RESEARCH ARTICLE

SAMRIDDHI Volume 14, Issue 3, 2022 Print ISSN: 2229-7111 Online ISSN: 2454-5767

Performance Evaluation of Indian Food Image Classification system using Transfer Learning with MobileNetV3

Jigar Patel[1*] Hardik Talsania[2], Kirit Modi[3]

[1]Ph.D. Scholar, Department of Computer Science & Engineering Parul University, Vadodara, India
[2]Ph.D. Scholar, Department of Computer Science & Engineering Parul University, Vadodara, India
[3]Head, Department of Computer Engineering Sankalchand Patel University, Visnagar, India

Abstract

Food image classification and recognition is an emerging research area due to its growing importance in the medical and health industries. As India is growing digitally rapidly, an automated Indian food image classification system will help in the development of diet tracking, calorie estimation, and many other health and food consumption-related applications. In recent years many deep learning techniques evolved. Deep learning is a robust and low-cost method for extracting information from food images, though, challenges lie in extracting information from real-world food images due to various factors affecting image quality such as photos from different angles and positions, several objects appearing in the photo, etc. We used Non-Standard dataset for Indian Food images of 13 different classes consist more than 3500 images. In this paper, we use CNN as our base model to build our system, which gives an accuracy of 47% of the system. After that, we deployed the transfer learning technique with MobileNetV3 for improvement in accuracy, which resulted in an improvement in accuracy of up to 90%.

Keywords: Indian Food dataset, CNN, Deep Learning,MobileNetV3, Image Processing, Transfer Learning,.

SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology (2022); DOI: 10.18090/samriddhi.v14i03.24

Introduction

Computer Vision is the disciplines of computer science and artificial intelligence that empowers computer systems to extract significant and valuable insights from visual data in the form of images. Traditional image analysis methods have historically exhibited low classification accuracy. Deep learning techniques, however, have demonstrated promising results in the identification of food types and its constituent parts. Automated food identification is a rapidly growing research topic with important implications not only for the online community, but also for the medical and healthcare field. It has the capacity to aid in the estimation of caloric intake, the determination of food quality, and the development of dietary tracking systems for addressing issues like obesity.[1-3] However, the inter-class variability and the large within-class variability make it difficult to identify the food in the pictures, making it impossible to capture the sophisticated aspects with standard techniques.[4] CNNs-Convolutional neural networks have emerged as a powerful way for classifying food images to address this issue because they can increase classification precision while still being able to recognize such minor traits on their own.[5]

Deep learning techniques, coupled with the availability of larger datasets and computational resources, have simplified image classification tasks. Among these techniques, CNNs have gained popularity and are extensively used for image classification.[6,7] The food products have complex semantically distorted entities, making it difficult to define their structure.[8] In this regard, deep learning techniques show excellent performance. Subsequently, CNNs were able to easily recognize such types of features autonomously, improving their classification accuracy.[9,10] Therefore, our research aims to classify food images using CNN. For example, the two foods shown in Figure 1 represent images of chapati and butter naan. Although they look the same, they are two different foods classes.

In this research, we demonstrated that rather than constructing models from scratch, transfer learning techniques are used with already established methodologies. While producing better results, this method reduces computing

Corresponding Author: Jigar Patel, Ph.D. Scholar, Department of Computer Science & Engineering Parul University, Vadodara, India, e-mail: jigarsharp@gmail.com

How to cite this article: Patel, J., Talsania, H., Modi, K. (2022). Performance Evaluation of Indian Food Image Classification system using Transfer Learning with MobileNetV3. *SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology,* 14(3), 398-405.

Source of support: Nil

Conflict of interest: None

Chapati **Butter Naan**

Figure 1: Identical food item of two different class

time as well as costs. For training and validation, the project makes use of a small & new food dataset with several classes and around 300 of photos for each class. The system can identify categories of images by utilizing pre-trained models.

After the introductory section, the subsequent sections of this paper are structured in the following manner: Section 2 provides an overview of the existing literature and research that is relevant to the topic at hand. Section 3 of this study pertains to the Methodology and Proposed System. Section 4 of the document presents the details pertaining to the implementation process and the subsequent results. Following this, Section 5 provides a comprehensive conclusion.

Related Work

This section provides an overview of the methodologies employed by researchers in the classification of food images across different food categories.

In their comprehensive analysis, the authors [11] examined a substantial number of studies that utilised deep learning methodologies in order to tackle various challenges within the food industry. These challenges encompassed a wide range of issues, including but not limited to food identification, calorie estimation, quality assessment of vegetables, fruits, meat, and aquatic food items, as well as concerns related to the food supply chain and the detection of contamination in food products. The study examined the potential of utilising deep learning as an advanced data mining technique in investigating the relationship between food consumption and sensory perception. The survey's findings indicate that deep learning outperforms alternative methods, such as manual feature extractors and traditional machine learning algorithms, in terms of its potential as a tool for assessing the quality and safety of food.

