Development of Image-based Food Recognition and Nutritional Status Evaluation System

by

Babatunde Onadipe

Department of Bioresource Engineering

Macdonald Campus of McGill University

Montreal, QC, Canada

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

November 2020

Abstract

Food or dietary intake is very essential for human life. However, dietary-related illnesses such as obesity, chronic heart diseases and diabetes are among the leading causes of deaths in the world today. In the past decade, dietary intake monitoring has increasingly become an important step in achieving a healthy lifestyle.

Conventional pen-and-paper methods of gathering and evaluating dietary intake data often used to monitor dietary intake include 24-Hour Dietary Recall (24HDR), Food Record (FR), Food Frequency Questionnaires (FFQs), and several others. The 24HDR method relies on a user's cognitive ability to remember and note down the food he or she consumed over the last 24 hours. The Food Record method requires that the user manually keep a record of the food consumed over a specified period of time. The FFQs method is designed to evaluate dietary intake patterns by gathering data about the frequency with which certain food items were consumed over a specified duration.

These approaches are expensive, tedious, time-consuming and prone to errors due to human subjectivity and hence, necessitate the need to develop automatic dietary or diet quality assessment and monitoring systems that are reliable, accurate and place a lower burden on users.

Image-based Artificial Intelligence (AI) techniques such as computer vision and machine learning offer a cornucopia of useful approaches that can help mitigate diet monitoring and assessment challenges that otherwise defy conventional methods. The aim of this study was to develop a system that employed image-based computer vision and machine learning techniques for automatic dietary intake assessment and nutritional or diet quality evaluation of food consumed by a user. To this end, two approaches were implemented. In the first approach, traditional computer vision features extraction algorithms were employed to extract high-level visual representations such as texture, color, as well as feature descriptors such as Scale Invariant Feature Transform (SIFT) features and orientation of gradients from captured

images of single foods (such as Bagel, Avocado and Croissant). These features were then analyzed in different combinations to build food recognition models using five different classical machine learning algorithms (K-Nearest Neighbors, Logistic Regression, Random Forest, Support Vector Machine, and Linear Discriminant Analysis) as well as an ensemble of all the algorithms.

Results showed that the ensemble model performed better than the individual models and was able to recognize test food images with an overall accuracy of 90.32% and with precision and recall values ranging from $85 - 95\%$. Following the recognition of the test food, the Score of nutritional adequacy of individual foods (SAIN) and score of nutrients to be limited (LIM) were computed using the SAIN-LIM nutrient profiling model. The SAIN-LIM scores measured the diet quality of the food and were computed using the weight and nutrients composition of the food retrieved from the Food and Nutrient Database for Dietary Studies (FNDDS). Based on the results, there were occasional misclassifications due to similarities in the color and textural patterns in the foods, which were the main factors that inhibited the system's performance.

The second system experimented on captured images of Composite or Mixed food with more complex and overlapping visual features. The methodology relied on Deep Convolutional Neural Networks (CNNs) that extracted a large number of abstract features from food images to produce better representation of the food. Concretely, a Transfer Learning approach was implemented to develop the food recognition model (using $15 - categories\ of$ composite food images containing $16,035$ training set and 2000 validation set) which achieved an accuracy of 98.10% on the held-out 2000 test food images. Finally, the SAIN-LIM nutrient profiling model was used to estimate the nutritional adequacy and healthiness significance of the food.

Although future plans must include increasing the number of food categories considered as well further developing the system into a mobile application, the results of this study have shown that the system can be used to adequately monitor and assess quality of diets consumed by individuals and communities.

Index Terms: Dietary Assessment, Computer Vision, Machine Learning, Deep Learning, Neural Networks, Transfer learning, Nutrient Profiling, Nutritional Status Evaluation, Nutritional Benefit Estimation

Résumé

La nourriture ou l'apport diététique est très essentiel à la vie humaine. Cependant, les maladies liées à l'alimentation telles que l'obésité, les maladies cardiaques chroniques et le diabète sont parmi les principales causes de décès dans le monde aujourd'hui. Au cours de la dernière décennie, la surveillance de l'apport alimentaire est devenue de plus en plus une étape importante vers un mode de vie sain.

Les méthodes conventionnelles de collecte et d'évaluation des données par papier et stylo sur l'apport alimentaire souvent utilisées pour surveiller l'apport alimentaire, comprennent le rappel alimentaire de 24 heures (24HDR), le dossier alimentaire (FR) et les questionnaires de fréquence alimentaire (FFQ). La méthode 24HDR repose sur la capacité cognitive d'un utilisateur à se souvenir et à noter la nourriture qu'il a consommé au cours des dernières 24 heures. La méthode d'enregistrement des aliments exige que l'utilisateur conserve manuellement un enregistrement des aliments consommés pendant une période de temps spécifiée. La méthode FFQ est conçue pour évaluer les schémas d'apport alimentaire en collectant des données sur la fréquence à laquelle certains aliments ont été consommés pendant une durée spécifiée.

Ces approches sont coûteuses, fastidieuses, chronophages et sujettes à des erreurs dues à la subjectivité humaine et nécessitent par conséquent la nécessité de développer des systèmes automatiques d'évaluation et de surveillance de l'alimentation qui soient fiables, précis et imposent moins de charge aux utilisateurs.

Les approches d'intelligence artificielle (IA) basées sur l'image telles que la vision par ordinateur et l'apprentissage automatique offrent une corne d'abondance d'approches utiles qui peuvent aider à atténuer les problèmes de surveillance et d'évaluation du régime alimentaire qui, autrement, défient les méthodes conventionnelles. Le but de cette étude était de développer un système utilisant des techniques de vision par ordinateur et d'apprentissage basées sur l'image pour l'évaluation automatique de l'apport alimentaire et l'évaluation de la qualité nutritionnelle des aliments consommés par un utilisateur. À cette fin, deux approches ont été mises en œuvre. Dans la première approche, des algorithmes d'extraction de caractéristiques de vision par ordinateur traditionnels ont été utilisés pour extraire des représentations visuelles de haut niveau telles que la texture, la couleur, ainsi que des descripteurs de caractéristiques tels que les caractéristiques de transformation de fonction invariante à l'échelle (SIFT) et

l'orientation des dégradés à partir d'images capturées d'un seul aliments (comme le bagel, l'avocat et le croissant). Ces caractéristiques ont ensuite été analysées dans différentes combinaisons pour créer des modèles de reconnaissance des aliments à l'aide de cinq algorithmes d'apprentissage automatique classiques différents (k plus proches voisins, régression logistique, forêt aléatoire, machine à vecteurs de support, et analyse discriminante linéaire) ainsi qu'un ensemble de tous ces algorithmes.

Les résultats ont montré que le modèle d'ensemble fonctionnait le mieux et était capable de reconnaître les images des aliments testés avec une précision globale de 90,32% et avec des valeurs de précision et de rappel allant de 85 à 95%. Suite à la reconnaissance de l'aliment testé, le score d'adéquation nutritionnelle des aliments individuels (SAIN) et le score des nutriments à limiter (LIM) ont été calculés à l'aide du modèle de profil nutritionnel SAIN-LIM. Les scores SAIN-LIM mesuraient la qualité nutritionnelle de l'aliment et ont été calculés à l'aide du poids et de la composition en nutriments de l'aliment extraits de la base de données sur les aliments et les nutriments pour les études diététiques (FNDDS). Sur la base des résultats, il y a eu des erreurs de classification occasionnelles dues à des similitudes dans la couleur et latexture des aliments. Ce sont les principaux facteurs qui ont inhibé les performances du système.

La deuxième approche a expérimenté des images capturées d'aliments composites ou mixtes avec des caractéristiques visuelles plus complexes et se chevauchant. La méthodologie s'est appuyée sur des réseaux neuronaux convolutifs profonds qui ont extrait un grand nombre de caractéristiques abstraites des images alimentaires pour produire une meilleure représentation de la nourriture. Concrètement, une approche d'apprentissage par transfert a été mise en œuvre pour élaborer un modèle de reconnaissance des aliments (en utilisant une formation composite de 15 classes, 16035 formations, 2000 images des aliments composites de validation) qui a atteint une précision de 98,10% sur les 2000 images des aliments d'essai. Enfin, le modèle de profilage nutritionnel SAIN-LIM a été utilisé pour estimer l'adéquation nutritionnelle et l'importance pour la santé de l'aliment.

Bien que les plans futurs doivent inclure l'augmentation du nombre de catégories d'aliments envisagées ainsi que le développement du système en une application mobile, les résultats de cette étude ont montré que le système peut être utilisé pour surveiller et évaluer adéquatement la qualité des régimes alimentaires consommés par les individus et les communautés.

Termes d'index: évaluation diététique, vision par ordinateur, apprentissage automatique, apprentissage profond, réseaux de neurones, apprentissage par transfert, profilage des nutriments, évaluation de l'état nutritionnel, estimation des avantages nutritionnels.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Notes of profound gratitude to:

God Almighty, *for His grace, favor, mercy and steadfast love haven brought me thus far through the thick and thin of my master's program. I am particularly grateful for the good health and wellbeing that were necessary to complete this thesis*

Professor Michael Ngadi, *for the moral and financial support as well as the opportunity he gave to me to attain this great height under his tutelage, guidance and supervision. God bless you sir.*

Food and Bioprocess Team, McGill University, *particularly to Mary Haruna**, Dr. Ebenezer Kwofie***, Dr. Ogan Mba, Christopher Kucha, Subeg Mahal, Serge, and the rest of the team for their collaboration, timely feedback advice and their in-depth and all-round support. God bless you all.*

Ibukunoluwa Akinsola, *for always being there. I must stop at this juncture, place every other thing on hold, and with a standing ovation express and place on record my deepest gratitude to her as a dear friend. Nothing kept me going in the pursuit of happiness and success more than her kind resounding words of encouragement and her relentless and unwavering love even in the trying times. She was always keen to listen and know what I was doing and how I was proceeding, even though it is likely that my Machine Learning stories and analogies were sometimes boring and unending, and she barely knew what it was all about, she still always patiently listened nevertheless. Many thanks to you my love, you'll always have a special place in my heart...*

My Parents and Siblings, *for their supports morally, financially, and spiritually. Nothing gives me more joy than the constant realization of how much of a blessing from God you all are to me.*

Restoration House, Montreal family, *for the love, prayers, supports and for making Canada feel like home to me.*

My Mentors, friends and everyone I have had the opportunity to work with, *for the encouragements, support and for accepting nothing less than excellence from me. I am especially grateful to Prof amd Mrs. Akinbode Adedeji, Dr. Mobolaji Omobowale, Mr & Mrs. Muyiwa Kolayemi***, Olubukola Alli,, Similoluwa Ayoola, Lynn Oben, Abisoye Olaomi, Samuel Adekanmbi**, Anthony Iheonye**, Troy Gu**, Lanre Adetunji**, Archit Gupta*, Olusanjo Ogunbayo, Okenna Obi-Njoku, Pastor Adesuwa Irekiigbe***, Mr and Mrs. Adeyemi Adegbenjo, Samuel Ebosele*, Tosin Oludare, Teddy Akwii, Christopher Nzediegwu, Kanyinsola Okafor, Bukky Alimi, God bless you all.*

Montreal, *for being such a diverse and vibrant international city and for connecting me to many wonderful people.*

Format of Thesis

This thesis is manuscript-based, and it is in harmony with the thesis preparation guidelines of Graduate and Postdoctoral Studies (GPS) of McGill University. It consists of two research manuscripts.

The study in the first manuscript (Chapter 3) described how the datasets used in this research were acquired and the steps involved in preprocessing and preparing the them for food image recognition model development.

The study in the second manuscript (Chapter 4) made use of traditional Computer Vision features extraction algorithms as well as the Ensemble of classical Machine Learning Algorithms to develop an image-based food recognition system in order to mitigate the challenges associated with conventional pen-and-paper based methods of dietary intake assessment. The dataset used and the developed models made the system capable of recognizing and assessing nutrients of single foods from images.

The study in Chapters 5 employed a more sophisticated approach, Deep Convolutional Neural Networks and how it can aid diet quality assessment especially of composite or mixed foods.

Chapters 1, 2 and 6 of this thesis are the introduction, review of literature, and conclusion and recommendations for future work, respectively.

Contribution of Authors

The primary supervisor of this thesis is Prof Michael Ngadi. He conceptualized the research problem and approved the suggested strategies and objectives presented by the author to tackle the problem. The author was responsible for the data collection and analysis, experimentation, and implementation of the algorithms, performing the computational analyses and wrote the entire thesis under the supervision of Prof Michael Ngadi.

Prof. Michael Ngadi provided financial support and was also involved in reviewing and editing this thesis.

Dr. Ebenezer Kwofie provided insightful advice during experiments and was also involved in reviewing and editing the thesis.

List of Abbreviations

ANN	Artificial Neural Network
CNN	Convolutional Neural Network
CUDA	Compute Unified Device Architecture
DALYs	Disability Adjusted Life Years
DRV	Daily Recommended Value
FCNN	Fully Convolutional Neural Network
GPU	Graphics Processing Unit
HOG	Histogram of Oriented Gradient
I.I.D	Independent and Identically Distributed
KNN	K-Nearest Neighbor
LBP	Linear Binary Patterns
LIM	Score of Nutrients to be Limited
LR; lr	Logistic Regression; learning rate
LDA	Linear Discriminant Analysis
MLP	Multi-Layer Perceptron
NB	Naive Bayes
NP	Nutrient Profile
ResNet	Residual Neural Network
RF	Random Forest
RNN	Recurrent Neural Network
SAIN	Score of Nutritional Adequacy of individual foods
SGD	Stochastic Gradient Descent
SIFT	Scale Invariant Feature Transform
SURF	Speedup Robust Features
SVM	Support Vector Machine
Top-K%	Top-K classification accuracy
UI	Uncertainty Interval
VGG	Visual Geometry Group Network

1 Introduction

1.1. BACKGROUND

Technology advancement has brought about several improvements in the agricultural and food related sectors such as plant breeding, mechanization of agriculture, food processing and storage, and several others. This made food readily available and less expensive especially in developed countries and as such hoisted such countries from the dismal era of food scarcity to one of food excess and food wastage *(Lee et al., 2013)*. Conversely, nutrient deficiency diseases became increasingly scarce while complex chronic diseases related to excess food consumption, tobacco and alcohol use, poor diets, and physical inactivity such as obesity, heart disease, diabetes, cancer, are now the mainstream leading causes of death and disability in many parts of the world *(Lee et al., 2013).*

It has become a salient fact that an increasing amount of people is becoming overweight or obese. Research has shown that there has been a massive global increase in body mass index from 1980 to 2010 (0.4 mg/kg^2 per decade (95% Uncertainty Interval (UI): $0.2 - 0.6$)) for men and $0.5\ mg/kg^2$ per decade (95% UI: $0.3 - 0.7$) for women *(Finucane et al., 2011).* In 2016, the global prevalence of obesity was reported to have approximately tripled that of 1975. In that same year, *WHO (2018)* reported that there were more than 1.9 billion adults aged 18 and over that are overweight, out of which more than 650 million ones were obese. This drift clearly demonstrates that access and consumption of calories have drastically increased.

