PithaNet: A transfer learning-based approach for traditional pitha classification

Shahriar Shakil, Atik Asif Khan Akash, Nusrat Nabi, Mahmudul Hasan, Aminul Haque
Department of Computer Science and Engineering, Faculty of Science and Information Technology, Daffodil International University, Dhaka, Bangladesh

Article Info

Article history:

Received Oct 5, 2022
Revised Jan 26, 2023
Accepted Feb 3, 2023

Keywords:

Bengali pitha
Convolutional neural network
Deep learning
Image classification
Transfer learning

ABSTRACT

Pitha, pithe, or peetha are all Bangla words referring to a native and traditional food of Bangladesh as well as some areas of India, especially the parts of India where Bangla is the primary language. Numerous types of pithas exist in the culture and heritage of the Bengali and Bangladeshi people. Pithas are traditionally prepared and offered on important occasions in Bangladesh, such as welcoming a bride grooms, or bride, entertaining guests, or planning a special gathering of family, relatives, or friends. The traditional pitha celebration and pitha culture are no longer widely practiced in modern civilization. Consequently, the younger generation is unfamiliar with our traditional pitha culture. In this study, an effective pitha image classification system is introduced. convolutional neural network (CNN) pre-trained models EfficientNetB6, ResNet50, and VGG16 are used to classify the images of pitha. The dataset of traditional popular pithas is collected from different parts of Bangladesh. In this experiment, EfficientNetB6 and ResNet50 show nearly 90% accuracy. The best classification result was obtained using VGG16 with 92% accuracy. The main motive of this study is to revive the Bengali pitha tradition among young people and people worldwide, which will encourage many other researchers to pursue research in this domain.

Corresponding Author:

Shahriar Shakil
Department of Computer Science and Engineering, Daffodil International University
Dhaka, Bangladesh
Email: shahriar15-8558@diu.edu.bd

1. INTRODUCTION

Pitha is a type of food that originated in the Indian subcontinent and is equally popular in Bangladesh and India. Pithas are similar to pancakes, dumplings, or fritters. The main ingredient of a pitha is a batter composed of rice flour or wheat flour, which is further moulded and could be filled with either sweet or savory fillings. Pithas, or traditional handcrafted cakes, are a wintertime treat that Bengalis are renowned for enjoying. In the winter, date juice and molasses made from sugarcane and dates are available, which are essential ingredients for pithas. Hence, this dish is immensely popular during winter in Bangladesh. When farmers harvest paddy from the field in the late autumn, Bangladesh's pitha season officially begins. In addition to the winter season, these pithas can be found on festive occasions such as weddings, Eid festivities, and puja celebrations. Many well-known pithas, including patishapta pitha, bhapa pitha, chitoi pitha, til er pitha, kolar pitha, tel er pitha, nakshi pitha, to name but a few, are seasonal during the winter. Food picture classifications, such as fast food, vegetables, fruits, cake, and so forth, have attracted a lot of interest in the field of research. Nevertheless, the classification of food images is still the most recent area of study. Therefore, the classification of traditional pithas has been studied in this research.

This study focuses on traditional food classifications. Pitha is a traditional food of the Indian subcontinent. Today, the tradition of pithas is all but extinct due to modern civilization. Through this study, the younger generations can be familiarised with the names of pithas, particularly, who are less informed about the varieties of pithas, and how it relates to the culture. People, specifically those residing in cities, do not have a lot of spare time to make pithas. Due to this, commercially produced cakes, pastries, and other foods are increasingly taking the place of traditionally handcrafted pithas, especially in urban areas.

The consumption of junk food is putting our young generation at serious health risk, which can be reduced if a healthy alternative pitha is revitalized. In this study, pre-trained models are used to classify the pithas. Over the past few years, the field of artificial neural networks (ANNs) s has advanced and developed rapidly. It is remarkable operational precision and accuracy shows promise while performing image processing tasks. The performance of neural networks is further improved by several factors, including unsupervised learning and rich feature extraction. The convolutional neural network (CNN) specialises in large image processing. The best of CNN models in recent years are AlexNet [1], VGG [2], and ResNet [3]. In image processing tasks such as object detection, semantic segmentation, and image recognition and classification, these models exhibit outstanding performance. Therefore, EfficientNetB6, ResNet50, and VGG16 models have been used for pitha classification in this research. ResNet is faster to improve and can achieve high accuracy as the depth of picture classifications increases [4]. The Visual Geometry Group at the University of Oxford and Google DeepMind collaboratively created VGGNet, a CNN model [2]. The architecture of VGGNet is an expanded version of AlexNet. A wide range of image classification tasks can be completed with EfficientNet. Performance will increase with the size of the EfficientNet model [5]. We use 10,000 pitha images of 8 classes in this work. VGG16 pre-trained model shows better performance. The contributions of this studies are as follows:

- A dataset of traditional pithas of Bangladesh has been constructed for this study.
- A comparative study has been shown among the popular transfer learning models such as EfficientNet, ResNet50, AlexNet, SqueezeNet and VGG16 for the classification, which to the best of our knowledge, no research studied earlier.
- It has been shown that the transfer learning models such as VGG16, ResNet50, and EfficientNetB6 can be fine-tuned for achieving better results.

2. LITERATURE REVIEW

Mahajan and Chaudhary [6] used a deeper CNN to classify hundreds of high-resolution photos into eight distinct classifications. A pre-trained representational deep neural network was used to extract picture features. The ResNet model was fitted using 8 community order dataset, which contains around 2,698 images and found accuracy of 93.57% on layer 18th layer. The authors designed a new TCNN(ResNet50) for defect identification that has 51 convolutional layers of depth. ResNet50, TCNN(ResNet50) employ transfer learning along with a feature extractor that is trained on ImageNet for defect diagnosis. Three datasets, such as the bearing damage dataset from KAT Datacenter, the motor bearing dataset from Case Western Reserve University (CWRU), and the self-priming centrifugal pump dataset, have been used to evaluate TCNN(ResNet50) [7].

In another study, researchers explored a quick method for helping psoriasis sufferers identify 15 foods that are known to cause inflammation. Using methods for image augmentation, a collection of 41,250 photographs of diverse inflammatory foods was produced from 10,000 original photos. AlexNet, VGG16, and EfficientNetB0 all used transfer learning, with EfficientNetB0 achieving the highest accuracy of 98.63% on the test set of 5,250 images. The accuracy rates for AlexNet and VGG16 are 87.22% and 93.79%, respectively [8]. The authors suggested a fruit and vegetable classification system that effectively draws object regions using image saliency and extracts image attributes using the CNN model. Vegetables and fruits classification images trained using a VGG architecture with tranfer learning approach. Meanwhile, they created a library of photographs of fruits and vegetables that spans 26 categories and includes the majority of real-world varieties. An excellent accuracy rate of 95.6% is achieved by the classification system, according to experiments done on an internal database [9].