The researchers [12] conducted a series of experiments to demonstrate the significance of the Bag of Visual Words (BoW) paradigm in accurately classifying different types of food items. The initial step involved the application of rotation and scale invariant filters to the images, resulting in the generation of a compact codebook of Textons specific to each food category. The Textons, which were derived from a class-based learning approach, were aggregated to construct a comprehensive visual dictionary. The food images were analysed as visual word distributions, specifically employing the Bag of Textons approach, during the classification phase with the utilisation of a Support Vector Machine (SVM). The results of the trials indicate that the utilisation of the Bag of Textons methodology for image representation yielded higher levels of accuracy compared to the existing approaches employed for categorising the 61 categories within the Pittsburgh Fast-Food picture Dataset.

[13] suggested using a genetic algorithm and a modified coding ant colony optimisation technique to simultaneously control the SVM's parameters and feature selection in order to categorise remote sensing photos using the best SVM. The results were compared to those of different evolutionary algorithms and swarm intelligence methodologies, such as GA, BPSO, BACO, BDE, and BCS. The proposed method performed better than the other research methods in terms of fitness values and was demonstrated to be durable and adaptive, indicating that it would be suitable for a few applications in daily life.

The predictive Thai fast food model [14] was constructed through a deep learning process, wherein it was trained using real-world pictures from the GoogLeNet dataset. The dataset, referred to as the TFF (Thai Fast Food) dataset by the researchers, comprised a total of 3,960 images. A total of eleven distinct sets of food images were generated from the dataset, with each set exclusively featuring a distinct Thai culinary dish. The final category encompasses food items commonly found in Thai fast-food establishments that do not fall under any of the previously mentioned ten categories (non-ten types). The results obtained from an alternative test set indicate that the classification average accuracy for predicting Proposed Thai fast food is 88.33%.

The FoodAI smart-food logging system was proposed by [15] in order to leverage the convenience provided by smartphones. This system offers state-of-the-art image recognition capabilities based on deep learning techniques. The development of FoodAI took place in Singapore, with a specific emphasis on the regional cuisines that enjoy significant popularity within the locality. The FoodAI models were trained using a corpus consisting of 400,000 food images, which were categorised into 756 distinct classes. A comprehensive examination and elucidation of the developmental process of this technology were also provided. The smartphone application Healthy 365, developed by the Health Promotion Board of Singapore, incorporated the utilisation of FoodAI, an API service.

In their study, the authors [16] presented a novel methodology that combines the Convolutional Block Attention Module (CBAM) with three well-known deep learning architectures, namely MobileNetV2, VGG16, and ResNet50. The objective of this approach was to effectively categorise images of Asian cuisine. The implementation of a Mixup strategy, specifically mixed data augmentation, was suggested as a means to improve the accuracy of discrimination by promoting smoother transitions between data points. The experimental comparison demonstrated the impact of implementing the Mixup technique and incorporating the CBAM mechanism on the outcomes. Upon considering these two criteria, the resulting Top-1 accuracy percentage for the final test set was 87.33%. The heat map's presentation was consistent with the emphasised aspects highlighted by CBAM. The findings yielded novel perspectives that enhanced the classification of photographs depicting Asian cuisine and validated the utilisation of the classification methodology.

In their study, the researchers [17] investigated the integration of an advanced food-matching technology with a widely recognised and validated method for studying meal choices, known as the "fake food buffet." The objective was to streamline the collection and analysis of data through automation. The approach integrates deep learning techniques for identifying fake-food images with natural language processing methods for matching and standardising meals. The performance of the system was evaluated by employing metrics that rely on the intersection over union principle, as well as the widely accepted industry standard for pixel accuracy. The deep learning model exhibited an overall accuracy of 92.18%, with a classification accuracy of 93% specifically for food matching. The deep learning model underwent training using synthetic food images that were collected by a total of 124 research participants. These images were categorised into fifty-five distinct food classes.

The study conducted by the researchers involved the utilisation of pre-trained ResNet-152 and GoogleNet convolutional neural networks (CNNs) to analyse food picture datasets, namely Food 5K, Food-11, RawFooT-DB, and Food-101. The researchers extracted various characteristics from these datasets as part of their analysis. [18] The networks were initially trained utilising the MatConvNet software and the ILSVRC dataset, which stands for ImageNet Large Scale Visual Recognition Challenge. In this study, deep features were extracted from Convolutional Neural Networks (CNNs) for the purpose of training machine learning classifiers, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest, and Naive Bayes. The findings indicate that the utilisation of ResNet-152 deep features in conjunction with SVM and RBF kernel yields accurate classification of food items, achieving a 99.4% accuracy rate. This outcome was observed through the analysis of the Food-5K validation food picture dataset. In a similar vein, the results indicate a high level of accuracy, specifically 98.8%, when employing the identical dataset and classifiers, namely Random Forest, SVM-RBF, and ANN.