While food wastage does not, right-off-the-bat have a direct impact on human health, unhealthy dietary intake poses greater risks to human existence. According to the report on the study of Global Burden of Disease published by *Afshin et al. (2019)*, globally, in 2017, risks related to poor diet were responsible for

11 *million* (95% *UI*: 10– 12) deaths (22% [95% *UI*: 21– 24] of all deaths among adults) and 255 *million* (234– 274) Disability Adjusted Life Years (DALYs) (15% [14– 17] of all DALYs among adults). Cardiovascular disease was the leading cause of diet-related deaths (10 million [9–10] deaths) and DALYs (207 million [192– 222] *DALYs*), followed closely by cancer (913, 090 [743 345– 1 098 432] deaths and 20 *million* [17– 24] *DALYs*) and type 2 diabetes (338, 714 *deaths* [244 995– 447 003] and 24 *million* [16– 33] *DALYs*). Over 5 *million* (95% *UI* 5– 5) diet-related deaths (45% [43– 46] of total diet-related deaths) and 177 *million* (163– 192) diet-related DALYs (70% [68– 71] of total diet-related DALYs) occurred among adults below 70 years of age. The above facts reinforced the importance of the need to develop strategies and tools as well as leverage existing initiatives that will facilitate effective dietary monitoring and maintenance of good healthy living.

Global trends of dietary intake monitoring and diet quality assessment are still a major challenge not only because of their inherent complexity, but also because of the limitations of the tools available for collecting dietary data *(Allman-Farinelli et al., 2017; Cade, 2017)*. Traditional dietary intake data collection and assessment tools such as 24-Hour Dietary Recall (24HDR), Food Record (FR), Food frequency Questionnaire (FFQ) as well as the Minimum Diet for Women (MDD-W) are some of the widely used available tools. These tools are designed to capture details about an individual's food intake in the past 24 hours (24HDR), record of food consumed over a prescribed duration (FR), capture an individual's usual food consumption by querying the frequency at which the user consumed certain food items based on a predefined food list (FFQ) *(Gibson, 2005; Lee et al., 2013),* as well as capture the diversity of the food consumed by the user (MDD-W) in the previous day or night *(Hanley-Cook et al., 2020)*. These methods of collecting dietary data are mostly deployed using traditional pen and paper-based approach and hence, highly reliant on the literacy and cognitive capacity of the user for quality data. This however, makes the methods very tedious, expensive, prone to reporting errors, time-consuming, biased and place a lot of burden on clients, and dietary researchers *(Amoutzopoulos et al., 2018; Cade, 2017; Illner et al., 2012)*.

Recent advancements in current and emerging technology present broad and vast opportunities to improve the methods of collecting dietary data with the aim to reduce cost, improve the quality and efficiency of the data collected and methods deployed respectively, as well as reduce the burden on dietary researchers and clients *(Amoutzopoulos et al., 2018)*. New prominent technology-based dietary assessment tools include web-based software, wearable devices, and smart mobile device applications according to literature have proven beyond reasonable doubts to have the capacity to improve accuracy and reduce cost

and burden associated with dietary data collection and processing and hence, overcome the inherent limitations of traditional methods *(Amoutzopoulos et al., 2018; Eldridge et al., 2019)*. In the past five years, global smartphone penetration rate has increased significantly with over 3.2 billion smartphone users in 2019 and projected to surpass 3.8 billion in 2021 *(Arne, 2019)*. The increased accessibility of smart mobile device has led to the exponential increase in the demand and prevalence of mobile applications. Consequently, this has expanded the possibility of the use of smart mobile devices in several ways including gathering dietary intake data in order to determine nutritional adequacy and deficiency of food materials, perform dietary researcher -led or self-administered nutritional, fitness and health status analysis as well as to conduct small or large-scale epidemiological studies and nutritional surveys *(Klurfeld et al., 2018)*. It is however expected that, the mobile device-based tools need be accurate, reliable and efficient to ensure confidence when gathering dietary data. This can only be achieved through intensified interest and research to understand and engineer the "core" entity of the mobile device-based dietary intake assessment spectrum – analysis and interpretation of collected data. Food attributes such as type, nutrients composition, volume/ portion size, ingredients, and processing methods are major concerns for a healthy diet *(Zhou et al., 2019).* Several smart mobile device-based dietary assessment tools have been developed to guide users to capture image(s), video or record voice notes of all the food and drinks consumed during eating occasions. *(Assessment, 2018; Boushey et al., 2015; Hemalatha et al., 2020; Illner et al., 2012; Khanna et al., 2010).* To this end, there has been a lot of successes in the usage of food images due to the readily availability of sophisticated image processing, computer vision and machine learning algorithms as well as high-end computing resources for analyzing image data *(Hemalatha et al., 2020; Min et al., 2019).* The rapid pace in the development of mobile-based dietary assessment tools can be attributed to the convergence of innovation in several disciplines including data mining, computer vision, machine learning, software engineering, human eating habits, food science, and nutrition.

1.2. HYPOTHESIS

Current conventional dietary assessment methods provide error-prone qualitative data which are highly reliant on the cognitive capacity of the user as well as the dietary diversity or group of the food consumed by the user. In this work the following hypothesis are made,

- An image-based system can be developed to identify, analyze, and profile the nutrients composition of simple and complex food products in order to provide an accurate quantitative diet quality assessment.

- A deep convolutional neural network system can be developed to produce accurate classification of captured images of consumed foods especially for complex or mixed food products and this system can then be incorporated into a self-administered image based dietary assessment tool.

1.3. RESEARCH OBJECTIVES

The overall objective of this research is to develop a food recognition and diet quality assessment system that can aid dietary intake monitoring, and evaluation of nutritional adequacy and health benefits or level of recommendation of the food consumed by a user. This objective entails the following sub-objectives:

1. Develop an automatic recognition and nutrient profiling system for single foods such as Bagel, Avocado and Croissant based on computer vision image analysis techniques and machine learning.
2. Develop a deep learning assisted techniques to extend the food recognition and diet quality assessment system to detect and classify mixed or complex composite foods such as beef salad, rice & beans, and lasagna, consumed by a user, and also evaluate the nutritional adequacy and the contribution of the food to the health of the user.

2 Current, Emerging, and Future Technology Trends in Dietary Assessment

ABSTRACT

Traditional dietary assessments can be very tedious, time-consuming and expensive because they are mostly conducted using pen and paper-based methods. In addition to the complexity and the burden on dietary researchers, and clients, the results generated can also be prone to variations in interpretation, under-reporting as well as recording errors. Technology advancement has been a major component in every research knowledge space and traditional dietary assessment methods have benefitted immensely from it. New technology-based methods such as mobile device-based, web-based and sensor-based 24-hour dietary recall, food record and food frequency questionnaire have emerged and have become increasingly popular and more accurate alternatives for collecting dietary intake data. However, these methods have inherent limitations in delivering the desired degree of accuracy. The need to reduce burden on users, increase dietary intake data gathering accuracy, and improve health or quality of life of growing population, necessitates the adoption of current and emerging technology concepts such as data analysis, computer vision, and machine learning to achieve a higher degree of accuracy and reliability in dietary data gathering processes. The purpose of this study is to review current developments in technology-based dietary assessment as well as explore the potential impacts of emerging technology trends.

Index Terms: Dietary assessment, Food Frequency Questionnaire, 24-hour Dietary Recall, computer vision, image-based, machine vision, smart mobile device, Artificial Intelligence

2.1. INTRODUCTION

Dietary behaviors and intake have been linked to nutritional-related illnesses, deaths, and several chronic diseases in the world today. Measuring dietary intake is a method used to evaluate eating patterns and behaviors, actual or usual intake, as well as to assess the adequacy of a person's diet. However, the method comes with several challenges. The task is accompanied by flaws in data-gathering techniques, human subjective attitude, intense burden on client or the dietary researcher, daily variations in a user's dietary intake and limitations of nutrient composition tables and databases. Conventional methods of gathering behavioral and dietary intake data such as 24-Hour Dietary Recall (24HDR), Food Record (FR), Food Frequency Questionnaires (FFQs) and Diet History are often used for research purposes. The methods make use of self-reporting and interviewing mechanism as well as require the user to be literate, perform difficult cognitive tasks and the dietary researcher to perform loads of manual analysis on the dietary data using information provided in the nutrient intake database *(Burrows et al., 2019; Vila-Real et al., 2016)*.

However, in recent years, advancement in technology has bolstered its application into the medium and methods of delivering conventional nutrition assessment. It has also brought about the development of several innovative methods through current and emerging technological concepts such as data analytics, computer vision, machine learning, internet of things, robotics, augmented reality and several others. Numerous technology-based methods for dietary assessment have emerged. The conventional methods have also been equipped with a fair share of technological applications to leverage existing initiatives. At present, there are several computer-based 24HDR, FR, and FFQ tools that have been developed, tested, validated and designed to be deployed through web-based platforms and as standalone software with cross-platform operating system compatibility *(Timon et al., 2016)*. The tools are either self-administered, whereby a user completes dietary information in the absence of a dietary researcher, or interviewer-administered, where the researcher uses the tool to collect/analyse dietary data in the presence of a client being examined. The new and innovative technology-based dietary assessment methods mostly are available as software application on smart devices such as smartphones, tablets, etc., and have been widely used for self-monitoring of dietary intakes and for communicating dietary researchers or clinicians. Others methods are available on, web-based platforms, scanner or sensors-based tools *(Burrows et al., 2019; Cade, 2017)*, Although these technology-based methods have played significant roles, they have not yet fully replaced some traditional methods because of their inherent limitations which include inadequate rendering or coding of data, internet bandwidth, usage complexity, etc. however, they have become increasingly popular and more accurate alternatives for collecting dietary intake data when compared with the traditional pen and paper-based

methods of assessment *(Ambrosini et al., 2018; Vila-Real et al., 2016)*. The purpose of this paper is to review recent developments in technology-based methods of dietary assessment as well as explore the potential impacts of emerging technology trends.

2.2. General Overview of Nutritional Assessment Methods

Over time, the concept of nutritional assessment has relied on four methods to determine the nutritional status of a person. These methods are based on several anthropometry, biochemical, clinical and dietary observations used either alone or in combination *(Gibson, 2005)*. These methods have further been improved with increasing emphasis on reduction in the risk of chronic disease, health maintenance, and effective decision-making.

2.2.1. *Anthropometry*

This method measures the physical dimensions and gross composition of the body which include, height, weight, head circumference, as well as taking the measurement of skinfold thickness, body density, air-displacement plethysmography, magnetic resonance imaging, and bioelectrical impedance in order to estimate fat and lean tissue proportion in the body *(Lee et al., 2013)*. The results obtained are often validated using reference data obtained from the measurement of many subjects. The measurements depend largely on age (and sometimes sex and race) and nutrition composition especially in cases where chronic imbalances of protein and energy are to be determined. Anthropometry measurement is a quick, easy and reliable method that has the capacity to identify mild and severe levels of malnutrition and also have the distinguishing potential of providing information about previous nutritional history, which can rarely be provided by other assessment methods *(Gibson, 2005)*. Anthropometric measurements are of two types, growth and body composition, and have been widely used for the assessment of the nutritional status of both children and adults. These measuring methods are prone to systematic and random error types which are usually as a result of inadequate and improper training of personnel, difficulties in measurement of certain anthropometric characteristics such as skinfolds, and instrumental or technical errors. The errors can be minimized by proper training of personnel to use standardized, validated techniques and by frequent calibration of instruments, thus improving the accuracy and precision of the measurement. The most common anthropometric measurements used for estimating nutritional status as described in Table 2.1 are weight, height and body composition *(Gibson, 2005; Lee et al., 2013)*.

Table 2.1: Description, benefits and challenges of anthropometric measurements

Anthropometry	Method (s)	Description	Pros/ Cons
Weight	Ideal Body Weight (Hamwi equations)	**For males**: 106 pounds for 5 feet plus 6 pounds per inch above 5 feet. **For females**: 100 pounds for 5 feet plus 5 pounds per inch above 5 feet. *Add 10% for large frame. Subtract 10% for small frame	Pro: Simple and quick to use at relatively low cost Con: not suitable for large group studies
	Adjustable Body Weight (ABW)	Actual Body Weight - Ideal Body Weight × .25 + Ideal Body Weight. Used to monitor obesity, i.e., when their body mass index (BMI) exceeds 30	Pro: Effective and easy, inexpensive for minimal population Con: not suitable for large group studies
Height	Measurement from head to feet of a person	Measured without shoe with the back against the wall or measuring board, standing erect and looking ahead	Pro: Simple and quick to use, effective when with large population Con: Highly prone to variations
Body Composition	Skinfold thickness	Quick and simple method of measuring the amount of subcutaneous body fat.	Pro: Simple and quick, effective with large population Con: not suitable for single persons and groups
	Body Mass Index (BMI)	Measures and specifies an individual's weight status as simply being underweight, average weight or overweight based on height. BMI = Weight (kg)/Height (m^2)	Pro: Quick and easy to use Con: not precise for assessing body composition
	Waist circumference	Measures waist circumference for assessing abdominal fat	Pro: Very effective for measuring abdominal fat Con: poorly measures internal visceral fat
	Bioelectrical Impedance Analysis (BIA)	Measures how the body resist flow of through it and also estimates body fat from body water using appropriate equations	Pro: Simple and quick to use, relatively low cost Con: but usually effective with large population and poorly effective with specific, poor accuracy in individuals and groups
	Hydrostatic Weighing	Uses the comparison between an individual normal bodyweight (outside water) bodyweight while completely underwater to estimate his/ her body density and body composition	Pro: Accurate and easy to use Con: expensive and not convenient.
	Dual-Energy X-Ray Absorptiometry (DXA)	Uses passage of high- and a low-energy X-ray beam to determine mineral density and body composition.	Pro: Very accurate and safe Con: weight limitations, expensive, need for regular cross-standardization
	3D Body Scan	Maps the body and its component parts and then estimate the body fat through a corresponding app	Pro: Accurate, easy to use Con: not very convenient as it requires the user to be naked.

Sources: *(Ball et al., 2014; Gibson, 2005; Lee et al., 2013)*

2.2.2. *Biochemical methods*

Biochemical method as a method for assessing nutritional status of an individual measures nutrients or its metabolite in blood, feces or urine. It also measures other blood components and tissues such as albumin and serum protein (protein indicator) and hemoglobin (iron indicator) *(Gibson, 2005; Lee et al., 2013)*. Biochemical measurements often give result of the most recent nutrient intake or the effect produced by prolonged nutrient deficiency. The result obtained is usually helpful in determining the extent of nutritional deficiency *(Burrows et al., 2017).* While the method is easy to perform, inexpensive and with a good degree of accuracy, it comes with its own limitations. The test lacks specificity and its result is easily rendered unusable by problems such as pathological conditions, usage of prescribed medication and human and technical error. Due to the drawback associated with biochemical test, it is often a good practice to use it in combination with other nutritional assessment tools in order to improve validity and confidence level of the nutritional data *(Temple et al., 2003)*.

2.2.3. *Clinical methods*

Physical signs and symptoms of nutritional deficiency associated with family and medical histories such as delayed growth, pallor of skin, palm surface and hair colour are indicators often used to detect health problems and nutritional deficiencies *(Indumathi et al., 2017; Lee et al., 1996)*. Examining nutritional status using the clinical method does not require much expertise because of the obvious nature of the signs and symptoms. Clinical method is termed to be generic as it fails to specify nutritional deficiency components *(Lee et al., 2013)*. Some of the problems encountered during clinical assessment of nutritional status include relatively low general occurrence in developed countries except in high-risk groups, non-specificity of clinical signs, and high susceptibility to errors due to human subjectiveness *(Indumathi et al., 2017; Lee et al., 2013)*. Used in a cautious manner, in conjunction with other nutrition assessment methods such as dietary and biochemical methods, it may aid in accurate and more elaborate assessment of nutritional status of the individual.

