An Efficient Food Image Classification By Inception-V3 Based Cnns

Viswanath.C.Burkapalli,Priyadarshini.C.Patil

Abstract: The food item recognition technique from the images is being the interesting area with several type of applications. Though the monitoring of food provides major part in the health related issues and very much needed in our everyday lives. In this paper, we adopted a Google Inception-V3 based convolutional neural networks (CNNs) model in order to classify food images. However, the artificial neural networks and CNNs have the proficiency to directly estimate score function from the image pixels, here we used convolution layer that is able to create its own convolution kernel in order to convolve with input layer to generate the tensor outputs. Moreover, the Max-Pooling function is used for features extraction from the data and help to train the CNN model. In addition, there are multiple number of layers has been considered and at last the obtained outputs are concatenated in order to generate the final outputs. Using our proposed implementation, we measured 92.89% of accuracy with considering the yummly and own dataset classes.

Index Terms: Convolutional neural networks (CNNs), Max-Pooling, machine learning, food classification, food image, segmentation.

1 INTRODUCTION

The present generation people are very conscious for their food in order to recover from diseases or to avoid the upcoming diseases. Technologies has evolved so much that the people required automatic labelling food application. For this purpose in several recent years many researchers has been aimed for automatic food recognition from using machine learning and computer vision techniques. Most of the people are having habit of over-eating that cause not being much active. Stressed and busy life of people causes unable to keep track the proper food dietary, which increases the significance of proper food classification and information about that. In present time, smart applications are required for the proper labelling food, which is only possible by using machine learning technique that is tremendously trending. Moreover these kinds of applications are proficient of balancing the user's food-habits and also warn them when unhealthy type of food is detected. The advancement in memory and processing technologies has made possible to train the machine learning model in reasonable time duration. There are number of classifiers are available which has been used for the classification process, whereas some better classifiers are available amongst various classifier such as Naïve Bayes, support vector machine, adaboost and artificial neural network. Furthermore, in [1] proposed the framework of pair-wise classification in order to improve the rate of recognition for food classification process. The bag of features ('BoF') is also most popular classifier which derived from bag of words and it is majorly used at natural language processing. Though it designed to acquire frequent appearing words through neglecting the appeared order [2] [3]. In addition, the food images are consist of some similar visualise pattern, which used to predict the type of food and itdecreases the process complexity issues via using methodology of direct image matching.

- *Vishwanath.C.Burkapalli., Professor, Dept. of Information Science and Engg., PDA College of Engg., Kalaburgi, Karnataka, India E-mail:vishwa_bs@rediffmail.com.*
- *Priyadarshini.C.Patil,Asst.prof.,Dept. of Computer Science and Engg., PDA College of Engg., Kalaburgi, Karnataka, India E-mail: priyadarshinicpatil@gmail.com.*

Therefore, several research work has been carried on BoF technique. Moreover, the texton based histograms has been used in order to perform the resemble process of BoF models. However, it concluded that BoF technique can carry less amount of information, also it shows failure when the image resolutions are high. However, the checker-board technique is utilized for colour capturing in order to apply at system for changing light conditions. Whenever, the numbers of classes are increasing, it has seen that the accuracy of performance is decreases. Though, the database is considered to be paramount characteristic for the food classification task and model should able to handle this on a real-time scenario. Henceforth, the real-time database of the food images has generated in [4] and further experiments has done on the same type of dataset through considering defined benchmark, while at starting stage they have used SIFT("scale-invariant-feature-transform") features and tested with the seven classes. In [5], shown the good performance at food replicas and lower performance efficiency at real images, also the capturing type and size of the images could be the cause for the degradation of performance. However, using the extracted SIFT features from the starter-food and foods in [6] shows better result but when the classes are less with several images. In [7] color and texture features are presented to the neural network for training and later of the unknown grain types mixed with foreign bodies. The combination of both color and texture features is employed in the work. In [8], considered the SVM classifier that provided the very less accuracy, furthermore, the SIFT feature extraction with the Gabor filter and the colour-histogram based feature extraction are also the option to get better result along with the SVM classifier. The k-means clustering approach has been used with 50 classes of food. In addition, it has aimed on gathering the several type of food dataset. In [9] proposed a novel dataset in order to evaluate the proposed algorithms to recognize the food type that will further used to monitor the food diets. The built database contains more than the 3,500 types of food through collecting food images of canteen with trays. Moreover, the global and local features set are extracted and further classified using classifiers such as; random forest (RF) [10], SVMs [11], neural networks [12] and k-NN classifier [5], which will provide proper knowledge on determining the appropriate classifier [13]. Whereas, CNN has got popularity in various domain of image classification such as remote sensing