Methodology and Proposed System

In this section, we describe the proposed Indian food image classification and recognition system based on CNN and Transfer Learning. The research Proposed model has been depicted in Figure 2..

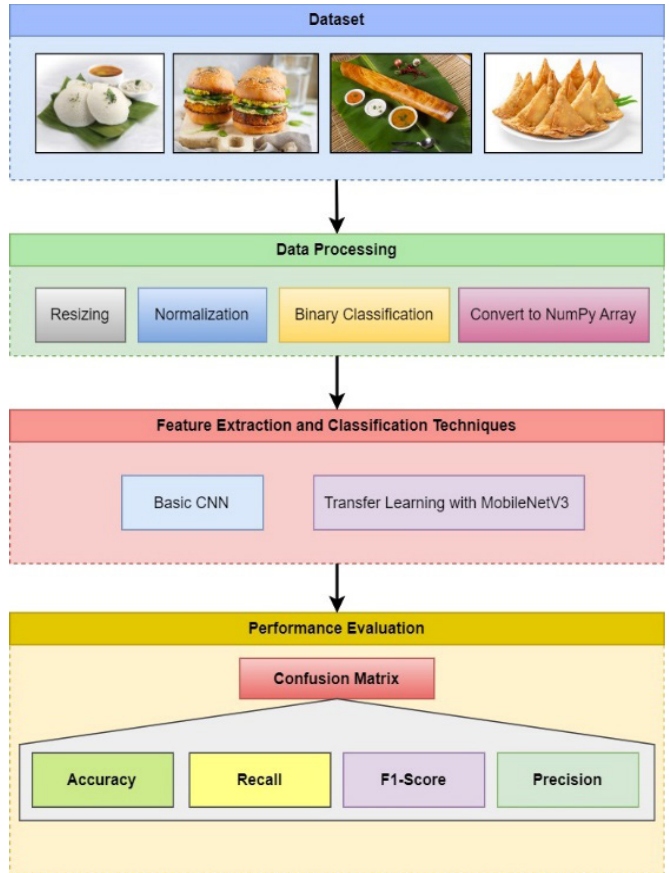

Figure 2: Proposed Indian Food Classification and Recognition System Architecture

Indian Food Image Dataset

The "13_Indian_food dataset" was utilised in the researcher's study. The dataset consists of over 3500 images. The dataset contains a comprehensive collection of food products, encompassing more than 13 distinct categories. Each category within the dataset contains a substantial number of images, exceeding 250 pictures per category. In this study, we will specifically focus on the food items indigenous to our local region. In order to mitigate the issue of overfitting, a portion of the training data, specifically 20%, is allocated for the purpose of validations. The proposed model for the research is illustrated in Figure 2.

Data Pre-Processing

Following the acquisition of the dataset, we proceeded to implement brightness normalisation. The aforementioned process involves altering the spectrum of pixel intensity values. One example of applications pertains to photographs exhibiting diminished contrast as a consequence of glare. Normalisation, also known as contrast stretching or histogram stretching, is a technique that is occasionally referred to by these alternative terms. . In order to gain insight into the distribution of the quantity of images within each category, a list was generated containing 13 categories, each accompanied by its respective number of images. In addition, the images underwent rescaling during the conversion process, achieved by dividing each pixel value by 255

Feature Extraction And Classification

For feature extraction and classification from the images, here we used Convolutional Neural Network method. We first build basic CNN model and then we used transfer learning method using MobileNetV3.

CNN (Base Model)

CNN, or Convolutional Neural Network, is a type of deep learning algorithm characterised by its feed-forward architecture. It is designed to process input images and effectively differentiate between various objects by assigning significance to specific features through the use of learnable weights. The development of CNNs is rooted in the biological principle of neurons found in the human brain, which possess the ability to acquire abstract characteristics and effectively categorise objects, demonstrating a notable capacity for generalisation. One of the primary rationales for selecting Convolutional Neural Networks (CNNs) lies in their capacity to facilitate weight sharing, thereby leading to a reduction in the number of parameters required for training. This approach facilitates seamless training and addresses the challenge of overfitting. The typical architecture of a Convolutional Neural Network (CNN) comprises several blocks of convolutional layers, activation functions, pooling layers, and fully connected layer. These components collectively contribute to the process of feature extraction and classification. [9, 19-21].

