Deep Learning Based On Food Image Recognition

Shreyaa Sridhar
School of Computing and Engineering
University of Missouri-Kansas City
ssn2m@mail.umkc.edu

Geovanni Nicque West
School of Computing and Engineering
University of Missouri-Kansas City
gwrn2@mail.umkc.edu

Khushbu Kolhe
School of Computing and Engineering
University of Missouri-Kansas City
kkvcm@mail.umkc.edu

Naga Venkata Satya Pranoop Mutha
School of Computing and Engineering
University of Missouri-Kansas City
nmgfg@mail.umkc.edu

Abstract

Good nutrition is an important part of leading a healthy lifestyle. Food choices that we make each day affect our health - how we feel today, tomorrow and in the future. In addition to that, predicting calories also plays an important role because it allows for you to have a tight handle on how many calories you are presumably taking in. We propose an application which recognizes food and predict calories, also a comparison of shallow and deep learning model accuracy based on the image taken.

Keywords: Deep learning, food image recognition, food-101 dataset

I. INTRODUCTION

Accurately recognizing food images is a particularly challenging task based on the nature of images, which is why approaches like Spark API models in the field have achieved a low classification accuracy. Deep neural networks have outperformed such solutions, and we present a approach where we are comparing the accuracy of different shallow and deep learning models. Here we are also comparing the accuracy our models with that of Clarifai API model.

II. RELATED WORK

NutriNet model [2] was built to recognize food and drinks. NutriNet+ was built on 6 convolutional layers when compared to AlexNet it has one extra layer.This convolutional layer was added to gain additional knowledge about the features in the higher resolution images. NutriNet model uses UNIMIB 2016 Food image dataset. NutriNet with the NAG solver was the best-performing model with a classification accuracy of 94.47%.This model majority of images are correctly classified in the recognition dataset, but not necessarily all of them. As a consequence, this could lower the classification accuracy for real-world images. Finally, since image segmentation was not performed, irrelevant items present in the images made the recognition difficult. When comparing NutriNet to NutriNet+, we can see that the extra convolutional layer did not yield any performance increase, as NutriNet+ models achieved results that are almost identical to the results by NutriNet models.

A food image recognition system that uses the multiple kernel learning method was introduced, which tested different feature extractors, and their combination, on a self-acquired dataset [5]. This proved to be a step in the right direction, as the authors achieved an accuracy of 26% to 38% for the individual features they used and an accuracy of 61.34% when these features were combined; the features include color, texture and SIFT information. Upon conducting a real-world test on 166 food images taken with mobile phones, the authors reported a lower classification accuracy of 37.35%, which was due to factors like occlusion, noise and additional items being present in the real-world images. The fact that the combination of features performed better than the individual features further hinted at the need for a more in-depth representation of the food images.

Pairwise local features method, which applies the specifics of food images to their recognition [6], analyzes the ingredient relations in the food image, such as the relations between bread and meat in a sandwich, by computing pairwise statistics between the local features. The authors performed an evaluation of their algorithm on the PFID dataset and achieved an accuracy of 19% to 28%, depending on which measure they employed in the pairwise local features method. However, they also noted that the dataset had narrowly-defined food classes, and after joining them into 7 classes, they reported an accuracy of 69% to 78%. This further confirmed the limitations of food image recognition approaches of that time: if a food image recognition algorithm achieved a high classification accuracy, it was only because the food classes were very general (e.g., "chicken").

III. APPROACH

In this project, we have decided to use deep learning technologies to design a model which is capable of predicting food.

A. Data Sources

- We have taken images of food from Food-101 dataset https://www.vision.ee.ethz.ch/datasets_extra/food-101/
- This dataset contains 101 food categories and 1000 images in each category.
- We have 10 food image categories - caesar salad, caprese salad, donuts, dumplings, french fries, greek salad, guacamole, hotdog, risotto and sushi.

B. Analytical Algorithms / Platform

For analysis we have used :

- Clarifai API Model.
- Shallow Learning : Spark API Model.
- Deep Learning : Inception V3 Model and MobileNet Model

Training and testing data is mostly considered as 70% for training data and 30% for testing.

C. Analytical Tools and IDE Used

- Apache Spark
- Android Studio
- Tensorflow
- IntelliJ
- PyCharm

D. Expected Inputs/Outputs

Input - Images from Food-101 dataset

Output - Image Categorization, Image Prediction, Statistical parameters like accuracy, precision, confusion matrix, etc.

