pISSN 1225-4339 eISSN 2287-4992
Korean J. Food Nutr. Vol. 37. No. 4, 197~210 (2024)
https://doi.org/10.9799/ksfan.2024.37.4.197

THE KOREAN JOURNAL OF
한국식품영양학회지
FOOD AND NUTRITION

Development of an Optimal Convolutional Neural Network Backbone Model for Personalized Rice Consumption Monitoring in Institutional Food Service using Feature Extraction

Young Hoon Park and †Eun Young Choi*

Full Professor, Dept. of Civil Engineering, Bucheon University, Bucheon 15073, Korea
**Assistant Professor, Dept. of Food and Nutrition, Bucheon University, Bucheon 15073, Korea*

Abstract

This study aims to develop a deep learning model to monitor rice serving amounts in institutional foodservice, enhancing personalized nutrition management. The goal is to identify the best convolutional neural network (CNN) for detecting rice quantities on serving trays, addressing balanced dietary intake challenges. Both a vanilla CNN and 12 pre-trained CNNs were tested, using features extracted from images of varying rice quantities on white trays. Configurations included optimizers, image generation, dropout, feature extraction, and fine-tuning, with top-1 validation accuracy as the evaluation metric. The vanilla CNN achieved 60% top-1 validation accuracy, while pre-trained CNNs significantly improved performance, reaching up to 90% accuracy. MobileNetV2, suitable for mobile devices, achieved a minimum 76% accuracy. These results suggest the model can effectively monitor rice servings, with potential for improvement through ongoing data collection and training. This development represents a significant advancement in personalized nutrition management, with high validation accuracy indicating its potential utility in dietary management. Continuous improvement based on expanding datasets promises enhanced precision and reliability, contributing to better health outcomes.

Key words: deep learning, neural networks, computer, rice, food services

Introduction

In institutional food service, the allocation and consumption of food portions often reflect individual preferences despite dietitians' efforts to ensure a balanced diet. This scenario presents a unique challenge in group meal provision, which aims to meet diverse dietary needs while maintaining nutritional balance (National Academies of Sciences, Engineering, and Medicine 2016; Peano et al. 2022; Yeom & Choi 2023).

Carbohydrates, as a major energy source, constitute approximately 55~65% of daily energy intake (Korean Nutrition Society 2020). This nutrient plays a critical role in the functioning of various organs, such as the brain, red blood cells, retina, lens, and renal medulla, which predominantly utilize glucose as their primary energy substrate. Therefore, maintaining a consistent blood glucose level is imperative for the optimal functioning of these organs, highlighting the necessity for regular carbohydrate consumption. This dietary requirement underscores the importance of carbohydrates in the human diet, particularly in relation to maintaining the energy demands of critical bodily functions (Kim MH 2013; Stubbs RJ 2021; Pan et al. 2023; Wali et al. 2023).

The proportion of energy intake from carbohydrates is closely associated with chronic diseases. Patients diagnosed with hypertension, metabolic syndrome, and diabetes tend to derive over 70% of their total energy from carbohydrates, a trend particularly pronounced among individuals over the age of 60. In a study targeting adults and the elderly over 40 in Korea, it was found that those with a carbohydrate energy intake ratio exceeding 65% had a 1.18 times higher likelihood of being at high risk for cardiovascular diseases compared to those with a ratio of 55~65% (Hou et al. 2022).

While many studies have reported positive correlations between low-carbohydrate, high-fat diets and health benefits, par-

† Corresponding author: Eun Young Choi, Assistant Professor, Dept. of Food and Nutrition, Bucheon University, Bucheon 15073, Korea. Tel: +82-32-610-3442, Fax: +82-32-610-3205, E-mail: eychoi@bc.ac.kr

ticularly in terms of lower overall calorie intake, recent research on long-term health maintenance has shown varying results. Consequently, the importance of adequate energy intake and appropriate carbohydrate consumption is being emphasized. This growing interest among those responsible for meal planning reflects an increased awareness of the role of balanced carbohydrate intake in overall health and nutrition (Yang et al. 2022; Santamarina et al. 2023).

There are various methods for measuring food intake. First, there is the traditional method of directly measuring with a scale. This method offers high accuracy and reliability, but it is time-consuming, has limitations in portability, and does not provide information beyond weight. Second, there are methods using machines. For example, scanning food with a photo to measure volume and convert it to weight is a hygienic, contactless method that can quickly process large quantities of food. However, it may lack accuracy, be costly, and be influenced by environmental factors. Additionally, there is the method of using trained models to perform automated measurements, which offers high scalability and precision but requires a large amount of data and a complex training process.

Currently, various deep learning models and systems are utilized for nutrition management worldwide. For example, studies utilizing FoodAI in the United States and the UECFood100 and UECFood256 datasets in Japan have been actively conducted, focusing primarily on food image recognition and classification (Sahoo et al. 2019; Kawano & Yanai 2014). However, these models are not capable of directly analyzing food intake, and additional data and analytical techniques are required to achieve this.

In Korea, certain services have gained attention for using computer vision to measure the image and volume of food. Computer vision employs traditional image analysis techniques, analyzing images through manually defined algorithms (Bolaños & Radeva 2016). In contrast, Convolutional Neural Networks (CNNs), a deep learning-based approach, automatically extract and analyze complex features from images through data training. Although CNNs require more data, they excel in solving more complex problems (Sandler et al. 2018).

Despite the availability of various tools for meal management and food intake surveys, most methods involve manually entering food names and quantities to calculate calorie and nutrient intake (Kalivaraprasad et al. 2021). As smartphones have become more widespread, capturing images has become easier, leading to increased demand for image-based calorie estimation. However, research on applying artificial intelligence to extract characteristics from images, estimate weight, and calculate nutrient intake remains insufficient (Mezgec et al. 2017; Vasiloglou et al. 2018; Cai et al. 2019; Lu et al. 2020; Fragopoulou et al. 2021; Matusheski et al. 2021).

CNNs play a crucial role in object detection (identifying specific objects within an image) and object segmentation (distinguishing objects from the background in an image). Despite these advancements, the efficiency of training and validation can vary significantly depending on the characteristics of the images. Therefore, it is challenging to determine the exact number of images required to achieve the desired accuracy before conducting a study. Nonetheless, having more data generally contributes to improving the accuracy of results.

