International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 03 | Mar 2021 www.irjet.net p-ISSN: 2395-0072

Adaptive Learning Rate Base Convolution Neural Network for Food Recognition

Urvashi Rakholiya[1], Ankita Gandhi[2], Hetal Bhaidasana[3]

[1]Post Graduate Student, PIET-CE, Parul University, Vadodara, Gujrat
[2]Asst. Prof, PIET-CSE, Parul University, Vadodara, Gujrat
[3]HOD & Asst. Prof, PIET-DS, Parul University, Vadodara, Gujrat

---

Abstract - *Picture recovery and order in the food field has gotten one of the increasingly more concerned examination themes in the field of mixed media investigation and applications. As of late, with the quick improvement of the Internet business and interactive media innovation, picture characterization and recovery innovation has become an exploration area of interest at home and abroad. Based on this, this paper proposes research on food image classification and image retrieval methods based on visual features, machine learning, CNN, R-CNN network and uses an adaptive learning rate is presented for training neural networks. Unlike most conventional updating methods in which the learning rate gradually decreases during training, the proposed method increases or decreases the learning rate adaptively so that the training loss decreases as much as possible. The experiments with some well-known datasets to train a multilayer perceptron show that the proposed method is effective for obtaining a better test accuracy under certain conditions.*

Key Words*: *Local Food Detection, Deep Learning, Convolutional Neural Network, Machine Learning.

I. INTRODUCTION

Image Classification

Image classification is the most important part of computer vision. In simple word, classification is the categorization of same kind of data in same category. Image classification is process that includes image preprocessing, image segmentation, key feature extraction [1]. There are three methods of classification supervised learning, unsupervised learning, semi-supervised learning.

Deep Neural Network

Deep neural network (DNN) is an artificial neural network with different film between the input and output layers. Deep neural network is also named as "stacked neural networks"; that is, networks composed of several layers. Deep neural network performs automatic feature extraction without human intervention.

Convolutional neural network

Convolutional neural network (CNN) is a specific type of deep neural network which is especially useful for image classification and image recognition. A convolutional neural network is also known as a convNet. CNN is capable of handling a large amount of data and can estimate the features automatically, and utilized for image classification [2]. Like other kinds of artificial neural networks; a convolutional neural network has an input layer, an output layer and many invisible layers. Some of these layers are convolutional layer, pooling layer and fully connected layer. In CNN, the layers are arranging in 3 dimensions: Width, height and depth. Main important of CNN, the neurons in one layer do not connect to all the neurons in the next layer but only to a small region of it. Last, the final output will be reduced to a single vector of probability scores, organized along the depth dimension. The CNN has excellent performance in machine learning problems [3]. Specially the applications that deal with image data, computer vision.

In any case, In the time of innovation, individuals utilizing a cell phone and they generally worry about their food [2]. On the off chance that they don't think about the name of food then they utilize the web to know the name of food. There are various sorts of classifier previously created to distinguish the food. There are a few groups who don't think about the new thing of food. From the start, we are spurred by the outsider individuals. Since consistently there are numerous vacationers or individuals go to our country for their work reason [3]. Greatest individuals don't think about our neighborhood food and they additionally don't have the foggiest idea about the name of the food. Likewise, it's useful for our neighborhood individuals. There are numerous individuals don't have the foggiest idea about the name all the food of our country [4]. Along these lines, it's useful for each individual not just the outsider. In our paper, we have proposed a model of Convolutional Neural Network (CNN).

During the time spent turn of events, we have utilized six extraordinary classes of nearby food of India. Whatever survives from the paper is figured out as seeks after: In Section II, we have given the relationship with some related papers. In Section III, we have given the square chart of our model, information portrayal with pre-handling, the

depiction of our proposed model with its preparation and furthermore test, the portrayal of Confusion Matrix for both paired and multi-class. In Section IV, we have given the outcome investigation in respects to our proposed model with other CNN [3] models. Finally, the end for certain future works has portrayed in Area V.