E. Evaluation/Validation

- **Clarifai API Model**

Accuracy of our model is 95.2%

- **Spark API Model**

Accuracy of our model is 44.5%

- **Inception V3 Model**

Accuracy of our model is 88.4%

- **MobileNet Model (TensorflowLite)**

Accuracy of our model is 83.8%

F. Model Architecture

The architecture diagram in figure 1 explains how our model works. Images from Food-101 dataset is provided as input to our model.

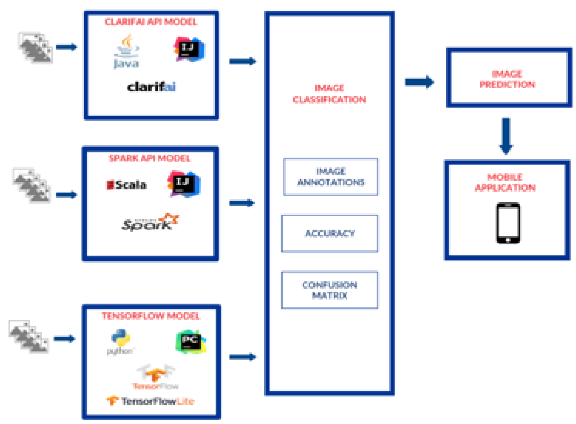


Figure 1 : Architecture Diagram

Here we have compared how our dataset works with different models such as Clarifai API model , Spark API model and Deep Learning models(SoftMax Regression Model, Convolutional Neural Network

Model and Inception V3 Model) . The next step is running our model to obtain annotations for each image, check accuracy and obtain confusion matrix.

G. Features, workflow, technologies

ACTIVITY DIAGRAM

The activity diagram figure 2 shows the workflow of our model.