Object Detection Models: SSD (Single Shot Multibox Detector) scans an image once and identifies objects within it efficiently. R-CNN (Region-based convolutional Neural Network) identifies potential areas where objects might be located and then closely examines these areas to identify the objects, akin to first spotting interesting regions in a photo and then zooming in to see what is there. These models use pre-trained CNNs such as VGGNet, ResNet, ResNeXt, MobileNet, and AlexNet as backbones, which demonstrate strong efficiency in extracting image characteristics due to their powerful and well-optimized architectures (Girshick et al. 2014; Simonyan & Zisserman, 2015; Szegedy et al., 2015; He et al. 2016; Liu et al. 2016; Chollet F 2017; Howard et al. 2017; Huang et al., 2017; Tan & Le 2019).

Object Segmentation Models: Models like U-Net, FCN (Fully Convolutional Network), and DeepLab are used to separate objects from the background in an image, such as distinguishing a person from the surrounding scenery. These models also utilize pre-trained CNNs like VGGNet, ResNet, EfficientNet, and Xception for feature extraction, directly influencing the performance of object detection and segmentation (Ronneberger et al. 2015; Chen et al. 2017).

The optimal structure of a CNN can be tailored according to the characteristics of the image, such as texture and color. More complex image features require deeper and more complex CNNs. Additionally, the optimal combination of filter size and number, pooling layers, types of optimizers, and the application of

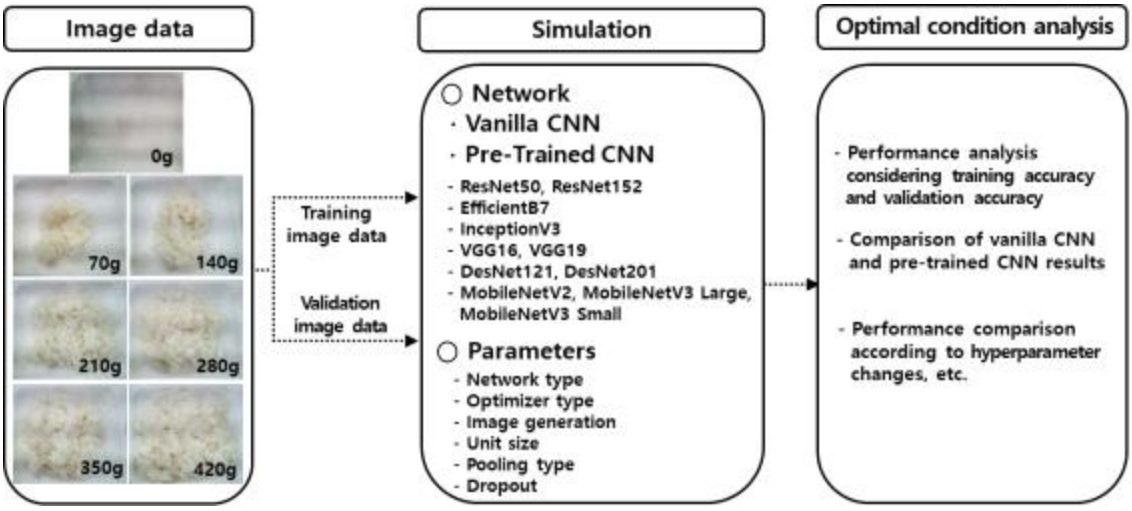


Fig. 1. Derivation of optimal convolutional neural network backbone for rice quantity detection.

dropout significantly enhances the accuracy of image feature extraction.

This study aims to identify the optimal CNN model for analyzing the amount of rice on a serving tray. To achieve this, two approaches are used as illustrated in Fig. 1: first, a vanilla CNN trained from scratch using new image datasets; and second, a pre-trained CNN fine-tuned with our specific data. The performance of image feature extraction in all networks varies significantly depending on the optimizer, type of pooling, application of image augmentation, dropout, and network size. Therefore, deriving optimal conditions tailored to specific image characteristics is crucial.

Such a personalized monitoring system is vital for efficient meal management. It helps reduce rice wastage, ensures that only the necessary amount is prepared, and achieves cost savings. Additionally, it enables the provision of customized diets that consider individual eating habits. This study focuses on developing the optimal neural network model to achieve these objectives.

Subjects and Methods

1. Building image datasets

For the purpose of deriving the optimal CNN for extracting features from images based on the grain serving sizes on a white tray, we have captured and secured image data for training and validating the CNN. As illustrated in Fig. 2, this dataset encompasses seven different labels corresponding to rice serving sizes of 0 g (0 kcal), 70 g (100 kcal), 140 g (200 kcal), 210 g (300 kcal), 280 g (400 kcal), 350 g (500 kcal), and 420 g (600kcal). Each label represents a distinct amount of grain served on the tray, allowing for a comprehensive evaluation of the CNN's ability to accurately extract features related to varying quantities of grain servings.

A total of 630 images were captured and used in this study, with each image being set to a resolution of 224×224 pixels. The entire image dataset was randomly split into training and validation sets at a ratio of 7:3, respectively.

2. Image feature extraction backbone

1) Vanilla convolutional Neural Network

In this study, a CNN was used to predict the quantity of cereal on a tray. The input image used for the experiment was reduced to a resolution of 13×13×256 pixels and fed into the neural network. A vector containing 256 elements extracted from each image was input into a Rectified Linear Unit (ReLU) layer, and finally, the softmax layer predicted the quantity of cereal, which could be 0 g, 70 g, 140 g, 210 g, 280 g, 350 g, or 420 g. The specific dimensions of each layer of the neural network and the applied hyperparameters are summarized in Table 1. This table provides a summary of the neural network's structure and hyperparameters, helping to understand and optimize the model's complexity and data processing flow. It also ensures the reproducibility of the research. The primary advantage of CNNs is their ability to extract local features, thereby reducing the size