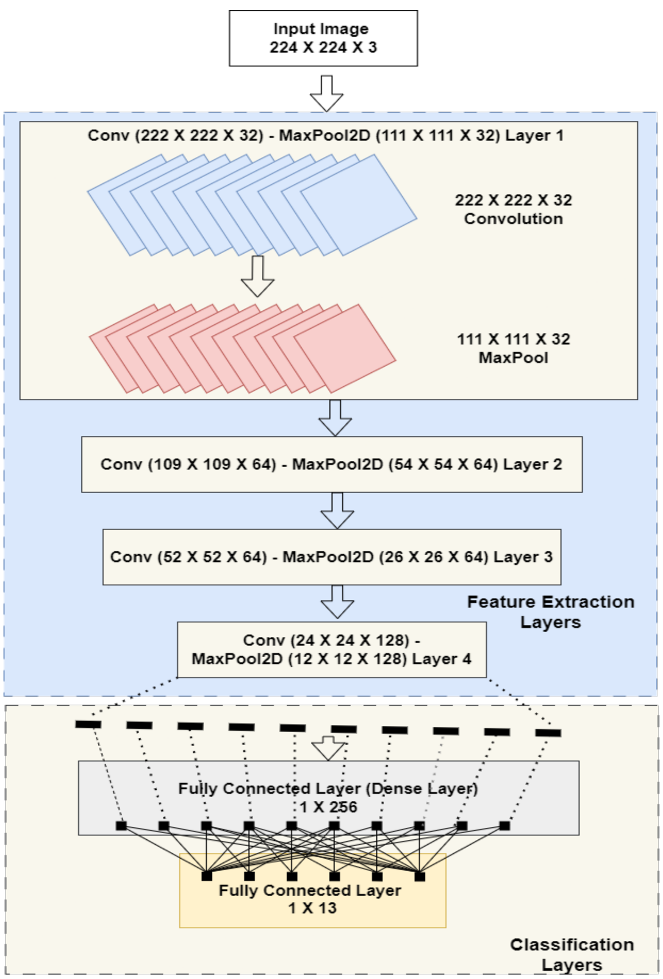

Figure 3: Architecture of Proposed CNN model

The convolutional neural network (CNN) model employed in this research was constructed de novo. After conducting multiple iterations, the most optimal outcomes were achieved for a model comprising of four convolution layers, each with a kernel size of 3, padding set to 'same', and utilising the rectified linear unit (ReLu) activation function.

The dimensions of the image input are 224 pixels in width, 224 pixels in height, and 3 channels. The number of filters employed in each convolutional layer increases twofold in the subsequent layer. Specifically, the initial layer comprises 32 filters, whereas the final layer comprises 128 filters. After each convolution layer, a subsequent maxpooling layer is applied with a pool size of 2. This process effectively reduces the dimensionality of the data while preserving the most significant features. The output obtained from the maxpooling layer is transformed into a flattened representation, converting the 3D matrix of features into a vector. This vector is then forwarded to a fully connected dense layer that consists of 256 units and utilises the rectified linear unit (ReLU) as its activation function. The final layer of the neural network consists of a dense layer with a number of units equal to the total number of food categories present in the dataset, which is 13. The activation function used in this layer is 'softmax', which is responsible for classifying the images. The softmax function was employed in this study due to the presence of a multiclass classification problem. Architecture of proposed CNN Model shown in Figure 3.

Transfer Learning with MobileNetV3

MobileNet is a simple, efficient and very computationally very light weight convolutional neural networks for mobile vision applications.

MobileNetV3 is a convolutional neural network architecture designed for efficient and lightweight mobile applications. It is the third generation in the MobileNet series, following MobileNetV1 and MobileNetV2. MobileNetV3 was introduced by Google in 2019 and aimed to improve both accuracy and efficiency compared to its predecessors.

The key idea behind MobileNetV3 is to use a combination of different techniques to achieve a good trade-off between model size, computational cost, and accuracy. Some of the main features of MobileNetV3 include:

Efficient inverted residuals

MobileNetV3 utilizes inverted residual blocks, which are a combination of a lightweight bottleneck layer and an expansion layer. This design reduces the computational cost while maintaining a high level of representational power.

Activation functions

MobileNetV3 introduces novel activation functions, such as the Hard Swish and the Hard Sigmoid, which are designed to strike a balance between non-linearity and computational efficiency. These activation functions help in reducing the number of operations required while maintaining good accuracy

In this experiment, MobileNetv3 was employed for the purpose of transfer learning. The outcomes of food categorization can potentially be obtained due to the utilisation of MobileNetV3, which has undergone pre-training on the ImageNet dataset. Hence, MobileNetV3 was employed for the purpose of feature extraction, followed by the application of a dense layer comprising 256 units, utilising the ReLU activation function. Subsequently, the photos were subjected to classification through the utilisation of a dense layer, wherein the number of units corresponded to the total count of food categories present within the dataset, which amounted to 13 in the current scenario. The softmax function was employed in this particular scenario as a result of the presence of a multiclass categorization problem. The architecture of the proposed Transfer Learning Model is depicted in Figure 4.

