IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111201

Real-time Food Recognition and Documentation Android System for the Learning of Nigerian Foods using Deep Learning Method

Nnamdi Johnson Ezeora[1], Ogbene Nnaemeka Emeka[2], Ejiofor Virginia Ebere[3], Ndubuisi John Ngene[4], Ozioko Ekene Frank[5], Asogwa T. C[6]

Department of Computer Science, University of Nigeria Nsukka Nsukka, Nigeria[1,2]

Department of Computer Science, Nnamdi Azikiwe University Awka Awka, Nigeria[3]

Department of Computer Science, Enugu State University of Sci. & Tech. Enugu, Nigeria[4, 5,6]

Abstract: Food Recognition is a computer vision application that has gained huge research interest. Food recognition and classification system help to understand foods' different diversity as different cultures entails different cuisines. There is need to recognize and document Nigerian Foods to avoid going into extinction. It is also necessary to understand the food specific estimated calorie contents. Existing food recognition systems are based on foods found in developed nations. This paper presents a real-time Android system for Nigerian food recognition and documentation system using finetuned MobileNetV2 model. The Convolutional Neural Network (CNN) is effective in image recognition and classification. Therefore, MobileNetV2, a CNN-based deep learning model, is utilized to recognize and classify 10 Nigerian food classes from 500 food images of our dataset developed from scratch known as the NaijaFood101. We evaluated the Nigerian food learning model's performance to determine the accuracy of the food recognition and classification system. The model achieved an average of 97% recognition accuracy on the evaluation or test data.

Keywords: Food recognition, calorie estimation, NaijaFood101 dataset, Nigerian foods, deep learning, convolutional Neural Network

I. OVERVIEW

This research paper presents a Nigerian food recognition and documentation android system for the learning of Nigerian foods and its calorie contents. Our custom Nigerian food known as the NijaFood101 dataset was developed to also help further research in this field. The theoretical introduction to Nigerian food recognition and documentation android system is presented in Section II. Section III reviews related works. Section IV describes the design approach to the developed system. Section V describes the methodologies applied in Nigerian food recognition and classification system. Section VI discusses the results from the experiments conducted. The summary was covered in Section VII.

II. INTRODUCTION

Deep Learning (DL) is an advanced Machine Learning technique that uses layers of neurons to mimic the thinking process, process data, and perform many other abstractions. Deep learning is widely used in computer vision problems, speech recognition and image processing. Input data or samples are processed through a series of layers called neurons. The previous layer's output serves as an input for the next layer. The first layer is referred to as the input layer, while the last layer is referred to as the output layer. All other layers between the first and last layers are known as the hidden ones. The result of the output layer after the summation of the input values and their corresponding weights is determined through an activation function. Object recognition is a trending topic in computer vision that is related to image or video processing to detect instances of a semantic object. The advancement in computer vision models has led to the vast application of Computer Vision (CV) in areas such as image and video recognition and classification problems. With the recent evolution of interest in deep learning methods, the learning algorithms outperformed previous state-of-the-art techniques in several tasks, and the abundance of complex datasets from different sources (e.g., visual, audio, medical, social, and sensor). The quest to develop a system that mimics the human brain fueled the initial development of neural networks. The frameworks are destined to run the inference task on mobile phones, which is suitable for scenarios that

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ≋ Impact Factor 7.918 ≋ Vol. 11, Issue 12, December 2022
DOI: 10.17148/IJARCCE.2022.111201

require real-time recognition tasks [3]. Nevertheless, running such CNN models on mobile devices is still challenging owing to the limited computing power and energy available [4]. Foods are nutritious and vital for human activity and growth. However, when the nutrients are not correctly balanced, that may lead to diseases like obesity, diabetes, and heart disease [5]. Therefore, studying food recognition and classification helps in food identification, dietary management, and food logging.

Food strongly correlates with people's culture, and most food recognition systems focused on food datasets found in developed countries [6]. However, as our languages are different in the same manner, our cuisines are different. Unfortunately, most Nigerian foods are gradually becoming extinct as globalization explodes, leading most Nigerians to eat imported foods, leaving the younger generation with little or no knowledge about their indigenous foods. Hence, food recognition helps to document the indigenous foods that will be recognized in real-time using a mobile android phone device. Furthermore, real-time recognition enables users to receive timely and effective recognition feedback to achieve valuable food detection and classification.