images [14], stem cell biology [15], and a survey of fruits and vegetables using deep learning is provided in [16]. A Back Propagation Neural Network (BPNN) is used to classify and recognize the fruit image samples, using three different types of feature sets, viz, color, texture, combination of both color and texture features. The study reveals that the combination of color and texture features are out performed the individual color and texture features in identification and classification of different bulk fruit image samples [17]. In [18] Deep Convolution Neural Network used to get segmentation feature maps and by 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 like shape, contextual information between regions, region connectivity etc. In recent years, many of the research efforts have been taken in the area of efficient food label prediction. Whereas the efficient information extraction from the food images are still challenging issue. Here, the major aim is to provide classification of food images by using CNNs. However, the CNNs are very much capable to handle huge amount of dataset and also it has capability to estimate important features which will be utilized in food classification process. Here we have considered the yummly dataset as well as real-time captured dataset, where most are from South-Indian food type such as Idly, Wada, Dosa and etc. in order to the provide classification. The remaining paper is organised in such a way that section 2 provides explains of proposed methodology with selected database and provides a proper description of CNN model. Section 3provides the observations and results, at lastly section 4 concludes the proposed system work and the future work.

2 PROPOSED METHODOLOGY

Fig .1 shows the proposed scheme for food images classification, and the each considered block explanation is given in this section. Here, we have considered food data acquisition, pre-processing, and CNN training. In addition, the obtained trained model further can be used to classify the food images through using test dataset.

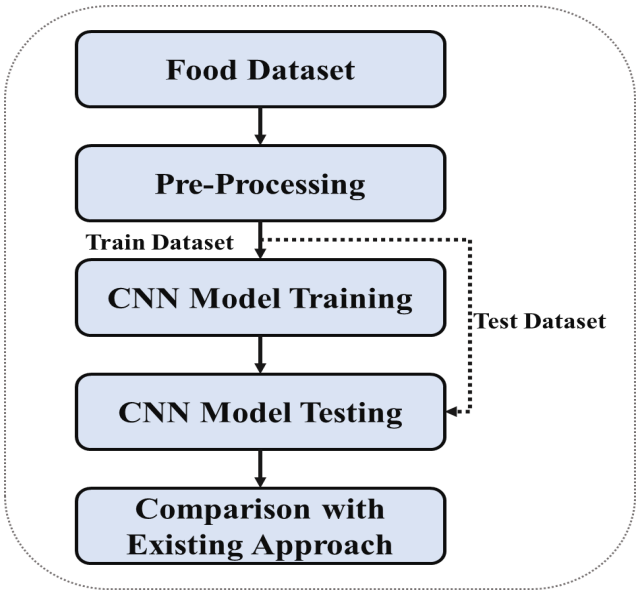

***Fig 1**: Proposed Scheme for food images Classification*

2.1 Acquired Food Dataset

Here, we have taken some data from the Yummly API [19] [20] and some real time south-Indian food data has considered, where some of the training and testing images has some noise, different color intensity and images with the wrong-labels. Which has taken care for proper the training and the testing phase and also we rescaled the images to a size of 299 x 299 dimensions.

2.2 CNN Structure

Google Inception-V3 based CNN model is considered for retrained, this model consist of AvgPool, Convolution, MaxPool, Concat Layer, Dropout, Fully Connected layer and Softmax Function.