2.2.4. *Dietary methods*

Dietary methods of assessment involve taking account of past or current nutrients intake from food by individuals or across large geographically dispersed populations in order to determine their nutritional status. Measuring dietary intake is the most commonly used method for determining an individual's nutritional status *(Lee et al., 1996)*. There are several methods of collecting dietary assessment data, which include face-to-face interviews, telephone interviews, by email or self-administrated. The deployed

method is usually a function of social and economic context, available resources, as well as the demographic characterization of the participants in question *(Vila-Real et al., 2016)*. Dietary assessment methods can be classified into two major groups: retrospective and prospective methods. The retrospective methods include the 24-Hour Dietary Recall (24HDR), the Food-Frequency Questionnaire (FFQ), and the Dietary History (DH), while prospective methods include Food Records (FR) *(Thompson et al., 1994)*. Their usage is often based on the purpose for which they are needed, the participant involved, as well as the validity and reliability of the tool *(Andreoli et al., 2011)*. The purpose of use may be to measure nutrients, foods or eating habits of a user, group of people or large population and if the purpose is to measure the nutrient or food intake, a diet record or 24HDR is most suitable. A diet record completed over several days has a higher potential of generating accurate intake data than a single 24HDR. On the other hand, if capturing eating habit especially in retrospect is the goal, the FFQ is more appropriate.

2.2.4.1 **24-HDR, Food-Frequency Questionnaire (FFQ) and Food Record (FR)**

24HDR and FFQ are the two commonly used dietary assessment tools and in some cases, they are often used in combination *(Vila-Real et al., 2016)*. 24HDR and FFQ, as the most used dietary methods, rely on the user's cognitive ability; hence, it is pertinent to first conduct a cognitive survey of the intended users before deploying the assessment tools *(Vila-Real et al., 2016; Wirfält, 1998)*. The 24HDR, in about 20 – 30 minutes, can collect data about foods and beverages consumed in the past 24 hours while the FFQ gives an account of a long-term assessment of how frequently certain food and beverage items were consumed. The Food Record (FR) is a very flexible and easy to use tool for gathering detailed information about food consumed with focus on short-term intake. Similar to 24HDR, it collects data such as food preparation methods, the kinds of ingredients used, amount consumed and the brand name or place of purchase *(Shim et al., 2014)*.

Collecting valid and reliable dietary data as well as analyses of dietary intakes especially for dietary assessment methods that rely on user's memory is a very sensitive task and needs to be administered by well-trained personnel with broad perception and knowledge-base *(Shim et al., 2014)*. In most cases, the interviewer needs to be a nutritionist, a dietary researcher or a nutrition student who had been previously trained by experts *(Vila-Real et al., 2016; Willett, 2012; Wirfält, 1998)*. Furthermore, a well-trained interviewer is equipped with skills to create a conducive atmosphere for the participant, as well as to ask relevant questions that improve the ability of the participants to remember their nutrient intake easily *(Willett, 2012)*.

In a study carried out by *Ma et al. (2009)*, to determine how many 24HDR data will be sufficient enough to analyze an individual's nutrient intake, their results showed that acquiring the data over three days is usually appropriate. They found out that if less than three, there would be significant differences in energy estimation and when more than three there was no significant improvement *(Ma et al., 2009; Vila-Real et al., 2016).* However, there are cases where there is difficulty in carrying out more than one recalls. For instance, when considering cases where participants' diets appear to be monotonous, or if the number of participants to be considered is large, a single recall would be enough to estimate a participant's dietary intake.

24HDR and FR have inherent advantages and disadvantages. While the FR does not require the users to recall information about food consumed, it places a lot of burden on them demanding them to self-record food consumed in real-time *(Shim et al., 2014; Vila-Real et al., 2016)*. On the other hand, the 24HDR also imposes a huge burden on the users because the data gathered relies on their cognitive ability as well as the skills of the interviewer. FFQ, though also has an absolute reliance on user's memory, when compared with 24HDR and FR, it gives a better idea of the user's typical nutrient intake because it has a larger retrospective period *(Vila-Real et al., 2016)*.

2.2.4.2 **Dietary History (DH)**

The dietary history tool was developed specifically by *Burke (1947)* in order to gather data about long-term dietary intake of a user *(Shim et al., 2014)*. This tool requires a very knowledgeable dietary researcher to engage the participant to complete a 24HDR, 3-day food diary, and checklist of foods usually consumed through an in-depth interview that can take nearly 90 minutes to complete *(Shim et al., 2014; Vila-Real et al., 2016)*.

Table 2.2 gives a summary of the above four dietary assessment methods with focus on their method and type of data collected, strengths and limitations.

Table 2.2: Summary of the Assessment Methods

	24-Hour Dietary Recall	Food Record	Food Frequency Questionnaire	Dietary History
Description	A subjective and open-ended dietary data collection tool that uses specially designed questionnaires and administered by well-trained interviewer	Subjective collection of dietary data collection using open-ended, self-administered questionnaires	Subjective measure of dietary intake using a predefined, self- or interviewer-administered survey tool	Subjective collection of dietary data collection using open- and closed-ended questionnaires administered by a trained interviewer
Type of data collected	Actual dietary intake data over the last 24 hours	Actual dietary intake information over a specific period	Estimate of usual intake over a relatively long period	Dietary intake over a relatively long period
Duration	Takes 20 – 30 minutes to complete; Often deployed over 1 – 3 days, but can be flexible depending on the study design	Often takes 3 – 7 days to complete, but can be flexible depending on the study design	Takes 30 – 60 minutes to complete; can deployed over a duration of 1 month to 1 year	Takes nearly 90 minutes to complete; deployed over a long time
Strengths	Collects extensive data about dietary intake; imposes relatively small burden on users; does not require literacy; easy-to-use and deploy; can be self-administered	Provides detailed dietary intake data in real time; no recall bias; no interviewer required	Simple and easy to deploy; cost-effective and timesaving; most suitable for large population study such as epidemiological studies	Gather data about long-term dietary intake of users
Limitations	Possibility of bias during recall; require trained interviewer; possible interviewer bias; expensive and time-consuming; multiple data collection days required to accurately measure actual dietary intake; tendency of altering user's eating habits	Places huge burden on users; require good level of literacy; required; there is possibility of under-reporting; very expensive and time-consuming; require multiple days for effective data collection; possibility of modifying user's diet due to repeated measures	Study- or research-specific; questionnaire is closed-ended; not very accurate due to recall bias; time-consuming; easily influenced by ethnicity, culture, user's economy etc.	Expensive to deploy; time-consuming; not suitable for population study such as epidemiological studies

Source: *(Shim et al., 2014; Vila-Real et al., 2016)*

2.3. TECHNOLOGY-BASED DIETARY ASSESSMENT

As reviewed in the previous section, many existing traditional methods of dietary assessment have inherent merits, associated errors and practical difficulties as encountered by dietary researchers and clients in their respective usage. Hence, this implies that there is a need for continuous innovative research and development in order to come up with systems with better efficiency, performance, and accuracy.

Advancement in technology has bolstered its application into many research fields and knowledge space including dietary assessment and this has brought about evolution and disruption of the existing initiatives. In recent years, there have been great improvements in the methods of collecting dietary intake data due to increased usage and access to the internet and the popularity of smart mobile devices *(Burrows et al., 2017)*. A lot of these have been and will continuously be made possible through current and emerging technological concepts such as computer vision, electronic sensor, data analytics, machine learning, cloud computing and internet of things.

In this section, we reviewed technological-based dietary assessment approaches, which leveraged existing traditional concepts (such as web-based 24HDR, FFQ, Food record, Diet history), as well as approaches which adopted new principles in their dietary data collection method. Several technology-based dietary assessment methods have been developed and continuously researched and been improved upon *(Illner et al., 2012)*. Publications on recent advanced technology applications in dietary assessment within 2008 – 2019 were examined and classified into three innovative categories namely: Scanner- /Sensor-based technologies, Web-/ Computer-based technologies, and. Mobile device-based technologies.

2.3.1. *Scanner- and sensor-based technologies*

Several scanner- and sensor-based tools for dietary recording exist today. The scanner can be used to scan barcodes of items purchased in stores. The wearable sensor is designed to record or monitor dietary intakes at set intervals which can then be used to analyze and estimate nutrient intake *(Forster et al., 2016; Stumbo, 2013)*. Microsoft SenseCam is a popular wearable sensor worn around the neck in order to record dietary intakes. Once turned on, it begins image capturing every 20 seconds when it detects movements, heat and light *(Boushey et al., 2017; Gemming et al., 2015)*. Evidence has shown that it is most effective when used in combination with other methods such as traditional 24HDR and food record methods of dietary assessment as a means of improving the accuracy by capturing incorrectly estimated or unreported information such as leftovers and unrecalled food *(Forster et al., 2016; Gemming et al., 2015; O'Loughlin et al., 2013)*. Some of the limitations faced with the usage of the SenseCam include poor image quality as

a result of inconsistency in the angle of the camera as well as low-light and hence, require further research *(Boushey et al., 2017)*. Another popular recently developed wearable sensor tool is eButton, a very small and lightweight device, worn on the chest, integrated with a camera to capture food images every 2 seconds *(Forster et al., 2016; Sun et al., 2014).* It has built-in food segmentation, volume estimation, modeling functionality and capability to automatically estimate the nutrient composition of food items *(Sun et al., 2014)*. Although e-Button has a promising future in estimating nutrient intake, it still needs a lot of research modifications that would limit the complexity and errors involved in automatically estimating nutrient intake from images of wide range of regularly and irregularly shaped foods items *(Forster et al., 2016)*. Automatic Dietary Monitoring (ADM) was developed by *Amft et al. (2009)* to estimate the weight of every food bites taken in order to reduce the burden associated with self-reporting. ADM was designed to use the body's sensors to monitor the weight of the user's bites of food through recording chewing cycles and food types. This is done through input data fed through a wrist-worn acceleration sensor, and a microphone in the external ear canal *(Amft et al., 2009)*. *Nishimura et al. (2008)* developed a wearable sensor system, which uses a microphone integrated into a Bluetooth headset to record chewing sounds in order to detect engineering properties of food such as crunchiness, and to also help users reduce their reliance on their cognitive abilities. Sensor systems generally face the limitation of being a burden to users because they find it rather uncomfortable to wear a sensor around their neck, chest or ear canal. They are however considered to be outdated or not widely used and have generally been replaced by more recent tools.

2.3.2. *Web-/ computer-based approaches*

As a substitute to the traditional pen-and-paper dietary assessment methods (24HDR, FFQ, Food record and diet history), researchers have developed several computer- and web-application-based versions which have proven to be less expensive, reduced user's burden, efficient, user-friendly and easy to self-administer, suitable for small and large-scale research purposes *(Subar et al., 2012; Timon et al., 2016)*. However, a greater percentage of the technology-assisted dietary assessment tools developed to date are centered around 24HDR and FFQ, and are web-based, due to global increases in internet access, usage and penetration *(Forster et al., 2016; Hutchesson et al., 2015)*.

Extensive research *(Cade, 2017; Forster et al., 2014; Hutchesson et al., 2015; Subar et al., 2012)* has shown that 24HDRs and FFQ best collect dietary data with high degree of accuracy, hence making them the most desired tool for monitoring the diets of populations, individuals, and more progressively for diet and diseases associated researches and control. The Web-based methods focus largely on 2HDR method and FFQ, and they

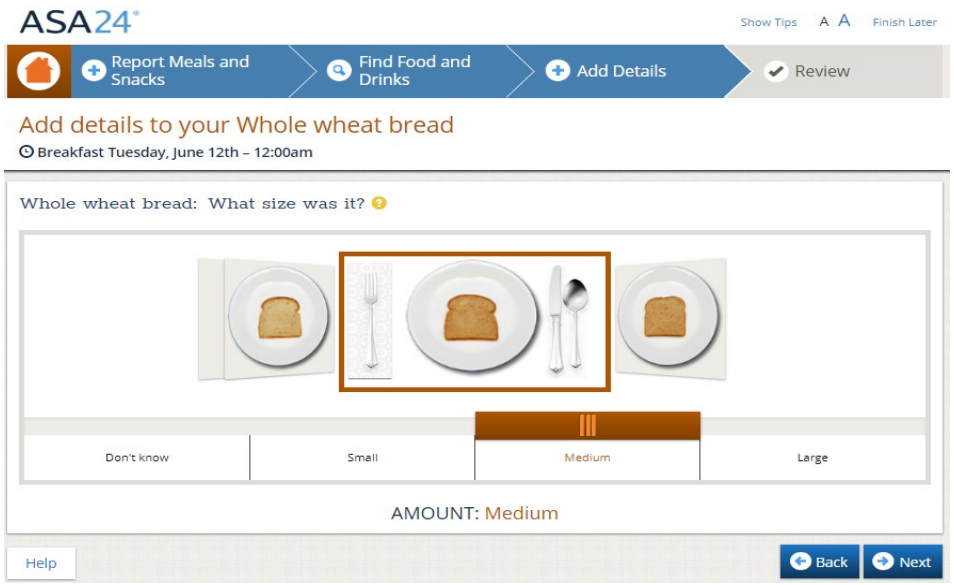
have been successful due to a number of factors including the ease of collecting data in a remote environment, the organised sequence of questioning, the use of digital portion size assessment graphical aids, automated analysis of data collected as well as the ability to generate quick dietary feedback *(Timon et al., 2016)*. The Web-based 24HDR tools have been used for dietary data collection for various population groups including young children, adolescents, and adults and it is modeled on the Automated Multiple Pass Method (AMPM) *(Subar et al., 2012; Timon et al., 2016)*. The tool guides and prompts the user to recall and record food and drink consumed in the last 24-Hour. The tool comprises of unique features such as, food and beverage lists, nutritional composition data, prompts and pictorial portion size assessment aids *(Subar et al., 2012)*. However, there are distinguishing characteristics incorporated in the tools by researchers based on target populations, demographics, and available resources.

The National Cancer Institute (NCI), in collaboration with *Subar et al. (2012)*, developed the Automated Self-Administered 24-hour dietary recall (ASA24). ASA24 is an open-source, web-based tool available to researchers, clinicians, and educators. It was developed to contain two web-based applications, the client and researcher websites. The client website is used by clients to complete 24HDRs using a dynamic and user-friendly web-interface. It consists of animated guides, audio and visual cues, eating occasion, time of consumption, eating-engagements (if the meal was consumed with a friend or if a computer or TV was used during the meal) selection mechanism. It is also possible for the clients to give account of the food consumed by browsing a food category or searching from a list of food and drink for keywords, as well as manually inputting desired food and drink at several stages of the recall session. ASA24 also consists of a researcher's (or clinicians, educators) website which is well-equipped with the capabilities to analyze recall data, generate reports, produce nutrient information data files as well as statistical schedules of complete, incomplete and upcoming recalls for clients *(CDC, 2006; Koegel et al., 2013; Subar et al., 2012; USDA-ARS, 2010)*.

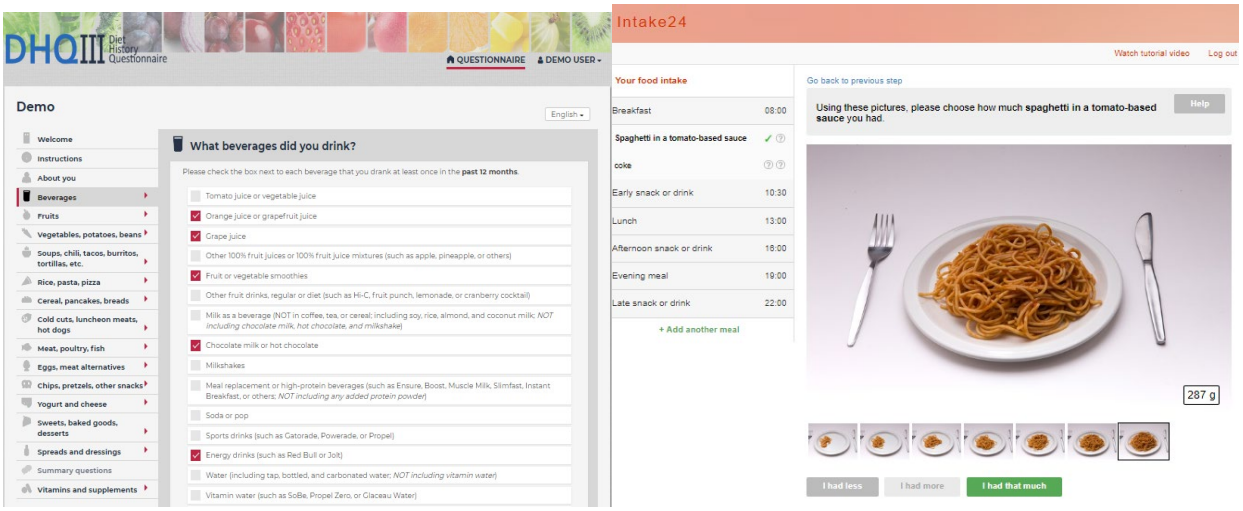
In a related development, some researchers *(Baranowski et al., 2014)*, came up with a self-administered multiple-pass-computerised 24HR using a tool they called the Food Intake Recording Software System (FIRSSt) *(Baranowski et al., 2002)*, and with its more recent version (FIRSSt4), renamed to ASA24-Kids but is no longer available for use.*(Baranowski et al., 2014; National Cancer Institute, 2018a)*. The tool was adapted from the ASA24, equipped with 10 000 + food images to quantify portion size estimation, and simplified specifically for children of age ten and above to comfortably self-administer a 24HDR. Table 2.3 and Figure 2.1 further describe some popular web-based dietary intake assessment tools such as online ASA24, INTAKE24 and DHQIII.

