International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Traditional Indian Food Classification Using Shallow Convolutional Neural Network

Bhoomi Shah[1*], Pratik Kanani[2], Poonam Joshi[3], Gayatri Pandya[4], Dhanashree Kulkarni[5], Nilesh Patil[6], Lakshmi Kurup[7]

Submitted: 25/11/2023 **Revised**: 03/01/2024 **Accepted**: 10/01/2024

Abstract: Food classification is a difficult challenge because there are many distinct categories, different foods look quite similar to one another, and there aren't enough datasets to train cutting-edge deep models. It will make improvements in computer vision models and datasets to test these models to solve this issue. This paper introduces Food10, a dataset of 10 Traditional Indian food categories with 5000 photos gathered from the web and concentrates on the second component of this study. We employ 4000 photos as a training set and 1000 images with human-validated labels for testing and validation. In the current study, we describe the steps involved in producing this dataset and offer pertinent baselines with deep learning models used in the Food10 dataset. Indian food is naturally oily and sweet hence contains lot of calories. Managing calorie intake is crucial for preventing obesity and mitigating the risk of numerous other diseases. The analysis of food images and calorie estimation can serve as a valuable tool to assist individuals in adhering to a healthy diet. Moreover, it can be beneficial for the general population in maintaining their everyday dietary choices. To calculate the calories food classification is the first step. In this research, a novel model was introduced with the aim of achieving enhanced accuracy and efficiency in the identification of Indian food, surpassing the performance of existing methodologies. The conventional models, such as AlexNet, VGG, and GoogleNet, were trained alongside the proposed model. On FOOD10 dataset, the proposed model Shallow Convolutional Neural Network (SCNN) gives a remarkable result with an average accuracy of 96%.

Keywords: *Convolution Neural Network, Dataset, Fine-tuning, Image classification, Transfer Learning, VGG16*

1. Introduction

People all over the world are becoming more health conscious as the world becomes more competitive and dynamic. The overweight problem is spreading rapidly around the world. It is the sole cause of some chronic conditions, such as diabetes, blood pressure, obesity, and a few cancers. Precise assessment of dietary calorie intake is important for assessing the viability of the weight loss process, which is why food calorie measurement is gaining popularity in today's society. To improve a person's eating habits, an evaluation of their dietary proclivities is required, which has previously been done physically, regularly, using self-detailing methods.

1 The Maharaja Sayajirao University of Baroda, Vadodara, India
ORCID ID: 0000-0001-6881-2919
2 Dwarkadas J. Sanghvi College of Engineering, Mumbai, India
ORCID ID: 0000-0002-6848-2507
3 Thakur College of Engineering and Technology, Mumbai, India
ORCID ID: 0009-0002-4671-6162
4 Dwarkadas J. Sanghvi College of Engineering, Mumbai, India
ORCID ID: 0009-0000-7135-0559
5 Mukesh Patel School of Technology Management and Engineering, NMIMS University, Mumbai, India
ORCID ID: 0009-0007-3643-6841
6 Dwarkadas J. Sanghvi College of Engineering, Mumbai, India
ORCID ID: 0000-0001-8335-4426
7 Dwarkadas J. Sanghvi College of Engineering, Mumbai, India
ORCID ID: 0000-0003-1579-2242
** Corresponding Author Email: bhoomi.shah.mca@gmail.com*

To achieve this, a dataset of 5000 photos is presented in this study that includes a variety of food positions captured by various cameras under various lighting conditions. In this research, we developed a novel strategy for more precise Deep Learning-based Indian food recognition. With the least amount of time and computational resources possible, the goal is to determine whether transfer learning could improve accuracy results.

The remainder of this paper is organized as follows. The methodology used in this work is detailed in Section 2, followed by descriptions of the experimental setup and datasets in Section 3 and a preliminary set of experiments and results in Section 4. Section 5 concludes the paper and suggests additional research based on the findings.

2. Related Study

The classification of food images has already been the subject of several research studies. one of the emerging applications of deep learning improvements is food image classification.

A study focused on the classification of images depicting Indonesian traditional foods utilizing classical machine learning techniques [1]. In this research, the authors undertook the extraction of distinctive features, specifically concentrating on color and texture attributes, from the images. These identified features were subsequently harnessed to facilitate the classification process, employing a selection of five well-established classical machine learning models as detailed in the study.