- **Average Pooling-** It is a 2-Dimentional (2D) function with a pool size of (8,8), which decreases the data variance as well as the computational complexity, also this layer allows to flows the outcome to the next layer.

- **Convolution-** A (299, 299, 3) input size is used by convolution function and this layer generate the feature-maps through convolving the input data.

- **Max-pool-** It is a two dimensional max-pooling function and it reduces the variance at data as well as the computational complexity. The Max pooling is used to extract the important features such as edges and average pooling is used to extract features smoothly.

- **Concatenation-** this layer is used to concatenates its several input blobs into a single blob of output and it takes list of tensors as the input, where all have the same type of shape expect concatenation axis. It returns an output of single tensor with concatenating all the inputs.

- **Dropout-** It is considered to be the regularization approach in order to minimize over-fitting in artificial neural network through overcoming complex co-adaptations from the training data. Here, we have considered 0.4 as the dropout scale and it is very effective method to performing averaging with the neural network model. Moreover, the term called dropout refers to drop-out the units such as visible and hidden side in a considered neural network.

- **Fully-connected-** This is used to connect all neuron to a one layer as well as to another layer, which works on the basis of traditional “multi-layer-preceptor” (MLP) ‘neural network’.

- **Softmax-** The softmax function is used like the output function, which works similar to max-layer when it is differentiable to train via a gradient descent. Though, the exponential function will cause increment in the probability of preceding layer and compare to the other value; correspondingly, all output summation will be always equal to one.

2.3 Pre-Processing of Image

Here, we ensure maximal efficiency by our proposed approach with considering pre-processing technique. This technique will also ensure that an image captured from any angle should

able for the classification. There are several important parameters which are considered in step of image pre-processing such as height-shift range, width-shift range, rotation angle, Horizontal flip, and fill mode. Height shift range and width shift range both are considered of 0.2 units, where in width shift range the images are generally shifted horizontally through 0.2 fraction and the obtained patterns will be different and allows to predict half or incomplete images. Whereas, in height shift range the images are generally shifted vertically through 0.2 fraction of total width and the purpose is similar to the horizontal shift. The 45 degree of rotation range is considered which rotate randomly and able to ensure that the images are taken from any type of angle should be predicted correctly with preserving the patterns diversity of feature maps. When horizontal flip is considered to be true then the images are horizontally flipped, this random flipping of image will help to identify various patterns and also helps to predict images accurately for upside down images. Moreover, the fill mode is set to be reflecting then the points which are outside the boundaries of images will going to be filled in according to the mode. In addition, random crop size is assigned in order to crop the images, which going to feed to neural network, where each images are forced to crop to size (299 x 299 x 3) and also ensure the linearity and compatibility at the neural network.

2.4 CNN Training Phase

This work utilizes the model of Google Inception-V3 that is pre-trained on the ImageNet [4], where the reshaped size “299 x 299 x 3” is considered for all the images. Moreover, the average-pooling function is considered on the food image dataset and takes average of all the image features. The space output dimensionality is defined through the dense-function. However, 0.5 dropout fraction-rate of input units is taken to overcome the issue of over-fitting. Additional, to decide the definite class from the several number of classes, the function of softmax is defined and it identify the maximum probability in order to obtain output for the particular class and neglect the rest of the classes. Here, a CNN is used to get effective food image classification, the stochastic Gradient-Descent is consider with a rapidly reducing learning to obtain a better performance. The 32 epochs has consider to train a model and have defined callbacks to record the growth using a log file. Moreover, a learning rate of scheduler is define, where it takes input of epoch index and afterwards provides an output of new learning rate. Check pointer callback is used to build the model checkpoints and these are saved in the format of .hdf5 files, where only best score is considered to save the learned models.