II. LITERATURE REVIEW

Pengcheng Wei, Bo Wang, Gao [1] based on Faster R-CNN network, implemented on food data set Dish-233 and 48, 189 images. It has proposed a simple convolutional neural network on image classification and also analyzed different methodsof learning rate set. Other compared with CNN-GF, the performance is improved by 5%. This additional is biased visible features, and improved in food image retrieval and grouping tasks.

Afsana Ahsan Jeny, Masum ShahJunayed, Ikhtiar Ahmed, Md. Tarek Habib [2] have proposed convolutional neural network, which classified food images. In approach, first pre-processing of images and then train the network. The images are actual preprocessor and all kinds of images are checked with CNN. From this,it is concluded that CNN is more suitable for classifying the images when the no of classes are more. The Food-101 dataset is used for experiment, which contains 101,000 images of 101 categories. Confusion matrix is used for evaluation of model. A confusion matrix is plotting each class label, and how many times it was correctly labeled vs. the other times it was wrongly non-identical as a different class. This proposed model is obtained 98% accuracy. In future, this classification task will be extended using prominent features that can categorize food images.

V. Hemalatha Reddy, Soumya Kumari, Vinitha Muralidharan, Karan Gigoo and Bhushan S. Thakare [3] proposed convolutional neural network, which classified food images. proposed classification methods like Stochastic gradient and region selection, K-mean clustering, and segmentation based on Graph Cut etc. These methods are used to detected the food items and calculate calories measurements and analyzed the food data set. Gradient. Stochastic gradient is normally used to reduce the error of a copy on the training data. To decrees the quadratic cost function and neural network is trained to observe the weight with bias. This was cost function as

$$\textbf{Cost } (w, b) \equiv 1/2 \,||\, y(x) - a \,||\qquad (1)$$

where, w → collection of all the weights in the network, b → biases, a → vector of outputs from the network, x → input. By using this function, cost function is improved by making it easier to make minor changes in the weights (w) and biases (b).
Deep learning is a growing method which helps a computer model learn in performing classification task directly from images. It's an detain of machine learning and has been implement with regard to locating of various layer of description are include.

Tomoumi Takase, Satoshi Oyama, Masahi to Kurihara. [4] proposed an approach to find prediction of food dishes and find the learning rate using Stochastic gradient descent Algorithm, ALR technique (Adaptable Learning Rate Tree algorithm). In proposed model, author compares the result of increases or decreases the learning rate adaptively so that the training loss decreases as much as possible. Comparison shows greater efficiency and accuracy can be obtained by ALR algorithm.

Weishan Zhang Dehai Zhao Wenjuan Gong Zhongwei Li Qinghua Lu Su Yang [5] used Support Vector Machine method. implemented on food data set UEC-Food101 dataset which is an open 100-class food image dataset including about 15000 images and the other is a fruit dataset that established by ourselves including over 40000 images. We have reach accuracy of 80.8% of fruit dataset with 60.9% on the multi-food dataset item.

Yi Sen Ng, Wanqi Xue, Wei Wang, Panpan Qi [6] used different classification algorithms Like Stochastic Gradient Descent (SGD), Confusion matrix, Augmentation technique to predict data from Food da. These algorithms are Adam optimization algorithm Confusion Matrix performed on Deep Residual Neural Network for grouping six classes of food also used two convolutional neural networks separately for grouping six classes of food images an achieved the highest accuracy of 98%.

Marc Bola˜nos, Marc Valdivia, and Petia Radeva, [7] proposed classification methods like CNN and RNN (Convolutional Neural Network and Recurrent Neural Network). These methods are used to detected the food items applying image food recognition by using the menu of the restaurants and improve the baseline by a 15%.

