

## RESEARCH ARTICLE

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

MINKI CHUN<sup>1</sup>, HYEONHAK JEONG<sup>1</sup>, HYUNMIN LEE<sup>1</sup>, TAEWON YOO<sup>2</sup>,  
AND HYUNGGU JUNG<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering, University of Seoul, Dongdaemun-gu, Seoul 02504, Republic of Korea

<sup>2</sup>Department of Software Convergence, Kyung Hee University, Giheung-gu, Yongin-si 17104, Republic of Korea

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

This work was supported by the National Research Foundation of Korea(NRF) Grant Funded by the Korea Government(MSIT) under Grant 2020R1G1A1009133.

**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 multiple 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 platform, a public food image dataset. The images were pre-processed and augmented for model training. The proposed model was evaluated in an experiment using a Korean food image dataset and 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 is 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 studies attempted to solve the difficult task of food image classification using traditional machine learning (ML) or deep learning (DL) methods. The traditional ML methods that have been used include (1) using 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) using support vector machine (SVM)-based image segmentation [6], and (5) using food images of specific domains [7], [8], [9], [10], [11]. DL methods are widely used to train food image classification models. The DL-based methods differ according to whether fine-tuning was performed [12] or not. Several researchers trained a classification system model using fine-tuning. Among them, some proposed a system using the model and others did not [13], [14].

The associate editor coordinating the review of this manuscript and approving it for publication was Jad Nasreddine<sup>1</sup>.

Nevertheless, none of these researchers 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 Korean food was developed in this study using a public food image dataset. Besides, although Korean foods that belong to the same food category have similar characteristics, the differences between food categories have been simplified or ignored in previous studies. Hence, we considered the differences between the 27 food categories defined by the Korea Culture and Information Service.<sup>1</sup> Analysis of the Korean food image dataset is still in its infancy. In this study, we examined the characteristics of Korean food image dataset 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 platform [15] operated by the National Information Society Agency.<sup>2</sup> AI-Hub platform is a data repository for sharing datasets with model developers training their artificial intelligence models. AI-Hub platform enables them to download publicly available data, such as text data (e.g., corpus data, machine reading comprehension data) and image data (e.g., Korean food image data, eye gaze data).

Second, we reduced model complexity and enhanced accuracy through image pre-processing. Subsequently, we augmented the pre-processed 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 pre-processed dataset for training Korean food image classification models.
- We present the level of 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 studies 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.

<sup>1</sup><https://www.kocis.go.kr/>

<sup>2</sup><https://eng.nia.or.kr/>

### A. TRADITIONAL ML-BASED FOOD IMAGE CLASSIFICATION

Prior studies created and evaluated ML-based food image classification models [1], [7], [8], [9], [10], [11], [16], [17], [18], [19], [20], [21], [22], [23]. To train food image classification models, ten studies used the images of food items whose origins are unknown (e.g., self-made dataset, PFID dataset) [1], [7], [8], [9], [10], [16], [17], [18], [19], [20]. Out of the ten studies, five proposed food image classification models using a dataset created with images from the web [7], [9], [18] and personal devices [16], [17]. For example, food images from the web were classified by the model that uses the histogram of oriented gradients, scale-invariant feature transforms, local binary patterns, and filter responses [17], while an SVM with the scatter search approach was used to classify food images from personal devices into multiple classes [9]. Meanwhile, the other studies created a food image classification model using a dataset that have already been created (i.e., Food-101 dataset [10], PFID dataset [19], Purdue dataset [20], UNICT-FD889 dataset [1], [8]). For instance, the UNICT-FD889 dataset contains 3,583 food images related to 889 different plates [8]. Farinella et al. [8] used the dataset for creating an image classification model that enables classifying images into food and non-food images.

On the contrary, four studies used the images of food items that belong to a specific country (e.g., China [11], Indonesia [22], [23], Japan [21]) when training their proposed food image classification models. They created food image classification models using images from the web [21] and personal devices [11], [22], [23], respectively. In particular, food images from the web were classified by the model that uses the bag-of-features [24], a color histogram, a gradient histogram, and the Gabor texture [23], while a random forest method [25] was used to classify food images from personal devices into multiple classes [21].