2.5 Classification Phase

In the classification phase, the problem comes when we considers the several cuisines types and dishes, which present in real world. While given the variety and size of the food cuisines in dataset, this will going to have a bit difficult task. Usage of the neural networks considers to be the better option to deal with difficulty of scaling and it’s mainly because of neural network ability to acquire the patterns which are not linear-separable. Along with this concepts it is capable of dealing with the other factors like noise that present in the images. The image-net database is very popular and easily available dataset in order to execute image classification that has been perfectly train in the CNN based Google Inception model and it has the several categories of existing classification. In order to generalize the system model, yummly_66 dataset and own South-Indian food recipe are added to CNN model via training it. Moreover, the model specifications are as follow, a size of “299x299x3” is considered for the input sensor with a 2 Max-pooling downscale in individual “spatial dimension”. With the function of softmax activation and 0.4 dropout rate. In addition, one-hot encode has been consider in order to get a binary features set from individual label, which is improved than single feature because it can take any of the required value from the several classes. A pipeline image augmentation has considered which provides cropping tools as well as the image preprocessor inception.

3. RESULTS AND ANALYSIS

This section provides the result and analysis using our proposed approach from the technique of performance measurement for the food classification. Here, we have used Python scripting language for the successive simulation of model, where the system configuration is of 12 GB RAM, Intel i5 processor and Windows 10 operating system. Model evaluation having multiple numbers of saved models, where it is accessibility to load and evaluate the models with highest accuracy and lowest loss. Furthermore, we also obtain a confusion matrix that based upon the outcome obtain through the CNNs. The confusion matrix is able to plot each of the class label, where it shows the correctly label prediction vs. the incorrectly label prediction for different class. Correspondingly, to validate the test set, there are multiple crops has considered instead of the single value, which increases the accuracy as comparatively to single-crop based evaluation technique. The outcome is generated from top predictions for individual crop that in turn is been used to deliver the top five predictions and the crops are produced for each item in a test set in order to get the predictions. Therefore, the predictions for individual image is obtained at this stage of process, afterwards the mapping approach is used for map the test element index in order to get the top predictions. There are total 15 classes has considered that contain 218 images and for testing we have used 30 images, where the testing images are chosen randomly. In training phase, several shifting, rotation and cropping operation has been done which multiplies the number of training images in a huge quantity. Here, we consider the segmented features (SF) that has feed as one of the input to classifier. The model includes; weights: - one-of-none for the 'imagenet' or random initialization for pre-training on the ImageNet, and Include_top: is used for whether to contain fully-connected layer at the topmost of network. The input_shape: is used as the optional shape tuple, where it needs to be specified if the include_top is taken as False and otherwise the shape of input is (299 x 299 x 3) (i.e., 'channels_last' format of data) or (3 x 299 x 299) (i.e., 'channels_first' format of data). Here, the pooling option is also can be considered when the include_top is False for the feature extraction and it should contain correctly 3 inputs channels with height and width. If the pooling function is considered to be “None” which means the model output will be 4D tensor outcome from the former convolutional layer. If the pooling function is considered to be “avg”, which means that the global average pooling function is applied to the outcome of preceding convolutional layer, therefore the model output will be the 2 dimensional tensor. When the “max” is used that

means the global-max pooling function will be applied. In addition, the classes function provide the optional classes number to classify the images, only if the include_top is specified to trueand there will be no weights argument should be specified. Finally, it returns an instance of Keras Model.

