University of Ibadan
Journal of Science and Logics in ICT Research

Food Components Recognition from Still Images Using Multi-Label Learning

[1]✉Woods, N. C., [2]Oladimeji, O. O. and [3]Fasola, O. O.

[1,2]Department of Computer Science, University of Ibadan, Ibadan, Nigeria
[2]Department of Computing and Electronic Engineering, Atlantic Technological University, Sligo, Ireland
[3]Department of Physical and Mathematical Sciences, Dominican University, Ibadan, Nigeria
Chyn.woods@gmail.com, S00243011@atu.ie, Sanjo.fasola@mbrcomputers.net

Abstract

Food recognition, a recent research area in image processing, helps identify food items to keep track of the food consumed, thereby maintaining a healthy diet. However, the task of food recognition is challenging due to the deformable nature of food items. Usually, there are more than one food item in a meal making the task more challenging. Therefore, the aim of this work is to develop a deep learning model to detect and enumerate visual food components present in a meal. In the multi-label learning approach, food images were collected to build a food image dataset, which comprised 2150 images. The images were pre-processed. Contrast Limited Adaptive Histogram Equalization was then applied followed by scaling to fit as input into the model for training/testing. Thereafter, Deep (VGG-16) and Dense (DenseNet50) models were used to extract deep features. The final layer of the model was applied with a multi-label technique to train on the selected features. The multi-label model was tested using appropriate metrics in which VGG-16 performed better than DenseNet50 with an accuracy of 91.90%, hamming loss of 8.10%, loss of 0.26%, precision of 73.49%. An independent test set was used on the model which showed impressive results. It was observed from this study that the proposed approach performed excellently well in predicting Nigerian Food components. It is recommended that this work be applied in real world this work in real world scenario such as dietary tracking to monitor food intake. Human-Computer Interaction with automatic purchasing systems at restaurants can be used to speed up services.

Keywords: *Food Recognition, VGG-16, DenseNet, Multi-label Learning, Deep Convolutional Neural Networks, Food Components*

1. INTRODUCTION

The integration of Artificial Intelligence (AI) with food recognition has been an area of research interest for the past few decades. However, with the advent of deep learning coupled with increasing computational power, the full potential of AI in food recognition is yet to be realized. Food is generally known to be fundamental to human existence; it always has and will always be essential in human life [1]. A healthy diet is vital to human health [2]; therefore, food plays a vital role in our daily lives [3]. Food does not only provide energy, but also gives us our cultural identity [4,5] and even our religious significance.

Food culture nowadays is spreading more than ever as a result of the digital evolution, with individuals sharing pictures of what they are eating on the Internet. Apart from digital evolution, our eating pattern and food preparation culture is also evolving. In retrospective times, it has been observed that food was mostly prepared at home; but currently we regularly eat food prepared by third parties such as takeaways and restaurants [6]. Hence, there is limited information about the food we eat; as a result, it is cumbersome to identify what we consume.

Therefore, we need to keep track of what we eat, which currently depends on human visual examination to assess the qualified food components and label them properly [1]. This method has been proven to be extremely painstaking, tedious and expensive [1], [7]. This has led to the development of several applications to manually monitor what we consume. Nevertheless, these applications

Woods N. C., Oladimeji O. O. and Fasola O. O. (2024). Food Components Recognition from Still Images Using Multi-Label Learning. *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR),* Vol. 11 No. 2, pp. 46 - 56

hardly provide mechanism for easy monitoring nutrition habits automatically [8]. However, people’s awareness about the food they consume and nutrition habits is increasing [9]. This is either because of certain kinds of food intolerance suffered by some people, mild or severe weight problems, or basically just interested in keeping a healthy diet.

Consequently, food recognition, which is the basic technology for such a kind of automatic dietary assessment tool, is now a trending research topic in recent years. Various technologies and algorithms allow us to guess the food component in food image, which is the most widely used approach [10]. Some of the previous works used these technologies to differentiate food images from other images like [11], [12],[13], [14], which is a binary classification approach. At the same time, some authors worked on multiclass approach which the food labels (names) are independent classes into which food images can be classified such as [15], [16],[17], [18]. However, some works focused on recognizing food ingredients present in single dish meals [9], [19]. Nevertheless, the main problem of these approaches is that they focused on major food component or single dish only and available data focused on food meals with food names as the label as shown in Figure 1. Presently, there are very few datasets on mixed dish food components available as shown in Figure 2. As a result, there is inadequate work on the multi-label classification of food components images in the literature. Another challenge is that various dishes/food items present on a plate are likely to overlap each other. Hence, there may be no clear boundaries between the food items. It is also important to note that shape of some dishes is not regular.

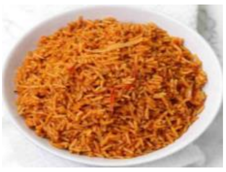


Figure 1. Example of Single Dish Meal

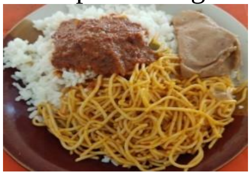


Figure 2. Example of Mixed dish.

