

Indian Food Image Classification and Recognition with Transfer Learning Technique using MobileNetV3 and Data Augmentation[†]

Jigar Patel ^{1*}, Kirit Modi ²

¹ Research Scholar, Parul University, Vadodara, India; jigarsharp@gmail.com

² Professor, Sankalchand Patel University, Visnagar, India; kiritmodi@gmail.com

* Correspondence: jigarsharp@gmail.com; Tel.: (+91-9662032881)

[†] Presented at the The 4th International Electronic Conference on Applied Sciences, 27 Oct–10 Nov 2023;

Available online: <https://asec2023.sciforum.net/>

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 recognition 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. In this paper, we use CNN as our base model to build our system, which gives an accuracy of 85% 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 93.3%. Furthermore, we applied data augmentation techniques in pre-processing phase and we train our model using transfer learning with MobileNetV3 and we got an accuracy of up to 95.3%. So, the accuracy of the model increases by applying the data augmentation technique on top of transfer learning.

Keywords: Indian Food dataset; Mobilenetv3; Data Augmentation; CNN; Transfer Learning; Deep Learning

1. Introduction

In recent years, there has been an increase in interest and research in the field of computer vision as a result of its wide range of applicability in various domains. Due to its potential impact on health monitoring, dietary assessment, and culinary recommendation systems, food recognition has attracted a lot of attention among these applications. The ability to accurately identify and categorise different types of food could be used for a wide range of things, such as dietary analysis, restaurant recommendation systems, and culture preservation. Indian food is known for its complex and culturally important culinary history. It has a wide range of flavours, ingredients, and regional differences that set it apart from other cuisines around the world [1]. The conventional methods employed for food classification have faced challenges in effectively categorising Indian cuisine due to its extensive diversity and nuanced nature of dishes. The emergence of deep learning and transfer learning has brought in a new era in the field of image classification [2]. These advancements have made it possible to extract hierarchical features from large datasets, hence enhancing the accuracy and effectiveness of the classification process. The MobileNetV3 architecture, which is widely known for its high efficiency and adaptability,

Citation: To be added by editorial staff during production.

Academic Editor: Alessandro Bruno

Published: date



Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

has demonstrated considerable potential in a range of image-related activities [3]. Indian cuisine comprises a diverse range of regional variations, each characterized by distinct ingredients, culinary methods, and presentations. Differentiating among these varied cuisines poses a significant problem that requires models capable of capturing both overarching characteristics and minor intricacies [4]. The utilization of pre-trained models on extensive datasets as a foundation for fine-tuning, known as transfer learning, has emerged as an effective approach for tackling these difficulties.

MobileNetV3 is renowned for its efficient architecture and exceptional performance[5], making it a highly suitable choice for transfer learning in the domain of Indian food classification. By leveraging transfer learning using the MobileNetV3 architecture, the model is able to efficiently acquire and apply its pre-existing knowledge, while also adjusting to the specific complexities and nuances inherent in Indian cuisine. Furthermore, the incorporation of data augmentation techniques is crucial in enhancing the model's resilience and ability to generalize by intentionally introducing variations into the training dataset.

The primary objective of this research work is to investigate the utilization of transfer learning techniques, specifically employing the MobileNetV3 architecture, for the purpose of classifying Indian food. Additionally, the study incorporates the integration of data augmentation methods to enhance the performance of the classification model.

2. Related Work

Rajayogi [6] proposed the implementation of a Convolutional Neural Network (CNN) to categorize photographs of Indian cuisine into their appropriate classes in their research. The methodology outlined in this study was trained and subsequently used to a dataset featuring Indian cuisine. The dataset consisted of 20 unique classes, each including 500 photos. Consequently, this methodology enabled the successful deployment of the InceptionV3, VGG16, VGG19, and ResNet methods. The experimental results provide evidence supporting the superiority of the proposed methodology, as it achieved an accuracy of 87.9% and a loss rate of roughly 0.5893, surpassing competing algorithms.

VijayaKumari G[7] developed a food recognition model employing transfer learning methods to accurately classify diverse food goods into their respective categories. The model employed the EfficientNetB0 architecture, a transfer learning method, to classify a total of 101 unique food categories. The resulting accuracy achieved by the model was 80%.

The study conducted by Reddy [8] introduced a computerized system that employs deep learning methodologies for the purpose of classifying diverse food categories (citation). The categorization of food images was conducted utilizing the VGG-16 and SqueezeNet models. The accuracy attained using SqueezeNet and VGG-16 was 77.20% and 85.07% respectively.