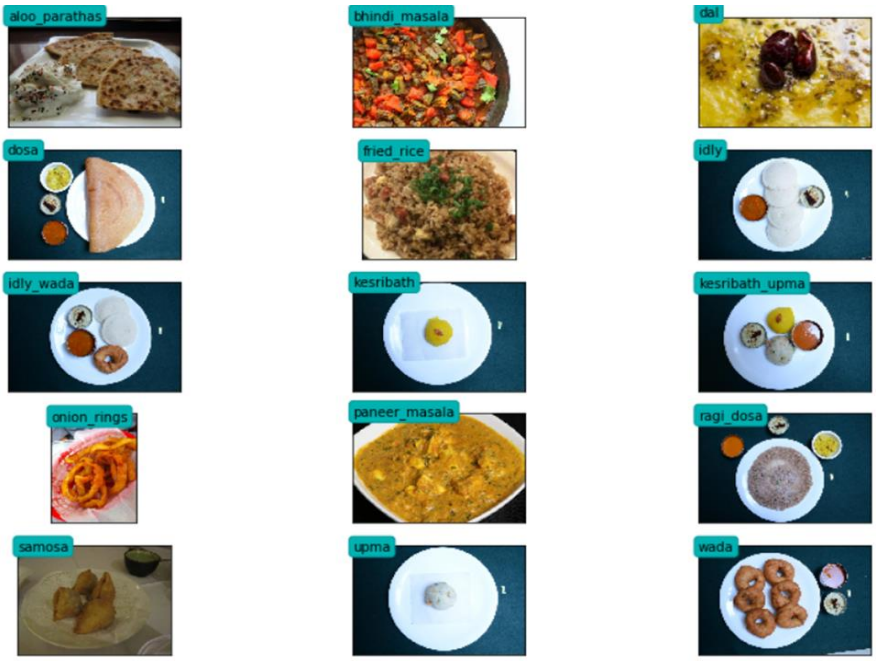

Fig 2. *Several Type of considered food classes.*

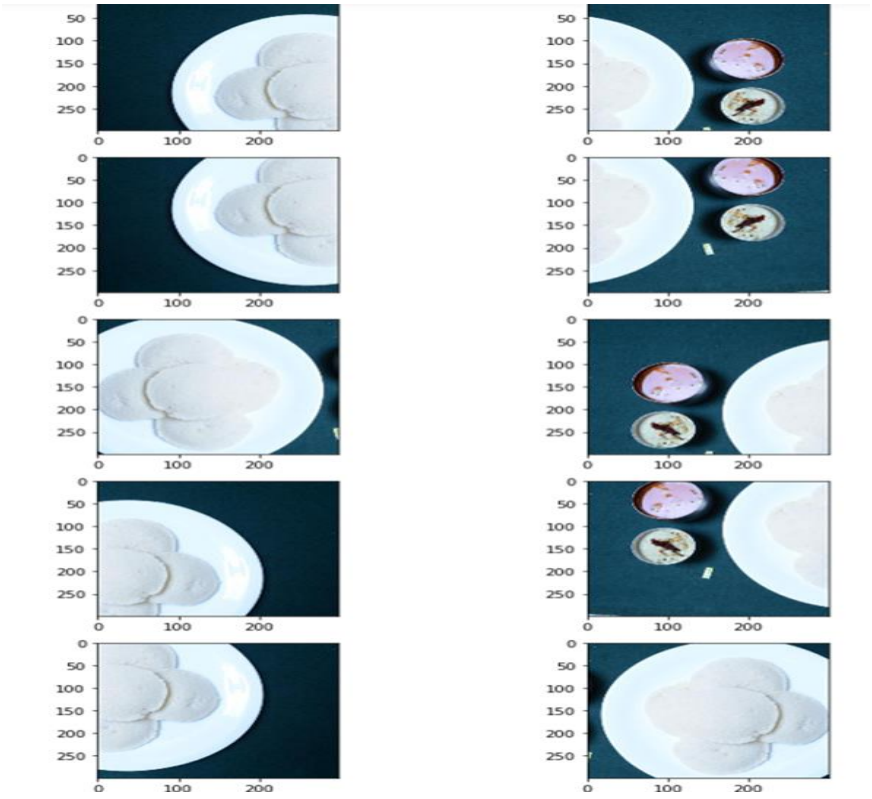

Fig 0. *Classification of idly image at different angle and position*

***Table 1**: Comparison with the existing Approaches*

Different Classification Approaches	Classification Accuracy (%)
Naive Bayes [21]	73.5
Random Forest [21]	76.2
Multinomial Logistic Regression [21]	78.8
Linear SVC[21]	79
Proposed CNN without SF	81.78
Proposed CNN with SF	92.89