Kumari *et al.* [10] focused on this study to develop a recognition model that classifies various food products into the proper categories by employing transfer learning techniques. The built model accurately categorized 101 different food types using the transfer learning method EfficientNetB0 with an accuracy of 80%. Their model outperformed other cutting-edge models with the highest level of accuracy. In another study, a prediction method was used for categorizing photos of Thai fast food is presented. The deep learning approach used by the model to create the predictive Thai fast food model was trained using real-world photos (GoogleNet dataset). The thai fast food (TFF) dataset contains 3,960 samples and achieved an accuracy of 88.33%, according to a separate test set's classification average [11].

In order to obtain information on Indonesian food for tourists, the authors use the CNN method in this study, which has proven to be quite fast and reliable in the process of classifying a complex and detailed item. Using the CNN method, the classification process may be carried out accurately, and data regarding food in the form of names or components can also be correctly gathered. Given that the categorization accuracy reached 70%, it is envisaged that this method will be implemented in the mobile-based system and provide a simple way for users to learn about Indonesian cuisine [12].

In this research, deep learning-based automatic food classification methods are described. For classifying food images, squeezenet and VGG16 CNNs are employed. SqueezeNet can reach quite a respectable accuracy of 77.20% even with fewer parameters. The accuracy of the projected VGG16 has significantly improved, reaching 85.07%, thanks to the deeper network [13]. For effective categorization of food photos, support vector machine (SVM) classifier, feature concatenation and deep feature extraction are employed in this research. The proposed model is performed using three openly accessible datasets FOOD-101, FOOD-5K and FOOD-11 and when evaluating performance, the accuracy metric is taken into consideration. According to the testing results, the accuracy for the FOOD-5K dataset is 99.00%, and for the FOOD-11 and FOOD-101 datasets, it is 88.08% and 62.44%, respectively [14].

Singla *et al.* [15] inn this study, the researchers tested a deep CNN-based GoogleLeNet model for classification and recognition of food and non-food objects. The photos used in the trials were gathered from two of their own image databases as well as from available image datasets, social sites, and imaging tools like smartphones and wearable cameras. The classification of food and non-food exhibits a high level accuracy of 99.2%, and the recognition of food categories exhibits an accuracy of 83.6%. In another study, the purpose of this study was to create a pre-trained structure for food recognition. Three different methods were applied to accomplish this goal, and their outcomes were evaluated to those of well-known pre-trained models AlexNet and CaffeNet. To apply these pre-trained models to their issue, the transfer learning technique was used. Test findings demonstrate that, as predicted, pre-trained models outperform suggested models in terms of output. The maximum improvement in classification performance achieved by the Adam technique was 32.85%; the maximum improvement achieved by Nesterov was 14.77% [16].

In throughout this paper, 1,676 datasets with 20% testing data and 80% training data that contain pictures of Indonesian traditional cakes will be subjected to the CNN Algorithm approach. Preprocessing, operational datasets, visualization datasets, modeling methodologies, performance evaluations, and errors analysis are all used in the stages, which led to the conclusion that performance evaluation has reached a level of 65.00% [17]. In order to classify images of Punakawan puppets, this work used a Gaussian filter as a preprocessing technique and a VGG16 learning architecture for classification. The study discovered that the maximum accuracy, 98.75%, was achieved while utilizing contrast limited the adaptive histogram equalization (CLAHE) + red, green, and blue (RGB) + Gaussian filter and thresholding images [18].

CNN breakthroughs in recent years suggested recognizing 12 different forms of illnesses affecting rice plants. Additionally, the performance of 8 various cutting-edge convolution neural network models has been assessed with a focus on diagnosing diseases of rice plants. The validation and testing accuracy of the proposed model are 96.5% and 95.3%, respectively, and properly diagnoses illnesses in rice plants [19]. This study proposes an effective CNNs-based fish categorization technique. Three types of splitting 80%-20%, 75%-25%, and 70% were tested on a novel dataset of indigenous fish species found in Bangladesh. The proposed model improved CNN's classification capacity with the highest accuracy of 98.46% [20]. The Kaggle-180-birds dataset was classified in this study using three different classifiers: the deep learning method (ResNet50), the classical machine learning (ML) algorithms (SVM, and decision tree (DT)), and the transfer learning-based deep learning algorithm (ResNet50-pretrained). The outcomes showed that the transfer learning classifier had the best classification effect, with a 98% to 100% accuracy rate [21]. Thfese various research methodologies provide us an assistance because our study is extremely unique and there hasn't been any significant research on this particular pitha classification topic.

Table 1 shows that most of the work is done on food image classification like fast food, cake, and other food using CNN models or pre-trained models. However, this work is done on traditional food which is pitha classification using transfer learning like EfficientNetB6, ResNet50, and VGG16. Among them, VGG16 showed better accuracy. There has been no significant work regarding the traditional pitha classification.

Table 1. Exclusive review of food image classification of recent work

Author’s Name	Used model	Accuracy	Year
Hridoy *et al.* [8]	AlexNet, VGG16, and EfficientNetB0	EfficientNetB0=98.63%, AlexNet=87.22%, VGG16=93.79%	2021
Kumari *et al.* [10]	EfficientnetB0	80%	2022
Yadav *et al.* [13]	SqueezeNet and VGG16	SqueezeNet=77.2%, VGG16=85.07%	2021
Kurnia *et al.* [17]	CNN	65.00%	2021

3. METHOD

In this section, we discuss about datasets, data pre-processing, model building, statistical analysis, and general architecture of our proposed model. The workflow diagram of this study is shown in Figure 1. Data collection and pre-processing is the first step of this research work. Here, we preprocess the data using augmentation, and image resize techniques. Afterward EfficientNetB6, ResNet50 and VGG16 are used for model building. Finally, we test all these models with real data and make predictions.

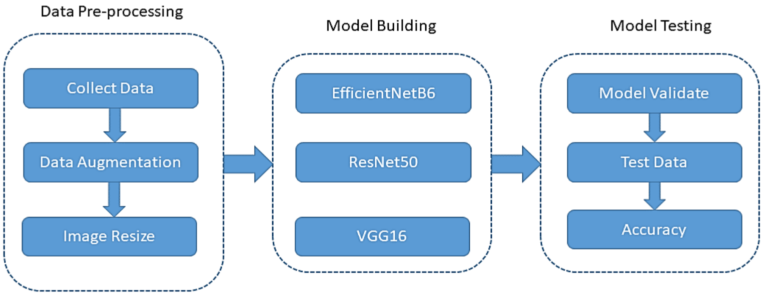


Figure 1. Workflow diagram of the study

3.1. Dataset

Samples of different types of traditional pitha play an important role in selecting features to differentiate them. For this work we gathered 5,240 primary images from different districts of Bangladesh of eight distinct types of traditional pithas are shown in Figure 2. All images are captuared by smartphone oneplus 6 with sensor size of 16 MP. Figures 2(a) bhapa pitha, 2(b) chitoi pitha, 2(c) jamai pitha, 2(d) nakshi pitha, 2(e) naru, 2(f) patishapta pitha, 2(g) puli pitha, 2(h) teler pitha.