Model Evaluation

A variety of standards for measuring the effectiveness of the classification system have been adopted by the scientific community. The confusion matrix is used to assess the research's success using the following key parameters: true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN). Validity metrices such as Accuracy, Recall, F1-score and Precision can be calculated using these parameters. Formulas of these validity matrices are as follows:

$$\text{Accuracy} = {}^{(TP+TN)}\!\big/_{(TP+FP+TN+FN)} \qquad (1)$$

$$\text{Precison} = {}^{TP}\!\big/_{(TP+FP)} \qquad (2)$$

$$\text{Recall} = {}^{TP}\!\big/_{TP+FN} \qquad (3)$$

$$\text{F1}-\text{score} = \frac{2*(\text{Precision}\times\text{Recall})}{(\text{Precision}+\text{Recall})} \qquad (4)$$

Implementation And Results

The experiment conducted in this study utilised Kaggle.com, an online platform specifically designed for deep learning purposes. In this study, our system is executed on a GPU P100, which has been made available through a portal, in order to enhance and expedite the computational processes. In this study, the implementation was carried out using the Tensorflow and Keras libraries. As previously mentioned, the dataset employed in this study comprises a collection of images depicting Indian cuisine, encompassing a total of 13 distinct food categories. The dataset contains a total of over 3500 images, with each class consisting of more than 250 images. The following are examples of images depicting various classes that are present within our dataset.

The data can be divided into three distinct categories, namely Training, Validation, and Testing. The training data will be utilised for the purpose of training the deep learning convolutional neural network (CNN) model. Subsequently, the performance of the model will be assessed by employing the test data, which consists of data that has not been previously encountered by the model. The images in this dataset were selected in a random manner using the ImageDataGenerator class from the keras library. Subsequently, in order to enhance the performance of the model, we conducted pre-processing on the images.

Implementation of CNN (Basic Model)

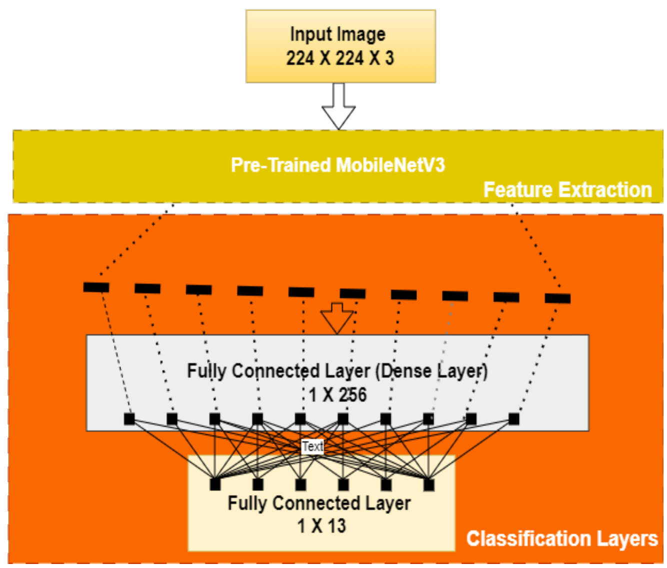

Figure 4: Architecture of Proposed Transfer Learning model

Layer (type)	Output Shape	Param #
conv2d_36 (Conv2D)	(None, 222, 222, 32)	896
max_pooling2d_36 (MaxPooling2D)	(None, 111, 111, 32)	0
conv2d_37 (Conv2D)	(None, 109, 109, 64)	18496
max_pooling2d_37 (MaxPooling2D)	(None, 54, 54, 64)	0
conv2d_38 (Conv2D)	(None, 52, 52, 64)	36928
max_pooling2d_38 (MaxPooling2D)	(None, 26, 26, 64)	0
conv2d_39 (Conv2D)	(None, 24, 24, 128)	73856
max_pooling2d_39 (MaxPooling2D)	(None, 12, 12, 128)	0
flatten_12 (Flatten)	(None, 18432)	0
dense_31 (Dense)	(None, 256)	4718848
dropout_6 (Dropout)	(None, 256)	0
dense_32 (Dense)	(None, 13)	3341

Total params: 4,852,365
Trainable params: 4,852,365
Non-trainable params: 0

Figure 5: Model Summary of CNN (Base Model)

Now our dataset is ready as an input to the model. Here We build our convolutional neural network model from scratch. In this we use following parameters:

Input Shape: (224, 224, 3)
Epoch: 100
Initial learning rate:0.0001
Batch size: 32

In feature extraction phase of CNN, we used ReLU activation function and at the time of classification used softmax activation function for proper classification of food images. As seen in figure 5, we have total trainable parameters are more than 4.8 million. Now after training this model, it's time to test our model and calculate the accuracy of our model. Here as you can see in figure, we got Test accuracy of our system is around 47%.

