HEALTHY VS UNHEALTHY: FOOD RELATED TWEETS AND IMAGE CLASSIFICATION

by

Tejaswini Oduru

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Technology

Charlotte

2020

Approved by:

_____

Dr. Albert Park

_____

Dr. Mohamed Shehab

_____

Dr. Yaorong Ge

ABSTRACT

TEJASWINI ODURU. Healthy vs Unhealthy: Food Related Tweets and Image Classification. (Under the direction of DR. ALBERT PARK)

Many studies have proved that the rise in obesity among the world population is due to an increase in calorie intake coupled with a lack of adequate physical activity. Food is an essential part of everyday life and has significant effects on our health and wellbeing. Although taking nutrients is primary, eating attitudes and behaviors also prevent chronic illness and mental health. There are many applications for keeping track of what we eat manually, but tools for detecting the healthy level of the food in the image are rare. People nowadays are influenced by social media and tend to post the images of food they consume on daily basis. These images represent the behavior and attitudes of users towards their health and calorie intake. Hence, tools for automatic food recognition could significantly alleviate the issue of maintaining a balanced diet not only at an individual level but also helps to understand general eating behavior of population. This study presents a deep learning architecture of food detection with levels of healthiness with transfer learning from a pre-trained classification model 152 residual layer network. It is performed in two steps. First, transfer learning is performed on the images to train the model with transferred features from the classification to boost the prediction. The model's accuracy was more than 80 percent for both multi-class classification. Second, we manually evaluated the performance of the model using Twitter images to better understand generalizability of our methods. The results show that the model is able to predict the images into their respective

classes including Definitely Healthy, Healthy, Unhealthy and Definitely Unhealthy with approximately 80 percent accuracy.

ACKNOWLEDGMENTS

I am highly indebted to Prof. Albert Park, and obliged for giving me the autonomy of functioning and experimenting with ideas. I would like to take this opportunity to express my profound gratitude to him not only for his academic guidance but also for his personal interest in my report and constant support coupled with confidence boosting which proved very fruitful and were instrumental in infusing self-assurance and trust within me. The nurturing and blossoming of the present work is mainly due to his valuable guidance, suggestions, astute judgment, constructive criticism and an eye for perfection. My mentor always answered myriad of my doubts with smiling graciousness and prodigious patience, never letting me feel that I am novices by always lending an ear to my views, appreciating and improving them and by giving me a free hand in my report. It's only because of his overwhelming interest and helpful attitude, the present work has attained the stage it has. Finally, I am grateful to our University and colleagues whose constant encouragement served to renew my spirit, refocus my attention and energy and helped me in carrying out this work.

Tejaswini Oduru

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

GPU	Graphic Processing Unit
DNN	Deep Neural Network
CNN	Convolutional Neural Network
SGD	Stochastic Gradient Descent

CHAPTER 1: INTRODUCTION

Obesity is defined as excess or abnormal fat, which builds up and poses a health risk. Being obese is associated with an increased risk of many diseases, including high cholesterol, stroke, type 2 diabetes, high blood pressure, heart attack, arthritis, gallbladder disease, and some cancers [1]. According to the estimates, 1.9 billion adults worldwide are reported to be either overweight or obese, in 2016. Obesity's health outcomes were responsible for 2.8 million preventable deaths per annum [2]. Monitoring dietary intake can play a prominent role in individual and public health. Food marketing on social media influences the intake of high-energy and low-nutrient foods such as fried foods, candy [3]. Several chronic illnesses and diseases such as cardiovascular disease, obesity, cancer is associated with increased consumption of these high caloric foods. This also leads to individual and country-level economic losses [4]. Due to this reason, understanding public food behavior is essential and can be performed using image-based food detection [5].

Food image recognition provides a means of determining the eating behavior of individuals by getting a healthy level of the food they consume. There is a need for an automatic method to ease the monitoring of the public's unhealthy and healthy eating habits [6]. However, due to the nature of food items, the problem of food image detection and recognition is complex [7]. Typically, foods are deformable objects, which makes it difficult to define their structure [7]. Also, certain types of food may have a high intra-class variance (related foods look very different) and a low inter-class variance (different foods look very similar), making the task of determining the type of food much more

complicated. Therefore, understanding the types of food can be just as informative understanding the exact ingredient for the purpose of predicting public heath trend.

Sixty five percent of all-American adults and 90% of young American adults now use social networking sites. One of the popular reasons for this usage is viewing and sharing pictures of food they take [6]. More than 10% of the images on social media sites are of food; pictures of delicious, enticing foods are pictures of food porn [7]. For example, in magazines, blogs, television, cookbooks, and social media networks like Pinterest, Twitter, Facebook, and Instagram, food porn involves images and videos of enticing food in the media. An increasing number of people share pictures of their cooked and taken food through social media such as Instagram and Twitter [5].

Computer vision is a field that can process images to uncover content of images. Recent efforts have been made to develop automatic object recognition [8] and image classification [8]. Food detection has not obtained adequate results, considering the significantly improved performance of the current state-of-the-art object detection approaches. The representation of food images plays a fundamental role in understanding engine retrieval and classification, as defined in a survey of studies in food image processing. Food retrieval and variety are difficult tasks because high variability and intrinsic deformability are present in the food [9]. The challenges come from different causes: the large variety of intraclass variants and the massive number of categories. In this context, our work focuses on bridging the gap between strategies for classification and detection to tackle the challenge of food classification. Understanding the types of food can be just as informative understanding the exact ingredient for the purpose of predicting public heath trend. We implement transfer learning from a pre-trained classifier, ResNet

[10], then generalize to our food type classifier by transferring features. Our experiment results show a significant result for classifying a range of definitely healthy to definitely non-healthy food. This research contributes as follows. First, we propose a transfer-learning-aided approach to improve the classifier performance. Second, we demonstrate the generalizability by conducting experiments by applying on social media images.

CHAPTER 2: BACKGROUND

2.1. Obesity and Its Consequences

While the effects on body weight are hereditary, behavioral, metabolic, and hormonal, obesity happens when there is a little activity, and daily activities surpass calorie intake. These calories do not burn, and the body accumulates these extra calories as fat [1]. Obesity has more than doubled globally in the years 1980 to 2014. Around 38 percent of men and 40 percent of women aged 18 or older were regarded as overweight in 2014. Besides, 11 percent of males and 15 percent of females were obese. At present, obesity is accountable for 5% of deaths, which also leads to a higher number of deaths worldwide than the underweight. If this continues at this rate of increment, studies show that obesity reduces life expectancy by eight years. It was estimated that the global economic impact of obesity was the US $2.0 trillion [2].

2.2. Social Media Impact

For several people, the influence of social media has advanced over the past decade from being merely an entertainment platform to a fully integrated part of almost every aspect of life [11]. On each platform, there is an innumerable number of results for hashtags related to the food, health and fitness. There are more than 93 million selfies worldwide, which are uploaded on social media platforms every day [3]. These images mainly contain the food images in restaurants, cooked food along with the tags and location. Below images show examples of twitter and Instagram posts.

