Indian Food Image Recognition with MobileNetV2

Priya N V[1], Preetam Kumari[2], Poorvika N[3], Sanjana R[4], Dr. Hema Jagadish[5]

[1-4]Student, Dept. of Information Science and Engineering, Bangalore Institute of Technology, Bengaluru, Karnataka, India
[5]Associate Professor, Dept. of Information Science and Engineering, Bangalore Institute of Technology, Bengaluru, Karnataka, India

---

Abstract - *Indian cuisine consists of a variety of regional dishes, due to which food image recognition is a necessity. Transfer Learning is widely used in multi-class image classification whereby a machine exploits the knowledge gained from a previous task to improve generalization about another. In this paper, we propose a custom-built CNN model and a transfer learning based MobileNetV2 model for the purpose of food recognition and classification. A calorie estimation algorithm based on image features and nutritional information is proposed, to estimate calories of the recognized food image. The dataset consists of 12 classes of Indian food images, with 100 images per class. After experimentation, it was found that the MobileNetV2 model outperformed the custom CNN model with an accuracy of 79.45%.*

***Key Words*: Convolutional neural network, Transfer Learning, MobileNetV2, Image Processing, Food recognition**

1.INTRODUCTION

The Indian cuisine consists of possibly hundreds of varieties of regional and traditional foods, due to which food image recognition becomes an essential task. The rising popularity of food blogging and image tagging pose requirements for a food recognition system customized to recognize regional foods. Due to the wide variety of Indian foods, the nutritional content varies largely. Not being conscious of our food intake can lead to diseases such as diabetes and obesity. A study reported that more than 650 million adults worldwide and 150 million adults in India are obese[1].

A few techniques that exist for multi-class image classification are SVM, KNN, and Artificial Neural Networks[4]. Transfer Learning technique has shown promising results in the field of image classification. Transfer learning is a deep learning technique where a model is trained to learn and store the knowledge from one problem and use the same model to other similar problems[2]. Convolution Neural Network is a deep learning technique that has gained popularity in image recognition tasks due to its high accuracy and robustness.

A study conducted by UNTWO suggested that in 2018 there were a record 1.4bn international tourist arrivals . A record number of tourists are choosing to travel to India for vacation , which leads to a requirement for a regional food classification system . Presently available systems are customized for international cuisines and mainstream foods.

This article proposes two methods to classify Indian food images: training a CNN from scratch and using a pre-trained MobileNetV2 model. A comparison of the above two models is made based on accuracy and loss. A calorie estimation algorithm based on image features is proposed to estimate the nutritional value of the recognized food. The food-20 dataset has been used to train the proposed models [6].

The organization of the paper is as follows: the related works are presented in Section 2. Section 3 deals with the proposed methodology followed by the experimental results, conclusion and future work in section 4 and section 5 respectively.

2. RELATED WORK

The application of CNN and transfer learning in image classification tasks such as food image recognition is gaining popularity due to its high accuracy , ability to deal with huge datasets and learn complex patterns. Most of the reviewed work used a very few classes and mainstream food items for the purpose of classification.

The work proposed by V Hemalatha Reddy et al . [5] designed a mobile application built on a dataset containing 20 classes using a custom trained 6 layer Convolutional Neural Network. Calories were estimated based on static nutritional values and input of quantity from users. The problem with this method is that the users may not know the quantity of the food hence leading to under or overestimation. Patanjali C et. al. [4] proposed a comparative study of indian food image classification using KNN and SVM. The study was performed on 5 classes on indian food images and experimental results showed SVM performed better.

Vishwanath C. Burkapalli et.al.[2] proposed a transfer learning based approach to identify and differentiate between under baked and over baked south indian food classes. The dataset was customly collected using smartphone cameras. The proposed model achieved a training accuracy of 93.2%.

