A Novel Dish Recognition Method Using Deep Learning

J. Krishna, S. Tejaswini, N. Viswa Sai Reddy, S. Susmitha, S. Sohail, and G. Prasanna

Abstract Understanding digital media dishes is a fascinating issue, but it also involves a lot of challenge. The dish's complicated ingredient list presents a hurdle. Due to the growth of deep learning, a number of efficient tools can partially resolve the issue. The job of dish recognition in this work is thought about. Based on the EfficientNet architecture and transfer learning, a unique dish recognition algorithm is proposed. First, add a number of significant layers to the EfficientNet-B0. Next, employ transfer learning to retrain the model using the best parameters that were learned during the first pre-training on ImageNet on the UEH-VDR dataset, a fresh batch of dish pictures. The UEH-VDR dataset includes pictures of Vietnamese food gathered from a variety of sources. According to experimental findings, the suggested approach can identify a dish with an accuracy of 92.33%. Additionally, it performs better than models built on well-known Space Invariant Artificial Neural Networks (SIANN) like VGG and residual neural network. On the basis of the training data, a mobile application is also created to assist tourists who wish to learn about Vietnamese cuisine.

Keywords Food acknowledgment identification · Deep learning · Transfer learning · Food tourism · Convolutional neural networks · EfficientNet

1 Introduction

A popular trend on the global tourist map is culinary tourism [1], sometimes known as food tourism. Dishes frequently reflect the traits of the locals and each culture. The geography, culture, religion, and climate [2] of any culture are just a few examples of the many distinct aspects that influence a culinary culture. For instance, olive oil

J. Krishna
Department of AI&ML, AITS, Rajampet, India

S. Tejaswini (✉) · N. Viswa Sai Reddy · S. Susmitha · S. Sohail · G. Prasanna
Department of Computer Science and Engineering, AITS, Rajampet, India
e-mail: falsekrishna.j.jk@gmail.com

A. Kumar and S. Mozar (eds.), *Proceedings of the 6th International Conference on Communications and Cyber Physical Engineering*, Lecture Notes in Electrical Engineering 1096, https://doi.org/10.1007/978-981-99-7137-4_35

and herbs are frequently used in Mediterranean cooking [3]. Steak and cutlets are staples of Western European cuisine. Additionally, cheese and wine are well-known elements in sauces. Rice is the staple cuisine in Eastern Asia and Indochina [4], while fish, shrimp, and soybean are typically used to make sauces.

Several more tourists are willing to pay to sample distinctive delicacies from the surrounding areas [5]. Therefore, there are numerous chances for economic growth with food tourism [6]. In order to enhance the culinary tourist experience, concentrating on creating a method to identify food. Transfer learning and Convolutional neural networks methods [7, 8] are used in the system's design, that is based on deep learning.

With nine well-known traditional dish classifiers [9], the structure concentrates on Vietnamese culinary culture: Banh-Chung, Bun, Banh-Mi, Com-Tam, Banh-Tet, Banh-Trang, Pho, Banh-Xeo and Goi-Cuon. The system is easy configurable to accommodate additional classifiers [10] and delicacies from different nations.

The following are the primary areas where this paper makes major contributions. In order to recognize food [11], a novel CNN is first created using EfficientNet [7] and the transfer learning method. Second, giving researchers and scientists access to dishes dataset [12], train the model, and acquired parameters so that can reuse and expand it with more dish classifiers. In order to tackle the problem of dish identification, a mobile application is finally built. It also gave some helpful information about foods.

The residual parts of the hypothesis are arranged as follows. A thorough analysis of similar techniques was presented in Sect. 2. Next Section presents the materials and the recommended procedure for classifying food into nine categories. The creation of a software platform to enhance the experience of tourists is presented in Sect. 4 together with the trial findings. Section 5 brings the process to a close.

