

## Enhancing Thai Food Classification: A CNN-Based Approach with Transfer Learning

Korakot Matarat 

Faculty of Science and Technology, Sakon Nakhon Rajabhat University, Sakon Nakhon 47000, Thailand

Corresponding Author Email: [korakot@snru.ac.th](mailto:korakot@snru.ac.th)

Copyright: ©2024 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.110625>

### ABSTRACT

**Received:** 15 January 2024

**Revised:** 9 April 2024

**Accepted:** 25 April 2024

**Available online:** 22 June 2024

#### Keywords:

*Thai cuisine, machine learning, food image classification, Convolutional Neural Network, Thai food recognition, image recognition*

In this research paper, we delve into the classification of Thai cuisine images. Despite Thailand's renowned reputation for its multicultural culinary landscape, there is a noticeable gap in dedicated studies on Thai food classification. This paper seeks to fill that void by applying deep learning methodologies, specifically Convolutional Neural Networks (CNNs), to the identification of Thai cuisine. Thai cuisine, shaped by regional and intra-regional variations, serves as a powerful cultural representation for the nation. The study employs image recognition through CNN and integrates transfer learning to enhance classification performance. The collaborative learning process between CNN and transfer learning contributes to achieving a noteworthy accuracy rate of 84%. While previous research has often overlooked the specificity of Thai cuisine, our aim is to shed light on the potential of deep classification networks, offering an engaging illustration for both researchers and food enthusiasts alike contributing to the broader field of food image classification.

## 1. INTRODUCTION

Thailand, a country celebrated for its rich cultural tapestry and vibrant traditions, stands as a culinary haven, offering array of delectable dishes that reflect its multicultural heritage. Thai cuisine, renowned for its intricate flavors and harmonious blend of spices, holds a distinct place in the global gastronomic landscape. Despite the nation's reputation for its culinary prowess and the significant role Thai food plays as a cultural ambassador, research specifically focused on the classification of Thai cuisine remains limited.

This paper, delving into the realm of Thai food classification using cutting-edge deep learning methodologies. Our exploration is fueled by the desire to shed light on the intricacies of Thai culinary artistry and contribute to the broader field of food image classification. The study acknowledges the unique regional and intra-regional variations within Thailand, recognizing the diverse nature of Thai gastronomy. Drawing inspiration from the success of deep learning techniques in various image recognition domains, we employ Convolutional Neural Networks (CNNs) as a robust methodology for Thai food image classification. Furthermore, we leverage the power of transfer learning to enhance the classification performance, acknowledging the potential of pre-trained models to discern of Thai culinary aesthetics.

As we embark on this, our primary goal is to not only achieve a significant accuracy rate in Thai food classification but also to provide a visually engaging illustration of the capabilities of deep classification networks. Through this research, we aim to offer technologists alike an insightful exploration of Thai cuisine, blending the realms of advanced machine learning. Join us in savoring the flavors of technology

and tradition, as we unravel the secrets of Thai food through the lens of Convolutional Neural Networks and transfer learning.

This paper's structure is as follows: The paper includes five sections. Section 2 provides a literature review that is pertinent to the subject matter. Section 3 elaborates on the proposed techniques. Section 4 presents experimental results. Finally, conclusions and prospective research can be found in Section 5.

## 2. LITERATURE REVIEW

The realm of food image recognition has undergone development in recent years, fueled by advancements in deep learning techniques, Convolutional Neural Networks (CNNs), and transfer learning. This literature review presents a comprehensive exploration of various methodologies and models proposed for accurate and efficient food recognition.

In the pursuit of improving Asian cuisine image classification, a method was proposed that integrates the Convolutional Block Attention Module (CBAM) with Mobile NetV2, VGG16, and ResNet50. Additionally, a Mix up data enhancement algorithm was employed to refine discrimination capabilities. The combined approach yielded a Top-1 accuracy rate of 87.33%, validating its effectiveness in categorizing Asian cuisine images [1]. The utilization of transfer learning in culinary image identification showcased promising results. A pre-trained Inception v3 CNN model was implemented to enhance the learning process, achieving an impressive accuracy of 97.00% across twenty food classes. This approach demonstrated the optimization potential of transfer learning in custom-built CNN frameworks [2]. In the context of pest