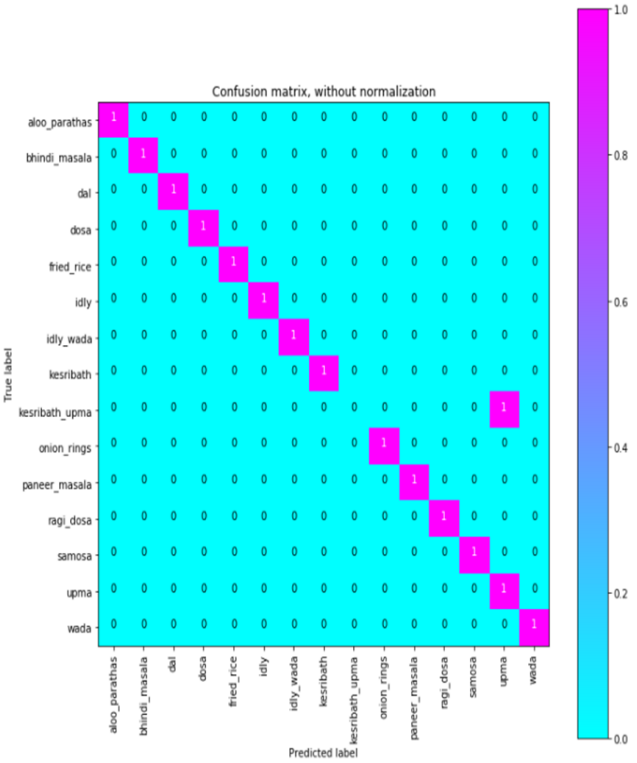

***Fig 4**. Confusion Matrix for the considered classes.*

Fig 2. shows the several type of considered food classes for image classification, Fig 3. shows the classification process by our proposed model at idly image under at different angle and position. Table 1 shows the comparison with the existing approaches; Naive Bayes, Random Forest, Multinomial Logistic Regression and Linear SVC models has been considered [21]. Where these models has been tested on various sources of dataset such as; Food.com, Yummly and Epicurious. Whereas, our model has been tested for several number of cuisine and it got 93% classification accuracy. The detailed information of classification has given through the confusion matrix that has shown in figure 3.3, where one class is incorrectly predicted and rest all are predicted efficiently.

4. CONCLUSION

In the above result and analysis section we seen our system model performance that is considerable high and as per simulation of the CNN we analyzed that it requires high-performable system for the large number of datasets. The capability of CNN for training is very high for non-linear data, but it requires more “computational time” in order to train model; though, the performance of model matters a lot, therefore once the system model is trained properly it takes very less time to produce the system output. All the considered images are appropriately preprocessed and every type of the images are tested using CNN model. Two type of training scenarios has consider with and without SF, where it analyzed that with SF our proposed model got 11.84% more accuracy as compared to without SF. As per the performed analysis, we can concluded that our proposed CNNs model are suitable for the food images classification. In future work, ingredient identification in the particular class of food can be obtain, calories estimation is also very useful for the dietary point of view and type of backing is also very important to predict.

REFERENCES

[1]. M. Puri, Z. Zhu, Q. Yu, A. Divakaran, and H. Sawhney, “Recognition and volume estimation of food intake using a mobile device,” in Applications of Computer Vision (WACV), 2009 Workshop on. IEEE, 2009, pp. 1–8.W.-K. Chen, Linear Networks and Systems. Belmont, Calif.: Wadsworth, pp. 123-135, 1993. (Book style)

[2]. L. Fei-Fei and P. Perona, “A bayesian hierarchical model for learning natural scene categories,” in Computer Vision and Pattern Recognition, 2005. IEEE Computer Society Conference on, vol. 2. IEEE, 2005, pp. 524–531.