Thus, the way forward is to define the problem as a food components recognition problem and having it in mind that the visual appearance of food items can vary from one food to another. Hence, along this line, this work proposes food components recognition from a multi-label by proposing a Convolutional Neural Network (CNN) model framework that allows us to determine food components present in food images.

The proposed multi-label learning approach offers clear advantages. It eliminates the need for bounding box or pixel-wise annotations, focusing solely on categories rather than the precise location of each dish. This approach significantly reduces the annotation workload and simplifies the network design. In contrast to detection and segmentation schemes, our approach boasts lower costs, reduced processing time, and potentially superior results. Moreover, when compared to other multi-label classification methods, our approach demonstrates significantly higher accuracy.

The rest of this paper is arranged as follows the review of existing works was done in Section 2 while the methodology used for this work is presented in Section 3, followed by the results obtained, while the discussion of the results follows, and the conclusion was drawn in Section 5.

2. RELATED WORKS

Based on cognitive uncertainty analysis, Aguilar *et al* [20] developed Forward Step-wise Uncertainty-Aware Model Selection (FS_UAMS) for food image classification. The approach identified the best combination of CNN models to optimize performance while circumventing the high computational resource demands typical of traditional model selection methods like random or exhaustive search. The outputs of two classical CNN models, ResNet 50 and InceptionV3 was modified, by introducing a dropout layer with a probability of 0.5 and an output layer with softmax activation. An accuracy of 89.26% was obtained. Similarly, Tahir *et al* [21] employed transfer learning to minimize computational costs and reduce the likelihood of generalization errors for food image analysis.

A Hamming loss of 0.0070 was achieved for food ingredients detection.

A food composition dictionary to identify visual regions in an image related to food composition and an attention mechanism to enhance the features of these visual regions was utilized by Wang *et al* [22]. A graph convolutional neural network (GCNN) was utilized to learn the constructed graph and aggregate semantic features with visual features. A macro-F1 of 90.82% was obtained. Similarly, Chen *et al* [23] introduced a method that identifies food ingredients using regional features in images by recognizing ingredients in local image regions and then pools these recognition results from different regions to determine the final identification. However, the approach necessitates complex network structures thereby preventing performance degradation.

A Wide Hierarchical Subnetwork-based Neural Network (WI-HSN) framework was developed by Zhang *et al* [24] to classify food types from images. The framework used a supervised subnetwork model for feature encoding and pattern classification. The WI-HSN achieved an accuracy of 90.8%.
These existing works have utilized deep learning, specifically the utilization of convolutional neural networks, transfer learning, and semi-supervised learning. However, to detect multiple food components this study aims to use a multi-label learning approach.

3. Methodology

This work presents an automatic multi-label classification of visual food components using deep learning. The framework presented in Figure 3 comprises three major modules: image gathering and processing, feature extraction using pre-trained CNN, and food components recognition.

3.1 Data Collection and Preparation

This work aims to use the state-of-the-art CNN framework for food components recognition, hence, there is a need for dataset that will be suitable to use. Since, this work aimed at working exclusively on Nigerian foods, there were constrains and limited sources from which the research data was gotten.

Hence, the images were gathered locally by taking pictures of food images and from the internet using google image search, Bing image search, Flickr, Facebook, Instagram using keywords such as "#Nigerianfood", "#Naijafoods", "#NigerianRecipe", "#food9ja". The dataset contains 2,081 images of Nigerian dishes that are common, including Southwestern Nigerian, comprising 26 food components. The following criteria were used for the food image selection: The image must be traditional and popular Nigerian dishes, the images cover a diverse range of Nigerian food categories and the images must be of high-resolution. Figure 4. Shows the distribution of the food components in the image dataset.

As observed in Figure 4, there is a class imbalance as some food components are consumed more than others ones.