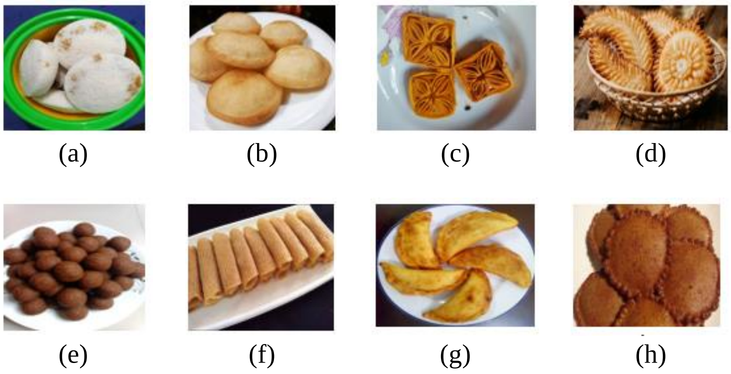


Figure 2. Sample from the pitha dataset (a) bhapa pitha, (b) chitoi pitha, (c) jamai pitha, (d) nakshi pitha, (e) naru, (f) patishapta pitha, (g) puli pitha, and (h) teler pitha

The sample size of each class in this dataset has been depicted in the pie chart shown in Figure 3. This chart illustrates that each of the classes has a similar number of samples, indicating that the dataset is balanced. There are no imbalance classes. Every class contains 655 images. The PithaNet dataset was randomly split among train, test, and validation, with around 70% of the photos used for training, 20% used for validation, and 10% used for evaluating the model. A treemap in Figure 4 provides a hierarchical view of the dataset, displaying the partitions of the dataset.

The overfitting issue with model training has been solved using image augmentation approaches, which also improves model performance [22]. By concentrating on horizontal flips, rotation, zooming, shear height-shift, width-shift, and rescale while augmenting, the CNN models will become less sensitive to the precise location of the item [23]. The rotation angle was 30, zoom range 0.2, height-shift, and width- shift range 0.2 and shear range was also 0.2. The augmentations were done in such a way that the quality of the images was not lost. As 10% (524) images were used to evaluate the model, the augmentation technique applies to the rest of the images which is 4716. We generated 10,000 images from 4,716 original images after augmentation. After that, all images were resized to 224×224. Figure 5 shows the differences between original images and augmented images.

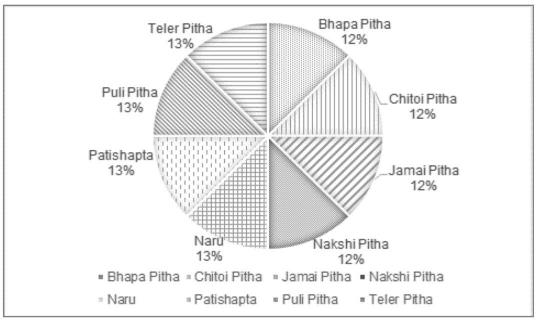


Figure 3. Representation of samples in every classes

■ Train Image ■ Validation Image ■ Test Image
Train Image
Validation Image
Test Image

Figure 4. Treemap of the size of train, validation, test

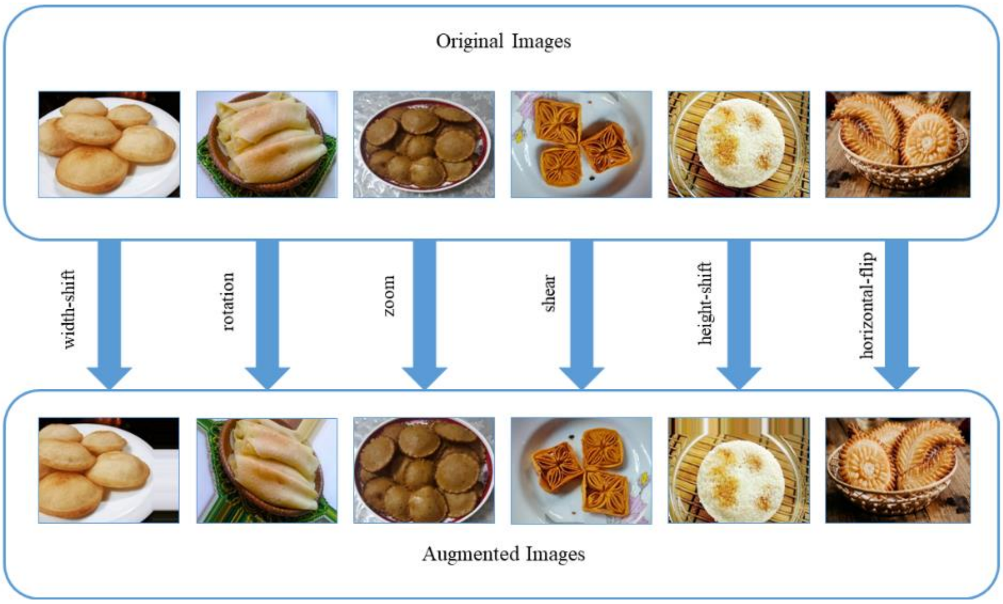


Figure 5. Original and augmented images

3.2. CNN

Most recently, CNNs have made significant advances in deep learning. It has resulted in a remarkable rise in the accuracy of image identification and recognition. A CNN is a sort of ANN that is mostly used for image detection and processing because of its capacity to spot patterns in visual data. CNN was the top contender for our experiment because feature engineering is not required there. If we used traditional ML approach in our pitha dataset, we also needed to use some feature extraction algorithms for feature selection in different types and shapes of pithas and it would take a lot of time. Also, we compare handcrafted features with CNN where CNN performs better.

3.2.1. Convolutional layer

Filtering actions are carried out mostly via convolutional layers. This layer is responsible for handling the vast majority of the computations of a CNN model. In order to construct the feature map, the set of images is fed into the surface. The kernel, also known as the feature detector, will look for features in the image. The kernel size can vary, but 3×3 or 5×5 matrices are commonly used. Following the convolution layer, the rectified linear unit (ReLU), which is frequently used in neural networks, performs the nonlinear activation process [24].