In order to achieve the objective of food classification, S. Yadav [9] proposed the utilization of the MobileNetV2 architecture in conjunction with a support vector machine. The findings of the simulation indicate that the classification accuracies of the Conv_1, out_relu, and Conv_1_bn layers of the MobileNetV2 model, when categorized using a Support Vector Machine, were found to be 84.0%, 87.27%, and 83.60% respectively.

3. Methodology

In the present research, the CNN [10] and transfer learning with MobileNetV3[11] is utilized. The proposed system architecture is illustrated in Figure 1.

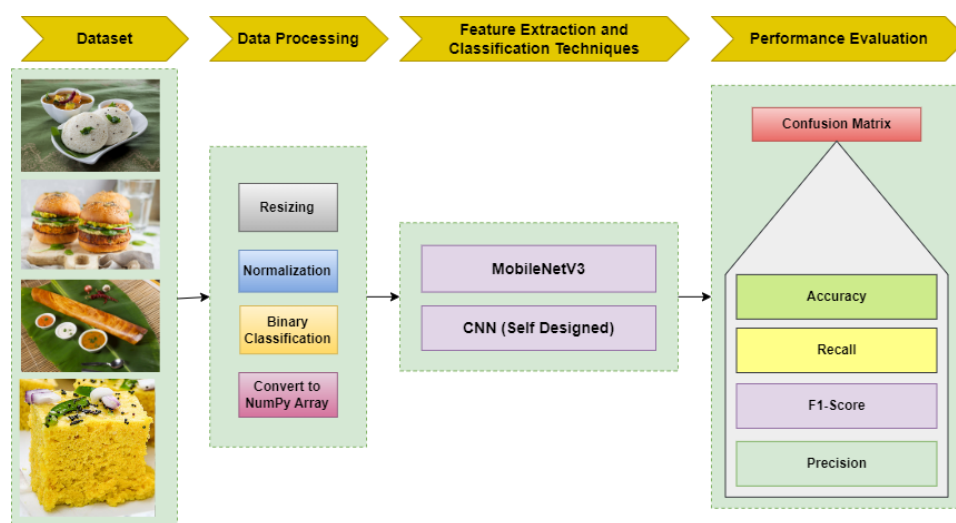


Figure 1. Indian Food Classification System Architecture.

3.1. Indian Food Dataset

The "9_Indian_food dataset"[12] was utilized in this study. The collection comprises a total of around 2700 photos. The dataset has a substantial quantity of images depicting Indian cuisine, encompassing over nine distinct categories. Each category within the collection has a substantial number of images, Almost 300 images per category. This is a balanced dataset. In order to mitigate the issue of overfitting, a portion equivalent to 20% of the training data is allocated for the purpose of validation.

3.3.1. Data Augmentation

The process of image data augmentation encompasses the application of diverse changes to images with the purpose of generating novel training instances that possess similarities to the original images while exhibiting minor variations [13]. This approach enhances the resilience and adaptability of machine learning models, specifically deep learning models, in the context of image data. Here we have limited data for training purpose so for the scope of enhancing performance of system we applied following data augmentation techniques: rotation image to 20-degree and Horizontal flip.

Upon acquiring the dataset, we proceeded to do pre-processing on the images contained inside the dataset. The utilization of preprocessing techniques on images holds the capacity to greatly improve their quality. Normalization, contrast enhancement, and image resizing are considered to be highly effective preprocessing techniques in the field of image processing. Also, as part of the conversion process, the images were rescaled by dividing each pixel value by 255.

3.2. Feature Extraction and classification

For feature extraction and classification from the images, here we used Convolutional Network method. We used CNN model (Self-Designed) and transfer learning method using MobileNetV3.

3.2.1. CNN (Self Designed)

The image input has a width of 224 pixels and a height of 224 pixels, as well as 3 channels. In each convolutional layer, the number of filters used in that layer doubles in the next layer. More specifically, the first layer has 32 filters and the last layer has 256 filters. After every convolution layer, a maxpooling layer with a pool size of 2 is added. This method reduces the number of dimensions in the data while keeping the most important traits. The result from the maxpooling layer is turned into a flattened representation by turning the 3D matrix of features into a vector. Then, this vector is sent to a fully

connected dense layer with 256 units and an activation function called the corrected linear unit (ReLU). The last layer of the neural network is a dense layer with as many units as there are food groups in the dataset, which is 9. In this layer, the 'SoftMax' activation function is used to group the images. Architecture of proposed CNN is shown in Figure 2.

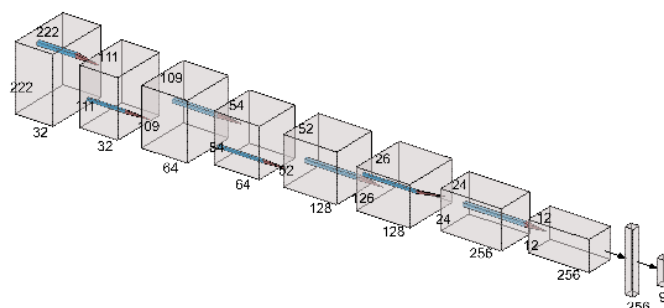


Figure 2. Architecture of Self-Designed CNN.