identification, the study focused on the ResNeXt-50 (32\*4d) model, exploring the impact of transfer learning and data augmentation. The findings emphasized the significant influence of transfer learning on classification accuracy, with a notable increase to 86.95% through the integration of fine-tuning [3]. Examining the broader scope of image classification, the study employed various pre-trained deep learning models, including AlexNet, GoogleNet, VGG16, DenseNet, and ResNet. Notably, the VGG16 model exhibited superior performance among the utilized models, showcasing the effectiveness of transfer learning in image classification scenarios [4]. The research categorized image classification algorithms into noise model-based and noise model-free approaches. While the former aimed at estimating noise structure, the latter leveraged learning paradigms like regularizers and robust losses to develop algorithms inherently resistant to noise. This classification provides insights into strategies for handling noisy labels in image datasets [5]. The study highlighted the challenges in recognizing and validating a vast variety of food products and employed CNN for food estimation and detection. The commendable performance of the CNN model, with an accuracy of 0.988 and low loss of 0.102, showcased its potential in addressing the complexities of culinary image processing [6]. This document provides a brief overview of how technology impacts human life, discussing both its positive and negative effects. It outlines the process of conducting a systematic review focused on food intake and health, particularly through image recognition and analysis. The document highlights the potential of technology in addressing dietary concerns and promoting healthier lifestyles [7]. The incorporation of multi-spectral images for automated food identification showcased the potential of Convolutional Neural Networks (CNNs). Achieving high accuracy (99.81%) in classifying food types with images acquired at various wavelengths, the study highlighted the significance of spectral information in enhancing precision [8]. Advancements in machine vision, image processing, and deep learning models were explored in the context of optimizing food processing efficiency. The paper provided a comprehensive examination of conventional and deep learning approaches, emphasizing their potential applications in the food processing domain [9]. Introducing optimization components, the study proposed a CNN-based culinary image recognition model with Particle Swarm Optimization (PSO) and Genetic Algorithm (GA). Achieving a classification accuracy of 82.3% using the UEC FOOD-100 dataset, the model showcased the potential of optimization techniques in improving accuracy [10]. The study focused on image processing and SVM-based separation of culinary products. Results demonstrated the accuracy and sensitivity of the regional measurement technique, providing a methodology for efficient surveillance and improvement of diverse parameters in culinary product separation [11]. Addressing the scarcity of food segmentation datasets, the study introduced a MobileNetV2-based model for food segmentation. With an ideal accuracy of 93.06% for food classification, the research contributed to the development of efficient segmentation techniques in culinary image processing [12]. Utilizing pretrained CNNs (ResNet-152 and GoogleNet) to extract features for culinary image classification, the study achieved 99.4% precision using ResNet-152 deep features with SVM and the RBF kernel. The research emphasized the effectiveness of deep CNN features in classifying various food items [13]. This document introduces a smartphone-based

system designed to aid children with visual impairments in recognizing food dishes and fruits in real-time. It presents a novel deep Convolutional Neural Network (CNN) model, developed through ensemble learning and trained on a customised dataset containing 29 varieties of food items. The model achieves a high accuracy of 95.55% in recognising food items. Additionally, the performance of various state-of-the-art CNN models for food recognition is evaluated using transfer learning, demonstrating the efficacy of the proposed approach [14]. This document surveys methods for food category recognition, focusing on core components like datasets and machine learning algorithms. Emphasizing deep learning techniques, it aims to advance research and industrial applications in this field [15]. Conducting a systematic review, the paper presented an overview of CNN models in culinary image processing. The review identified challenges, available datasets, optimization techniques, and proposed resolutions for each investigation encountered, contributing to a comprehensive understanding of the field [16]. Applying transfer learning techniques to classify images in an Indian culinary dataset, the study achieved significant cost and labor savings. InceptionV3 outperformed other models with an accuracy of 87.9%, showcasing the advantages of leveraging pre-trained models for specific cultural cuisines [17]. The research focused on caloric estimation in Malaysian dishes using the AlexNet CNN with transfer learning. Achieving a precision of 91.43%, the study highlighted the potential of deep learning in estimating nutritional content based on image recognition of specific culinary items [18]. Replicating CNNs on the RaspberryPi 3B platform for mobile-like environments, the study aimed to accelerate the processing of culinary images. The implementation reduced processing time to 3.3 seconds per image using PeachPy, showcasing the feasibility of deploying CNNs on resource-constrained devices [19]. Employing transfer learning techniques for image classification on a diverse culinary dataset, the study achieved an approximate accuracy of 84%. The results emphasized the significance of utilizing transfer learning to enhance the accuracy of culinary image classification [20]. Introducing the DenseFood model based on a densely connected Convolutional Neural Network, the study employed a combination of softmax loss and center loss. Experimental results demonstrated the DenseFood model's significantly higher accuracy (81.23%) compared to alternative models (DenseNet121 and ResNet50) [21]. This document introduces the concept of food computing, which involves analysing large-scale food data acquired from various sources for multiple purposes such as perception, recognition, and recommendation. It highlights the importance of computational approaches in addressing food-related challenges across different fields. Additionally, it provides a comprehensive overview of emerging concepts, methods, and tasks in food computing while identifying key challenges and future directions. Overall, it serves as a valuable resource for researchers and practitioners working in food-related areas [22]. Conducting a survey of 11 food object recognition systems implemented on mobile devices, the study categorized distinctive features and stages of the object recognition process. The survey aimed to guide developers and researchers working on food object recognition for specific use cases [23]. Comparing the VGG16, VGG19, and ResNet50 architectures for product classification, the study concluded that the ResNet50 architecture outperformed others. Achieving accuracies of 0.9733 at epoch 20, the research provided

