I.J. Education and Management Engineering, **2020, 3, 1-11**
Published Online June 2020 in MECS (http://www.mecs-press.net)
DOI: 10.5815/ijeme.2020.03.01

Available online at **http://www.mecs-press.net/ijem**

Cuisine Detection Using the Convolutional Neural Network

Dipti Pawade[a], Ashwini Dalvi[a], Dr. Irfan Siddavatam[b] , Myron Carvalho[c] , Prajwal Kotian[c] , Hima George[c]

[a]Assistant Professor, IT, K. J. Somaiya College of Engineering, Vidyavihar, Mumbai, India
[b]Associate Professor, IT, K. J. Somaiya College of Engineering, Vidyavihar, Mumbai, India
[a]Student, IT, K. J. Somaiya College of Engineering, Vidyavihar, Mumbai, India

Received: 16 March 2020; Accepted: 25 March 2020; Published: 08 June 2020

Abstract

In today's fast world, everyone wants the information in one click. The same rule applies when you have some food items in front of you. In social events, few cuisines are known to us while some are not. Also, in a few cases, we know the cuisine name, but we are not aware of its nutritional value. This motivated us to develop a system that can identify the cuisine name from the image and gives the nutrient value for the same. Here Convolutional Neural Network (CNN) is used to predict the cuisine name present in an image and then further its nutritional value is calculated based on the information present in a database. User needs to click the image of the cuisine; the application will identify the cuisine name and its nutrition value for standard serving amount considering the cuisine is prepared using the standard recipe.

Index Terms: Convolutional Neural Network, Cuisine Identification, Nutrition, ReLU.1

* Corresponding author.
E-mail address:

1. Introduction

In a diverse country like India, we have a wide variety of food. In addition to this, globalization has brought continental food in our dish. Many times, in a buffet or a social gathering, the food which is being served is not known to us. In this situation, the common thing which one does is asking about the dish name to nearby people, friends or the one who serves it. But yet the nutritional value of that dish remains unknown. Thus there is the need for a system that can not only provide the automatic recognition of cuisine but also help in estimating nutritional values, making them useful for dietary assessment and planning. This helps people with specific dietary limitations. There are many approaches, which can predict the name of the cuisine from the image. These approaches are discussed in detail in section II. But the major challenge faced by almost all the system is identifying the individual cuisine names from a dish containing many food items. To address this paper, we have provided a Convolutional Neural Network (CNN) model which takes the cuisine as an input, process it and gives the probability which is then compared against the probability matrix to predict the cuisine name. This name is then parsed to the database to fetch its nutritional value.

The primary objectives of our study are:

- Processing real-time cuisine image
- Identify the cuisine name using CNN
- Predicting carbohydrate, energy, mineral, fat, fibre, protein and calories contained in a standard serving amount of the cuisine identified.

Our main focus is on accurately identifying the multiple cuisine items present in the dish. We also tried to improve the performance of the model when to identify the closely related food items like biryani and rice or different types of curries.

The rest of the paper is organized as follows. Section 2 gives an overview of background work carried out by various researchers in this area. In part 3, the model architecture is discussed. In Section 4, the results are discussed. Here we have considered multiple parameters to evaluate our system. Finally, in section 4, we have concluded stating the worth of our research.

2. Literature Survey

In this section, we have discussed the various system and methodologies proposed out by various researchers. NutriNet [1] is based on deep CNN architecture [2,3]. It uses about 520 food images of different food classes, and provides an accuracy of around 86.72%, along with an accuracy of 94.47% when tested for 130,517 images. The NutriNet can identify food and drink item when presented separately. Identifying the two or more food item in a single image is not possible. Also, for noisy or blur input image, it fails to give an accurate result. Bossard et al. [4] have implemented the process of food capture by mining only the discriminative parts with the help of Random Forests, due to which it can mine all classes at the same time also sharing knowledge among them. It uses components that are patches aligned to the image superpixel. The Dataset used in this model has 101 food categories within a total of 101000 images of food. This proposed model provides an accuracy of 50.76% which is better than alternative methods of classification except for CNN. The key features of their studies are food identification through Random Forests Mining, superpixel based patch sampling which reduces the number of sliding windows, a free to use and an extensive dataset for food recognition, it outperforms Improved Fisher Vectors classifier. Another system [5] uses Fuzzy C-means Clustering for Segmentation and Morphological operations to identify and measure the volume, mass, and density of food items in an image. The calorific value is then calculated in the MatLab environment. The process incorporates Fuzzy C-implies Clustering Segmentation which permits every data point to have a place with various groups with differing degrees of membership. Then it is gone through Morphological activities, boundary extraction, enlargement, disintegration, and highlight extraction. This system can identify the raw vegetables and fruits. It does not deal with the cooked food item. Also, it specifies the calorific value of each item separately and does not give the composite calorific value of all items present in the dish.