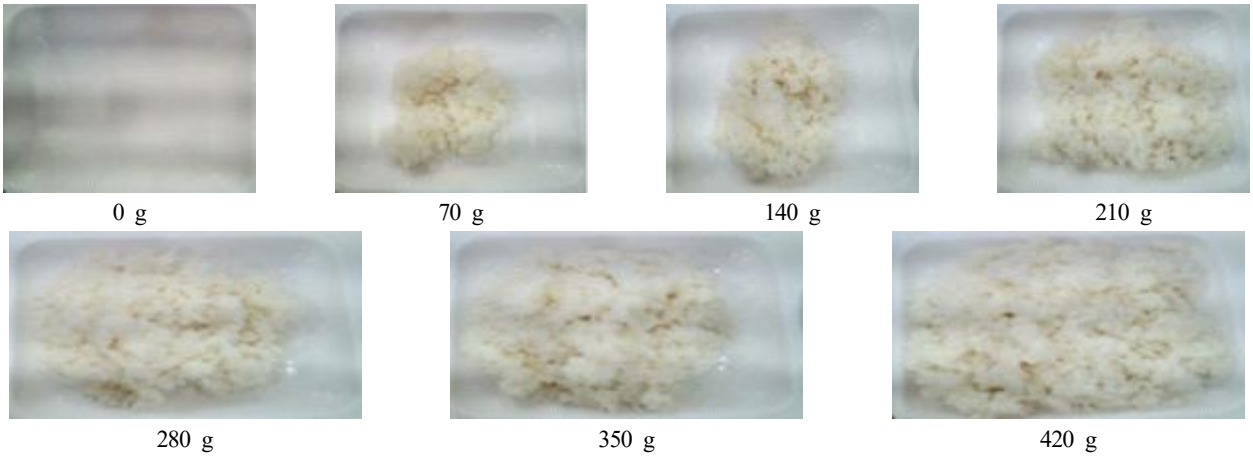


Fig. 2. Examples of images for rice quantity detection.

Table 1. Structure and hyperparameters of a vanilla convolutional neural Network used for rice quantity detection

Layer type	Output shape[4)]	Number of parameters (Total number of weights)
Conv2D_1[1)]	223,223,32	416
MaxPooling2D_1	111,111,32	0
Conv2D_2	110,110,64	8,256
MaxPooling2D_2	55,55,64	0
Conv2D_3	54,54,128	32,896
MaxPooling2D_3	27,27,128	0
Conv2D_4	26,26,256	131,328
MaxPooling2D_4	13,13,256	0
Flatten	43,624	0
Dense (ReLU)	256	11,075,840
Dense (softmax)	7	1,799

· Activation function: ReLU / Softmax
· Rate of learning: 1×10^{-5}
· Loss function: Categorical cross entropy
· Optimizer: RMSprop[2)] / Adam[3)]

[1)] Conv2D: convolutional 2D.
[2)] RMSprop: root mean square propagation.
[3)] Adam: adaptive moment estimation.
[4)] Output shape: The dimensions of the image data after it passes through each layer of the network.

of the input data and consequently decreasing the computational load. In this study, the pooling layers reduced the spatial size of the input array through a process known as downsampling. Max pooling selects the maximum value from a subset of the input array, whereas average pooling computes the average value. Based on the study by Dominik Scherer et al. (Scherer et al. 2010), which demonstrated that max pooling performs better than average pooling on image datasets, max pooling was employed in this study. Additionally, instead of sigmoid-shaped activation functions like y=tanh(x), which amplify nonlinearity and increase computational time, the ReLU activation function was adopted for its efficiency in reducing computational burden (Krizhevsky et al. 2017). Given that this study involves multi-class classification, the softmax function was utilized.

The detailed structure of the neural network and the specifics of each layer are as follows:

1. **Input Layer**: Receives image data with a resolution of 13×13×256 pixels.
2. **convolutional Layer**: Uses multiple filters to extract features from the image.
3. **Pooling Layer**: Reduces the size of the feature map using max pooling.
4. **Activation Layer**: Applies the ReLU activation function to introduce nonlinearity.
5. **Output Layer**: Utilizes the softmax function to predict the final quantity of cereal.

This study designed and trained a neural network using CNN to predict the quantity of cereal on a tray by applying appropriate techniques and hyperparameters at each layer.

2) Pre-trained convolutional Neural Networks

Pre-trained CNNs are networks that have been previously trained on large datasets, allowing them to act as generalized models that efficiently perform tasks even on images that are completely different from the ones they were originally trained on. Feature extraction involves using a network system that has been pre-trained on a large dataset to extract features from a new image dataset. Based on these extracted features, a custom classifier for the new image dataset is trained. Feature extraction can be further divided into fast feature extraction and feature extraction (Lin et al. 2011).

Fast feature extraction involves running a pre-trained CNN on a new image dataset and saving the output as a NumPy array on disk for use as input to a separate fully connected classifier. This method is efficient and cost-effective because it requires executing the computationally intensive convolution operations only once. However, it does not allow for the application of data generation to minimize overfitting.

Feature extraction involves extending a pre-trained CNN by stacking dense layers on top and then running the entire model end-to-end on new image data. This approach allows for the use of image generation, as all input images exposed to the model pass through the convolution base layers every time.

As of January 12, 2022, the Keras website lists 39 pre-trained deep learning models available for use alongside pre-trained weights. Among these, the pre-trained deep learning models widely used in the field of computer vision include ResNet50, EfficientNet, and InceptionV3. The ResNet50 model, developed by Microsoft, addresses the vanishing gradient problem and is composed of up to 100 layers. EfficientNet, a state-of-the-art (SOTA) model developed by Google, and InceptionV3, also developed by Google, are evaluated as efficient in terms of the number of parameters they generate and the computational cost incurred.

In this study, we utilize pre-trained CNN models to extract features from a new image dataset and train a custom classifier based on these features. We compare the differences between fast feature extraction and feature extraction methods and analyze the advantages and disadvantages of each approach.

3) Simulation methodology

In Convolutional Neural Networks (CNNs), optimizers are algorithms that modify network attributes such as weights and learning rates to reduce loss. In this study, the RMSprop and Adam optimizers were applied. The RMSprop optimizer is an extension of the gradient descent algorithm and uses the decaying average of partial gradients to adjust the step size for each parameter. Using decaying moving averages overcomes the limitations of adaptive gradient algorithms, where the algorithm forgets the initial gradient and focuses only on the most recent gradient during the search process (Kurbiel & Khaleghian 2017).

The Adam optimizer is an extension of stochastic gradient descent widely adopted in recent deep learning applications in the fields of computer vision and natural language processing. It combines the advantages of the adaptive gradient algorithm, which maintains the learning rate for each parameter to improve performance on sparse gradients, and the RMSprop algorithm, which adjusts the learning rate for each parameter based on the average of recent gradient sizes (Kingma & Ba 2014).