insights into the effectiveness of different CNN architectures for practical issues in product classification [24]. Proposing an AI-based image classification system for expedited invoicing in the retail sector, the study compared various classifiers. The proposed system, incorporating preprocessing with the KNN classifier, exhibited enhanced precision (93.103%) in classifying fruit images compared to other classifiers [25]. Focused on enhancing the accuracy of culinary image classification, the study developed a recognition model using Efficientnetb0 and achieved an accuracy rate of 80% for 101 unique culinary varieties. The model demonstrated superior accuracy compared to other cutting-edge models [26]. Introducing AlsmViT, a vision transformer-based approach for food image classification, the study emphasized feature enhancement and data augmentation. Achieving high accuracies (95.17% and 94.29%) on Food-101 and Vireo Food-172 datasets, AlsmViT addressed challenges in handling foods with similar appearances but different nutritional values [27]. Exploring CNN applications in food detection and analysis, the study compared CNN performance with other methods. The research identified the potential of CNN, combined with nondestructive detection techniques, to effectively detect and analyze complex food matrices, presenting future trends in the domain [28]. Proposing a dual-stage learning paradigm for YOLO-SIMM in culinary image recognition, the study achieved overall recognition accuracy and introduced a method for detecting foreign bodies in food. The proposed technique effectively separated foreign bodies from food using threshold segmentation technology [29]. Introducing a novel framework for ingredient segmentation, the study utilized a CNN-based Single Ingredient Classification Model. Through comprehensive experiments, the proposed method demonstrated effective segmentation with optimal results on the FoodSeg103 dataset. The methodology set the groundwork for subsequent ingredient identification [30].

In conclusion, the literature review highlights the diverse methodologies and advancements in food recognition research. From transfer learning and CNN-based models to innovative

segmentation and object recognition approaches, these studies collectively contribute to the ongoing development of accurate and efficient systems for culinary image processing. The continuous evolution of deep learning techniques and the integration of novel methodologies promise further breakthroughs in the field of food recognition.

### 3. PROPOSED TECHNIQUE

In this investigation into the classification of food images, a CNN model utilizing transfer learning was proposed. Presently, a variety of deep Convolutional Neural Network architectures are available for implementation by anyone and vary in complexity of number of layers. In recent years, transfer learning concepts implemented in CNN architectures, such as fine-tuning and have outperformed conventional machine learning models for image classification tasks in terms of performance. The implementation of transfer learning techniques could potentially enhance the performance of our dataset's very deep network architecture.

#### 3.1 Techniques

Convolutional Neural Networks (CNNs) are extensively employed in image recognition due to their classification as deep neural networks. Comprising foundational layers, including hidden and fully connected layers, CNNs utilize hidden layers for feature extraction from training images and fully connected layers for image classification [20]. In this study, the Conv2D layer is employed to generate a feature map or kernel through a comprehensive scan of the input image. The pooling layer, while retaining information from the previous layer, reduces the size of the output. The flatten layer plays a crucial role in preparing the multidimensional output for input to fully connected layers, transforming it to a single dimension. This layer accepts data from each input, maintains connections with each output node, assigns unique weights to connections, and has the ability to assign the appropriate activation to each output node, as illustrated in Figure 1.

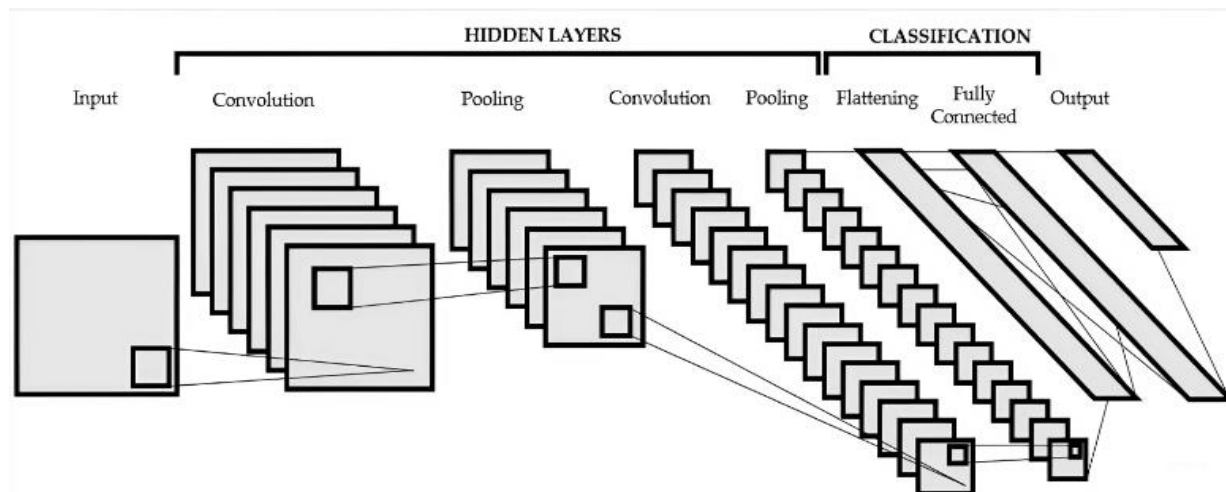
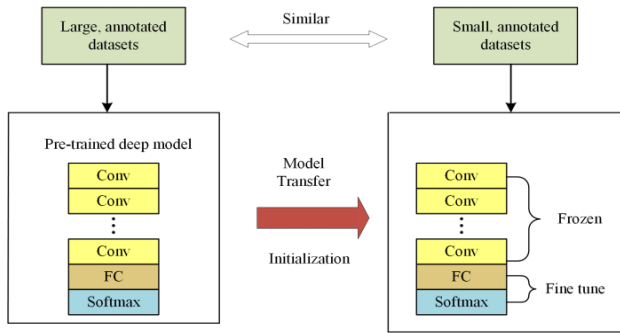


Figure 1. CNN architecture [6]

Transfer learning is a strategy that enhances efficiency by repurposing a pre-existing model previously trained on a large dataset for a related. The core idea is to leverage knowledge gained from the original domain to improve performance in a new domain. In this approach, a pretrained model from the

