ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2020.0321

TRANSFER LEARNING: INCEPTION-V3 BASED CUSTOM CLASSIFICATION APPROACH FOR FOOD IMAGES

Vishwanath C. Burkapalli[1] and Priyadarshini C. Patil[2]

[1]*Department of Information Science and Engineering, Poojya Doddappa Appa College of Engineering, India*

[2]*Department of Computer Science and Engineering, Poojya Doddappa Appa College of Engineering, India*

Abstract

Deep-learning approach has become more popular in the field of image processing. When it is concerned with health issues, there are lots of improvements in the applications of food image classification by deep learning methods. Transfer learning has become one of the popular techniques used in inception V3 for image classification, it is the reuse of a pre-trained model on a new model, where it uses a small amount of dataset to reduce the training time and increases the performance. In this paper, the Google Inception-V3 model is considered as a base, in top of that fully connected layer is built to optimize the classification process. In the model building process, convolution layers are capable to enough learn on its own convolution kernel to produce the tensor outputs. In addition, the separately obtained segmented features are concatenated with our custom model before the classification phase. It enhances the capability of important features and utilizes in the process of food classification. Here, the dataset of 16 class food images is considered and it contains thousands of images. The 96.27% classification accuracy has been obtained at the testing phase which is compared with different state-of-art techniques.

Keywords:

Deep Learning, Transfer Learning, Convolutional Neural Networks (CNNs), Food Classification, Calories Estimation, South Indian Dataset, Inception Model

1. INTRODUCTION

Food is always being one of the important parts of human life and, got lots of attention in the present research field. In general, supplies of food depend upon the human eye verification in order to validate the food quality ingredients. Ingredient determining, calories estimation, and proper labeling is a very tough task even for the human eyes as well as for machines and this process is very tedious, costly, and laborious [1]. Therefore, an efficient food detection model which can be able to automatically classify the food type is impressive. There are several recognition techniques and image processing technique that has got speedy enhancement in several applications [2]-[5] such as; remote sensing, medical imaging, surveillance systems and etc. Many researchers worked on the importance of data mining and machine learning techniques that can be used to classify food images [6]. For the assessment of properly dietary intake and accurate calorie estimation of food is very important paramount. However, advancement in processing technique and memory has allowed to, train the model of machine learning in limited time duration. Moreover, the build machine learning models are proficient in order to practice in the real-time data, and some traditional model was incapable of label correctly. Therefore, images are used to send towards “high-processing” servers, which increase the communication cost as well as delay in many applications.

Presently available systems are able to handle large amounts of high-quality images, which will be very helpful in order to provide food classification. In addition, it is focused on building a real-time type of application that able to capture the food images and instantly train the model of machine learning. Whereas, it helps us to take preclusion in order to avoid blood-pressure and diabetes type of diseases. Several types of methods currently used for the dietary assessment consist of manually recording instruments and self-reporting. The major problem with these types of methods is an assessment where the calorie consumption evaluation via the participant is susceptible to bias such as under-reporting and underestimating of intake food [7]. While to achieve enhanced accuracy and reduction in bias is required in current methods, the automatic analysis of calorie information and ingredient present in the particular cuisines, as well as food classification, is our major objective. In addition, proper baked, unbaked, and over baked information also provided.

The pair-wise classification framework is proposed in [8], which is used to increase the recognition rate for the process of food classification. Several studies have used a bag of features (BoF) methodology [9], in some approach additionally texton based histograms are used to perform the assembling procedure of the BoF models. But it is observed that BoF methodology carries lesser information, also when the high image resolutions are considered it shows failure. [10] Proposed a methodology for recognition and classification of fruit juice image. Their RGB color features are extracted from the images from the knowledge base. A 3-Sigma-based classifier got the classification efficiency of around 98%. The work finds applications in restaurants, malls, motels, and the like where the service robots automatically serve the food.

The real-time food image database is generated in [11], several experiments have been conducted at the same dataset with defining several benchmarks. Moreover, in the starting phase, they come across using Scale-Invariant-Feature-Transform (SIFT) features, which are further tested at seven classes. The optimized performance at the food replicas shows in [12] but performance efficiency was lower at the real-time images, where the type of capturing and image size could be the major reason for the performance degradation. While using the acquired SIFT features from the food database shows the improved result in [13] but when the numbers of classes are less in the database.

The Support Vector Machine (SVM) is considered in [14], which provides less accuracy, and further, Gabor filter can be used with pre-processing of SIFT feature extraction. The color-histogram method with feature extraction is also a good option to obtain a better outcome along with the SVM classifier. A food dataset is accumulated in [15], and the approach is proposed for food type recognition, also it capable to monitor dietary food plans. Also, the local features and global features are extracted to provide input to the classifier for the classification process. Other several types of classifiers that can be used for classification such as; k-NN [12], SVM [16], Neural Network (NN) [17], and Random Forest [18].

In recent years, there are several efforts have been made in the research field of food class prediction, where to extract information about food in a particular image is a challenging role. In this paper, the primary main aim is to collect real-time food images, which can be collected via camera, mobile camera, food blog website, and other internet resources. Most of these classes are of south Indian cuisine and these are considered because of less datasets available in open source platforms. Then, CNN based Inception-V3 architecture is considered as the base of our model because of its capability to get important features that can be utilized in process of food classification. On top of that, fully interconnected classification layer were used to get optimize classification accuracy.

The paper is arranged in such a way, where section 2 consists of information about related work, section 3 provides an architectural description of the proposed model with different phase information. Section 4 gives detailed information of result and analysis, the conclusion of our work is provided in section 5.