2 Review of Similar Methodology

Food or dish characterization is a fascinating issue with numerous practical applications. It can also be used to examine the calories in foods to govern the best meal in terms of calories for persons. There are frequent works that address the issue. A Chinese food recognition [13] and value assessment system constructed on multi-design support vector machines (SVM) and Adaboost has been proposed [14] by Chen et al. With the use of SVM, Kawano and Yanai [15] created mechanism for sensing meals on the go. The device is able to measure calories and nourishment as well as identify dishes. Another food image recognition system was created by [16] Yanai and Kawano utilising a pre-trained convolutional neural and tuned deep convolutional neural system.

Yadav and Chand [17] created a different automated method for classifying food images using the VGG network. Martinel et al. [4] presented using broad range hysteresis networks to identify food based on residual learning. A new CNN acknowledged as a personalised classification approach [18] for nutrition recognition was

proposed by Horiguchi et al. [19] Zahisham created a different way for identifying food using the ResNet-50 model. The above-mentioned techniques have a few major downsides. The approaches SVM or Artificial neural strategies have constraints effectiveness attributable to the dish identification obstacle requires various learning tools [20] and sources. Other CNN-related techniques are more efficient, but need a large amount of training data. CNN models have this quality. Additionally, some techniques, like VGG, that were trained on intricate CNN models consumed a lot of system resources. Due to their high efficacy, CNN models are a major focus of this paper. Additionally, observe CNN model performance that is appropriate for mobile applications. As a result, use the Model EfficientNet and transfer learning method to create a dish identification system.

3 Components and Techniques

A. Components

The collection of 7848 photos of food known as UEHVDR was compiled from various online sources. Each image is in RGB colour mode and comes in different sizes. The picture database is divided into three sections: a practise set of 6273 pictures (about 80% data at irregular intervals), an authentication set of 780 pictures (about 10% data at irregular intervals), and a test database 795 photos (about 10% data at irregular intervals).

Vietnamese cuisine has a wide variety of foods, yet data may be categorised into a number of fundamental groups. For illustration. Numerous variations of pho, including pho bo, pho ga, pho chay, and others, are included in the pho group. Just to make things easier, refer to them all as Pho. International visitors from all over the world are also familiar with this name. Banh-Mi, another well-known Vietnamese meal, is the same in this instance. Nine well-known and traditional groups—Banh-Chung, Bun, Banh-Mi, Com-Tam, Banh-Tet, Banh-Trang, Pho, Banh- Xeo and Goi-Cuon—were employed in the study.

It should remain highlighted the titles of the Dishes are only available in Vietnamese because interpretations might affect their popularity or meaning. For tourists, learning about a culture through the names of regional foods is also interesting. Materials from the UEH-VDR collection, in component are shown in Fig. 1. Figure 2 illustrates the number of photos in individual class for the training dataset, a validating test, and a training and testing sets.

B. Techniques

(1) **EfficientNet-B0**: An optimized collection of basis systems was created using neural network—based using convolution layers by uniformly scaling width, height, and resolution over whole parameters. Figure 3 displays EfficientNet-B0, a member of the EfficientNet [7] family of designs. Layers in the EfficientNet-B0 configuration include:

Fig. 1 A portion of the dataset of Vietnamese cuisine images

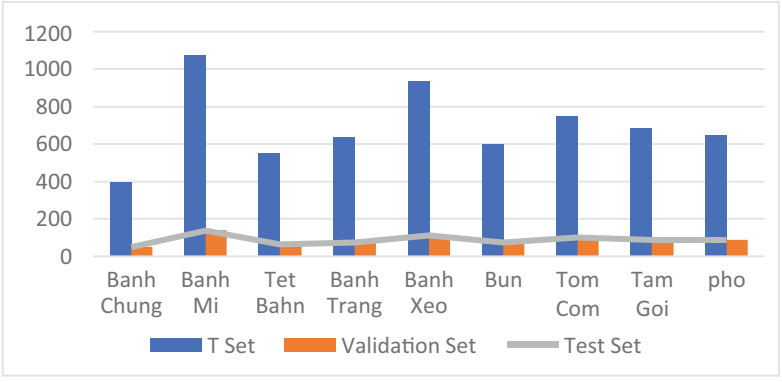

Fig. 2 Number of photos in each class's training, testing, and evaluation datasets

A 2-D convolution layer is called the conv layer (Conv2D). It creates a kernel to generate the output, integrating it with data from the previous level. **Depthwise Conv layer** Each input channel will receive a single convolutional filter from the depth-wise convolution layer.