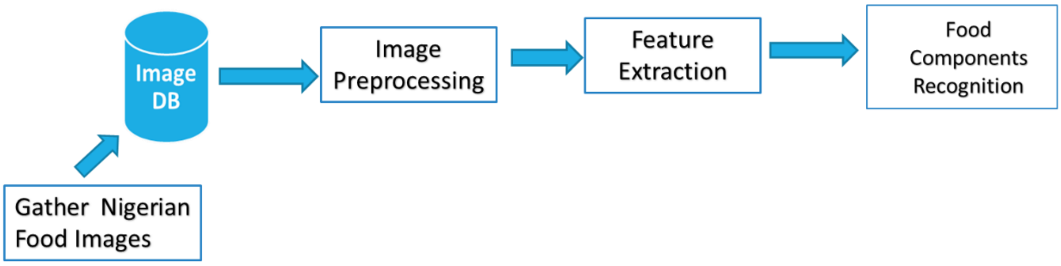


Figure 3. Framework for the Multi-label food components recognition

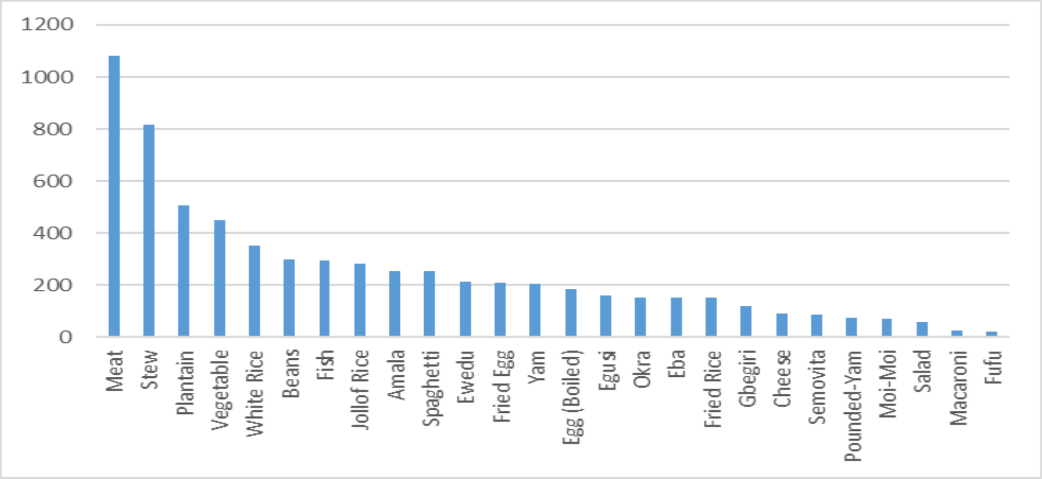


Figure 4. Distribution of the food components in image dataset

3.2 Image Preprocessing

Cropping: The cropping process attempts to exclude or reduce the image's unwanted area (image background). Therefore, the resultant image would have the necessary part of the image (food image) which would form the features for the food component recognition. The image cropping was performed manually on the dataset to take out irrelevant backgrounds to ensure that the needed area (food image) was given as input to the model. Cropping also ensured that important features were extracted after the irrelevant background (non-food components) had been cropped out.

Contrast Enhancement: Image contrast is a vital factor used to determine the quality of image [16]. It helps to differentiate an object from another as well as background. In image processing, contrast enhancement is used to augment the visual appearance of an image for human visual analysis or subsequent machine analysis. In this work, Contrast Limited Adaptive Histogram Equalization (CLAHE), an algorithm used to adjust too bright or too dark images, was used for contrast enhancement [25]. This was done using a threshold of 3.0 to clip the histogram of the intensity of an image and redistribute the histogram to adjust the contrast in the image.

Resizing and Normalization: Since the necessary portion of the food images were obtained via cropping, each of the images was of varying sizes; hence, there was a need to resize the images to ensure that the images were configured (in height and width) for the input layer of the training model. Therefore, the images were resized to 224x224. For the normalization, the RGB values were divided by 255 to obtain normalized values (between 0-1), following the standard practice in machine learning for easy computation.

Labelling: Since, this work aimed at working exclusively on Nigerian foods, there were constrains and limited sources from which the research data was gotten. In essence, the value of one will be assigned to a food component that is present in an image and zero to a food component that is not in the image.

Figure 5 depicts the sample of the food image dataset while Figure 6 highlights some of the labeling of the dataset.

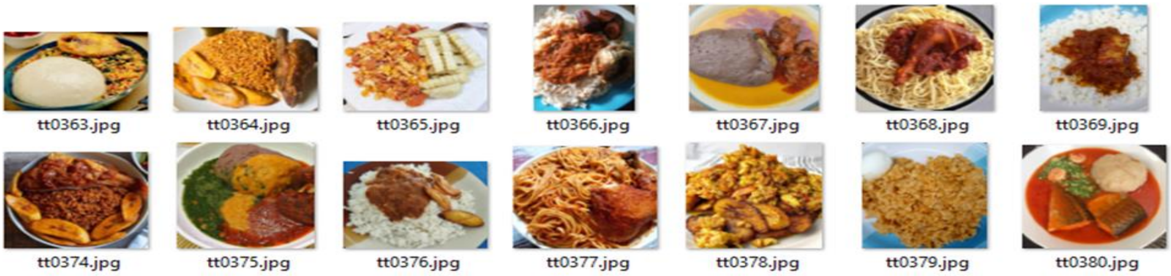
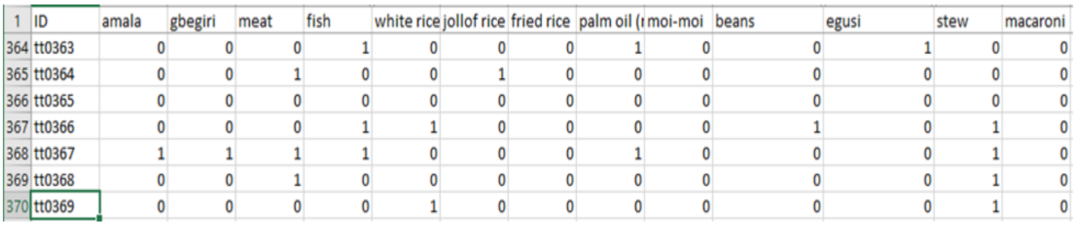