Traditionally, CNN models run-on high-performance computing servers due to hardware requirements and are unavailable to operate on smartphones and to perform automatic real-time recognition. To overcome these issues, there is a vital need for a machine learning model suitable for smartphones to perform computing-intensive Computer vision tasks, such as object recognition. Hence, we used MobileNetV2, an optimized deep-learning model for Nigerian food recognition and classification. TensorFlow Lite provides a framework for executing the MobileNetV2 model on mobile processors. Our paper implements an android system for Nigerian food recognition and classification to detect Nigerian foods using the deep learning pre-trained model MobileNetV2 algorithm [7]. The NaijaFood101 datasets used in this research were developed from scratch, and a dietician was employed to classify the food datasets. The dataset consists of foods found in the South-eastern region of Nigeria

III. LITERATURE REVIEW

As seen in [8], a Nigerian indigenous food image recognition system was developed. The dataset used for the study consists of 12 food classes and 400 images of indigenous Nigerian foods. The images were pre-processed using the median filter and Gray-Level co-occurrence matrix. The features extracted from the images were classified using the Convolutional Neural Network algorithm. They achieved a 73% level of accuracy of correct recognition with their model. Their findings also suggested that CNN has a higher level of accuracy than other traditional algorithms in automatic segmenting and extracting features and providing accurate classification. However, the system is not a mobile application and cannot perform real-time recognition. Also, the recognition accuracy achieved in the work is 76%.

The work in [9] introduced 1000 food images from the Chinese food dataset called the ChinFood1000. The proposed ChinFood1000 dataset currently contains the most significant number of food categories among all publicly available food datasets. The food classes of the ChinFood1000 dataset were carefully selected to include the most popular Chinese dishes. The dataset also consists of 852 categories of Chinese dishes and 91 classes of drinks and snacks. Also, 26 kinds of fruits and 31 kinds of other food are in the dataset. The images in the dataset comprise both significant inter-class affinity and high intra-class variance. They applied a CNN model to train the dataset properly. The experiments evaluate the baseline approach on the three most widely used food datasets and achieve the best performance. The baseline approach was also applied to the ChinFood1000 dataset, with reasonable accuracy reported.

The researchers in [10] proposed a new deep convolutional neural network configuration to detect and recognize local food images. They suggested that similarity of color and texture of food types makes food image recognition challenging. However, the application of deep learning in food recognition is effective in image recognition, and CNN is the contemporary approach for deep learning to be implemented. In the study, they optimized CNN for food detection and recognition tasks with few modifications.

Furthermore, they presented a new dataset of the most consumed local Malaysian food items collected from publicly available Internet sources, including but not limited to image search engines. In evaluating the performance, CNN achieved significantly higher accuracy than traditional approaches with manually extracted features.

According to [11] addressed the problem of food recognition by extracting the discriminative food regions. Using this technique, they proposed a novel network architecture in which a primary network maintains the base accuracy of

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁑ Impact Factor 7.918 ⁑ Vol. 11, Issue 12, December 2022
DOI: 10.17148/IJARCCE.2022.111201

classifying an input image. In contrast, an auxiliary network extracts the discriminative food regions, followed by the regional network that classifies the extracted regions. The original input image and the extracted regions representations are integrated for the final recognition. In addition, they introduced a new fine-grained food dataset named Sushi-50. Such-50 consists of 50 different sushi categories. Finally, they conducted extensive experiments to evaluate the proposed model on three food datasets, Food-101, Vireo-172, and Sushi-50, which achieved the following testing accuracy of 90.4%, 90.2%, 92.0%, respectively, compared with other existing approaches.

The authors in [12], in their study, presented a novel image retrieval approach for small and medium-scale food datasets. They increased the number of the dataset by utilizing image transformation techniques and image augmentation. The technique helped them promote the average food recognition accuracy with deep learning technologies.

The phases in their system development are; image transformation techniques were adopted to augment food images. Secondly, transfer learning based on deep learning was applied to extract image features, and finally, a food recognition algorithm, ResNet feature vectors, was applied to the extracted deep feature vectors. The presented image-retrieval architecture is analyzed on a food dataset composed of 41 food classes of food ingredients and 100 food mages for each category. Extensive experiments conducted in the study showed the advantages of image-augmentation architecture for moderate datasets using deep learning.