An indigenous programming instrument to quantify calorie and supplement by utilizing a cell phone is proposed by Pouladzadeh [6], which uses Gabor Filter for a surface division of pictures. It shows an accuracy of around 86% in non-blended food items. It utilizes 3000 images captured under various conditions. It categorizes food items in multiple subcategories, like solid or fluid food, and blended or unmixed food. In another approach given in [7], the food item is classified through SVM. But the CNN indicated altogether higher accuracy than other vector-machine-based strategies with handmade features and accomplished precision of over 70%. Amatul et al. [8] have utilized a pre-prepared Convolutional Neural Network (CNN) as a component extractor to prepare a picture classification classifier. A multiclass straight Support Vector Machine (SVM) classifier developed with separated CNN highlights is utilized to characterize cheap food pictures to ten distinct classes. They have achieved a pace of 89.5%, which is higher than the exactness accomplished using a sack of highlights (BoF) and SURF. The scope of their study is limited to fast food only, and they have considered an image containing only one fast food item.

Another system is proposed [9] to build up an application for evaluating food calories and improve individuals' wellness. It has 79.2% grouping precision. Convolutional Neural Network controlled this framework. The strategy for food detection is applied through Tensor Flow. The output is produced by characterizing the picture and gives helpful data to clients, for example, Name, Calories, and Nutrition. This study is limited to fruits only. Park et al. [10] have a developed Korean food detection application using deep CNN for the mind-boggling acknowledgement model. It had a test accuracy of 91.3%. The focus of this study is on Korean food only. Another methodology [11] is based on the utilization of Faster R CNN, which is used to recognize a specific dish through a captured image and the calories in each distinguished dish are evaluated. It utilizes UEC FOOD-1003, which is a Japanese nourishment photograph dataset having multi-name pictures. This Dataset incorporates more than 100 single-mark photos for every classification and afterwards 11566 single-name images in total and can distinguish food items with 90.7% accuracy. It utilized 80% of the food images in the manufactured Dataset for preparing, and the rest 20% for execution assessment. The Personalized Classifier [12] for food image recognition actualizes a personalized structure, called a sequential personalized classifier that is a blend of closest predicted classes. It utilizes 1,508,171 images transferred by 20,820 clients from the overall population. This study involves an assigned class NCM classifier that performs exactly with CNN and a client explicit classifier which then learns the client's information steadily, assigns more weight to the next info. As opposed to existing examinations directed utilizing counterfeit situations, it acquainted another dataset with assessing customized grouping execution in sensible circumstances. It further designs to quicken this work on personalization by using data from differ rent clients whose marking inclinations are like that of the objective client.

Table 1 summarizes the literature survey discussed so far. For summarization, the dataset used, implemented methodology, type of content in input image and accuracy is considered. From the table, most of the researcher has used the CNN approach for food identification. Few researchers have used the input image with a single food item while others have an image with two or more food item and predicted their names. We observed that there scope of designing a system which can predict the name of the food item with higher accuracy when multiple items with similar appearance are present in the dish. We took this challenge and designed a model for the same.