ManpreetKaur Basantsingh Sardar et.al.[3]present a system for fruit recognition and its calorie estimation based on the shape, color and texture along with histogram of gradients and GLCM. With the help of nutritional look up tables these features are fed to multi SVM classifier. The dataset consists of 5 classes of fruit images captured using a mobile phone camera. The problem with this method is that volume based calorie estimation does not provide high accuracy.Hemraj Raikwar et.al.[7] proposed a system for calorie estimation from fast food images using Support Vector Machine.The model was trained on 5 classes of fast food. Histogram Of Oriented Gradients (HOG) method is used for calorie estimation of detected fast food image.The model achieves an accuracy of 72%.

Kiran Ambhore et.al.[8] proposed a system for food image recognition using KNN. The image colour, shape, size and texture features are extracted and it is given to the K-nearest neighbour (KNN) for recognizing the food and then the calorie value is measured with the help of a nutrition table. Shamay Jahan et.al.[9] proposes a custom CNN based approach to classify indian snack images. The dataset consists of 60,000 images of 10 classes of indian snack foods. The model achieved a test accuracy of 95% using backpropagation algorithm. Diksha Solanki et.al.[10] proposes a food image classification and calorie estimation system based on the Food-101 dataset. The dataset consists of 101 categories of mainstream food images and 3 classes have been selected for classification. A custom CNN model is used for the purpose of classifying food. Nutrionx dataset has been used to estimate calorie in food image[11]. The model achieved peak validation accuracy of 78.9%. The disadvantage with the proposed method is that only 3 classes have been used for the purpose of classification.

3. PROPOSED METHODOLOGY

The proposed work is implemented in Python using a Convolutional Neural Network (CNN) model and Transfer Learning[1][2][5]. The models were trained on 1000 images for 12 classes and then used to predict food class. A new input goes through stages of image processing like resizing and colour space conversion etc, before it is fed to the trained model. After comparing the features of the input with the features of each trained class, the output is predicted.

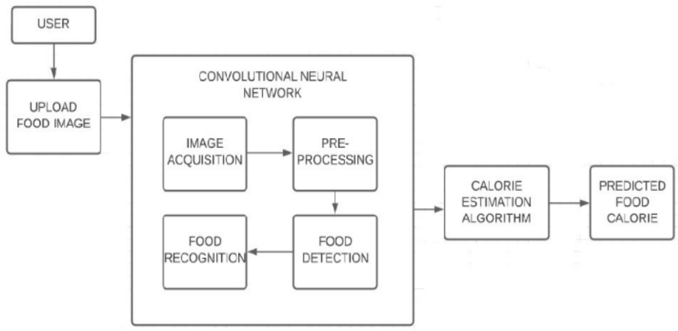


Fig -1: System Architecture

The User uploads a food image from the system using a GUI. The image goes through the stages of image pre-processing and is then passed on to the CNN model for classification. After the model recognizes the food image , a calorie estimation algorithm is used to detect the calories in the uploaded food item.

3.1 Dataset

Based on the original food-20 dataset with 20 indian food categories[6]. A subset of the twelve food categories[Dosa, Puri, Upma, Chapati, Samosa, Chaat, Idly, Biryani, Dhokla, Bisibelebath, Buternaan and Noodles] is used. The data consists of two main subfolders: training, testing. The training data consists of 65-70 images per class, and 30 validation images per class.

Fig -2: Sample Images of the Dataset

3.2 Pre-Processing

For the purpose of CNN and DNN modelling, the training images have been resized to 50x50[5]. The OpenCV module reads images in BGR format (Blue-Green-Red) and for the purpose of training the classifier, Image conversion to RGB has been done[4]. For the purpose of Transfer learning the images are resized to 256x256 and data augmentation techniques have been used.

3.3 Convolutional Neural Network(CNN)

The convolutional neural network (CNN) is a class of deep learning neural networks[5][9]. The model consists of

Convolution, MaxPool, Conv2D Layer, Dropout, Fully Connected layer and ReLu Function. Deep Neural Networks(DNN) have an input layer, an output layer and a few hidden layers between them. These networks not only have the ability to handle unstructured data, unlabeled data, but also non-linearity as well. TFLearn provides a model wrapper 'DNN' that can automatically perform neural network classifier tasks, such as training, prediction, save/restore. The proposed CNN Structure consists of:

Conv2D - The first Convolutional 2D layer consists of 32 kernels of 3x3. Takes an input of size 50x50x3 where 50x50 is the rescaled size of images from the dataset. RGB, the color aspect of the image is denoted by 3.

