# Automated Food image Classification using Deep Learning approach

Sapna Yadav
Jawaharlal Nehru University
New Delhi, India
sapnayadav1990@gmail.com

Alpana
Jawaharlal Nehru University
New Delhi, India
alpana.srk@gmail.com

Satish Chand
Jawaharlal Nehru University
New Delhi, India
schand20@gmail.com

2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) | 978-1-6654-0521-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICACCS51430.2021.9441889

***Abstract*- Food image classification is an emerging research field due to its increasing benefits in the health and medical sectors. For sure, in the future automated food recognition tools will help in developing diet monitoring systems, calories estimation and so on. In this paper, automated methods of food classification using deep learning approaches are presented. SqueezeNet and VGG-16 Convolutional Neural Networks are used for food image classification. It is demonstrated that using data augmentation and by fine-tuning the hyperparameters, these networks exhibited much better performance, making these networks suitable for practical applications in health and medical fields. SqueezeNet being a lightweight network, is easier to deploy and often more desirable. Even with fewer parameters, SqueezeNet is able to achieve quite a good accuracy of 77.20%. Higher accuracy of food image classification is further achieved by extracting complex features of food images. The performance of automatic food image classification is further improved by the proposed VGG-16 network. Due to increased network depth, proposed VGG-16 has achieved significant improvement in accuracy up to 85.07%.**

***Keywords- Food classification, Image processing, Machine Learning, Deep Learning, Transfer Learning, VGG-16, SqueezeNet.***

## I. INTRODUCTION

Automatic food recognition is an emerging research topic not only for the social network domain aspect. Indeed, researchers are focusing on this area because of its increasing benefits for medical point of view. Automatic food recognition tools will help in facilitating the decision-making process of calories estimation, quality detection of food, build diet monitoring systems to combat obesity and so on [1].

On the other hand, food is inherently deformable and shows high divergence in appearance. Since food images have high intraclass variance and low inter-class variance due to which classic approaches do not recognize complex features. This makes food recognition a difficult task for which complex features are not recognized by classic approaches. CNNs can easily identify these features automatically, thus increasing classification accuracy [2]. Therefore, this paper attempts to use CNNs for food image classification. Fig. 1 shows images looking very similar to each other but both images belong to different food classes. This is due to low inter class variance among food items.


(a) Cup-cake (b) Red velvet cake

Fig. 1. Different types of food classes

Convolution Neural Network has basically three layers: convolution layers, pooling layers and fully connected layers. Convolution layer assigns learnable weights and biases to input image. Pooling layer down-samples the input data by summarizing the features thus reducing trainable parameters. At the end fully connected layer is present having full connections to all neurons. Softmax activation function calculated the probability of the image belonging to a particular class.

Since food images have high intraclass variance and low inter-class variance due to which some of the complex features are not recognized by Machine Learning methods, but CNNs can easily identify these complex features. These network models based upon deep learning has achieved significant success by automatically discovering very high-level features, thus increasing classification accuracy. Therefore, the proposed work intends to use CNNs for food image classification. These networks extract features automatically by applying convolution operation in certain layers on the input data using a convolution filter to produce a feature map.

These networks contain millions of parameters and their training needs a huge amount of data and high computational resources. Hence, researchers preferred utilizing pre-trained networks by fine-tuning on domain-specific data. Knowledge learned by the pre-trained models can be utilized on related data using the transfer learning approach [3]. In this paper, food image classification has been investigated using SqueezeNet and VGG-16 models of CNN. These are pre-trained networks trained on more than a million images of the ImageNet dataset consisting of 1000 classes. The learned weights and features from pre-trained deep convolution

978-1-6654-0521-8/21/$31.00 ©2021 IEEE

neural networks have been used in SqueezeNet and VGG-16 on the Food-101 dataset through transfer learning.

To classify image traditional techniques and deep learning techniques are used. Through traditional techniques only basic features of images like color, shape, texture [4] etc. can be detected. Classical machine learning algorithms like support vector machine (SVM) [5], random forests [6] and artificial neural networks [7] can also be implemented for image classification with lesser accuracy compared to deep learning methods. With deep learning techniques deep and complex features can be identified easily making recognition task better. Deep learning techniques like convolution neural network (CNN), transfer learning, data augmentation and deep feature fusion network are used widely [8].

The rest of the paper is organized as follows: In section 2, details of our methodology are presented. Section 3 discusses food the classification models used in this paper. Results from experimentation are illustrated in section 4. Section 5 concludes experimental results.

## II. METHODOLOGY

This section describes the proposed framework to recognize the food items from the food image dataset. Fig. 2 shows the proposed framework for food image classification.