2. RELATED WORK

In the food recognition process, there have been several enhancements in state of the art technique in recent years. The local and global features are extracted from the food images in order to execute classification tasks [19] [20], where they make use of vocabulary trees and k-nearest neighbors and afterward adding the structural information and local appearance of the food objects to complete the classification task. In [16], used a bag of textons in order to represent the food images and SVM classifier used to classify those. In [22] considered the context of food as the feature, which is further used for the classification of food that has been consumed. Whereas, the food image datasets consist of a variety of food across from several countries.

In [23], considered Japanese food items for the recognition process, also presented multiple kernel learning in order to get features from several images, where features are SIFT, texture, and color. They achieved a classification accuracy of 61% with considering 50 hand-selected images of food from the internet. In [24], recognition of 85 food types was considered with the classification accuracy of 62.5% through recognition of the Japanese food item images. Moreover, Pittsburgh Fast-food (PFF) dataset [11] is the first openly available type of food dataset which consists of 101 number of classes, also has 3examplesat each class. In [29] proposed an approach of food recognition by considering small dataset that was envisioned to use in a smart-phone based food-recognition model, which was a part of their dietary assessment development.

In [12], the authors proposed a methodology to mine simultaneously discriminative food components through using random forest (RF) classifier, which is evaluated using the Food-101 food dataset. In recent years the deep learning method has been recognized as a very effective approach in order to provide recognition for large-scale objects and leveraged extensively in food recognition applications. A CNN approach for the food ingredient classification is proposed in [1], where the proposed framework shows a fine-tuning methodology with the CNNs in order to provide an online forecast of the food ingredients. In [6], the authors considered the 1000 categories of food, where it pre-trained using deep CNNs (DCNNs), also get an improved performance at a food classification process. Initially, 1000 food categories have been considered from the ImageNet with the 2000 categories, which are made by combined with ILSVRC 1000 number of ImageNet categories. Afterward, Caffe is used for the DCNNs as the pre-trained process for two datasets respectively, and it fine-tunes the pre-trained DCNNs.

Moreover, the experimental outcome shows the optimal result by using DCNN at 2000 food categories, where it is necessary to fine-tune the DCNN at the pre-processing step. In paper [17], shows the food recognition and detection process using CNN, also they considered only ten food classes in order to train the model and the dataset is publicly available such that the food-logging system. Leveraged CNN is used to classify the food and non-food images taken from the three different datasets [29]. In [30], food recognition is considered through patch-wise and voting methods via six layers of CNN. A novel methodology has been proposed for food recognition using a food image dataset called UNIMIB2016, which has 73 classes of food and 3616 instances of food. Where they used multi-variant features in order to provide a classification of food, the experimental outcome shows the significance of features extraction based CNNs model.

The deep learning methodology has evolved as a very popular in the research and development area that comes from the traditional artificial NNs (neural networks). At this time, CNN’s are frequently used in the computer vision, and specifically, the deep CNNs have headed in the process of image classification [25]. Though the AlexNet is first deep CNNs architecture for the image classification, where the framework was developed by

Alex Krizhevsky that performed better than some different advanced methodologies in the ImageNet. Whereas, it includes some pooling layers and convolutional layers that are arranged on each of the top rather than like traditional approach of a convolutional layer that followed via a pooling layer. In [26] Deep Convolution Neural Network used to get segmentation feature maps. Using this the absence of shape and edge constraints are solved by a post-processing phase with edge adaptive model. This model helps to get important characteristics of shape, pattern, etc.

GoogleNet is developed by Google that also provides the optimal deeper CNNs, where its main model is known as Inception and uses less number of parameters in CNNs. The major is to eliminate more parameters by GoogleNet is that it uses the function of average pooling as a replacement of fully-connected layers in the CNN that causes of network parameters elimination. In [27], also proposed another deep architecture known as a deep residual learning approach, furthermore Microsoft has also developed a residual network called ResNet for ImageNet localization, segmentation, and detection. In addition, this ResNet is a way of the deeper network compared to above other discussed network frameworks.

CNN's are very well designed for the application of image processing and it can achieve improved accuracy. Whereas, very large datasets are required to train via a deep-CNN approach. However, labeling and collecting such types of datasets are very difficult in process, and training deep CNN using some small type of image datasets is essential but it is very challenging. This study presents multi-class classification using CNNs architecture for food image datasets. In [28], proposed a low dimensional code which can be able to handle large dimensional data via training deep neural network also with the small central layer in order to develop large dimensional input vectors. Therefore, the deep learning approach can be used in several applications, also it gets nice continuous attention in the industrial area as well as academic area, due to its impressive deep learning performance in the field of image recognition. Here, deep learning methods for multi-class food classification were applied and some studies in [25] [28] show the various improvements in the deep artificial neural network, further, it also can be enhanced.

In some classification tasks, the depth of the residual neural network can reach up to 150 layers which are eight times more-deeper than the VGG networks. The extension of width and depth in the neural network allows extracting the more high-level type of features as compared to light neural networks. However, the major difficulty is to train a deep learning model at large-scale images from the beginning; like as ImageNet dataset that have a million number of labeled images.

3. PROPOSED METHODOLOGY

The proposed system model is shown in Fig.1 for the food image classification. Initially, the dataset collection procedure has been done, where the majority of classes belong from South-Indian cuisines that are collected by our self. All the food images are separated into their respective class folder and labeled properly. In order to process further, the size of images is considered to be colored images, because all of the images are resolution/size was very different from each other. The selection of this particular size is done by performing analysis, where it observed that majority of data images are collected via internet source that causes the image resolution to be low as compared to self-taken food images. Therefore, in order to maintain the relationship between neighborhood pixels it was important to consider the generic size of the image, so can each of the images can contribute to the model training process. If the size of the image is lesser than the predefined size, then it requires re-sizing of image and if the size of the image is more than the pre-defined size then it crops down to the pre-defined size of the image. Moreover, pre-processing, CNN model training, and classification phase were used; where the classification process has done using the test data that is unseen to a built model.