Table 2.3: Summary of some popular web-based dietary intake assessment tools

Platform	Description	Method	Features	Limitation	References
ASA24	Freely available web-based tool for epidemiologic, interventional, behavioral, or clinical research that allows for multiple automatically coded self-administered 24-hour recalls and food records	24HDR, Food Record	Enable researchers, clinicians, and teachers to register a study, set study parameters, manage study logistics, and obtain output files; guides participant using visual aids and other memory prompts to complete 24HDR or single or multiple day food records and to collect details about food form, preparation, portion size, and other additions; allows for ease of adding and modifying food, drink, and supplement choices at multiple points during the recall or record; available in 3 languages: English, Spanish and French; allows researchers to monitor study progress and obtain a variety of reports, including statistics for complete, incomplete, and upcoming recalls or records for users; contains about 65 nutrients and 37 food groups	Requires stable and reliable internet; require users to be literate. Depends on the cognitive ability of the user	*(National Cancer Institute, 2018b; Subar et al., 2012)*
INTAKE24	Intake24 is a self-completed computerized 24HDR system developed for use with users aged 11 years and over (including older adults). Its development was an iterative made up of a 4-stage user interaction and refinement.	24HDR	Online, engaging and intuitive 24-hour dietary recall system based on the multiple pass 24-hour recall; Database contains more than 2500 foods and over 2500 portion size images which have been extensively validated in a feeding study and against 4-day weighed diaries; 20-minute average completion time; Contains custom search algorithms that are highly tolerant to spelling mistakes; Equipped with extensive range of prompts for items commonly forgotten and consumed together; Can be access on desktop, laptop, tablet & mobile devices; Fully customizable visual style (using CSS) allowing addition of study logos; Simple integration of customized additional questions; Automatic coding to nutrient data	Requires stable and reliable internet; require users to be literate; Depends on user's memory	*(Simpson et al., 2017)*
DHQIII	Diet History Questionnaire (DHQ) is an open source food frequency questionnaire (FFQ) used by researchers, clinicians, or educators to assess food and dietary supplement intakes.	FFQ	DHQIII consists of 263 food and beverage line items and 26 dietary supplement questions; contain questions on cooking methods, frequency, portion size information about food and beverage consumed over a period of 1 month or year; takes about 30 to 60 minutes to complete	Requires stable and reliable internet; require users to be literate; time-consuming	*(National Cancer Institute, 2018c)*



(a)



(b) **(c)**

Figure 2.1: (a) Automated Self-Administered 24-Hour (ASA24) (National Cancer Institute, 2018b), (b) Diet History Questionnaire III (National Cancer Institute, 2018c), (c) Intake24 (Simpson et al., 2017)

2.3.3. *Mobile device technologies*

The use of mobile devices such as Personal Digital Assistants (PDAs), smartphones, and smart tablets, for dietary assessment has gained wide recognition in recent years. They have been very efficient in solving dietary data gathering accuracy limitations such as recall bias, data errors, delayed data entries and missing information, thus allowing for graphical, real-time and on-demand data assessment *(Seebregts et al., 2009).* Mobile devices are also known to provide hitch-free communication platform between a client and dietary researcher as well as a user-friendly and comfortable tool for children *(Oliver et al., 2013)*.

2.3.3.1 **Personal Digital Assistants (PDAs)**

The use of PDAs in dietary assessment has evolved in its concepts, architecture (software and hardware integration), and its applications over the last two decades since its first usage in the mid 1990s *(Comrie et al., 2009; Illner et al., 2012)*. The device has been well-equipped and upgraded from the older versions which were integrated with about 180 food items to the much recent versions with food items up to 4000. The system is also equipped with some portion-size tools that users can use to estimate their food intake *(Illner et al., 2012)*. PDA consists of specialized software with the integrated food items in colored images and different portion-sizes which help the user to effectively quantify the amounts of food intake. One variant of PDA is the Japanese 'Well-Navi' instrument, which guides the users to take digital photos of their foods and drinks before and after consumption and afterward send the pictures to a dietary researcher via internet for further analysis *(Fowles et al., 2008; Illner et al., 2012)*. Studies have shown that PDA is a reliable tool for gathering dietary data *(Acharya et al., 2011)*, however, progressive increase in the usage and popularity of smartphones have resulted in the disruption of PDA as a foremost tool for dietary assessment since it can render the same functionalities and much more *(Illner et al., 2012; Recio-Rodríguez et al., 2014)*.

2.3.3.2 **Smartphones, tablets and notebooks**

The evolution of mobile phones is by far one of the fastest iterative advancements that mankind has ever witnessed. The growing needs and expectations of users across a wide age-range facilitated the evolution of the smartphone and its functional features from being just a gadget for communication to becoming a personal companion and a tool for executing basic tasks such as financial transaction, navigation, multimedia and most importantly dietary assessment. The capacity and capabilities of smartphones have led to a great deal of research focused on the development of innovative solutions for the use of smartphones to quantify dietary intakes using mobile applications (mobile apps) and digital image-based

approaches *(Weiss et al., 2010).*

Mobile application monitoring approach

There are several diet apps available for use on smartphones, however, very limited studies have examined their use as tools for dietary assessment *(Jospe et al., 2015)*. The apps available for monitoring dietary intake can be classified as a standalone dietary record app or as one being integrated into other apps. The dietary record apps are basically the digital versions of their traditional counterparts but have several key features for both users and dietary researchers to record food consumption, and to review and analyze recorded intakes respectively. Some of these app features include well-integrated food composition databases, memory prompters, search functionality, and suggested food lists, saved favorite foods and recipes, and barcode scanners *(Allman-Farinelli et al., 2017)*. On the other hand, where the dietary record is combined with capacities to tracking multiple aspects of health. They have also integrated features such as physical activity monitors and body weight scales, as well as tracking of health parameters such as blood glucose, blood pressure, and sleep in for example, MyFitnessPal, MyFatSecret, LoseIt, and MyPlate. *(Allman-Farinelli et al., 2017; Forster et al., 2016; Gilhooly, 2017).* Although there is a growing availability of apps for tracking dietary intake, many are still limited and targeted to some specific group of people e.g. people with diabetes mellitus *(Allman-Farinelli et al., 2017; Rusin et al., 2013)*. Other limitations also include burden on user, database limitation, incomplete nutrient composition, difficulty in getting database for food groups and supplements, difficulty in exporting data from the mobile apps for research purposes, difficulty in estimating portion size, lack of flexibility on how portions are entered, incomplete nutrient profiles, presence of unnecessary or confusing information that are unhelpful for research purpose, etc. Hence, there is a need for more studies to further expand the usage and capabilities of mobile apps for dietary assessment. Table 2.4 below further describes some existing dietary assessment mobile applications.

Table 2.4: Summary of some popular dietary assessment mobile applications

Mobile App	Features	Limitation	Reference
MyFitnessPal	Barcode scanner; Recipe importer; Track all major ingredients; Built-in step tracker; Exercise entry; 20 languages supported; Personalized goals; Easy links with other apps and devices; 5 million+ food items	Adjusting serving sizes time-consuming; manually entering nutrition information is prone to error; multiple repetitions of food items in the database	*(Chen et al., 2017; MyFitnessPal, 2018)*
MyPlate	Barcode scanner; Personalized goals for major nutrients; Meal time reminders; Track water intake; Easy creation of custom foods and meals; Easily integrates with Apple's health app; Exercise entry; Social support features; Data exporting; Bilingual (English and Spanish); 2 million+ items	Diet logging is tedious; insufficient food varieties in database	*(MyPlate, 2018); https://www.everydayhealth.com/diet-nutrition/experts-what-are-the-flaws-of-myplate.aspx*
Lose It	Barcode scanner; Calorie tracking; Exercise Tracking; Community Access; Apple Health & Google Fit Sync; Wi-Fi Scale Support; Activity Tracker Support (Fitbit, etc.); Fitness App Support (Nike+ Run Club, etc.); Macronutrient Goal Setting & Tracking; Nutrition Insight Reporting; Data Analysis & Recommendations; Meal Planning; Meal Plan, Recipe & Workout Library; Water Tracking; Custom Themes; 7 million+ food items	Doesn't keep track of common vitamins and minerals; Popular food brands missing in the database	*(LoseIt, 2018); https://www.healthline.com/nutrition/10-best-weight-loss-apps#section1*
Healthwatch 360	Barcode scanner; Track 500+ symptoms and health conditions; Track 30+ nutrients; Nutrition score; Recipe builder; Reports and trends; Data collection and export with researcher portal	Insufficient food varieties in app database	*(HealthWatch 360, 2018)*
Easy Diet Diary	Barcode scanner; extensive and easy-to-browse database; easily track calorie and energy intake	Insufficient food in database; lacks integration with other apps	*(Easy Diet Diary, 2018)*
Carbs and Cals – Diabetes and Diet	Personalized goals for major nutrients; Contains easy-to-browse 3500+ photos of food & drink with up to 6 portions for each food; Personalized goals for major nutrients; Visual format makes diet monitoring easy; Customizable food database	Insufficient food in database	*(Carbs & Cals, 2018)*

.

Digital imaging approach

The advancement in smartphone technology, especially in the computing power and digital camera integration, has facilitated its usage as a tool for dietary assessment. According to a review by *Sharp et al. (2014)*, the focus of most of the studies since 2002 on smartphone usage for dietary assessment has been largely on digital imaging or image-based approaches. When dealing with the image-based approach, there are usually two major steps carried out on the image of the food captured namely: food recognition & classification and volume/ portion size estimation. The two steps involved in imaged-based approach can either be one that requires manual intervention of a trained image analyst or one that is automated/ technology-assisted i.e. does not require human intervention. The manual intervention requires the dietary researcher to carefully analyze the food images using pre-existing templates in order to obtain information such as class of food, volume of food, nutrients composition, etc. The automated approach relies on computer vision and machine learning assisted techniques to identify the class and volume of food in the images. The image-based approaches are known to reduced burden to users when compared to other forms of dietary intake recording methods, however, they are still liable to errors due to under-reporting especially if users capture poor quality images or forget to capture images of the food before consumption *(Forster et al., 2016).*

The idea behind the manual or human-assisted method is that, users take pictures of their food using their smartphone and send the pictures remotely to a secure central server application where a trained image analyst can access and compare the sent images with stored images of food items in order to analyze and classify the food, and further estimate the portion size/ volume of food intake and nutrient intake information *(Boushey et al., 2017; Martin et al., 2012).* Some drawbacks of this approach are that it would be very unproductive if the user lacks internet and the excessive time it will take if the study were to be for a large epidemiological study. In addition, it is also very expensive because it requires the services of a well-trained and knowledgeable dietary researcher.

The automated/ technology-assisted method uses built-in image classification, analysis and visualization as well as volume estimation tools to automatically estimate the portion size and volume of food consumed from images of food taken before and after consumption, with a fiducial marker placed near the food. This principle was adopted in the work of some researchers and yielded reliable accuracy *(Almaghrabi et al., 2012; Anthimopoulos et al., 2015; Zhu et al., 2010)*. They developed an interactive system that requires the user to capture an image (at $angle\ 45°\ or\ 60°$) of their food before and after meal through the

integrated camera on their smartphone using a specially developed mobile application for their smartphone. The image is then sent to a dedicated backend server were a special software analyzes the image using pre-configured computer vision and machine learning models to identify the type of food, estimate the portion size and estimate the energy and nutrient intake from corresponding nutrient databases *(Boushey et al., 2017).*

One crosscutting limitation of the automated/ technology-assisted method is its inability to differentiate between food with similar shape and color, for example, butter and margarine, brownies and chocolate *(Forster et al., 2016; Zhu et al., 2010)*. In order to mitigate errors associated with automated food item classification, there are several research considerations to incorporate features such as voice recognition which will allow the user give further voice description of the food during consumption. With the growing number of researches in the use of image-based methods in smartphones, there is no doubt that it would soon be widely and commercially available for use by the general public and dietary researchers.

2.4. DIGITAL IMAGE-BASED APPROACH FOR DIETARY ASSESSMENT

2.4.1. *Food recognition*

Developing a reliable Food recognition or classification system strongly depends on the availability of food image databases. The quality and size of food image dataset determine the performance and robustness of food recognition algorithms, therefore building a reliable database cannot be overemphasized. Several food image datasets have been developed and published in literature for different applications such as dietary assessment *(Boushey et al., 2017; Liu et al., 2016)*, food recognition *(Ciocca et al., 2017; Farinella et al., 2016; Subhi et al., 2019)*, food ingredients prediction *(Bolaños et al., 2017)*, food quantity or volume estimation *(Chen et al., 2012; Subhi et al., 2018),* calorie estimation *(Liang et al., 2018)* and several others.