By analysing the dimensions average and standard deviation, the **BatchNormalization** layer normalises input components. It then determines the normalized [21] activation function by:

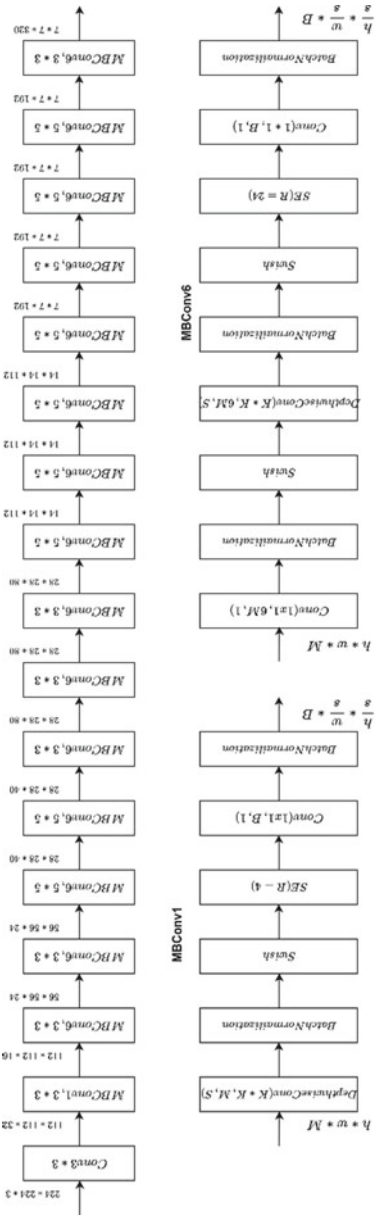

Fig. 3 EfficientNet-B0 structure, where K × K is the dimensions of the filter, S is the stride, B is the feature maps, h × w is the size of the image

$$x_i = \frac{x_i - \mu B}{\sqrt{\sigma_B^2 + g}}, \tag{1}$$

where xi is the quantity of information items, μB and σB^2 are dimensions average and standard deviation, and 1 is the constant of quantitative consistency.

As for the input, following is a flat and non-monotonic activation function that the Swish layer will use:

$$swish(x) = \frac{x}{1 + \exp(-x)} \tag{2}$$

For each channel, the SE layer will apply distinct weights rather than the same.

In addition to functioning well and persistently on ImageNet, EfficientNet-B0 may also be used satisfactorily to other datasets.

(2) EfficientNet-B0 transfer learning

Any machine-learning model can apply the transfer learning technique, but deep learning is where it first gained popularity. Neural networks using convolution layers is viewed as having training in particular database sets in order to retrieve characteristics. The prototype is able to forecast outcomes based on learnt attributes. A lot of data must be used to train the model for CNNs in order to increase their accuracy. There are certain challenges to gathering a lot of data in practise. Additionally, if anything in the obtained information is lacking, the data is still not trustworthy sufficient to be implemented with CNN models. The transfer learning method is the answer to these problems. An examination of the transfer learning approach is shown in Fig. 4.

Transfer learning allows a model to keep its ideal parameters after being tested on a well-known dataset like ImageNet. The following learning challenge uses the previously taught characteristics. As a result, the model's overall accuracy will increase.