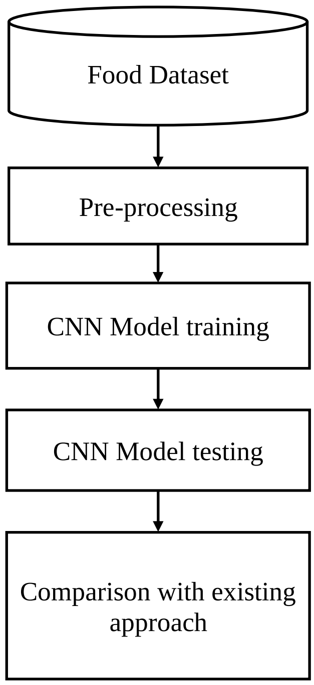


Fig.1. Proposed Model Approach

3.1 CNN STRUCTURE

Convolution neural network is a multi-layer artificial NN that incorporates both feature extraction as well as classification; also it takes the input of multiple raw images and produces an output of classification. However, the neurons present in a layer are organized in 3-dimensions with image input of height, width, and depth. Where it is all connected to every area of its preceding layer and the last layer decreases the whole image to a vector in order to provide class scores. Here the Inception-V3 model of CNN is considered and it retrained using our own accumulated datasets. The proposed Inception-V3 model is shown in Fig.2, which consists of AvgPool, MaxPool, Convolution, Concat Layer, Fully Connected layer, Dropout, and Softmax Function. The characterization of the convolution layer is provided via the weights sharing and sparse connectivity, where it calculates the neuron's outcome that is associated with present local regions from the preceding layer. In addition, it shares the weights of the neurons under individual feature maps and corresponds to kernels at the same layer. In the part of the classification, a fully connected layer takes the Inception-V3 feature outputs, as well as the custom generated segmented feature.

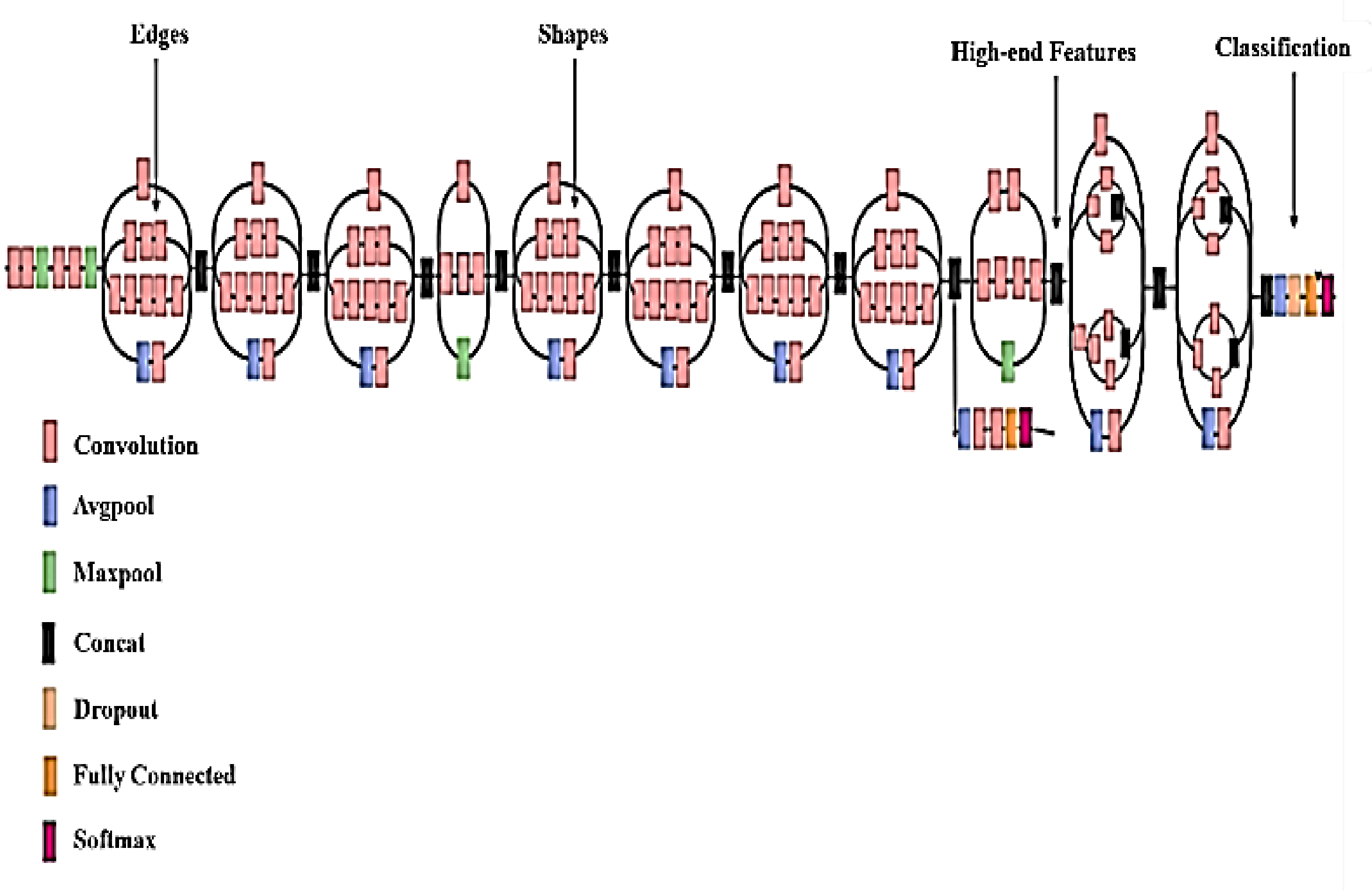


Fig.2. Inception V3 Model

The food image classification model consists of two main parts such as; feature extraction using CNN and classification process by using the softmax layer and fully connected layers. The convolution function takes input and it generates feature maps via convolving of input data.