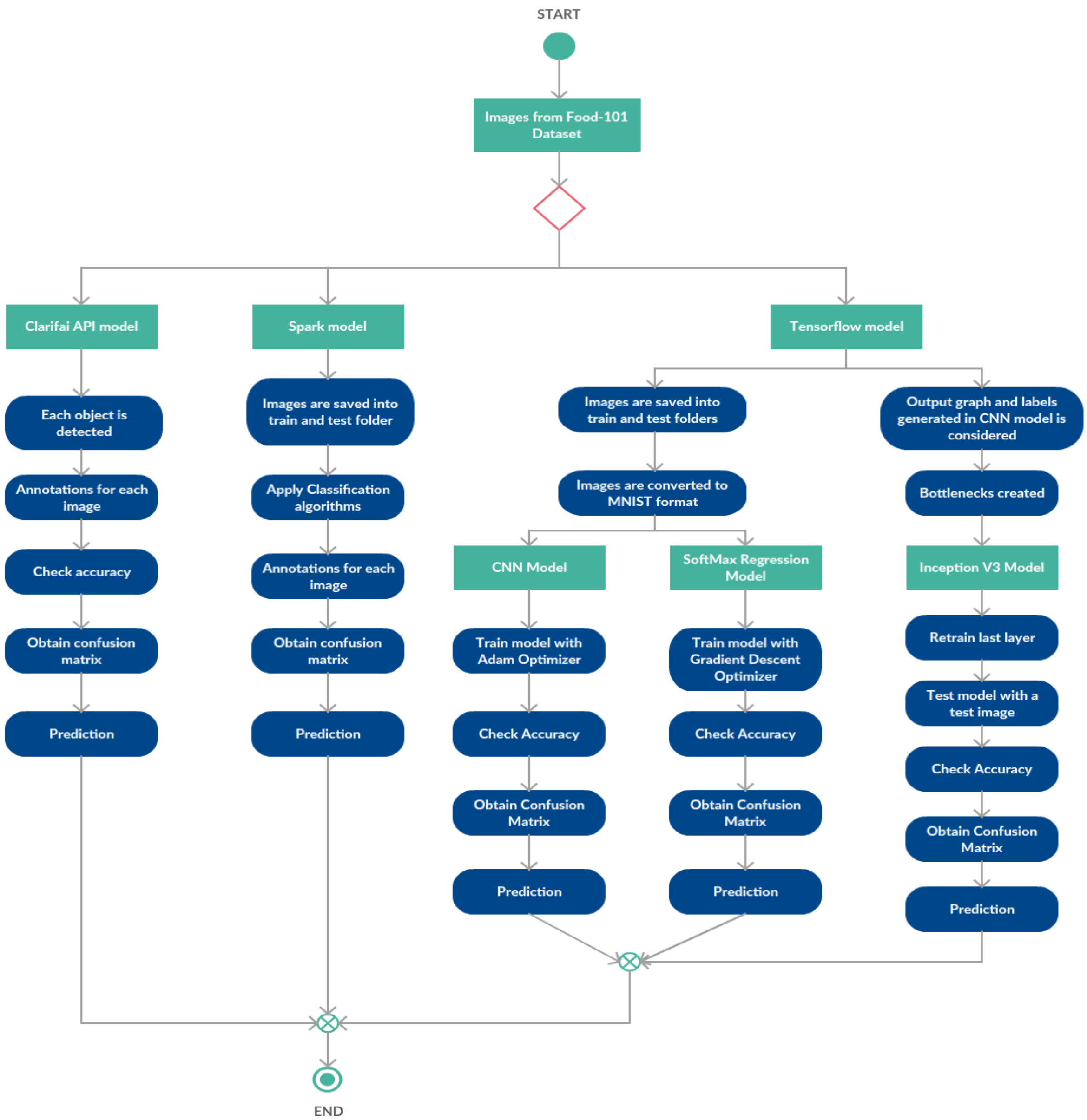


Figure 2 : Activity Diagram

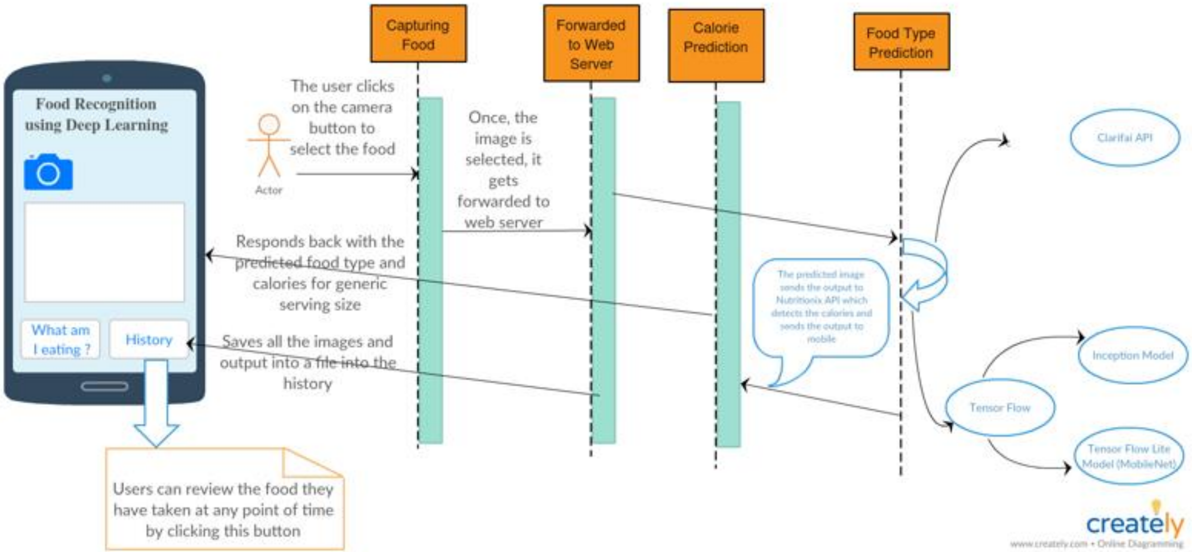


Figure 3 : Sequence Diagram

SEQUENCE DIAGRAM

The sequence diagram figure 3 shows the interaction or collaboration within our model.

FEATURE SPECIFICATION

- Food Image Recognition
- Predict Calories using Nutritionix API

H. Existing Applications/Services Used

- We have used **Clarifai API service** which automatically tags the images or video file in such a way ensuring a quick organization, management and searching throughout the content.
  URL: https://www.clarifai.com/
- **Spark API** offers easy-to-use APIs, for operating on large datasets, across languages: Scala, Java, Python, and R.
- We have used the **Tensorflow** for Deep learning models such as Convolutional Neural Network Model, SoftMax Regression Model and Inception V3 Model.
- **TensorflowLite** uses MobileNet Model, a family of mobile-first computer vision models designed to effectively maximize accuracy while being mindful of the restricted resources for an on-device or embedded application.

