*i*Journals: International Journal of Software & Hardware Research in Engineering (IJSHRE)
ISSN-2347-4890
Volume 8 Issue 12 December 2020

IMPLEMENTATION OF NIGERIAN INDIGENOUS FOOD IMAGE RECOGNITION SYSTEM

Dr Mrs. F.A Ajala; Folowosele A.O; Jeremiah Y.S; Atanda O.G; Adigun E.B. and

(Computer Science, Ladoke Akintola Univerisity of Technology, Ogbomoso, Nigeria),

Abdulkareem Q.B

(Computer Science, Kwara State Polytechnic, Ilorin, Nigeria)

Email: *faajala@lautech.edu.ng[1]; boye4christ@yahoo.com[2]; teejabar5@gmail.com[3]; dayo.oladayoo@gmail.com[4]; adigmanuel@hotmail.co.uk[5]; quadribolajiabdulkareem@gmail.com[6]*

DOI< 10.26821/IJSHRE.8.12.2020.81204 >

ABSTRACT

Food Image Recognition is one of the promising applications of visual object recognition in computer vision. A couple of Nigerian indigenous mealsare on the verge of going extinct in the near future, it is imperative to work towards preserving the knowledge of these indigenous meals. The aim of this study is to implement a Nigerian Indigenous Food Image Recognition System. In this study, a dataset consisting of12 categories and 400 images of indigenous Nigerian meals was downloaded from Kaggle.com, the images were pre-processed using the median filter and Gray-Level co-occurrence matrix, then, the features were extracted and classified using the Convolutional Neural Network algorithm.A 73% level of accuracy of correct recognition was achieved with the model.Based on our findings, Convolutional Neural Network has a higher level of accuracy than other traditional algorithms in automatically segmenting and extracting features and providing accurate classification.

KEY WORDS: Indigenous, Image Recognition, Algorithm, model.

1.0 INTRODUCTION

Computer vision is an interdisciplinary scientific field that deals with how computers can be made to gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do. Computer vision tasks include methods for acquiring, processing, analyzing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information, e.g. in the forms of decisions. Understanding in this context means the transformation of visual images (the input of the retina) into descriptions of the world that can interface with other thought processes and elicit appropriate action. This image understanding can be seen as the disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistics, and learning theory [1].

The scientific discipline of computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-

dimensional data from a medical scanner. The technological discipline of computer vision seeks to apply its theories and models to the construction of computer vision systems [2].

Food is anything that can be eaten and would supply the desired nutrients meant for growth, development and good living in both plants and animals. Food is an essential part of human life and hence humans have found creative, easy and conducive ways to satisfy their eating urge by manipulating available plant and animal produce for meat, grains etc. to their satisfaction to get desired nature, quality and quantity of food. (Nelson,2000).

Owing to the rapid rate of civilization, many indigenous foods from different cultures in Nigeria are going into extinction with the advent of modern-day fast foods, therefore the need arises to preserve the knowledge of these meals and make their recipes available for the present age and years to come using artificial neural networks to correctly recognize them.

The aim of this paper is to implement a Nigerian Indigenous Food Image Recognition System.

The objectives of this work include:

i. Creation of an indigenous Nigerian foods' dataset
ii. A framework for Food Image recognition using deep learning.
iii. Implementation of proposed framework using a proof-of-concept

II. RELATED WORK

Joutou & Yanai (2009) proposed an SVM using multiple kernel learning technique for food image recognition, they achieved an accuracy of 61.34%. Lu (2016) used CNN and a bag of feature support vector machine (SVM) for food image recognition, the CNN model provided a higher accuracy if 74%. Farooq & Sazanov (2017) used SVM and CNN to classify different categories of fast foods with a classification accuracy of 94.01%. Kawano & Yanai (2014) achieved classification accuracy of 72.26% for the UEC-FOOD100 dataset using deep CNN.

Regarding food image recognition, Zhu *et al*(2013) described food recognition using a small dataset, which was intended to be used in a smartphone-based food-logging system as part of their Technology Assisted Dietary Assessment project. Kagaya*et al*(2014)examined 85 food items, achieving 62.5% accuracy for the recognition of Japanese food images collected from the Web. They used multiple kernel learning for feature fusion as their machine learning method. The Pittsburgh Fast-Food Image Dataset is a dataset of American fast-food images, which was used to evaluate a food-recognition method in. Food balance, an aspect of nutritional content, was estimated by image processing. Image retrieval was applied to food recording.