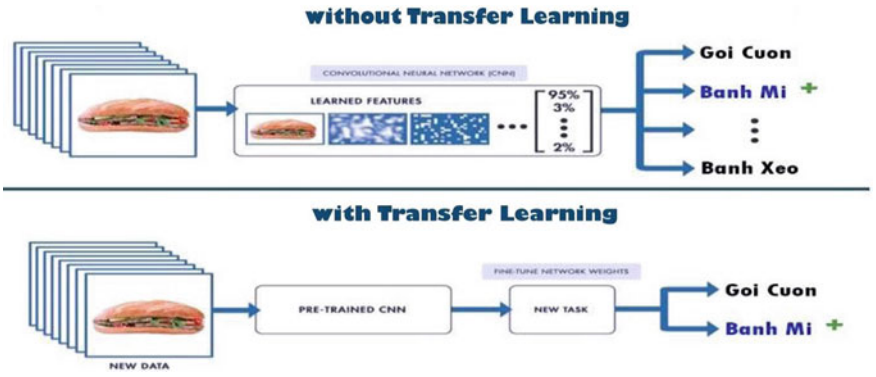

Fig. 4 A CNN model transfer learning approach

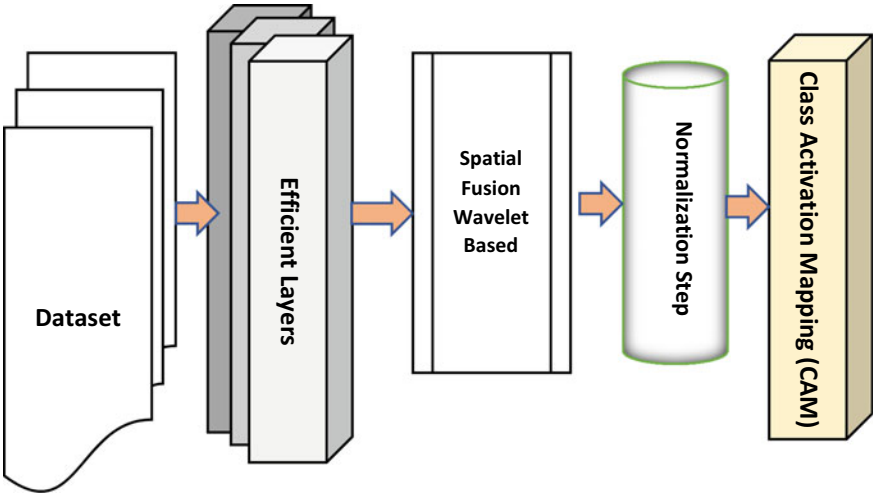

Fig. 5 Architecture of EfficientNet-B0

Transfer learning will be used in this project for EfficientNet-B0, which was trained on ImageNet. Figure 5 shows the suggested model's architecture utilising EfficientNet-B0 and transfer learning. The design of the suggested model is essentially the same as that of EfficientNet-B0.

In other words, it gets information from just about every neuron in the layer before it. By monitoring the dimensions average and standard deviation, the Batch-Normalization layer normalises input components before estimating the standardized activation.

C. Error metrics

The following metrics will be employed to gauge how to see how the dish categorization is doing and how it stacks up against additional links with a similar focus:

$$accuracy = \frac{Tr_p + Tr_n}{Fa_n + Fa_p + Tr_p + Tr_n}, \tag{3}$$

$$precision = \frac{Tr_p}{Tr_p + Fa_p}, \tag{4}$$

$$recall = \frac{Tr_p}{Tr_p + Fa_n}, \tag{5}$$

$$F1\ score = \frac{Tr_p}{2Tr_p + Fa_p + Fa_n}, \tag{6}$$

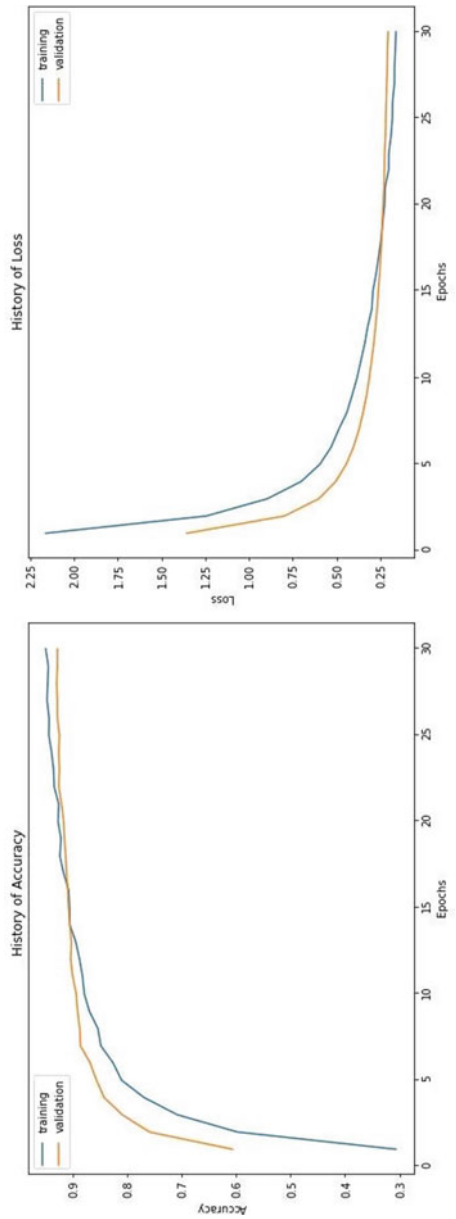

Fig. 6 Process of accuracy and loss training

where, in turn, Trp , Trn, Fap, Fan (stands for) True Positive, True Negative, False positive and False Negative. Additionally, employ region under the Relative Operating Characteristic curve (AUC) measures. The ROC curve's total area, together with the vertical and horizontal axes, are used to evaluate AUC.

For TPR and FPR, the ROC curve is marked at various minimal values:

$$TPR = \frac{Tr_p}{Tr_p + Fa_n}, \tag{7}$$

$$FPR = \frac{Fa_p}{Fa_p + Tr_n} \tag{8}$$

Additionally considered is Cohen's Kappa (Kappa score). The Kappa rating is determined by:

$$k = \frac{p_0 - p_e}{1 - p_e}, \tag{9}$$

where pe stands for anticipated agreement and p0 refers to the empirical probability.

4 Results of Experiment and Discussions

A. Model training and configuration

On the Google CODLAB platform, develop the conceptual system using the training data, validation datatset and test set in the same ratio of 8:1:1. 30 epochs are the predetermined number. The categorical cross-entropy is what is employed for the loss function:

$$Loss = -\sum_{c=1}^{M} y_{0,c} \ln(p_0, c), \tag{10}$$

where c is a class tag, po, c denotes the frequency that observation o relates to class c, yi, c signifies a binary pointer (0 or 1), which yields 1 if class tag c accurately distinguishes observation o, and M denotes the number of classes.

Through using Keras framework's default configuration and the optimum Adam technique [23], model is trained.

All of the representations analyzed in the study VGG16, ResNet50, and EfficientNetB0 are used using the above- mentioned settings. In Fig. 6, the accuracy and loss training graphs are displayed.

	Banh_Chung	Banh_Mi	Banh_Tet	Banh_Trang	Bun	Pho	Goi_Cuon	Com_Tam	Banh_xeo
Banh_xeo	0.02	0.00	0.06	0.00	0.04	0.00	0.06	0.02	**0.81**
Com_Tam	0.00	0.01	0.01	0.01	0.01	0.01	0.01	**0.96**	0.00
Goi_Cuon	0.00	0.01	0.01	0.00	0.00	0.01	**0.86**	0.01	0.09
Pho	0.00	0.00	0.05	0.03	0.01	**0.85**	0.01	0.05	0.00
Bun	0.01	0.00	0.02	0.00	**0.92**	0.03	0.01	0.01	0.00
Banh_Trang	0.01	0.00	0.04	**0.92**	0.01	0.01	0.00	0.00	0.00
Banh_Tet	0.00	0.00	**0.98**	0.02	0.00	0.00	0.00	0.00	0.00
Banh_Mi	0.00	**0.97**	0.00	0.01	0.01	0.00	0.00	0.01	0.00
Banh_Chung	**0.96**	0.00	0.00	0.02	0.00	0.00	0.00	0.01	0.00

Fig. 7 The confusion matrix

B. Outcomes and hypotheses

The trained system is exported and use the model on the test dataset to gauge how well the suggested strategy works. Figure 7 displays the confusion matrix using the data source.