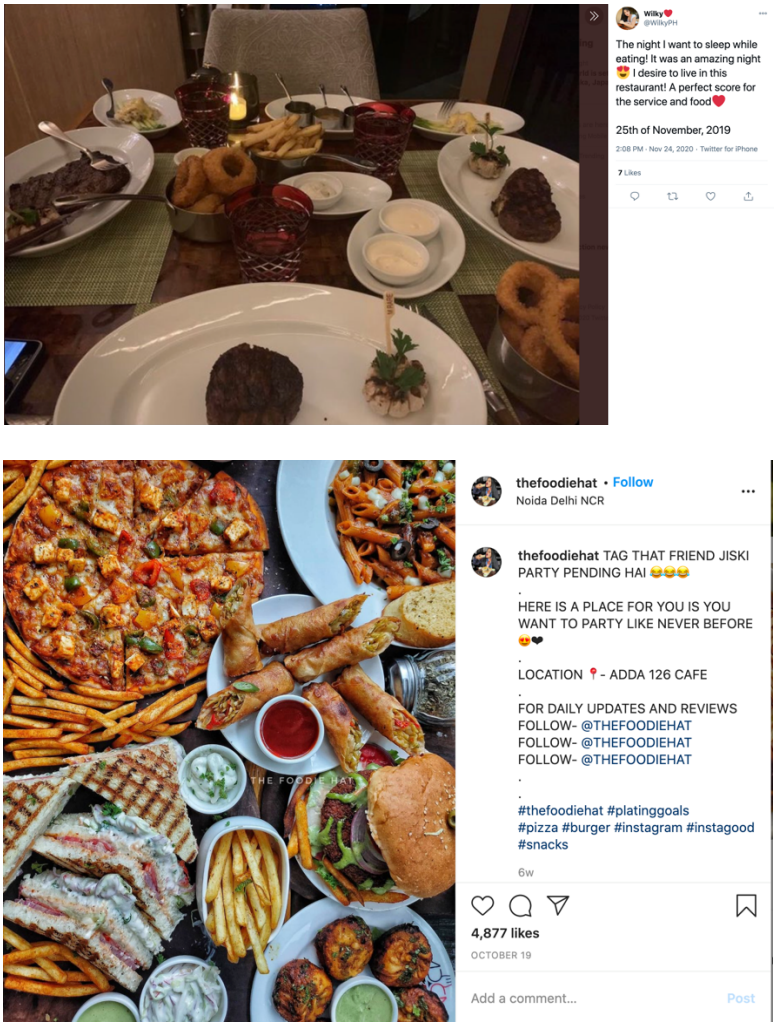

Figure 1: Example Images of restaurant food on twitter and Instagram

Motivated by the obesity epidemic in the United States, Instagram pictures taken at 164,753 restaurants by millions of users were analyzed [11] to understand 1. the relationship between fast food and chain restaurants and obesity, 2. people's thoughts on and perceptions of their daily dining experiences, and 3. the nature of social reinforcement and approval in the context of dietary health on social media. When the prominence of fast-food restaurants in US counties correlates with obesity, the Foursquare data shows a more significant correlation at 0.424 than official survey data from the County Health Rankings would show. The analysis further reveals a relationship between small businesses and local

foods with better dietary health, with such restaurants getting more attention in areas of lower obesity. However, social approval favors unhealthy foods high in sugar even in such places, with donut shops producing the most liked photos. They found that both safe and unhealthy tags trigger more likes, reflecting the discovery that social forces are both for and against unhealthy lifestyles. Good and unhealthy tags are associated with similar tastes in areas of low obesity, although this distinction is more pronounced in areas of high obesity. Therefore the dietary environment exposed by the study is a dynamic ecosystem, with fast food playing a role alongside social interactions and personal preferences, which can often be at odds [11].

In another study, Twitter's ability is analyzed to provide insight into US-wide dietary preferences by associating the tweeted dining experiences of 210K users to their choices, social networks, and demographics [12]. They verify this methodology by examining the caloric values of the foods mentioned in the tweets to the state-wide obesity rates, obtaining a Pearson correlation of 0.77 across the 50 US states and the District of Columbia. A model based on a combination of demographic variables and food names listed on Twitter is designed to forecast county-wide obesity and diabetes statistics. In addition, this data is relevant to economic and social variables, such as wages and education. This is further linked to societal and economic factors, such as education and income. This indicates that places with higher education levels, for example, tweet about food that is slightly less caloric [12].

Authors present the first large scale content analysis of Instagram posts, discussing both the image and the related associated hashtags, aimed at understanding the content of partially labelled images taken in-the-wild and the relationship with hashtags that people

use as noisy labels in this study [13]. In particular, they study the feasibility of learning to recognize food image content in a data driven way, exploring both the categories of food, and how to recognize them, purely from social network data. Even without using manual annotation, this approach to food recognition can often achieve accuracies higher than 70% in identifying popular food-related image categories.

In a study, initiatives to decrease neighborhood-based health inequalities need access to meaningful, timely, and local information concerning health behavior and its determinants were taken [14]. The authors primarily examine Twitter's validity as a source of information for the analysis of dietary patterns and attitudes. The healthiness quotient of food-related tweets and sentiment regarding those tweets from metropolitan Detroit is analyzed. Findings show the feasibility of using Twitter to understand neighborhood characteristics of food attitudes and potential use in studying neighborhood-based health disparities.

2.3. Computer Vision

Computer vision has progressively become a widely recognized technology in diverse fields such as image processing [15], face recognition [16], object detection [17], and medical research [18]. In contradiction to the conventional methods that need a long time to execute, computer vision has been ventured into a branch of artificial intelligence for profound learning, quick and better analysis. It supports researchers in analyzing photographs and videos to obtain valuable information, unique patterns, and understand descriptions [19].

The advancements in machine learning and vision-based computer applications have paved the way for more efficient dietary evaluation methods. The surge of interest in deep learning methods is because they have been shown to outperform previous state-of-the-art techniques in several tasks and the abundance of complex data from multiple sources. The general purpose of these vision-based methods is to identify the food. The general role of recognizing the food is to provide these vision-based approaches. With deep learning algorithms, food detection and recognition precision have been significantly improved to more than 70 percent accuracy [20].

2.4. Deep Learning

Deep learning is a way of representation-learning that improves multilevel design by using the deep artificial neural network made of multiple layers of neurons. Deep learning models show strong capabilities in classification tasks, provided that sufficient data were available, representing a particular problem [17]. With the powerful automatic feature learning ability, the deep learning method starts to be applied in food science, mainly referring to food category recognition, fruit, and vegetable quality detection, food calorie estimation.