2.4.2. *Features extraction and classification*

The efficiency and accuracy of a good food recognition system depend on the quality and relevance of the selected visual features of the food. An essential step in solving food recognition problem is to adequately represent the extracted visual information and store them as feature space or vector in an appropriate database. As evidenced in a number of research publications *(Choras, 2007; He et al., 2013; Nguyen et al., 2014; Subhi et al., 2019)*, Global features such as Color, shape, texture, etc. and local features such as Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), Local Binary pattern

(LBP) are crosscutting feature descriptors that are often extracted from food images and used to learn or train a food classification or recognition models. This is a machine learning technique that gives the model the capabilities to identify and predict the classes to which a set of unknown food belongs to, based on the food image features used as the input data in the learning phase. The feature descriptors used are largely due to their inherent properties such as invariance to geometric and photometric transformation, translation, rotation and scaling *(P. Pouladzadeh et al., 2014).* A study carried out by *Chen et al. (2012)* employed LBP and SIFT features individually on a food image dataset, their results showed that the accuracy of 53% was achieved with using SIFT features alone while using the LBP features only resulted in 46% accuracy. However, combining both features, along with additional Gabor filter and color features, increased the accuracy to 68%. A different study *(Beijbom et al., 2015),* using the same dataset, extracted SIFT, LBP HOG, MR8 filter, and color features, and using a SVM classifier, obtained an accuracy of 77.4%. In more recent research works, deep learning-based methods, a subset of machine learning, have been employed to learn and train more robust and effective neural networks. Convolutional Neural Network (CNN) is a considerably stable and broadly used deep learning-based algorithm, and it has been employed for food recognition by several researchers, however, it requires huge amounts of data to build *(Aguilar et al., 2017; Christodoulidis et al., 2015b; Hassannejad et al., 2016; Mezgec et al., 2017a; Yanai et al., 2015a).* CNN significantly improved classification accuracy on large food image datasets. An accuracy of 89% was achieved on Food101 *(Bossard et al., 2014)* and 83.15% on UECFood256 *(Martinel et al., 2018).*

2.4.3. *Food weight and nutrient estimation*

Before the corresponding nutrients information of food in a recognized food image can be estimated, the volume or weight has to be known or calculated *(Subhi et al., 2019).* The nutrient content can be estimated by computing the actual mass of the food based on the estimated volume and the density of the recognized food as well as calorie information obtained from nutritional databases such as the USDA Food Composition Database *(Liang et al., 2017; Subhi et al., 2019; USDA, 2010).* A method of estimating the amount of nutrients in a food consumed by an individual is to express the food in terms of the serving or portion size *(Abramovitch et al., 2012; Lu et al., 2018; Yang et al., 2018)*. The nutrient contents of nearly all popular, commonly consumed, and publicly sold food items for instance, the ones contained in food and nutritional databases such as United States Department of Agriculture Food and Nutrient Database for Dietary Studies (FNDDS), *(Montville et al., 2013; USDA, 2010)*, Canada Nutrient File *(Health Canada, 2015)*, and several others were computed based on a per 100g serving size or ready-to-eat form of the food. Food and Nutrient Database for

Dietary Studies (FNDDS), the database used to code food intake and calculate nutrients intake for the What-We-Eat-In-America (WWEIA), dietary component of the National Health and Nutrition Examination Survey (NHANES). The FNDDS also consist of portions and weights of commonly consumed foods and beverages as reported in the WWEIA database. FNDDS contains over 8000 food and beverages, 65 nutrient components for each of the food and over 30000 typical portion weights *(Montville et al., 2013; Rhodes et al., 2017)*. The weight and nutrient composition of foods obtained from the FNDDS can further be used for various applications such as being an integral part of an image-based nutritional status evaluation system.

Estimating food volume from 2-dimensional image(s) can be very challenging because food portion comes in different sizes and shapes, or served as single or mixed portion, thus, contribute to variations in extracted image features. In a research carried out by *Sun et al. (2008)*, a reference card is placed in the field of view of the camera while capturing the image of the food items. While the reference card helped to estimate the pose and scaling factor, the user still needs to manually select the food area upon capturing the food image. Once the food area is known, the volume can then be estimated based on multiple camera parameters which included the focal length, position, and other intrinsic parameters. *Sun et al. (2015)* presented a virtual reality (VR) approach which uses pre-built 3D models of food with known volumes. In its application, a user needs to superimpose the model onto the food items in the real scene and by scaling, translating, and rotating the models to fit the food items in the image, the food volumes can then be estimated, and nutrient content further computed. Advancements in computing power and artificial intelligence tools and algorithms have paved way for the application of deep learning approaches to estimate food volume from food images. Recently, a research team at Google developed a deep learning-based framework for estimating food volume. The idea behind their work relies on reconstructing 3D models of single RGB image of food based on inferred depth maps trained by convolutional neural networks (CNN). An advantage of the proposed work is the fact that it does not require the user to capture multiple images of food or to place a fiducial marker in the field of view of the camera as proposed in other research. Despite several achievements presented in literature to estimate nutrient content of food from single RGB image, there are still several issues such as occlusion, blurred inferred depth map, and several others *(Lo et al., 2018; Meyers et al., 2015).* This indicates that there is need for more work to be done.

2.5. NUTRIENT PROFILING FOR NUTRITIONAL STATUS EVALUATION

Nutrient profiling (NP) is a scientific approach for categorizing or ranking foods based on their nutritional composition for the purpose of preventing disease and promoting health *(WHO, 2010)*. Nutrient profiling has in the past decade leveraged technology to support dietary assessment. It has been used for several applications including health and nutrition claims, product labelling logos or symbols, information and education, provision

of food to public institutions, and as a self-administered dietary intake monitoring tool *(Koen et al., 2016; WHO, 2010)*. As a dietary intake monitoring tool, common applications of nutrient profiling include description of nutrient levels in foods (e.g. high fat, low fat, source of fiber, energy dense, nutrient poor, etc.), or description of the effects of nutrient consumption (e.g. healthy, healthier option, less healthy, etc.).

In designing a reliable nutrient profiling model, two major approaches exist which are based on carefully set parameters *(Drewnowski et al., 2008; Scarborough et al., 2007):*

1. Across-the-board approach; one which food are scored or classified using the same algorithm in an account to identify healthiness in foods.
2. Category-/ food-group specific approach; one which unique algorithms are adopted for different food groups in order to identify healthy diets within the group.

In the past few years, a number of nutrient profiling algorithms have been developed and validated following globally acceptable best practices and standards all of which take into cognizance certain crosscutting characteristics, namely:

1. The types and number of nutrients selected for usage in nutrient profiling algorithms in an effort to define food healthiness. These can be identified as *qualifying nutrients* (or positive nutrients, which are of good benefits to one's health), and *disqualifying nutrients* (or negative nutrients or nutrients to be limited).
2. The recommended values of the selected nutrients for nutrient profiling. This is often subject to national or international nutritional regulatory standards *(Masset, 2012; Tharrey et al., 2017).*
3. The optimal reference base upon which the nutrient content is computed. It is often expressed per 100 g or 100 kcal of food or per standard serving size which indicates the quantity of food considered for the computation *(AFSSA, 2008)*. Research have shown that references based on 100 kcal and on serving sizes better represent positive nutrients while the negative nutrients were preferably expressed as per 100 g *(AFSSA, 2008).*
4. The nutrient profiling algorithm to be used for effective combination of the recommended reference base and the nutrient content information *(Drewnowski et al., 2008).*
5. The threshold to distinguish healthy and unhealthy foods for effective nutrient profiling *(AFSSA, 2008; Masset, 2012)*

Table 2.5 below highlights a cross-section of some popular published nutrient profiling models which adopted a blend of the characteristics stated above.

Table 2.5: A cross-section of some published "across-the-board" nutrient profiling models (Masset, 2012; Tharrey et al., 2017)

Name	Algorithm	Reference Base	Qualifying Nutrients (+ve)	Disqualifying Nutrients (-ve)	References
Nutritious Food Index (NFI)[a]	$\mathbf{NFI} = \sum(\mathrm{w.\%DV_{positive}} + \mathrm{w.\%DV_{negative}})$	Serving	Fibre, Calcium, Iron, Zinc, Magnesium, Potassium, Phosphorus, Niacin, Folate and Vitamins A, C, B1 and B2.	Total fat, SFA, Cholesterol, Sodium	*(Gazibarich et al., 1998)*
Ratio of recommended to restricted food components (RRR)[b]	$\mathbf{RRR} = \frac{\sum\left(\mathrm{Nutrients_{Good}}/_{6}\right)}{\sum\left(\mathrm{Nutrients_{restricted}}/_{5}\right)}$	Serving	Protein, fibre, Calcium, Iron and Vitamins A and C.	Energy, SFA, total sugar, cholesterol, Sodium.	*(Scheidt et al., 2004)*
Calories for Nutrient (CFN)[d]	$\mathbf{CFN} = \frac{\mathrm{ED}}{\left(\sum_1^{13} \mathrm{\%DV_{100g}}\right)/_{13}}$	1000kcal	Protein, Calcium, Iron, Zinc, Magnesium, folate, niacin and Vitamins A, C, B1, B2, B6 and B12		*(Zelman et al., 2005)*
Food Quality Score 12, and 3 (FQS 1,2,3)	$\mathbf{FQS}_{1/2/3} = \frac{\sum_1^{n1/n2/n3} \mathrm{\%DV_{1/2/3}}/_{n1/n2/n3}}{\sum_1^5 \mathrm{\%DV}/_5}$	2000kcal	$\mathbf{n_1}$: fibre, Vitamins A, C, E, D, and B12, folate, Calcium, Magnesium, Iron, Potassium. $\mathbf{n_2}$: same, but category specific. $\mathbf{n_3}$: $\mathbf{n_1}$ + Protein, Phosphorous, Zinc, Copper, niacin, Pantothenic acid, Vitamins B1, B2, Kand B6, Manganese, Selenium.	Denominator: energy, SFA, cholesterol, sodium, and energy from fats	*(Kennedy et al., 2008)*
SAIN, LIM	$\mathbf{SAIN} = \frac{\sum_1^i \mathbf{ratio_i}}{\mathbf{i}} \times 100;\ \mathrm{ratio_i} = \left[\frac{\mathrm{nutrient_i}}{\mathrm{RV_i}}\right] \times \frac{100}{\mathrm{E}}$ $\mathbf{LIM} = \frac{\sum_1^3 \mathbf{ratio_j}}{\mathbf{3}} \times 100;\ \mathrm{ratio_j} = \left[\frac{\mathrm{nutrient_i}}{\mathrm{MRV_j}}\right] \times 100$	100kcal / 100g	5 nutrients from Protein, fibre, Calcium, Iron, ALA, MUFA, Vitamins C, D, and E	Sodium, SFA, added sugar	*(Darmon et al., 2009; Tharrey et al., 2017)*
Nutrient Rich Food (NRF9.3)	$\mathbf{NRF9.3} = \frac{\sum_1^9 \mathbf{\%DV}}{9} - \mathrm{LIM}$	100kcal or RACC	Protein, fibre, Calcium, Iron, Magnesium, Potassium and Vitamins A, C, E and B12.	SFA, added (or total) sugar, Sodium	*(Fulgoni et al., 2009)*

__Abbreviations__: $\mathbf{\%DV_i}$, percent of daily value (recommended intake) for a nutrient in the reference amount or in amount __i__; __SFA__, saturated fatty acids; __MUFA__, Mono-unsaturated fatty acids; __ALA__, α-linolenic acid. [a] __w__, weight given to individual nutrients. [b] Nutrient: nutrient content per serving. [c] __ED__, energy density. [d] __RACC__, reference amount customarily consumed, __LIM__, the score of nutrients to be limited, __SAIN__, score of nutritional adequacy of individual food (Masset, 2012; Tharrey et al., 2017)

The key components required to obtain the nutrient profile in food include the type and composition of nutrients in the food, the weight or portion size of the food, as well as the daily recommended values (DRV) of nutrients present is such food. While the nutrient composition, standard weight/ portion size and DRVs of a large variety of most common foods can be retrieved from publicly available food databases such as USDA Food and Nutrient Database for Dietary Studies (FNDDS) *(Montville et al., 2013)*, Canadian Nutrient File *(Health Canada, 2015)*, food weight or volume can also be computed manually *(Dehais et al., 2017; Yang et al., 2018).* However, estimating volume from food image is still a challenging task *(Min et al., 2019)*.

2.5.1.1 **The SAIN, LIM model**

The SAIN, LIM model is one of the most popular nutrient profiling models that defines food products in terms of their nutritional adequacy and healthiness based on an individual's recommended consumption. The SAIN, LIM nutrient scoring system which was developed by the French food safety agency, *AFSSA (2008)* has proven to be an highly effective tool among researchers. This system is based on two previously published indicators: the Nutrient Density Score (NDS), based on qualifying nutrients (i.e. positive nutrients), and the LIM score, based on disqualifying nutrients (i.e. the nutrient to be limited) *(Darmon et al., 2005; Maillot et al., 2007)*. A primary threshold value was assigned to the two scores which further defined a four "*nutrient profile classes"* for the model. The SAIN score, computed (using Equation 2.1) for 100 kcal of food, is an un-weighted arithmetic mean of the percentage adequacy for five qualifying nutrients (plus 1 optional nutrient) in the food composition tables and for which a daily recommended value (DRV) existed.

$$\textbf{\textit{SAIN Score}} = \left(\frac{\left(\frac{\left(\frac{Protein}{65} + \frac{Fibre}{25} + \frac{Vit\ C}{110} + \frac{Ca}{900} + \frac{Fe}{12.5} + \frac{Vit\ D}{5} - \min\ ratio \right)}{5} \times 100 \right)}{E} \right) \times 100 \qquad (2.1)$$

Where:

Protein = protein content in g/100g;

Fibre = fibre content in g/100 g;

Vit C = vitamin C content in mg/100g;

Ca = Calcium content in mg/100g;

Fe = Iron content in mg/100 g;

Vit D = vitamin D content in µg/100g

E = energy density in kcal/edible 100g;

Minimum ratio (min ratio) = the lowest of the 6 [nutrient/DRV] ratios.

The LIM score (computed using Equation 2.2) is the mean of the percentages by which a particular food exceeds the recommended nutritional value (Table 2.6) for each of the nutrients taken into account in the food, namely: Sodium, added sugars, and saturated fatty acids (SFA) and it is expressed per 100g of cooked or rehydrated food *(Darmon et al., 2009; Tharrey et al., 2017).*

$$\boldsymbol{LIM\ Score} = \left(\frac{\frac{Na}{3153} + \frac{SFA}{22} + \frac{Added\ Sugar}{50}}{3} \right) \times 100 \qquad (2.2)$$

Table 2.6: Daily Recommended values (DRVs) used to compute each food's nutrient density score (SAIN) and limited nutrient score (LIM), respectively (Darmon et al., 2009)

Score	Nutrient	DRV
Nutrient density score (SAIN)	Protein (g)	65
	Fibre (g)	25
	Vitamin C (mg)	110
	Calcium (mg)	900
	Iron (mg)	12.5
	Vitamin D (µg)	5
	Vitamin E (mg)	12
	α-linolenic acid (g)	1.8
	Mono-unsaturated fatty acids (g)	44.4
Limited nutrient score (LIM)	Saturated fatty acids (g)	22
	Added sugars (g) [a]	50
	Sodium (mg)[b]	3153

These values are based on French *(Martin, 2001)* and European *(Eurodiet Core Report, 2000)* nutritional recommendations. [a] If added sugars are not available, "free sugars", as defined by the WHO are used *(World Health Organization, 2003).* [b] Salt added at the table was not included.

2.5.1.2 Defined threshold values for SAIN and LIM score

SAIN score was computed based on 2000 $kcal$ reference daily energy intake. The optimum value for the SAIN is 100% for 2000 $kcal$, which corresponds to 5% for 100 kcal $(100/2000)$ food. Hence, a SAIN value ≥ 5 indicates a good nutrient density. On the other hand, the LIM was calculated for 100 g while the reference value used to compute the threshold value is based on food intake rather than on energy intake. The LIM score was computed based on 1340 g/d mean daily food intake. The maximal value for the LIM score is 100% for 1340 g $(100/1340)$, which is equivalent to 7.5% for 100 g food. Therefore,

a LIM value < 7.5 indicates a minimal amount of nutrients to be limited *(AFSSA, 2008; Darmon et al., 2009).*

Based on its SAIN and LIM values and the thresholds defined for each score, each food can be classified into 1 of 4 possible SAIN, LIM classifications as shown in Figure 2.2. Class 1 SAIN ≥ 5 and LIM < 7.5; class 2, SAIN < 5 and LIM < 7.5; class 3, SAIN ≥ 5 and LIM ≥ 7.5; and class 4, SAIN < 5 and LIM ≥ 7.5. Class 1 represent foods with the most recommendable nutrient profile (i.e. high nutrient density and low content of limited nutrients), whereas class 4 comprised of foods with the least recommendable nutrient profile (low nutrient density and high content of nutrients worth limiting). Foods from class 2 and class 3 are intermediate in terms of nutritional quality *(Darmon et al., 2009; Tharrey et al., 2017)*.