### B. DEEP LEARNING-BASED FOOD IMAGES CLASSIFICATION

DL-based food image classification methods are widely used because they enable various systems to extract and classify image features. Prior studies created and evaluated DL-based food image classification models [12], [13], [14], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38]. To train food image classification models, researchers used the images of food items whose origins are unknown (e.g., Food-475 dataset, Food-524 dataset) [12], [13], [14], [26], [27], [28], [29], [30], [31], [32], [33]. Out of eleven studies, eight proposed food image classification models using a fine-tuning method in order to reduce training time and enhance model accuracy [13], [14], [28], [29], [30], [31], [32], [33]. For example, Islam et al. [13] used Inception V3 and a fine-tuning method to enhance the accuracy of the proposed model by about 16 percent over the model accuracy of the previously proposed convolutional neural network or

AlexNet [39]. Similarly, Pan et al. [29] used ResNet-50 [40] and a fine-tuning method to enhance the accuracy of the proposed model by about 10 percent over the model accuracy of AlexNet [39].

While a fine-tuning method is being widely used for training image classification models, some researchers proposed food image classification models without using the fine-tuning method [12], [26], [27]. For example, Samraj et al. [12] developed a visual food classification model to prevent errors in handling large amounts of food. They created an image classification model that classifies 170 food image classes using a dataset of food items such as Asian vegetarian meals and Bland meals. Similarly, Christodoulidis et al. [26] proposed a method of extracting patches from food images and then feeding them to a CNN. They used a dataset of 246 images of food items served in the restaurants of Bern University hospital. In addition, Memiş et al. [27] compared the performance of several DL 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 [41].

In contrast with the aforementioned studies, five studies used the images of food items that belong to a specific country (e.g., Australia [34], Brazil [35], Japan [36], Korea [37], [38]) when training their proposed food image classification models. While three studies proposed classification models without using a fine-tuning method [35], [36], [37], for training a food image classification model, two studies used a fine-tuning method [34], [38]. For example, Islam et al. [34] used a fine-tuning method to enhance the accuracy of the proposed model by about four percent over the model accuracy of the previously proposed AlexNet [39].

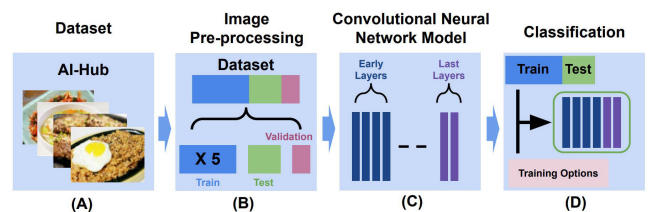
### C. LIMITATIONS OF PRIOR STUDIES

Based on the aforementioned studies, we found a few limitations as follows. First, although previous studies proposed image classification models trained using food from various countries, only few used the Korean food image dataset. Moreover, the researchers who developed the Korean food image classification models pointed out that the dataset was insufficient [37], [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. Second, understanding of the classification performance for each Korean food category is limited. Although Korean foods in the same food category have similar characteristics, food categories 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 [42]. 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

the researchers with opportunities to improve Korean food classification models.

### 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 dataset to train a model for classifying Korean food images. We obtained image data for model training through image pre-processing 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.



**FIGURE 1.** Training process of classification models. (A) Collect Korean food images from the AI-Hub platform. (B) Pre-process 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.

#### A. DATASET

We used a Korean food image dataset provided by the AI-Hub platform. Each image of the dataset was collected by researchers using a web crawler from Korea Institute of Science and Technology. As reported by the Ministry of Agriculture, Food and Rural Affairs and the Korean Food Foundation, the AI-Hub platform enabled researchers to navigate 27 categories of 150 food classes (see Table 1). As a result, we obtained an average of 1,004 images ( $SD=17.02$ ) per class and a total of 150,610 images from the platform.

#### B. IMAGE PRE-PROCESSING

We reduced the complexity of the model network and increase classification accuracy by pre-processing 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  $\times$  331 pixels to fit the input of the model.