2.5. ImageNet

The explosion of image data on the Internet can encourage more sophisticated and robust models and algorithms to retrieve, index, organize and interact with images and multimedia data [21]. But a fundamental issue is to determine how such knowledge can be utilized and regulated precisely. Hence, a new database called ImageNet has been introduced, a large-scale ontology of images built upon the WordNet structure's backbone. ImageNet aims to populate most of the 80,000 sets of WordNet with an average of 500-

1000 clean and full resolution images, which results in tens of millions of annotated images organized by the semantic hierarchy of WordNet [22]. ImageNet is proved to be much larger in scale and diversity and much more reliable than the existing image datasets. Constructing such a large-scale database is a challenging task. ImageNet is useful in three simple applications in object recognition, image classification, and automatic object clustering.

In a research study, a deep convolutional neural network is trained to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes [23] . The top-1 and top-5 error rates of 37.5% and 17.0% are achieved on the test data, which is considerably better than the previous state-of-the-art. The neural network has 60 million parameters and 650,000 neurons, consists of five convolutional layers, followed by max-pooling layers, and three fully connected layers with a final 1000-way SoftMax. To make training faster, they used non-saturating neurons and a very efficient GPU implementation of the convolution operation.

2.6. ResNet

ResNet is the abbreviated form for Residual Network. Deep convolutional neural networks have made a series of breakthroughs in image classification and recognition. It has become a trend to go deeper to solve more complex tasks and improve classification or recognition accuracy. However, due to vanishing gradient problems and degradation problems, training deeper neural networks has been difficult [24]. Residual learning tries to solve both these problems. Using skip connections or shortcuts to jump over specific layers is an artificial neural network that creates a deeper neural network. Skip connection is a mechanism by which more in-depth layer activations of a particular layer are applied

directly to some other layer's activation in the network. As it is represented in the Figure 1, the activation of layer l-2 is added directly to the activation of layer l which is deeper in the network. This process continues throughout the network, and thus, the activations of the layers are pushed deep into the network. This helps us solve the gradient vanishing issue as the layers' activations towards the beginning of the network have been added directly to the deeper layers. There are different versions of ResNet, including ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ResNet-152 [25]. The numbers denote layers, although the architecture is the same. For our experiment, we have used the ResNet-152 model with 152 layers.

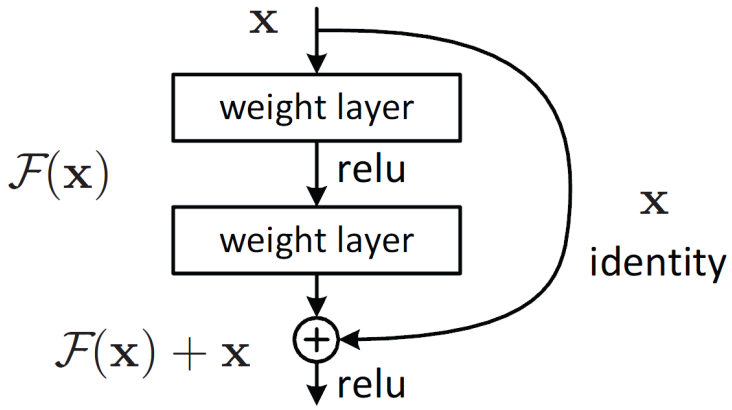

Figure 2: Residual learning: a building block

In neural networks, every layer learns low- or high-level features while being trained for the task at hand. In residual learning, instead of attempting to learn features, the model tries to retain some residual. As we can see in Fig. 1, the input 'x' is being added as a residue to the weight layers' output, and the activation is carried out. Relu activations are being used in the ResNet model [26].

A residual learning framework has been developed to ease the training of networks that are considerably deeper than those used previously [25]. This study provides

comprehensive empirical evidence showing that these residual networks are more comfortable optimizing and can gain accuracy from considerably increased depth. On the ImageNet dataset, residual nets were evaluated with a depth of up to 152 layers---8x deeper than VGG nets but still having lower complexity. An ensemble of these residual nets achieves a 3.57% error on the ImageNet test set.

In this research, the authors examine the convolutional network's effect on its accuracy in the large-scale image recognition environment [27]. The key contribution is a detailed evaluation of networks of increasing depth using an architecture with very small (3x3) convolution filters, which shows that a substantial improvement in prior-art configurations can be achieved by pushing the depth to 16-19 weight layers. They illustrate that representations generalize well to other datasets, where state-of-the-art outcomes are achieved. To encourage more study on the use of deep visual representations in computer vision, they then made two of the best performing ConvNet models publicly accessible. Many classification models were introduced for food detection.

2.7. Food Recognition

Research works in the literature have often focused on various aspects of the food recognition problem. Many papers discuss the challenges of food identification by developing recognition methods that vary in terms of features and classification methodologies [28]. Below are some of the studies on food detection and classification algorithms. An approach to food and drink image detection and recognition that uses a newly defined deep convolutional neural network architecture, called NutriNet, was developed [7]. This architecture was tuned on a recognition dataset containing 225,953 512 × 512-pixel images of 520 different food and drink items from a wide range of food groups,

on which they have achieved a classification accuracy of 86.72%, along with an accuracy of 94.47% on a detection dataset containing 130,517 images. A real-world evaluation on a dataset of self-acquired images, combined with pictures from disease patients, all taken using a smartphone camera, achieved a top-five accuracy of 55% [7].

The technique of automatically building a food diary that records the ingredients consumed can help individuals adopt a balanced diet. A method for adapting a high performing state of the art CNN to serve as a multi-label predictor for learning recipes in terms of their list of ingredients has been proposed in view of the issue of identification of food ingredients as a multi-label learning problem [29]. Given a picture, this model is able to predict its list of ingredients, even if the model has never seen the recipe corresponding to the image. Furthermore, this model trained with a high variability of recipes and ingredients is able to generalize better on new data and visualize how it specializes each of its neurons to different ingredients.

CNN-based food calorie estimation is developed for multiple-dish food images in an approach to food calorie estimation. By multi-task learning of food calorie estimation and food dish detection with a single CNN, this approach estimates food calories while simultaneously detecting dishes [30]. It is anticipated to achieve high speed and save memory by simultaneous estimation in a single network. They use two types of datasets for training a single CNN. For the two types of datasets, multiple-dish food photos with bounding-boxes attached and single-dish food photos with food calories are used. Results show that the multi-task approach achieved higher speed and a smaller network size than a food detection and calorie estimation sequential model.

In general, Dish recognition is very challenging due to different cuisines, cooking styles, and the intrinsic difficulty of modeling food from its visual appearance. A large number of these images are taken in restaurants [31]. However, contextual information can be essential to improve recognition in such a situation. In particular, geo context has been widely exploited for outdoor landmark recognition. Similarly, they also use knowledge about menus and the location of restaurants and test images. First, a framework is adapted based on discarding unlikely categories located far from the test image. Then, the problem is reformulated using a probabilistic model connecting restaurants, dishes, and locations. This model is applied in three different tasks: restaurant recognition, dish recognition, and location refinement. Experiments on six datasets show that the method can boost all task performance by integrating multiple evidence pieces.