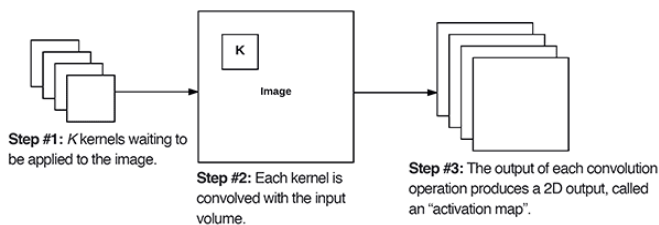
source domain is selected, serving as the starting point for a model in the target domain. Both the entire pretrained model or specific parts of it can be utilized. Fine-tuning is a common practice, involving adjustments to weights and biases in certain layers while keeping others frozen. This ensures

effective adaptation of the pretrained model to the specific input-output pairs of the new domain during training, as illustrated in Figure 2.



**Figure 2.** Transfer learning by using fine-tuning [26]

Fine-tuning is a technique in neural network modeling that involves initiating a new model for a specific task while benefiting from the knowledge acquired through a pre-trained deep neural network. The pre-trained model encapsulates learned features, weights, and biases [26]. In our approach, we employed fine-tuning by incorporating the Conv2D layer and the MobileNet model, reshape each image into a 256×256 shape. We also perform the normalization of pixel values. The primary goal was to leverage knowledge obtained from a previous task for a new one. Our strategy involved using the convolutional base to extract features from input images, which were then applied to a multi-class classification task. To align the model with our specific application, we maintained most of the pre-trained model's architecture unchanged. This process is illustrated in Figure 3. In essence, fine-tuning allows us to build upon the expertise gained from a broader task and tailor it to the nuances of a more specific application.



**Figure 3.** Keras Conv2D

### 3.2 Adam optimization algorithm for deep learning

The Adam optimization algorithm, widely used in deep learning, combines principles from Momentum and RMSprop. It efficiently adjusts model parameters during training by computing running averages of gradients and squared gradients for each parameter. Adaptive modification of learning rates is based on these averages. The update rule for each parameter involves a combination of the momentum term and a scaled gradient, adjusted by the square root of the exponentially decaying moving average of squared gradients. This adaptive approach makes Adam a favored choice in optimizing machine learning models. This is a component of the investigation.

### 3.3 Classification

The model undergoes 10 training epochs, employing Keras's Adam Optimization Algorithm with a determined

optimal learning rate of 0.0001. The training utilizes a categorical cross-entropy loss function, specifically chosen for the model's task of multi-class image classification, allowing fine-tuning of the final layer to adapt it to our specific classification needs.

### 3.4 Performance evaluation

We use the assessment process to see how well we're doing. The confusion matrix is a tool for determining how well a model performed. Eqs. (1)-(4) contain the formulas for this evaluation confusion matrix.

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

Recall is the proportion of genuine positives that were accurately detected. It was determined using the formula:

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (2)$$

Precision is the proportion of correct affirmative identifications. It was determined using the formula:

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (3)$$

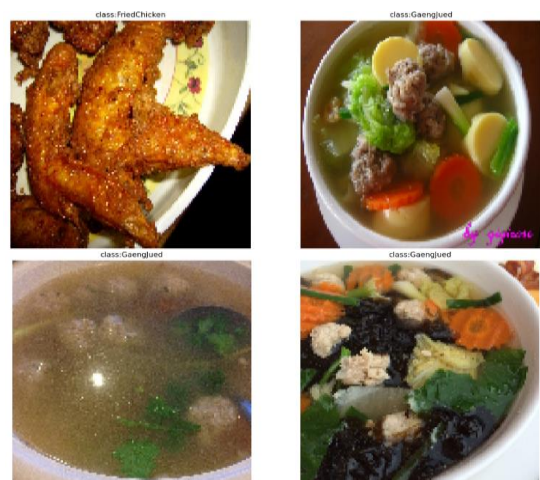
where, TP=True Positive, sentiments that are positive and are actually classified as 1; TN=True Negative, sentiments that are negative and are actually classified as 0; FP=False Positive, sentiments that are negative but are classified as 1; FN=False Negative, sentiments that are positive but are classified as 0;

The F1 score is the harmonic mean of precision and recall. It is given by the formula [2]:

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

## 4. EXPERIMENTAL RESULTS

For the research, Thai culinary dataset, which consists of twenty distinct culinary classes. The dataset comprises 3,677 training preview image samples and 1,097 testing image samples. Notably, these example images are rich in color. Figure 4 visually presents representative color images of culinary samples from the dataset, providing a glimpse of the diverse and colorful nature of the included images [31].



**Figure 4.** Sample food images used in the dataset

The input image data undergoes a series of two-dimensional convolutions and pooling operations. In a Convolutional Neural Network (CNN) designed for classifying 20 types of food images, the process begins with input images of size 256×256 pixels in RGB format. We then flatten the output of convolution and pooling operations to construct a dense network. The SoftMax Layer, the network's final layer for classification into 20 categories, transforms the outputs from the previous dense layer into probabilities for each class, determining its parameter count based on the number of inputs from the last dense layer and the total number of output classes (20). The specific number of parameters at each stage depends on the architecture's detailed configuration, such as the size and number of filters in convolutional layers and the size of dense layers. The image below Figure 5 depicts the number of parameters we need to learn at each layer of our CNN model.

