International Journal of applied exercise physiology

ISSN: 2322-3537
Vol-14 Issue-01 June 2025

Machine Learning-Based Automated Food Image Classification

[1] P. Premchand, [2] M. Soumya

[1]Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.
[2] MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

Abstract-

The growing number of applications of food picture categorization in healthcare has made it a promising new area of study. Diet tracking systems, calorie estimators, and similar applications will undoubtedly benefit from automated food identification techniques in the future. The article introduces automatic food categorization systems that make use of deep learning techniques. For the purpose of food picture categorization, SqueezeNet and VGG-16 Convolutional Neural Networks are used. It is shown that these networks performed much better after data augmentation and hyperparameter tweaking, which makes them applicable to real-world medical and health applications. since of its small size, SqueezeNet is often preferred since it is simpler to set up. With only 77.20 percent of the parameters used, SqueezeNet still manages to get a respectable result. By extracting complex aspects of food photographs, we may further obtain higher accuracy in food image categorization. With the suggested VGG-16 network, automated food picture categorization becomes much more efficient. With its deeper network, the projected VGG-16 has significantly improved accuracy, reaching 85.07%. cake in a cup cake made with red velvet

Introduction

Research into automatic food identification is a relatively new area of study, and it's not only in the realm of social networks. Indeed, this field is receiving a lot of attention from academics due to the growing number of medicinal advantages associated with it. Diet monitoring systems to battle obesity, automatically identify food quality, and aid in calorie calculation are just a few examples of how automatic food identification technologies will revolutionize the food industry [1]. Contrarily, food exhibits a great deal of visual variation and is intrinsically malleable. Traditional methods fail to identify intricate details in food photos because of their large intraclass variation and little interclass variance. Because of this, food identification is a challenging problem, and traditional methods fail to identify complex aspects. Convolutional neural networks (CNNs) are able to automatically detect these properties, leading to improved classification accuracy [2]. This work so endeavors to use CNNs for the categorization of food images. Even though they come from distinct food groups, the two photos in Fig. 1 seem quite similar. Reason being, there isn't much variation among food classes.

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

Fig. 1. Different types of food classes

Convolution, pooling, and fully linked layers make up a convolution neural network. The convolutional layer takes an input picture and applies biases and learnable weights to it. In order to reduce the number of trainable parameters, the pooling layer down-samples the input data by aggregating the features. A completely connected layer, which has connections to every neuron in the network, is present at the very end. The likelihood of an image's class membership

International Journal of applied exercise physiology
ISSN: 2322-3537
Vol-14 Issue-01 June 2025

was determined using the Softmax activation function. Food categorization, picture processing, ML, DL, TL, VGG-16, SqueezeNet are some of the keywords.

Convolutional neural networks (CNNs) are able to detect intricate elements in food photographs that Machine Learning approaches miss because of the images' high intraclass variation and low inter-class variance. By enhancing classification accuracy via the automated discovery of extremely high-level features, these deep learning-based network models have achieved great success. Consequently, convolutional neural networks (CNNs) are going to be used for food picture categorization in our suggested study. A feature map is automatically generated by these networks by performing a convolution operation at certain layers to the incoming data using a convolution filter. Training one of these networks requires a mountain of data and powerful computers due to the millions of parameters it contains. Researchers favored using pre-trained networks that had been fine-tuned using domain-specific data. By using the transfer learning strategy, previously trained models may be applied to new, relevant data [3]. Using the SqueezeNet and VGG-16 convolutional neural network (CNN) models, this research explores food picture categorization. These networks have already been trained using the ImageNet dataset, which contains over a million pictures and 1000 classifications. For the Food-101 dataset, SqueezeNet and VGG-16 have used transfer learning to apply the acquired weights and features from pre-trained deep convolution neural networks. Both conventional and deep learning methods are used for picture classification. Images can only have their most fundamental characteristics, such as color, form, texture [4], etc., identified using conventional methods. Image categorization using deep learning techniques is more accurate than using classical machine learning algorithms such as support vector machines (SVMs) [5], random forests [6], and artificial neural networks [7]. Deep learning algorithms greatly improve identification tasks by making it easy to identify deep and complicated characteristics. Common deep learning methods include CNNs, TL, data augmentation, and DFRNs (deep feature fusion networks) [8]. Here is how the remainder of the paper is structured: Our technique is laid forth in section 2. The classification models used in this work are covered in Section 3. Section 4 displays the results of the experiments. Section 5 presents the findings from the experiments.

Methodology

In this section, the suggested framework for food item recognition from the food picture collection is detailed. Figure 2 displays the suggested structure for food picture categorization.