Then, assess the mistake metrics built on the confusion matrix. The suggested method's accuracy, recall, precision, AUC, Kappa scores and F1 score are shown in Table 1 together with the values for VGG-16, EfficientNet-B0 and Resnet-50. The EfficientNetB0 based on transfer learning produced the greatest results across all statistics, as can be shown.

C. Mobile application development

The training data is used to create an application called Dish Recognition in order to apply the findings to the recognition of Vietnamese dishes (VDR). Based on

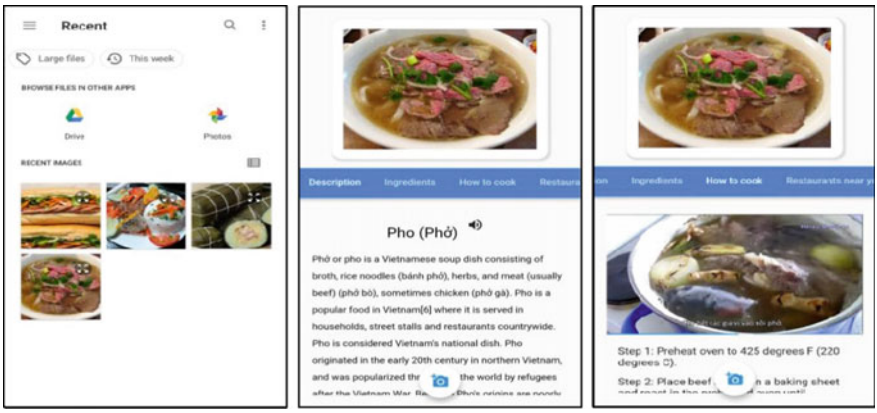

Fig. 8 The API UEH-VDR application: showing the recipe (right), identifying the food and showing its properties (middle), and synchronising an input photo (left)

Table 1 Efficiency evaluation of the CNN model

	Accuracy	Precision	Recall	AUC	F1-score	Kappa
ResNet [22]	81.51	89.86	73.58	97.28	80.91	78.96
Efficient Net-B0 [7]	82.98	88.59	78.11	97.75	83.02	80.57
VGG-16 [17]	76.23	82.88	69.43	95.76	75.56	72.92
Efficient Net-B0 with transfer learning	**92.33**	**94.13**	**90.82**	**99.08**	**92.45**	**91.28**

The bold indicates the proposed method values which are superior to previous values in Table 1

the Flutter framework, the application was created. Because of this, the VDR app may run on many operating systems, as for Android and iOS, using just an individual source code. Visit the website:https://play.google.com/store/apps/details?id=com.ueh.vdr to access the Google Play Store and download the Android app. The application will eventually be released on the Apple App Store.

The application's user interface is depicted in Fig. 8. Users of the application can upload images from the local computer memory or take new ones with the integrated camera. The software will identify the dish categorizer in any of the nine available classifiers and display some helpful information, like the dish's name, pronunciation in Vietnamese, an explanation of its ingredients, preparation instructions, and locations where it may be eaten.

5 Conclusion

In the essay, it has put forth a brand-new approach to dish recognition that combines the EfficientNet model with transfer learning. A number of significant layers, including BatchNormalization, GlobalAveragePooling2D, Dropout, and Dense are added to the EfficientNetB0. Then, transfer learning was applied to make advantage of the knowledge gained throughout the ImageNet pretraining procedure. The suggested strategy delivered excellent performance and accuracy for dish recognition. Additionally, a mobile application is created to use the training dataset to support food tourism. While the suggested approach operates successfully and consistently, some shortcomings must be addressed in the forthcoming, including expanding the quantity of categories for food and integrating with dish calorie evaluation.