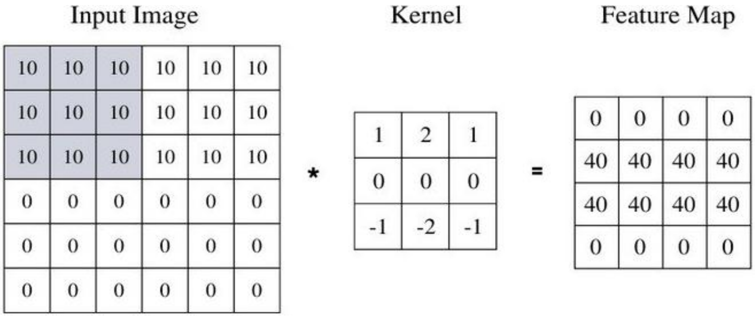
$$R(z) = max\ (0, z) \tag{1}$$

Any negative input causes the function to return 0, while any positive value z causes it to return that value. Thus, it may be expressed as (1).

3.2.2. Feature map

CNNs are unreasonably expensive due to the depth and numerous parameters. Therefore, dimension reduction between the layers is required. The dimension is reduced in the pooling layer by a down sampling process [25]. The pooling layer compiles the trait features in a specific area of the feature map. If the input is represented by f of an image $m\ x\ n$ dimension, h represents filter, and the result matrix is marked with $M\ [m\ x\ n]$ and j and k represent padding size and stride size respectively, then the formula to calculate feature map values is given in (2) and Figure 6 shows the calculations behind the feature map [26].

$$G[m,n]=(f * h)\ [m,\ n]=\textstyle\sum_j \sum_k h[j,\ k]\ f[m\ \text{-}j,\ n\text{-}\ k] \tag{2}$$



Input Image

10	10	10	10	10	10
10	10	10	10	10	10
10	10	10	10	10	10
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

Kernel

1	2	1
0	0	0
-1	-2	-1

Feature Map

0	0	0	0
40	40	40	40
40	40	40	40
0	0	0	0

Figure 6. Calculator of feature map of 6×6 input image with 3×3 kernel

3.3. Pre-trained models

Pre-trained models include network weights that have undergone training. Utilizing an architecture that has already been learnt from a classification task thereby minimizes the number of steps necessary for the output to converge. It is because, in general, the features that are collected for the classification task will be comparable. Initializing the model with the pre-trained weights saves time during training and is therefore more effective than starting with random weights. According to a survey in the literature on similar classification tasks, researchers have utilized different pre-trained models for food classification tasks, including AlexNet, VG16, EffecientNet, and SqueezeNet. In this paper we have selected three popular pre-trained models VGG16, ResNet50, and EffecientNetB6. A brief explanation of those three base models will be described in this section:

3.3.1. VGG16

VGG-16 is a 16 layers-deep CNN. It was first introduced in the ILSVRC-2014 by Simonyan and Zisserman [2]. On the ImageNet dataset, which has roughly 138 million parameters, this model won the top prize by achieving 92.7% accuracy and ranked in the top 5. The input in this architecture is 224×224. It has a convolution layer with 3×3 kernel size with stride 1, a max-pooling layer that employs 2×2 kernel with stride 2, and a total of three fully connected layers connected with SoftMax activation function. It is currently one of the most popular options in the community for extracting features from images. In ImageNet, which contains more than 14 million images from close to 1,000 classes, the VGG16 model scores about top-5 test accuracy. As our pitha dataset will have various classes, we put it on our list in the hopes that it will perform well.

3.3.2. ResNet50

A well-known neural network called residual networks, often known as ResNet, provides the basis for numerous computer vision tasks. [3] ResNet for image recognition was initially developed by Zhang [3], in a research paper titled “deep residual learning for image recognition”, which took first place in the ILSVRC-2015 contest. ResNet's ability to train incredibly deep neural networks was its core innovation. In the 34-layer net, every block of 2-layers is swapped out with this 3-layer bottleneck, block to produce a 50-layer ResNet, which has a fixed input size of 224×224 pixels. In this architecture, the number of parameters is almost 23 million. The reason of choosing ResNet50 was we wanted to test a deeper network than VGG16 and VGG19 in this experiment. ResNet50 contains more layers, but because global average pooling is used rather than fully-connected layers in the architecture, the overall size is really significantly smaller.

3.3.3. EfficientNet B6

Making use of a compound coefficient, the EfficientNet architecture and scaling algorithm consistently scales all depth, breadth, and resolution dimensions. Tan and Le [27] from Google Research, came up with this EfficientNet architecture in the research paper titled "EfficientNet: Rethinking model scaling for CNN." A presentation of this paper was made at the 2019 International Conference on ML. They suggested a brand-new scaling technique that scales the network's depth, width, and resolution equally. With the help of the neural network search, they developed a new baseline model and increased it up to produce the EfficientNets family of deep learning models, that beat the earlier CNNs in terms of performance and accuracy. The input size of EfficientNetB6 is 224 x 224 and it has more than 40M parameters [28]. Beside VGG16, which is a more popular and established model and ResNet50 for its deeper layer and different architecture, we choose EfficientNet hoping to obtain a good result with reasonable parameters and cost.

3.4. Transfer learning

Conventional ML models need to be trained from scratch with a substantial amount of data, which is computationally costly. Transfer learning is a powerful deep learning approach in which a network learned for a task "A" is repurposed for a new task "B". In a CNN model, the initial layers typically work to identify the common features from each image. The model only makes an attempt to differentiate between different classes in the final few layers. Because the higher layers deal with specific features, we influence the dichotomy by deciding how much to change the network's weights. Because PithaNet has a moderate number of samples and the dataset is comparable, we can avoid a lengthy training process by using the model's prior knowledge. Therefore, training the classifier and the upper layers of the convolutional base should be sufficient. As a result, this paper concentrates on preserving the original configuration of the convolutional base and more layers frozen to avoid overfitting. The typical approach for solving image classification-related problems is to use a set of fully connected layers come after two or more densely connected layers where the final dense layer has a commonly used activation function SoftMax, if the problem is multiclass classification. To minimize overfitting global average pooling layer is added which will minimize the overall number of parameters. Figure 7 depict the architecture of the proposed modified transfer learning model.