Table 1 Summary of the Literature Survey

Sr. No.	Author name	Year of publication	Dataset	Methodology	Accuracy	Input image contains single/Multiple food item
1.	Simon Mezgec et. al. [1]	2017	225,953 512 × 512 pixel images of 520 different food	NutriNet AlexNet, GoogLeNet and ResNet	86.72%,	Single
2.	Lukas Bossard et. al. [4]	2014	101 food categories, with 101'000 images.	Random Forests (rf) including SVM classification	50.76%,	Single
3.	R.Kohila et. al. [5]	2017	NA	Fuzzy c-means clustering.	82%	Multiple
4.	Parisa Pouladzadeh et. al. [6]	2014	3000 images.	Gabor filter	86%	Multiple
5.	Kiyoharu Aizawa et. al. [7]	2014	85 food items, Japanese food images	CNN	70%	Single
6.	Amatul et. al [8]	2018	4545 images of 101 different food items,	Multiclass linear SVM classifier trained with extracted CNN	89.5%	Multiple
7.	Anita Chaudhari et. al [9]	2019	1000 images of per object .	CNN, Deep Learning, Image Recognition.	79.2%	Multiple
8.	Seon-Joo Park et. al. [10]	2019	92,000 images categorized in 23 groups of Korean food.	Deep Convolutional Neural Networks (DCNN)	91.3%	Single
9.	Takumi Ege et. al. [11]	2017	UEC FOOD-1003 Japanese food photo	Faster R-CNN	90.7%	Multiple
10.	Sosuke Amano [12]	2018	1,508,171 food images uploaded by 20,820 users	sequential personalized classifier (SPC) 1-nearest neighbor	71%	Single

3. Implementation Overview

Figure 1 depicts the flow of the overall process which comprises three steps viz.

- Preprocessing,
- Training and Validation,
- Testing.

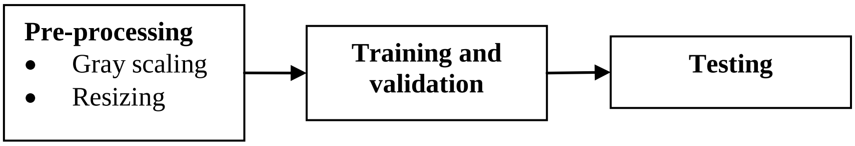


Figure1 Implementation Overview

3.1 Pre-processing

- The fundamental aim of this step is extracting the most relevant data from the input image. It also helps to curtail the parameters of data, consequently resulting in facilitating the learning process by providing the necessary attributes and getting rid of the majority of the outliers hence enhancing the results. Thus this step focuses on identifying and filtering the essential attributes that will be considered during the training of the model. These attributes must be chosen such that it accurate detection of meals when selected while filtering out the majority of the noise. The selection of features should be made such that those that contribute to precise food detection are retained, thus resulting in the elimination of noise to a great extent. The preprocessing step comprises the following sub-steps:
- Reading the images: Firstly, the images are loaded from a directory that forms the training and validation dataset.
- Grayscaling the images: Next, the images are scanned in a grayscale format. This helps in reducing the complexity compared to colour images for processing.
- Resizing the images: Then, the dimensions of all the input images are normalized. All the images are resized to 50*50 pixels. The attributes extracted from this are then stored in an array. It's accompanied by their respective labels and stored into a different array. This is used as the training dataset.

3.2 Training and Validation

The next step is to train the model. For this, there are different methods, each leverage the output of the model and also its accuracy on the testing dataset. Parameters like the size of input batch, number of epochs, etc. highly influence the result of the output model. Thus we divided the training step into three parts, first shuffle the data, then pick a random batch out of it and fit the model in the last step.

- Shuffle Data: Shuffling ensures that during training, the model reads the data in a random order, thereby cutting down the bias in the learning phase. This also provides a significant improvement in efficiency.
- Select a random batch: Selecting a batch randomly ensures the elimination of bias due to order. For that, we randomly pick a batch of 32. The batch is then processed through the model and after every batch of iteration, the model is rectified and hence it tries to cut down the error throughout iterations of a batch. This helps the model in learning at a fast pace and decreases the overall learning and building time of the model.

Fit Model: Figure.2 represents the architecture of our model. It comprises multiple convolutional layers to elicit the features. The softmax layer then predicts the food image classification. The model is fed with input images of size 50*50. A convolutional layer of 32 filters of size 3x3 forms the initial layer, followed by the second layer which is a max pooling of order 2x2, and a convolutional layer of 64 filters of order 3x3 which forms the third layer. Another max-pooling layer of order 2x2 forms the fourth layer. The fifth and sixth layers are similar to the third and fourth layers. At each layer rectified linear unit (ReLU) [13, 14, 15] is used as the activation function. To avoid overfitting, a dropout layer of 0.2 is applied after every max pooling layer.

