer than normal.	17		Jang	Fermented savory condiments such as red pepper paste, soybean paste, and soy sauce
4		Twigim	Food cooked in boiling oil.	18		Bab	Rice and barley grains are boiled in water such that they do not lose moisture or become loose.
5		Haemul	Food made with animals or plants from sea or the river.	19		Kimchi	Vegetables marinated in brine and stored for a long time.
6		Jjim	Food made by boiling various ingredients such as seafood, meat, and vegetables, in a little water for a long time or by boiling them with hot steam.	20		Hangwa	Confectionery made by sweetening grain flour, fruit, edible roots or leaves with honey, syrup, sugar.
7		Muchim	Food cooked with various condiments such as vegetables, dried fish or seaweed.	21		Tteok	Food made by grinding, steaming, boiling, or frying grains in oil
8		Jolim	Food made by boiling seasoned meat, fish, or vegetables such that almost no broth remains.	22		Jeok	Fish or meat seasoned and grilled on a skewer
9		Mandu	Food made by kneading wheat flour with stuffing, steamed, boiled or fried.	23		Jangajji	Food made with vegetables in paste or sauce and preserved for a long time to eat.
10		Guk	Food made by adding a lot of water to vegetables, meat or fish.	24		Hoe	Food made to be eaten raw such as fish, shellfish, beef flesh, or liver
11		Ssam	A food made by wrapping meat, rice, miso in leafy vegetables	25		Eumcheong-ryu	A general term for beverages other than alcohol
12		Gui	Seasoned and grilled meat or fish.	26		Namul	Food prepared by seasoning plants or vegetables harvested from the mountains or fields.
13		Others	Food not included in 27 categories	27		Juk	Food made with grain as the main ingredient, boiled in water to make it semi-fluid.
14		Jeon	Thinly sliced or minced fish, meat, or vegetables, seasoned, coated with flour and egg, and then fried in oil.				

We applied data augmentation using random seeds to our training dataset to obtain additional data. The images obtained by data augmentation were as follows: 1) 105,427 original training images; 2) 105,427 images with brightness changed between 0.5 and 1.5 times using random seeds; 3) 105,427 images left in which we changed the saturation from 0.5 to 0.8 using random seeds; 4) 105,427 images in which we changed the contrast from 1.2x to 1.7x using a random seed; 5) 105,427 images flipped horizontally. The training data used for model training were a total of 527,135 images, a five-fold increase from 105,427 images selected for use in model training.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3227796

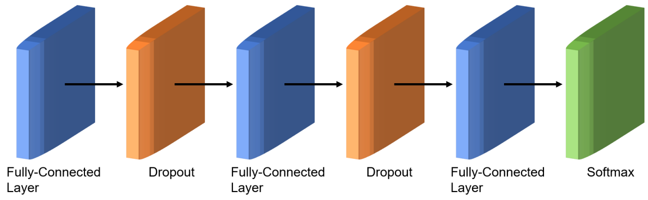


FIGURE 2. Architecture of the Classifier Layer.

TABLE 2. Evaluation Accuracy of Each Model used in this Study. InceptionResNetV2 and ResNet-50V2 showed the highest and lowest evaluation accuracy, respectively.

Model	Accuracy	Training Time (s)
InceptionResNetV2	**0.8191**	436182
NasNetLarge	0.7791	294853
MobileNetV2	0.7536	82366
ResNet-101V2	0.737	51649
ResNet-152V2	0.7131	82458
ResNet-50V2	0.6827	42915

C. CONVOLUTIONAL NEURAL NETWORK MODEL

We employed a CNN, which is widely used to classify objects in an image, to extract image features. In this study, ResNet-50V2, ResNet-101V2, ResNet-152V2 [45], InceptionResNetV2 [46], NasNetLarge [47], and MobileNetV2 [48] models pretrained with ImageNet data were used to classify Korean food. We connected the pretrained feature extraction network to the classifier network, which consists of global average pooling layers, fully connected layers, and dropout layers. The classifier structure is as follows (see Figure 2).