Figure 5. Sample of the food image dataset



1	ID	amala	gbegiri	meat	fish	white rice	jollof rice	fried rice	palm oil (r	moi-moi	beans	egusi	stew	macaroni
364	tt0363	0	0	0	1	0	0	0	1	0	0	1	0	0
365	tt0364	0	0	1	0	0	1	0	0	0	0	0	0	0
366	tt0365	0	0	0	0	0	0	0	0	0	0	0	0	0
367	tt0366	0	0	0	1	1	0	0	0	0	1	0	1	0
368	tt0367	1	1	1	1	0	0	0	1	0	0	0	1	0
369	tt0368	0	0	1	0	0	0	0	0	0	0	0	1	0
370	tt0369	0	0	0	0	1	0	0	0	0	0	0	1	0

Figure 6. Sample of the food image dataset labels in CSV file

3.2 Experimental Analysis

To maximize the use of software and hardware resources made available for researchers by Google, Google Colab was used for the training and testing of the model. Keras library with Tensor Flow was used for model implementation in the python programming language. To augment the limited dataset, the least ten instances of the dataset were augmented using flipping (vertically and horizontally) and rotation (90 and 180 degrees) techniques. Hence, the image summed up to 10,405 food images. The dataset was then partitioned into 70/30% for training and validation, respectively while 7283 instances were used for training, the remaining 3122 instances were used for the validation.

Feature Extraction

It has been proven that DCNN architecture performs better when used to extract features. Therefore, this study employed two types of convolutional neural architectures— deep (VGG-16) [26], and dense (DenseNet50) [27] learning models to extract deep features. instead of building a deep learning model from scratch. The network was initialized with the weights trained on ImageNet [28], and we fine-tuned it with the experimental datasets. Comparative analysis was also conducted to determine the suitability and appropriateness of the deep models for food components recognition.

Recognition

Having extracted the deep features using VGG-16 and DenseNet50, the next process is to train the extracted deep features with our multi-label algorithm. Most of the previous works aimed at binary or multi-class classification tasks, but real-life problems sometimes call for multi-label classification. These types of problems can be solved by developing a multi-label model framework which was done in this research work. From our research we note that, multi food components can be predicted as seen in Equation (1).

$$f(x) \rightarrow (\widehat{y_1}, \widehat{y_2}, \ldots \widehat{y_n}) \qquad (1)$$

Where n is the number of output labels. f might return values (from negative infinity to infinity) as seen in Equation (2).

$$f(x) \in (-\infty, \infty) \qquad (2)$$

However, the model should produce each $\hat{y}$ (predicted value) as binary where each $\hat{y}$ is either 1 or 0 depending on if the food component is present or not present in the image. Hence, Sigmoid (σ) is applied to produce values between 0 and 1, allowing multiple highly activated outputs (Equation (3)).

$$(Sigmoid)\sigma\,(z_i) = \frac{e^{z_i}}{e^{z_i}+1} \qquad (3)$$

Where z_i= output value of the ith node in the output layer.

Since mixed dish recognition is a multi-label classification task, binary cross-entropy will be used as the loss function referred to as summation of binary cross entropy over all classes as depicted in Equation (4).

$$L(y,\hat{y}) = -\sum_{i=1}^{N}\sum_{c=1}^{N_c}[y_{i,c}\log(\widehat{y_{i,c}}) - (1-y_{i,c})\log(1-(\widehat{y_{i,c}})] \qquad (4)$$

Where N is the total number of samples, N_c is the total number of food components (classes), y is the ground truth.

During the backpropagation the weights of the network updated to optimize the recognition performance.

The algorithm was extended to accommodate 26 food components and plugged in three additional dense layers to recognize food components. As food components entail multiple labels, a threshold is essential to control the prediction/selection of labels. The threshold was set to 0.5 (50%) by the deep learning-based multi-label recognition standards.

Table 1: The Model Training Parameters

Parameter	**Values**
Learning Rate	0.001
Batch Size	32
Optimizer	Adam
Epochs	30

Table 2: The Performance Evaluation of the Food Components Recognition Model

Metric	**Score**	
	VGG16	**DenseNet50**
Hamming Loss	0.081	0.102
Precision	0.73	0.65
Training Accuracy	99.98%	89.94%
Validation Accuracy	91.90%	89.71%
Training Loss	0.0037	0.2444
Validation Loss	0.2296	0.2667