Here in Figure 6, we can see sample results of prediction, Train and validation accuracy and loss graph shown in Figure 7. Confusion matrix is shown in Figure 8 and classification report of system build with CNN (Base Model) shown in

Figure 6: Sample of Prediction Result

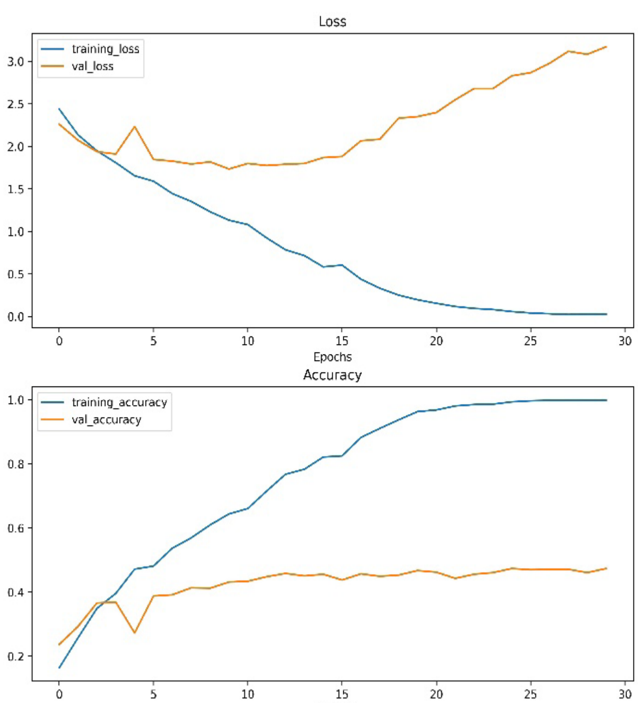

Figure 7: Graph of Prediction Accuracy and Loss vs Epochs

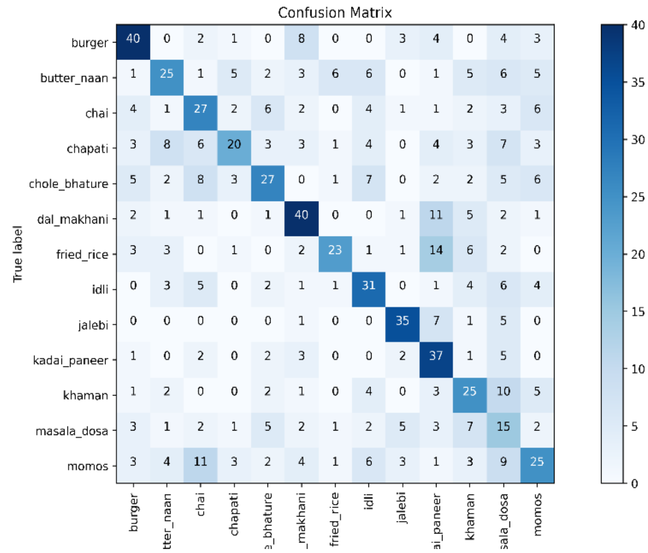

Figure 8: Confusion Matrix of system with CNN (Base Model)

	precision	recall	f1-score	support
burger	0.61	0.62	0.61	65
butter_naan	0.50	0.38	0.43	66
chai	0.42	0.46	0.44	59
chapati	0.56	0.31	0.40	65
chole_bhature	0.52	0.40	0.45	68
dal_makhani	0.57	0.62	0.59	65
fried_rice	0.68	0.41	0.51	56
idli	0.48	0.53	0.50	58
jalebi	0.69	0.71	0.70	49
kadai_paneer	0.42	0.70	0.52	53
khaman	0.39	0.47	0.43	53
masala_dosa	0.19	0.31	0.23	49
momos	0.42	0.33	0.37	75
accuracy			0.47	781
macro avg	0.49	0.48	0.48	781
weighted avg	0.50	0.47	0.47	781

Figure 9: Classification Report of CNN (Base Model).