After training the CNN model, we look at two key things: the loss and the accuracy. The loss tells us how well the model is doing during training, while the accuracy tells us how accurate its predictions are. In this case, after 5 training cycles (epochs), the loss doesn't really change much as shown in Figure 6. This suggests that the model isn't getting much better at learning from the training data. The accuracy of the validation data, which is new data the model hasn't seen before, only changes a little after 8 epochs. This means the model's performance on new data isn't improving much beyond that point. So, it seems like the model isn't improving much after a

certain point in training. This information helps us understand how well our CNN model is learning and where it might need adjustments.

```
Model: "sequential"
-----
```

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 256, 256, 128)	3584
max_pooling2d (MaxPooling2D)	(None, 128, 128, 128)	0
conv2d_1 (Conv2D)	(None, 128, 128, 32)	36896
max_pooling2d_1 (MaxPooling2D)	(None, 64, 64, 32)	0
conv2d_2 (Conv2D)	(None, 64, 64, 64)	18496
max_pooling2d_2 (MaxPooling2D)	(None, 32, 32, 64)	0
conv2d_3 (Conv2D)	(None, 32, 32, 64)	36928
max_pooling2d_3 (MaxPooling2D)	(None, 16, 16, 64)	0
dropout (Dropout)	(None, 16, 16, 64)	0
flatten (Flatten)	(None, 16384)	0
dense (Dense)	(None, 256)	4194560
dense_1 (Dense)	(None, 20)	5140

```
-----
Total params: 4295604 (16.39 MB)
Trainable params: 4295604 (16.39 MB)
Non-trainable params: 0 (0.00 Byte)
-----
```

Figure 5. CNN architecture with parameter count

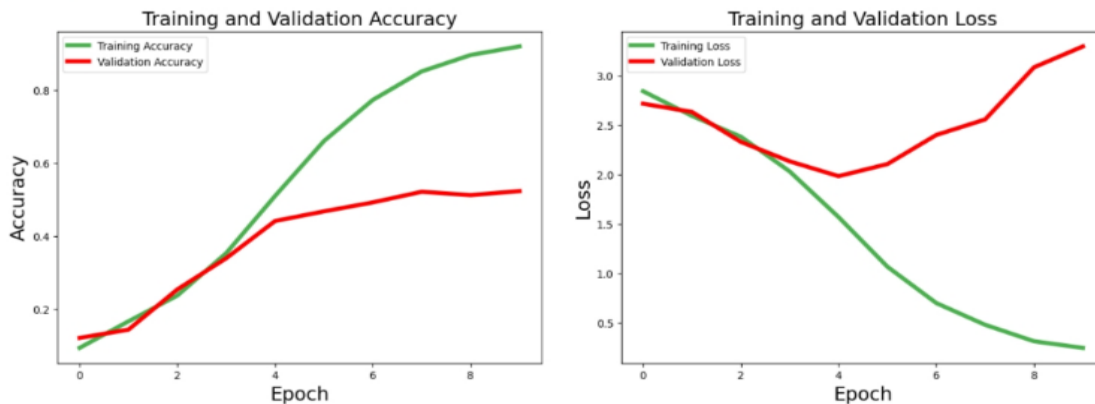


Figure 6. Training and validation accuracy and loss with Conv2D

```
Model: "sequential_1"
-----
```

Layer (type)	Output Shape	Param #
mobilenetv2_1.00_224 (Functional)	(None, 8, 8, 1280)	2257984
global_average_pooling2d (GlobalAveragePooling2D)	(None, 1280)	0
dropout_1 (Dropout)	(None, 1280)	0
dense_2 (Dense)	(None, 20)	25620

```
-----
Total params: 2283604 (8.71 MB)
Trainable params: 25620 (100.08 KB)
Non-trainable params: 2257984 (8.61 MB)
-----
```

Figure 7. Model architecture with transfer learning

Leveraging Transfer Learning (TL) to Drive Enhancement:

A mere average performance has been attained with our rudimentary CNN model. TL involves the application of knowledge acquired during the resolution of one problem to a distinct yet interconnected problem. When training requires a negligible quantity of data, TL is of assistance. As an example, for training purposes, we have only 200 images of each

culinary item. Potentially beneficial for enhancing classification performance is TL. In this instance, pre-trained weights from a model trained on the ImageNet dataset are utilized. Our input and output layers are utilized in lieu of those of the pre-trained model.

This architecture enhances a rudimentary CNN model for culinary item classification using transfer learning (TL). TL leverages knowledge from a pre-trained model dataset to improve performance despite limited training data (200 images per category). The pre-trained model's convolutional layers are used, while the input and output layers are customized for the culinary item classification task. This adaptation boosts CNN's performance, addressing the initial model's average performance due to its simplicity and lack of specialized design. In this model, there will be no retraining of the current model weights. We will train only the final stratum. Figure 7 illustrates the tally of learnable and non-learnable parameters.

Once the TL-based model has been trained, we conduct a training performance comparison between our CNN model

and the TL-based model. The training accuracy and loss values for the simple CNN and TL-based models with respect to epochs are depicted in the following graphs. The training accuracy of the TL-based model is greater during the validation phase. The training accuracy of TL-based models

increases progressively, in contrast to simple CNN models. The loss values exhibit a continual decrease in the TL-based model.