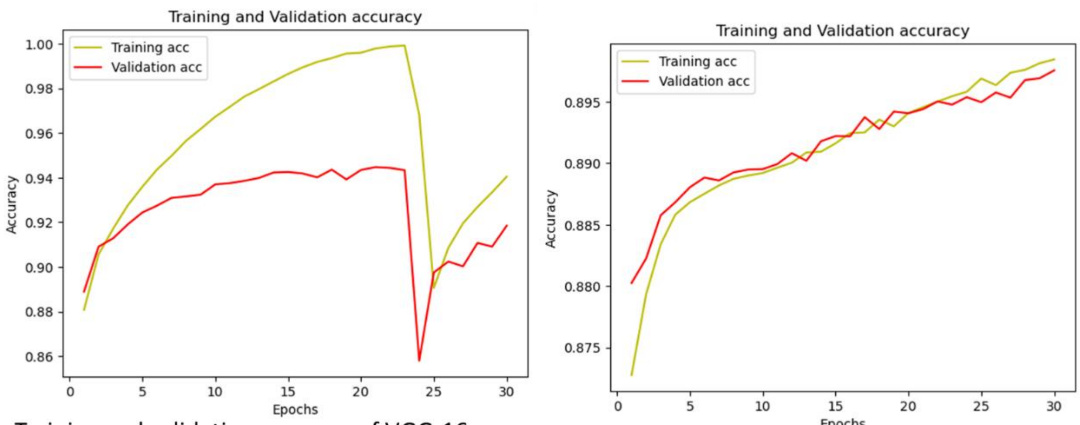
In essence, only labels with prediction scores higher than the threshold are considered as being identified. Table 1 gives the details of the training parameters.

However, to address class imbalance within the training data for our multi-label image classification task, we implemented a class-weighting strategy during the training of our neural network model. The following steps outline the procedure used: To account for class imbalance, class weights using the ‘compute class weight function [29], was computed. For classes absent in the computed class weights, we set default weights to 1.0. This step ensures that all classes are considered during training, even if they are not present in the initial class-weight computation. The class weights were then incorporated during training to mitigate the impact of class imbalance.

4. Results and Discussion

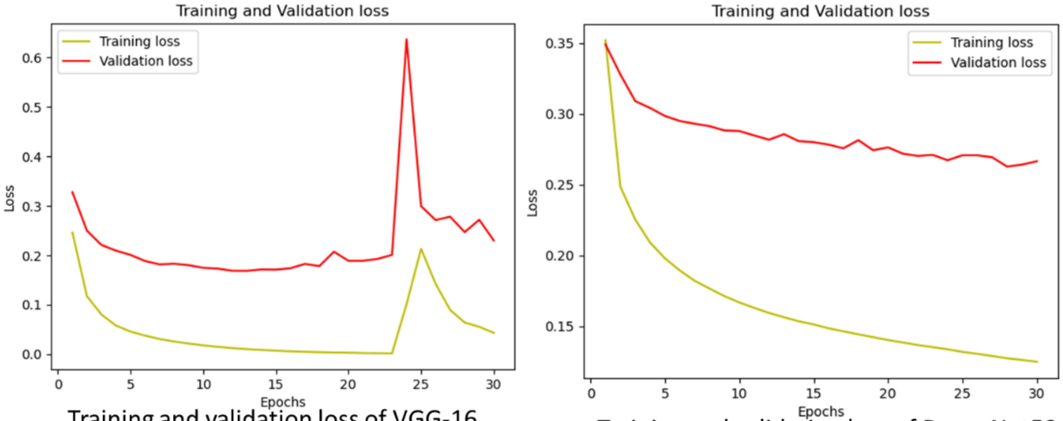
The food components recognition model results are hereby presented in this section and will be described with the following metrics: Accuracy, Hamming loss and precision. The results were validated with test dataset and independent test set. It is important to note that the closer to 1 the metric score is, the better it is, except for the value of hamming loss. Figures 7 and 8 illustrate the performance of the models while Table 2 showcases the performance of the optimal predictive models of VGG16, and DenseNet50 for comparative analysis. VGG-16 demonstrated notable results.

It was observed that VGG16 had better performance than DenseNet in this food component recognition task. The VGG-16 model was used to make predictions on the independent test set. The output of these predictions were probabilistic scores, with values ranging between 0 and 1; hence to obtain the percentage of the prediction for each visual food component, the predicted results are multiplied by 100. Table 3 shows some of the predicted results obtained. From image (a-e), it can be seen that the Model prediction of the food components was high and accurate. It confirms that what was established in the ground truth holds. Furthermore, the aggregated confusion matrix was visualized as shown in Figure 9 to provide an overall view of the models’ performance across all labels.