D. CLASSIFICATION

The classification models were trained through two steps: initial training and fine-tuning. For the initial training step, we trained only the parameter of the classifier; the parameter of the feature extraction network was fixed. After the initial training, we trained all the parameters of the classification model, including the feature extraction network and classifier, in the fine-tuning step. The Adam optimizer was adopted for the objective function of each training. Through the processes mentioned in the previous sections, we developed a model that classifies 150 classes of Korean food images.

IV. RESULTS

A. MODEL EVALUATION

The trained model was evaluated using images from the evaluation set. We used ten percent of the Korean food images collected from the AI hub for model performance evaluation. The accuracy was measured using the following formula:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

where TP is the true positive, TN is the true negative, FP is the false positive, and FN is the false negative.

TABLE 3. Classification Recall, Precision and F1-score of 27 Korean food groups in descending order based on recall.

Group Name	Recall	Precision	F1-score
Juk	0.958	0.942	0.95
Namul	0.939	0.908	0.924
Eumcheong-ryu	0.938	0.916	0.927
Hoe	0.936	0.86	0.896
Jangajji	0.935	0.899	0.917
Jeok	0.918	0.885	0.901
Tteok	0.903	0.914	0.909
Hangwa	0.889	0.888	0.889
Kimchi	0.879	0.878	0.878
Bab	0.874	0.885	0.879
Jang	0.835	0.852	0.844
Bokkeum	0.834	0.784	0.808
Myeon	0.833	0.826	0.829
Jeon	0.829	0.795	0.812
Others	0.814	0.867	0.84
Gui	0.804	0.792	0.798
Ssam	0.795	0.621	0.697
Guk	0.792	0.816	0.804
Mandu	0.785	0.84	0.811
Jolim	0.78	0.79	0.785
Muchim	0.777	0.816	0.796
Jjim	0.762	0.806	0.784
Haemul	0.743	0.935	0.828
Twigim	0.731	0.755	0.742
Tang	0.73	0.751	0.74
Jjigae	0.683	0.671	0.677
Jeongol	0.58	0.686	0.629

TABLE 4. Number of accurate predictions from the evaluation result of the model based on InceptionResNetV2. The number of test samples for each food class is 200.

Food Class	Correct Predictions
Spinach Salad	200
Beef Tartare	196
Braised Lotus Roots	199
Pollack Jjigae	99
Maeuntang	95
Gopchang jeongol	116

B. OBTAINED RESULTS

We evaluated the classification performance of each model using an evaluation image set and found out commonalities and differences between the models. We report accuracy of each model and accuracy for each food group. The average top-1 recall of all the transfer-learning-based models was 0.6. The model based on InceptionResNetV2 outperformed the other pretrained models (see Table 2).

We found that each model, using different feature extraction networks, shared a common result. According to the evaluation results, the recall and precision for each food group showed large differences (see Table 3). The recall of the 27 main categories, including the rest of the food types,

TABLE 5. Recall of specific food class from the evaluation result of InceptionResNetV2 model using original and augmented dataset.

Food Class	Top-1 recall before augmentation	Top-1 recall after augmentation
Jjamppong	0.830	0.660
Ueongjorim	0.875	0.815
Dakbokkeumtang	0.720	0.650