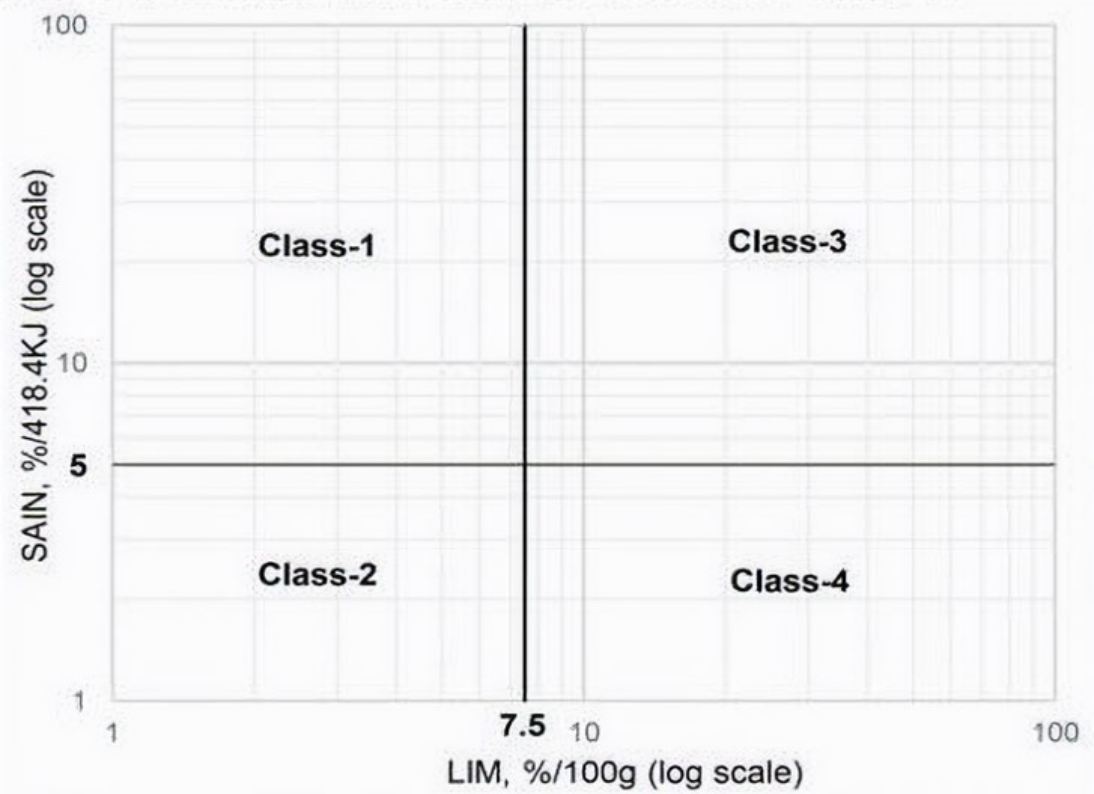


Figure 2.2: SAIN, LIM Food distribution chart

2.6. HOW EMERGING TECHNOLOGIES CAN IMPACT DIETARY ASSESSMENT

The age-old understanding and recognition of the limitations associated with self-reporting measures of dietary assessment has facilitated extensive application of current and emerging technology concepts such as computer vision and machine learning in order to continually improve accuracy, as well as reduce researcher and client burden. This growing success is making self-reporting easy and usable across all ages and genders, for instance, development of web app platforms and smart mobile device apps for food record and diet monitoring that can be easily used by adolescents and aged users for estimation of portion size, food recognition, nutrient profiling, etc. through analysis of food images. These emerging technology concepts can be classified as, (i) Techniques (for extracting desired features from food images) and (ii) Tools (driving potential).

2.6.1. *Techniques*

2.6.1.1 Artificial Intelligence (AI) techniques and Algorithms

In the past decade, Artificial Intelligence (AI) techniques such as computer vision, machine learning (including sub-fields such as deep learning), and natural language processing have been increasingly used for several applications such as face & voice recognition, spam email filtering, real-time navigation, etc. on popular web-platforms such as Facebook, Google, as well as in robotics, health care systems, drones and self-driving cars. Conventional computer vision techniques comprised of using feature extraction algorithms to extract desired visual features from captured images of food items. Global features such as color, shape, texture, etc. and local features such as Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), Local Binary pattern (LBP) are common feature descriptors that are often extracted from food images and used to learn or train a food classification or recognition models. This is a machine learning technique used to develop classification model which can be used to identify and predict the classes to which a set of unknown food belongs to, based on the food image features used as the input data to train the model. The feature descriptors used is largely due to their inherent properties such as invariance to geometric and photometric transformation, translation, rotation and scaling *(P. Pouladzadeh et al., 2014).* A study carried out by *Chen et al. (2012)* employed LBP and SIFT features individually on a food image dataset, their results showed that the accuracy of 53% was achieved with using SIFT features alone while using the LBP features only resulted in 46% accuracy. However, combining both features, along with additional Gabor filter and color features, increased the accuracy to 68%. A different study *(Beijbom et al., 2015),* using the same dataset, extracted SIFT, LBP HOG, MR8

filter, and color features, and using a SVM classifier, obtained an accuracy of 77.4%.

In the last decade, advancement in computing and processing power available in recent computers have largely contributed to the development of powerful deep learning *(Goodfellow et al., 2016; LeCun et al., 2015; Schmidhuber, 2015)* algorithms such as Deep Convolutional Neural Network (CNN), Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), and Autoencoders. These algorithms have found successful applications in food image classification and has proven to be very promising for image-based dietary assessment *(Ahn et al., 2019; Mandal et al., 2018; Mezgec et al., 2017b)*.

2.6.1.2 **Cloud computing**

Cloud computing generally refers to physical data centers which provides computing power, data storage, software, etc. to users all over the world, and are available over the internet (cloud). The rapid growth in cloud computing in recent decade can be attributed to the advancement in computing and storage equipment production. This has made it easy and inexpensive to host mobile and web applications in the cloud in place of on-premise hosting which are very expensive to set up, scale, and maintain. This has in-turn led to rapid increase in the development of several dietary data gathering and analysis applications which relies on the cloud for reliable storage, compute, and analysis of dietary data, thereby taking the burden off the mobile devices. There is no doubt that cloud computing will continue to provide long-lasting solutions to dietary assessment problems.

2.6.1.3 **Augmented Reality (AR) and Virtual Reality (VR)**

Augmented and virtual reality are two emerging technologies that are rapidly gaining popularity in dietary assessment. They have recently been deployed as techniques for estimating the portion size or volume of food in real time. While augmented reality tries to integrate digital information with the user's environment i.e. overlays new information on top of the environment in real time, virtual reality attempts to create a completely artificial environment. Though still in early research and development stages, several researchers have employed augmented and virtual reality for food volume estimation. In researches carried out by *Domhardt et al. (2015)* and *Stütz et al. (2014)*, mobile augmented reality applications which can assist users in self-reporting their nutrient intake were developed. *Yang et al. (2018)* also tried to solve problems associated with perception while capturing images of food for nutrient intake estimation. They developed a virtual reality food volume estimation method that would not require the user to place a

fiducial marker in the field of view while capturing images of food. The complexity and error margins showed that there are still much more improvements and research to be done.

2.6.2. *Resources and tools*

Resources such as image dataset and emerging computing tools and platforms such as Graphical Processing Unit (GPU), Quantum Computing, 5G, and Internet of Things (IoT) are driving potential or wheel upon which the emerging techniques will thrive in order to deliver desired overarching goals and objectives fast-advancing technological era.

2.6.2.1 **Image dataset**

Data has never been more available, and it has progressively become the '***plethoric oil***'. Particularly, ease of generating image data from different kind of devices have witnessed tremendous growth in recent years which is indicative of the leap into the Big Data and Internet of Things (IoT) era of technology where data and computing capabilities are readily available and easily accessible. This recent growth have spurred a great deal of researches toward the application of deep learning to solve problems in various fields such as robotics navigation *(Pierson et al., 2017),* self-driving cars *(Bojarski et al., 2016)*, agricultural production *(Kamilaris et al., 2018)*, remote sensing *(Cheng et al., 2016)*, medical imaging analysis *(Shen et al., 2017)*, and food detection and recognition *(Aguilar et al., 2017; Christodoulidis et al., 2015a).* Popular image database, ImageNet, with over 14 million images gave rise to the development of several well-known open source CNN architectures including, AlexNet *(Krizhevsky et al., 2012)*. VGGNet *(Simonyan et al., 2014)*, GoogLeNet *(Szegedy et al., 2015)*, and ResNet *(He et al., 2016)*. The impressive classification accuracies derived from these architectures subsequent to being trained on huge dataset motivated the concept of Transfer Learning *(Pan et al., 2009)* which deals with repurposing the knowledge (learned parameters) acquired from training a CNN on huge dataset to a different but related task with smaller dataset.

In food images recognition tasks, the major challenges are the large variations in food (both single and composite foods) due to shape, color, texture, volume, ingredients, composition, as well as image background noise *(Zhou et al., 2019)*. The successes and popularity of CNN architectures as well as the increasingly publicly available food image datasets contributed to the growing success in the application of deep learning for food recognition, detection, and to aid dietary intake assessment *(Min et al., 2019).* Furthermore, researchers can now (e.g. using deep learning API libraries (Tensorflow, Pytorch, Keras, etc.) implement transfer learning by fine-tuning and retraining open source pre-trained models (with

inherent ability to extract features such as texture, color, high-level abstract representations, etc.) on their own food image dataset. This approach has proven to significantly reduce training time and increase accuracy *(Sahoo et al., 2019; Zhou et al., 2019)*. Table 2.7 describe performance of popular deep learning architectures on publicly available dataset.

Table 2.7: Popular benchmark food image datasets

Dataset	Dataset Description and Source	Research work ref	DL Architecture	Top-1% & Top-5% accuracies
UECFood-256: (Yanai et al., 2015b)	• Popular foods in Japan and other countries • 256/ 31397 images each with a bounding box indicating the location of the food item in the image. • http://foodcam.mobi/dataset256.html	*Martinel et al. (2018)*	WISeR	83.15 | 95.45
Food-101: (Bossard et al., 2014)	• Popular food in USA • 101/ 101,000 images • https://data.vision.ee.ethz.ch/cvl/datasets_extra/food-101/	*Martinel et al. (2018)*	WISeR	90.27 | 98.71
UECFood-100: (Matsuda et al., 2012)	• Popular Japanese food • 100/9060 images each with a bounding box indicating the location of the food item in the image. • http://foodcam.mobi/dataset100.html	• *Martinel et al. (2018)* • *H. Hassannejad et al. (2016)*	• WISeR • Inception v3	• 89.58 | 99.23 • 81.50 | 97.30
Food-475: (G. Ciocca et al., 2017)	• Popular food in USA, China and Japan • 475/ 247,636 • http://www.ivl.disco.unimib.it/activities/food475db/	*Ciocca et al. (2018)*	ResNet-50	81.59 | 95.50
VIREO Food172: (Chen et al., 2016)	• Popular Chinese dishes • 172/ 110241 • http://vireo.cs.cityu.edu.hk/VireoFood172/	*Ciocca et al. (2018)*	ResNet-50	85.86 | 97.32

2.6.2.2 Graphical Processing Unit (GPU)

GPU is a programmable logic chip (processor) specialized for display functions. It is optimized exclusively for data computations and to renders images, animations and video for the computer's screen. The inherent ability of GPUs to perform massive parallel computations made it possible for it to effectively handle the popular dense linear algebra matrix-vector multiplication steps in neural networks *(Pierson et al., 2017).* The growing availability and application of GPUs makes the parallel processing faster, cheaper, and more efficient and it is responsible for the persistent breakthroughs in deep learning algorithms and models. With these, deep learning training that could take weeks, days and hours can be significantly reduced to take just few hours and minutes. Further considerations of deep learning applications in image-based dietary assessment may lead to faster and better accuracy in image-based food classification and recognition, as well as in determining the nutrient information of food from captured images.

2.6.2.3 **Quantum computing**

Another promising emerging technological advancement that would in the nearest future revolutionize our world and by extension improve technological applications in dietary assessment is the Quantum Computing. The concept of Quantum Computing involve developing computers including smart mobile devices that adopt quantum physics principles and capable of seamlessly performing computing operations in billion-fold realms and beyond *(Rouse et al., 2010)*. Quantum computers will have the capacity to analyze data in order to provide feedback much faster and efficiently than regular classical computers. This would in no doubt truncate the learning curve for artificial intelligence machines *(IBM Research Lab, 2018)*. Though still in conceptual and research stages, Quantum Computing is predicted to mainstream in 2023 and hence, might be deployed to solve complex dietary assessment problems, develop intuitive algorithms to accurately predict nutrient information and several other dietary challenges.

2.6.2.4 **5th Generation (5G) mobile networks technology**

The 5th generation mobile networks (5G) are the next generation of mobile internet connectivity, offering faster speeds, seamless, low-latency and more reliable connections on smartphones and other smart devices than ever before *(Qualcomm, 2018; Techradar, 2018)*. With its expected launch across the world by 2020, its application in dietary assessment would be very mind-blowing. Web-based dietary assessment methods such as FFQs and 24HDR would be significantly transformed through seamless and low-latency interactions with food images and nutrient information databases. Image-based dietary assessment methods deployed on smart mobile devices would experience flawlessness in data query from cloud databases regardless of the location around the world.

2.6.2.5 **Internet of Things (IoT)**

Dietary assessment is also expected to benefit extensively from the limitless possibilities of the emerging Internet of Things (IoT). With its global widespread expectation by 2025, thanks to cheap processors and wireless networks, IoT promises seamless communication and interconnectivity capabilities that add layers of digital intelligence to all passive and active things which may include computing devices, mechanical and digital machines, objects, animals or people that have been provided with unique identifiers and the ability to transfer data over a network without specifically requiring human-to-human or human-to-computer interaction. With IoT application in dietary assessment, for instance, complex nutrient information determination steps such as volume estimation might be discarded due to tendencies that food plates might be equipped with sensors that can accurately and automatically measure the volume

of food served on the plate.

These technological tools and techniques and the ones yet to be developed are expected to revolutionize our digital society and our daily activities. However, much more research work needs to be done to enable technology-aided dietary assessment methods to harness and integrate the capabilities of the emerging technologies. In addition, despite the promising benefits of these emerging technologies, there are still challenges mostly common in developing countries, and unserved or underserved parts of developed countries that need to be addressed, such as low internet penetration, poor literacy level, and data security implications.

2.7. CONCLUSION

Continuous improvement on the existing methods of dietary assessment remains an active area that requires constant research and development. Further validation of usability as well as enhancement of newly adopted methods using current, emerging and future technology concepts certainly requires more adept research focus, given researcher's primary aim to reduce burden on users and dietary researchers and also to improve nutrient data collection accuracy. Smart mobile devices have established a solid bedrock for personalized dietary assessment and have also created room for the inclusion of exciting new ideas such as personalized nutrition advice, dietary lifestyle monitoring, physical fitness guide, and several others. It is unequivocally clear that tech-based dietary assessment methods such as the web-based and smart mobile device-based methods are on a path to a very promising and productive future. However, as technology advancement further ripples across the existing technologies, recursive integration and iteration of innovative studies, coupled with well-designed experimentation are highly recommended in order to have seamless alignment between the existing and future technologies.

Preface to chapter 3

This chapter describes the image dataset used in this thesis in terms of how they were acquired, the steps involved in processing the data and how they were utilized. The dataset contains a total of $\sim 18,500$ food images out of which ~ 5500 are images of single foods belonging to 10 food categories ($Food10$ dataset). The remaining ~ 13000 images are of mixed or composite food product and are categorized into 15 types of food ($Food15$ dataset). The $food10$ dataset was used to traditional train machine learning algorithms while the $food15$ dataset was used to train a deep convolutional neural network.