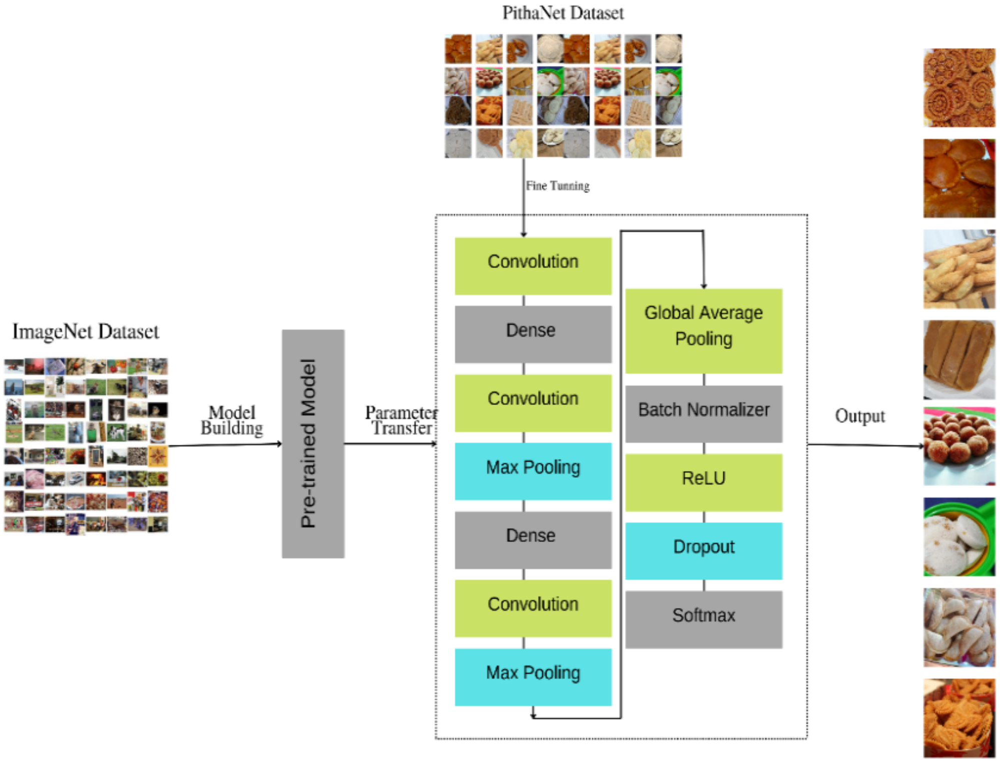


Figure 7. Proposed transfer learning architecture

4. RESULTS AND DISCUSSION

Since the PithaNet dataset involves 8 groups, in this study, multi-class classification has been carried out using EfficientNetB6, ResNet50, and VGG16. Every model mentioned above has been assembled into Google Colab with GPU support. Moreover, the models have been trained and fitted using the training and validation sets. Besides, a test set has been used to assess how well the model performed in classifying photos.

Table 2 summarises the input size, learning rate, epoch count, and parameter count of the adjusted model. As EfficentNetB6 and VGG 16 have significantly more parameters than ResNet50, therefore, this study used a learning rate of 0.1 for those two and 0.001 for ResNet50. Figure 8 displays the training, validation, and test accuracies on 8,000 training photos, 2,000 validation images, and 524 testing images. The EfficientNetB6 model train accuracy is about 96.5%, validation accuracy is 87.5%, and test accuracy is 90%. In comparison, ResNet50 performs slightly better than EfficientNetB6. It provides 97.6% train accuracy, 89.6% validation accuracy, and 90% test accuracy. Nonetheless, the VGG16 gives a satisfactory result with an accuracy of about 95.5%, a validation accuracy of 91.5, and a test accuracy of 91.5%.

Table 2. The input size, learning rate, number of epochs, and different parameters of the modified models

Model name	Input size	Learning rate	Number of epochs	Number of parameters	
				Total params	Trainable params
EfficentNetB6	224 * 224	0.01	50	98,767,511	57,807,368
ResNet50	224 * 224	0.01	50	74,972,552	51,384,840
VGG16	224 * 224	0.001	50	27,564,360	12,849,672

The number of epochs is the sum of all the entire iterations over the training dataset. The batch size is the quantity processed prior to model modification. All the models were trained with 50 epochs and the batch size was 16. For optimizing the algorithm Adam was used. As a loss function, the categorical cross entropy was applied, and metrics were utilized for accuracy. Figure 9 displays the model's training and validation accuracy and loss of EfficientNetB6, ResNet50, and VGG16 model. Also, it indicates that this model is not over fitted or under fitted. It performs well in the real test datasets images. This can be better understood from the confusion matrix.