The development of automatic nutrition diaries, which would allow keeping track of everything we eat objectively, could enable a whole new world of possibilities for people concerned about their nutrition patterns [32]. A method for simultaneous localization and recognition of food is established. First, generate a food activation map for developing bounding box proposals on the input image (i.e., a heat map of probabilities). Second, identify each of the food types or food-related items present in each bounding box. They show that the proposal can achieve high precision and appropriate recall levels with just a few bounding boxes, compared to the most similar problem nowadays - object localization.

Food image recognition tasks are currently being tested against fixed datasets [33]. However, in real-world conditions, there are cases in which the number of samples in each class continues to increase and samples from novel classes appear. In particular, dynamic datasets in which each individual user creates samples and continues the updating process

often has content that varies considerably between different users, and the number of samples per person is very limited. A single classifier familiar to all users cannot control such dynamic data. Linking the gap between the laboratory environment and the real world has not yet been accomplished on a large scale. Personalizing a classifier incrementally for each user is an assuring way to do this. A personalization problem that involves adjusting to the user's domain incrementally using a very limited number of samples is employed [33]. They proposed a useful personalization framework, a combination of the nearest class mean classifier and the 1-nearest neighbor classifier based on deep features. A new dataset of daily food images collected by a food-logging application is used to conduct realistic experiments. Experimental results show that the proposed method significantly outperforms existing methods.

The types of fruits they consume and their nutrients are essential to individuals because consuming fruits and vegetables is essential in maintaining a balanced diet [34]. This study introduces an automatic way to detect and recognize the fruits in an image to enable keeping track of daily intake automatically using images taken by the user. The proposed method uses state of the art deep-learning techniques for feature extraction and classification. Deep learning methods, especially convolutional neural networks, have been widely used for various classification problems and have achieved promising results. The trained model has performed an accuracy of 75% in the classification of 43 different fruit types. Similar methods have achieved up to 70% with fewer classes.

2.8. Transfer learning of ResNet

First, a base network is trained on a base dataset during transfer learning, and the features learned from the first task are repurposed or transferred to a second network to

train on a second dataset and job. If the features are sufficient for both base and target tasks, this method will work instead of the only simple task [35]. Deploying pre-trained models on similar data has shown strong results in tasks related to image classification. ResNet Model takes weeks to use modern hardware to train [10]. This model can be downloaded and integrated with new models that bring better results with the image as input. The primary objective of transfer learning is to apply a model rapidly and to increase efficiency. Instead of building a new DNN model, the model will transfer the features it has learned from the numerous datasets that have done the same task to solve the current problem. This transaction is also known as knowledge transfer.

A food-specialized detection deep learning architecture with knowledge transferred from a pre-trained food/non-food classification model is developed [36]. Because of their incompatible outputs, existing techniques in object detection all distinguish it from image classification. This work bridges the gap by using transferred features between the two most basic computer vision topics. This work gives a new perspective on object detection. Experiments are performed in two parts. First, transfer learning quantification experiments show that initializing a network with transferred features from classification task can surprisingly produce a boost to generalization for the detection task. Second, experiments on three state-of-the-art neural network backbones show that the approach enables rapid progress and improved performance. The results significantly surpass all original plain networks with more than 10% precision improvement.

2.9. Optimizer

The optimizer and the loss function are the key elements that enable the network to work on the data. Optimizer, in simple terms, sets the learning rate of a neural network.

The optimizer used for this model is Stochastic Gradient Descent (SGD) [37]. SGD optimizer has proven to be performing better than many other optimizers. Choosing a loss function can also be a difficult task. Loss is a measure of the performance of a model. The lower, the better. When learning, the model aims to get the lowest loss possible. It takes in the input from the SoftMax function output and the true label. The trained SoftMax layer is stacked on top of the layers transferred from ResNet-152 to build our deep learning model, which is used to classify the test set [26]. The loss function used to find the loss for this model is Categorical-Cross Entropy. Cross-entropy is a measure of the difference between two probability distributions for a given random variable or set of events. These functions are directly imported from PyTorch packages, torch.nn and torch.optim. torch.optim is a package implementing various optimization algorithms. torch.nn module to help us in creating and training the neural network.

CHAPTER 3: METHODS

3.1. Proposed Method

We developed a food image recognition system that uses deep convolutional neural networks with the image dataset. To classify the image, we have used the transfer learning approach on a pre-trained model. A pre-trained model is a model that was trained on a large benchmark dataset to solve a problem similar to the one that we want to solve [29]. The architecture of the pre-trained model remains the same and it is trained according to the given dataset. The pretrained model used in the experiment is ResNet152, 152-layer residual net is the deepest network ever presented on ImageNet, while still having lower complexity than VGG nets [25]. A food image is provided to the recognition model as the input, and the output is a text class label describing the food item. We have trained the model for multi-class (four class labels: definitely healthy, healthy, unhealthy, and definitely unhealthy). Depending on the model, the class label for the image is predicted as one of the classes. Fig. 2 shows the flow of the experiment.

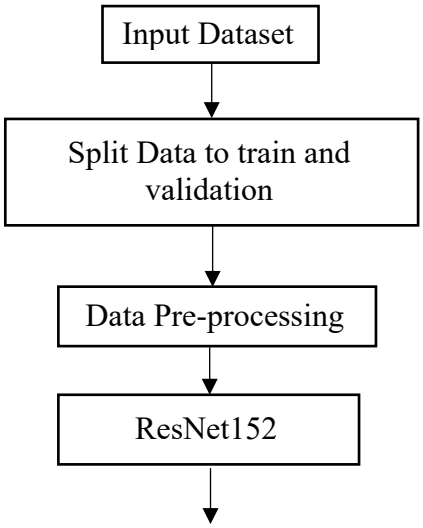

Figure 3: Proposed Architecture

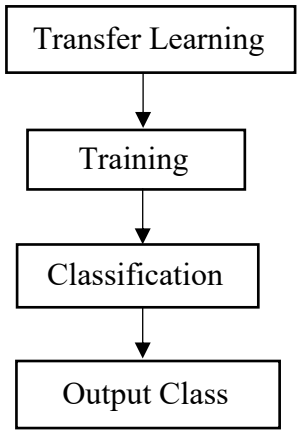

Figure 3: Proposed Architecture (Continued)

The method consists of the following basic steps:

- Given a dataset of labeled food images.
- Center crop each image to 224 × 224 dimension to feed it to the deep neural network;
- Build a deep neural network (DNN) based on the ResNet architecture with 152 layers (ResNet-152), transferring the parameters of ResNet-152 convolutional layers to the convolutional layers of the DNN model.
- Freeze the transferred layer's parameters and train the DNN model to classify each sample into its category.