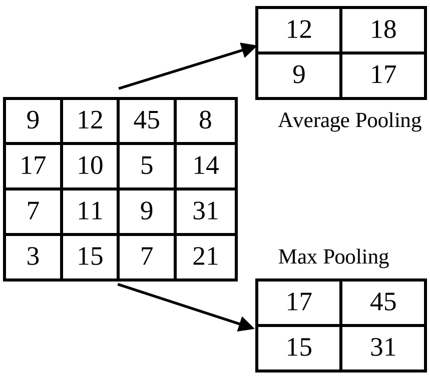


Fig.3. Operation of Average pooling and Max pooling

In order to build the obtained features from the convolution robust to counter noise, the pooling function is applied where the feature resolution is decreased through pooling function. The operation of pooling can be of two types such as; average pooling and max pooling as shown in Fig.3. Here average pooling is a two-Dimensional (2D) function that has (8, 8) size and minimizes the variance of data via minimizing computational complexity. Furthermore, it allows transferring the output to the next layer for the next operation. And max pooling is also a 2D pooling function which reduces the computational complexity and data variance. In terms of feature extraction, max-pooling extract important features of edges and smooth features are extracted by average pooling.

The concatenation is a process to concentrate its input blobs to a single output blob, where the input is in the form of a tensor list and returns an output of tensor with concatenating all tensor inputs. The dropout function is used as the regularization method to decrease the over-fitting in a model of the convolutional neural network via overcoming multifaceted co-adaptations of the training data. In this model, 0.4 dropout scales are considered and it is a nice effective methodology to perform averaging at the CNN model. The fully connected function is used to link all the neurons to one layer also with other layers and the working procedure is similar to the old "multi-layer-perceptron" (MLP) NN. The operation of softmax function is used to get the output and works similarly to the max-layer when it trains by gradient descent.

3.2 PRE-PROCESSING OF IMAGE

To get the maximum efficiency through our proposed scheme, pre-processing technique is considered and it will ensure that the image taken from different sides and angles should classify properly. Several necessary parameters have been considered such as fill mode, Horizontal flip, rotate angle, height-shift range, and width-shift range. The height shift range is considered as 0.2 and the width shift range is also considered as 0.2. In width shift range, the images are shifted horizontally by 0.2 fractions and allow to predict the different half/ incomplete images. Moreover, in the height-shift range, the images are shifted vertically by 0.2 fractions and the purpose is the same as a horizontal shift. The rotation range of 45º is taken that rotates randomly, it also able to guarantee that an image is captured from a different angle to predict appropriately with conserving the variety of the patterns of the feature maps.

The horizontal flip is true then images are flipped horizontally and random image flipping will help to recognize various patterns, which helps to accurately predict images for the upside-down images. The fill-mode is considered to set the points that are outside of image boundaries and the process is to be filled in accordance with the considered mode. Also, the random crop size is taken to crop images that further will be the input to the neural network. Where the images are forced to a proper size, and also ensure compatibility and the linearity at NNs.

3.3 CNN TRAINING PHASE

Here, the proposed model utilizes the custom Inception-V3 weights which are pre-trained using ImageNet [10] and it considers the reshaped size of 150×150×3 for all images. The function of average-pooling is considered at the food image dataset, where it takes the average of image features and the dimensionality of space output is defined via the dense-function. Here, a dropout fraction-rate is taken to handle the over-fitting issue. Moreover, the definite class is defined from the number of classes, the softmax function is used to identify the determined probability to get output for a particular class and at that time the rest of classes are neglected.

A convolution neural networks classifier is used to obtain an effective classification of food image; the RMSprop optimizer is considered with a learning rate of 2e-5 to achieve better performance. A total of 100 epochs has been taken to train a CNN model and also have callbacks to record the progress by log file. However, a scheduler learning rate is also defined to take the epoch index as input. The interface of the checked pointer is used to provide checkpoints of the model and saved in the .hdf5 file format, where this considers only the best score to save the learned convolution neural network models.

3.4 CLASSIFICATION PHASE

During the classification phase difficulty arises while considering several types of cuisines and dishes that exist in the real world. The different size and variety of dishes in the dataset causes difficulty in the classification process. The CNNs are considered to be a better approach to overcome the scaling problem due to its capability to capture the patterns associated with the images. In addition, it is also capable to deal with noise that existing in the images.

The database of image-net is a very common available dataset to provide image classification using the inception model by an effective training process. But here we have the dataset of our own South-Indian food images to train the convolution neural network model, where some classes are of un-baked and over-baked. The specifications of the model are considered as follows size "150×150×3" is consider, 2 Max-pooling downscale is used for the individual "spatial dimension" and the softmax function is activation and dropout rate 0.4 is considered.

4. RESULTS AND ANALYSIS

The result and analysis of our proposed model is given in this section, which shows the performance evaluation of food classification process. The data used for training our model consist of thousand numbers of images; with a batch size of 128 and 100 the number of epochs with 1719 images, Nvidia Pascal Titan Graphical Processing Unit. The input is converted into tensor objects, which have a float 32 representation. Python language has been considered in this work for model development. The testing has been done in the standalone system under Anaconda prompt, the system is configured with 12 GB RAM, operating system Windows 10 and processor of Intel i5.