The model was trained for 10 epochs, and the training results, i.e., train accuracy and loss, are presented in Figure 8.

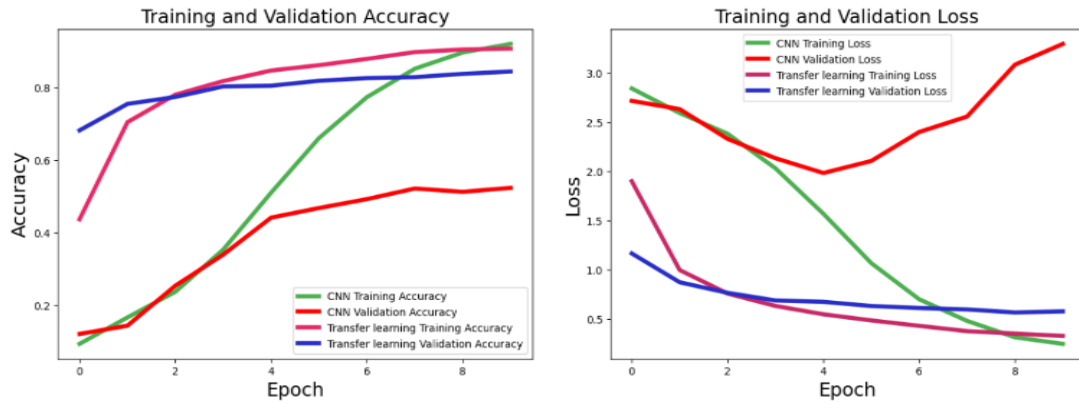


Figure 8. Training and validation accuracy and loss

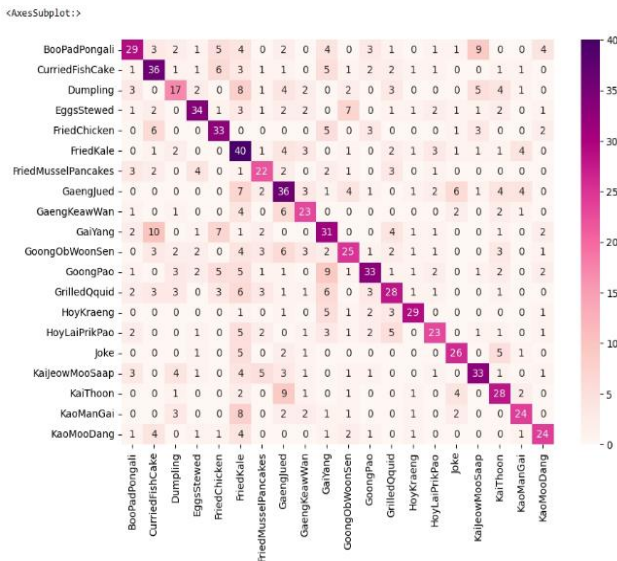


Figure 9. Confusion matrix of fine-tuned Conv2D

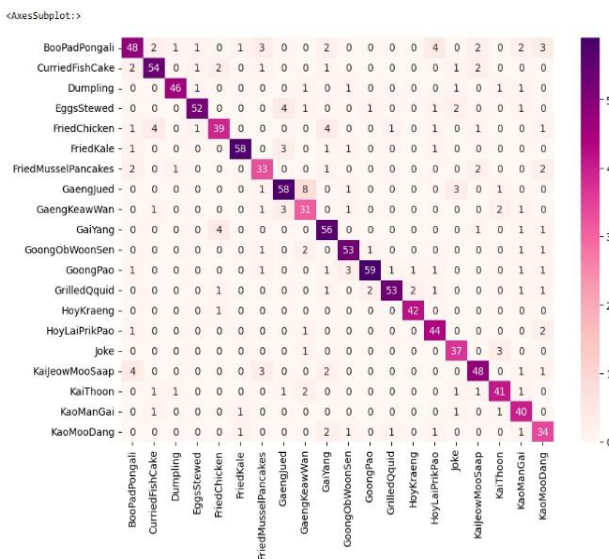


Figure 10. Confusion matrix of fine-tuned MobileNet

The accuracy of the training improved as the number of epochs increased until the 9 epoch, at which point it remained nearly constant. The epoch with the lowest train loss and highest train accuracy was identified as the 8. Subsequent to that, a marginal decline in train loss and an increase in train accuracy were attributed to overfitting. Training for more than ten epochs would result in significant overfitting. Following training, the accuracy and loss of the test more than 80 and 0.57, respectively.

To enhance the proposed Convolutional Neural Network (CNN) model, various strategies can be considered, including augmenting training data, adjusting image dimensions, increasing the number of deep layers, epochs, and scaling parameters. However, caution is needed as raising the number of epochs and layers may lead to overfitting, diminishing the model's generalization capability. Additionally, increasing image scale and the number of layers can extend training time, affecting computational efficiency. Striking a balance between these factors is crucial for optimizing the CNN model's performance without introducing potential drawbacks.

The outcome of evaluating the model is presented in Table 1. It is observed that the accuracy and f1-score values of MobileNet increase when transfer learning is applied to the network. Furthermore, following fine-tuning with the new dataset, the loss experienced a marginal decrease.