Sandhya Arora, Gauri Chaware, Devangi Chinchankar, Eesha Dixit and Shevi Jain,[8] used to SIFT and HOG. BOF (Bag-of-Features) and DCNN. In proposed model, author compares the result of increases or decreases accuracy and use of hand-crafted features s. Food items are being recognized using the cutting sounds, acoustic sensors, so on.

Lei Zhou, Chu Zhang, Fei Liu, Zhengjun Qiu, and Yong He [9] used Faster R-CNN, and Image retrieval method based on BOW (Bag of Words) classified SIFT features and implemented on food data set Food101 the combination of deep learning and multisource data fusion including RGB images, spectra, smell, taste, and considered to make a more comprehensive assessment of food, analyzed obtained the best classifier.

Abdulkadir ŞENGÜR, Yaman AKBULUT, Ümit BUDAK, [10] used different classification algorithms like Support Vector

Machine (SVM). and Fine tuning of a pre trained CNN. These techniques used to quality of size 4096 are remove from fc6 with fc7 layers and sequence with multiple mixture to control best of deep factor of stepwise food image classified. Achieved the accuracy of 79.85%.

Mohammed A. Subhi, Sawal Md. Ali, [11] proposed classification methods like CNN (Convolution Neural Network) These methods are used to detected the food items on food dataset and calculate the accuracy.

Niki Martinel, Gian Luca Foresti, Christian. [12] proposed an approach to using computer vision and pattern recognition algorithms and find prediction of food dataset-100. Achieved the accuracy of 90.27%.

III. RESEARCH METHODOLOGY

First, we give the block diagram of our proposed model; second, we portray our dataset; third, we talk about the pre-processing of our dataset; fourth, we quickly depict our proposed model with architecture; fifth, we example the train model with train, test and validation split; and then describe various classes.

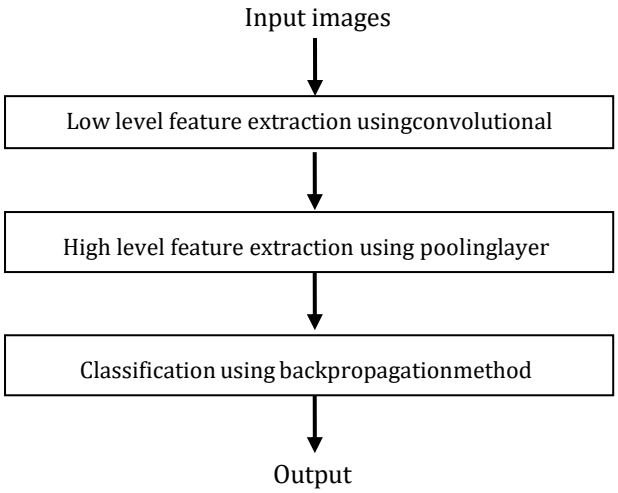


Figure 1. Abstract model of proposed approach

We have to used TensorFlow and Kera's both are the high-level framework of deep learning. We have also have used more layer of Convolution Neural Network (CNN). Which are Convolution layers, pooling layers, fully connected layers, flatten layers. We have used an activation function to make our output non-linear. In this paper used ReLU activation function. This output is

$$F(x) = \max(0, x) \qquad (2)$$

A. Dataset

In our observation, we need a too much food images. We have visited different places, restaurant and taken some pictures of local food of Gujrat and also collect some food images from FOOD-101. We have to managed 100 classes for our observation. So basically, total real images of our observation are 10, 1000 and per classes has 1000 images.

B. Data Processing

At the point when we caught the pictures and assembled the pictures from the web, all pictures weren't a similar size. For removing this issue, as an issue of first significance, we have resized the information. In the wake of resizing all the picture size is 224 x 224 pixel. By then, we have applied 5 expanded techniques with the objective that we don't go up against the overfitting [2] issue. These techniques are: turn right +30 degree, turn left - 30 degree, flip evenly, concealing and interpretation. To apply these techniques, we have gotten all out 3600 pictures for our test. It is amazingly difficult to exhibit all of the pictures. The following is an example dataset that has 30 pictures to investigate (Fig. 2).