During the evaluation of the model, it has several number saved models, which allows providing loading accessibility with the lowest loss and high accuracy in process of model evaluation. Consistently, the validation of the test set is done by using multiple crops instead of a single value, which raises the accuracy to compare to a single-crop based technique. Moreover, the confusion matrix is generated to plot each class label and it shows the true label prediction vs. false label prediction for the several classes. Here, a total of 16 classes are considered which contain 2149 number of images, for the testing purpose 20% of total images are used (i.e., randomly selected from all classes) and 80%

remaining image for training purposes. Equation (1) provides the accuracy function, where correctly predicted is divided by total testing class and multiplied by 100 in order to get accuracy in percentage.

$$Accuracy = \left(\frac{Correctly\ Predicted\ Class}{Total\ Testing\ Class}\right) \times 100 \qquad (1)$$

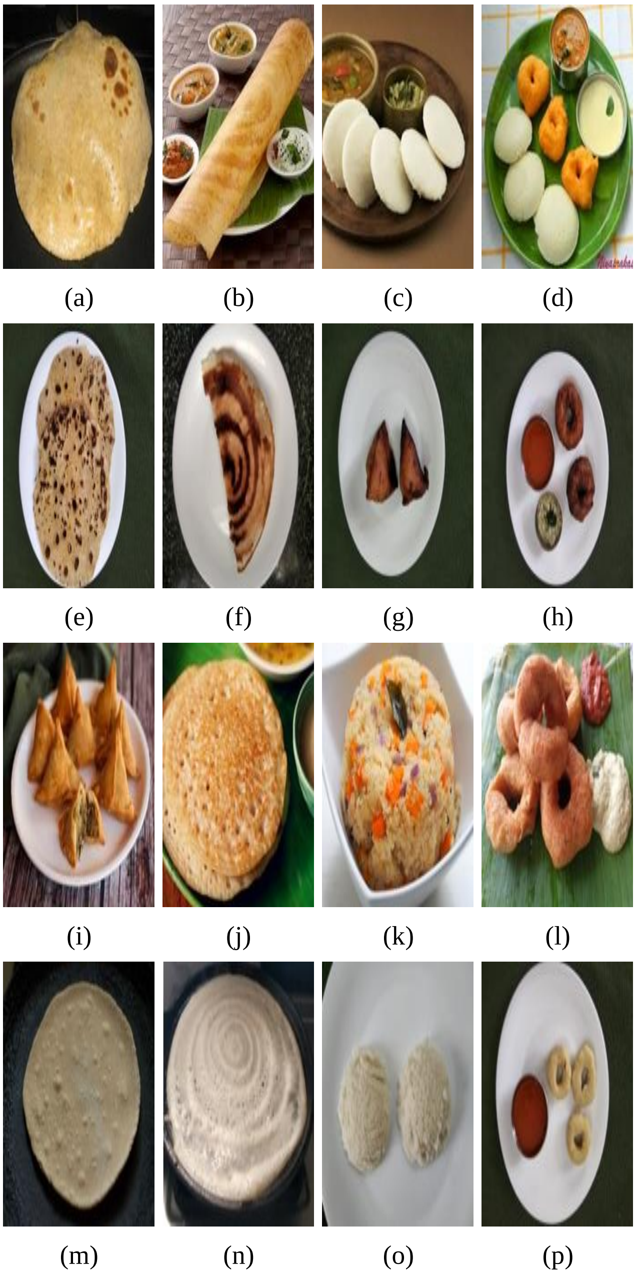


Fig.4. Several Type of considered cuisine(a)Chapatti, (b) Dosa, (c) Idly, (d) Idly and vada,(e) Overbaked Chapatti, (f)Overbaked Dosa, (g) Overbaked Samosa, (h) Overbaked Vada, (i) Samosa, (j) Set-Dosa, (k) Upma, (l) Vada, (m) Less-Baked Chapatti, (n) Less-Baked Dosa, (o) Less-Baked Idly and (p) Less-Baked Vada

Here total of 16 classes were considered and type of these classes has shown in Fig.4 (i.e., (a) Chapatti, (b) Dosa, (c) Idly, (d) Idly and vada, (e) Overbaked Chapatti, (f) Overbaked Dosa, (g) Overbaked Samosa, (h) Overbaked Vada, (i) Samosa, (j) Set-Dosa, (k) Upma, (l) Vada, (m) Less-Baked Chapatti, (n) Less-Baked Dosa, (o)Less-Baked Idly and (p) Less-Baked Vada).