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) original training images; (2) images with brightness changed between 0.5 and 1.5 times using random seeds; (3) images left in which we changed the saturation from 0.5 to 0.8 using random seeds; (4) images in which we changed the contrast from 1.2x to 1.7x using a random seed; (5) images flipped horizontally. The training data used for model training were a total of

**TABLE 1.** List of food categories in Korean food data.

#	Image	Category Name	Description	#	Image	Category 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		Bap	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		Jorim	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		Eumcheongryu	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.				

527,135 images, a five-fold increase from 105,427 images selected for use in model training.

**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 [43], InceptionResNetV2 [44], NasNetLarge [45], and MobileNetV2 [46] models pre-trained with ImageNet data [39] were used to classify Korean food. We connected the pre-trained feature extraction network to the classification model network, which consists of global average pooling

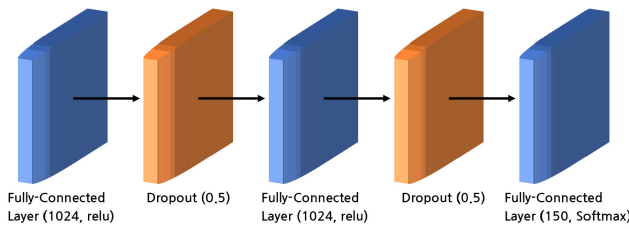


FIGURE 2. Architecture of the classification model layer.

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	<b>0.8191</b>	436,182
NasNetLarge	0.7791	294,853
MobileNetV2	0.7536	82,366
ResNet-101V2	0.737	51,649
ResNet-152V2	0.7131	82,458
ResNet-50V2	0.6827	42,915

layers, fully connected layers, and dropout layers. Figure 2 describes the classification model structure.

D. CLASSIFICATION

We trained each Korean food image classification model. Each classification model was trained through two stages: initial training and fine-tuning stages. The aim of the initial training stage was to obtain pre-trained classification models using ImageNet data [39]. During the initial training stage, we trained only the parameters of each classification model. To fix the parameters of the feature extraction network, we set multiple values: the batch size to 128, the training epoch value to 1,000, and the early-stopping value to 50, respectively. We used the Adam optimizer with a 0.001 learning rate and adopted categorical cross-entropy as a loss function. To fit each trained classification model to the Korean food image data, we trained each model again by setting the learning rate to 0.0001 instead of 0.001.

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 platform for model performance evaluation. The accuracy was measured using the following formula:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

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

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

TABLE 3. Classification recall, precision and f1-score of 27 Korean food categories in descending order based on recall.

Category Name	Recall	Precision	F1-score
Juk	0.958	0.942	0.950
Namul	0.939	0.908	0.924
Eumcheongryu	0.938	0.916	0.927
Hoe	0.936	0.860	0.896
Jangajji	0.935	0.899	0.917
Jeok	0.9108	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
Bap	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	0.840
Gui	0.804	0.792	0.798
Ssam	0.795	0.621	0.697
Guk	0.792	0.816	0.804
Mandu	0.785	0.804	0.811
Jorim	0.780	0.790	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	0.730	0.751	0.74
Jjigae	0.683	0.671	0.677
Jeongol	0.580	0.686	0.629