Table 1. Evaluation result

Conv2D		MobileNet	
Accuracy	f1-score	Accuracy	f1-score
0.52	0.53	0.84	0.84

A CNN architecture in which, for a specific task or dataset, the parameters of a pre-trained Conv2D layer within a neural network are fine-tuned. A confusion matrix is a tabular depiction of the performance of the classification model. To every non-diagonal value in the ideal confusion matrix must be zero. By utilizing the confusion matrix provided, one can examine the misclassified images. As an illustration, the majority of GaiYang images are erroneously categorized as CurriedFishCake, KaiThoon as GaengJued, GoongPao as GaiYang, and so forth, are presented in Figure 9.

In contrast to the preceding perplexity matrix, the current one exhibits a greater abundance of diagonal values. GaiYang, Fried Chicken, and Eggs Stewed are the only categories that still contain a number of misclassified images. Finally, we have effectively completed the task of classifying images of Thai cuisine. The efficacy of transfer learning has been demonstrated in this experiment, are presented in Figure 10.

## 5. CONCLUSIONS AND FUTURE WORK

The research utilizes a Convolutional Neural Network (CNN) with transfer learning to classify Thai cuisine images, achieving an accuracy rate of 84%. Monitoring the model's performance over epochs revealed stable training loss after 5 epochs, while validation accuracy fluctuated within a narrow range even after 8 epochs. Epoch 8 showed the lowest training loss and the highest training accuracy. We attributed subsequent marginal changes to overfitting, suggesting that training beyond ten epochs could result in significant overfitting. Following training, the model achieved a test accuracy of more than 80 and a test loss of 0.57, indicating satisfactory generalization performance. We propose various strategies to enhance the CNN model, such as data augmentation, image dimension adjustment, increasing the number of deep layers, epochs, and scaling parameters. However, caution is advised to prevent overfitting and maintain computational efficiency, while the current CNN model shows promising performance, there are opportunities for enhancement through careful optimization strategies and exploring larger datasets for improved accuracy and broader applicability.

Future research can focus on eliminating dataset disturbances and using larger datasets with more classes and images per class to improve accuracy and decrease loss rates. We could also explore the retention of model weights to develop applications such as mobile or web-based image classifiers, which could include additional features like calorie extraction from classified food items.

## REFERENCES

- [1] Xu, B., He, X., Qu, Z. (2021). Asian food image classification based on deep learning. *Journal of Computer and Communications*, 9(03): 10-28. <https://doi.org/10.4236/jcc.2021.93002>
- [2] Chaitanya, A., Jayashree, S., Priyamvada, C., (2023). Food image classification and data extraction using Convolutional Neural Network and web crawlers. *Procedia Computer Science*, 218: 143-152. <https://doi.org/10.1016/j.procs.2022.12.410>
- [3] Li, C., Zhen, T., Li, Z. (2022). Image classification of pests with residual neural network based on transfer learning. *Applied Sciences*, 12(9): 4356. <https://doi.org/10.3390/app12094356>
- [4] Cengil, E., Çinar, A. (2019). Multiple classification of flower images using transfer learning. In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), pp. 1-6. <https://doi.org/10.1109/IDPA.20198875953>
- [5] Algan, G., Ulusoy, I. (2021). Image classification with deep learning in the presence of noisy labels: A survey. *Knowledge-Based Systems*, 215: 106771. <https://doi.org/10.1016/j.knosys.2021.106771>
- [6] Lubura, J., Pezo, L., Sandu, M.A., Voronova, V., Donsi, F., Šic Žlabur, J., Ribić, B., Peter, A., Šurić, J., Brandić, I., Klõga, M., Ostojić, S., Pataro, G., Virsta, A., Oros, A.E., Micić, D., Đurović, S., De Feo, G., Procentese, A., Voća, N. (2022). Food recognition and food waste estimation using Convolutional Neural Network. *Electronics*, 11(22): 3746. <https://doi.org/10.3390/electronics11223746>
- [7] Kaur, R., Kumar, R., Gupta, M. (2023). Deep neural network for food image classification and nutrient identification: A systematic review. *Reviews in Endocrine and Metabolic Disorders*, 24(4): 633-653. <https://doi.org/10.1007/s11154-023-09795-4>
- [8] Lee, K.S. (2023). Multispectral food classification and caloric estimation using Convolutional Neural Networks. *Foods*, 12(17): 3212. <https://doi.org/10.3390/foods12173212>
- [9] Zhu, L., Spachos, P., Pensini, E., Plataniotis, K.N. (2021). Deep learning and machine vision for food processing: A survey. *Current Research in Food Science*, 4: 233-249. <https://doi.org/10.1016/j.crfs.2021.03.009>
- [10] Chopra, M., Purwar, A. (2022). Food image recognition by optimizing CNN with PSO and GA. In Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing, pp. 37-42. <https://doi.org/10.1145/3549206.3549214>
- [11] Pillai, S.M., Mohideen, S.K. (2020). Food calorie measurement and classification of food images. *International Journal of Pharmaceutical Research* (09752366), 12(4): 262. <https://doi.org/10.31838/ijpr/2020.12.04.262>
- [12] Farooqui, M., Rahman, A.U., Alorefan, R., Alqusser, M., Alzaid, L., Alnajim, S., Althobaiti, A., Ahmed, M.S. (2023). Food classification using deep learning: Presenting a new food segmentation dataset. *Mathematical Modelling of Engineering Problems*, 10(3): 1017-1024. <https://doi.org/10.18280/mmep.100336>
- [13] McAllister, P., Zheng, H., Bond, R., Moorhead, A. (2018). Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets. *Computers in Biology and Medicine*, 95: 217-233. <https://doi.org/10.1016/j.combiomed.2018.02.008>
- [14] Fakhrou, A., Kunhoth, J., Al Maadeed, S. (2021). Smartphone-based food recognition system using multiple deep CNN models. *Multimedia Tools and Applications*, 80(21): 33011-33032. <https://doi.org/10.1007/s11042-021-11329-6>
- [15] Zhang, Y., Deng, L., Zhu, H., Wang, W., Ren, Z., Zhou, Q., Gorriz, J.M., Wang, S. (2023). Deep learning in food category recognition. *Information Fusion*, 98: 101859. <https://doi.org/10.1016/j.inffus.2023.101859>
- [16] Kaur, R., Kumar, R., Gupta, M. (2023). Deep neural network for food image classification and nutrient identification: A systematic review. *Reviews in Endocrine and Metabolic Disorders*, 24(4): 633-653. <https://doi.org/10.1007/s11154-023-09795-4>
- [17] Rajayogi, J., Manjunath, G., Shobha, G. (2019). Indian Food image classification with transfer learning. In International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), Bengaluru, India, pp. 1-4. <https://doi.org/10.1109/CSITSS47250.2019.9031051>