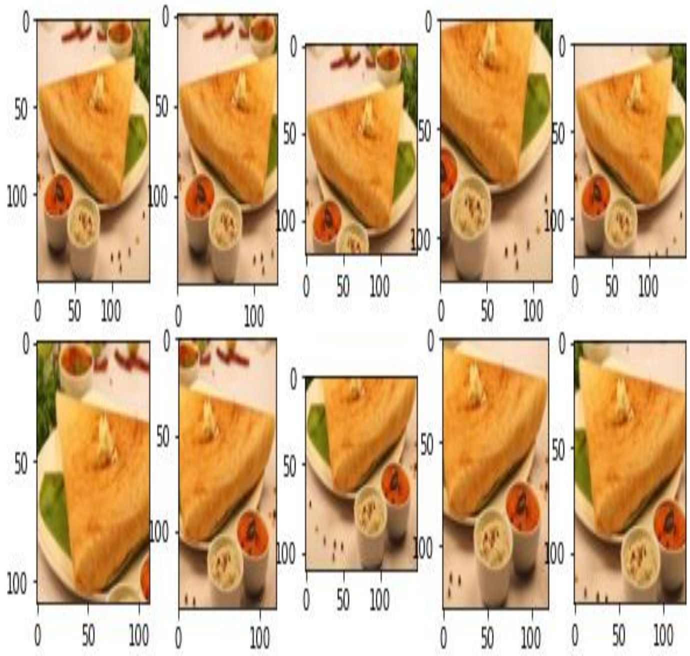


Fig.5. Top 10 classification using dosa image at different angle and position

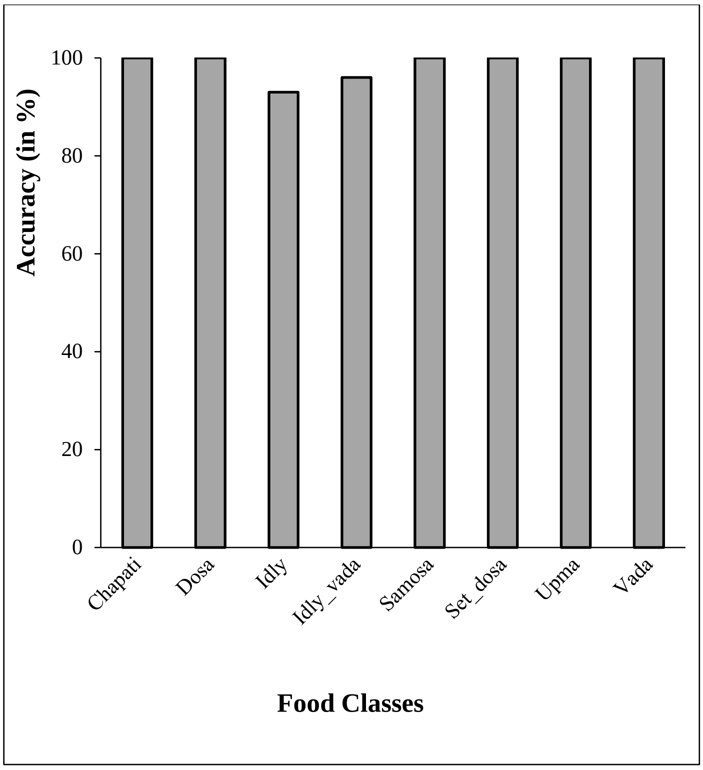
In that a total of 8 classes for proper baked, 4 classes for over baked and other 4 classes of less baked cuisines are considered. In addition, to check the performance of food images from different angles and sizes, the output is generated from the top prediction value for each crop that considered delivering the top five estimations. Therefore, the individual image predictions as shown in Fig.5 (classification of dosa image at different position and angle) are executed at this process level, so on the mapping method is utilized to map the index of test element to acquire the top estimates.

The base model InceptionV3 with pre-trained weights of image-net is included with input shape (150,150,3). In classification phase two more dense layers are added, where the first layer has 128 units, activation function Rectified linear unit (ReLu) with 0.5 dropouts, and Batch Normalization is considered. In the second layer 64 units, the activation function of ReLu with 0.5 dropouts is considered. The considered optimizer is RMSprop which is a momentum-based optimizer and has similar functionality of Ada-delta optimizer. The accuracy metrics and loss function is of ‘binary cross-entropy’. The training process batch size is 128 and the number of total images of training is 1719. Training logs were generated for each iteration/ epoch and loss is monitored. Current loss is checked with respect to previous loss and less loss model is getting saved, here we have run the model to 100 epochs.

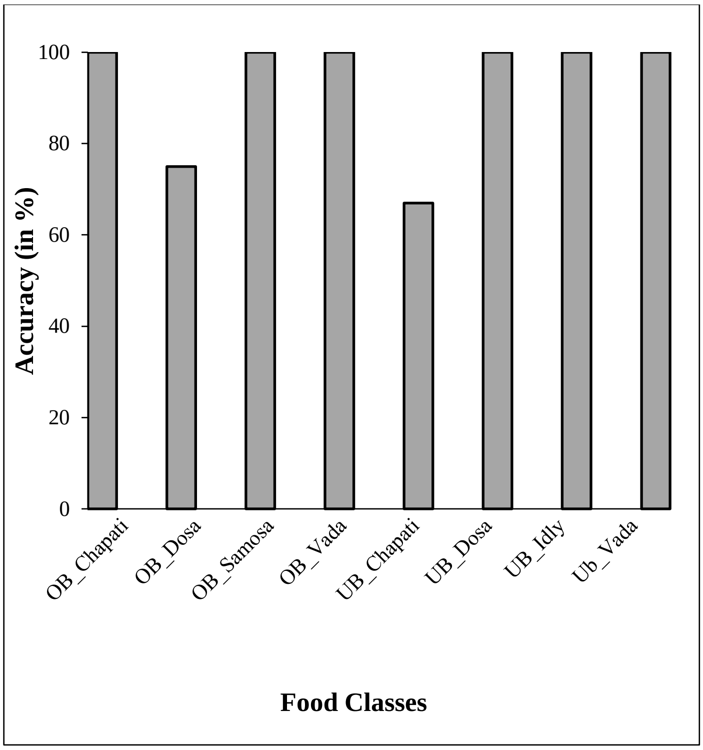
Predicted Label – Dosa	Predicted Label – Samosa
	
(a)	(b)
Dosa Recipe contains: urad dal, fenugreek, raw rice, par-boiled rice, beaten rice, ghee. Number of Calories/unit is 120 Properly Baked	Samosa Recipe contains: medium potatoes, oil, cumin seeds, fennel seeds, crushed coriander seeds, finely chopped ginger, green chili, teaspoon hing, asafoetida, green peas, coriander powder, garam masala, amchur, red chili powder, salt. Number of Calories/unit is 91 Properly Baked
(c)	(d)
Predicted Label – UB_Vada	Predicted Label – Set_dosa
	