The next two layers are hidden dense layers, each one followed by a dropout layer of 0.2. The features extracted from the above CNN helps the model to learn with softmax activation in the final layer. Cross-entropy loss function is used for our training.

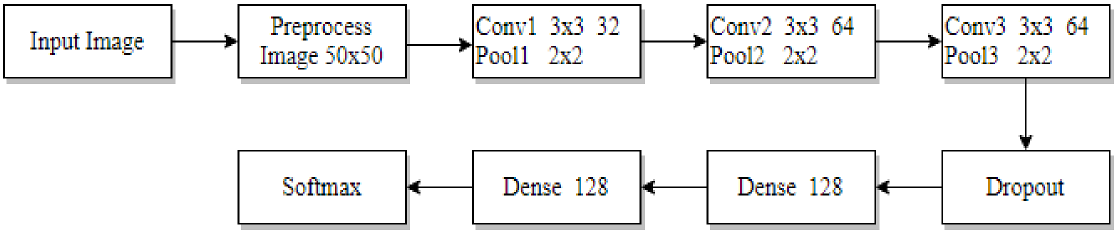


Figure.2 Model Architecture

In validation, the trained model's performance will be evaluated against a dataset the model hasn't seen before, i.e. the testing dataset. The primary data set is split into two parts, the former used in training and the latter in testing. We had specified a validation split of 0.1 during our training process. The testing dataset is used to identify and test the ability of the model to classify food images that weren't present in the training dataset.

3.3 Testing

In this phase, we check the performance of the model on a real-world (unseen) data. The success of the model is measured based on how well it performs in the testing phase. In this step, we provide an image of a food item and get its name based on the prediction of the model. The primary subtasks of this phase are as follows:

- Preprocessing: The sequence of input data is preprocessed by passing it through a series of preprocessing steps discussed above.
- Apply Model: The preprocessed image is given to the model, and its prediction is recorded. The probability distribution of prediction spans among ten classes; consequently, the food item with the most considerable probability is taken as predicted food. Consider an example of an image containing dal and rice. This image is fed to the model and based on the features extracted by the model, and it gives a list of values, each value corresponding to the 10 food items in our dataset. From this list of values, the index of the maximum value is selected. Now that index is mapped with the food item in the categories list to find the name of the predicted food item and thus the cuisine is predicted. For our example, the index that was returned is 1. And the index of 1 corresponds to the food item Dal and Rice, which is the correct prediction in this case. The output is shown in Figure 3.

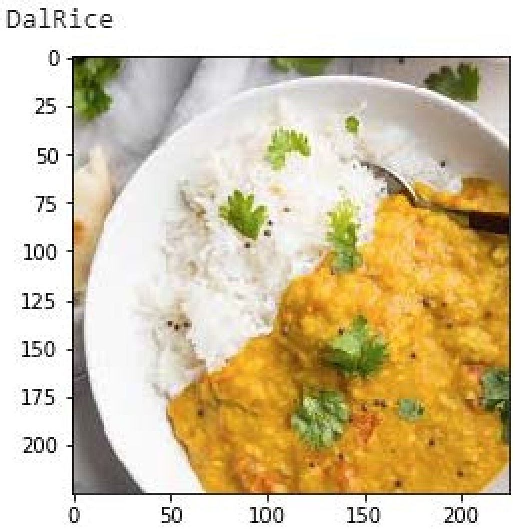
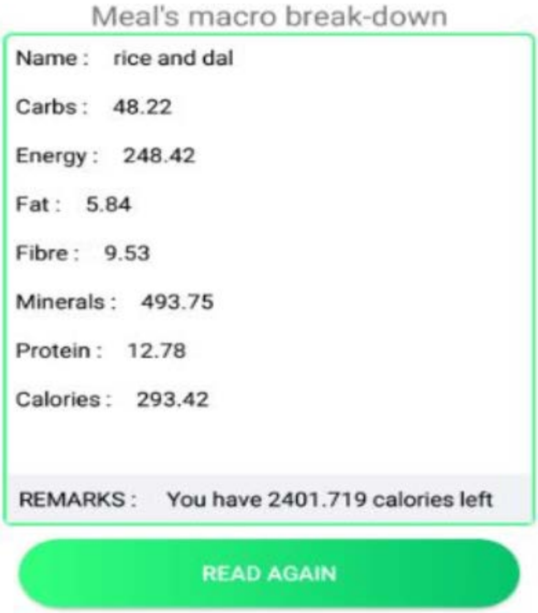