Deep learning has recently been used in image recognition. Deep learning is a collective term for algorithms having a deep architecture that solves complex problems. The most distinctive characteristic is that better image features for recognition are automatically extracted via training.

III. METHODOLOGY

The main steps of the proposed methodology to Food Image classification are listed below. The block diagram consists of five phases:

i.) Data Acquisition
ii.) Image Pre-Processing;
iii.) Feature extraction;
iv.) Classification using convolutional neural network;

Data Acquisition

The dataset created consists of 40 images each under 12 distinct categories. The data was acquired through data scraping and some were acquired by preparing the meals and taking pictures manually. Food images were also downloaded from Kaggle.com. The dataset will be partitioned in this fashion: 70% for training the model and 30% for testing the model

Image Preprocessing:

Twelve common food images are selected as the main research objects, which are Efo, Amala, Egusi, Eba, Iyan, Ikokore, Akara, Moi Moi, Ewedu, Ekuru Ofada and Fufu. Due to differences in ways of acquiring the source images of the Food and in resolution and size, image preprocessing on source images are needed for the sake of subsequent vertical image segmentation.

The specific processing is as follows.Considering that noise constantly exerts a negative impact on acquired samples of Food source images, it is necessary to denoise through median filtering for the reduction of the impact on Food segmentation and

identification brought by the irrelevant background in the images. The median filter is a popular way to remove "salt-and-pepper" noise from an image and at the same time preserve edges and keep useful information. In this paper, the median filter is adopted to preprocess and smoothen the source images.

The main idea of the median filter algorithm is to run through the signal gray value by gray value, replacing each gray value with the median of neighboring gray values. The pattern of neighbors is called the "window," which slides, gray value by gray value, over the entire signal. By using the median filtering algorithm, the original images are denoised and their qualities are enhanced. The formula is as follows:

$F'(x_0, y_0) = \frac{sortF\,(X,Y)(N+1)}{2}$ (median filter equation)

Feature Extraction

Feature Extraction is part of the process of dimensionality reduction by which an initial set of raw data is reduced, while still accurately and completely describing the original dataset. For this work, two features are considered below.

Texture Feature Extraction:

Gray-Level Co-occurrence Matrix(GLCM) is an effective techniquefor extracting texture features. The textures of different food image can be obtained, such as contrast, correlation, entropy, uniformity, and energy. In this project 12 types of food are selected as the main research objects, which are Efo, Amala, Egusi, Eba, Iyan, Ewa Agoyin, Akara,Moi Moi, Ewedu, Ofada and Fufu, ekuru, Ikokore.

i. **Contrast**

It refers to the gray level difference between adjacent pixels; is the distribution probability of gray-level difference between adjacent pixels; and refers to the contrast, mainly used to describe the degree of depth of the image textile grooves. The higher the contrast value goes, the deeper the grooves, and vice versa.

ii. **Correlation**

It is the mean value of the sum of the elements in each column in the square, the standard deviation of the sum of the elements in each column, and the standard deviation of the sum of the elements in each row. Correlation is mainly used to describe the details of relevant elements in each row and column in the process of vertical image segmentation.

iii. **Entropy**

Entropy, which measures the quantity of information within the image and is changed with different textures. As it increases, the textures of speck would be arranged sparsely, and vice versa. When the entropy becomes zero, there is no texture.

iv. **Uniformity**

Where stands for the moment of inertia that is used for the description of the roughness of image texture.

v. **Energy**

Where the energy value mainly applied to describe the thickness of texture, is the quadratic sum of elements of gray-level co-occurrence matrix in the horizontal and vertical directions.

Color Feature Extraction:

In order to improve identification food can be extracted on the basis of identifying the texture features.Resnet32 architecture was applied for image segmentation. ResNet proposed a solution to the "vanishing gradient" problem. Neural networks train via back-propagation, which relies on gradient descent to find the optimal weights that minimize the loss function. When more layers are added, repeated multiplication of their derivatives eventually makes the gradient infinitesimally small, meaning additional layers won't improve the performance or can even reduce it.

ResNet solves this using "identity shortcut connections" – layers that initially don't do anything. In the training process, these identical layers are skipped, reusing the activation functions from the previous layers. This reduces the network into only a few layers, which speeds learning. When the network trains again, the identical layers expand and help the

network explore more of the feature space. ResNet was the first network demonstrated to add hundreds or thousands of layers while outperforming shallower networks. A primary strength of ResNet is its ability to generalize well to different datasets and problems.