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3227796

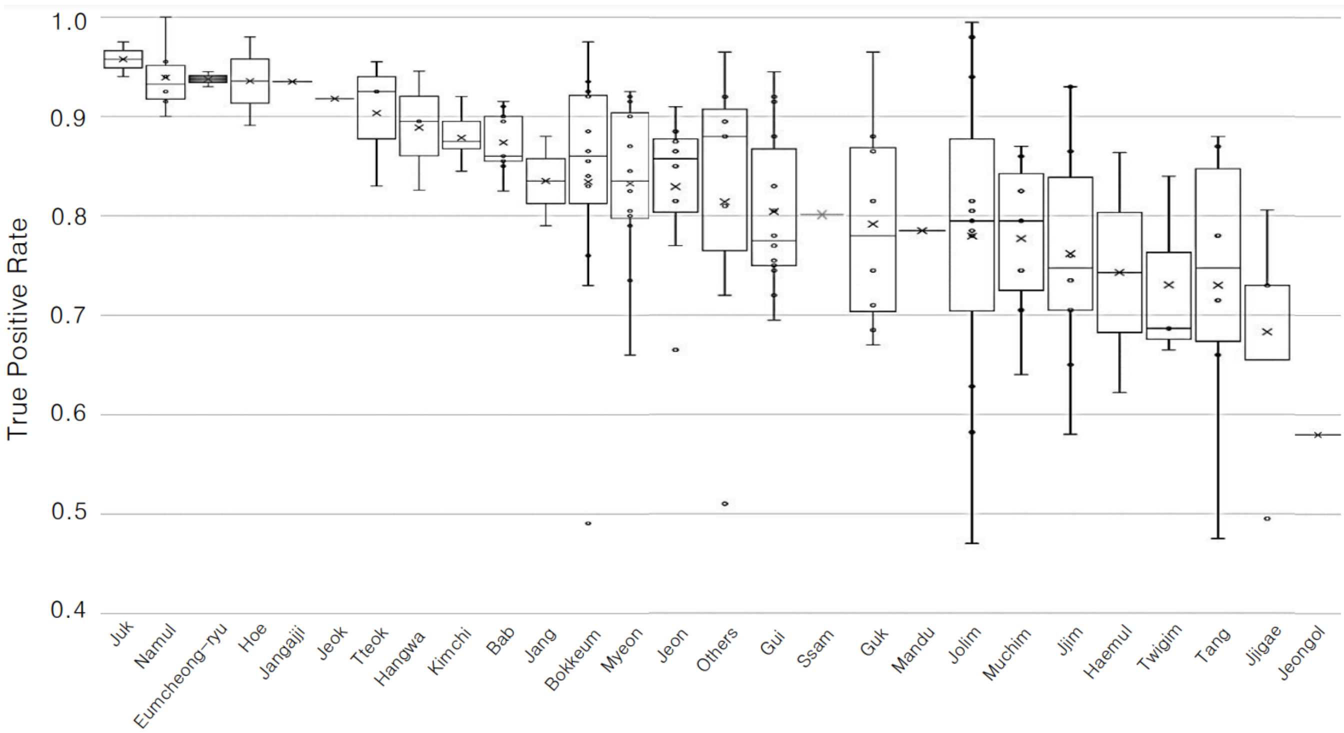


FIGURE 3. A box-plot describing evaluation results of InceptionResNetV2 using 27 group of Korean food data. Circles represent the TP rate of each food class. X-marks represent the average value of the TP rate of each food group. Boxes represent first and third quartile value of the TP rate of each food group. A horizontal line in each box represents the middle value of the TP rate of each food group. A horizontal line at the end of each whisker represents the lowest or highest TP rate of each group, excluding any outliers. The whiskers of the box-plot are drawn within the 1.5 interquartile range.