Figure.3 Cuisine Prediction


Meal's macro break-down

Name : rice and dal

Carbs : 48.22

Energy : 248.42

Fat : 5.84

Fibre : 9.53

Minerals : 493.75

Protein : 12.78

Calories : 293.42

REMARKS : You have 2401.719 calories left

Figure.4 Nutrient Value Prediction

Once the cuisine name is predicted, the breakdown of the nutrients for that meal will be calculated using Nutritionx [16], which is a natural language processing API. As shown in Figure 3, we got composite cuisine as dal and rise. These names are passes to fetch the nutrition value. We have also extracted the approximate quantity of cuisine and then protein, carbohydrate, fat, calorie etc. are displayed, as shown in Figure 4. Along with this in Figure 4, there is a remark stating the calories left. For this in the beginning user's body mass index is calculated and accordingly, daily calories consumption is predicted. So after every food consumption remark is generated for remaining calorie consumption.

4. Results

For our system, we prepared a custom dataset of images of food items. For this, we collected 100 images each of 10 dishes and then we used this dataset to train our model.

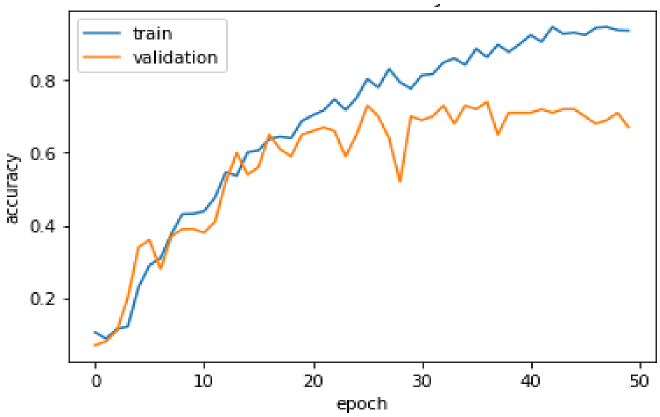


Figure 5 Model Accuracy

Out of these 1000 images, 900 were used for training, and the remaining 100 were used for validation. We trained our model with 50 iterations at a default learning rate. Figure 5 shows the plot of accuracy during training and validation phases of the model. The plot shows a training accuracy of 99.68% and a validation accuracy of 73%. The accuracy changes over the 50 iterations are shown in the Figure 5.