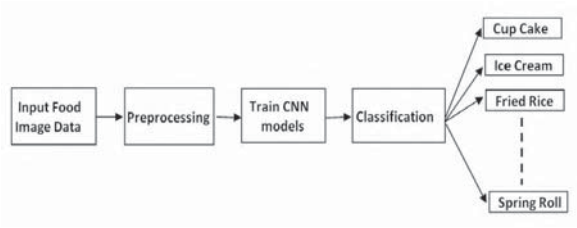


Fig. 2. Proposed framework for food recognition from image dataset.

### A. Dataset

The food-101 dataset contains 101,000 images of 101 categories, each category of food containing 1000 images. 5000 images belonging to 10 food categories have been considered for classification. These 10 food categories from the Food-101 dataset are Cup_cakes, French_fries, Fried_rice, Greek_salad, Ice_cream, Omelette, Pizza, Red_velvet_cake, Samosa and Spring_rolls. These 10 categories are chosen by keeping in mind Indian food items. In this dataset, 30% of training data is used for validation to help to prevent overfitting.

### B. Image Preprocessing

Image preprocessing improves image features by suppressing unwanted distortions and enhancing some important image features. Also, the effectiveness of the dataset can be improved using some data augmentation techniques described below.

- *Data Augmentation:* Data augmentation acts as a regularizer that helps prevent overfitting in neural networks and improves imbalance class problems [9]. Mostly, techniques such as cropping, resizing, rotating, translating and flipping etc., are applied to the original images. Here, rotation range and random reflection techniques have been used, which are briefly described below.

*Rotation Range*: 45 degrees. Images are randomly rotated in this range. This helps to get better performance by considering all appearances of food images for this range.

*Random Reflection*: True, Random X-Translation and Y-Translation in pixel range: [-30 30].

All of the transformations are the affine transformation of the original image that takes the following form:

$$I(x,y) \xrightarrow{T} J(x',y') \qquad (1)$$

A geometric operation transforms a given image $I$ into a new image $J$ by modifying the image points coordinates. T is a mapping function that is a continuous coordinate transform [10].

### C. Network Parameters

SqueezeNet and VGG-16 CNN architectures have been used for training the food dataset. Some of the training parameters taken for both the models for better performance are shown in the table below.

TABLE 1: TRAINING PARAMETERS USED IN SQUEEZENET AND VGG-16

| Parameters | Value |
|---|---|
| Solver | SGDM |
| MiniBatchSize | 64 |
| InitialLearnRate | 0.001 |
| Dropout | 0.5 |

*a) SqueezeNet:* Traditional machine learning classification models are not efficient in recognizing complex image features. CNNs can automatically extract complex features used in computer vision tasks because of a large number of parameters and more depth. SqueezeNet is one of the deep neural networks released in 2016 with the goal of few parameters and small size. Comparing the size of SqueezeNet with AlexNet (winner of ILSVRC 2012), AlexNet has a size of 227 MB with 61.0 million parameters. While SqueezeNet has a much lesser size of 4.60 MB with only 1.24 million parameters having almost the same accuracy on the ImageNet dataset [11].

SqueezeNet has a depth of 18 with 68 layers and 75 connections. This pre-trained network employs design strategies to reduce the number of parameters, notably using fire modules that *squeeze* parameters using [1 x 1] convolutions. Also, this model takes the input of size [227 x 227 x 3] and

[3 x 3] convolution with stride [1 x 1]. For dimension reduction [3 x 3] max pooling with stride [2 x 2] is used. Dropout layers with a probability 0.5 are added after the fire9 module to reduce overfitting. There are no fully connected layers in this network. The softmax layer calculates the probability, followed by the Classification layer, which gives the class output.

SqueezeNet is trained with a learning rate of 0.001 and batch size of 64 with a validation frequency of 1. The network is trained for 810 iterations with a 50% dropout rate. SqueezeNet was designed as a compact replacement for AlexNet, having 50 times fewer parameters, yet it performs almost 3 times faster [12].

*b) VGG-16:* The second model, VGG-16, has been chosen which is the best performing architecture in ILSVRC-2014. It is a deeper yet simpler variant of the convolution neural networks. VGG-16 has a depth of 16 with size 515 MB having 138 million parameters. Having 41 layers, input to this pertained model is images of dimensions [224 x 224 x 3]. Convolution layers have filters of size [3 x 3] with stride [1 x 1]. Max Pooling is performed over a [2 x 2] pixel window with stride [2 x 2]. The last 3 layers are fine-tuned for the new classification problem [13]. ReLU activation function is applied in all hidden layers to decrease the likelihood of vanishing gradient problem. VGG-16 is trained with a learning rate of 0.001 and a batch size of 64 with a validation frequency of 50. The network is then trained for 408 iterations with a 50% dropout rate.