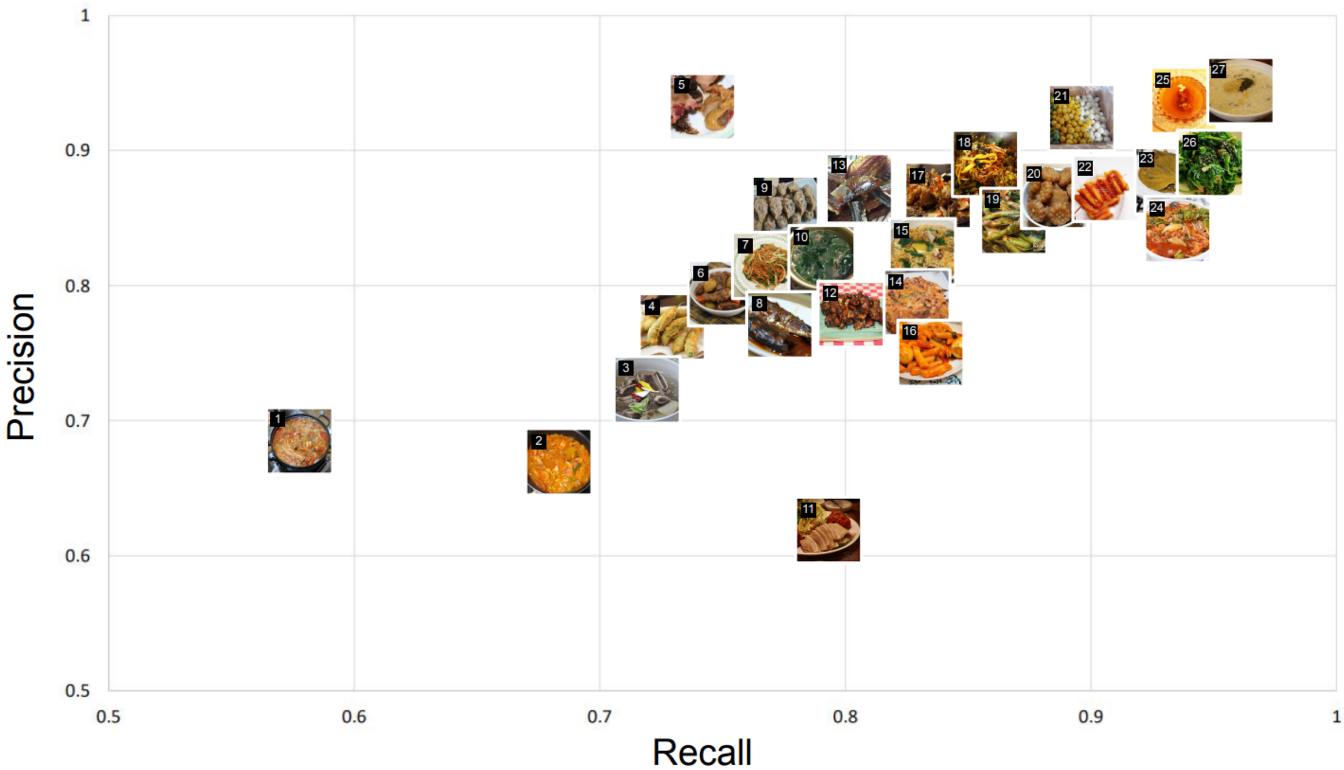


FIGURE 4. Figure showing recall and precision of each 27 groups of Korean food data on the coordinate plane. The X- and Y-Axes represents the recall and precision, respectively, of each group. The food images placed at each coordinate are representative of the corresponding food group.

		Predicted Label								
		Potato Pancake	Rolled Omelet	Sunny side up egg	Kimchi Pancake	Meat Fritters	Pan-fried Battered Fish Fillet	Green Onion Pancake	Pan-fried Battered Zucchini	Others
True Label	Potato Pancake	154	3	2	1	10	10	4	7	9
	Rolled Omelet	3	177	0	0	0	4	1	0	15
	Sunny side up egg	3	0	170	0	0	0	0	0	27
	Kimchi Pancake	4	0	0	175	4	0	10	1	6
	Meat Fritters	9	1	0	1	133	15	0	24	17
	Pan-fried Battered Fish Fillet	8	1	0	0	6	173	1	1	10
	Green Onion Pancake	7	0	0	2	0	0	182	4	5
	Pan-fried Battered Zucchini	9	1	0	10	4	4	6	163	3

FIGURE 5. The classification result of Jeon type food images using model based on InceptionResNetV2.

is shown in Figure 3, and the precision and recall of each food type are shown in Figure 4 with representative images of each food. The classification results for the images of foods belonging to Jeongol/Jjigae/Tang had a low recall and precision. On the other hand, food images in major classes such as Juk/Namul/Eumcheong-ryu showed the highest recall and precision of all the food classes, relatively.