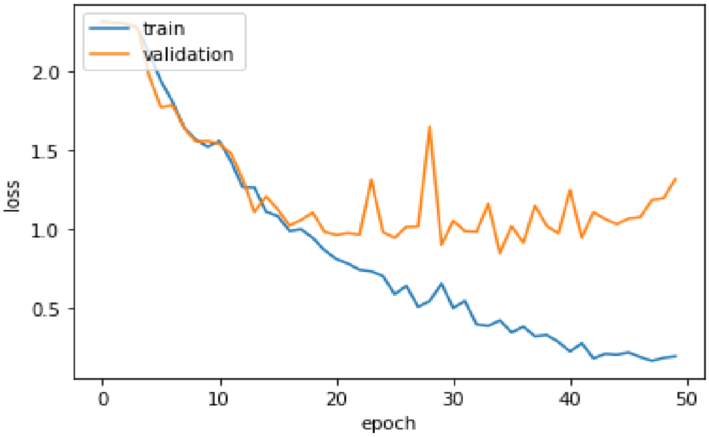


Figure. 6 Model Loss

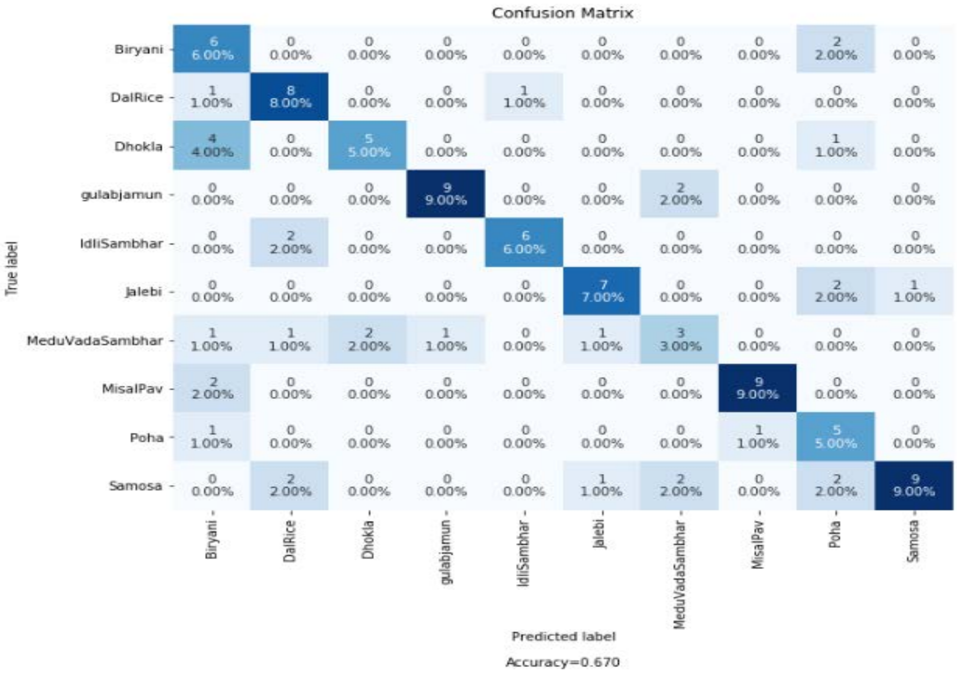


Figure 7 Confusion Matrix

Figure. 6 shows the plot for loss of model during the two phases (training and validation). The plot shows a loss of 0.0164 during the training phase and 0.91 during the validation phase. The changes in model loss over the 50 iterations are shown in the figure 6. Some other parameters related to the model were also calculated such as the average precision which was found out to be 0.72. Similarly, the average recall and average F1 score were 0.67 and 0.68 respectively. The performance of the model on the test data was evaluated by plotting a confusion matrix which is shown in Figure 7. It is a 10X10 matrix. The confusion matrix shows the predicted labels for each of the food item images. The values along the diagonal of the confusion matrix are the True Positives which signifies the percentage of the food item being identified correctly. So by taking the summation of the diagonal value for each class (here each food item represents the individual class) the accuracy of the model is calculated as 0.67 or 67%.

5. Conclusion and Future Scope

The present work offers two-fold objectives of recognition of food items on plate and calculation of the nutritional value of these items. The work is unique in its way for mentioned objectives as there is no such combined attempt discussed in the literature to the best of the author's knowledge. Also, work can be extended by integrating the user's diet plan and suggesting whether the user should consume food items on a plate or not. In this paper, we have discussed the Convolutional Neural Network Based model to which one needs to feed the preprocessed, resized, grayscale cuisine image as an input. The model then predicts the dishes present in the image and then the system calculated the nutrition parameters like carbohydrate, protein, fat, and calories. For evaluating the model, we have accuracy, loss graph and also considered the confusion matrix. The training accuracy of our model is 99.8%. Average precision, recall, and F1 score is 0.72, 0.67 and 0.68, respectively.

Looking at the diverse food culture all over the world, in the future, we will work on different classes of food so that our system can identify any dish over the globe. Also, predicting the exact quantity of food is a challenge that needs to be addressed in the future.

References

[1] P. Chen, Y. Lu, V. W. Zheng, X. Chen, and B. Yang, “KnowEdu : A System to Construct Knowledge Graph for Education,” IEEE, vol. 6, pp. 31553–31563, 2018. Pawade D., Sakhapara A., Shah C., Wala J., Tripathi A., Shah B. (2019) Text Caption Generation Based on Lip Movement of Speaker in Video Using Neural Network. In: Advances in Computing and Data Sciences. ICACDS 2019. Communications in Computer and Information Science, vol 1046. Springer, Singapore

[2] Pawade, A. Sakhapara, M. Jain, N. Jain and K. Gada, "Story Scrambler – Automatic Text Generation Using Word Level RNN-LSTM", International Journal of Information Technology and Computer Science (IJITCS), Vol.10, No.6, pp.44-53, 2018. DOI: 10.5815/ijitcs.2018.06.05

[3] Mezgec, B. Koroušić Seljak, "Nutrinet: A deep learning food and drink image recognition system for dietary assessment", Nutrients, vol. 9, no. 7, pp. 657, 2017.