As seen in [13], designed and implemented a mobile food recognition system aimed at typical food in China. The system also provides a reference for the nutritional intake of the population. The similarity in the food image recognition found in China posed a challenge to the implementation, such as the diverse appearance of the same food and the complex food background. Furthermore, they built a dataset, ChineseFood80, with quality data and used transfer learning to train models by importing pre-training models. The system performance was compared with the baseline method; the recognition accuracy improved by more than 10%.

The work in [14] present a new dataset of food images used to evaluate food recognition and dietary assessment systems. They developed a Mediterranean Greek food (MedGRFood) dataset consisting of food images from Mediterranean cuisine, mainly from Greek cuisine. The dataset contains 42,880 food images belonging to 132 food classes collected from the web. Based on CNN, an EfficientNet model variant, specifically the EfficientNetB2, they proposed a new deep learning schema that achieves 83.4% top-1 accuracy and 97.8% top-5 accuracy in the MedGRFood dataset for food recognition.

The researchers in [15] empirically evaluated recent transfer learning models for deep learning feature extraction for a food recognition model. They utilized the VIREO-Food172 Dataset and a newly established Sabah Food Dataset to evaluate the food recognition model. Afterward, they integrated the model into a web application system to automate food recognition. The model consisting of a fully connected layer with 11 and 10 Softmax neurons is used as the classifier for food categories in both datasets.

They further evaluated Six pre-trained Convolutional Neural Network (CNN) models as the feature extractors to extract essential features from food images. From the evaluation, the research found that the EfficientNet feature extractor-based and CNN classifier achieved the highest classification accuracy of 94.01% on the Sabah Food Dataset and 86.57% on VIREO-Food172 Dataset.

IV. PROPOSED USE CASE DESIGN

A use case diagram is a behavior diagram and exposes the observable interactions between actors and the system under development. A use case diagram is made up of the system, actors and how they relate to each other. The actors in this system are divided into two, user and administrator. The administrator is responsible for preparing the datasets, dataset pre-processing, training the system recognition model and preparing recognized food report.

The administrator is also responsible for deleting the dataset and updating food records. The sole responsibility of the user is to capture a real-time image of Nigerian foods to be recognized to ascertain the food's glycemic index level. Figure 1, illustrates the Use case diagram of the proposed system.

DOI: 10.17148/IJARCCE.2022.111201

V. RESEARCH METHOD

The research method describes how the information system will be coded or programmed, deployed, installed, and delivered into an operating system. The research method employed for the proposed system for recognizing and documenting Nigerian food for learning purposes consists of five stages, namely:

1. Dataset preparation
2. Dataset pre-processing
3. Developing TFRecord
4. Finetuning the model
5. Training and Evaluation
6. Deploying TFlite model to Android

Dataset Preparation

Dataset acquisition is one of the fundamental and difficult tasks of developing a food recognition system. Part of the objective of this work is to develop a Nigerian food dataset. A custom image dataset is required to obtain accurate food item recognition results since foods vary by region. Our findings show no publicly available image dataset for Nigerian food item recognition. Therefore, we collected more than 500 food images by taking pictures in the market and farm and searching the Internet for web-based photos. In our quest to build an adequate dataset suitable for training the complex recognition model for Nigerian food images, 35 food groups based on Nigeria's frequently consumed food list were selected, and documented. Finding the desired images of the food objects require that they be collected by taking photographs, that is, collecting them manually or by hiring a third party to take the pictures.

Dataset pre-processing

Preprocessing of dataset involves enhancing the extracted features of the image for improved recognition. After datasets are collected, the next step requires that the collected data are labelled. A labelled dataset generates to a file in an XML format that must be converted to a CSV file. Next after a successful conversion is image segmentation. Image segmentation is a method of breaking down an image into regions that comprise a subgroup of images or areas of interests. The subgroups are also referred to as image segments. They help reduce the image's complexity to make further processing or analysis of the image easier. In this work, the subgroups are the image and background areas. There are three (3) major common types of segmentation, namely:

1. Thresholding segmentation,
2. Edge-based segmentation
3. Region-based segmentation

Developing TFRecord

The CSV file output for the dataset is then converted to TF-Records format using TF-Record python code. There are several things to take note of over here. While creating the TF-Records file, the label map needs to be generated. The label map is essentially an encoder for the classes to be trained as the neural net predicts only in numbers after the final sigmoid layer.

Finetuning the model

Finetuning the model involves modifying the models' hyper parameters such as batch size, number of steps, and the learning rater. TensorFlow's object detection pipeline comes with a pre-configured file that optimizes most of the configuration selection for efficient model training. Before configuration, the path to the label map which points to the path of the training TF-Records files is configured. Next, the number of the food classes to be recognized are set, together with the batch size.