In the classification of 150 classes of food images, the classification model exhibited high accuracy in classifying images of spinach salad, beef tartare, and braised lotus roots. However, the accuracy of the classification model was significantly lower when classifying images of some foods such as Pollack Jjigae, Maeuntang, and Gopchang Jeongol (see Table 4). Similarly, images of other food groups, such as Jeon/Jjim/Jjigae were incorrectly predicted as other food images in the same group. According to the evaluation results of the model based on InceptionResNetV2, only nine out of 46 incorrect predictions of potato pancakes belonged to different groups except Jeon, whereas the remaining 37 incorrect predictions belongEd to the same group, Jeon (see Figure 5). Furthermore, we found that the classification accuracy decreased for some foods when the augmented dataset was used. For example, the accuracy of Jjamppong, Ueongjorim, Dakbokkeumtang, and Haemuljjim was less accurate after data augmentation (see Table 5).

V. DISCUSSION

Some interesting observations made during the training of our classification model, the questions that followed, and other studies that yielded similar results are discussed in this section.

FIGURE 6. Images of Kimchi Jjigae(left), Pollack Jjigae(middle), and Maeuntang(right).

A. DATASET CHARACTERISTICS AFFECTING CLASSIFICATION ACCURACY

We found that the classification accuracy of the models was consistent for all the food classes regardless of the feature extraction networks used. One of the main reasons for the low classification accuracy in certain food image classes was the complex mixing of foods in an image and high similarities between some image classes. The classification model exhibited low performance when classifying the images of food that was a combination of different food classes or food with similar shape and color [3].

Our findings regarding incorrect predictions within the same group reaffirm the fact that image recognition of Korean food is difficult because Korean food images are complex, and images of foods within the same food class may appear different [37]. Because images in the same food groups may have similar color or pattern characteristics, the integration of a hierarchical deep learning model could improve classification accuracy in the same group of foods [38]. However, foods in the unsalted vegetable group showed a lower classification accuracy than other food groups. Moreover, 32 out of the 200 test images of Pollack Jjigae were incorrectly predicted as Maeuntang which belonged to a different food group (see Figure 6). Because the hierarchical classification model required high accuracy of the first-layer classifier [38], a hierarchical approach was necessary to distinguish food classes sharing similar features.

We compared the classification accuracy between the augmented data and the original data and found that the classification accuracy for some food images decreased after augmentation for all the models. Some augmentation processes, such as saturation modification and risk discarding important color information [49] for food recognition. We assumed that color information was more important than shape or patterns into recognizing food classes, such as Jjamppong, Ueongjorim, and Dakbokkeumtang. In addition, instead of using unified and randomized data augmentation, data augmentation with stochastic gradient descent has the potential to improve precision and robustness [50].

B. BACKBONE CHARACTERISTICS AFFECTING MODEL PERFORMANCE

We found interesting elements that affect the classification performance in feature extraction networks and data augmentation. Among the feature extraction networks we used, InceptionResNetV2 with data augmented showed the highest TP rate. Our results align with prior studies which established that InceptionResNetV2 was superior to ResNet-50 and DenseNet-201 in extracting features from food images [51].

The food classification accuracy of the models based on six particular feature extraction networks (e.g., ResNet50V2) decreased when using our data augmentation process. We assumed that augmenting food image data by modifying color information affected classification accuracy.

C. LIMITATIONS AND FUTURE WORK

We trained a model that classified Korean food images; however our study has several limitations. First, although we built a model that classifies Korean images, we did not resolve the poor classification of certain food classes. Resolving the poor performance of the classification model for some food classes is essential for increasing the average performance. Second, we attempted to improve the performance of the model by augmenting the image datasets, but we did not explain why the classification performance for some classes of foods decreased after data augmentation. Determining the cause of the classification performance degradation for some classes of food images after data augmentation is essential for improving model performance. Third, six backbone classes were used for model training. Nevertheless, we need to use the backbones of more than six classes for model training because an architecture such as AlexNet used for food image classification in previous studies may be more suitable for use as a feature extraction network for our model [38], [41]. Furthermore, although our model classified 150 food class images, we need to train other food classes by collecting additional food image data. Collecting more food images and additional training enable the classification model to classify various food classes. Finally, although we built a Korean image classification model, we did not propose and evaluate a system that uses the classification model. The development of a system that uses the classification model enables users to use the model and evaluate their satisfaction with the performance of the classification model.