[4] Bossard L., Guillaumin M., Van Gool L. (2014) Food-101 – Mining Discriminative Components with Random Forests. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8694. Springer, Cham.

[5] R. Kohila and R. Meenakumari, "Predicting calorific value for mixed food using image processing," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2017, pp. 1-4.

[6] P. Pouladzadeh, S. Shirmohammadi and R. Al-Maghrabi, "Measuring Calorie and Nutrition From Food Image," in IEEE Transactions on Instrumentation and Measurement, vol. 63, no. 8, pp. 1947-1956, Aug. 2014.

[7] Kagaya, Hokuto & Aizawa, Kiyoharu & Ogawa, Makoto. (2014). Food Detection and Recognition Using Convolutional Neural Network. 10.13140/2.1.3082.1120.

[8] Amatul Bushra Akhi, Farzana Akter, Tania Khatun & Mohammad Shorif Uddin “Recognition and Classification of Fast Food Images” Global Journal of Computer Science and Technology 2018

[9] Anita Chaudhari, Shraddha More, Sushil Khane, Hemali Mane, Pravin Kamble “Object Detection using Convolutional Neural Network in the Application of Supplementary Nutrition Value of Fruits”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-8 Issue-11, September 2019.

[10] Seon-Joo Park, Akmaljon Palvanov, Chang-Ho Lee, Nanoom Jeong, Young-Im Cho, and Hae-Jeung Lee, “The development of food image detection and recognition model of Korean food for mobile dietary management”, Nutr Res Pract. 2019 Dec; 13(6): 521–528.

[11] Ege, T., Yanai, K.: Estimating food calories for multiple-dish food photos. In: Proceedings of Asian Conference on Pattern Recognition (ACPR) (2017).

[12] Horiguchi, S. Amano, M. Ogawa, and K. Aizawa. 2018. Personalized classifier for food image recognition. IEEE Trans. Multimed. 20, 10 (2018), 2836—2848

[13] Gangming Zhao, Zhaoxiang Zhang, He Guan, Peng Tang, Jingdong Wang, “Rethinking ReLU to Train Better CNNs”, 31 Aug 2018. Available on : https://arxiv.org/pdf/1709.06247.pdf

[14] Abien Fred M. Agarap, “Deep Learning using Rectified Linear Units (ReLU)” 7 Feb 2019.Available on : https://arxiv.org/pdf/1803.08375.pdf

[15] Ide, Hidenori & Kurita, Takio. (2017). Improvement of learning for CNN with ReLU activation by sparse regularization. 2684-2691. 10.1109/IJCNN.2017.7966185.

[16] I https://www.nutritionix.com/

Author's Profile

Dipti Pawade received B.E. degree in Computer Science and Engineering from Sant Gadge Baba University in 2009 and M.E. degree in Embedded System and Computing from G. H. Raisoni College of Engineering, Nagpur in 2012. Since 2012 she is an Assistant Professor in the Department of Information Technology at K J Somaiya College of Engineering, Vidyavihar, Mumbai. Her interest includes Machine learning, Web Security, and Web Application Development.

Ashwini Dalvi is pursuing her PH.D. Degree from VJTI, affiliated to Mumbai University. She joined the Department of Information Technology, K. J. Somaiya College of Engineering, Mumbai in 2006 as an Assistant Professor. She has published over 25 journal and conference papers in the areas of Security, Intelligent applications.

Irfan Siddavatam has received the PH.D. Degree from VJTI, affiliated to Mumbai University, in 2018. In 2001, he joined the Department of Information Technology, K. J. Somaiya College of Engineering, Mumbai, as an Associate Professor. His research interests include Cyber Physical System Security, Artificial Intelligence and Internet of Things.

How to cite this paper: Dipti Pawade, Ashwini Dalvi, Dr. Irfan Siddavatam, Myron Carvalho, Prajwal Kotian, Hima George. "Cuisine Detection Using the Convolutional Neural Network ", International Journal of Education and Management Engineering(IJEME), Vol.10, No.3, pp.1-11, 2020. DOI: 10.5815/ ijeme.2020.03.01