In reference [2], the primary emphasis was directed towards the application of computer vision techniques for the extraction of pertinent features from an assemblage of food-related images. These extracted features were subsequently employed as input for training machine learning models, thereby enabling the classification of distinct food items present within the images. The investigation encompassed a series of experiments involving 30 distinct food categories, each containing a corpus of 100 images.

In recent times, deep learning frameworks, particularly Convolutional Neural Networks (CNNs), have gained significant prominence for their widespread application in the realm of image classification tasks across diverse domains [3]–[5].

In the context of reference [6], the researchers employed the MLC-41 dataset, meticulously composed of 41 distinct food labels, each accompanied by a curated collection of 100 images. The study systematically distinguished between various feature sets, feature selection methods, and classifiers employed within their DeepFood framework. Notably, the study highlighted the superiority, in terms of accuracy, of ResNet deep feature sets, the Information Gain (IG) technique for feature selection, and the SMO classifier for the recognition of food ingredients.

In reference [7], the study was focused on the classification of images depicting Indian cuisine, with the aim of assigning them to their appropriate categories. This classification task was achieved through the implementation of transfer learning using the Inception v3 model. The research also encompassed a thorough comparison of multiple models, considering aspects such as accuracy and validation loss, to provide a comprehensive evaluation.

Similarly, in reference [8], the researchers assembled a dataset consisting of 3960 images showcasing Thai cuisine. They proceeded to fine-tune their model utilizing Google's iterations of the Inception V3 architecture, implementing a learning rate of 0.001 and conducting training over 4,000 iterations. This meticulous approach led them to achieve an impressive average accuracy of 88.33% on their specific dataset.

Furthermore, documented in reference [9], the researchers employed the Inception v3 model on the FOOD-101 dataset sourced from Kaggle, encompassing a remarkable 101,000 images categorized into 101 classes. Across this dataset, the researchers meticulously evaluated diverse models including SVM, neural networks, RFDC-based strategies, Resnet18, and CNN. After thorough experimentation, the study advocates for the utilization of the CNN approach, yielding the most noteworthy accuracy of 83.97% in comparison to alternative methodologies.

Described in reference [10], the researchers pursued a distinct avenue by crafting a fresh deep learning framework termed NutriNet. This innovation involved a tailored adaptation of the well-known AlexNet architecture. Notably, the approach commenced with an augmentation of the input image dimensions from 256x256 to 512x512 pixels. Moreover, a supplementary convolutional layer was seamlessly integrated, augmenting the architecture in comparison to the original AlexNet. This strategic enhancement resulted in the development of a customized model boasting an impressive accuracy of 86.72%.

Elaborated in reference [11], the study introduced a sophisticated deep learning framework comprising a convolutional neural network designed for the purpose of categorizing various food items into distinct classes. Notably, the study advocated for the adoption of the Inception v3 and v4 models, which exhibited superior accuracy in accurately identifying these food items.

Elucidated in reference [12], the study made use of the VIREO-172 dataset, comprising an extensive compilation of 110,241 images representative of 172 distinct classes, each averaging around 641 images per class. In pursuit of performance evaluation, the researchers trained three distinct models: DenseFood, DenseNet121, and ResNet50, utilizing the dataset above. Remarkably, the DenseFood model demonstrated notable prowess by achieving an accuracy rate of 81.23%, surpassing its counterparts and emerging as the most effective model in the comparative analysis.

3. Methodology

In the context of this study, a dataset has been created, comprising images representing 10 distinct categories of Indian food items. Given that larger images tend to consume more memory and computational resources, necessitating a larger neural network, it has been deemed essential to undergo a series of preprocessing steps prior to employing these images for subsequent image classification tasks. Furthermore, augmentation techniques have been judiciously applied to enhance the model's ability to generalize effectively.

The subsequent phase of this approach involves feature extraction, facilitated by Convolutional Neural Network (CNN) layers, encompassing input layers, convolutional layers, pooling layers, and fully connected layers. These CNN layers play a pivotal role in learning hierarchical features from the preprocessed images.

The SCNN model has been implemented using the Python programming language. The model has been implemented within the TensorFlow environment. Keras is widely regarded as one of the premier libraries for convolutional neural networks (CNNs), and it is seamlessly integrated within the TensorFlow environment. Several libraries, including pandas, NumPy, OpenCV, matplotlib, and scikit, were employed in the implementation of the convolution-based Night-CNN. For feature extraction 11x11, 3x3, and 1x1 filters within the convolutional layers has been used by the SCNN model. Following each convolutional layer, a max-pooling layer is applied to reduce the dimensionality of the images.