3.3.2. Transfer Learning with MobileNetV3

The MobileNetV3 was implemented in order to facilitate the process of transfer learning. MobileNetV3, which has undergone pre-training on the ImageNet dataset, has the potential to good outcomes in food categorization. Therefore, the utilization of MobileNetV3 was implemented for the objective of extracting features, then applying a dense layer consisting of 960 units and employing the ReLU activation function. Following that, the images underwent categorization using a dense layer. The number of units in the last layer is equivalent to the total number of food categories in the dataset, which was 9 in this particular case. Here we train this model with our Indian food image dataset and take weight reference to "ImageNet". Model Summary is shown in Table 1.

Table 1. Model summary using Transfer Learning with MobileNetV3.

Layer (type)	Output Shape	Param #
MobileNetV3Large (Functional)	(None, 960)	2996352
Flatten (Flatten)	(None, 960)	0
dense (Dense)	(None, 960)	922560
dropout (Dropout)	(None, 960)	0
dense_1 (Dense)	(None, 9)	8649

Total params: 3,927,561.

Trainable params: 3,903,161.

Non-trainable params: 24,400.

3.3. Evaluation

The scientific community has widely adopted many criteria to assess the effectiveness of the classification system. The evaluation of the research's effectiveness is performed by employing a confusion matrix, which encompasses crucial variables including true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN). Validity measures such as Accuracy, Recall, F1-score, and Precision can be calculated with these characteristics.

4. Implementation and Results

The experiments undertaken in this study used Kaggle.com, an online platform specifically built for the purpose of deep learning. Following Hyper-parameters used for the experiments.

Image Size: 224x224

Epoch: 50 for CNN (Self Designed) & 10 for (Transfer Learning with MobileNet)

Optimizer: Adam, Loss Function: categorical_crossentropy, Batch size: 32

The testing accuracy was determined to be 87% for CNN(Self Designed) Model. The testing accuracy achieved a value of 93.5% when evaluated on the testing dataset with transfer learning using pre-trained MobileNetV3 model and it further improved to 95.3% when data Augmentation applied on Training dataset. Comparative results are shown in Figure 3. Proposed method’s result comparison with past related work is shown in Figure 4.

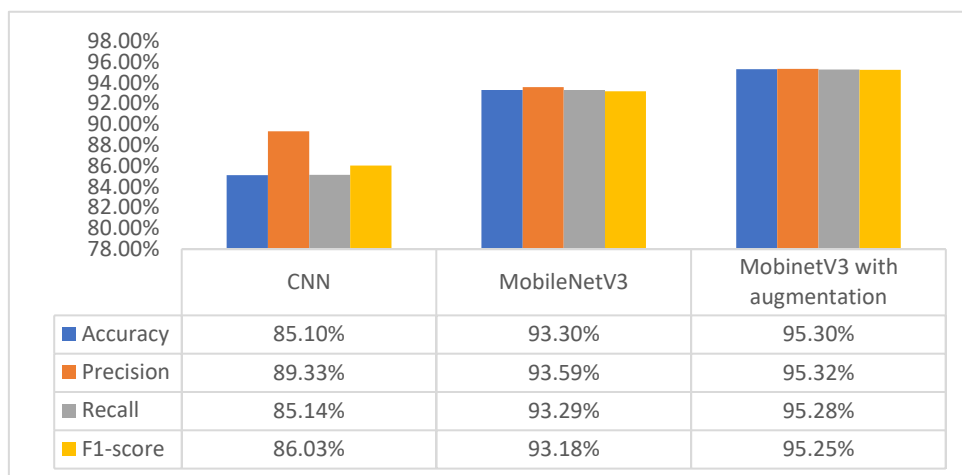


Figure 3. Result Comparisons of all models.