IV. IMPLEMENTATION DETAILS

A. Experimentation results of Clarifai API, Spark API and Tensorflow

CLARIFAI API MODEL

In Clarifai API,the type of prediction is based on what model you run the input through. For example, if you run your input through the 'food' model, the predictions it returns will contain concepts that the 'food' model knows about. Clarifai API has predicted an accuracy of 95.2%.

Figure 4 : Clarifai API image prediction

```
[44, 1, 0, 1, 0, 3, 1, 0, 0, 0]
[0, 50, 0, 0, 0, 0, 0, 0, 0, 0]
[0, 0, 47, 2, 1, 0, 0, 0, 0, 0]
[1, 0, 0, 48, 0, 0, 0, 0, 0, 1]
[0, 0, 0, 0, 50, 0, 0, 0, 0, 0]
[0, 2, 0, 0, 0, 48, 0, 0, 0, 0]
[1, 0, 0, 0, 0, 0, 47, 0, 0, 2]
[0, 0, 2, 0, 2, 0, 0, 46, 0, 0]
[1, 0, 0, 0, 0, 0, 0, 0, 49, 0]
[3, 0, 0, 0, 0, 0, 0, 0, 0, 47]
```

Figure 5 : Confusion matrix of Clarifai API

SPARK API PREDICTION

Spark API works well on large dataset and by using the dataset on different models like Random Forest Model, Decision Tree Model and Naive Bayes Model , we found that Random Forest Model works well for our model. The accuracy of our model using Spark API is 44.04%

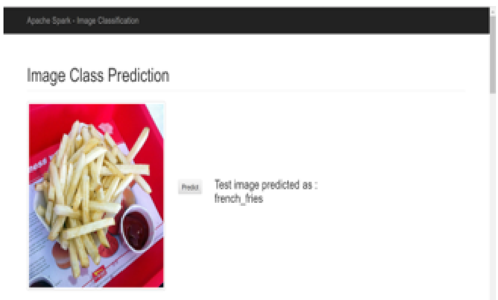


Figure 6 : Spark API Prediction

TENSORFLOW MODEL

In our project, we are using Inception V3 Model for food image recognition. When comparing the accuracies of Softmax Regression, CNN, Inception V3 and MobileNet models, we found Inception V3 and MobileNet models works better and chose that as our Deep Learning models. Our Inception V3 model has predicted an accuracy of 88.4% and MobileNet models predicted an accuracy of 83.8%.

```
[43, 0, 0, 0, 0, 2, 3, 0, 1, 1]
[0, 38, 0, 0, 0, 6, 0, 1, 1, 4]
[0, 0, 46, 0, 1, 1, 0, 1, 0, 1]
[1, 0, 0, 47, 0, 1, 0, 0, 1, 0]
[0, 0, 0, 0, 50, 0, 0, 0, 0, 0]
[6, 2, 0, 0, 0, 41, 0, 0, 0, 1]
[2, 0, 0, 0, 0, 0, 47, 0, 1, 0]
[0, 0, 0, 0, 3, 0, 1, 45, 0, 1]
[2, 0, 1, 0, 0, 4, 1, 0, 42, 0]
[1, 1, 2, 0, 0, 1, 1, 0, 1, 43]
```

Figure 7 : Confusion matrix for Inception V3 model

```
[38, 2, 0, 1, 0, 6, 2, 0, 0, 1]
[1, 43, 0, 0, 0, 3, 0, 0, 0, 3]
[0, 3, 43, 1, 0, 1, 0, 1, 0, 1]
[1, 0, 0, 47, 0, 1, 0, 1, 0, 0]
[0, 0, 0, 0, 49, 0, 0, 1, 0, 0]
[5, 6, 0, 0, 0, 38, 1, 0, 0, 0]
[4, 1, 0, 0, 0, 3, 38, 1, 2, 1]
[0, 0, 0, 3, 2, 0, 0, 43, 0, 2]
[2, 0, 2, 1, 0, 3, 2, 0, 38, 2]
[1, 1, 2, 1, 1, 1, 0, 0, 1, 42]
```

Figure 8 : Confusion matrix for MobileNet model

B. Mobile Application UI

User selects the camera icon to take picture of the food. The food image is sent to the server which handles the image recognition . Here we have included history which displays all the food history the user has checked previously. Calories are obtained by Nutritionix API considering any food image as for one serving.