[3]. T. Joachims, “Text categorization with support vector machines: Learning with many relevant features,” Machine learning: ECML-98, pp. 137–142, 1998.

[4]. M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar, and J. Yang, “Pfid: Pittsburgh fast-food image dataset,” in Image Processing (ICIP), 2009 16thIEEEInternationalConferenceon. IEEE,2009,pp.289–292.

[5]. 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.

[6]. F.KongandJ.Tan,“Dietcam: Automatic dietary assessment with mobile camera phones,” Pervasive and Mobile Computing, vol. 8, no. 1, pp. 147–163, 2012.

[7]. B. S. Anami, Dayanand G. Savakar ,“Effect of Foreign Bodies on Recognition and Classification of Bulk Food Grains Image Samples”, Effect of Foreign Bodies on Identification and Classification of Bulk Food Grains Image Samples, Journal of Applied Computer Science and Mathematics, Volume 3(6), Pages: 77- 83, 2019.

[8]. T. Joutou and K. Yanai, “A food image recognition system with multiple kernel learning,” in Image Processing (ICIP), 2009 16th IEEE International Conference on. IEEE, 2009, pp. 285–288.

[9]. 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.
[10]. L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101–mining discriminative components with random forests," in Lecture Notes in Computer Science, vol. 8694. Springer, 2014, pp. 446–461.
[11]. G. M. Farinella, M. Moltisanti, and S. Battiato, "Classifying food images represented as bag of textons," in Image Processing (ICIP), 2014 IEEE International Conference on. IEEE, 2014, pp. 5212–5216.
[12]. H. Kagaya, K. Aizawa, and M. Ogawa, "Food detection and recognition using convolutional neural network," in Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014, pp. 1085–1088.
[13]. Y. He, C. Xu, N. Khanna, C. J. Boushey, and E. J. Delp, "Analysis of food images: Features and classification," in Image Processing (ICIP), 2014 IEEE International Conference on. IEEE, 2014, pp. 2744–2748.
[14]. Baoxuan Jin, Peng Ye, Xueying Zhang, Weiwei Song, Shihua Li, "Object-Oriented Method Combined with Deep Convolutional Neural Networks for Land-Use-Type Classification of Remote Sensing Images", Journal of the IndianSociety of Remote Sensing, June 2019, Volume 47, Issue 6, pp 951–965.
[15]. Dai Kusumoto, Shinsuke Yuasa, "The application of convolutional neural network to stem cell biology", Inflammation and Regeneration, July 2019, 39:14.
[16]. Dayanand Savakar, "Identification and Classification of Bulk Fruits Images using Artificial Neural Networks", International Journal of Engineering and Innovative Technology (IJEIT), Volume 1, Issue 3, March 2012.
[17]. Vishwanath.C. Burkapalli, Priyadarshini.C.Patil, "Fine Grained Food Image Segmentation through EA-DCNNs", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-1, November 2019.
[18]. Anuja Bhargava, Atul Bansal, "Fruits and vegetables quality evaluation using computer vision: A review", Journal of King Saud University - Computer and Information Sciences, Available online 5 June 2018.
[19]. W. Min, S. Jiang, J. Sang, H. Wang, X. Liu, and L. Herranz. "Being a super cook: Joint food attributes and multi-modal content modeling for recipe retrieval and exploration", IEEE Transactions on Multimedia, 19(5):1100 C 1113, 2017.
[20]. S. Sajadmanesh, S. Jafarzadeh, S. A. Ossia, H. R. Rabiee, H. Haddadi,Y. Mejova, M. Musolesi, E. De Cristofaro, and G. Stringhini. Kissing cuisines: Exploring worldwide culinary habits on the web.
[21]. arXiv preprint arXiv:1610.08469, 2016.
[22]. S. Jayaraman, T. Choudhury and P. Kumar, "Analysis of classification models based on cuisine prediction using machine learning", International Conference On Smart Technologies For Smart Nation (SmartTechCon), Bangalore, pp. 1485-149, 2017..