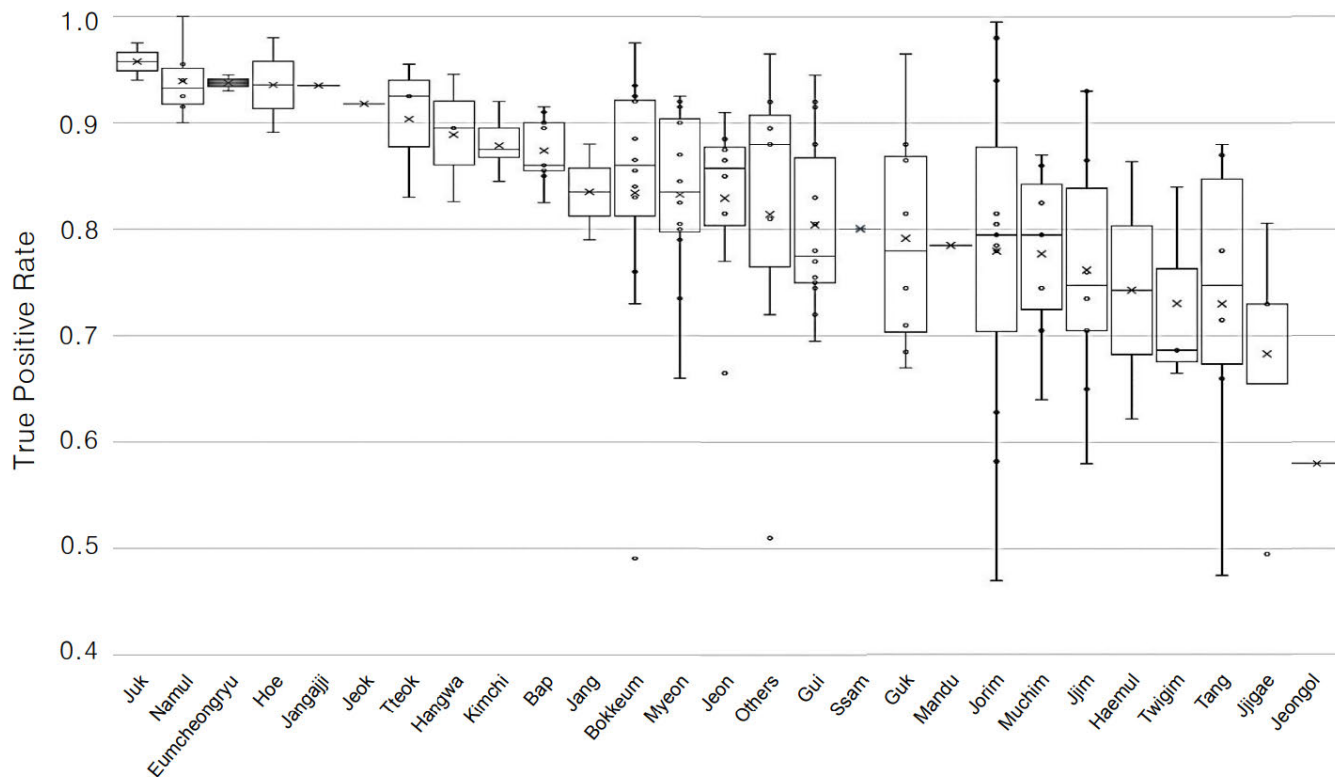
TABLE 4. Number of accurate predictions from the evaluation result of the model based on InceptionResNetV2. The number of test samples for each food class is 200.

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

each model and accuracy for each food category. The average top-1 recall of all the transfer-learning-based models was 0.6. The model based on InceptionResNetV2 outperformed the other pre-trained 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 category showed large differences (see Table 3). The recall of the 27 main categories, including the rest of the food types, 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/Eumcheongryu 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



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

**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

as Dongtae Jjigae, Maeuntang, and Gopchang Jeongol (see Table 4). Similarly, images of other food categories, such as Jeon/Jjim/Jjigae were incorrectly predicted as other food images in the same category. According to the evaluation results of the model based on InceptionResNetV2, only nine out of 46 incorrect predictions of potato pancakes belonged to different categories except Jeon, whereas the remaining 37 incorrect predictions belonged to the same category, 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.

**A. DATASET CHARACTERISTICS AFFECTING CLASSIFICATION ACCURACY**

We discovered 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 category 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 categories may have similar color or pattern characteristics, the integration of a hierarchical DL model could improve classification accuracy in the same category of foods [38]. However, foods in the unsalted vegetable category showed a lower classification accuracy than other food categories. Moreover, 32 out of the 200 test images of Dongtae Jjigae