3.2. Data Collection

To develop a model which can predict the healthy level of the food items, we have considered previous literature [14] where authors have given an exclusive list of food items with respect to their health components by analyzing the tweets for tags and food vocabulary. Hence, we have collected dataset that is representative of definitely healthy, healthy, unhealthy, and definitely unhealthy food images (Table 1). We use Google image search and crawl the search results. We manually verify the images and combined each

food item according to their respective categories: healthy, healthy, unhealthy, and definitely unhealthy. These images can be trained with the ResNet classifier by transferring the features from ImageNet dataset. Each category contains ten types of food images, where each food type had 700 of images on average. The total number of images in the dataset are 37,224. The split was done with 80, 20 percent for train and validation.

Table 1: Image number for each category of food items for classification

Unhealthy				Healthy			
Def. unhealthy		Unhealthy		Healthy		Def. healthy	
Food Type	Number of images	Food Type	Number of images	Food Type	Number of images	Food Type	Number of images
starbucks	1047	pizza	1100	coffee	1099	sushi	841
ice cream	983	grill	833	tea	742	apple	896
chocolate	955	taco bell	961	coconut	892	fish	986
cake	1060	fries	478	rice	834	salad	1124
bacon	873	tacos	1019	turkey	983	pumpkin	1047
cookies	1050	sauce	945	potatoes	738	pineapple	687
icing	1164	steak	1120	chili	887	fruit	921
McDonalds	951	taco	1012	protein	564	eggs	874
coney island	973	oil	952	roasting	871	orange	947
candy	957	chipotle	954	baking	996	oyster	908

3.3. Environment Setup

The experiment was carried out on Jupyter Notebook, in an environment of Pytorch with CUDA 10.1architecture. The System in which the experiment was carried out runs on Ubuntu and Graphic Processing Unit's (GPU's) were used in this experiment. As for hardware support, graphics processing unit (GPU) coupled with Compute Unified Device Architecture (CUDA) Toolkit and the NVIDIA CUDA Deep Neural Network library

(cuDNN, a GPU-accelerated library of primitives for DNNs) produced by NVIDIA company can provide hardware and software acceleration for deep learning computation. NVIDIA's most advanced data center GPU is the Tesla V100. It's hardware is optimized for CUDA software technology. CUDA excels in parallel computing and deep learning algorithms. Graphics processing units (GPUs) have become the main tools for speeding up general purpose computation in the last decade [24]. They offer a massive parallelism to extend algorithms to large-scale data for a fraction of cost of a traditional high-performance CPU cluster, allowing scalability over incomputable datasets through traditional parallel approaches. These toolkits accelerate widely used deep learning frameworks mentioned above. Software and hardware acceleration tools greatly shorten the computing time and have the potential to meet the requirements of real-time data processing. We have used PyTorch, a Python library for GPU-accelerated deep learning. It supports tensor computation with strong GPU acceleration, and DNNs built on a tape-based auto grad system [24].

3.4. Data Pre-processing

In order to use our images with a network trained on the ImageNet dataset, we need to preprocess our images in the same way as the ImageNet network. For that, we need to rescale the images to 224×224 and normalize them as per ImageNet standards. We can use the torch vision transforms library to do that. Here we take a Center Crop of 224×224 and normalize as per ImageNet standards [23]. We resize the images by Center Crop which crops the given image at the center as 224 by 224 pixels square image. Crop can be square or rectangle in shape depending on the size parameter dimensions. We have given the batch size to be 128. Steps per epoch for training are calculated by dividing the total objects in

training with batch size. Hence 37,224 divided by 128 gives approximately 290 steps for each epoch.

CHAPTER 4. RESULTS

4.1. Train and validate classification model

The metrics measured during the training of the dataset were Accuracy and Loss. These metrics were measured for both training and validation data. The results below are shown for both multi-class classification. The accuracy and loss of both training and validation data are tabulated in Table 2 for multi-class classification. Fig. 3 shows the graph of accuracy while training multi-class classification. As shown in the Fig. 3, the number of epochs for training is 15.

Table 2: Training Performance for multi-class classification

Metrics	Value
train-loss	2.5334
train-acc	87.9832
validation loss	2.4864
validation acc	80.6089

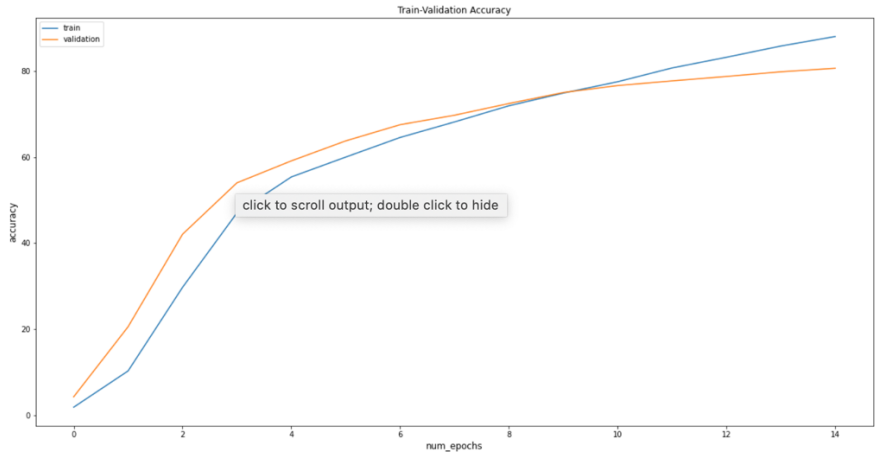

Figure 4: Graph of accuracy for multi-class classification

The trained model is tested on the validation images for all the categories of food. The figures below show the prediction of food images in their respective classes.