### D. Transfer Learning

The network is not built from scratch rather a pre-trained model will be used applying transfer learning. Transfer learning utilizes the pre-trained network, trained on the ImageNet dataset with the learned weights to obtain features that are used in our dataset [14]. Learning is improved in the new task through knowledge transfer from related tasks. Only the last few layers are trained which make a prediction by recognizing specific features of the images.

### E. Food Classification

The network configuration of CNN layers used in food image classification is as follows:

- *Convolution Layer:* The convolution operation is performed in this layer to extract a feature map. Its parameters include filter size and stride. ***ReLU*** is an activation function used in hidden layers used to make all negative values to zero. It adds non-linearity. ReLU is used because it is easier to train and often achieves better performance [15].

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

- *Pooling Layer:* CNNs are computationally expensive due to a large number of parameters and depth. So, there is a need for dimensionality reduction in between the layers. The pooling layer does this by down sampling operation.
- *Fully Connected Layer:* This layer is usually present towards the end of CNN architecture. This layer converts input into single dimensional vector where each input is connected to all the neurons.

## III. RESULTS

The experiment conducted in this research used MATLAB as a programming language on intel core i7 2.60 GHz processor, NVIDIA GeForce GTX GPU, 8.00 GB RAM on Windows 10 64-bit Operating system.

The SqueezeNet model has been trained with an initial learning rate of *0.001* and batch size of *64* for *810 iterations* on GPU. The SqueezeNet fine-tuned model's performance has been evaluated in terms of classification accuracy. The model has performed much better compared to traditional machine learning models with a training accuracy of 93.47 % and validation Accuracy of 77.20%. Fig. 3 shows the training progress of SqueezeNet classifier.

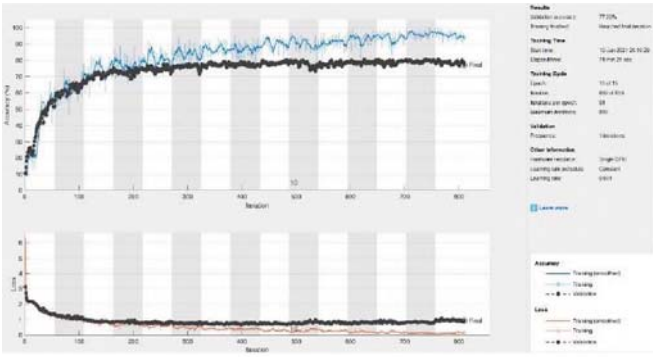

Fig. 3. Training progress of SqueezeNet classifier

The second model chosen for food image classification on the Food-101 dataset of preferred 10 classes of Indian food is VGG-16. It advances the SqueezeNet model with increased classification accuracy. Proposed VGG-16 model has been trained with the same learning rate and a batch size of 0.001 and 64 respectively as SqueezeNet. The model has been trained for 408 iterations on CPU with a validation frequency of 50 for 823 minutes and achieved a classification accuracy of 85.07%. Proposed VGG-16 showed much improvement in the accuracy due to the increased depth of the network with a higher number of parameters. Fig. 4 shows the training progress of proposed VGG-16 classifier.

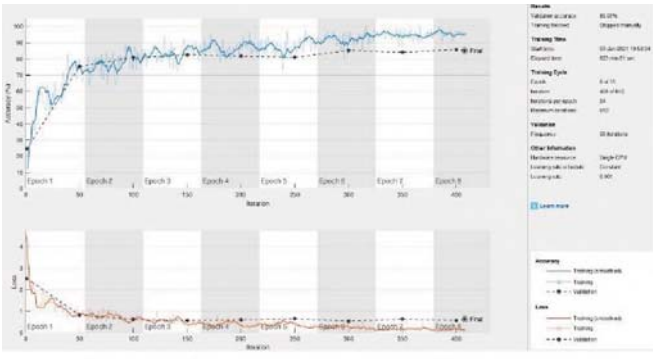

Fig. 4. Training progress of proposed VGG-16 classifier

Fig. 5 shows the predicted class labels of randomly chosen food images

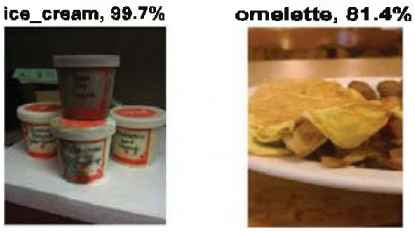


Fig. 5. Predicted class labels of randomly chosen food images

## IV. CONCLUSION

In this paper, automatic food image classification techniques based on deep learning approaches have been presented. Better performance of food image classification was achieved by extracting high-level complex features. For this, the SqueezeNet and VGG-16 deep learning models have been used. In designing these networks, data augmentation techniques have been used and hyperparameters were fine-tuned to improve network performance. It is observed that SqueezeNet having a much lesser model size and fewer parameters performed well with an accuracy of 77.20%. As compared to SqueezeNet, proposed VGG-16 is a deeper network with more parameters. Therefore, proposed VGG-16 has achieved much better performance and was able to classify food images more accurately with higher accuracy of 85.07%.