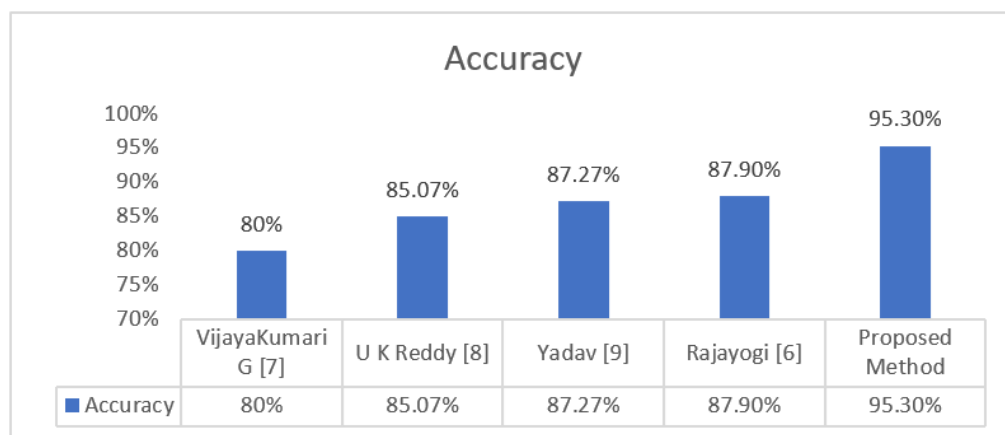


Figure 4. Result Comparisons with Past Related Work.

5. Conclusion

In this study, we presented Transfer Learning for classifying and recognising Indian food image datasets using MobileNetV3. With even fewer epochs than CNN (Self Designed Model), we achieved a better accuracy of the system, up to 93.3%. Even in a shorter amount of time and with less computational resources, we nearly achieve an 8% increase in accuracy in the model where we applied Transfer Learning with MobileNetV3. Based on this, we may conclude that pre-trained, modern models like MobileNetV3 effectively transfer learning. For performance improvement we used data augmentation technique on our training data then system accuracy increased to 95.3%. The proposed model can be utilised to construct dietary systems, calorie estimation applications, and healthcare applications for Indian people because MobileNetV3 is specifically built to function on less computation-powerful platforms like smartphone, tablet, smart watch, etc.

Author Contributions: Conceptualization, methodology, software, validation writing, visualization- Jigar Patel.; supervision, Kirit Modi

Funding: This research received no external funding.

Institutional Review Board Statement: “Not applicable” for studies not involving humans or animals.

Informed Consent Statement: “Not applicable” for studies not involving humans.

Conflicts of Interest: Declare conflicts of interest or state “The authors declare no conflict of interest.”

Data Availability Statement: Dataset is available on: <https://www.kaggle.com/datasets/jigarsharp/indian-food-9-class/> and Trained Proposed model is available on: https://github.com/jigarsharp/Indian_food_classification.git

References

1. Sadler CR, Grassby T, Hart K, Raats M, Sokolović M, Timotijević L. Processed food classification: Conceptualisation and challenges. *Trends in Food Science & Technology*. 2021 Jun 1;112:149-62.
2. Islam MT, Siddique BN, Rahman S, Jabid T. Food image classification with convolutional neural network. In 2018 international conference on intelligent informatics and biomedical sciences (ICIIBMS) 2018 Oct 21 (Vol. 3, pp. 257-262). IEEE.
3. Liu Y, Li Y, Yi X, Hu Z, Zhang H, Liu Y. Lightweight ViT model for micro-expression recognition enhanced by transfer learning. *Frontiers in Neurorobotics*. 2022 Jun 30;16:922761.
4. Manjunath swamy BE, Shreyas G, Tejaraj MP, Shafaq AF. Indian Food Image Classification with Transfer Learning. In *Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 1* 2022 Nov 1 (pp. 91-100). Singapore: Springer Nature Singapore.
5. Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*. 2021 Jun 10.
6. Rajayogi JR, Manjunath G, Shobha G. Indian food image classification with transfer learning. In *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS) 2019 Dec 20* (pp. 1-4). IEEE.
7. VijayaKumari G, Vutkur P, Vishwanath P. Food classification using transfer learning technique. *Global Transitions Proceedings*. 2022 Jun 1;3(1):225-9.
8. Reddy KU, Swathi S, Rao MS. Deep Learning Technique for Automatically Classifying Food Images. *Mathematical Statistician and Engineering Applications*. 2022 Sep 10;71(4):2362-70.
9. Yadav, S. and Chand, S. Automated food image classification using deep learning approach. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 542-545)*. IEEE.
10. Zhang W, Wu J, Yang Y. Wi-HSNN: A subnetwork-based encoding structure for dimension reduction and food classification via harnessing multi-CNN model high-level features. *Neurocomputing*. 2020 Nov 13;414:57-66.
11. Qian S, Ning C, Hu Y. MobileNetV3 for image classification. In *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE) 2021 Mar 26* (pp. 490-497). IEEE.
12. <https://www.kaggle.com/datasets/jigarsharp/indian-food-9-class> (Accessed on: 25 June 2023)
13. Phiphiphatphaisit S, Surinta O. Food image classification with improved MobileNet architecture and data augmentation. In *Proceedings of the 3rd International Conference on Information Science and Systems 2020 Mar 19* (pp. 51-56).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.