3 Food Image Dataset Preparation for Diet Quality Assessment Research

ABSTRACT

This paper introduces a food image dataset suitable for development of image-based diet quality assessment systems. The dataset was derived partly from publicly available food images datasets as well as downloaded from the web. The dataset contains a total of $\sim 18,500$ food images out of which $\sim 5{,}500$ are images of single foods belonging to 10 food categories ($Food10$ dataset). The remaining $\sim 13{,}000$ images are of mixed or composite food product and are categorized into 15 types of food ($Food15$ dataset). In addition, the paper also described how images were acquired and the steps involved in preprocessing and preparing the dataset for food image recognition model development. The dataset has the potential of being used for several image-based dietary related researches. The dataset will further be made publicly available via the following cloud API: https://food-image-dataset.s3.amazonaws.com/Composite_food_dataset.zip

Index Terms: Food image dataset, image-based, computer vision, machine learning, diet quality assessment

3.1. INTRODUCTION

Image-based food recognition is a very beneficial approach to improve diet quality assessment systems. However, it can be a challenging task since food is intrinsically deformable and complex as it contains a lot of visual variabilities such as color, texture, shape, etc. In addition, developing a food image recognition model require a fairly large amount of image dataset with adequate visual representation of the food. The work presented in this paper described how images were acquired and the steps involved in preprocessing and preparing the dataset for food image recognition model development. The dataset contains a total of $\sim 18,500$ food images out of which $\sim 5{,}500$ are images of single foods belonging to 10 food categories ($Food10$ dataset). The remaining $\sim 13{,}000$ images are of mixed or composite food product and are categorized into 15 types of food ($Food15$ dataset). The $food10$ dataset was used for the development of food image recognition model as a component part of an image-based nutrient scoring system (*see Chapter* 4), by employing computer vision features extraction algorithms such as SIFT, HOG, LBP, Color Histogram to build a robust feature vector which can then be used to train machine learning algorithms including random forest, KNN, LDA, SVM, as well as the ensemble of the 5 machine learning algorithms. In a similar context, the $Food15$ dataset was used to train a deep convolutional neural network as a component part of diet quality assessment system (*see Chapter* 5). Table 6.1 (*Appendix C*) shows some popular benchmark publicly available dataset. The entire dataset is further be made available in the cloud for further image-based diet quality assessment researches (*see Appendix D)*.

3.2. MATERIALS AND METHODS

3.2.1. *Dataset preparation pipeline*

3.2.1.1 Image data acquisition

The dataset contains a total of $\sim 18,500$ colored food images out of which $\sim 5{,}500$ are images of single foods belonging to 10 food categories ($Food10$ dataset). The remaining $\sim 13{,}000$ images are of mixed or composite food product and are categorized into 15 types of food ($Food15$ dataset). There are three resources for the dataset namely: (i) web images (images downloaded from the internet), (ii) single-serving food images selected from the publicly available UECFOOD256 dataset *(Kawano et al., 2014)*, and (iii) randomly selected composite food images from the Food101 *(Bossard et al., 2014).* The $Food10$ dataset and its respective number of images per class are given as follows: C_1: Avocado (777), C_2: Bagel (324), C_3: Banana (790), C_4: Cheeseburger (601), C_5: Coconut (791), C_6: Cooked beans (322), C_7: Cooked rice (603), C_8: Croissant (257), C_9: Pizza (312), and C_{10}: Spaghetti (536). The $Food15$ dataset contained

13,030 composite food images with ~870 *images per class* as given below*: Lasagna, Steak_with_mashed_potatoes, Spaghetti_beef_tomato-sauce, Macaroni_and_cheese, Fried_rice, Fish_and_chips, Chicken_curry, Hot_dog, Rice_and_beans, Beef_salad, Pizza, Egusi_Soup, Pad_thai, Waffles_and_fruits, Cheeseburger.* Figure 3.1 shows examples of images in the Food15 dataset.

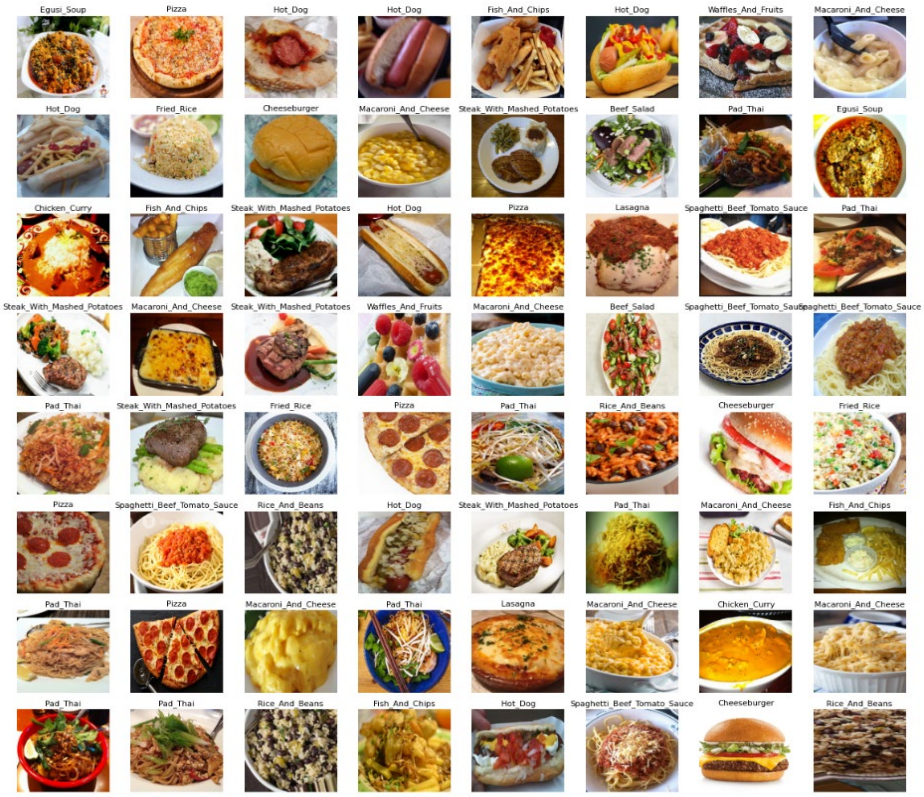


Figure 3.1: Examples of images in the Food15 dataset

3.2.1.2 **Data pre-processing**

Data cleaning, formatting and labelling

Since the dataset was acquired from different sources, it contained different inconsistencies such as background noise, wrong labels, unsupported image formats, duplicated images, variations in image size and dimensions, etc. In this step, images with irregular height or width (too large or too small) which usually are irrelevant images were first removed. The cleaning steps also involved manually going through all the images in the dataset in order to get rid of false positive images, that is images that were likely to confuse the machine/ deep learning algorithms or compromise the quality of the dataset. These false positive images included very blurry images, corrupt images, images with unsupported format, images with excessive background noise, etc.). The images also consisted of several kinds of image formats *(.jpg, .tiff, .png, .gif, .webp*, etc.) which were not all suitable for recognition model development or might slow

down the computation processes. In order to maintain uniformity, all images were converted to the supported ***.png*** image format due to the image quality preservation capacity of the image format. Particular attention was placed on processes involved in the image data labelling workflow. In order to automate the labelling process and to also make sure the labels do not have any noise in them, a script was written in python which helped to reduce the amount of time that would have been spent on manual labelling.

Image resizing

The dataset consisted of images with different resolutions, however for effective performance of the algorithms, it was important to establish a base image size that ensured that images with constant input dimensionality were used for model development. To ensure that the images had the same size and aspect ratio, the $Food10$ and the $Food15$ images were resized to a fixed resolution of 128×128 pixels and $300 \times 300\ pixels$ respectively.

Image denoising and smoothening

The database contained images with random variation of brightness or color information (image noise). The two types of image noise observed in the dataset were, *i) Gaussian noise* (resulting from poor illumination and/ or high temperature due to faulty camera sensor or circuitry during digital image acquisition or image transmission; *ii) Salt and pepper noise* (random speckles of dark (with 0 pixel value) and random bright (with 255 pixel value) colors all over an image resulting from sharp and sudden disturbances in the image signal due to faulty memory location or malfunctioning of camera's sensor cell during image acquisition). The gaussian noise and the salt and pepper noise were removed by filtering all the images in the dataset using gaussian filter and mean filter respectively. The gaussian filter also serve the purpose of smoothening the images.

Data splitting

The $Food10$ dataset was partitioned into training and held-out test set in the ratio $\mathbf{80}$: $\mathbf{20}$, respectively while the $Food15$ was partitioned into training, validation and test sets in the ratio $[70\%: 15\%: 15\%]$ respectively. The training set was further split into training and validation set. The training was used to fit the model in order to learn unique patterns peculiar to the dataset. The validation set was used for tuning the parameters of the model, evaluate the predictive quality of the trained model, as well as select the best performing model(s). The held-out test set was used to report the performance of the resulting machine learning model.

Feature scaling

The images in the $Food15$ dataset were scaled to be homogeneous i.e. scaled to take small values in order to improve the numerical operations and performance of the optimization function during the training of the neural network. For the feature scaling technique called the *min-max scaling* or *normalization* (computed using Equation 3.1), the features (image pixels) were casted to *float32* and then scaled from range of $[0 - 255]$ to have a range of [0 and 1]. This was done by subtracting the min value and dividing by the difference between the max and min values. The image normalization was carried out at the point of feeding the images to the network algorithm.

$$z = \frac{x - min\ (x)}{max(x) - min\ (x)} \tag{3.1}$$

Where $z\ is\ the$ normalized value for pixel value $\boldsymbol{x}$, $min(x)\ and\ \ max(x)$ are the minimum and maximum values in $\boldsymbol{x}$ given its range

3.3. DISCUSSION

$Food10$ dataset was used to develop recognition model by employing computer vision and classical machine learning algorithms. Global and local features such as color, shape, and texture were extracted from every image in the dataset using color histogram, histogram of oriented gradients (HOG), and local binary patterns (LBPs) feature descriptor algorithms, respectively. Each feature descriptor stored the extracted features as feature vectors. These vectors were then combined or concatenated together into a single robust feature vector which was then used to train the machine learning algorithms during the model development.

$Food15$ dataset was used to train a deep convolutional neural network using the transfer learning approach. Five different pre-trained networks namely: VGG16, VGG19, ResNet101V2, InceptionV3, and Xception (*see Appendix C*) were repurposed and trained on the 70% $Food15$training set using the *Fine-tuning* approach and their performance were evaluated on the 15% validation set.

3.4. CONCLUSION

A robust dataset containing single-serving food images and as well as mixed or complex food images ($Food15$) has been presented. The dataset is capable of promoting open research in training image recognition models for the development of image-based dietary quality assessment systems. The dataset can also be expanded to accommodate more food categories and hence used to create larger and more accurate food recognition model.

Preface to chapter 4

A comprehensive review of literature showed that image-based approaches to monitoring and assessing dietary intake area rapidly evolving area of study with a lot of promising potential.

This chapter explores hand-engineered Computer Vision features extraction algorithms to extract high-level visual representation of food such as texture, color, as well as feature descriptors such as Scale Invariant Feature Transform (SIFT) features and orientation of gradients. This chapter describes how these features, classified in different combinations, using machine learning algorithms as well as their ensemble can be assessed to build a food recognition system that can aid dietary intake monitoring and profiling of a user.

4 Development of an Image-based Food Recognition and Nutrient Profiling System

ABSTRACT

Diets have high impact on nutrition-related illnesses (e.g. diabetes, obesity, cancer) and incidence of mortality among different population groups around the world. Dietary and nutritional status assessment currently rely on monitoring procedures that are very prone to flaws from data-gathering practices, human subjective attitude, and daily variations in a user's dietary intake. The procedures are also expensive, tedious and time-consuming. In this paper, an approach of dietary intake assessment and nutritional status evaluation based on artificial intelligent (AI) tools such as computer vision and machine learning is presented. A combination of techniques such as preprocessing, features extraction, classification, and nutrient profiling were deployed to estimate the nutrient factor from users' meal. In the approach, unique features such as color, gradients of orientations and texture were extracted from food images database belonging to 10 classes. These features were used as training and validation dataset for the classification model. In its application, the user captures image of the food before consumption using a smart mobile device. The developed classification model identifies the class of the food in the image as well as its nutrient composition. The nutritional profiling score were then computed using the SAIN-LIM nutrient profiling model and data obtained from dietary composition database. The developed system can serve as a useful tool that enables users to self-administer and evaluate their single or multi-day food records as well as enables dietary researchers to track and analyze nutrition goals of clients and population groups. Results showed that assessing and estimating the SAIN-LIM scores of user's meal from its image will improve and facilitate proper control of dietary intake and overall maintenance of healthy diet.

Index Terms: Dietary assessment, computer vision, machine learning, nutrient profiling, nutritional status evaluation

4.1. INTRODUCTION

Inappropriate dietary intake has come into focus in the past decade as part of the leading causes of nutrition-related illness and death in the world. For instance, it has been associated as a major contributor to the development of chronic heart diseases, diabetes, and other vascular syndromes. The entry steps in addressing these challenges are to measure, analyze and monitor dietary intakes in order to determine an individual's nutritional status. The procedures often employ either traditional methods of gathering dietary intake data such as 24-Hour Dietary Recall (24HDR), Food Record (FR), Food Frequency Questionnaires (FFQs) or a recent approach which involves examination of images of consumed foods sent to a dietary researcher by a client in order for the researcher to carefully analyze the food images using pre-existing templates to obtain information such as class of food, volume of food, nutrients composition, etc., *(Ainaa et al., 2018; Ambrosini et al., 2018; Lee et al., 2013).* There are several challenges associated with these methods. These include but are not limited to flaws in data-gathering techniques, intense burden on client or the dietary researcher, and daily variations in client's dietary intake. The variations may be attributed to the fact that they utilize or rely on tedious, subjective, error-prone and time-consuming self-reporting, interviewing, and data-gathering mechanisms that require the user to be literate, perform difficult cognitive tasks and the dietary researcher to perform loads of manual data entry and analysis on the dietary data using the information provided in a nutrient intake database *(Vila-Real et al., 2016)*. The need to improve the accuracy of dietary intake data and data-gathering processes as well as mitigate the evolving nutrition-related chronic diseases has continued to gain significant attention in the nutrition and health research community. For these reasons, there has been a progressive realization and incessant demand for the development of a sophisticated systems to automatically carry out tasks associated with nutrient intake and nutritional status determination, such as food recognition, food type classification, volume estimation, and, nutrient profiling *(Subhi et al., 2019)*. These have been the principal theme of several research efforts in the nutrition and health research knowledge space.

So far, broad spectrum of innovative and advanced technology-assisted solutions have been developed, validated and presented in literature. These solutions have been seen to leverage the recent increasing computational efficiency in mobile devices, advancement in computer vision and

machine learning algorithms as well as breakthroughs in cloud computing and cloud-based mobile technology.

The computer vision frontier is still in its developmental stage. There are still several challenges such as occlusion, variations in photometric properties and viewpoint especially when applied to food and nutrition assessment. The problem has further been exacerbated by the complexity and variations in the geometrical structure and appearance (color, shape, and texture) as well as intrinsic deformability of food. This makes recognition or classification of food images a difficult tasks for several classification models, and hence a subject of concern for computer vision researchers. A popular recognition or classification technique that has in the past decade gained popularity and veritable applications due to advancement in computational efficiencies of computer hardware is the Ensemble Method. Single models derived from training machine learning algorithms (such as logistic regression, decision tree, etc.) on some data often do not yield good results, especially when dealing with a dataset with dynamic features. Ensemble method takes several single models as inputs and combines them into one single predictive model in order to achieve more reliable and accurate results.

The main objective of this study was to develop an automatic recognition and nutrient profiling system for single foods such as bagel, avocado and croissant based on computer vision image analysis techniques and machine learning. The study also covered an objective to develop a food image database comprising of single foods which can be further used in subsequent supervised learning studies.