TABLE 2: TRAINING AND VALIDATION RESULTS OF SQUEEZENET AND PROPOSED VGG-16

| Models | Training Accuracy | Validation Accuracy | Validation Loss |
|---|---|---|---|
| SqueezeNet | 92.83% | 77.20% | 0.9490 |
| **VGG-16** | **94.02%** | **85.07%** | **0.7435** |

## Acknowledgment

The authors would like to thank Vivek Yadav (Programmer, DoIT), Prativa Das (Research scholar, JNU).

## References

[1] Zhou, L., Zhang, C., Liu, F., Qiu, Z., & He, Y, "Application of Deep Learning in Food: A Review," Comprehensive Reviews in Food Science and Food Safety, vol. 18, pp. 1793-1811, 2019.

[2] Farinella, G. M., Moltisanti, M., & Battiato, S., "Classifying food images represented as Bag of Textons," IEEE International Conference on Image Processing (ICIP), Paris, pp. 5212-5216, doi: 10.1109/ICIP.2014.7026055, 2014.

[3] Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A., "Learning deep features for scene recognition using places database," Proceedings of the 27th International Conference on Neural Information Processing Systems, vol. 1, pp. 487-495, ACM, 2014.

[4] Rahmani, G. A.,"Efficient Combination of Texture and Color Features in a New Spectral Clustering Method for PolSAR ImageSegmentation,"National Academy Science Letters, vol. 40, pp. 117–120, 2017, https://doi.org/10.1007/s40009-016-0513-6.

[5] Wang, M., Wan, Y., Ye, Z., & Lai, X.,"Remote sensing imageclassification based on the optimal support vector machine andmodified binary coded ant colony optimization algorithm,"Information Sciences, vol. 402, pp. 50-68, 2017, https://doi.org/10.1016/j.ins.2017.03.027.

[6] Xia, J., Ghamisi, P., Yokoya, N., & Iwasaki, A., "Random Forest Ensembles and Extended Multiextinction Profiles for Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 56 , pp. 202-216, 2018, doi:10.1109/TGRS.2017.2744662.

[7] Kaymak, S., Helwan, A., & Uzun, D., "Breast cancer image classification using artificial neural networks," Procedia Computer Science, vol. 120, pp. 126-131, 2017, https://doi.org/10.1016/j.procs.2017.11.219.

[8] Chaib, S., Liu, H., Gu, Y., & Yao, H., "Deep Feature Fusionfor VHR Remote Sensing Scene Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 55, pp. 4775-4784, 2017, doi:10.1109/TGRS.2017.2700322.

[9] Simard, P. Y., Steinkraus, D., & Platt, J. C., "Best Practices for Convolutional Neural Networks," 12th International Conference on Document Analysis and Recognition, vol. 2. IEEE Computer Society, 2003.

[10] Bazargani, Anjos, M. &., Lobo, A. &Mollahosseini, F. &,Shahbazkia, A. &, & Hamid, "Affine Image RegistrationTransformation Estimation Using a Real Coded," Proceedings of the 14th annual conference companion on Genetic and evolutionary computation, pp. 1459–1460, ACM, 2012, https://doi.org/10.1145/2330784.2330990).

[11] Keutzer, F. N., "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," ICLR, 2016.

[12] Kurama, V. (2020, June 5). *A Review of Popular Deep Learning Architectures: ResNet, InceptionV3, and SqueezeNet.* Retrieved January 25, 2021, June 5, 2020 from PaperspaceBlog: https://blog.paperspace.com/popular-deep-learning-architectures-resnetinceptionv3squeezenet/#:~:text=The%20SqueezeNet%20architecture%20is%20comprised,3%20%C3%97%203%20convolution%20filters.&text=Meanwhile%2C%20a%20Deep%20Compression%20SqueezeNet,and%20a.

[13] Simonyan, K., & Zisserman, A., "Very Deep Convolutional Networks for Large-Scale Image Recognition," ICLR. arXiv, 1409.1556, 2015.

[14] Ghazia, M. M., Yanikoglu, B., & Aptoula, E., "Plant identification using deep neural networks via optimization of transfer learning parameters," Neurocomputing, vol. 235, pp. 228-235, 2017, https://doi.org/10.1016/j.neucom.2017.01.018.

[15] Brownlee, J., "A Gentle Introduction to the RectifiedLinear Unit (ReLU)," retrieved January 24, 2021, from MachineLearning Mastery:https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural networks/#:~:text=The%20rectified%20linear%20activation%20function,otherwise%2C%20it%20will%20output%20zero.