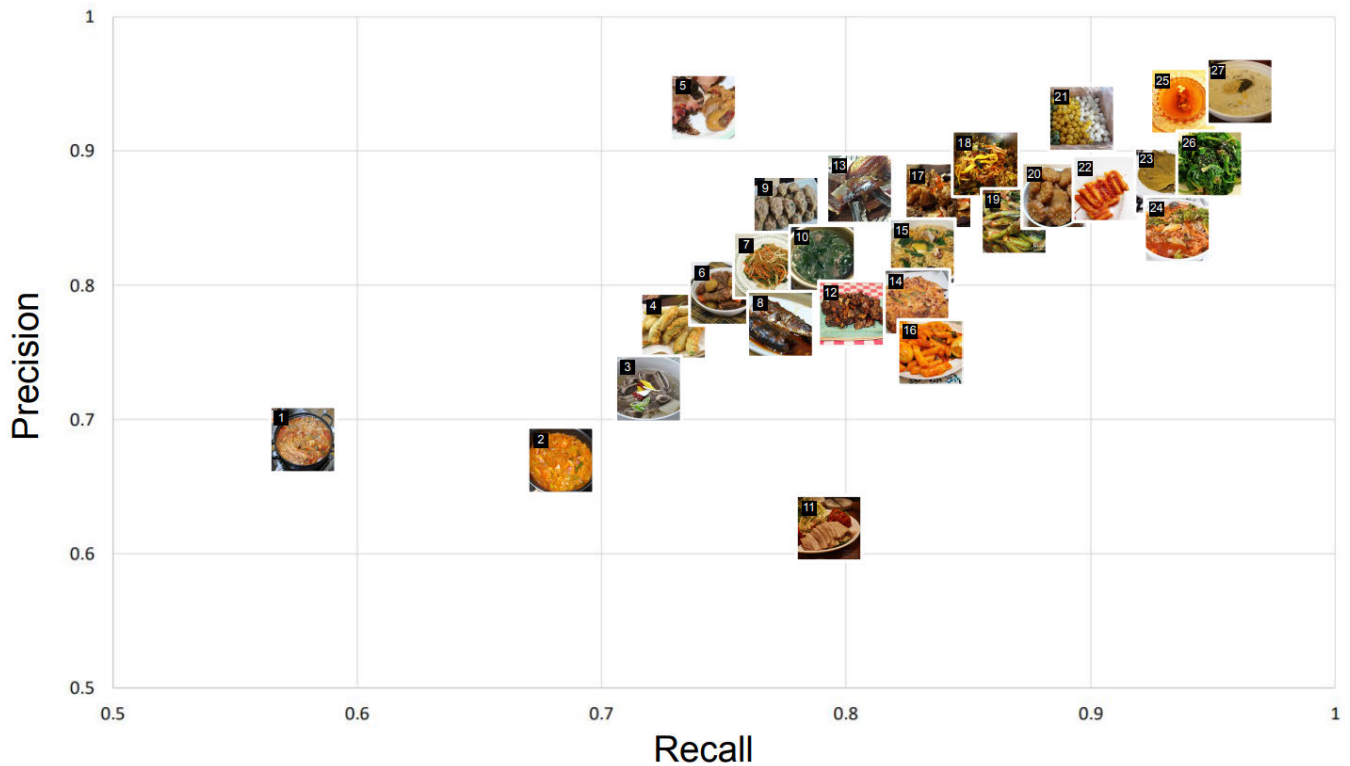


FIGURE 4. Recall and precision of each 27 categories of Korean food data on the coordinate plane. The X- and Y-Axes represents the recall and precision, respectively, of each category. The food images placed at each coordinate are representative of the corresponding food category.

		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. Classification result of Jeon type food images using model based on InceptionResNetV2.

were incorrectly predicted as Maeuntang which belonged to a different food category (see Figure 6). Because the hierarchical classification model required high accuracy of the first-layer classification model [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



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

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 [47] 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 [48].

### **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 [49]. 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 [39] and CaffeNet [50] used for food image classification in previous studies may be more suitable for use as a feature extraction network for our model [38], [39]. 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.

Future studies should focus on improving the classification model performance and evaluating the feasibility of the classification models. First, to improve the classification model performance, we may increase the classification accuracy of some food classes (e.g., Jjigae, Jeongol) for which the model showed poor performance. For example, we may consider using hierarchical ensemble structures [38] for model training for such classes with poor performance. Also, it would be worthwhile to convert the color space from RGB to HSV and use histogram normalization during the pre-processing process. Additionally, we may train the classification models using additional backbones (e.g., CoCa, efficient adaptive ensembling, InceptionResNetV3 [51]) not used in this study. Second, to confirm the feasibility of the classification models, we would need to conduct usability studies with target users using our proposed classification models. For instance, during the usability study, we may calculate the inference time from the difference between the measured times by measuring the time when the models receive input data and the time when the models generate output results, respectively. In addition, we may determine the level of users' satisfaction with our proposed classification models.