The number of training steps, which refers simply to the number of training steps, is then set to a desired number. The hyper parameters are well determined to avoid model overfitting or underfitting. The MobilenetV2 model is then downloaded, and the location of the checkpoint file is also incorporated in the config file.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⋇ Impact Factor 7.918 ⋇ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111201

Here are required file paths to be modified in the model configuration file:

1. Modify the *num_classes* to the number of classes of food to be recognized.
2. Set and save model *checkpoint* file path

The overall system architecture is illustrated in Figure 2.

Training and Evaluation

Training simply means learning the dataset extracted features. After successfully completing the previous steps, the Python train command is engaged for learning to start using MobileNetV2. Training time varies depending on the computing power of your machine and the number of training steps. The higher the batch size and number of steps is the most suitable to for accurate training but can take days to train depending on device computing power. Evaluating the model training performance can be done in parallel with the training process. The performance of the training can be visualized or monitored using Tensor-board application interface.

Deploying TFLite Model to Android

TensorFlow Lite provides an efficient gateway to run TensorFlow models on Android and micro-controller devices. To run the conversion scripts, the TensorFlow Object Detection API on our machine is set up in computer. The TFLite model serves as the Android inference engine. It is copied to the Android project along with the labelmap.txt and calori_info.txt. The Python command below generates the TfLite model for Android.

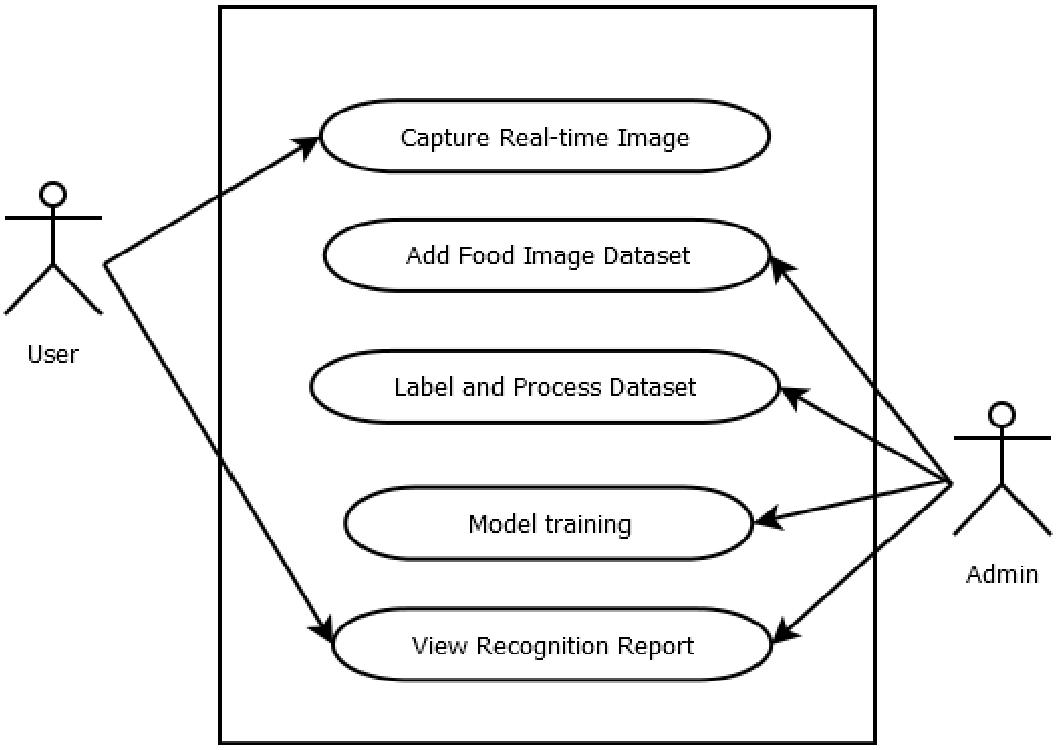


Figure 1: Use Case Diagram of the Proposed System

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁘ Impact Factor 7.918 ⁘ Vol. 11, Issue 12, December 2022
DOI: 10.17148/IJARCCE.2022.111201