- [18] Rafhan, A., Suhaili, B. (2021). Malaysian food recognition using AlexNet CNN and transfer learning. In 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, pp. 59-64. <https://doi.org/10.1109/ISCAIE51753.2021.9431833>
- [19] Rao, N.S., Soujanya, P., Praneeth, N.R., Kumar, Y.R. (2020). Food categories classification and nutritional value estimation using CNNs on Raspberry Pi 3B. *International Journal for Research in Applied Science & Engineering Technology*, 8(5): 5504. <https://doi.org/10.22214/ijraset.2020.5504>
- [20] Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., Miao, Y. (2021). Review of image classification algorithms based on Convolutional Neural Networks. *Remote Sensing*, 13(22): 4712. <https://doi.org/10.3390/rs13224712>
- [21] Metwalli, A.S., Shen, W., Wu, C.Q. (2020). Food image recognition based on densely connected Convolutional Neural Networks. In 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), pp. 027-032. <https://doi.org/10.1109/ICAIIIC48513.2020.9065281>
- [22] Min, W., Jiang, S., Liu, L., Rui, Y., Jain, R. (2019). A survey on food computing. *ACM Computing Surveys (CSUR)*, 52(5): 1-36. <https://doi.org/10.1145/3329168>
- [23] Knez, S., Šajn, L. (2020). Food object recognition using a mobile device: Evaluation of currently implemented systems. *Trends in Food Science & Technology*, 99: 460-471. <https://doi.org/10.1016/j.tifs.2020.03.017>
- [24] Sheldon, M., Mukul, A. (2021). A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. In *International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications*, pp. 96-99. <https://doi.org/10.1109/CENTCON52345.2021.9687944>
- [25] Subarna, S. (2020). Analysis of artificial intelligence based image classification techniques. *Journal of Innovative Image Processing*, 2(1): 44-54. <https://doi.org/10.36548/jiip.2020.1.005>
- [26] VijayaKumari, G., Priyanka, V., Vishwanath, P. (2022). Food classification using transfer learning technique. *Global Transitions Proceedings*, 3(1): 225-229. <https://doi.org/10.1016/j.gltp.2022.03.027>
- [27] Gao, X., Xiao, Z., Deng, Z. (2024). High accuracy food image classification via vision transformer with data augmentation and feature augmentation. *Journal of Food Engineering*, 365: 111833. <https://doi.org/10.1016/j.jfoodeng.2023.111833>
- [28] Liu, Y., Pu, H., Sun, D. W. (2021). Efficient extraction of deep image features using Convolutional Neural Network (CNN) for applications in detecting and analysing complex food matrices. *Trends in Food Science & Technology*, 113: 193-204. <https://doi.org/10.1016/j.tifs.2021.04.042>
- [29] Wang, Y., Wu, J., Deng, H., Zeng, X. (2021). Food image recognition and food safety detection method based on deep learning. *Computational Intelligence and Neuroscience*, 2021: 1268453. <https://doi.org/10.1155/2021/1268453>
- [30] Zhu, Z., Dai, Y. (2023). A new CNN-based single-ingredient classification model and its application in food image segmentation. *Journal of Imaging*, 9(10): 205. <https://doi.org/10.3390/jimaging9100205>
- [31] Chakkrit, T., Surachet, K. (2018). Nu-ResNet: deep residual networks for Thai food image recognition. *Journal of Telecommunication, Electronic and Computer Engineering*, 10(1-4): 29-33.

## NOMENCLATURE

CNN	Convolutional Neural Network
CBAM	Convolutional Block Attention Module
VGG16	Visual Geometry Group 16
VGG19	Visual Geometry Group 19
SVM	Support Vector Machines
RBF	Radial Basis Function
KNN	K-Nearest Neighbors
GA	Genetic Algorithm
PSO	Particle Swarm Optimization
TL	Leveraging Transfer Learning
RMSProp	Root Mean Square Propagation
Adam	Adaptive Moment Estimation