(e)	(f)
Vada Recipe contains: urad dal, green chilli, ginger, few curry leaves, coriander, dry coconut/ kopra, hing/asafoetida, rice flour, salt, oil for deep frying. Number of Calories/unit is 150 Under Baked	Set-dosa Recipe contains: parboiled rice, whole urad dal, fenugreek seeds, Ghee or Oil, Water, Garnishing → well chopped Coriander leaves, carrot, onion. Number of Calories/unit is 117 Properly Baked
(g)	(h)

Fig.6. Predicted class label with image and corresponding console

Here, Fig.6 shows some of the predicted class labels with images and the corresponding console screenshot. The Fig.6(a) shows the correctly classified output of dosa (properly baked), it shows the contained recipe and calories per unit as shown in Fig.6(c). In Fig.6(b) shows the correctly classified output of samosa (properly baked) with the contains recipe and calories per unit in Fig.6(d), In Fig.6(e) shows the correctly classified output of samosa (under baked) with the contains recipe and calories per unit as shown in Fig.6(g). In addition, Fig.6(f) and Fig.6(h) contain the output details of properly baked set-dosa with their recipe and calories.



(a)



(b)

Fig.7. Recognition Accuracy for each considered classes (a) properly cooked food images, (b) over/un cooked food Images

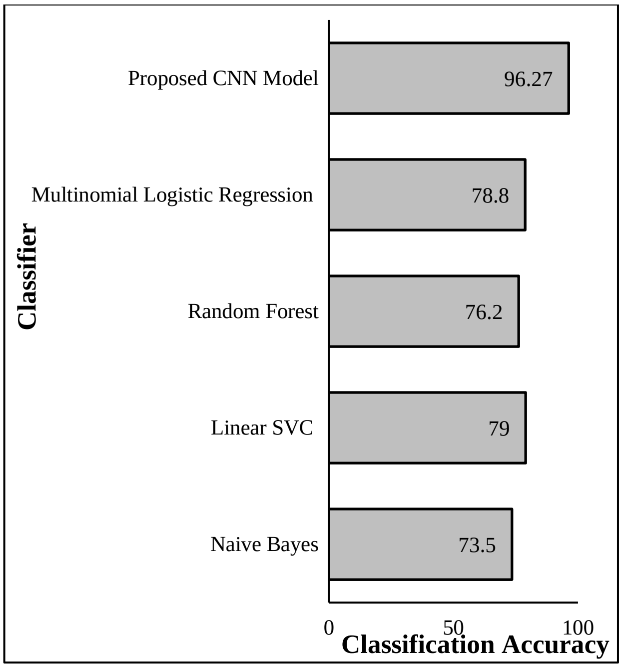


Fig.8. Classification accuracy comparison w.r.t existing Approaches

Our work considers 80% of images from dataset such as 1719 images are selected for the training phase and got training accuracy of 98.77%, which means out of 1719 images the model has fitted correctly at 1698 images. However, 20% of images such as 430 images have kept for testing, and out of this 414 images where correctly classified as shown in Fig.7, it describes the confusion matrix for the considered classes. It can be analyzed that miss-classification happened for some images whose texture, shape and color are very identical; otherwise it has given optimal results at the unseen dataset. The Fig.8 shows the classification accuracy compared with respect to existing approaches such as Naive Bayes, Linear SVC, Random Forest and Multinomial Logistic Regression models have been considered in [21]. These models were tested on different sources of food datasets such as; Yummly, Food.com, and Epicurious. Whereas, our proposed model has been tested on self-collected cuisine dataset and manages to get 96.27% classification accuracy.

5. CONCLUSION

The work started with collecting real-time food images from various resources such as cameras, mobile cameras, food blog website, and other internet resources. The majority of collected classes belong from south Indian cuisine. The custom convolution neural network-based model is trained over more than thousands of food images at the top of imageNet weights, which enhances our model capability to get important features quickly. In the result analysis, it has been seen that model accuracy at the training phase is good about 98.77% at 100 epochs. To optimize the model performance in the testing phase the individual image predictions are executed using the mapping method where it maps the index of test element to acquire the top estimates. So we got 96.27% classification accuracy in testing which is the much better result as per considered existing approaches.

Our proposed custom convolution neural networks model is much optimal for the classification of food images. And it is able to provide us calories estimation, ingredient, and as well as, correctly able to differentiate between properly baked, less baked, and over baked food images. In future work, different frameworks such as; TFOD, YOLO, Dectoron2 and etc. can be utilized for more enhancement.

REFERENCES

[1] H. Chen, J. Xu, G. Xiao, Q. Wu, and S. Zhang, "Fast Auto-Clean CNN Model for Online Prediction of Food Materials", *Journal of Parallel and Distributed Computing*, Vol. 117, pp. 218-227, 2017.

[2] S. Pouyanfar and S.C. Chen, "Semantic Concept Detection using Weighted Discretization Multiple Correspondence Analysis for Disaster Information Management", *Proceedings of 17th IEEE International Conference on Information Reuse and Integration*, pp. 556-564, 2016.

[3] X. Chen, C. Zhang, S.C. Chen and S. Rubin, "A Human-Centered Multiple Instance Learning Framework for Semantic Video Retrieval", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Vol. 39, No. 2, pp. 228-233, 2009.

[4] C. Chen, Q. Zhu, L. Lin and M.L. Shyu, "Web Media Semantic Concept Retrieval via Tag Removal and Model Fusion", *ACM Transactions on Intelligent Systems and Technology*, Vol. 4, No. 4, pp. 1-22, 2013.