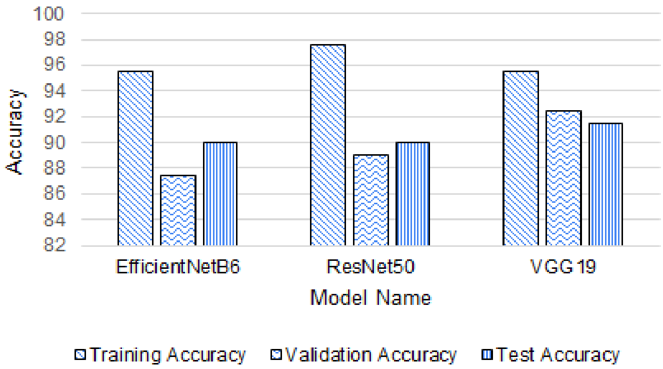


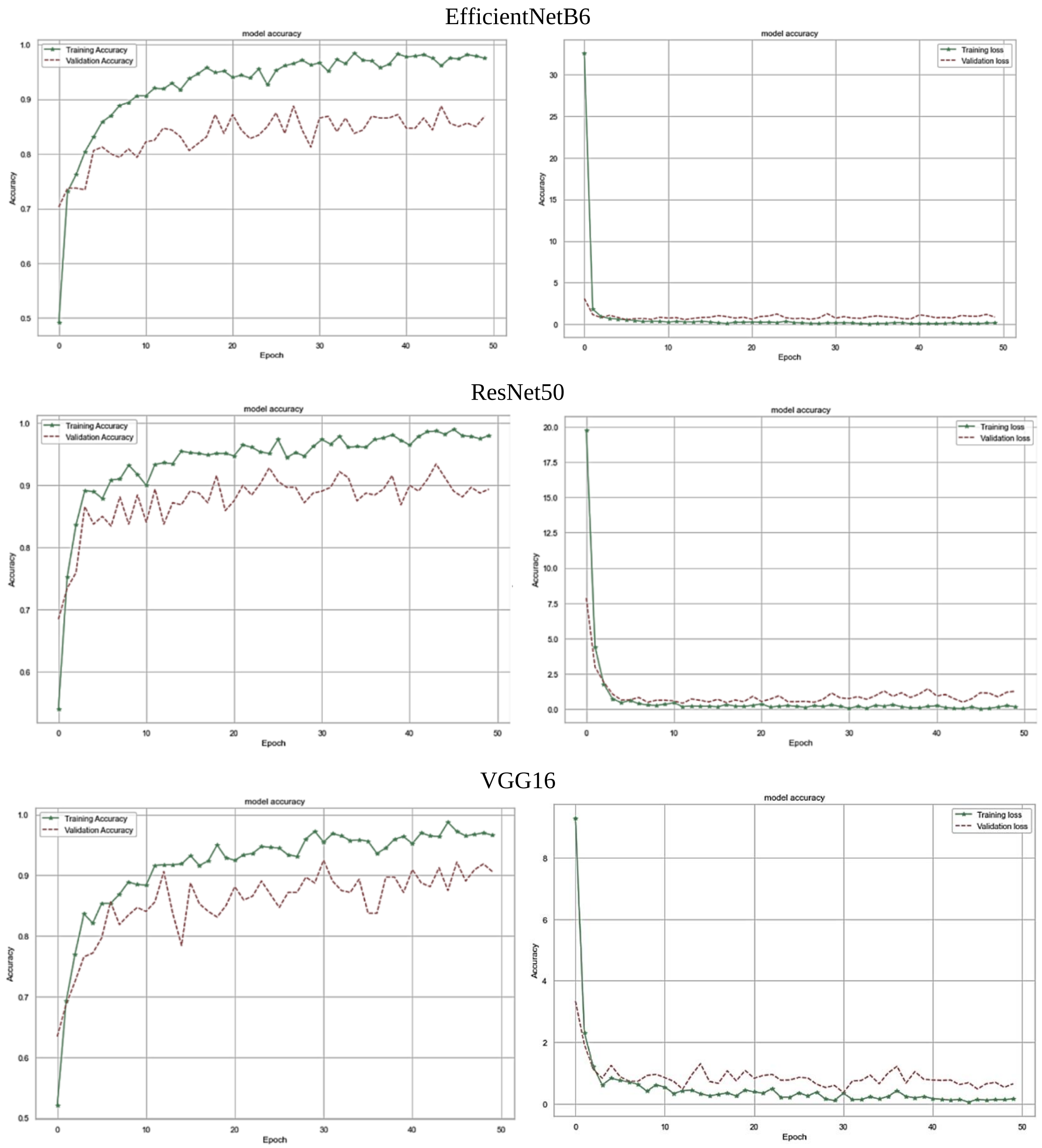
Figure 8. Accuracy of CNN models during training and testing

Figure 10 shows the confusion matrix of EfficientNetB6, ResNet50, and VGG16 model. Accuracy is a key component of every ML software. It is necessary to filter data through a categorization procedure called Confusion Matrix in order to make it trustworthy for subsequent operations. Machine learning assessment measures, such as machine learning specificity, machine learning accuracy, and machine learning sensitivity, are included in the confusion matrix.

Table 3 demonstrates the accuracy of 8 classes for EfficientNetB6, ResNet50, and VGG16 model. The percentage of accurately anticipated data points across all data points is what is known as accuracy. Following the (3), it can be calculated. Table 4 demonstrates the precision and recall of 8 classes for EfficientNetB6, ResNet50, and VGG16 model. Precision is defined as the ratio of correctly diagnosed positive samples to all samples that are positively classified. One metric for measuring the effectiveness of a machine learning model is precision. How many of the discovered objects are actually relevant depends on a model's precision. The (4) can be used to find it.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{4}$$

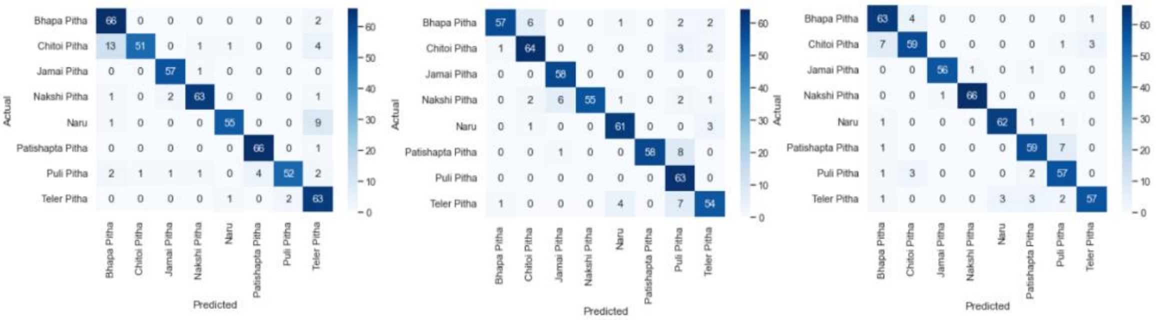


EfficientNetB6

ResNet50

VGG16

Figure 9. Training, validation accuracy and loss graph



EfficientNetB6 ResNet50 VGG16

Figure 10. Confusion matrix on 524 testing image set

Table 3. Accuracy (%) efficiency of EfficientNetB6, ResNet50, and VGG16 for each class

Class name	EfficientNetB6	ResNet50	VGG16
Bhapa Pitha	96.37	97.52	96.95
Chitoi Pitha	96.18	97.13	96.56
Jamai Pitha	97.38	98.66	99.43
Nakshi Pitha	98.66	97.71	99.62
Naru	97.71	98.09	98.85
Patishapta	99.05	98.28	97.14
Puli Pitha	97.52	98.30	96.76
Teler pitha	95.80	96.18	97.52

Table 4 Precision and recall (%) performance of EfficientNetB6, ResNet50, and VGG-16 for each class

Class name	Precision			Recall		
	EfficientNetB6	ResNet50	VGG16	EfficientetB6	ResNet50	VGG16
Bhapa Pitha	79.52	96.61	85.12	97.06	83.82	92.65
Chitoi Pitha	98.08	87.67	89.39	72.86	91.43	84.29
Jamai Pitha	95.00	89.23	98.25	83.83	1.00	96.55
Nakshi Pitha	95.55	1.00	98.51	94.03	82.09	98.51
Naru	96.49	91.04	95.38	84.62	93.85	95.38
Patishapta	94.29	1.00	89.39	98.51	86.57	88.06
Puli Pitha	96.30	87.50	83.82	82.54	1.00	90.48
Teler pitha	95.45	87.10	93.44	76.83	81.82	86.36

Recall or sensitivity refers to the quantity of positive records that were accurately anticipated. The proportion of positive samples to all positive instances is used to compute the recall that were correctly classified as positive. Recall measures how well the model can differentiate positive samples. The recall increases as more positive samples are found. It can be formulated by (5),

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{5}$$

Table 5 demonstrates the F1-score and specificity of 8 classes for EfficientNetB6, ResNet50, and VGG16 model. The F1-Score or F-measure is an evaluation metric for classifications, where it is specified as the harmonic average score of recall and accuracy. It is a metric used in statistics to assess how accurate a test or model is. It is represented as follows in mathematics (6). Specificity may be defined as the ability of something like the algorithm or system to predict a genuine negative of each available category. It is frequently referred to as the real negative rate in literature. It may be calculated using the (7).