Overfitting occurs when there is insufficient training data, making it challenging to train a model that can generalize well to new data. Image generation is a technique used to increase the diversity of the dataset by generating similar image samples, thereby reducing overfitting. In this study, image generation was applied by varying the image rotation by ±20°, image height by ±10%, image width by ±10%, and image size by ±10%.

Dropout is one of the most effective and widely used regularization techniques for neural networks. Applying dropout to a network layer randomly excludes some of the layer's output features during training. The dropout rate is typically set between 0 and 0.5 (0 to 50%) (Srivastava et al. 2014). In this study, a dropout rate of 50% was applied to evaluate the impact of

dropout.

To derive the maximum performance of the vanilla CNN, we evaluated the changes in performance based on the optimizer, image generation, and dropout (50%). Similarly, to ascertain the maximum performance of 12 pre-trained CNNs, we assessed the variations in performance due to the optimizer, dropout (50%), image generation, feature extraction methods, and fine-tuning. The considered cases are presented in Table 2.

Results

1. Vanilla CNN

The changes in training and validation accuracy for the vanilla CNN, according to the type of optimizer, image generation, and dropout, are depicted in Fig. 3. Examining the training and validation accuracy graphs in Fig. 3 reveals that cases V1, V2, V5, and V6, where image generation was applied, show a phenomenon of underfitting, with validation accuracy higher than training accuracy. This indicates that the vanilla CNN did not fully learn the characteristics of the training image data. On the other hand, cases V3, V4, V7, and V8, where either dropout was applied or neither image augmentation nor dropout was applied, display overfitting, with training accuracy exceeding validation accuracy. The top-1 validation accuracy from Fig. 3 have been compiled in Table 3. Validation accuracy is measured as the ratio of predictions that match the true values across the entire validation dataset when the trained architecture predicts weight based on image features. The closer the predicted rice weight is to the true value, the higher the accuracy. Validation accuracies are automatically computed and presented during the training process.

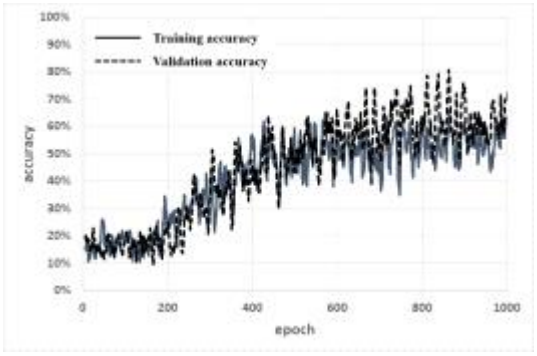
Generalization refers to how well a trained neural network performs on new data. The top-1 validation accuracy of V1, V2, V5, and V6 in Table 3 are results derived from underfitting, indicating insufficient generalization, which suggests a high level of uncertainty when applying to an actual dispensing system. Conversely, the top-1 validation accuracy of V3, V4, V7, and V8 in Table 3, which have relatively higher generalization from overfitting, were analyzed to be low at 55~60%. There is a significant difference between training accuracy and validation accuracy, indicating a gap in performance.

Table 2. Simulation cases of pre-trained and vanilla convolutional neural networks for rice quantity detection

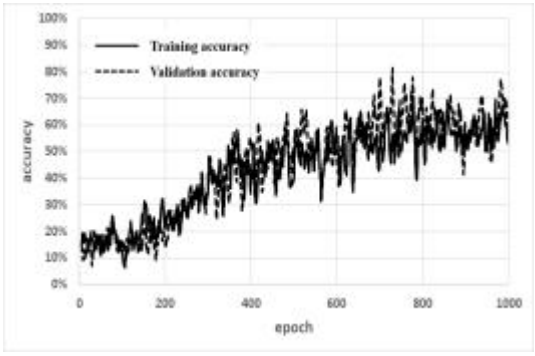
Case		Optimizer	Dropout	Image generation	Extraction type	Tunning
Vanilla CNN	Pre-trained CNN					
V1	P1	RMSprop[1)]	Yes	Yes	Feature extraction	No
-	P2					Yes
V2	P3			No	Fast feature extraction	No
-	P4					Yes
V3	P5		No	Yes	Feature extraction	No
-	P6					Yes
V4	P7			No	Fast feature extraction	No
-	P8					Yes
V5	P9	Adam[2)]	Yes	Yes	Feature extraction	No
-	P10					Yes
V6	P11			No	Fast feature extraction	No
-	P12					Yes
V7	P13		No	Yes	Feature extraction	No
-	P14					Yes
V8	P15			No	Fast feature extraction	No
-	P16					Yes

[1)] RMSprop: root mean square propagation.
[2)] Adam: adaptive moment estimation.