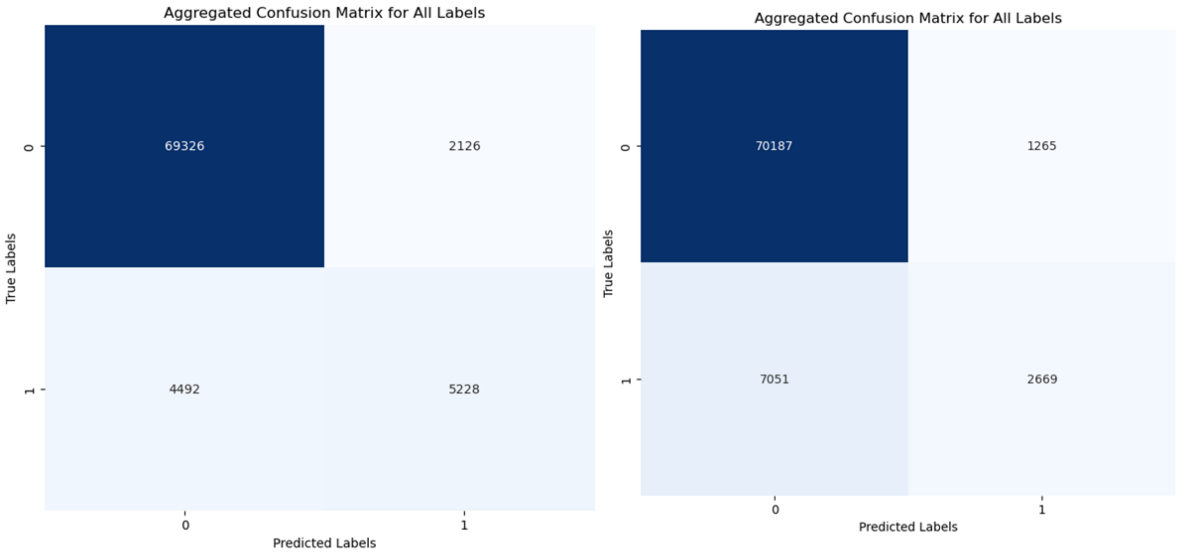
Training and validation accuracy of VGG-16 Training and validation accuracy of DenseNet50

Figure 7. The Training and Validation Accuracy of the Two Models



Training and validation loss of VGG-16 Training and validation loss of DenseNet50

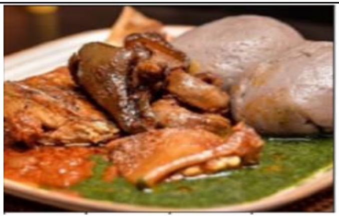
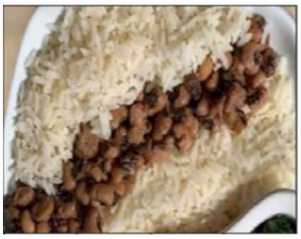
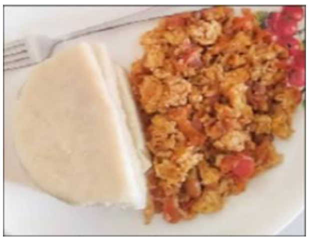
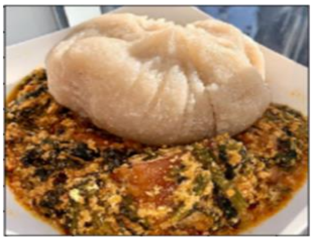
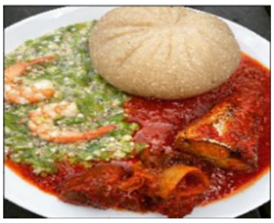
Figure. 8. The Training and Validation Loss of the Two Models



Aggregated Confusion Matrix for VGG-16 Aggregated Confusion Matrix for Dense50

Figure 9 The Aggregated confusion matrix for the two Models

Table 3: Some Samples of Food Component Recognition on the Independent Test

	Image	Ground Truth	Model Prediction
a		Stew Amala Beef Ewedu	Stew (0.999) Amala (0.985) Beef (0.983) Ewedu (0.965)
b		White Rice Beans	White Rice (0.993) Bean (0.895)
c		Yam Fried Egg	Yam (0.967) Fried Egg (0.744)
d		Eba Vegetable Egusi Soup Beef	Beef (0.993) Vegetable (0.967) Egusi Soup (0.945) Eba (0.905)
e		Eba Stew Fish Okra Beef	Stew (0.999) Beef (0.998) Eba (0.964) Okra (0.905) Fish (0.715)

The key result of this work is in two parts: the deep convolutional neural network model for food components recognition and the food image recognition dataset, which has more number of different food items compared to previous works in the field like [30] and, unlike this work, also contains a wide variety of food components

This work has an accuracy of 91.90%, which is better than the accuracy values presented by the other deep convolutional neural network techniques on this subject matter like [31]. Nevertheless, the results of these (previous) studies cannot be directly compared to ours because testing was done on different datasets. Furthermore, it is worthy of note that the accuracy of the classification reduces typically as the number of classes (food components) in the dataset increases, this makes the results much more promising, based on the number of classes (food components) in our dataset which surpasses the previous works cited above.