$$F1 - Score \ = \frac{2(recall * precision)}{recall + precision} \tag{6}$$

$$Specificity = \frac{TN}{TN+FP} \tag{7}$$

Table 5. F1-score and specificity (%) performance of EfficientNetB6, ResNet50, and VGG-16 for each class

Class name	F1-Score			Specificity		
	EfficientNet6	ResNet50	VGG16	EfficientNetB6	ResNet50	VGG16
Bhapa Pitha	87.42	89.76	88.73	96.27	99.56	97.59
Chitoi Pitha	83.61	89.51	86.77	99.78	98.01	98.64
Jamai Pitha	89.06	94.31	97.39	99.36	98.49	99.79
Nakshi Pitha	94.74	90.16	98.51	99.34	1.00	99.78
Naru	90.16	92.42	95.38	99.56	98.69	99.35
Patishapta	96.35	92.80	88.72	99.12	1.00	98.47
Puli Pitha	88.89	93.33	87.02	99.57	98.07	97.61
Teler pitha	85.14	84.38	89.76	99.32	98.25	99.13

Figure 11. show the number of misclassified images, where out of 525 images EfficientNetB6 misclassified 51, ResNet50 misclassified 54, and VGG16 performed better with only 44 misclassified images. VGG16 founds a little difficulty in class “Chitoi Pitha” with 11 misclassifications but in the rest of the classes, the misclassification rate is significantly low. After reviewing all of these, it’s clear that VGG16 performs well on PithaNet dataset.

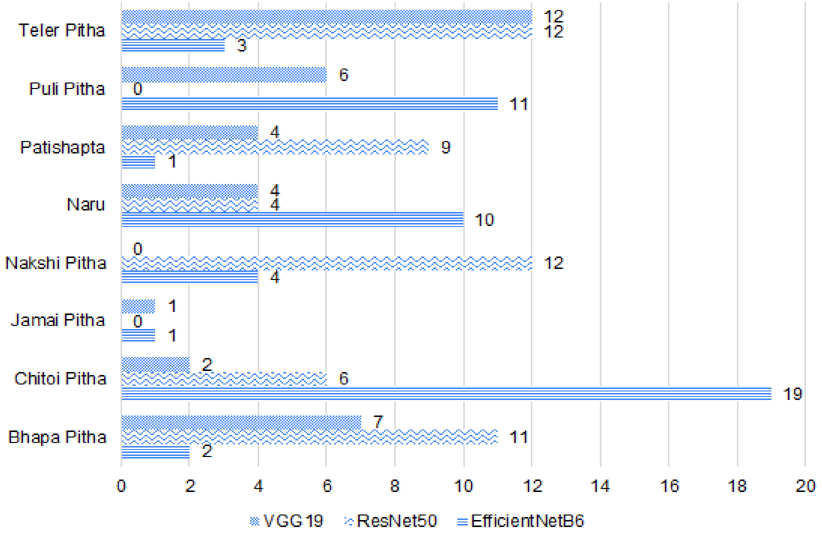


Figure 11. Total number of misclassified images

5. CONCLUSION

The intended outcome of this research is to classify eight different varieties of Bangladeshi pitha using a pre-trained CNN model. The study shows that VGG16, ResNet50, and EfficientNetB6 performs well. Among these, the model VGG16, based on a CNN, provides a higher accuracy of about 92%. Although image classification is a typical task in computer vision, it becomes quite complex while classifying multiple classes. Some pithas usually differ from one another in terms of shape. Even though it might be challenging to operate under certain constraints, we made an effort to get through these obstacles despite our lack of resources. Most importantly, the F1-score of the VGG16 model for each class is nearly 90%, which indicates that almost all types of classes are predicted successfully.

In the future, we would like to develop an automated application that can identify not only the pithas, but also their ingredients, and calorie information. Anyone using that app will be able to see the name and features of pithas by clicking on a particular image of a Pitha. This will be beneficial for future generations in order to gain more knowledge regarding our traditional pithas, which will enhance their cultural awareness as well. Future addition to this study will also include new datasets and pitha variations.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[2] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *Prepr. arXiv1409.1556*, Sep. 2014.

[3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Computer Vision and Pattern Recognition*, pp. 1–9, Dec. 2015.

[4] Z. Zahisham, C. P. Lee, and K. M. Lim, “Food recognition with ResNet-50,” in *2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Sep. 2020, pp. 1–5. doi: 10.1109/IICAIET49801.2020.9257825.

[5] Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, “Plant leaf disease classification using EfficientNet deep learning model,” *Ecological Informatics*, vol. 61, Mar. 2021, doi: 10.1016/j.ecoinf.2020.101182.

[6] A. Mahajan and S. Chaudhary, “Categorical image classification based on representational deep network (RESNET),” in *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, Jun. 2019, pp. 327–330. doi: 10.1109/ICECA.2019.8822133.

[7] L. Wen, X. Li, and L. Gao, “A transfer convolutional neural network for fault diagnosis based on ResNet-50,” *Neural Computing and Applications*, vol. 32, no. 10, pp. 6111–6124, May 2020, doi: 10.1007/s00521-019-04097-w.

[8] R. H. Hridoy, F. Akter, M. Mahfuzullah, and F. Ferdowsy, “A computer vision based food recognition approach for controlling inflammation to enhance quality of life of psoriasis patients,” in *2021 International Conference on Information Technology (ICIT)*, Jul. 2021, pp. 543–548. doi: 10.1109/ICIT52682.2021.9491783.

[9] G. Zeng, “Fruit and vegetables classification system using image saliency and convolutional neural network,” in *2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC)*, Oct. 2017, pp. 613–617. doi: 10.1109/ITOEC.2017.8122370.

[10] V. Kumari G., P. Vutkur, and P. Vishwanath, “Food classification using transfer learning technique,” *Global Transitions Proceedings*, vol. 3, no. 1, pp. 225–229, Jun. 2022, doi: 10.1016/j.gltp.2022.03.027.

[11] N. Hnoohom and S. Yuenyong, “Thai fast food image classification using deep learning,” in *2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON)*, Feb. 2018, pp. 116–119. doi: 10.1109/ECTI-NCON.2018.8378293.

[12] R. P. Prasetya and F. A. Bachtiar, “Indonesian food items labeling for tourism information using convolution neural network,” in *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, Nov. 2017, pp. 327–331. doi: 10.1109/SIET.2017.8304158.

[13] S. Yadav, Alpana, and S. Chand, “Automated Food image classification using deep learning approach,” in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Mar. 2021, pp. 542–545. doi: 10.1109/ICACCS51430.2021.9441889.

[14] A. Sengur, Y. Akbulut, and U. Budak, “Food image classification with deep features,” in *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, Sep. 2019, pp. 1–6. doi: 10.1109/IDAP.2019.8875946.