Classification

To classify the pre-processed images, we used a convolutional neural network model which consists of an input layer, multiple hidden layers and an output layer. The hidden layers of this model consist of convolutional layers that convolve. We used the Fast.Ai deep learning framework. Fast.Ai is a low code deep learning framework which is built on top of PyTorch.PyTorchis an open source machine learning library based on the Torch library, used for applications such as computer vision and natural language processing. The FastAi library simplifies training fast and accurate neural networks using modern best practices.

The FastAi library allows training a Model in a certain data block very easily by binding them together inside a learner object. It structures its training process around a learner class, whose object binds together a PyTorch Model, a data set, an optimizer and a loss function, the entire learner object then will allow us to launch training and then by using the rational kernel function of FastAi, the classification Model can be established.

THE PROPOSED FRAMEWORK

The proposed framework for the indigenous food recognition system

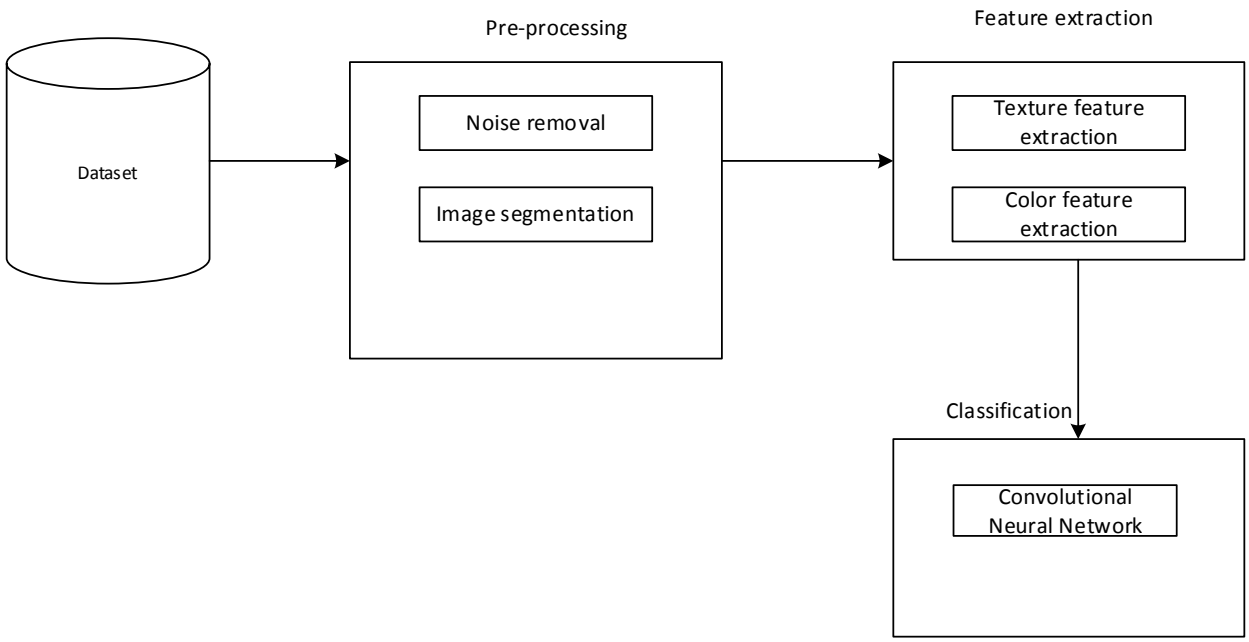

Figure 1 Proposed framework for Food image recognition

IV. RESULTS AND DISCUSSION

In our experiment, the data used was acquired from Kaggle.com and freshly taken pictures after manually preparing some of the meals. This dataset consists of 12 categories and each category has about 40 images. Among the 40 images, around 70% of them were used for training and the remaining 30% were used for testing. There are 400 images in total in this dataset. Each image only contains the labeled information indicating the food type. Most of the images are popular Indigenous Yoruba food images. Some of these food images are shown in Figure 3 below.

Figure 2: Dataset Samples

Performance Evaluation

Performance evaluation shows how well our model is doing and also it tells us the accuracy and precision of our prediction.

The Table 1 below shows the pre-trained model with domain specific fine-tuning such as the number of iterations (Epoch), Train_loss, Valid_loss, Error_rate, Accuracy and Time.