As illustrated in Table 1, various CNN architectures like AlexNet, VGG, and GoogleNet exhibit different layer configurations and parameter sizes.

Table 1. Comparison of existing models with SCNN based on parameters and layers

Year	Name of Architecture	No. of Layers	Parameters	Limitation
2012	Alexnet	8	60M	Utilizing high-resolution images can pose challenges. Suffering from overfitting problem
2014	VGG	16	134M	Takes more time to run the model and occupies more disk space. Suffering from vanishing gradient problem and complex than Alexnet.
2014	GoogleNet	22	360M	Requires more time for training the model hence very expensive.
2023	SCNN	5	22M	Identify the Indian food accurately.

The standard architecture model employs convolutional layers, max-pooling layers, and fully connected layers for food identification. However, this model uses a considerable number of layers for food identification, resulting in a significant increase in parameter size. Consequently, the model demands more execution time and substantial computational resources.

Therefore, the SCNN model was designed with a reduced number of layers, resulting in fewer parameters. This design choice aims to reduce computation time and resource requirements. In this research, we compared the results obtained with the SCNN model to those of the standard model.

3.1 Preprocessing

Image preprocessing constitutes a pivotal phase in the pipeline, encompassing a series of operations designed to impart specific transformations to the images before commencing model training and inference. In this research paper, with the overarching aim of enhancing image features to facilitate a more robust deep learning approach, these essential methods have been categorized into two distinct segments: preprocessing and augmentation.

Within the domain of preprocessing, several crucial steps have been undertaken. Initially, a fundamental task involved standardizing the dimensions of the images within the dataset. Given that the original images exhibited variations in size, they were uniformly resized to dimensions of 224x224 pixels. This resizing operation served the dual purpose of streamlining computational requirements and ensuring a consistent input size for subsequent processing.

Additionally, the images in the dataset presented color information in the form of RGB coefficients, with pixel values spanning the range from 0 to 255. To expedite computational processing while preserving data integrity, a rescaling operation was performed. This operation entailed dividing the pixel values of the entire dataset by a factor of 255.0, effectively transforming the pixel values into a normalized range of [0, 1].

Fig. 1 shows the sample images from the dataset.

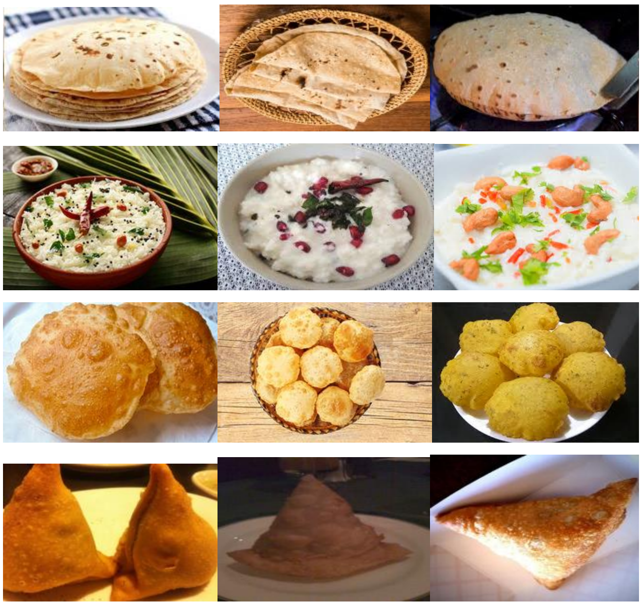

Fig 1. Sample images from the dataset Food10

4. Experiments and Results

The entire research described in this paper was executed on a robust computational platform, harnessing the capabilities of a multicore processor paired with a Tesla K80 GPU boasting an impressive 12 GB of memory. The central processing unit (CPU) of this system operated at a clock speed of 2.3 GHz, providing the necessary computational power to handle the intricacies of deep learning tasks.

For the implementation of the deep learning components, the authors leveraged the popular Keras framework, seamlessly integrated with Tensorflow version 2.0. These frameworks, coupled with Python version 3.6, served as the primary tools for the development and execution of the deep learning models and associated algorithms. The default configuration of optimizer and learning rate in Keras has been considered i.e., RMSprop optimizer with a learning rate of 0.002. Categorical cross-entropy loss is applied in this case also to measure the performance of the model. The hyperparameters of the model is defined in Table 2.