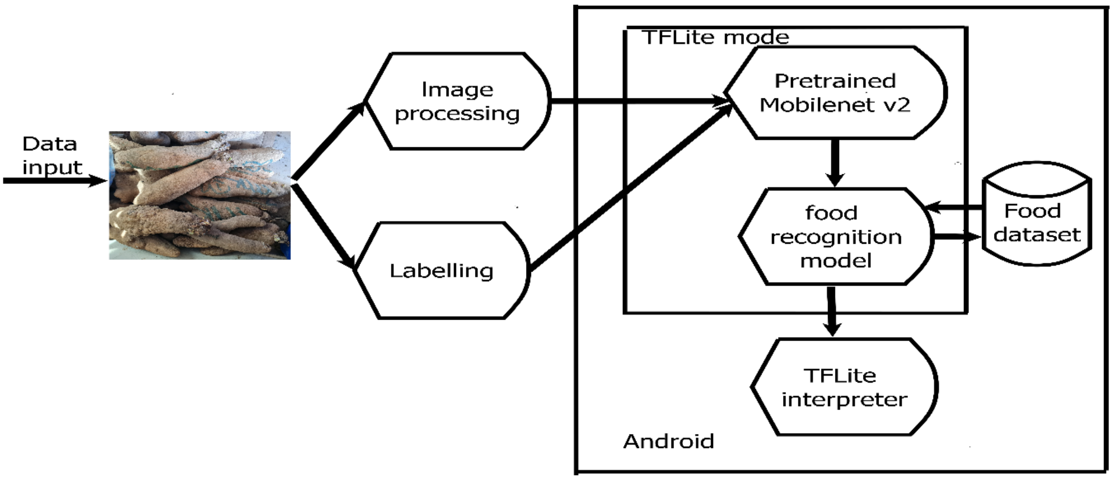


Figure 2: System Architecture

VI. RESULTS

The recognition test accuracy is the ability of a model to recognize test data examples using trained datasets. Test datasets are not part of trained datasets or datasets the model has never seen before. The datasets are generally divided into train and test data in deep model learning training stages. The train data comprises 80% of the dataset, while the test data makes up the remaining 20%. The model does not learn from the test dataset. Hence, the test datasets are used to determine the accuracy of the trained data. Results generated from our experiments as shown in Figure 3 and as visualized on TensorBoard shows that the model achieved an average test accuracy of 97%. This an improvement over the 76% recognition accuracy achieved in [8]. TensorBoard is a web application tool that provides model performance measurements and visualizations needed during the model learning pipeline. TensorBoard helps researchers to track experiments' metrics such as accuracy, loss and visualization model graph.