Four iterations were carried out. The train loss is the number indicating the how bad the model prediction was on a single sample. The validation loss is the individual loss function based on the difference between the predicted value and the target value.

The Error rate shows the frequency of error occurred during each iteration of the model.

The Accuracy of a deep learning classification is the measure of how often the model classifies a data correctly, it is one metric for evaluating classification model. Informally, it is the fraction of prediction our model got right. Formally,

$$Accuracy = \frac{\text{No of correct prediction}}{\text{Total number of predictions}}$$

For this work, there was a 73% accuracy level.

Time(in secs) is the take taken for each iteration.

Table 1: Training and Iteration of results.

Epoch	Train_loss	Valid_loss	Error_rate	Accuracy (%)	Time (secs)
0	3.016038	1.884488	0.535354	0.464646	03.12
1	1.747200	1.687740	0.353535	0.646465	02.46
2	1.207421	1.463481	0.272727	0.727273	02.46
3	0.927073	1.242190	0.262626	0.737374	02.49
Mean	1.760519	1.569475	0.356061	0.643939	02.63

Confusion Matrix:

It is a table that describes the performance of a classification model on a set of test data in which the true values are known. It is also known as error matrix with two dimensions whereeach row in the matrix represents the instances in a predicted class while each column represents the instances in an actual class or vice versa. It makes it easy to see if the system is confusing two classes.

In the figure 4 below, *Amala* gave the highest recognition rate which was mainly because the category had a large number of training images while *Efo/egusi* were the two hardest category, the majority of which were misclassified due to similar features.

The recognition results are as shown in the figure 5 below.