Layer (type)	Output Shape	Param #
MobilenetV3large (Functional	(None, 1280)	4226432
dense (Dense)	(None, 256)	327936
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 13)	3341

Total params: 4,557,709
Trainable params: 331,277
Non-trainable params: 4,226,432

Figure 10: Summary of model transfer learning with MobileNetV3

Figure 11: Sample of the Prediction Results

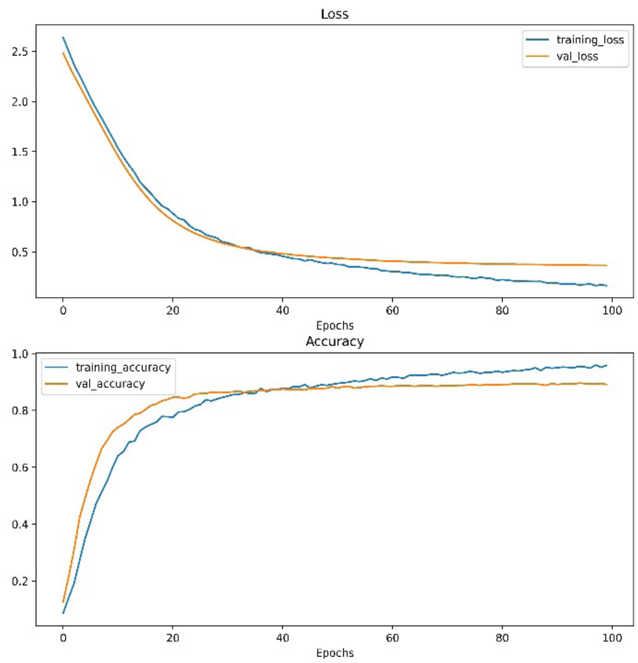

Figure 12: Graph of Prediction Accuracy and Loss vs Epochs

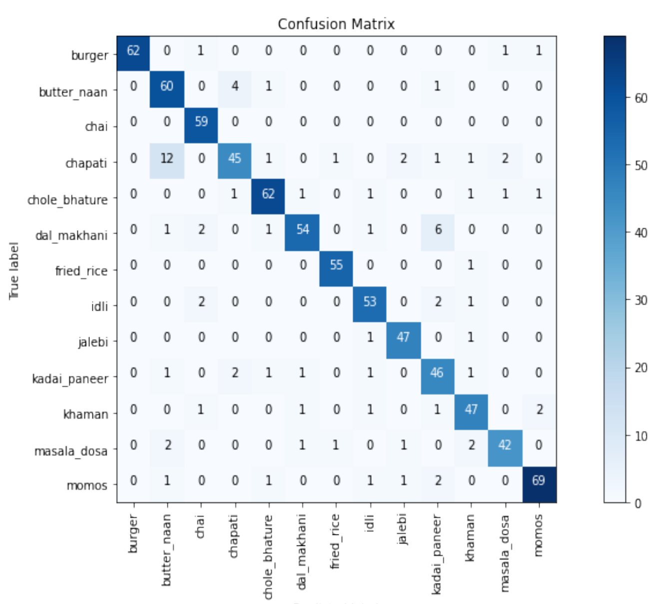

Figure 13: Confusion Matrix of system with Transfer Learning Method

	precision	recall	f1-score	support
burger	1.00	0.95	0.98	65
butter_naan	0.79	0.91	0.85	66
chai	0.94	1.00	0.97	59
chapati	0.89	0.77	0.83	65
chole_bhature	0.95	0.93	0.94	68
dal_makhani	0.95	0.83	0.89	65
fried_rice	0.93	0.98	0.96	56
idli	0.87	0.90	0.88	58
jalebi	0.94	0.96	0.95	49
kadai_paneer	0.83	0.85	0.84	53
khaman	0.87	0.91	0.89	53
masala_dosa	0.84	0.88	0.86	49
momos	0.94	0.91	0.93	75
accuracy			0.90	781
macro avg	0.90	0.91	0.90	781
weighted avg	0.91	0.90	0.90	781

Figure 14: Classification Report of Transfer Learning Model.

Table 1: Comparison of Accuracy

Model	*Accuracy*	*Precision*	*Recall*	*F1-score*
CNN (Basic Model)	49%	50%	47%	47%
MobileNetV3	90%	91%	90%	90%

Figure 9.

Implementation of MobileNetV3 (Transfer Learning Model)

To boost the performance of our system, we then used Transfer Learning method using MobileNetV3. As MobileNetV3Large is pre-trained model on ImageNet, we freeze upper layers and just train lower layers in model using our Indian food image dataset. Here in figure 10, you can see by freezing upper layers non-trainable parameters are comes around 4.2 million and trainable parameters are just 0.3 million. So, our system will become fast because of less no. of trainable parameters

In Transfer Learning with MobileNetV3 we use following parameters:

Image Size: 224x224

Epoch: 100

Initial learning rate:0.0001

Batch size: 32

After using this method, we got Test accuracy around 90%. Here in Figure 11, we can see sample results of prediction, Train and validation accuracy and loss graph shown in Figure 12. Based on Confusion Matrix showed in Figure 13, we calculated Precision, Recall and F1-score for each class. Here for calculation these we used classification_report class of sklearn library. Classification report is shown in Figure 14.