Convolutional Layer - A (50,50,3) input size is used by convolution function and this layer generates the feature maps through convolving the input data.

Pooling Layer - The second layer with a pool size of 2x2 is the max-pooling layer. For better feature extraction, these layers are repeated once again. Then, to get more filtered images for the fully connected layers, the kernel's size is increased from 32 to 64.

Fully Connected Layer - This is used to connect all neurons to a one layer as well as to another layer, which works on the basis of traditional multi-layer-preceptor(MLP) neural networks. Two fully connected layers are used next with 128 and 90 neurons respectively.

Dropout layer - To prevent overfitting, dropouts have been added in between the dense layers.

ReLu - Activation functions are mathematical equations that determine the output of a neural network model. All the convolutional 2D layers and the fully connected layers have an activation function of Rectified Linear Unit (ReLu).

The network is configured to output 12 values, one for each class in the classification task, and the softmax function is used to normalize the outputs. Adam optimization algorithm was used and the learning rate was configured as 0.0001.

3.4 Transfer Learning using MobileNetV2

Transfer Learning is a technique of CNN which uses feature extraction from a pre-trained model and then trains a classifier using the extracted features of the pre-trained model[2].

MobileNetV2 is a convolutional neural network architecture that seeks to perform well on mobile devices. It is based on an inverted residual structure where the residual connections are between the bottleneck layers.MobileNet has been pre-trained on the ImageNet database which contains more than a million images. It captures the edges, color and pattern in initial layers and complex patterns related to our task in the final layers.

The proposed work uses a mobilenet model for the purpose of classification.

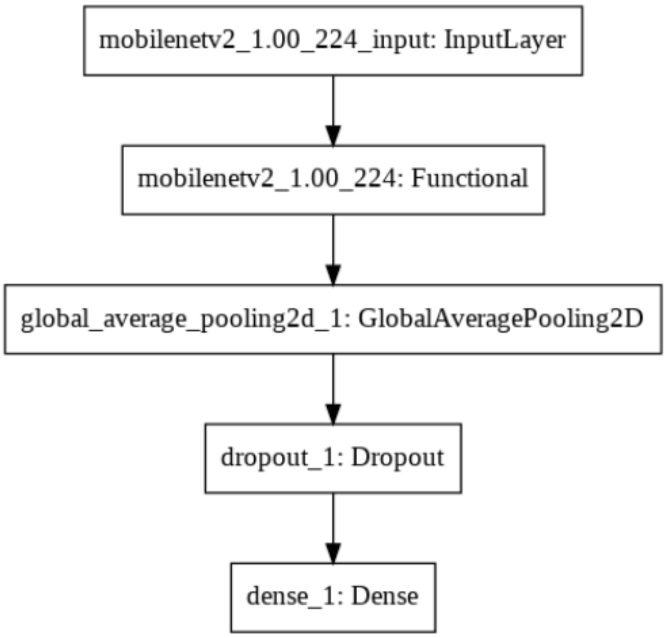


Fig -3:Visualization of proposed MobileNetV2 Architecture

The architecture of MobileNetV2 contains the initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers. The primary network (width multiplier 1, 224 × 224), has a computational cost of 300 million multiply-adds and uses 3.4 million parameters. Instead of adding fully connected layers on top of the feature maps, an average of each feature map is taken, and the resulting vector is fed directly into the softmax layer.