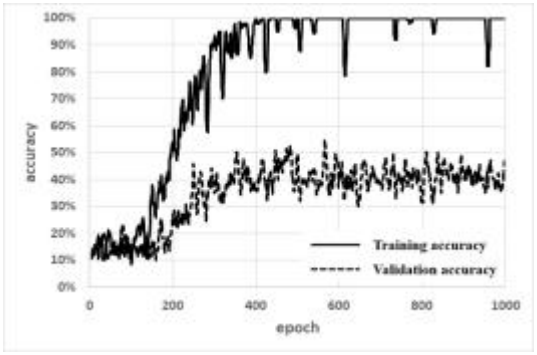
(A) Case V1



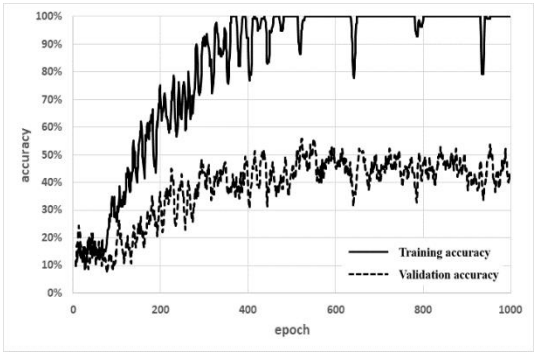
(B) Case V2



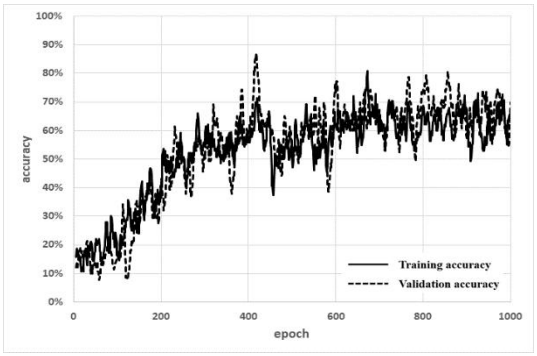
(C) Case V3



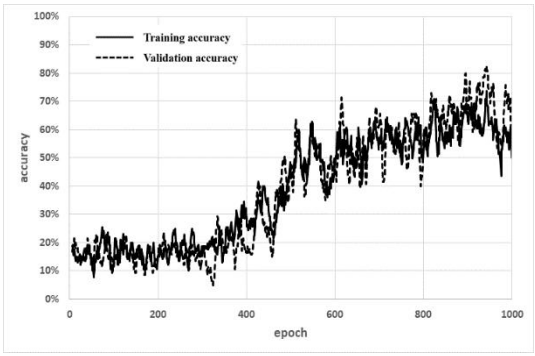
(D) Case V4



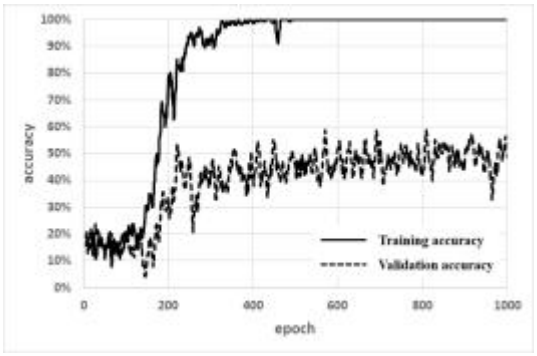
(E) Case V5



(F) Case V6



(G) Case V7



(H) Case V8

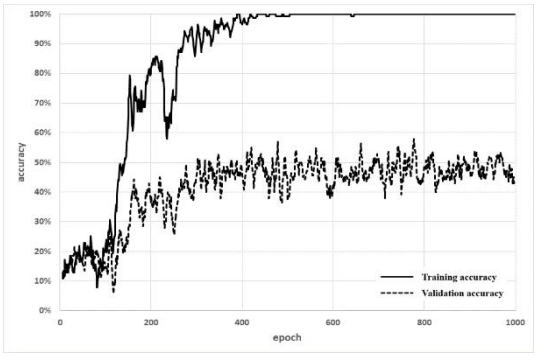


Fig. 3. Training and validation accuracy curves of vanilla convolutional neural network for rice quantity detection.

Table 3. Maximum validation accuracy[1] of vanilla convolutional neural network for rice quantity detection

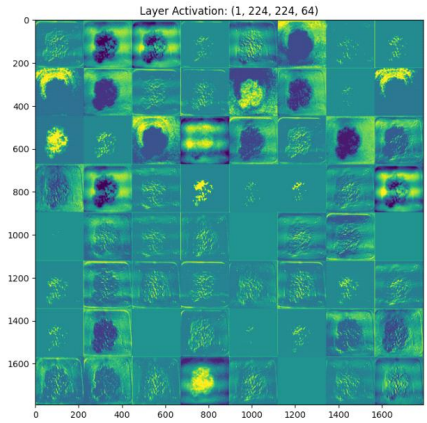
Remark	Case ID							
	V1	V2	V3	V4	V5	V6	V7	V8
Top-1 validation accuracy (%)	82%*	82%*	55%	55%	88%*	72%*	60%	58%

[1] Accuracy=Number of correct predictions/total number of predictions.
*Underfitting.

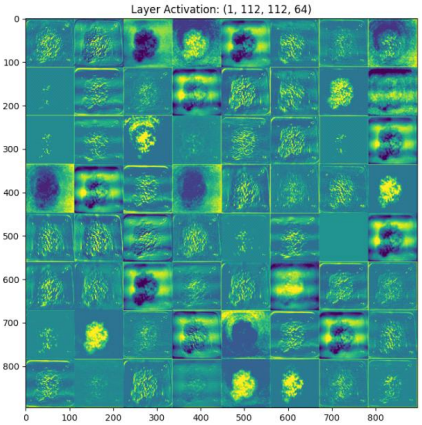
2. Pre-trained CNNs

As previously discussed, simulations were conducted for 16 scenarios involving twelve pre-trained CNN, including VGG19, considering variables such as image generation, fine-tuning, the application of dropout, optimizer types, and methods of feature extraction. CNN iterate through convolutional and pooling layers, generating feature maps at each stage. The activation results within the VGG19 pre-trained CNN based on the image input in Fig. 2 are summarized in Fig. 4. After passing through the first layer of the VGG19 pre-trained CNN, the feature map illustrated in Fig. 4A is derived from the results of 64 filters, maintaining all information of the initial input image. However, as we ascend to the higher layers of the VGG19 pre-trained CNN, the activations become increasingly abstract and visually challenging to comprehend. This transition results in a gradual reduction of information about the visual content of the image,

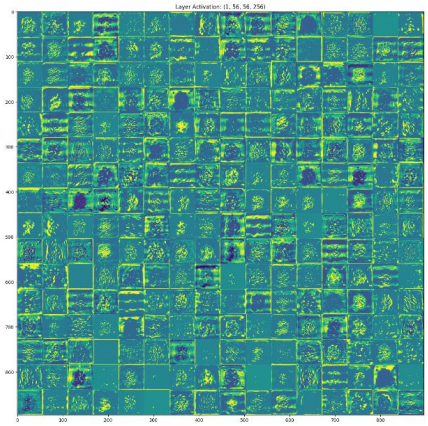
(A) After passing the block1_conv2 layer



(B) After passing the block1_pool layer



(C) After passing the block3_conv4 layer



(D) After passing the block3_pool layer

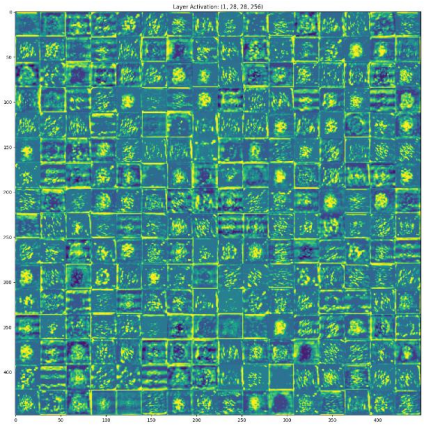


Fig. 4. Interlayer feature map of VGG[1]19 pre-trained convolutional neural network for rice quantity detection. [1] VGG: visual geometry group.

while information pertaining to the image's class progressively increases, as demonstrated in Fig. 4B, 4C, and 4D.