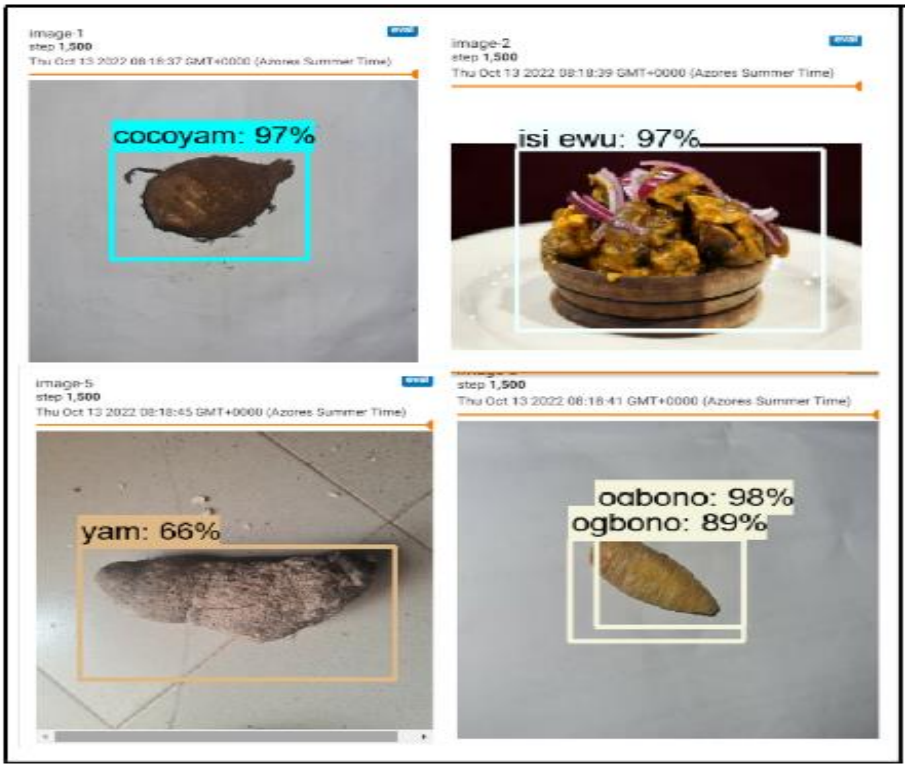


Figure 3: Test Accuracy Results

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 12, December 2022

DOI: 10.17148/IJARCCE.2022.111201

The main goal of training a model is to find a set of weights and biases that have low training loss on average across all dataset examples, i.e., as the loss gets closer to Zero, the better. Therefore, experiments were conducted, the batch size was set to 10, and the number of training steps was set to 1500. The learning rate was kept at 0.004, and 500 dataset images (train data = 400, test data = 100) were used for the training. As a result, the training classification and total loss obtained are 1.13 and 1.52, respectively, as shown in Figures 4 and 5.

The screenshots of the training results with corresponding estimated calorie content, when deployed to an Android smartphone, as shown in Figure 6, indicate that the optimized MobileNetV2 model learned the dataset’s features.