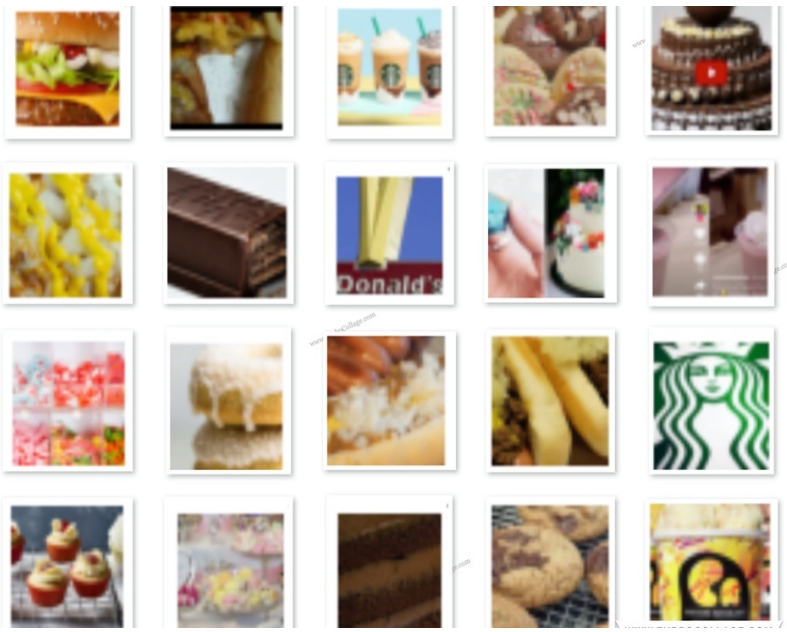

Figure 5: Images predicted as Definitely Unhealthy

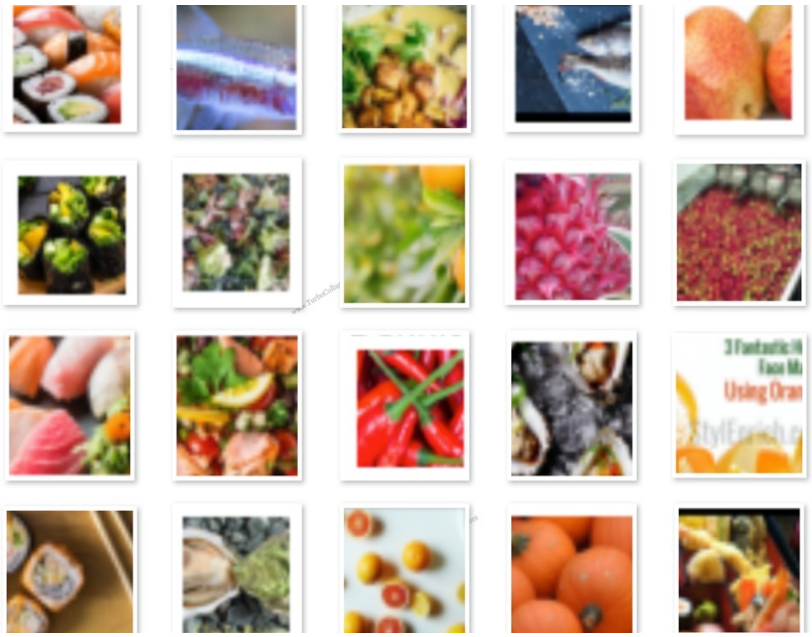

Figure 6: Images predicted as Definitely Healthy

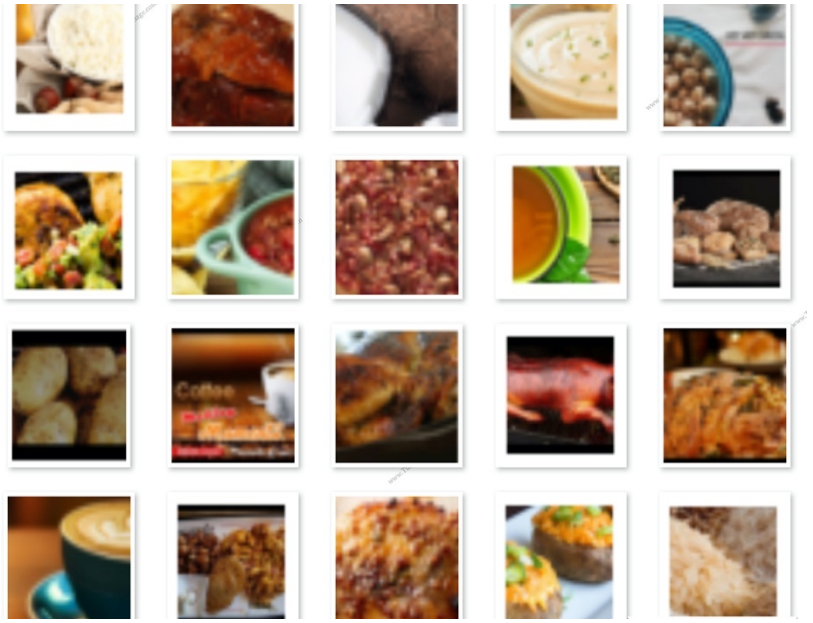

Figure 7: Images predicted as Healthy

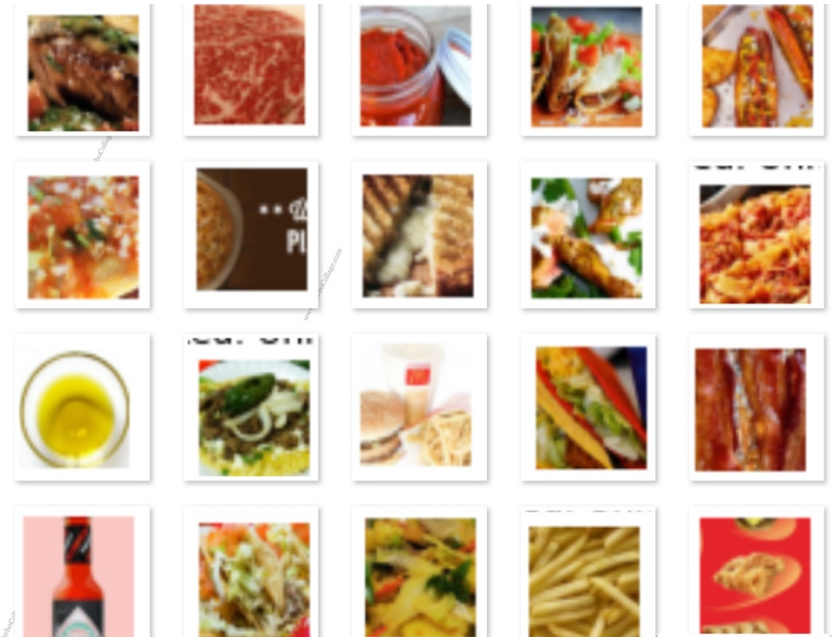

Figure 8: Images predicted as Unhealthy

4.2. Testing on real dataset

The model is further tested on twitter images. Around 20,000 images are collected from the tweet related images using the twitter API. We then manually selected useful

images from the pool of images, where most of the images were of the category's pumpkin, fruit in Definitely Healthy, cake, burger in Definitely Unhealthy, pizza, fries in Unhealthy, coffee, turkey in Healthy. The prediction is shown in the images below.

Figure 9.1: Definitely Unhealthy

Figure 9.2: Definitely Healthy

Figure 9.3: Healthy

Figure 9.4: Unhealthy

We have manually evaluated the twitter images in 10-fold cross-validation. We isolated every set into ten subsets, then every subset is utilized one by one to check the execution of the classification algorithm that is produced, prompting ten independent execution prediction for each image. Prediction accuracy have been utilized as the primary basis for

their grading. All accuracies are acquired by taking the average of the outcomes from 10 executions of 10-fold cross-validation. The average of all these accuracies is found as 73 percent.

We also validated the performance of our classification by calculating the individual performance for each class as shown in Table 3. The performance was evaluated using measures of precision, recall, F1 score and overall accuracy, defined as follows:

$$
\begin{aligned}
precision &= \frac{TP}{TP + FP} \\
recall &= \frac{TP}{TP + FN} \\
F1 &= \frac{2 \times precision \times recall}{precision + recall} \\
accuracy &= \frac{TP + TN}{TP + FN + TN + FP}
\end{aligned}
$$

Figure 10: Metrics Formulae

where TP (true positive), TN (true negative), FP (false positive) and FN (false negative). Recall is also known as the ‘true positive rate’. Precision is also called ‘positive predictive value’, which is the ratio of true positives to combined true and false positives.