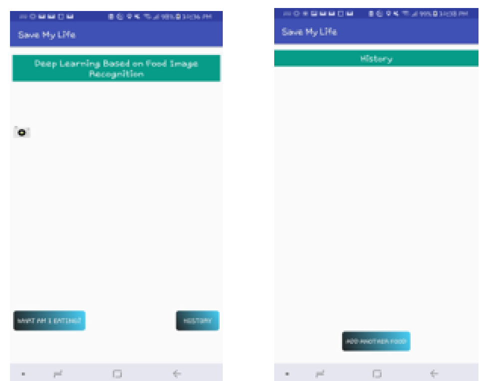


Figure 9 : Before taking picture UI of Main Page and History Page

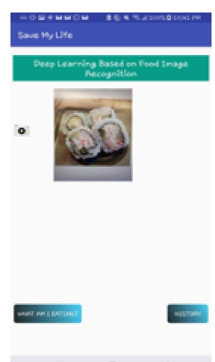


Figure 10 : Picture taken using Camera

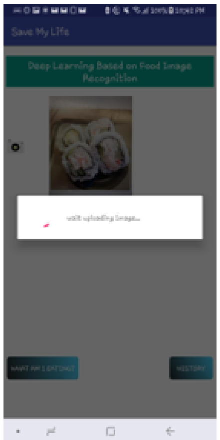


Figure 11 : Computing Prediction

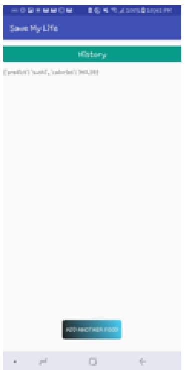


Figure 12 : Food Image Prediction

V. RESULTS

A. Datasets

FOOD-101 DATASET

We have taken images of food from Food-101 dataset https://www.vision.ee.ethz.ch/datasets_extra/food-101/

This dataset contains 101 food categories and 1000 images in each category. We have considered 10 food image categories for our project - caesar salad, caprese salad, donuts, dumplings, french fries, greek salad, guacamole, hotdog, risotto and sushi. We have considered similar categories like three different salads and our models predicts well all the categories.