The results of simulations with 12 pre-trained CNN indicated that models such as VGG16 are suitable for image feature extraction in this study, with key outcomes displayed in Fig. 5. The top-1 validation accuracy of all pre-trained CNN are summarized in Table 4.

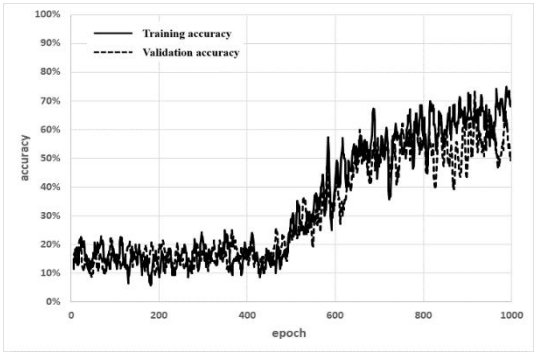
In Fig. 5, training and validation accuracy increases with learning iterations, and overfitting, which is important for generalization, is investigated. In Table 4, it was analyzed that ResNet50, ResNet152, EfficientB7, and MobileNetV3_small pre-trained CNN were not suitable for detecting the amount of grain ration on a white serving tray, which is the subject of this study.

InceptionV3, MobileNetV2_256, and MobileNetV2_1024 pre-trained CNN were found to faithfully extract grain image characteristics in all cases in Table 4. VGG16, VGG19, DesNet121, DesNet201, MobileNetV2_Large pre-trained CNN were evaluated to faithfully extract grain image characteristics only under certain conditions. , taking into account overfitting for generalization and minimizing the difference between training and verification accuracy.

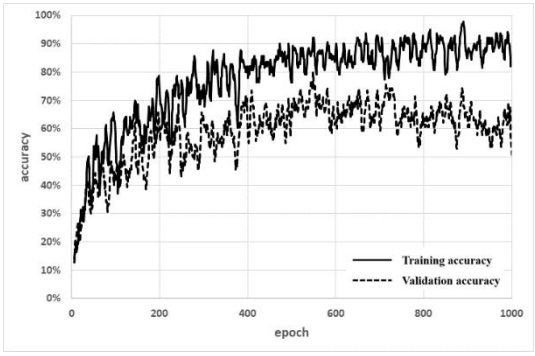
Table 5 shows that when applying the P5 and P13 conditions in Table 2 to the DesNet121 pre-trained CNN, top-1 verification accuracy of 90% is secured. P5 and P13 are cases where dropout and fine tuning are not applied, but image generation and general feature extraction are applied. The optimizer RMSprop and Adam types were analyzed to have no effect.

Pre-trained CNN exhibit excellent capabilities for image feature extraction but are associated with high computational demands. This presents a challenge when deploying these networks on mobile devices, where power consumption becomes a critical concern. In response to the design requirements of mobile and embedded vision applications, Andrew G. Howard and colleagues developed the MobileNetV2, MobileNetV3Large, and MobileNetV3Small pre-trained CNN, tailored to efficiently address these constraints (Howard et al. 2017; Sandler et al. 2018; Howard et al. 2019).

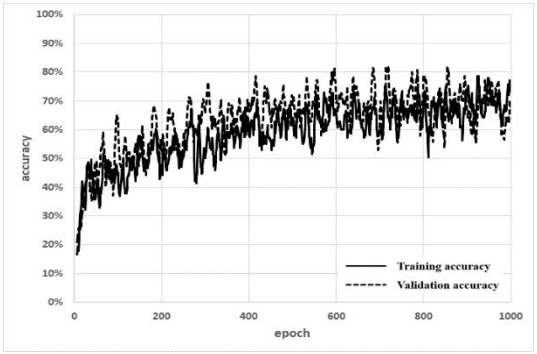
(A) VGG[1]19, Case P5



(B) DesNet[2]121, Case P5



(C) DesNet201, Case P1



(D) MobileNet[3]V2_1024, Case P5

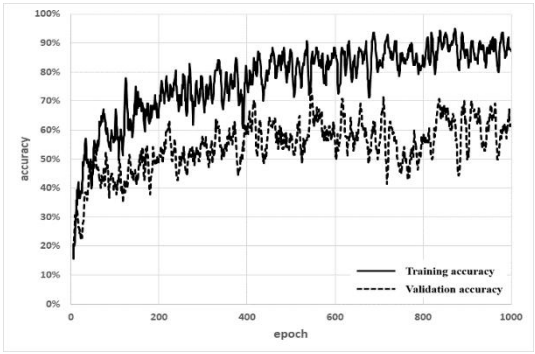


Fig. 5. Suitable pre-trained convolutional neural network for image feature extraction for rice quantity detection. [1] VGGnet: visual geometry group network. [2] DesNet: densely connected convolutional network. [3] MobileNet: mobile network.

Table 4. Top-1 validation accuracy[1] of pre-trained convolutional neural network for rice quantity detection

Case	ResNet[2]50	ResNet152	EfficientNetB7	InceptionV3	VGG[3]16	VGG19
P1	-	-	-	60%	-	-
P2	-	-	-	50%	-	-
P3	-	-	-	50%	-	-
P4	-	-	-	50%	-	-
P5	-	-	-	65%	72%	75%
P6	-	-	-	40%	-	-
P7	-	-	-	45%	55%	48%
P8	-	-	-	45%	50%	48%
P9	-	-	-	70%*	-	-
P10	-	-	-	75%*	-	-
P11	-	-	-	56%	58%	48%*
P12	-	-	-	55%	58%	55%
P13	-	-	-	65%	-	-
P14	-	-	-	45%	-	-
P15	-	-	-	45%	48%	51%
P16	-	-	-	40%	51%	41%

Case	DesNet[4]121	DesNet201	MobileNet[5]V2 256	MobileNetV2 1024	MobileNetV3 Large	MobileNetV3 Small
P1	75%*	80%*	60%	65%	-	-
P2	-	-	29%	25%	38%	-
P3	55%	60%	55%	58%	-	-
P4	55%	60%	55%	58%	-	-
P5	80%	73%	70%	75%	-	-
P6	40%	-	40%	21%	23%	-
P7	51%	60%	50%	48%	-	-
P8	58%	55%	50%	52%	-	-
P9	82%*	78%*	68%*	76%	-	-
P10	58%*	-	31%	28%	22%	-
P11	65%	60%	52%	58%	-	-
P12	65%	60%	52%	58%	-	-
P13	80%	90%	72%	73%	-	-
P14	32%	50%	30%	38%	22%	-
P15	55%	55%	50%	55%	-	-
P16	66%	50%	45%	48%	-	-

[1] Accuracy=Number of correct predictions/total number of predictions.
[2] ResNet: residual network.
[3] VGGNet: visual geometry group network.
[4] DesNet: densely connected convolutional network.
[5] MobileNet: mobile network.
-: unresponsive, *: underfitting.