Table 3: Individual class performance for multi-class classification

Class	TP	FN	TN	FP	Precision	Recall	Accuracy	F1 Score
Healthy	38	12	35	15	71.6	76	73	73.73
Unhealthy	33	17	30	20	62.2	66	63	64.04
Definitely Healthy	39	11	36	14	73.58	78	75	75.72
Definitely Unhealthy	38	12	36	14	73.07	76	74	74.50

The overall accuracies for all the 200 images are as shown in the Table 4.

Table 4: Overall performance for the proposed method

TP	FN	TN	FP	Precision	Recall	Accuracy	F1 Score
144	56	142	58	71.28	72	71.5	71.63

To better understand the model accuracy, we have specified an individual analysis for each of the class. The result of false negatives for Healthy 8 images out of 12 images were predicted as Definitely Healthy. Similarly, for false negatives of unhealthy class, out of 17 images, 12 images were wrongly classified as Definitely unhealthy. For definitely unhealthy, out of 12 false negatives, 11 of them were incorrectly predicted as unhealthy. In case of definitely healthy class 10 out of 11 false negatives were wrongly classified as healthy. With false positives, out of 15 times where the prediction was wrongly classified as healthy, 10 images were actually belonging to definitely healthy. Out of 20 times where the prediction was wrongly predicted as unhealthy, 14 images were actually belonging to definitely unhealthy. Out of 14 times where the prediction was wrongly classified as Definitely Healthy, 10 images were actually belonging to healthy. Out of 14 times where the prediction was wrongly classified as Definitely Unhealthy, 11 images were actually belonging to unhealthy. In addition, there are some outliers and these images were incorrectly predicted in false positives. The system confuses with the healthy and definitely healthy, unhealthy and definitely unhealthy. In some situations, the model confuses baking with cake and detects definitely unhealthy as healthy. Below figure 11 show example where the model was confused with baking in healthy and cake for definitely unhealthy. Table 5 shows the false negatives and false positives individually for the classes. The number of test images for cake and baking were 15 out of which 3 images in baking were

confused with cake and 4 images in cake were confused with baking in healthy. Table 6 shows the number of images and the number of false positives and false negatives for the cake and baking food items. In our analysis, we found one case that was responsible for lower accuracy, because the google search images for cake and baking are mostly similar because of the recipe of cake.

Table 5: False Negatives and False Positives for the individual classes

	Predicted Healthy		Predicted Unhealthy		Predicted Definitely Unhealthy		Predicted Definitely Healthy	
	FN	FP	FN	FP	FN	FP	FN	FP
Healthy	-	-	3	3	3	3	9	10
Unhealthy	4	4	-	-	-	-	1	5
Definitely Healthy	10	10	1	3	1	3	-	-
Definitely Unhealthy	1	4	11	10	11	10	-	-

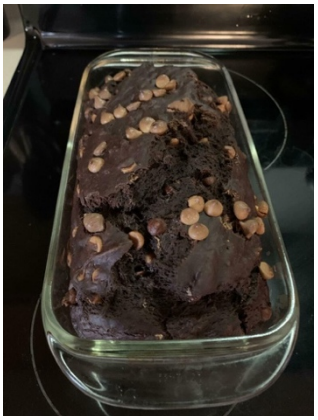

Figure 11: Example image for cake and baking categories

Table 6: False positives and False negatives for the cake and baking

	Predicted as Cake	Predicted as Baking
Cake	3	4
Baking	3	5

CHAPTER 5: DISCUSSION

5.1 Principal Findings

The main result of our research is two-fold: the transfer learning of residual networks and the food image recognition dataset, which contains several different food types and, mainly concentrates on health of the social media users and their behavior. We achieved a promising classification accuracy of 80.60% for the recognition task, which is higher than the accuracy values reported by most of the other deep convolutional neural network approaches in the field.

To test how the trained model, perform in practice, we built a testing dataset containing real-world food images from twitter and the model was able to classify the images into the classes with around 76% accuracy. Ten manual evaluations were made to find the accuracy on the real-world social media images. Each image prediction is manually evaluated to find how many images were correctly predicted out of the 10 images. This process is repeated ten times and finally an average of the accuracies is found. The images were sometimes incorrectly detected as healthy for unhealthy and definitely unhealthy for unhealthy images. To perform the experiments, we have used GPU enabled Ubuntu system and executed them on Jupyter notebook.

5.2 Public Health Implication

This research is used to determine whether the food image is healthy or unhealthy. It helps individuals to maintain their healthcare by monitoring the food they eat. Results of this paper can be used to understand the food attitudes of the people on social media data. Such results illustrate the possibility of using social media to signal the possible intake of

healthy and unhealthy food and related attitudes. This study helps in understanding obesity patterns among social users and also in monitoring the dietary patterns of an individual as well as the general public. This approach is useful for analyzing the user's interests and incentives when sharing food images and, consequently, helps understand an individual's perception of visually appealing and the associated health.

5.3 Limitations and Future direction

A shortcoming of our food recognition system is that the deep learning model is limited to one output per image, which means that not every item gets successfully recognized in images with multiple food items. Overfitting could also be one of the reasons why the classification accuracy is lower on real-world images than on images from the testing subset, with other possible reasons being added noise and occlusion in real-world images and the fact that our recognition dataset could still contain some irrelevant images. We can see clearly in Fig. 3 that the training accuracy surpasses the validation accuracy after 9 epochs which is described as the over-fitting condition. Consequently, this could lower the classification accuracy for real-world images. Finally, since we do not perform image segmentation, irrelevant items present in the training dataset make the recognition task more challenging.

Next step in this study will be to further modify the dataset and test for different social media images other than twitter. Also, a sentiment analysis can be carried on the tweets extracted related to food to find the sentiment of the users towards healthy and non-healthy eating. In the current state of the recognition system, we are classifying 40 different food items. This can be further improved by adding more food items to the dataset in their respective categories.

CHAPTER 6: CONCLUSION

We proposed a food image detection and recognition system that we built, in the scope of which we made use of transfer learning to train the pre-trained residual network with the custom dataset. The use of Transfer learning for food image classification has brought good results. To provide a higher classification accuracy for recognizing food images from the dataset that we acquired using Google image searches, we have used the top classifier with 152 layers. Social media users can use our recognition system to monitor their health. We have also experimentally quantified how transferability benefits object detection tasks from image classification. We found that initializing with transferred features can significantly improve generalization performance. Our work is an intuitive combination of image classification and object detection. With the focus on obesity in relevance to food, we provide a solution to obtain healthy food levels present in the image and monitor the diet and health behaviors of social media users.

REFERENCES

[1] A. Alnuaimi, S. Rawaf, S. Hassounah, and M. Chehab, "Use of mobile applications in the management of overweight and obesity in primary and secondary care," *JRSM Open*, 2019, doi: 10.1177/2054270419843826.