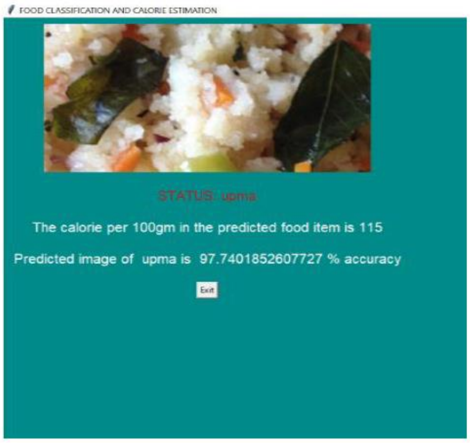
The input images are preprocessed to 256x256 size. Learning rate is kept at a minimum value to improve training at 0.0001. Dropout of 0.2 is used. The activation function used is softmax. The optimizer used is Adam optimizer. SparseCategoricalCrossEntropy is used as a parameter for loss values.

3.5 Calorie Estimation

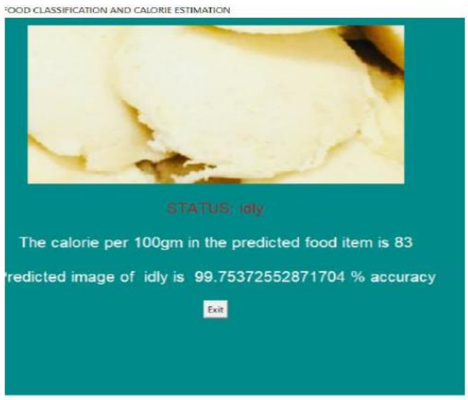
The image sequences are captured with BGR colour space by default. In BGR(Blue,Green,Red) Red occupies the least significant area, green the second and blue the third. The BGR colour space is converted to HSV (Hue Saturation Value) space.This is achieved by a function of the OpenCV module called cvtColor( ) using COLOR_BGR2HSV.The algorithm uses the concept of indexed images which is a direct mapping of pixel values to colormap values. The calories range of each food class is obtained from nutrition websites [11] and then used along with the HSV method to estimate the calories of the food image.

4. RESULTS

A graphic user interface is developed using Tkinter. The user uploads a food image from the system and the system then displays the food class and estimated calories.



(a)



(b)

Fig -4: Outputs showing predicted food class and calories for (a)Food Class Upma (b) Food Class Idly

The dataset was divided into training and test images.The model was trained on 70% of the images and was tested on 30% of the images.

Table -1: Accuracy Comparison of the models

MODEL	EPOCHS	ACCURACY
CNN	100	88.42 % Train accuracy 73% Test accuracy
MobileNetV2	50	89% Train accuracy 79.44% Test accuracy

The CNN model was trained for 50 epochs and achieved a test accuracy of 73%.

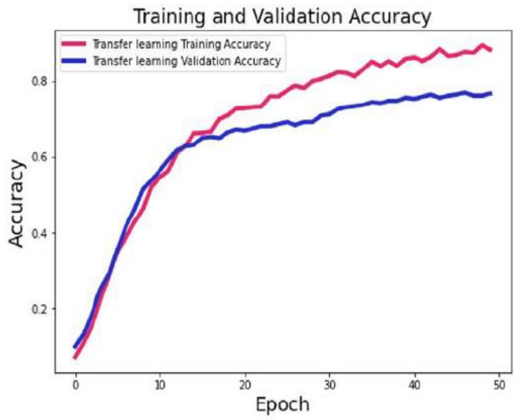


Fig -5 : Training and Validation Accuracy graph for MobileNetV2 model.

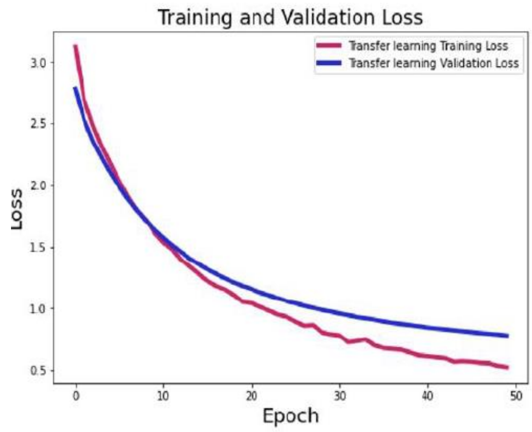


Fig -6 : Training and Validation Loss graph for MobileNetV2 model.