Table 5. Top-1 validation accuracy[1] of pre-trained convolutional neural network for rice quantity detection

Network	Maximum validation accuracy	Case
InceptionV3	65%	P5
VGG[2]16	72%	P5
VGG19	75%	P5
DesNet[3]121	**80%**	**P5, P13**
DesNet201	**90%**	**P13**
MobileNet[4] V2 256	**72%**	**P13**
MobileNetV2 1024	**76%**	**P9**
MobileNetV3 Large	38%	P2

[1] Accuracy=Number of correct predictions/total number of predictions.
[2] VGGnet: visual geometry group network.
[3] DesNet: densely connected convolutional network.
[4] MobileNet: mobile network.

Our evaluation revealed that, among the MobileNet architectures, the MobileNetV3Large and MobileNetV3Small networks were not suitable for detecting rice serving amounts, as indicated in our results. Conversely, the MobileNetV2 network demonstrated a promising application in this context, achieving a top-1 validation accuracy of 76% for rice serving amount detection on mobile platforms, as assessed in our study.

Discussion

This study investigates the application of CNN not only for image classification but also as backbones for object detection and segmentation, focusing on extracting features from images of rice serving amount changes on white serving trays. We specifically used white trays because training a model to detect white rice on a white tray is considered more challenging than detecting white rice on trays of other colors or materials, such as aluminum. From this, we inferred that the model could achieve similar or even higher accuracy when detecting white rice on trays of different colors or materials.

The study further examines the training and validation accuracy of both vanilla and pre-trained CNNs, considering factors such as the type of optimizer, the application of image augmentation and dropout, and different methods of feature extraction. By analyzing these factors, we aim to identify the optimal conditions for accurately detecting and segmenting rice servings, which could then be generalized to various tray types and conditions.

In the application of vanilla CNN, the implementation of image generation has been observed to increase validation accuracy. However, an instance of underfitting is identified where the validation accuracy exceeds the training accuracy significantly at the number of iterations required to achieve top-1 validation accuracy, rendering it unsuitable for generalization necessary for real-world applications. Conversely, when only dropout is applied or neither image augmentation nor dropout is implemented, a typical case of overfitting occurs. Not only is there a significant difference between training and validation accuracy, but the top-1 validation accuracy is also found to be limited to a range of 55-60%, indicating suboptimal performance.

In the application of twelve pre-trained CNN, it was analyzed that ResNet50, ResNet152, EfficientB7, and MobileNetV3_small pre-trained CNNs are not suitable for detecting rice serving amounts. Conversely, the InceptionV3, MobileNetV2_256, and MobileNetV2_1024 pre-trained CNNs have been evaluated as effectively extracting the characteristics of rice serving amount images. The VGG16, VGG19, DesNet121, DesNet201, and MobileNetV2_Large pre-trained CNN were found to faithfully extract the characteristics of rice serving amount images only under specific conditions.

The derived top-1 validation accuracy of pre-trained CNN reached 90% in the case of the DesNet121 network when the RMSprop or Adam optimizer was applied, without the application of dropout and fine-tuning, and with the inclusion of image generation and general feature extraction techniques. For detecting rice serving amounts on mobile devices, the

MobileNetV2 network, among the MobileNets, was evaluated as highly suitable due to its ability to minimize the use of resources such as power on mobile devices.

The successful creation and deployment of a deep learning model for quantifying rice servings in institutional foodservice represent significant progress in the field of personalized nutrition management. The model's high validation accuracy suggests promising potential for effective diet management. In traditional Korean meals, the quantities of specific components such as rice, soups, main courses (primarily protein-based), side dishes (e.g., vegetables), and kimchi are often influenced by individuals' nutritional knowledge and health awareness. This variability underscores the need to analyze the nutritional content across various food groups.

It is thought that future research should expand to include main courses and side dishes to develop a comprehensive understanding of the entire meal's nutritional composition. Such technological advancements will enhance the ability to detect the nutritional content of the entire tray, thereby supporting efforts to tailor dietary intake to meet individual nutritional requirements and promoting personalized nutrition management.

Conflicts of Interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding Source

This work was supported by 2024 Bucheon University Research Grant.

References

Bolaños M, Radeva P. 2016. Simultaneous food localization and recognition. In 2016 23rd International Conference on Pattern Recognition (ICPR). pp.3140-3145. IEEE

Cai Q, Li J, Li H, Weng Y. 2019. BTBUFood-60: Dataset for object detection in food field. Available from https://doi.org/10.1109/BIGCOMP.2019.8678916 [cited 20 March 2024]

Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. 2017. Deeplab: Semantic image segmentation with deep convolution nets, atrous convolutional, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell* 40:834-848

Chollet F. 2017. Xception: Deep learning with depthwise separable convolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp.1251-1258. IEEE

Fragopoulou E, Detopoulou P, Alepoudea E, Nomikos T, Kalogeropoulos N, Antonopoulou S. 2021. Associations between red blood cells fatty acids, desaturases indices and metabolism of platelet activating factor in healthy volunteers. *Prostaglandins Leukot Essent Fatty Acids* 164:102234

Girshick R, Donahue J, Darrell T, Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp.580-587. IEEE

He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp.770-778. IEEE

Hou W, Han T, Sun X, Chen Y, Xu J, Wang Y, Yang X, Jiang W, Sun C. 2022. Relationship between carbohydrate intake (quantity, quality, and time eaten) and mortality (total, cardiovascular, and diabetes): Assessment of 2003-2014 National Health and Nutrition Examination Survey participants. *Diabetes Care* 45:3024-3031