Therefore, future studies should focus on improving the classification model performance, expanding classifiable food classes, and performing user studies. For model performance, it is important to improve the classification accuracy of some food classes for which the model showed poor performance. Moreover, additional tasks, such as clarifying the reason why the classification accuracy for images of some food classes decreased after data augmentation, testing and applying other backbone networks to improve model performance further, and collecting Korean food images in addition to the 150 classes of Korean food images, still remain. Korean food has more than 150 classes; therefore, collecting more images of different classes allows training a model that classifies more classes of foodimages. For user study, proposing a system that tracks records of food intake to help users maintain proper food balance is essential Researchers should evaluate users' satisfaction and the performance of the food image classification model using this system.

VI. CONCLUSIONS

We aimed to train a classification model that classifies images of Korean food. We believe that our model can be used in a system that automatically classifies the type of food that the user has consumed. Our work will benefit the community of users and researchers using trained models. It also benefits users of systems that automatically classify and record the types of food they take. Users can to classify images of similar Korean food into 150 classes using the model and record what they eat. The key contribution of this study to the research community is that we created a preprocessed dataset for training a Korean food image classification model. We also used several pre-trained deep neural networks to train models that classify Korean food images into 150 classes and evaluated the classification accuracy and time required for model training.

REFERENCES

[1] P. McAllister, H. Zheng, R. Bond, and A. Moorhead, "Towards personalised training of machine learning algorithms for food image classification using a smartphone camera," in *International Conference on Ubiquitous Computing and Ambient Intelligence*. Springer, 2016, pp. 178–190.

[2] G. Waltner, M. Schwarz, S. Ladstätter, A. Weber, P. Luley, M. Lindschinger, I. Schmid, W. Scheitz, H. Bischof, and L. Paletta, "Personalized dietary self-management using mobile vision-based assistance," in *International Conference on Image Analysis and Processing*. Springer, 2017, pp. 385–393.

[3] A. Singla, L. Yuan, and T. Ebrahimi, "Food/non-food image classification and food categorization using pre-trained googlenet model," in *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, 2016, pp. 3–11.

[4] X.-J. Zhang, Y.-F. Lu, and S.-H. Zhang, "Multi-task learning for food identification and analysis with deep convolutional neural networks," *Journal of Computer Science and Technology*, vol. 31, no. 3, pp. 489–500, 2016.

[5] S. Mezgec and B. Koroušić Seljak, “Nutrinet: a deep learning food and drink image recognition system for dietary assessment,” *Nutrients*, vol. 9, no. 7, p. 657, 2017.

[6] H. Hoashi, T. Joutou, and K. Yanai, “Image recognition of 85 food categories by feature fusion,” in *2010 IEEE International Symposium on Multimedia*. IEEE, 2010, pp. 296–301.

[7] G. M. Farinella, D. Allegra, F. Stanco, and S. Battiato, “On the exploitation of one class classification to distinguish food vs non-food images,” in *International Conference on Image Analysis and Processing*. Springer, 2015, pp. 375–383.

[8] S. J. Minija and W. S. Emmanuel, “Food image classification using sphere shaped—support vector machine,” in *2017 International Conference on Inventive Computing and Informatics (ICICI)*. IEEE, 2017, pp. 109–113.

[9] S. Inunganbi, A. Seal, and P. Khanna, “Classification of food images through interactive image segmentation,” in *Asian Conference on Intelligent Information and Database Systems*. Springer, 2018, pp. 519–528.

[10] H. Yang, D. Zhang, D.-J. Lee, and M. Huang, “A sparse representation based classification algorithm for chinese food recognition,” in *International Symposium on Visual Computing*. Springer, 2016, pp. 3–10.

[11] A. Samraj, D. Sowmiya, K. Deepthisri, and R. Oviya, “Food genre classification from food images by deep neural network with tensorflow and keras,” in *2020 Seventh International Conference on Information Technology Trends (ITT)*. IEEE, 2020, pp. 228–231.