Comparison of Accuracy for both method is shown in Table 1.

Conclusion

In this paper, we proposed Transfer Learning using MobileNetV3 on Indian food image dataset for classification and recognition. We got higher accuracy of the system up to 90% with even a smaller dataset. We almost got 43% more accuracy in model where we used Transfer Learning with

MobileNetV3 even in less time and with less computational power. Based on this we can say that really transfer learning with pre-trained model like MobileNetV3.Here we train our model with almost 3000 different food images and we feel that if we train these models with more bigger size dataset then system's accuracy may increase. Even Data Augmentation techniques can be applied here for increase system's accuracy. As MobileNetV3 is specially designed for running on less computation powerful platforms like mobile, tablet, smartwatch etc., proposed model can be used for building dietary system, calorie estimation application and healthcare applications for Indian people.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

For this research work all authors' have equally contributed in Conceptualization, methodology, validation, resources, writing original draft preparation, writing review and editing.

References

[1] Resende Silva, B. V., & Cui, J. (2017). A Survey on Automated Food Monitoring and Dietary Management Systems. Journal of Health & Medical Informatics,8(3), 272

[2] Ayon, S. A., Mashrafi, C. Z., Yousuf, A. B., Hossain, F., & Hossain, M. I. (2021). FoodieCal: A Convolutional Neural Network Based Food Detection and Calorie Estimation System. 2021 National Computing Colleges Conference (NCCC),1-6.

[3] Ege, T., Ando, Y., Tanno, R., Shimoda, W., & Yanai, K. (2019). Image-Based Estimation of Real Food Size for Accurate Food Calorie Estimation. 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR).

[4] Aguilar, E., Nagarajan, B., Khatun, R., Bolaños, M., & Radeva, P. (2021). Uncertainty Modeling and Deep Learning Applied to Food Image Analysis. Biomedical Engineering Systems and Technologies, 3–16.

[5] Ciocca, G., Napoletano, P., & Schettini, R. (2017). Food Recognition: A New Dataset, Experiments, and Results. IEEE Journal of Biomedical and Health Informatics, 21(3),588–598.

[6] Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review, 53(8), 5455–5516.

[7] Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Modi, K., & Ghayvat, H. (2021). CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope. Electronics,10(20), 2470.

[8] Kong, J., Wang, H., Wang, X., Jin, X., Fang, X., & Lin, S. (2021). Multi-stream hybrid architecture based on cross-level fusion strategy for fine-grained crop species recognition in precision agriculture. Computers and Electronics in Agriculture,185, 106-134.

[9] Rajayogi, J. R., Manjunath, G., & Shobha, G. (2019). Indian Food Image Classification with Transfer Learning. 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), 1-4.

[10] Yadav, S., Alpana, & Chand, S. (2021). Automated Food image Classification using Deep Learning approach.7th International Conference on Advanced Computing and Communication Systems (ICACCS), 542-545.

[11] Zhou, L., Zhang, C., Liu, F., Qiu, Z., & He, Y. (2019). Application of Deep Learning in Food: A Review. Comprehensive Reviews in Food Science and Food Safety,18(6),1793–1811.

[12] Farinella, G. M., Moltisanti, M., & Battiato, S. (2014). Classifying food images represented as Bag of Textons. 2014 IEEE International Conference on Image Processing (ICIP), 5212-5216.

[13] 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. Information Sciences ,402,50–68.

[14] Hnoohom, N., & Yuenyong, S. (2018). Thai fast food image classification using deep learning. 2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON), 116-119.

[15] Sahoo, D., Hao, W., Ke, S., Xiongwei, W., Le, H., Achananuparp, P., Lim, E.-P., & Hoi, S. C. H. (2019). FoodAI. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2260-2268.

[16] Xu, B., He, X., & Qu, Z. (2021). Asian Food Image Classification Based on Deep Learning. Journal of Computer and Communications,09(3),10–28.

[17] Mezgec, S., Eftimov, T., Bucher, T., & Koroušić Seljak, B. (2018). Mixed deep learning and natural language processing method for fake-food image recognition and standardization to help automated dietary assessment. Public Health Nutrition, 1-10.

[18] 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.

[19] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET), 1-6.

[20] Sae-Lim, W., Wettayaprasit, W., & Aiyarak, P. (2019). Convolutional Neural Networks Using MobileNet for Skin Lesion Classification.16th International Joint Conference on Computer Science and Software Engineering (JCSSE), 242-247.

[21] Classification," In: *Proc. of 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 242-247, 2019,