## **VI. CONCLUSION**

We aimed to train a Korean food image classification model. 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 pre-processed dataset for training a Korean food image classification model. Also, we 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.

## **ACKNOWLEDGMENT**

The authors acknowledge the Urban Big Data and AI Institute of the University of Seoul supercomputing resources (<http://ubai.uos.ac.kr>) made available for conducting the research reported in this paper. The authors appreciate HCAIL members for their constructive feedback on our initial manuscript.



## 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 *Proc. Int. Conf. Ubiquitous Comput. Ambient Intell.* Cham, Switzerland: 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 *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: 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 *Proc. 2nd Int. Workshop Multimedia Assist. Dietary Manag.*, Oct. 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," *J. Comput. Sci. Technol.*, vol. 31, no. 3, pp. 489–500, May 2016.
- [5] S. Mezgec and B. K. Seljak, "NutriNet: A deep learning food and drink image recognition system for dietary assessment," *Nutrients*, vol. 9, no. 7, p. 657, Jun. 2017.
- [6] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1997.
- [7] H. Hoashi, T. Joutou, and K. Yanai, "Image recognition of 85 food categories by feature fusion," in *Proc. IEEE Int. Symp. Multimedia*, Dec. 2010, pp. 296–301.
- [8] 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 *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2015, pp. 375–383.
- [9] S. J. Minija and W. R. S. Emmanuel, "Food image classification using sphere shaped—Support vector machine," in *Proc. Int. Conf. Inventive Comput. Informat. (ICICI)*, Nov. 2017, pp. 109–113.
- [10] S. Inunganbi, A. Seal, and P. Khanna, "Classification of food images through interactive image segmentation," in *Proc. Asian Conf. Intell. Inf. Database Syst.* Cham, Switzerland: Springer, 2018, pp. 519–528.
- [11] H. Yang, D. Zhang, D.-J. Lee, and M. Huang, "A sparse representation based classification algorithm for Chinese food recognition," in *Proc. Int. Symp. Vis. Comput.* Cham, Switzerland: Springer, 2016, pp. 3–10.
- [12] A. Samraj, D. Sowmiya, K. A. Deepthisri, and R. Oviya, "Food genre classification from food images by deep neural network with tensorflow and Keras," in *Proc. 7th Int. Conf. Inf. Technol. Trends (ITT)*, Nov. 2020, pp. 228–231.
- [13] M. T. Islam, B. M. K. Siddique, S. Rahman, and T. Jabid, "Food image classification with convolutional neural network," in *Proc. Int. Conf. Intell. Informat. Biomed. Sci. (ICIBMS)*, Oct. 2018, pp. 257–262.
- [14] 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 Trans. Services Comput.*, vol. 11, no. 2, pp. 249–261, Mar./Apr. 2018.
- [15] N. I. S. Agency. (2021). *Ai Hub*. [Online]. Available: <https://aihub.or.kr/>
- [16] Y. Wang, Y. He, F. Zhu, C. Boushey, and E. Delp, "The use of temporal information in food image analysis," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2015, pp. 317–325.
- [17] O. Beijbom, N. Joshi, D. Morris, S. Saponas, and S. Khullar, "Menu-Match: Restaurant-specific food logging from images," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Jan. 2015, pp. 844–851.
- [18] K. Dalakleidi, M. Sarantea, and K. Nikita, "A modified all-and-one classification algorithm combined with the bag-of-features model to address the food recognition task," in *Proc. 10th Int. Joint Conf. Biomed. Eng. Syst. Technol.*, 2017, pp. 284–290.
- [19] G. M. Farinella, M. Moltisanti, and S. Battiato, "Classifying food images represented as bag of textons," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 5212–5216.
- [20] 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 *Proc. SPIE*, vol. 7873, Feb. 2011, Art. no. 78730B.
- [21] T. Joutou and K. Yanai, "A food image recognition system with multiple kernel learning," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 285–288.
- [22] P. K. Fahira, Z. P. Rahmadhani, P. Mursanto, A. Wibisono, and H. A. Wisesa, "Classical machine learning classification for javanese traditional food image," in *Proc. 4th Int. Conf. Informat. Comput. Sci. (ICICoS)*, Nov. 2020, pp. 1–5.
- [23] 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 *Proc. 3rd Int. Conf. Informat. Comput. Sci. (ICICoS)*, Oct. 2019, pp. 1–5.
- [24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Feb. 2004.
- [25] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [26] S. Christodoulidis, M. Anthimopoulos, and S. Mouggiakakou, "Food recognition for dietary assessment using deep convolutional neural networks," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2015, pp. 458–465.
- [27] S. Memis, B. Arslan, O. Z. Batur, and E. B. Sonmez, "A comparative study of deep learning methods on food classification problem," in *Proc. Innov. Intell. Syst. Appl. Conf. (ASYU)*, Oct. 2020, pp. 1–4.
- [28] G. Ciocca, P. Napoletano, and R. Schettini, "CNN-based features for retrieval and classification of food images," *Comput. Vis. Image Understand.*, vols. 176–177, pp. 70–77, Nov. 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 *Proc. IEEE 3rd Int. Conf. Collaboration Internet Comput. (CIC)*, Oct. 2017, pp. 181–189.
- [30] G. O. Yigit and B. M. Özyildirim, "Comparison of convolutional neural network models for food image classification," *J. Inf. Telecommun.*, vol. 2, no. 3, pp. 347–357, Jul. 2018.
- [31] F. Zhou and Y. Lin, "Fine-grained image classification by exploring bipartite-graph labels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1124–1133.
- [32] H. Kagaya, K. Aizawa, and M. Ogawa, "Food detection and recognition using convolutional neural network," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 1085–1088.
- [33] A. Myers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, S. Guadarrama, G. Papandreou, J. Huang, and K. Murphy, "Im2Calories: Towards an automated mobile vision food diary," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 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 *Proc. Digit. Image Comput., Techn. Appl. (DICTA)*, Dec. 2018, pp. 1–5.
- [35] C. N. C. Freitas, F. R. Cordeiro, and V. Macario, "MyFood: A food segmentation and classification system to aid nutritional monitoring," in *Proc. 33rd SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, Nov. 2020, pp. 234–239.
- [36] T. Ege and K. Yanai, "Image-based food calorie estimation using recipe information," *IEICE Trans. Inf. Syst.*, 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 Res. Pract.*, 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 Trans. Internet Inf. Syst.*, vol. 14, no. 4, pp. 1704–1720, 2020.
- [39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012.
- [40] D. Eigen and R. Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2650–2658.
- [41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [42] *200 International Korean Menu Guide*, Korean Food Found., Korean Food Service Inf. Ltd., 2014.
- [43] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 630–645.
- [44] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-V4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 1–7.
- [45] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710.