[12] M. T. Islam, B. N. K. Siddique, S. Rahman, and T. Jabid, “Food image classification with convolutional neural network,” in *2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, vol. 3. IEEE, 2018, pp. 257–262.

[13] C. Liu, Y. Cao, Y. Luo, G. Chen, V. Vokkarane, M. Yunsheng, S. Chen, and P. Hou, “A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure,” *IEEE Transactions on Services Computing*, vol. 11, no. 2, pp. 249–261, 2017.

[14] N. I. S. Agency, “Ai hub,” 2021. [Online]. Available: https://aihub.or.kr/

[15] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[16] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[17] Y. Wang, Y. He, F. Zhu, C. Boushey, and E. Delp, “The use of temporal information in food image analysis,” in *International Conference on Image Analysis and Processing*. Springer, 2015, pp. 317–325.

[18] O. Beijbom, N. Joshi, D. Morris, S. Saponas, and S. Khullar, “Menu-match: Restaurant-specific food logging from images,” in *2015 IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2015, pp. 844–851.

[19] K. Dalakleidi, M. Sarantea, and K. S. Nikita, “A modified all-and-one classification algorithm combined with the bag-of-features model to address the food recognition task.” in *HEALTHINF*, 2017, pp. 284–290.

[20] G. M. Farinella, M. Moltisanti, and S. Battiato, “Classifying food images represented as bag of textons,” in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 5212–5216.

[21] F. Zhu, M. Bosch, T. Schap, N. Khanna, D. S. Ebert, C. J. Boushey, and E. J. Delp, “Segmentation assisted food classification for dietary assessment,” in *Computational imaging IX*, vol. 7873. International Society for Optics and Photonics, 2011, p. 78730B.

[22] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[23] T. Joutou and K. Yanai, “A food image recognition system with multiple kernel learning,” in *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2009, pp. 285–288.

[24] P. K. Fahira, Z. P. Rahmadhani, P. Mursanto, A. Wibisono, and H. A. Wisesa, “Classical machine learning classification for javanese traditional food image,” in *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*. IEEE, 2020, pp. 1–5.

[25] P. K. Fahira, A. Wibisono, H. A. Wisesa, Z. P. Rahmadhani, P. Mursanto, and A. Nurhadiyatna, “Sumatra traditional food image classification using classical machine learning,” in *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*. IEEE, 2019, pp. 1–5.

[26] S. Christodoulidis, M. Anthimopoulos, and S. Mougiakakou, “Food recognition for dietary assessment using deep convolutional neural networks,” in *International Conference on Image Analysis and Processing*. Springer, 2015, pp. 458–465.

[27] S. Memiş, B. Arslan, O. Z. Batur, and E. B. Sönmez, “A comparative study of deep learning methods on food classification problem,” in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, 2020, pp. 1–4.

[28] G. Ciocca, P. Napoletano, and R. Schettini, “Cnn-based features for retrieval and classification of food images,” *Computer Vision and Image Understanding*, vol. 176, pp. 70–77, 2018.

[29] L. Pan, S. Pouyanfar, H. Chen, J. Qin, and S.-C. Chen, “Deepfood: Automatic multi-class classification of food ingredients using deep learning,” in *2017 IEEE 3rd international conference on collaboration and internet computing (CIC)*. IEEE, 2017, pp. 181–189.

[30] G. Özsert Yiğit and B. M. Özyildirim, “Comparison of convolutional neural network models for food image classification,” *Journal of Information and Telecommunication*, vol. 2, no. 3, pp. 347–357, 2018.

[31] F. Zhou and Y. Lin, “Fine-grained image classification by exploring bipartite-graph labels,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1124–1133.

[32] H. Kagaya, K. Aizawa, and M. Ogawa, “Food detection and recognition using convolutional neural network,” in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 1085–1088.

[33] A. Meyers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, S. Guadarrama, G. Papandreou, J. Huang, and K. P. Murphy, “Im2calories: towards an automated mobile vision food diary,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1233–1241.

[34] K. T. Islam, S. Wijewickrema, M. Pervez, and S. O’Leary, “An exploration of deep transfer learning for food image classification,” in *2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2018, pp. 1–5.