Table 2. Hyperparameter of SCNN Model

Hyperparameter	Description
No. of Convolution layer	4
No. of Fully Connected layer	1
Activation Function	LeakyRelu
Learning rate	0.002
Dropout rate	0.6
Batch size	32
Number of epochs	500

After that, the combined model is compiled using LeakyRelu optimizer with a learning rate of 0.002 and categorical cross-entropy as a loss function.

The architecture of SCNN model has been shown in Fig 2.

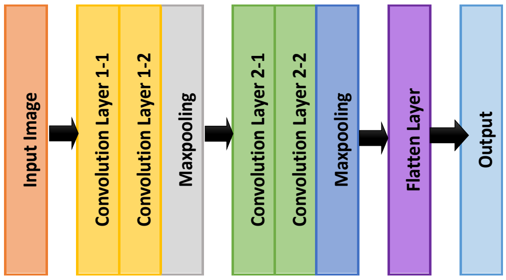

Fig 2. Architecture of SCNN

Conventional CNN models such as VGG, Alexnet and Googlenet have also been implemented for FOOD10 dataset. The accuracy of the SCNN with existing models has been compared Table 3.

Table 3. Comparison of existing models with SCNN on FOOD10 dataset

Model	Accuracy(%)	Error Rate(%)
Alexnet	80	42.10
VGG	76	40.12
GoogleNet	85	15.79
SCNN	96	9.84

It is observed from Table 3 that SCNN outperforms other models in terms of both accuracy and error rate. Due to less number of layers it will also takes less time to run the model.

Table 4 shows the class-wise performance of the SCNN model with splendid f1 scores and an accuracy of 96%. It is evident from the figure that, this approach performed pretty well in the case of classifying Indian food images. TABLE I CLASS-WISE PERFORMANCE OF THE FINE-TUNED VGG16 Classes Accuracy Precision Sensitivity F1-Score Biryani 0.98 0.95 0.95 0.95 Chapati 1.00 1.00 1.00 Curd rice 0.97 0.94 0.96 Dosa 0.96 1.00 0.98 Idly 0.96 1.00 0.98 Pongal 1.00 0.95 0.97 Puri 1.00 1.00 1.00 Indian food images. ‘Salad’, ‘Sambar’ and ‘Samosa’ are classified perfectly.

Table 4. CLASS-WISE PERFORMANCE OF THE SCNN

Classes	Accuracy	Precision	Sensitivity	F1-Score
Biryani	0.96	0.95	0.95	0.95
Chapati		1.00	1.00	1.00
Curd rice		0.97	0.94	0.96
Dosa		0.96	1.00	0.98
Idly		0.96	1.00	0.98
Pongal		1.00	0.95	0.97
Puri		1.00	1.00	1.00
Salad		0.85	0.69	0.78
Sambar		0.65	0.69	0.65
Samosa		0.78	0.89	0.85

Fig 3 shows the graphical representation of training and testing accuracy.

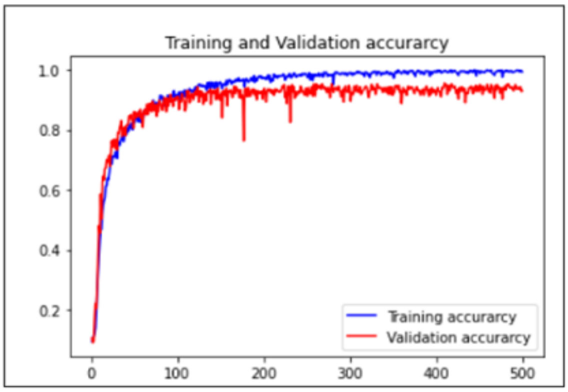

Fig 3. Training and validation accuracy of SCNN for 500 epochs

5. Conclusion

This paper encompasses two significant contributions centered around the utilization of Convolutional Neural Network (CNN) architecture for the classification of Indian foods. A primary focus of this research effort has been the creation of a novel dataset FOOD10, consisting of ten distinct food classes, namely Biryani, Chapati, Curd rice, Dosa, Idly, Pongal, Puri, Salad, Sambar and Samosa. The genesis of this dataset was prompted by the scarcity of publicly available image data on the internet, necessitating extensive efforts to gather sufficient and pertinent information to ensure computational robustness.