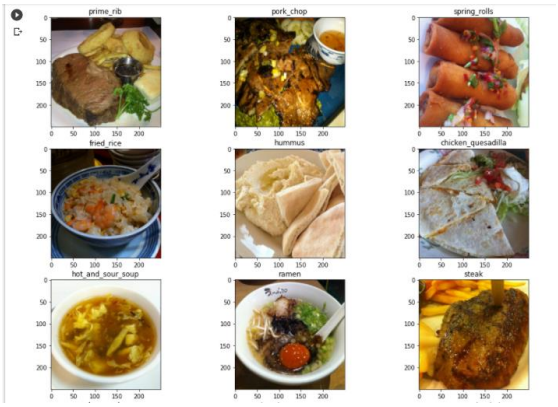


Figure 2. Dataset sample

C. Proposed Model

The proposed model has 48 convolution layers with pooling layers, enactment capacities, smooth layers, and at long last, completely associated layers and has additionally a few regularization techniques like dropout and bunch standardization. The guideline layer of CNN [3] is the convolutional layer. In request to make a component map, utilizing the channel or piece, the execution level is run in the info information. For each layer, the convolution shape is 64, 128, 256, 512, 1024 and 2048.

1. level feature extraction using convolutional layer

The first layer of proposed approach is convolutional layer. The primary purpose of convolution is to extract features from an input image and get more abstract information of image. The convolutional layer in line of a set of filters and every filter is separately convolved with image. Convolution preserves the relationship between pixels by learning features using small squares of input data. It is mathematical operation that takes two inputs such as image matrix and filter or kernel. Filter is used only in the first layer of a convolutional neural network to extract features from image like edge detection etc.

2. High level feature extraction using pooling layer

The main idea of pooling is down-sampling in order to reduce the complexity for further layer [10]. Pooling layer is placed in between successive convolution layers. The pooling layer has the effect of the secondary feature extraction. Pooling (also known as sub sampling) layer reduces the dimensionality of each feature map (convolved feature) but retains the most important information. In CNN, pooling is generally performed with 2*2 window, stride 2 and no padding.
The output size of pooling layer is calculated by equation 3.

$$\text{Output} = (I - Ps)/S + 1 \qquad (3)$$

Where I =Input image size Ps=window (pool) sizeS=stride

Max pooling is one of the most common types of pooling methods. It partitions the imageinto sub-region rectangles, and it only returns the maximum value from sub-region. One of the most common sizes is used in max-pooling is 2*2 [10]. In example, the maximum value within the window (sub-region of image) is pooled into an output matrix with (2*2*1).