As a result of the food images' complication, several previously-proposed food recognition techniques had poor classification accuracy, this is where deep learning is beneficial. Food components have challenging features to describe, making automatic feature definition a more suitable technique. The results of this research further affirm this. Nevertheless, overfitting remains an issue with deep learning; in this case, the problem is that there are various diverse classes of food components, and due to the class imbalance, the scarcer classes create fewer images, which presents a bigger risk of overfitting on the little images of that class that are in the dataset. Overfitting might be a contributing factor to the training accuracy being lower than the accuracy on the testing subset; this was also affirmed by Mezgec and Seljak [15] in their study.

Furthermore, the better performance of VGG16 compared to DenseNet50 could be due to the nature of the data and the complexity of the task. DenseNet50 is a more complex model than VGG-16, and with increased complexity comes an increased demand for computational resources. The dataset used in the study is relatively small. Hence, VGG-16 has better generalizability due to its simplicity. In this current study, 26 different food components were classified. While that number is significantly more than what has been previously published in the field like in Deng [30] and Wang and Chen [31], it is still insignificant when compared to the number of food options. Future work will use Grad-CAM to perform visual analysis, so as to understand what information these models actually learn.

5. Conclusion

This work aimed to develop a multi-label learning framework that could enumerate food components in food images. For this reason, a model was developed using a multi-label learning approach. The model was trained on a food image dataset collected both locally and from the internet, summing up 2150 images and increased to 10,405 after augmentation; to the best of our knowledge, this is the first Nigerian food image dataset on this subject matter. This research holds significant promise for improving dietary analysis, supporting food logging applications, and preserving cultural culinary traditions. The contributions of this research include, it could be used to assist persons that are visually impaired to know what they are eating, cultural specificity, West-African food image dataset was created, to the best of our knowledge, this is the first west-African food image dataset., and integration of cultural knowledge, the challenge of recognizing mixed dishes through the lens of multi-label learning was explored. Our framework is designed to recognize dishes at the region level with multiple granularities. while limitations include data availability, and generalization to other cuisines.

In future it would be interesting to incorporate the food components recognition into purchasing or point of sales systems for restaurants. It could also be used as food components recognition in some mobile and web applications for food ordering system. As a result of the tedious task of gathering datasets, this work was able to gather over 2000 images. Thus, it would also be recommended that more data (food images including other food components) be added to the dataset. It would be interesting to propose more approaches to food components recognition, which will also compare this work in the future. Also, the application of association rule mining techniques for improved prediction since most people have a common way for combining the food components (items) to form a meal. Hence, association rule mining will help unveil these food components' combination pattern.

References

[1] L. Pan, S. Pouyanfar, H. Chen, J. Qin, and S. C. Chen, "DeepFood: Automatic Multi-Class Classification of Food Ingredients Using Deep Learning," in *Proceedings - 2017 IEEE 3rd International Conference on Collaboration and Internet Computing, CIC 2017*, 2017, pp. 181–189. doi: 10.1109/CIC.2017.00033.

[2] L. Zhou, C. Zhang, F. Liu, Z. Qiu, and Y. He, "Application of Deep Learning in Food : A Review," *Compr. Rev. Food Sci. Food Saf.*, vol. 0, 2019, doi: 10.1111/1541-4337.12492.

[3] G. G. C. Lee, C. W. Huang, J. H. Chen, S. Y. Chen, and H. L. Chen, "AIFood: A Large Scale Food Images Dataset for Ingredient Recognition," in *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, IEEE, 2019, pp. 802–805. doi: 10.1109/TENCON.2019.8929715.

[4] C. Fischler, “Food, self and identity,” *Soc. Sci. Inf.*, vol. 27, no. 2, pp. 275–292, 1988, doi: 10.1177/053901888027002005.

[5] W. Min, B. K. Bao, S. Mei, Y. Zhu, Y. Rui, and S. Jiang, “You Are What You Eat: Exploring Rich Recipe Information for Cross-Region Food Analysis,” *IEEE Trans. Multimed.*, vol. 20, no. 4, pp. 950–964, 2018, doi: 10.1109/TMM.2017.2759499.

[6] A. Salvador, M. Drozdzal, X. Giro-I-Nieto, and A. Romero, “Inverse cooking: Recipe generation from food images,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10445–10454. doi: 10.1109/CVPR.2019.01070.

[7] H. Chen, J. Xu, G. Xiao, Q. Wu, and S. Zhang, “Fast auto-clean CNN model for online prediction of food materials,” *J. Parallel Distrib. Comput.*, vol. 117, pp. 218–227, 2018, doi: 10.1016/j.jpdc.2017.07.004.

[8] M. Bolanos and P. Radeva, “Simultaneous food localization and recognition,” in *Proceedings - International Conference on Pattern Recognition*, 2016, pp. 3140–3145. doi: 10.1109/ICPR.2016.7900117.

[9] M. Bolaños, A. Ferrà, and P. Radeva, “Food Ingredients Recognition Through Multi-label Learning,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, pp. 394–402. doi: 10.1007/978-3-319-70742-6_37.

[10] N. Martinel, G. L. Foresti, and C. Micheloni, “Wide-slice residual networks for food recognition,” *Proc. - 2018 IEEE Winter Conf. Appl. Comput. Vision, WACV 2018*, vol. 2018-Janua, pp. 567–576, 2018, doi: 10.1109/WACV.2018.00068.