[46] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[47] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.

[48] A. Fawzi, H. Samulowitz, D. Turaga, and P. Frossard, "Adaptive data augmentation for image classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3688–3692.

[49] G. A. Tahir and C. K. Loo, "Progressive kernel extreme learning machine for food image analysis via optimal features from quality resilient CNN," *Appl. Sci.*, vol. 11, no. 20, p. 9562, Oct. 2021.

[50] 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 *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 675–678.

[51] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.



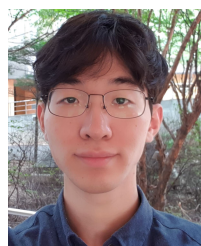
**HYUNMIN LEE** is pursuing his B.S. degree in computer science from the University of Seoul, Republic of Korea. His research interests include natural language processing, human-computer interaction, and machine learning.



**TAEWON YOO** is pursuing his B.S degree in software convergence from Kyung Hee University, Republic of Korea. His research interests include human-computer interaction and responsible AI.



**MINKI CHUN** is pursuing his M.S. degree in computer science from the University of Seoul, Republic of Korea. He received his B.S. degree in information communication engineering from Dongguk University, Republic of Korea, in 2020. His research interests include deep learning, explainable AI, and user experience.



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



**HYUNGGU JUNG** is an Associate Professor of Computer Science and Engineering and an Associate Professor of Artificial Intelligence at the University of Seoul and directs the Human-Centered Artificial Intelligence Laboratory (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. degree 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 include lie at the intersection of human-centered AI, health informatics, social computing, and accessibility and 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/>.

...