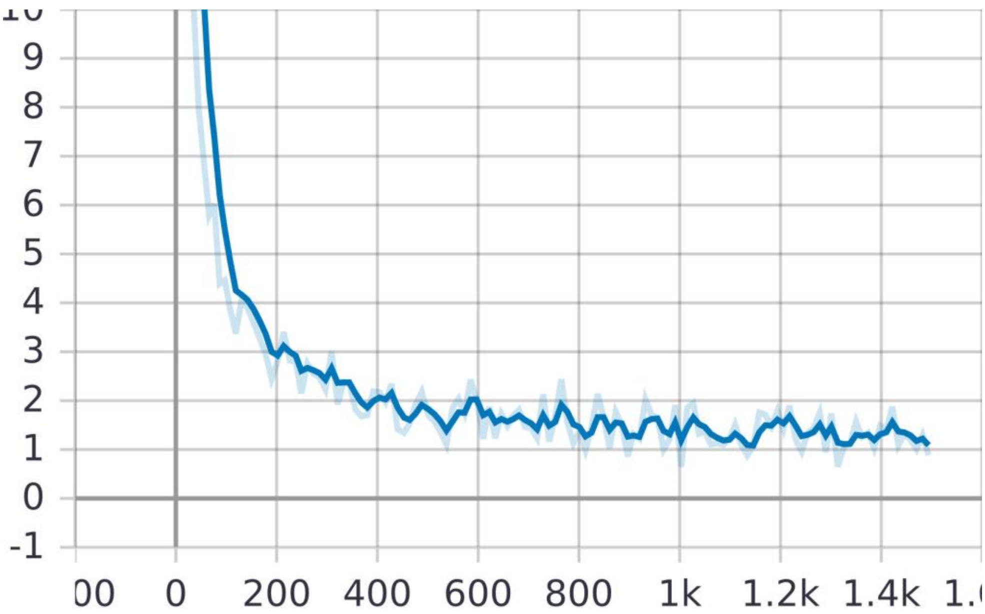


Figure 4: Classification Loss Graph

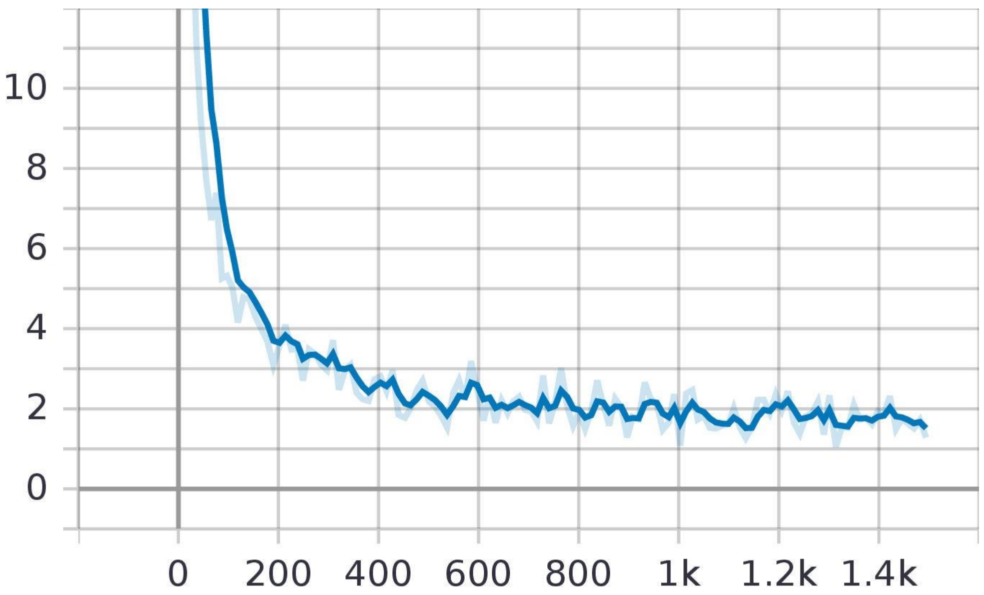
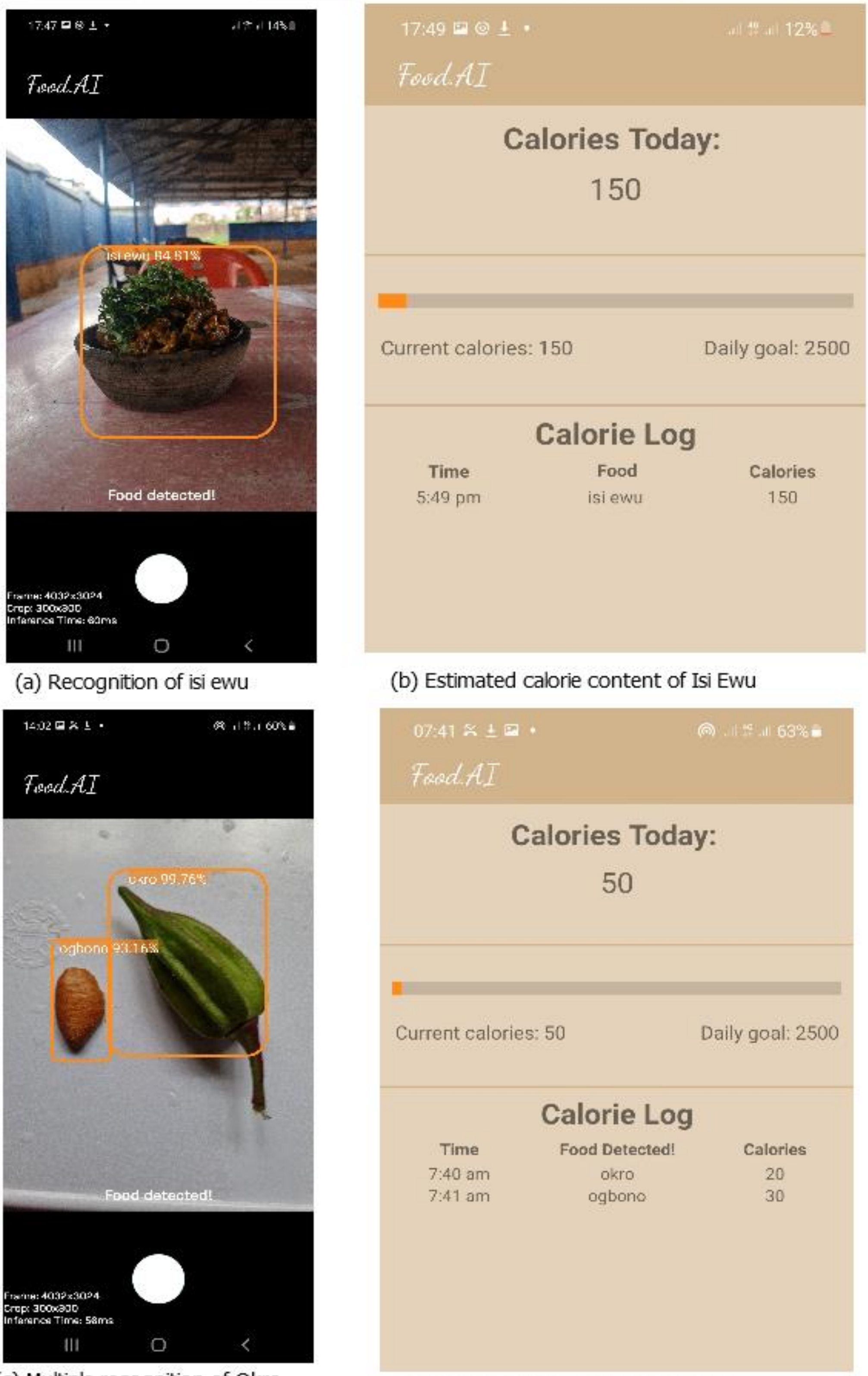