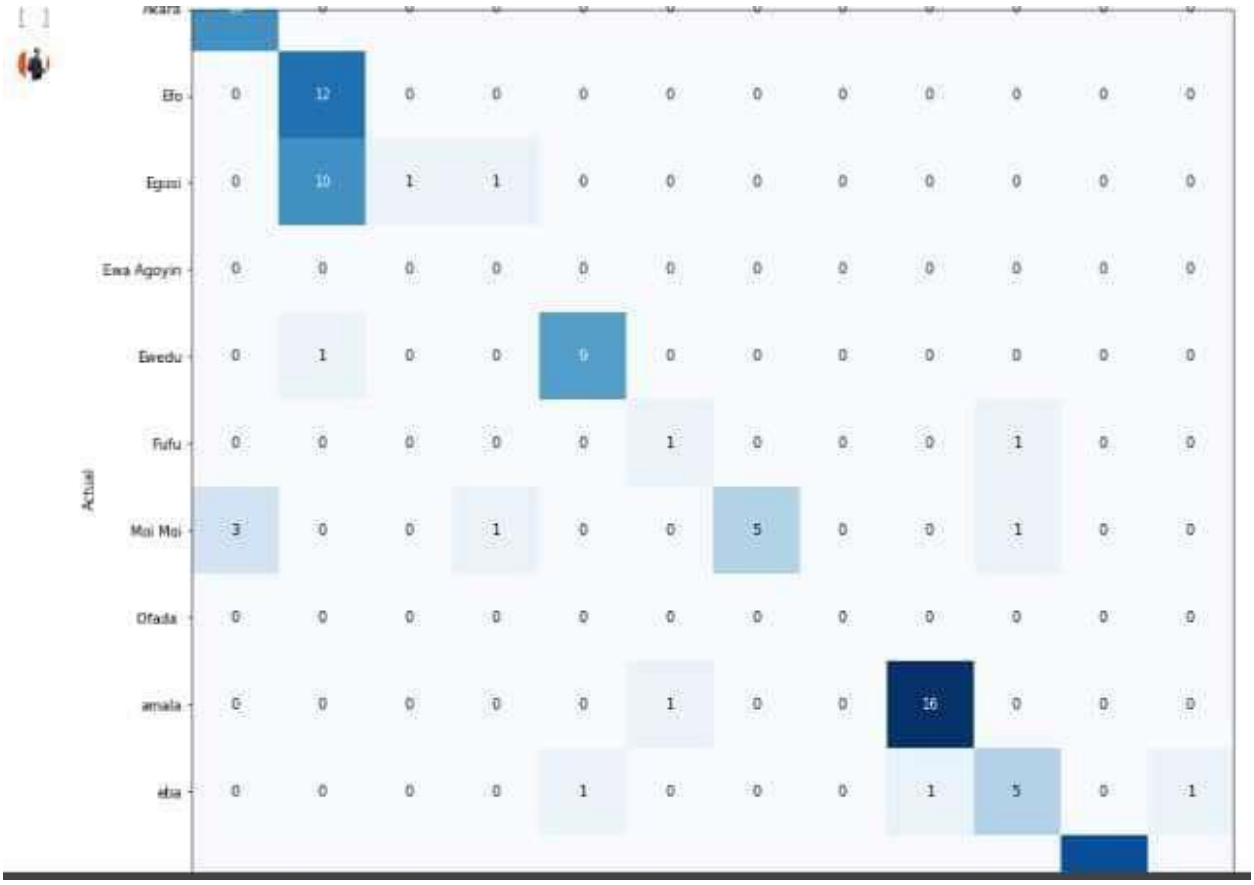

Figure 3: Confusion Matrix from result

Figure 4: Shows recognition result

Proof-of-concept

The image below shows a welcome page to the project tagged YORU-FOOD built with HTML which is standard mark-up language for document design to be displayed in a web browser, it was assisted by technology such as Cascading Style Sheets (CSS) and scripting language called JavaScript, which introduces users to what the website is all about, that is, Yoruba food indigenous food classifier and also includes the recipe for preparing them.

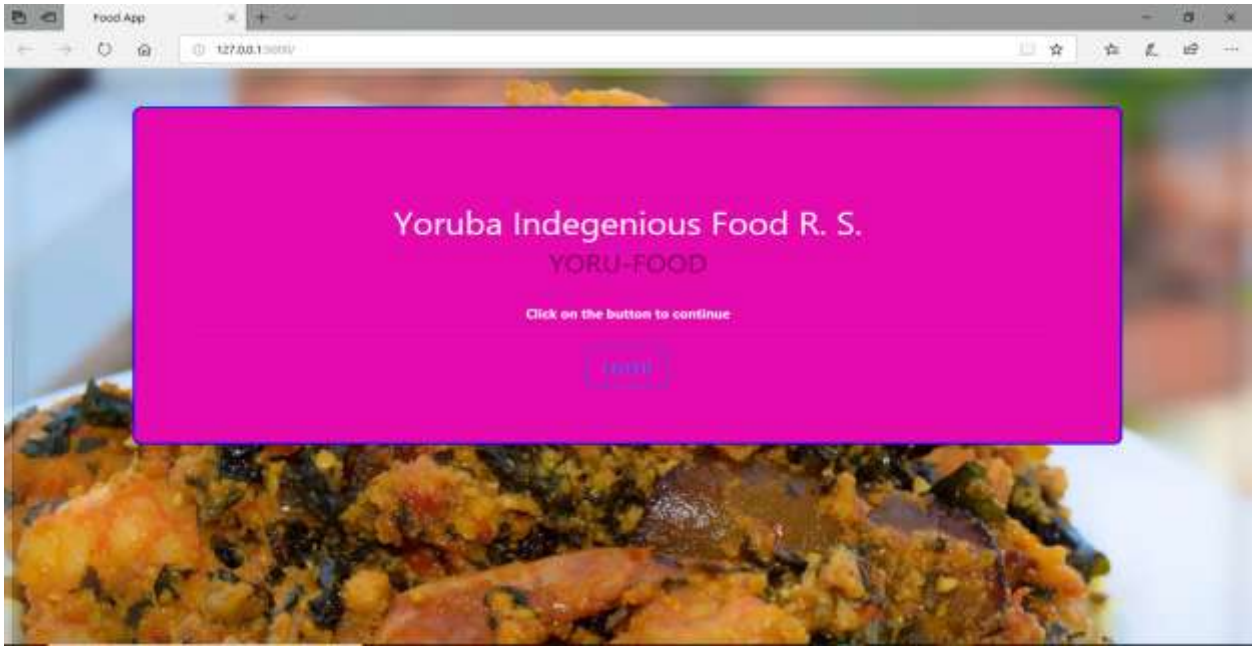

Figure 5: Home Page

On clicking the enter button on the Welcome page, A new page is displayed that allows the user to upload an image of a Nigerian indigenous food and on pressing the classify button, the food name will be displayed and the recipe will be made available for viewing.