[2] M. Tremmel, U. G. Gerdtham, P. M. Nilsson, and S. Saha, "Economic burden of obesity: A systematic literature review," *International Journal of Environmental Research and Public Health*. 2017, doi: 10.3390/ijerph14040435.

[3] N. Serrano Fuentes, A. Rogers, and M. C. Portillo, "Social network influences and the adoption of obesity-related behaviours in adults: A critical interpretative synthesis review," *BMC Public Health*. 2019, doi: 10.1186/s12889-019-7467-9.

[4] Y. Qutteina, L. Hallez, N. Mennes, C. De Backer, and T. Smits, "What Do Adolescents See on Social Media? A Diary Study of Food Marketing Images on Social Media," *Front. Psychol.*, 2019, doi: 10.3389/fpsyg.2019.02637.

[5] V. Bettadapura, E. Thomaz, A. Parnami, G. D. Abowd, and I. Essa, "Leveraging context to support automated food recognition in restaurants," in *Proceedings - 2015 IEEE Winter Conference on Applications of Computer Vision, WACV 2015*, 2015, doi: 10.1109/WACV.2015.83.

[6] N. Martinel, G. L. Foresti, and C. Micheloni, "Wide-slice residual networks for food recognition," in *Proceedings - 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018*, 2018, doi: 10.1109/WACV.2018.00068.

[7] S. Mezgec and B. K. Seljak, "Nutrinet: A deep learning food and drink image

recognition system for dietary assessment," *Nutrients*, 2017, doi: 10.3390/nu9070657.

[8] H. Kagaya, K. Aizawa, and M. Ogawa, "Food detection and recognition using convolutional neural network," in *MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia*, 2014, doi: 10.1145/2647868.2654970.

[9] G. M. Farinella, D. Allegra, M. Moltisanti, F. Stanco, and S. Battiato, "Retrieval and classification of food images," *Comput. Biol. Med.*, 2016, doi: 10.1016/j.compbiomed.2016.07.006.

[10] R. U. Khan, X. Zhang, R. Kumar, and E. O. Aboagye, "Evaluating the performance of ResNet model based on image recognition," in *ACM International Conference Proceeding Series*, 2018, doi: 10.1145/3194452.3194461.

[11] Y. Mejova, H. Haddadi, A. Noulas, and I. Weber, "#FoodPorn: Obesity patterns in culinary interactions," in *ACM International Conference Proceeding Series*, 2015, doi: 10.1145/2750511.2750524.

[12] S. Abbar, Y. Mejova, and I. Weber, "You tweet what you eat: Studying food consumption through twitter," in *Conference on Human Factors in Computing Systems - Proceedings*, 2015, doi: 10.1145/2702123.2702153.

[13] J. Rich, H. Haddadi, and T. M. Hospedales, "Towards bottom-up analysis of social food," in *DH 2016 - Proceedings of the 2016 Digital Health Conference*, 2016, doi: 10.1145/2896338.2897734.

[14] V. G. V. Vydiswaran *et al.*, "'Bacon bacon bacon': Food-related tweets and sentiment in metro detroit," in *12th International AAAI Conference on Web and Social Media, ICWSM 2018*, 2018.

[15] D. Oliva, M. Abd Elaziz, and S. Hinojosa, “Image Processing,” in *Studies in Computational Intelligence*, 2019.

[16] S. Singh and S. V. A. V. Prasad, “Techniques and challenges of face recognition: A critical review,” in *Procedia Computer Science*, 2018, doi: 10.1016/j.procs.2018.10.427.

[17] L. Zhou, C. Zhang, F. Liu, Z. Qiu, and Y. He, “Application of Deep Learning in Food: A Review,” *Comprehensive Reviews in Food Science and Food Safety*. 2019, doi: 10.1111/1541-4337.12492.

[18] F. Altaf, S. M. S. Islam, N. Akhtar, and N. K. Janjua, “Going deep in medical image analysis: Concepts, methods, challenges, and future directions,” *IEEE Access*. 2019, doi: 10.1109/ACCESS.2019.2929365.

[19] V. Wiley and T. Lucas, “Computer Vision and Image Processing: A Paper Review,” *Int. J. Artif. Intell. Res.*, 2018, doi: 10.29099/ijair.v2i1.42.

[20] M. A. Subhi, S. H. Ali, and M. A. Mohammed, “Vision-Based Approaches for Automatic Food Recognition and Dietary Assessment: A Survey,” *IEEE Access*, 2019, doi: 10.1109/ACCESS.2019.2904519.

[21] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” 2010, doi: 10.1109/cvpr.2009.5206848.

[22] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int. J. Comput. Vis.*, 2015, doi: 10.1007/s11263-015-0816-y.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, 2017, doi: 10.1145/3065386.

[24] G. Nguyen *et al.*, “Machine Learning and Deep Learning frameworks and

libraries for large-scale data mining: a survey," *Artif. Intell. Rev.*, 2019, doi: 10.1007/s10462-018-09679-z.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, doi: 10.1109/CVPR.2016.90.

[26] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognit.*, 2018, doi: 10.1016/j.patcog.2017.10.013.

[27] 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.

[28] G. Ciocca, P. Napoletano, and R. Schettini, "Food Recognition: A New Dataset, Experiments, and Results," *IEEE J. Biomed. Heal. Informatics*, 2017, doi: 10.1109/JBHI.2016.2636441.

[29] 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, doi: 10.1007/978-3-319-70742-6_37.

[30] T. Ege and K. Yanai, "Multi-task learning of dish detection and calorie estimation," in *ACM International Conference Proceeding Series*, 2018, doi: 10.1145/3230519.3230594.

[31] L. Herranz, S. Jiang, and R. Xu, "Modeling Restaurant Context for Food Recognition," *IEEE Trans. Multimed.*, 2017, doi: 10.1109/TMM.2016.2614861.

[32] M. Bolanos and P. Radeva, "Simultaneous food localization and recognition," in

Proceedings - International Conference on Pattern Recognition, 2016, doi: 10.1109/ICPR.2016.7900117.

[33] S. Horiguchi, S. Amano, M. Ogawa, and K. Aizawa, "Personalized Classifier for Food Image Recognition," *IEEE Trans. Multimed.*, 2018, doi: 10.1109/TMM.2018.2814339.

[34] M. A. Fard, H. Haddadi, and A. T. Targhi, "Fruits and vegetables calorie counter using Convolutional neural networks," in *DH 2016 - Proceedings of the 2016 Digital Health Conference*, 2016, doi: 10.1145/2896338.2896355.

[35] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in Neural Information Processing Systems*, 2014.

[36] J. Sun, K. Radecka, and Z. Zilic, "Exploring better food detection via transfer learning," in *Proceedings of the 16th International Conference on Machine Vision Applications, MVA 2019*, 2019, doi: 10.23919/MVA.2019.8757886.

[37] N. S. Keskar and R. Socher, "Improving generalization performance by switching from ADAM to SGD," *arXiv*. 2017.