[5] T. Meng and M.L. Shyu, "Leveraging Concept Association Network for Multimedia Rare Concept Mining and Retrieval", *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 860-865, 2012.

[6] K. Yanai, and Y. Kawano, "Food Image Recognition using Deep Convolutional Network with Pre-Training and Fine-Tuning", *Proceedings of IEEE International Conference on Multimedia and Expo Workshops*, pp. 1-6, 2015.

[7] P. Pouladzadeh, S. Shirmohammadi and R. Al-Maghrabi, "Measuring Calorie and Nutrition from Food Image", *IEEE Transactions on Instrumentation and Measurement*, Vol. 63, No. 8, pp. 1947-1956, 2014.

[8] M. Puri, Z. Zhu, Q. Yu, A. Divakaran and H. Sawhney, "Recognition and Volume Estimation of Food Intake using a Mobile Device", *Proceedings of IEEE International Conference on Applications of Computer Vision*, pp. 1-8, 2009.

[9] L. Fei Fei and P. Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories", *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 524-531, 2005.

[10] Basavaraj Anami and Vishwanath C. Burkpalli, "Recognition and Classification of Images of Fruits Juices Based on 3-Sigma Approach", *International Journal of Computational Vision and Robotics*, Vol. 1, No. 2, pp. 206-217, 2010.

[11] M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar and J. Yang, "PFID: Pittsburgh Fast-Food Image Dataset", *Proceedings of IEEE International Conference on Image Processing*, pp. 289-292, 2009.

[12] F. Zhu, M. Bosch, I. Woo, S. Kim, C.J. Boushey, D.S. Ebert and E.J. Delp, "The Use of Mobile Devices in Aiding Dietary Assessment and Evaluation", *IEEE Journal of Selected Topics in Signal Processing*, Vol. 4, No. 4, pp. 756-766, 2010.

[13] F. Kong and J. Tan, "Dietcam: Automatic Dietary Assessment with Mobile Camera Phones", *Pervasive and Mobile Computing*, Vol. 8, No. 1, pp. 147-163, 2012.

[14] T. Joutou and K. Yanai, "A Food Image Recognition System with Multiple Kernel Learning", *Proceedings of IEEE International Conference on Image Processing*, pp. 285-288, 2009.

[15] G. Ciocca, P. Napoletano and R. Schettini, "Food Recognition: A New Dataset, Experiments, and Results", *IEEE Journal of Biomedical and Health Informatics*, Vol. 21, No. 3, pp. 588-598, 2017.

[16] G.M. Farinella, M. Moltisanti and S. Battiato, "Classifying Food Images Represented as Bag of Textons", *Proceedings of IEEE International Conference on Image Processing*, pp. 5212-5216, 2014.

[17] H. Kagaya, K. Aizawa and M. Ogawa, "Food Detection and Recognition using Convolutional Neural Network", *Proceedings of ACM International Conference on Multimedia*, pp. 1085-1088, 2014.

[18] L. Bossard, M. Guillaumin and L. Van Gool, "Food-101-Mining Discriminative Components with Random Forests", *Proceedings of IEEE International Conference on Image Processing*, pp. 446-461, 2014.

[19] Y. He, C. Xu, N. Khanna, C. Boushey and E. Delp, "Analysis of Food Images: Features and Classification", *Proceedings of IEEE International Conference on Image Processing*, pp. 2744-2748, 2014.

[20] D.T. Nguyen, Z. Zong, P.O. Ogunbona, Y. Probst and W. Li, "Food Image Classification using Local Appearance and Global Structural Information", *Neurocomputing*, Vol. 140, pp. 242-251, 2014.

[21] S. Jayaraman, T. Choudhury and P. Kumar, "Analysis of Classification Models based on Cuisine Prediction using Machine Learning", *Proceedings of IEEE Conference on Smart Technologies for Smart Nation*, pp. 1485-1490, 2017.

[22] V. Bettadapura, E. Thomaz, A. Parnami, G.D. Abowd, and I. Essa, "Leveraging Context to Support Automated Food Recognition in Restaurants", *Proceedings of IEEE International Conference on Applications of Computer Vision*, pp. 580-587, 2012.

[23] T. Joutou and K. Yanai, "A Food Image Recognition System with Multiple Kernel Learning", *Proceedings of IEEE International Conference on Image Processing*, pp. 285-288, 2009.

[24] H. Hoashi, T. Joutou and K. Yanai, "Image Recognition of 85 Food Categories by Feature Fusion", *Proceedings of IEEE International Conference on Multimedia*, pp. 296-301, 2010.

[25] K. He, X. Zhang, S. Ren and J. Sun. "Deep Residual Learning for Image Recognition", *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.

[26] Vishwanath. C. Burkapalli and Priyadarshini. C. Patil, "Fine Grained Food Image Segmentation through EA-DCNNs", *International Journal of Innovative Technology and Exploring Engineering*, Vol. 9, No. 1, pp. 1-12, 2019.

[27] G.E. Hinton and R.R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks", *Science*, Vol. 313, No. 5786, pp. 504-507, 2006.

[28] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed and D. Anguelov, "Going Deeper with Convolutions", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9, 2015.

[29] H. Kagaya and K. Aizawa, "Highly Accurate Food/Non-Food Image Classification based on a Deep Convolutional Neural Network", *Proceedings of IEEE International Conference on Image Analysis and Processing*, pp. 350-357, 2015.

[30] S. Christodoulidis, M. Anthimopoulos and S. Mougiakakou, "Food Recognition for Dietary Assessment using Deep Convolutional Neural Networks", *Proceedings of IEEE International Conference on Image Analysis and Processing*, pp.458-465, 2015.