3. Classification using backpropagation method

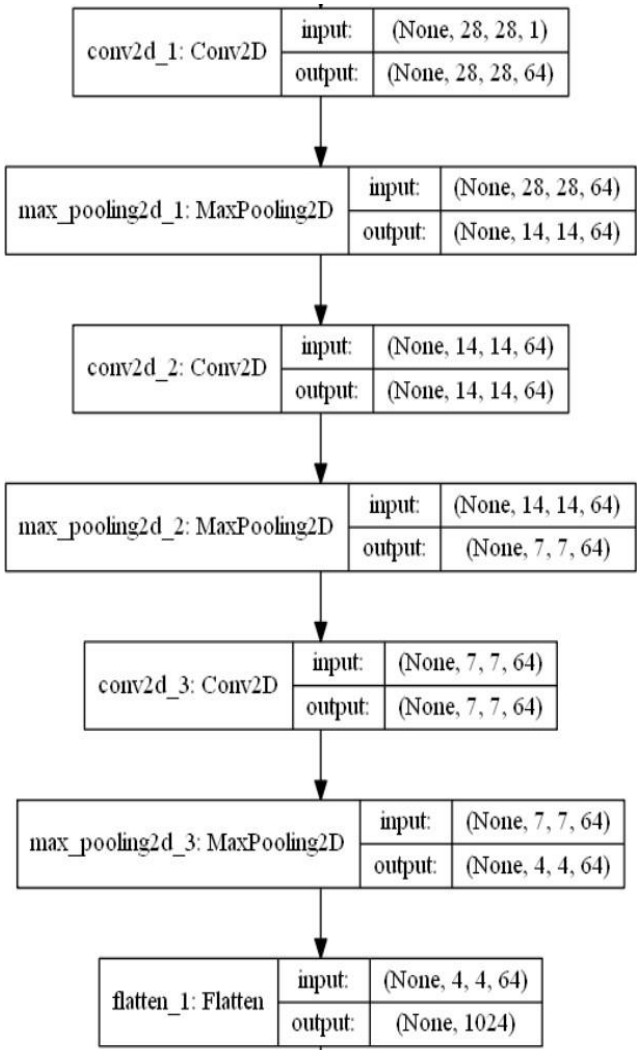
In fully connected layer, backpropagation method is used for classification. The fully connected layer is last layer, which uses a SoftMax activation function in the output layer. The fully connected layer is a similar to the way that neurons are arranged in a traditional neural network.
Therefore, each node in a fully connected layer is directly connected to every node in next layer [10]. Convolutional and pooling layers produce output as high-level features of the input image. The fully connected layer uses these features and classifies the input image into various classes given on the training dataset. The SoftMax activation function is used in output layer. This function is used for calculate the probabilities of the target classes.

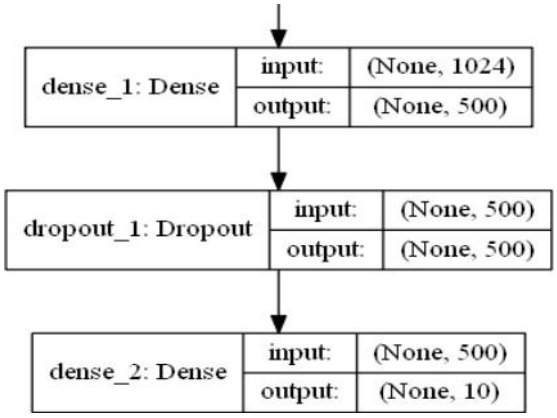
D. Training Model

In our present method, we used three CNN different model. In Proposed_CNN used all basic CNN layers are used and this model is analyzed on different optimization algorithms such as Adam, SGD and AdaGrad with different learning rates. In Proposed_CNN_Backtracking, apart from above three-layer batch-normalization layer and data augmentation are added for improving the performance and stability of network and L2-regularization technique is used to augment performance in Proposed_CNN_Regularization. As per our observation Proposed_CNN is the best performance and give accurate accuracy.

In our model, we have used 30 epochs. And in every epoch, our model has performed very well. The proposed model has applied in food-101 dataset and found a good result on train, test, and verification set. Data sample training used to provide a neutral assessment to fit a final model of test dataset. We have used 80% of our data for testing, testing, and verification, and 20% of data was used for testing. This means we have a total of 10,1000images. So, 75,250 images were used for training and validation and 25,250 images were used for testing. Now there are 1000 images in each class.

Proposed_CNN Architecture





IV. RESULT ANALYSIS

In Table 1, we have calculated these equations of confusion matrix and f1-Score.