Figure 5: Total Loss Graph

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁂ Impact Factor 7.918 ⁂ Vol. 11, Issue 12, December 2022
DOI: 10.17148/IJARCCE.2022.111201



(a) Recognition of isi ewu

(b) Estimated calorie content of Isi Ewu

(c) Multiple recognition of Okro and Ogbono

(d) Estimated contents of Ogbono and Okro

Figure 6: Screenshot of Nigerian Food Recognition on Android Phone

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁑ Impact Factor 7.918 ⁑ Vol. 11, Issue 12, December 2022
DOI: 10.17148/IJARCCE.2022.111201

VII. CONCLUSION

In this work, an android food recognition and documentation of Nigerian foods have been designed and implemented using an finetuned MobileNetV2 model. A new Nigeria food dataset is also developed. The developed system shall aid in the recognition and documentation system for the learning of Nigerian foods to avoid its extinction. Also, the system shall enable the users to gain knowledge of the estimated calorie contents of recognized food. The developed model was assessed to test the performance in aspects like test accuracy. The model achieved an average 97% recognition test accuracy on the test data.

For future works, advanced pre-processing of Nigerian food images should be explored to improve extracted features with similar physical attributes like shape and colour. Also, brief history of recognised food should be incorporated to the recognition system.

REFERENCES

[1] Jason Brownlee. “A gentle introduction to object recognition with deep learning”. In: Machine Learning Mastery 5 (2019).
[2] Gary Bradski and Adrian Kaehler. Learning OpenCV: Computer vision with the OpenCV library.” O’Reilly Media, Inc.”, 2008.
[3] Mengwei Xu et al. “A first look at deep learning apps on smartphones”. In: The World Wide Web Conference. 2019, pp. 2125–2136.
[4] Andrey Ignatov et al. “Ai benchmark: Running deep neural networks on android smartphones”. In: Proceedings the European Conference on Computer Vision (ECCV) Workshops. 2018, pp. 0–0.
[5] Dario Allegra et al. “A review on food recognition technology for health applications”. In: Health Psychology Research 8.3 (2020).
[6] Nnamdi Ezeora et al. “Real-time food recognition and classification system to aid diabetic patients - systematic review”. In: International Journal of Advanced Research in Computer and Communication Engineering 11.9 (2022).
[7] Wei Liu et al. “Ssd: Single shot multibox detector”. In: European conference on computer vision. Springer. 2016, pp. 21–37.
[8] Mrs FA Ajala et al. “IMPLEMENTATION OF NIGERIAN INDIGENOUS FOOD IMAGE RECOGNITION SYSTEM”. In: International Journal of Software and Hardware Research in Engineering (2020).
[9] Zhihui Fu, Dan Chen, and Hongyu Li. “Chinfood1000: A large benchmark dataset for chinese food recognition”. In: International Conference on Intelligent Computing. Springer. 2017, pp. 273–281.
[10] Mohammed A Subhi and Sawal Md Ali. “A deep convolutional neural network for food detection and recognition”. In: 2018 IEEE-EMBS conference on biomedical engineering and sciences (IECBES). IEEE. 2018, pp. 284–287
[11] Jianing Qiu et al. “Mining discriminative food regions for accurate food recognition”. In: arXiv preprint arXiv:2207.03692 (2022).
[12] Lili Pan et al. “Image augmentation-based food recognition with convolutional neural networks”. In: Comput-Mater. Continua 59.1 (2019), pp. 297–313.
[13] Yue Geng. “Development of Mobile Food Recognition System Based on Deep Convolutional Network”. In: 3D Imaging—Multidimensional Signal Processing and Deep Learning. Springer, 2022, pp. 77–89.
[14] Fotios S Konstantakopoulos, Eleni I Georga, and Dimitrios I Fotiadis. “Mediterranean Food Image Recognition Using Deep Convolutional Networks”. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE. 2021, pp. 1740–1743.
[15] Mohd Norhisham Razali et al. “Indigenous food recognition model based on various convolutional neural network architectures for gastronomic tourism business analytics”. In: Information 12.8 (2021), p. 322.