[15] A. Singla, L. Yuan, and T. Ebrahimi, “Food/non-food image classification and food categorization using pre-trained GoogLeNet model,” in *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, Oct. 2016, pp. 3–11. doi: 10.1145/2986035.2986039.

[16] G. Ö. Yiğit and B. M. Özyildirim, “Comparison of convolutional neural network models for food image classification,” *Journal of Information and Telecommunication*, vol. 2, no. 3, pp. 347–357, Jul. 2018, doi: 10.1080/24751839.2018.1446236.

[17] D. A. Kurnia, A. Setiawan, D. R. Amalia, R. W. Arifin, and D. Setiyadi, “Image processing identifacation for Indonesian cake cuisine using CNN classification technique,” *Journal of Physics: Conference Series*, vol. 1783, no. 1, Feb. 2021, doi: 10.1088/1742-6596/1783/1/012047.

[18] K. Kusrini, M. R. A. Yudianto, and H. Al Fatta, “The effect of Gaussian filter and data preprocessing on the classification of Punakawan puppet images with the convolutional neural network algorithm,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, pp. 3752–3761, Aug. 2022, doi: 10.11591/ijece.v12i4.pp3752-3761.

[19] M. S. I. Prottasha and S. M. S. Reza, “A classification model based on depthwise separable convolutional neural network to identify rice plant diseases,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, pp. 3642–3654, Aug. 2022, doi: 10.11591/ijece.v12i4.pp3642-3654.

[20] A. AL Smadi, A. Mehmood, A. Abugabah, E. Almekhlafi, and A. M. Al-smadi, “Deep convolutional neural network-based system for fish classification,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 2, pp. 2026–2039, Apr. 2022, doi: 10.11591/ijece.v12i2.pp2026-2039.

[21] M. Alswaitti, L. Zihao, W. Alomoush, A. Alrosan, and K. Alissa, “Effective classification of birds’ species based on transfer learning,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, pp. 4172–4184, Aug. 2022, doi: 10.11591/ijece.v12i4.pp4172-4184.

[22] R. Yunus *et al.*, “A framework to estimate the nutritional value of food in real time using deep learning techniques,” *IEEE Access*, vol. 7, pp. 2643–2652, 2019, doi: 10.1109/ACCESS.2018.2879117.

[23] M. Arar, A. Shamir, and A. Bermano, “InAugment: Improving classifiers via internal augmentation,” *Prepr. arXiv.2104.03843*, Apr. 2021.

[24] N. A. Mohammed, M. H. Abed, and A. T. Albu-Salih, “Convolutional neural network for color images classification,” *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 11, no. 3, pp. 1343–1349, Jun. 2022, doi: 10.11591/eei.v11i3.3730.

[25] W. Xiong, L. Zhang, B. Du, and D. Tao, “Combining local and global: Rich and robust feature pooling for visual recognition,” *Pattern Recognition*, vol. 62, pp. 225–235, Feb. 2017, doi: 10.1016/j.patcog.2016.08.006.

[26] D.-X. Zhou, “Theory of deep convolutional neural networks: Downsampling,” *Neural Networks*, vol. 124, pp. 319–327, Apr. 2020, doi: 10.1016/j.neunet.2020.01.018.

[27] M. Tan and Q. V. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *Proceedings of the 36 International Conference on Machine Learning*, Long Beach, California, PMLR 97, 2019.

[28] B. Koonce, “EfficientNet,” in *Convolutional Neural Networks with Swift for Tensorflow*, Berkeley, CA: Apress, 2021, pp. 109–123. doi: 10.1007/978-1-4842-6168-2_10.

BIOGRAPHIES OF AUTHORS

Shahriar Shakil received the B.Sc. degree in Computer Science and Engineering from Daffodil International University, Dhaka, Bangladesh in 2021. Currently, he is working as a Lecturer in Computer Science and Engineering department at Daffodil International University. Previously, he worked at Daffodil International University as a Teaching Assistant, and at Expert Consortium Ltd. as a Machine Learning Programmer. He has published the results of his study in international peer-reviewed publications. He published several courses at the International Online University (IOU) related to Python. His research interests are machine learning, deep learning, computer vision, natural language processing, and data mining. He can be contacted at email: shahriar15-8558@diu.edu.bd.

Atik Asif Khan Akash received his Bachelor of Science degree in Computer Science and Engineering from Daffodil International University, Bangladesh in 2021. His current field placement is with Daffodil International University (DIU). He is currently working as a Lecturer at the Department of Computer Science and Engineering, Daffodil International University (DIU), Daffodil Smart City, Dhaka, Bangladesh. Before that he had done a one-year long fellowship program called “Teaching Apprentice Fellowship -2022”, where he has contributas Teaching Assistant at DIU. His area of research interest includes Machine learning, Computer Vision, Image Processing, and Object Recognition. Atik has facilitated a number of workshops and seminars on machine learning and data science in collaboration with various clubs and non-profit organizations that aimed to promote skill development. He also contributed programming and data science-related courses to online course providers such as International Online University (IOU). He can be contacted at email: akash.cse@diu.edu.bd.

Nusrat Nabi has completed B.Sc. degree in the department of computer science and engineering (CSE), Daffodil International University, Dhaka, Bangladesh. She is currently working as a Lecturer at DIU. She worked as a document processing engineer at Apurba Technologies, Ltd. It is a USA-based software firm specializing in providing Machine learning solutions for expertise. She worked as Vice President of Research and Journal wing at Daffodil International University Girls' Computer Programming Club-DIU GCPC. Her current research interests include Machine learning, Natural language processing, Computer vision, Data mining and Deep learning. She is an active researcher and reviewer for Q3 journals. She can be contacted at email: nusrat15-10524@diu.edu.bd.

Mahmudul Hasan is working as Young Professional (Technology Management) at Aspire to Innovate (a2i) Program of ICT Division, Bangladesh. His major areas of interest in study are data mining, machine learning, image processing, and natural language processing. Additionally, data science and computer vision are among his areas of interest. He can be contacted at email: mahmudul15-8991@diu.edu.bd.

Aminul Haque is currently working as Associate Professor at the Department of Computer Science and Engineering, Daffodil International University (DIU), Daffodil Smart City, Dhaka, Bangladesh. Before joining DIU, he completed his Ph.D. from MONASH University. Dr. Haque completed his B.Sc. from Shahjalal University of Science and Technology, Bangladesh. His area of research interest includes Data mining, Machine learning and Distributed computing. He has published his research outputs in several international peer reviewed journals and conferences. He also contributed data science related courses to online platforms such as International Online University (IOU). Recently he contributed to developing a skill-based national curriculum on big data and data science related courses. He can be contacted at email: aminul.cse@daffodilvarsity.edu.bd.