As a culmination of these data collection endeavors, a noteworthy total of 5000 images were meticulously assembled. This dataset is a substantial and valuable contribution to the domain of Indian food image classification.

Furthermore, to classify the FOOD10 dataset, an entirely new CNN model named SCNN has been devised, achieving a commendable accuracy rate of 96% when tested on independent data. It is important to note that this SCNN model gives highest accuracy as compared to all the existing models with less number of parameters.

These experimental results underscore the viability and effectiveness of the SCNN approach for the classification of Indian food. Additionally, the versatility of the SCNN model suggests its potential applicability in diverse food classification scenarios. Looking ahead, future research endeavors may expand into the realm of food ingredient classification for dietary purposes. With an expanded dataset, advanced data augmentation techniques, and superior regularization methods, the SCNN model promises to yield even more impressive results. Furthermore, the possibility of fine-tuning various other pre-trained models using the novel dataset represents an avenue for future exploration and enhancement.

Another future scope is to calculate calories from the food images which can be helpful to people to maintain their day-to-day diet. Also estimating calorie from food images helps people to reduce the risk of serious health conditions like hypertension, chronic diseases, and heart disease.

Author contributions

Bhoomi Shah and Pratik Kanani contributed to the study conception and design. Material preparation, data collection and analysis were performed by all authors. The manuscript was written and revised by all authors on previous versions of the manuscript. All authors read and approved the final manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] P. K. Fahira, A. Wibisono, H. A. Wisesa, Z. P. Rahmadhani, P. Mursanto,and A. Nurhadiyatna, "Sumatra traditional food image classificationusing classical machine learning," in *3rd International Conferenceon Informatics and Computational Sciences (ICICoS)*. IEEE, 2019, pp.1–5.

[2] P. McAllister, H. Zheng, R. Bond, and A. Moorhead, "Towards personalised training of machine learning algorithms for food image classification using a smartphone camera," in *International Conference onUbiquitous Computing and Ambient Intelligence.* Springer, 2016, pp.178–190.

[3] M. S. Sarma, K. Deb, P. K. Dhar, and T. Koshiba, "Traditional bangladeshi sports video classification using

deep learning method," *Applied Sciences*, vol. 11, no. 5, p. 2149, 2021.
[4] S. Deepak and P. Ameer, "Brain tumor classification using deep cnn features via transfer learning," *Computers in biology and medicine*, vol.111, p. 103345, 2019.
[5] R. Vaddi and P. Manoharan, "Hyperspectral image classification using cnn with spectral and spatial features integration," *Infrared Physics &Technology*, vol. 107, p. 103296, 2020.
[6] L. Pan, S. Pouyanfar, H. Chen, J. Qin, and S.-C. Chen, "Deepfood:Automatic multi-class classification of food ingredients using deep learning," in *IEEE 3rd international conference on collaboration and internet computing (CIC).* IEEE, 2017, pp. 181–189.
[7] J. Rajayogi, G. Manjunath, and G. Shobha, "Indian food image classification with transfer learning," in 4th International Conference on Computational Systems *and Information Technology for Sustainable Solution (CSITSS)*, vol. 4. IEEE, 2019, pp. 1–4.
[8] N. Hnoohom and S. Yuenyong, "Thai fast food image classification using learning," in *2018 International ECTI Northern Section Conference on electrical, electronics, computer and telecommunications engineering(ECTI-NCON)*. IEEE, 2018, pp. 116–119.
[9] D. J. Attokaren, I. G. Fernandes, A. Sriram, Y. S. Murthy, and S. G.Koolagudi, "Food classification from images using convolutional neuralnetworks," in *TENCON 2017-2017 IEEE Region 10 Conference*. IEEE,2017, pp. 2801–2806.
[10] S. Mezgec and B. K. Seljak, "Using deep learning for food and beverage image recognition," in *2019 IEEE International Conference on Big Data(Big Data). IEEE*, 2019, pp. 5149–5151.
[10] Z. Shen, A. Shehzad, S. Chen, H. Sun, and J. Liu, "Machine learning based approach on food recognition and nutrition estimation," *Procedia Computer Science*, vol. 174, pp. 448–453, 2020.
[12] A. Metwalli, W. Shen, and C. Q. Wu, "Food image recognition based on densely connected convolutional neural networks," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC).* IEEE, 2020, pp. 027–032