[35] C. N. Freitas, F. R. Cordeiro, and V. Macario, “Myfood: A food segmentation and classification system to aid nutritional monitoring,” in *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2020, pp. 234–239.

[36] T. Ege and K. Yanai, “Image-based food calorie estimation using recipe information,” *IEICE TRANSACTIONS on Information and Systems*, vol. 101, no. 5, pp. 1333–1341, 2018.

[37] S.-J. Park, A. Palvanov, C.-H. Lee, N. Jeong, Y.-I. Cho, and H.-J. Lee, “The development of food image detection and recognition model of korean food for mobile dietary management,” *Nutrition research and practice*, vol. 13, no. 6, pp. 521–528, 2019.

[38] H. Yang, S. Kang, C. Park, J. Lee, K. Yu, and K. Min, “A hierarchical deep model for food classification from photographs,” *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 14, no. 4, pp. 1704–1720, 2020.

[39] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[40] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[42] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 675–678.

[43] D. Eigen and R. Fergus, “Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2650–2658.

[44] *200 International Korean Menu Guide*. Korean Food Promotion Institute, 2014.

[45] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in *European conference on computer vision*. Springer, 2016, pp. 630–645.

[46] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in *Thirty-first AAAI conference on artificial intelligence*, 2017.

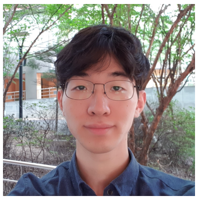
[47] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, “Learning transferable architectures for scalable image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.

[48] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

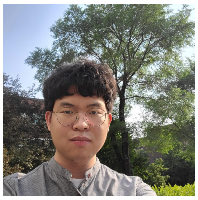
This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3227796

[49] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.
[50] A. Fawzi, H. Samulowitz, D. Turaga, and P. Frossard, "Adaptive data augmentation for image classification," in *2016 IEEE international conference on image processing (ICIP)*. Ieee, 2016, pp. 3688–3692.
[51] G. A. Tahir and C. K. Loo, "Progressive kernel extreme learning machine for food image analysis via optimal features from quality resilient cnn," *Applied Sciences*, vol. 11, no. 20, p. 9562, 2021.

MINKI CHUN received a B.S. degree in information communication engineering from Dongguk University, South Korea, in 2020. He is currently pursuing an M.S. degree in Computer Science from the University of Seoul, South Korea. His research interests include deep learning, explainable AI, and user experience.



HYEONHAK JEONG received a B.S. degree in computer science from the University of Seoul, South Korea, in 2021. He is currently pursuing an M.S. degree. His research interests include human-computer interaction, usable security, accessibility, and deep learning.



HYUNMIN LEE is pursuing a B.S. degree in computer science from the University of Seoul. His research interests include natural language processing, human-computer interaction, and machine learning.

TAEWON YOO is currently pursuing a B.S degree in the Kyung Hee University. His research interests include human-computer interaction and responsible AI.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3227796

HYUNGGU JUNG is an associate professor of Computer Science & Engineering and an assistant professor of Artificial Intelligence at the University of Seoul and directs the Human-Centered Artificial Intelligence Lab (HCAIL). He was an assistant professor of Software Convergence at Kyung Hee University from April 2018 to August 2019. He received his B.S., M.Math, and M.S. degrees from KAIST, the University of Waterloo, and Stanford University, respectively, all in Computer Science. He also received a B.S., with a minor in Business Economics from KAIST, and his Ph.D. in Biomedical and Health Informatics from the University of Washington School of Medicine. Furthermore, he worked at Microsoft Research and PARC, a Xerox company, as a research intern. His research interests lie at the intersection of human-centered AI, health informatics, social computing, and accessibility & aging. His research aims to advance AI research through design and engineering to support individuals with special needs (e.g., older adults, streamers with visual impairments, and North Korean defectors with depression) across multiple domains: health, social media, sharing economy, and education. He is a recipient of the Korean Government Scholarship and the Mogam Science Scholarship. Further information about his recent research activities is available at https://hcail.uos.ac.kr/.