References

1. Okumus B (2021) Food tourism research: a perspective article. Tourism Rev 76(1):38–42
2. Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46

3. Zhou L, Zhang C, Liu F, Qiu Z, He Y (2019) Application of deep learning in food: a review. Comprehen Rev Food Sci Food Safety 18:1793–1811
4. Martinel N, Foresti GL, Micheloni C (2018) Wide-slice residual networks for food recognition. In: Proceeding IEEE winter conference application computer vision (WACV), pp 567–576
5. Park SJ, Palvanov A, Lee CH, Jeong N, Cho YI, Lee HJ (2019) The development of food image detection and recognition model of Korean food for mobile dietary management. Nutri Res Pract 13(6):521–528
6. Liberato P, Mendes T, Liberato D (2020) Culinary tourism and food trends. In: Advances in tourism,technology and smart systems, Springer.https://doi.org/10.1007/978-981-15-2024-2_45
7. Tan M, Le Q (2019) EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proceeding 36th international conference machine learning, Long Beach, CA, USA, pp 6105–6114
8. Zhang Z, Sabuncu MR (2018) Generalized cross entropy loss for training deep neural networks with noisy labels. In: Proceeding 32nd international conference neural information process system, Montreal, QC, Canada, pp 1–11
9. Farinella GM, Moltisanti M, Battiato S (2014) Classifying food images represented as Bag of Textons. In: IEEE International conference in image processing (ICIP), Paris, pp 5212–5216. https://doi.org/10.1109/ICIP.2014.7026055, 2014
10. Taha AA, Hanbury A (2015) 'Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool.' BMC Med. Imag. 15(1):1–29
11. Zhou B, Lapedriza A, Xiao J, Torralba A, Oliva A (2014) Learning deep features for scene recognition using places database. In: Proceedings of the 27th ICNIPS, vol 1, pp 487–495, ACM
12. Krishna J, Rupesh Kumar Reddy M, Rudra Kumar M (2017) Efficient high utility top-k frequent pattern mining from high dimensional datasets. IJSRCSEIT 2(4):625–631
13. Rahmani GA (2017) Efficient combination of texture and color features in a new spectral clustering method for PolSAR image segmentation. Nat Acad Sci Lett 40:117–120. https://doi.org/10.1007/s40009-016-0513-6
14. Chen MY, Yang YH, Ho CH, Wang SH, Liu SM, Chang E, Yeh CH, Ouhyoung M (2012) Automatic Chinese food identification and quantity estimation. In: Procceding SIGGRAPH Asia technical briefs (SA), pp 1–4
15. Kawano Y, Yanai K (2013) Real-time mobile food recognition system. In: Proceeding IEEE conference computer vision pattern recognition workshops, pp 1–7
16. Yanai K, Kawano Y (2015) Food image recognition using deep convolutional network with pre-training and fine-tuning. In: Proceeding IEEE international conference multimedia explosion workshops (ICMEW), 2015, pp 1–6
17. Yadav S, Chand S (2021) Automated food image classification using deep learning approach. In: Proceeding 7th international conference advance computer communication system (ICACCS), pp 542–545
18. Bolanos M, Radeva P (2016) Simultaneous food localization and recognition. In: Proceeding 23rd International Conference Pattern Recognition (ICPR), pp 3140–3145
19. Horiguchi S, Amano S, Ogawa M, Aizawa K (2018) Personalized classifier for food image recognition. IEEE Trans Multimedia 20(10):2836–2848
20. Wang M, Wan Y, Ye Z, Lai X (2017) Remote Sensing image classification based on the optimal support vector machine and modified binary coded ant colony optimization algorithm. Inform Sci 402:50–68
21. Kingma DP, Ba LJ (2015) Adam: a method for stochastic optimization. In: Proceeding international conference learn represent (ICLR), San Diego, CA, USA, pp 1–15
22. Csurka G, Larlus D, Perronnin F (2013) What is a good evaluation measure for semantic segmentation? In: Proceeding British machine vision conference, p 5244
23. Zahisham Z, Lee CP, Lim KM (2020) Food recognition with ResNet50. In: Proceeding IEEE 2nd International Conference Artificial Intelligent Engineering Technology (IICAIET), pp 1–5