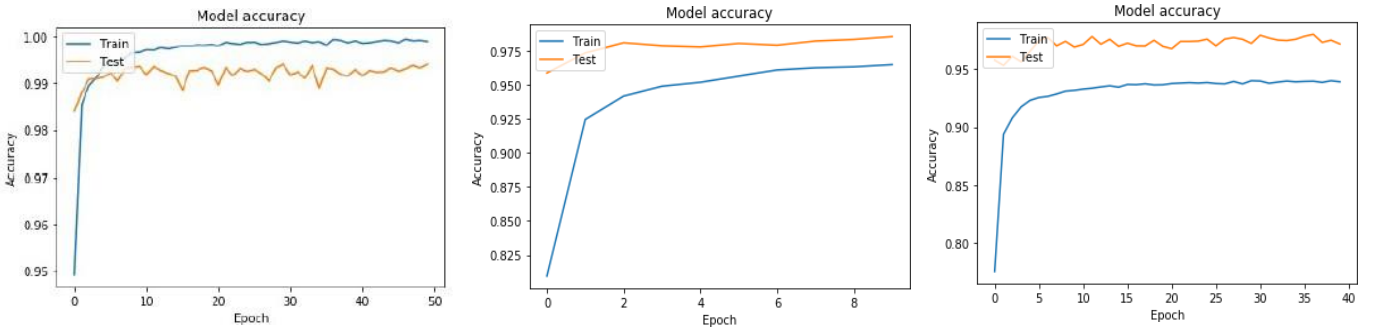
Class	Precision	Recall	F1-score
0	1.00	0.99	0.99
1	1.00	0.99	1.00
2	0.99	1.00	1.00
3	0.99	1.00	1.00
4	0.99	0.99	0.99
5	0.99	1.00	0.99
6	0.99	0.99	0.99
7	1.00	0.99	0.99
8	0.99	1.00	0.99
9	0.99	0.99	0.99

After creating 30 epochs, we can see that our model got the training accuracy is 96.56%, the validation accuracy is 93.31% and the testing accuracy of 96.26% on our food-101 dataset.

Here, Table 2 is showing the comparison among Proposed_CNN, proposed_CNN_Backtracking, Proposed_CNN_Regularization.

Model Name	Training Accuracy	Testing Accuracy
Proposed_CNN	96.56%	96.26%
Proposed_CNN_Backtracking	93.53%	95.50%
Proposed_CNN_Regularization	95.20%	96.26%

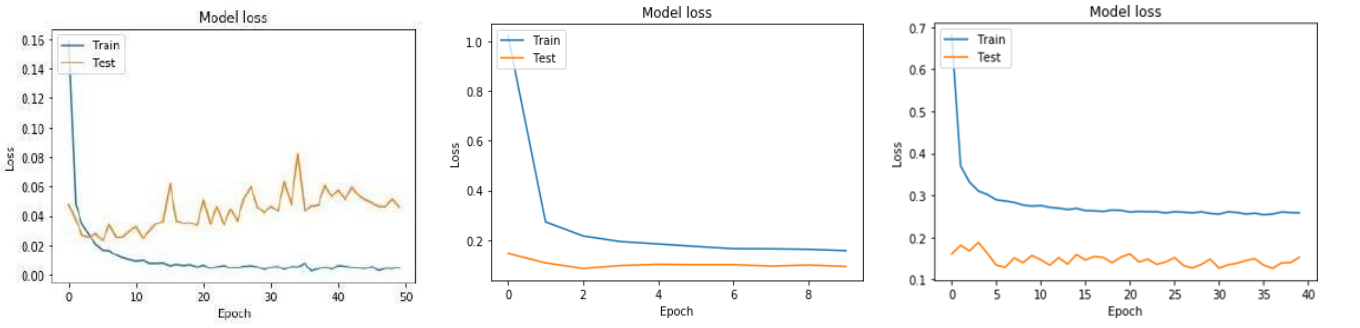
Performance comparison of FOOD-101 Dataset for all models (accuracy)



(A) Proposed_CNN **(B) Proposed_CNN_Backtracking** **(C) Proposed_CNN_Regularization**

Figure 3. Performance comparison of FOOD - 101 dataset for all Proposed_CNN (Accuracy)

Performance comparison of dataset for all models (loss)



. **(A) Proposed_CNN** **(B) Proposed_CNN_Backtracking** **(C) Proposed_CNN_Regularization**