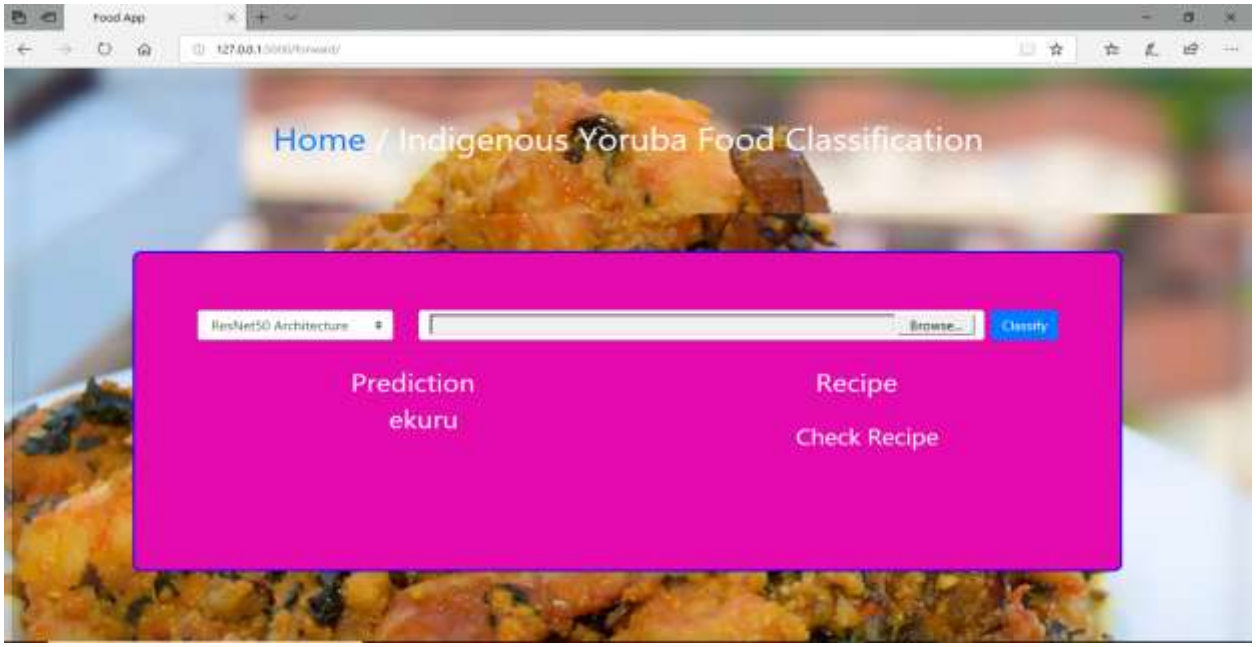

Figure 6: Classification Page

V. CONCLUSION AND RECOMMENDATION

Conclusion

In this study, we proposed a framework for food image recognition. An accuracy of 73% was achieved on the fourth iteration which was an improvement on the 46%, 64%, 72% which was recorded at iterations 0-3 respectively. The results obtained revealed that the framework is capable of producing state-of the-art results due to a high level of accuracy of the classifications obtainedgiven the relatively limited training dataset used.

The result also showed how our model recognizes similarities between some data with similar properties like color, texture etc. Certain foods like *Iyan* and *Semo* possesssimilar propertieswhich could even be mistaken by a normal human brain.

Recommendations

i. The CNN model accuracy could be further improved by increasing the training iterations.
ii. Using a semi-supervised approach to train the dataset in the future.
iii. Separate images example Semovita or Efo not Semovita and Efo. The image must capture a distinct data.

References

1. Dana H. Ballard; Christopher M. Brown (1982). *Computer Vision*. Prentice Hall. ISBN 978-0-13-165316-0.
2. Reinhard Klette (2014). *Concise Computer Vision*. Springer. ISBN 978-1-4471-6320-6.
3. Joutou, T., & Yanai, K. (2009, November). A food image recognition system with multiple kernel learning. In 2009 16th IEEE International Conference on Image Processing (ICIP) (pp. 285-288). IEEE.
4. Farooq, M., & Sazonov, E. (2017, April). Feature extraction using deep learning for food type recognition. In *International conference on bioinformatics and biomedical engineering* (pp. 464-472). Springer, Cham.
5. Kagaya, H., Aizawa, K., & Ogawa, M. (2014, November). Food detection and recognition using convolutional neural network. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 1085-1088).
6. Kawano, Y., & Yanai, K. (2014, September). Automatic expansion of a food image dataset leveraging existing categories with domain adaptation. In European Conference on Computer Vision (pp. 3-17). Springer, Cham.
7. Lu, Y. (2016). Food image recognition by using convolutional neural networks (cnns). arXiv preprint arXiv:1612.00983.