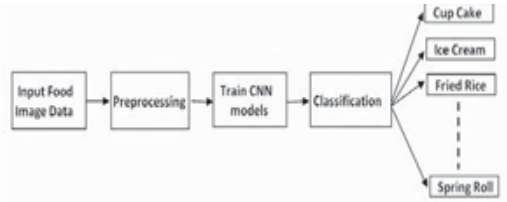


Fig. 2. Proposed framework for food recognition from image dataset.

Table A. Each of the 101 food categories in the food-101 dataset has 1000 photographs. In all, there are 101,000 images in this collection. For this categorization, 5,000 photos from 10 different food categories were taken into account. The Food-101 dataset includes ten food categories: Cup cakes, French fries, fried rice, Greek salad, ice cream, omelette, pizza, red velvet cake, samosas, and spring rolls. With Indian cuisine in mind, we settled on these 10 categories. For the purpose of avoiding overfitting, validation accounts for 30% of the training data in this dataset. A. Image Pretreatment By removing artifacts and boosting key details, picture preprocessing makes a big difference in the final product. Some data augmentation approaches are also detailed below, which may increase the dataset's efficacy.

Enhancement of Data: In order to address imbalance class issues and avoid neural network overfitting, data augmentation is used as a regularizer [9]. Many common editing methods are used on the source photographs, including cropping, resizing, rotating, translating, and flipping. The following is a synopsis of the methods used: rotation range and random reflection. Angle of Rotation: 45\. The images in this range are rotated at random. Taking into account all instances of food photos within this range helps to get better performance. X-Translation and Y-Translation generated at random between the range of [-30 30]

International Journal of applied exercise physiology
ISSN: 2322-3537
Vol-14 Issue-01 June 2025

pixels; random reflection: true. Every one of these changes is an affine transformation of the initial picture, and it looks like this:

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

In order to create a new picture J from an existing one I, a geometric operation changes the coordinates of the image points. According to [10], T is a continuous coordinate transform that maps to other functions. Section C: Network Settings When training on the food dataset, SqueezeNet and VGG-16 CNN architectures were used. The following table displays some of the training settings used to improve the performance of both models. The training parameters used by SqueezeNet and VGG-16 are shown in Table 1.

Parameters	Value
Solver	SGDM
MiniBatchSize	64
InitialLearnRate	0.001
Dropout	0.5

When it comes to identifying complicated picture attributes, SqueezeNet shows that traditional ML classification models fall short. Due to their extensive parameter sets and increased depth, CNNs are able to autonomously extract complicated features used in computer vision applications. Among the deep neural networks introduced in 2016, SqueezeNet aimed to minimize size and parameters. The winner of ILSVRC 2012, AlexNet, has 61.0 million parameters and a size of 227 MB, far larger than SqueezeNet. On the other hand, SqueezeNet maintains almost same accuracy on the ImageNet dataset while having a much smaller size of 4.60 MB and just 1.24 million parameters [11]. The 68 layers and 75 connections make up SqueezeNet's 18-depth architecture. Using design methods, such as fire modules that compress parameters using [1 x 1] convolutions, this pre-trained network reduces the amount of parameters. The input dimensions of this model are [227 x 227 x 3] as well.

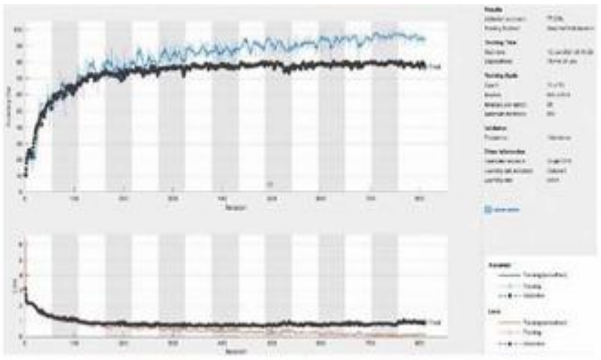
Three-by-three convolution with a one-by-one stride. Reducing dimensions to [3 x 3] requires max pooling with a stride of [2 x 2]. In order to decrease overfitting, dropout layers are inserted after the fire9 module with a probability of 0.5. This network does not have any completely linked layers. The probability is calculated by the softmax layer, and the class output is provided by the classification layer. SqueezeNet is trained using 64 batches and a validation frequency of 1 using a learning rate of 0.001. With a 50% dropout rate, the network is trained for 810 iterations. Developed as a more efficient alternative to AlexNet, SqueezeNet boasts three times the performance with fifty times less parameters [12]. b) VGG-16: The top-performing architecture in ILSVRC-2014, the second model, has been selected. The convolution neural networks have been simplified while being made deeper. With 138 million parameters, VGG-16 is 515 MB in size and has a depth of 16. Pictures of dimensions [224 x 224 x 3] are fed into this pertained model's 41 layers. Size [3 x 3] and stride [1 x 1] are the filters used in convolution layers. A 2x2 pixel window with a 2x2 stride is used for Max Pooling. To address the novel categorization issue, the last three layers are fine-tuned [13]. All hidden layers use the ReLU activation function to reduce the probability of the vanishing gradient issue. For VGG-16's training, we used a validation frequency of 50, a batch size of 64, and a learning rate of 0.001. Next, we train the network for 408 iterations, dropping out 50% of the time. D. Knowledge Transfer Instead of starting from scratch, the network will use transfer learning to apply to a pre-trained model. As an example, our dataset makes use of features derived from a pre-trained network that was trained using the learnt weights on the ImageNet dataset [14]. Knowledge transfer from related tasks improves learning in the new task. Predictions are only made by the last few layers that have been taught to recognize certain visual aspects. E. Categorization of Foods This is the setup of the convolutional neural network (CNN) layers used for food picture categorization: at order to get a feature map, the convolution process is carried out at the convolution layer. The stride and filter size are its parameters. To convert all negative values to zero, hidden layers employ the ReLU activation function. It introduces non-linearity. Because it is often more effective and simpler to teach, ReLU is commonly utilized [15].