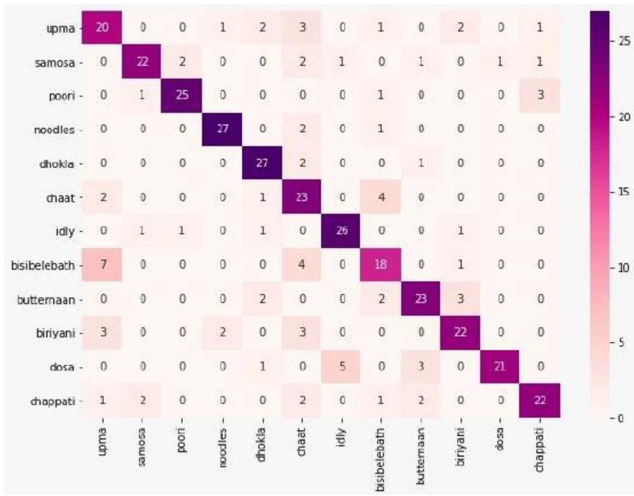


Fig -7 : Confusion Matrix

The MobileNetV2 model outperformed the CNN model and achieved an accuracy of 79.44% at 50 epochs. Fig -6 shows the loss(0.76) over training and validation set.

5. CONCLUSIONS

In this study, a comparison is made between conventional CNN model and MobileNetV2 model based on their accuracy in multi-class classification for indian food image dataset. It is seen that the MobileNetV2 model outperforms the CNN model in terms of accuracy. It can be concluded that when a large dataset is not available it is better to use Transfer Learning than Conventional CNN.

In future work, ingredient identification in the particular class of food can be obtained, calories estimation can be improved using volume estimation and other extracted features. A more sophisticated tool for image classification can be developed using more than 12 classes.

REFERENCES

[1] Rajeev Ahirvar, “Prevalence of obesity in India: A systematic review”, NCBI, 2019.

[2] Vishwanath C. Burkapalli, Priyadarshini C. Patil, “TRANSFER LEARNING: INCEPTION-V3 BASED CUSTOM CLASSIFICATION APPROACH FOR FOOD IMAGES”, ICTACT, 2020.

[3] ManpreetKaur Basantsingh Sardar, Dr. Sayyad D. Ajij, “Fruit Recognition and its Calorie Measurement: An Image Processing Approach,”International Journal Of Engineering And Computer Science, 2016.

[4] Pathanjali C, Latha A, Jalaja G, “A Comparative Study of Indian Food Image Classification Using K-Nearest-Neighbour and Support-Vector-Machines”, International Journal of Engineering &Technology , 2018.

[5] V Hemalatha Reddy , Soumya Kumari , Karan Gigoo , “Food Recognition and Calorie Measurement using Image Processing and Convolutional Neural Network“, IEEE, 2019.

[6] Kaggle,<https://www.kaggle.com/cdart99/food20dataset>

[7] Hemraj Raikwar , Himanshu Jain ,“Calorie Estimation from Fast Food Images Using Support Vector Machine”,IJFRCSCE,2018.

[8] Kiran Ambhore, “Measuring Calories and Nutrition from Food Image”,IJARCCE,2016.

[9] Shamay Jahan, Shashi Rekha, Shah Ayub Quadri, “Bird’s Eye Review on Food Image Classification using Supervised Machine Learning”,IJLTEMAS,2018.

[10] Diksha Solanki, Ankit Anurag, Dr. Amita Goel, Ms. Vasudha Bahl4, Ms. Nidhi Sengar , “ Ankit Anurag2, Dr. Amita Goel3, Ms. Vasudha Bahl4, Ms. Nidhi Senga “, IRJET,2021.

[11] Nutrionx,< https://www.nutritionix.com/ >