[11] A. Singla and L. Yuan, “Food / Non-food Image Classification and Food Categorization using Pre-Trained GoogLeNet Model,” in *2nd International Workshop on Multimedia Assisted Dieatry Management*, 2016. doi: 10.1145/2986035.2986039.

[12] H. Kagaya and K. Aizawa, “Highly Accurate food/non-food image classification based on a deep convolutional neural network,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9281, pp. 350–357, 2015, doi: 10.1007/978-3-319-23222-5_43.

[13] F. Ragusa, V. Tomaselli, A. Furnari, S. Battiato, and G. M. Farinella, “Food vs Non-Food Classification,” in *2nd International Workshop on Multimedia Assisted Dieatry Management*, 2016, pp. 77–81. doi: 10.1145/2986035.2986041.

[14] P. McAllister, H. Zheng, R. Bond, and A. Moorhead, “Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets,” *Comput. Biol. Med.*, vol. 95, pp. 217–233, 2018, doi: 10.1016/j.compbiomed.2018.02.008.

[15] S. Mezgec and B. K. Seljak, “Nutrinet: A deep learning food and drink image recognition system for dietary assessment,” *Nutrients*, vol. 9, no. 7, pp. 1–19, 2017, doi: 10.3390/nu9070657.

[16] A. B. Akhi *et al.*, “Recognition and Classification of Fast Food Images,” *Glob. J. Comput. Sci. Technol.*, vol. 18, no. 1, 2018.

[17] Z. Shen, A. Shehzad, S. Chen, H. Sun, and J. Liu, “Machine Learning Based Approach on Food Recognition and Nutrition Estimation,” *Procedia Comput. Sci.*, vol. 174, pp. 448–453, 2020, doi: 10.1016/j.procs.2020.06.113.

[18] S. J. Park, A. Palvanov, C. H. Lee, N. Jeong, Y. I. Cho, and H. J. Lee, “The development of food image detection and recognition model of Korean food for mobile dietary management,” *Nutr. Res. Pract.*, vol. 13, no. 6, pp. 521–528, 2019, doi: 10.4162/nrp.2019.13.6.521.

[19] J. Chen and C. W. Ngo, “Deep-based ingredient recognition for cooking recipe retrieval,” in *MM 2016 - Proceedings of the 2016 ACM Multimedia Conference*, 2016, pp. 32–41. doi: 10.1145/2964284.2964315.

[20] E. Aguilar, B. Nagarajan, and P. Radeva, “Uncertainty-aware selecting for an ensemble of deep food recognition models,” *Comput. Biol. Med.*, vol. 146, 2022, doi: 10.1016/j.compbiomed.2022.105645.

[21] G. A. Tahir and C. K. Loo, “Explainable deep learning ensemble for food image analysis on edge devices,” *Comput. Biol. Med.*, vol. 139, 2021, doi: 10.1016/j.compbiomed.2021.104972.

[22] Z. Wang *et al.*, “Ingredient-Guided Region Discovery and Relationship Modeling for Food Category-Ingredient Prediction,” *IEEE Trans. Image Process.*, vol. 31, 2022, doi: 10.1109/TIP.2022.3193763.

[23] J. Chen, B. Zhu, C. W. Ngo, T. S. Chua, and Y. G. Jiang, “A Study of Multi-Task and Region-Wise Deep Learning for Food Ingredient Recognition,” *IEEE Trans. Image Process.*, vol. 30, 2021, doi: 10.1109/TIP.2020.3045639.

[24] W. Zhang, J. Wu, and Y. Yang, “Wi-HSNN: A subnetwork-based encoding structure for dimension reduction and food classification via harnessing multi-CNN model high-level features,” *Neurocomputing*, vol. 414, 2020, doi: 10.1016/j.neucom.2020.07.018.

[25] S. M. Pizer *et al.*, “Adaptive histogram equalization and its variations,” *Computer Vision, Graphics, and Image Processing*, vol. 38, no. 1. p. 99, 1987. doi: 10.1016/s0734-189x(87)80156-1.

[26] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015, pp. 1–14.

[27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017. doi: 10.1109/CVPR.2017.243.

[28] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015, doi: 10.1007/s11263-015-0816-y.

[29] G. Lemaître, F. Nogueira, and C. K. Aridas, “Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning,” *J. Mach. Learn. Res.*, vol. 18, 2017.

[30] L. Deng *et al.*, “Mixed-dish recognition with contextual relation networks,” in *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 112–120. doi: 10.1145/3343031.3351147.

[31] Y. Wang, J. J. Chen, C. W. Ngo, T. S. Chua, W. Zuo, and Z. Ming, “Mixed dish recognition through multi-label learning,” in *CEA 2019 - Proceedings of the 11th Workshop on Multimedia for Cooking and Eating Activities*, 2019, pp. 1–8. doi: 10.1145/3326458.3326929.