Howard A, Sandler M, Chu G, Chen LC, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V, Le QV, Adam H. 2019. Searching for MobileNetV3. Available from https://arxiv.org/abs/1905.02244 [cited 20 March 2024]

Howard A, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. 2017. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint. Available from https://arxiv.org/abs/1704.04861 [cited 20 March 2024]

Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. 2017. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp.4700-4708. IEEE

Kalivaraprasad B, Prasad MVD, Vamsi R, Tejasri U, Santhoshi MN, PramodKumar A. 2021. Analysis of food recognition and calorie estimation using AI. In 1st International Conference on Advances in Signal Processing, Vlsi, Communications and Embedded Systems: Icsvce-2021

p.020020. AIP Publishing

Kawano Y, Yanai K. 2014. Food image recognition with deep convolutional features. In UbiComp '14 Adjunct: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp.589-593. Association for Computing Machinery

Kim MH. 2013. Characteristics of nutrient intake according to metabolic syndrome in Korean elderly-using data from the Korea national health and nutrition examination survey 2010. *Korean J Food Nutr* 26:515-525

Kingma DP, Ba J. 2014. Adam: A method for stochastic optimization. Available from https://arxiv.org/abs/1412.6980 [cited 20 March 2024]

Korean Nutrition Society. 2020. Dietary reference intakes for Koreans 2020. Available from https://www.kns.or.kr/fileroom/fileroom.asp?BoardID=Kdr [cited 20 March 2024]

Krizhevsky A, Sutskever I, Hinton GE. 2017. ImageNet classification with deep convolutional neural networks. *Commun ACM* 60:84-90

Kurbiel T, Khaleghian S. 2017. Training of deep neural networks based on distance measures using RMSProp. Available from https://arxiv.org/abs/1708.01911 [cited 20 March 2024]

Lin Y, Lv F, Zhu S, Yang M, Cour T, Yu K, Cao L, Huang T. 2011. Large-scale image classification: Fast feature extraction and SVM training. In 2011 IEEE Conference on on Computer Vision and Pattern Recognition CVPR 2011. pp.1689-1696. IEEE

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. 2016. SSD: Single shot multibox detector. In Computer Vision - ECCV 2016: 14th European Conference on Computer Vision. pp.21-37. Springer

Lu Y, Stathopoulou T, Vasiloglou MF, Pinault LF, Kiley C, Spanakis EK, Mougiakakou S. 2020. goFOOD™: An artificial intelligence system for dietary assessment. *Sensors* 20:4283

Matusheski NV, Caffrey A, Christensen L, Mezgec S, Surendran S, Hjorth MF, McNulty H, Pentieva K, Roager HM, Seljak BK, Vimaleswaran KS, Remmers M, Péter S. 2021. Diets, nutrients, genes and the microbiome: Recent advances in personalised nutrition. *Br J Nutr* 126:1489-1497

Mezgec S, Koroušić Seljak B. 2017. NutriNet: A deep learning food and drink image recognition system for dietary assessment. *Nutrients* 9:657

National Academies of Sciences, Engineering, and Medicine. 2016. Review of WIC Food Packages: Proposed Framework for Revisions: Interim Report. The National Academies Press

Pan B, Zhao N, Xie Q, Li Y, Hamaker BR, Miao M. 2023. Molecular structure and characteristics of phytoglycogen, glycogen and amylopectin subjected to mild acid hydrolysis. *npj Sci Food* 7:27

Peano C, Girgenti V, Sciascia S, Barone E, Sottile F. 2022. Dietary patterns at the individual level through a nutritional and environmental approach: The case study of a school canteen. *Foods* 11:1008

Ronneberger O, Fischer P, Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention: MICCAI 2015. pp.234-241. Springer

Sahoo D, Hao W, Ke S, Xiongwei W, Le H, Achananuparp P, Lim E, Hoi SC. 2019. FoodAI: Food image recognition via deep learning for smart food logging. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp.2260-2268. Association for Computing Machinery

Sandler M, Howard AG, Zhu M, Zhmoginov A, Chen LC. 2018. MobileNetV2: Inverted residuals and linear bottlenecks. Available from https://arxiv.org/abs/1801.04381 [cited 20 March 2024]

Santamarina AB, Mennitti LV, de Souza EA, de Souza Mesquita LM, Noronha IH, Vasconcelos JRC, Prado CM, Pisani LP. 2023. A low-carbohydrate diet with different fatty acids' sources in the treatment of obesity: Impact on insulin resistance and adipogenesis. *Clin Nutr* 42:2381-2394

Scherer D, Müller A, Behnke S. 2010. Evaluation of pooling operations in convolutional architectures for object recognition. In Artificial Neural Networks - ICANN 2010. pp.92-101. Springer

Simonyan K, Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition. Available from https://arxiv.org/abs/1409.1556 [cited 20 March 2024]

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 15: 1929-1958

Stubbs RJ. 2021. Impact of carbohydrates, fat and energy density

on energy intake. *Nat Med* 27:200-201

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. 2015. Going deeper with convolution. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp.1-9. IEEE

Tan M, Le QV. 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. Available from https://arxiv.org/abs/1905.11946 [cited 20 March 2024]

Vasiloglou MF, Mougiakakou S, Aubry E, Bokelmann A, Fricker R, Gomes F, Guntermann C, Meyer A, Studerus D, Stanga Z. 2018. A comparative study on carbohydrate estimation: GoCARB vs. dietitians. *Nutrients* 10:741

Wali JA, Ni D, Facey HJW, Dodgson T, Pulpitel TJ, Senior AM, Raubenheimer D, Macia L, Simpson SJ. 2023. Determining the metabolic effects of dietary fat, sugars and fat-sugar interaction using nutritional geometry in a dietary challenge study with male mice. *Nat Commun* 14:4409

Yang Q, Lang X, Li W, Liang Y. 2022. The effects of low-fat, high-carbohydrate diets vs. low-carbohydrate, high-fat diets on weight, blood pressure, serum liquids and blood glucose: A systematic review and meta-analysis. *Eur J Clin Nutr* 76:16-27

Yeom MY, Choi EY. 2023. Correlation of the nutrition quotient between parents and picky eaters in preschoolers. *Korean J Food Nutr* 36:103-113

Received 22 July, 2024
Revised 12 August, 2024
Accepted 14 August, 2024