Figure 4. Performance comparison of FOOD - 101 dataset for all Proposed_CNN (loss)

V. CONCLUSIONS

In this thesis, convolutional neural network is used for food image classification. Contrary to traditional methods that use hand-crafted feature, the proposed approach is used automatic feature extraction. In first phase, layers of convolutional neural network and various parameters have been deeply studied. Simple convolutional neural network model is also implemented. In Second phase, various CNN models are proposed for image classification and improve learning rate. For evaluating the performance of the convolutional neural network of dataset: 101 category Food dishes. Proposed models are trained on these four dataset and results are observed. Among them 96.40% accuracy for FOOD-101 dataset are archived. These results are compared with existing CNN models.

In these models, optimization of parameters and modification of layers can be done to archive better accuracy. In this work, improve learning rate can also be used for better results.

REFERENCES

[1] Pengcheng Wei, Bo Wang, 'Food image classification and image retrieval based on visual features and machine learning,'Computer Science and Technology-2020, IEEE.

[2] Afsana Ahsan Jeny, Masum ShahJunayed, Ikhtiar Ahmed, Md. Tarek Habib 'FoNet-Local Food Recognition Using Deep Residual Neural Networks', International Conference on emerging Reseach in Electronics, Computer Science and Technology-2019, IEEER. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[3] V. Hemalatha Reddy, Soumya Kumari, Vinitha Muralidharan, Karan Gigoo and Bhushan S. Thakare, 'Literature Survey—Food Recognition and Calorie Measurement Using Image Processing and Machine Learning Techniques', International Conference on emerging Reseach in Electronics, Computer Science and Technology-2020, IEEE.

[4] Tomoumi Takase, Satoshi Oyama, Masahi toKurihara, 'Effective neural network training with adaptive learning rate based on training loss', International Conference on emerging Reseach in Electronics, Computer Science and Technology-2018, IEEE.

[5] Weishan Zhang Dehai Zhao Wenjuan Gong Zhongwei Li Qinghua Lu Su Yang, 'Food Image Recognition with Convolutional Neural Networks', International Conference on emerging Reseach in Electronics, Computer Science and Technology-2016, IEEE.

[6] Yi Sen Ng, Wanqi Xue, Wei Wang, Panpan Qi,' Convolutional Neural Networks for Food Image Recognition: An Experimental Study', International Conference on emerging Reseach in Electronics, Computer Science and Technology-2019, IEEE.

[7] Marc Bola˜nos, Marc Valdivia, and Petia Radeva,' Where and What Am I Eating? Image-Based Food Menu Recognition', International Conference on emerging Reseach in Electronics, Computer Science and Technology-2015, IEEE.

[8] Sandhya Arora, Gauri Chaware, Devangi Chinchankar, Eesha Dixit and Shevi Jain,' Survey of Different Approaches Used for Food Recognition', International Conference on emerging Reseach in Electronics, Computer Science and Technology-2019, IEEE.

[9] Lei Zhou, Chu Zhang, Fei Liu, Zhengjun Qiu, and Yong He,' Application of Deep Learning in Food: A Review', International Conference on emerging Reseach in Electronics, Computer Science and Technology-2019, IEEE.

[10] Abdulkadir ŞENGÜR, Yaman AKBULUT, Ümit BUDAK,' Food Image Classification with Deep Features', International Conference on emerging Reseach in Electronics, Computer Science and Technology-2018, IEEE

[11] Mohammed A. Subhi, Sawal Md. Ali, 'A Deep Convolutional Neural Network for Food Detection and Recognition', International Conference on emerging Reseach in Electronics, Computer Science and Technology-2018, IEEE

[12] Niki Martinel, Gian Luca Foresti, Christian Micheloni' Wide-Slice Residual Networks for Food Recognition', International Conference on emerging Reseach in Electronics, Computer Science and Technology-2016.