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

Pooling Layer: Convolutional neural networks (CNNs) have a lot of parameters and depth, which makes them computationally costly. Hence,

International Journal of applied exercise physiology
ISSN: 2322-3537
Vol-14 Issue-01 June 2025

dimensionality reduction between the layers is necessary. This is accomplished by a down sampling procedure by the pooling layer. Near the very end of a convolutional neural network (CNN) design is the fully connected layer. Each input is coupled to all of the neurons in this layer, which turns it into a one-dimensional vector. III. Outcomes The study lab ran on an Intel Core i7 2.60 GHz, NVIDIA GeForce GTX graphics processing unit (GPU), 8.00 GB of random-access memory (RAM), and Windows 10 64-bit operating system for the experiments. Using a GPU for 810 iterations and an initial learning rate of 0.001, the SqueezeNet model was trained. Batch size was set to 64. We measured the performance of the SqueezeNet-tuned model by looking at how well it classified data. With a validation accuracy of 77.20% and a training accuracy of 93.47%, the model outperforms conventional machine learning methods. Figure 3 displays the SqueezeNet classifier's training progress.



With a preference for the Food-101 dataset's 10 categories of Indian cuisine, VGG-16 was the runner-up model for food picture categorization. This improves the SqueezeNet model by making classifications with more precision. The training parameters for the proposed VGG-16 model were identical to those for SqueezeNet: a batch size of 64 and a learning rate of 0.001. With 408 CPU iterations and 50 validation frequencies across 823 minutes, the model was trained to an 85.07% classification accuracy. With a larger number of parameters and a deeper network, the proposed VGG-16 achieved much better accuracy. Figure 4 displays the training status of the VGG-16 classifier that has been suggested.

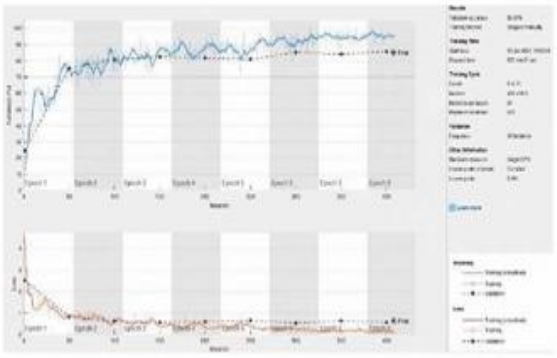


Fig. 5 shows the predicted class labels of randomly chosen food images

Conclusion

This article presents methods for automatically classifying food images using deep learning algorithms. Extrapolating complicated, high-level information improved the accuracy of food picture categorization. This has made use of the deep learning models SqueezeNet and VGG-16. Utilizing data augmentation approaches and fine-tuning hyperparameters, these networks were designed to perform better. An accuracy of 77.20% was achieved with SqueezeNet, which had a far smaller model size and fewer parameters. The proposed VGG-16 network has more parameters and is deeper than SqueezeNet. As a result, the suggested VGG-16 performed much better and obtained an accuracy of 85.07% when it came to food picture classification. Results of SqueezeNet and the Proposed VGG-16 in Training and Validation (Table 2)

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

References

[1]. Zhou, L., Zhang, C., Liu, F., Qiu, Z., & He, Y, "Application of Deep Learning in Food: A Review," Comprehensive Reviews in

ISSN: 2322-3537
Vol-14 Issue-01 June 2025

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/10.1016Zj.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]. 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: ttps://blog.paperspace.com/popular-deep-learning-architectures- resnetinceptionv3 queezenet/# : ~: text=The%20 S queezeNet%20archite

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

[13]. 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.

[14]. 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.