B. Evaluation

ZENHUB BOARDS

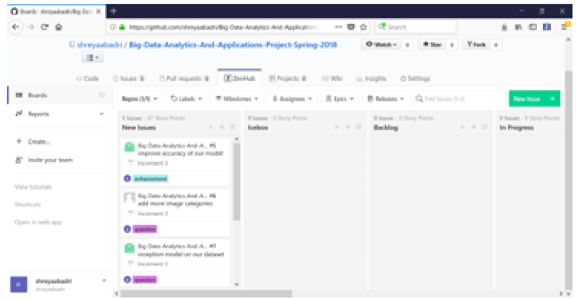


Figure 13 : New Issues for Increment 3 created

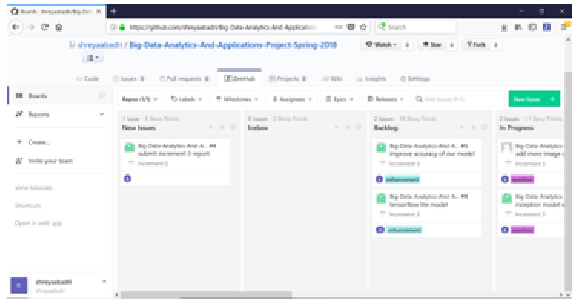


Figure 14 : Zenhub Board progress

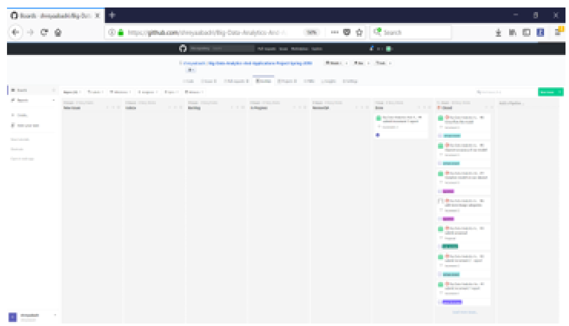


Figure 15 : Increment 3 final board

BURNDOWN CHART

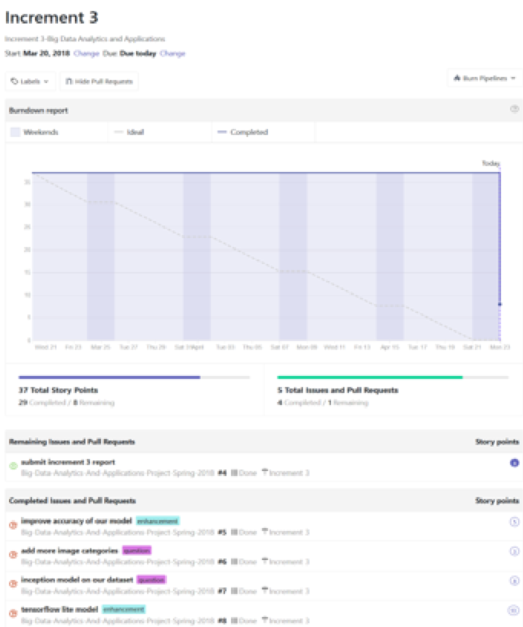


Figure 16 : Burndown chart for increment 3

COMPARE ACCURACY OF CLARIFAI API , SPARK API AND TENSORFLOW MODELS

Model		Accuracy
Clarifai API Model		95.2%
Spark API Model (Random Forest Model)		44.04%
Tensorflow Model	Inception V3 Model	88.4%
	MobileNet Model	83.8%

Figure 17 : Accuracy of our model

Based on the dataset considered, the Clarifai API model was the best model based on the accuracy of 95.2%. Whereas, the Inception V3 model was found to be the best of Deep Learning Models with an accuracy of 88.4%.

VI. DISCUSSION & LIMITATIONS

This data set contains of 101 food types,each food type consisting of 1000 images. In comparison among different models used for our dataset, we have got the best accuracy for Clarifai API Model and least for Spark API model. Though, the accuracy looks very low except on Clarifai, it is mainly because we restricted input to only ten food types, which is minute compared to the data set we have taken for reference. We tested with multiple food types, but taking five or six food types at a time. We observed that the accuracy varies for the models (i.e not one model stays low every time, they keep altering). We have used similar types of foods like three different types of salads to check the efficiency of our model. In addition, we also try to display the calories present in each food type, and the accuracy comparison for each of the models. The solution to lost data is still under study. The provisions for editing or making changes in the app are not yet to be figured out.

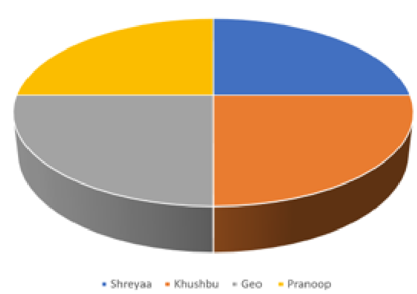
VII. CONCLUSION

In this paper, we have described our study of developing a food image recognition mobile application. Our dataset was trained on across different shallow and deep learning models. After comparing the accuracies of different models, we inferred that Inception V3 model works best and provides accurate result. As future work, we plan to determine portion size of food images.

VIII. APPENDIX

Plan & Project Timelines, Members, Task Responsibility

Shreyaa Sridhar (21) - 25%
Khushbu Kolhe (9) - 25%
Geovanni Nicque West (23) - 25%
Naga Venkata Satya Pranoop Mutha (15) - 25%



IX. REFERENCES

(1) Yuzhen Lu, "Food Image Recognition by Using Convolutional Neural Networks (CNNs)", Cornell University Library,

Computer Vision and Pattern Recognition,Dec 2016, arxiv:1612.00983

(2) Simon Mezgec and Barbara Korousic Seljak, “NutriNet: A Deep Learning Food and Drink Image Recognition System for Dietary Assessment” NCBI, Nutrients, v.9(7), 2017 July, ,PMC5537777

(3) Chen M., Dhingra K., Wu W., Yang L., Sukthankar R., Yang J. PFID: Pittsburgh Fast-Food Image Dataset; Proceedings of the ICIP 2009; Cairo, Egypt. 7–10 November 2009; pp. 289–292.

(4) 5. Lowe D.G. Object Recognition from Local Scale-Invariant Features; Proceedings of the ICCV’99; Corfu, Greece. 20–21 September 1999; pp. 1150–1157.

(5) 6. Joutou T., Yanai K. A Food Image Recognition System with Multiple Kernel Learning; Proceedings of the ICIP 2009; Cairo, Egypt. 7–10 November 2009; pp. 285–288.

(6) 7. Yang S., Chen M., Pomerlau D., Sukthankar R. Food Recognition using Statistics of Pairwise Local Features; Proceedings of the CVPR 2010; San Francisco, CA, USA. 13–18 June 2010; pp. 2249–2256.

(7) Artifact from the Future: GOTTA EAT 'EM ALL http://www.iftf.org/future-now/article-detail/artifact-from-the-future-gotta-eat-em-all/