4.2. MATERIALS AND METHODS

4.2.1. *Dataset protocol*

4.2.1.1 **Image data acquisition for the food recognition system**

The food image dataset used in this study comprised of 5313 color images of single-serving food belonging to 10 categories or classes of food. There are two resources for the dataset namely: web images (images downloaded from the internet) and a cross-section of the publicly available UECFOOD256 dataset *(Kawano et al., 2014)*. The 10 food classes and their respective number of images per class are given as follows: C_1: Avocado (777), C_2: Bagel (324), C_3: Banana (790), C_4: Cheeseburger (601), C_5: Coconut (791), C_6: Cooked beans (322), C_7: Cooked rice (603), C_8: Croissant (257), C_9: Pizza (312), and C_{10}: Spaghetti (536). The images obtained were from different internet sources and contained a lot of visual differences which potentially introduced photometric and geometric variabilities (such as multiple scaling, rotation, difference in illumination, difference in viewpoint, image blurring, compression and background noise) in the dataset.

4.2.1.2 **Data preprocessing**

Pre-processing referred to the set of transformations and feature engineering practices applied to the dataset in order to enhance the features of the images before feeding them to the machine learning algorithm. The following preprocessing steps were carried out on the dataset.

Data cleaning, formatting, and labelling

Since the dataset was acquired from different sources, it contained different inconsistencies such as background noise, wrong labels, unsupported image formats, duplicated images, variations in image size and dimensions, etc. In this step, images with irregular height or width (too large or too small) which usually are irrelevant images were first removed. The cleaning steps also involved manually going through all the images in the dataset in order to get rid of false positive images, that is images that were likely to confuse the machine learning algorithms or compromise the quality of the dataset. These false positive images included very blurry images, corrupt images, images with unsupported format, images with excessive background noise, etc.). The images also consisted of several kinds of image formats *(.jpg, .tiff, .png, .gif, .webp*, etc.) which were not all suitable for computer vision algorithms or might slow down the computation processes of the

computer vision and machine learning algorithms. In order to maintain uniformity, all images were converted to the supported ***.png*** image format due to the image quality preservation capacity of the image format. Another foreseen challenge that might hinder the performance of the machine learning model was the problem of incorrect image labelling. Particular attention was placed on processes involved in the image data labelling workflow. In order to automate the labelling process and to also make sure the labels do not have any noise in them, a script was written in python which helped to reduce the amount of time that would have been spent on manual labelling.

Cropping, resizing and patch selection

The dataset consisted of images with different resolutions, however for effective performance of the algorithms, it was important to establish a base image size that ensured that images with constant input dimensionality were fed to the machine learning algorithm. Firstly, the important portion of the image often called ***region of interest (ROI)*** was cropped out while discarding the excess pixels and unwanted background noise. This in turn will contribute to rapid processing of the image data by the computer vision algorithms. To ensure that the images had the same size and aspect ratio, the images were further resized to a fixed resolution of 128×128 pixels. Concretely, given a sample rectangular image, the ROI was first cropped and then resized such that the shorter side was of length 128 pixel, resulting into a central image patch with resolution of 128×128 pixels.

Image denoising and smoothening

The database contained images with random variation of brightness or color information (image noise). The two types of image noise observed in the dataset were, *i) Gaussian noise* (resulting from poor illumination and/ or high temperature due to faulty camera sensor or circuitry during digital image acquisition or image transmission; *ii) Salt and pepper noise* (random speckles of dark (with 0 pixel value) and random bright (with 255 pixel value) colors all over an image resulting from sharp and sudden disturbances in the image signal due to faulty memory location or malfunctioning of camera's sensor cell during image acquisition). The gaussian noise and the salt and pepper noise were removed by filtering all the images in the dataset using gaussian filter and mean filter respectively. The gaussian filter also serve the purpose of smoothening the images.

Data splitting

Dataset splitting was necessary in order to eliminate bias while training the machine learning algorithms. The classification accuracy of a machine learning algorithm depends largely on the size of

the training dataset, that is, larger numbers of training dataset would usually enable the classifiers to produce better generalization. The dataset was partitioned into training and held-out test set in the ratio **80**: **20,** respectively. The training set was further split into training and validation set. The training was used to fit the model in order to learn unique patterns peculiar to the dataset. The validation set was used for tuning the parameters of the model, evaluate the predictive quality of the trained model, as well as select the best performing model(s). The held-out test set was used to report the performance of the resulting machine learning model. The predictive performance of the model was evaluated by comparing predictions on the test set with true values (known as ground truth) using the accuracy, precision and recall performance metrics.

Some of the images in the database used in this study are shown in Figure 4.1.

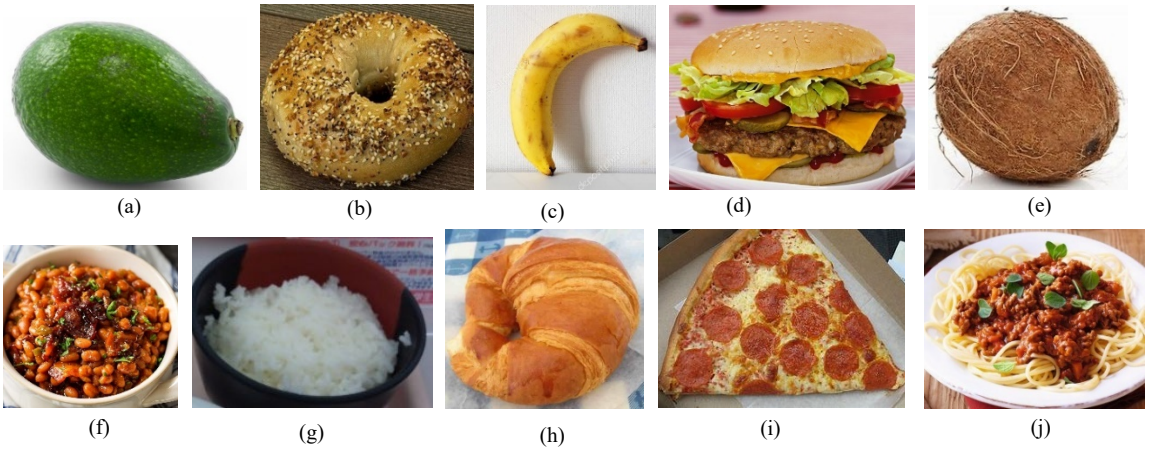


Figure 4.1: Sample of the images in the dataset. The dataset consists of 5313 food images organized into 10 food classes: (a) Avocado, (b) Bagel, (c) Banana, (d) Cheeseburger, (e) Coconut, (f) Cooked beans, (g) Cooked rice, (h) Croissant, (i) Pizza, and (j) Spaghetti. All images in the dataset are color RGB images.

4.2.2. *System architecture and model development*

The development of the machine learning model to be used for classification in the food recognition system followed a series of steps as to make up the proposed system as shown in Figure 4.2.

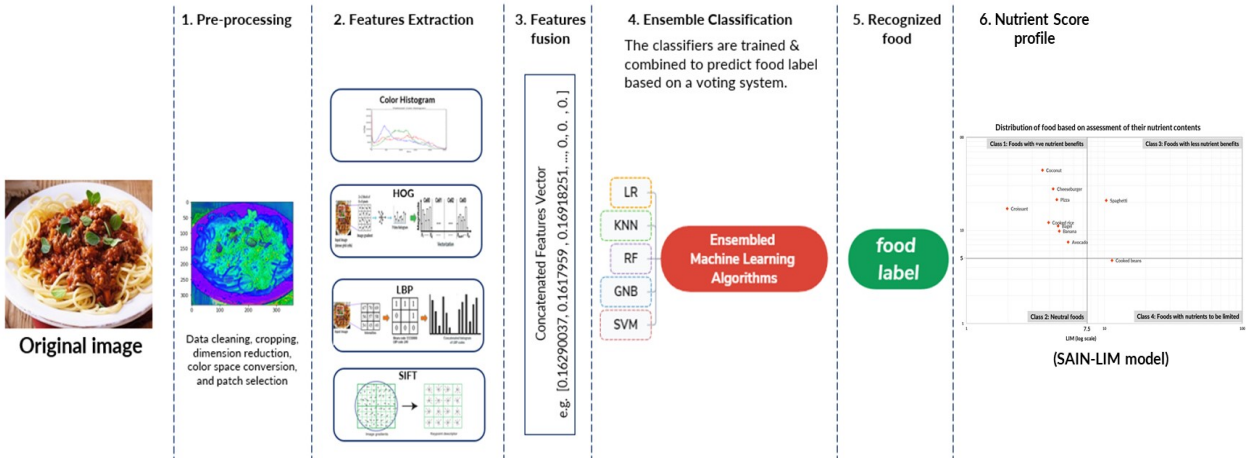


Figure 4.2: Overview of the Proposed System Architecture

4.2.2.1 **Features extraction**

The feature extraction procedure was responsible for the extraction of useful visual features that was used for image perspective understanding, interpretation, and classification. Global and local features such as color, shape, and texture were extracted from every image in the dataset using color histogram, histogram of oriented gradients (HOG), and local binary patterns (LBPs) feature descriptor algorithms, respectively. Each feature descriptor stored the extracted features as feature vectors. These vectors were then combined or concatenated together into a single robust feature vector which was then used for training the machine learning algorithms. The three feature extraction algorithms are further described below.

Color histogram

Color features was extracted from the images by first finding the color space which best described the color array. Hue, Saturation, and Value (HSV) color features was used to describe colors in terms of the degree of color, vibrancy and brightness of the image which conforms seamlessly with how the human eye tend to perceive colors *(Singha et al., 2011)*. In order to extract the color histogram in the HSV color space, the images were first transformed from the default RGB color space into HSV color space *(Su et al., 2011)* using Equation 4.1. Each of the images were then divided into $4 \times 4\, blocks$ and a $96 - bin$ HSV color histogram was extracted for each image. Each HSV channel was then quantized into $32 - bin$ resulting into an $8 \times 8 \times 8\, bins$. In total, $1536 - dimension$ $(32 \times (4 \times 4)\, blocks \times 3\, channels)$ color feature vector was extracted from each image.

$$H = \cos^{-1}\left\{\frac{\frac{1}{2}\left[(R-G) + (R+B)\right]}{\sqrt{(R-G)^2 + (R-B)(G-B)}}\right\}; \; S = 1 - \frac{3[\min(R,G,B)]}{V}; \quad (4.1)$$

$$V = \frac{1}{3}(R + G + B)$$

Where,

H = Hue; represent the true color, e.g., red, yellow, green, cyan, blue, magenta, etc. and it is measured in degrees from 0 and 360 with 0 being red.

S = Saturation; represent the amount of true color used. A color with 100% saturation will be the purest color possible, while 0% saturation produces grayish color

V = The Value; represent an analog of brightness of the color. A color with 0% brightness is pure black while a color with 100% brightness has no black mixed into it

Histogram of Oriented Gradients (HOG)

In order to detect precise edges and other relevant dense features within the images, 2×2 *block size* which covers an 8×8 *pixels* cell neighborhood with 50% overlap between a block and the next block and binned over 8 angular directions histogram (every 45° along a unit circle) spanning from 0 *to* 180 *degrees* was used in the study. This resulted in a HOG feature descriptor with 2048 *feature dimension* ($16 \times 16 \times 8$ *feature size*). HOG feature descriptor takes into consideration the occurrence of gradient orientation in local regions of an image and hence, it is invariant to geometric and photometric transformations *(Dalal et al., 2005)*.

Local Binary Patterns (LBPs)

Local Binary Patterns (LBP) *(Ojala et al., 2002)* is a powerful and efficient non-parametric representation of the texture component of an image as a texture descriptor. The following steps were implemented in constructing LBP texture descriptor for each image as:

i. The image was first converted to grayscale
ii. For each pixel p in the grayscale image, a neighborhood of size $\boldsymbol{r}$ (3×3) surrounding the center pixel was selected as shown in Figure 4.3, the center pixel was thresholded against its remaining 8 neighborhood pixels.
iii. The LBP value for the center pixel was then computed and stored in a 2D array output with the same width and height as the input image. It then followed that, if the intensity of the center pixel was

greater-than-or-equal to that of its neighbor, then its value was set to 1; else, it was set to 0. This generated a set of binary values stored as an 8-bit array, which was then converted to a decimal value. A total of $2^8 = 256$ possible combinations of LBP codes was obtained from the 8 surrounding pixels. This step of thresholding, coding binary values, and storing the output decimal value in the LBP array was iterated for every pixel in the input image.

iv. Finally, the histogram was then computed over the output LBP array. Since a 3×3 neighborhood had $2^8 = 256$ possible patterns, the LBP 2D array thus had a minimum value of 0 and a maximum value of 255, which yielded a $256 - bin$ histogram of LBP codes as the final feature vector.

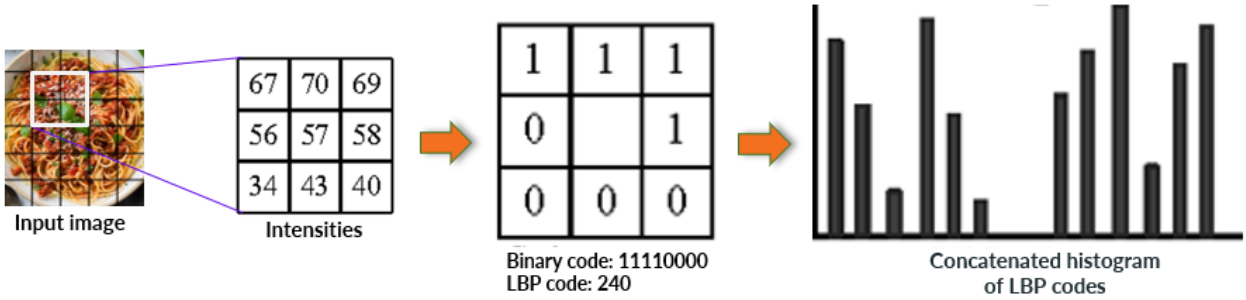


Figure 4.3: An illustration of the LBP descriptor

Scale Invariant Feature Transform (SIFT)

The Scale Invariant Feature Transform (SIFT) descriptor extracts local features at key points which have high informative content, stable under local and global noise, as well as substantially robust to a wide range of image translation, scaling, rotation, changes in 3D viewpoint, and partially invariant to illumination changes, and affine distortion *(Lowe, 1999, 2001; Lowe, 2004)* The SIFT features were extracted using the following four steps:

i. **Scale-Space Extrema Detection**: This is a filtering stage that involved the use of Difference of Gaussian function to detect stable keypoint locations in scale space which are invariant to changes in scale and orientation. The scale-space extrema, $D(x, y, \sigma)$, was detected by computing the Difference of Gaussian of two images, one with scale k times the other as given in the Equation 4.2 *(Gonzalez et al., 2018; Lowe, 2004)* below,

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] \star I(x, y)$$
$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (4.2)$$

Where " $\star$ " is the convolution operator, $G(x, y, \sigma)$ is a variable-scale Gaussian, $L(x, y, \sigma)$ is the scale-space representation and $I(x, y)$ is the input image.

ii. **Keypoint Localization:** The keypoint localization was performed in order to reject unstable points from the list of keypoints by eliminating those that have low contrast or are poorly localized on an edge.

iii. **Orientation Assignment:** A consistent orientation was then assigned to each keypoints based on local image properties. This allowed for ease of representing a keypoint, relative to its orientation, and hence achieving invariance to image rotation.

iv. **Keypoint Descriptor:** Finally, the, feature descriptor vector was computed for each keypoint. The gradient information was rotated to align with the orientation of the keypoint and then weighted by a Gaussian weighting function with a variance of $1.5 \times keypoint\ size$. This data was then used to generate multiple histograms over a window centered on the keypoint. The descriptor computed uses a set of 16 histograms, aligned in a 4×4 subregion, each with 8-directional bins (the bins are multiples of 45°), one for each of the main gradient directions and one for each of the mid-points of the directions. This resulted in a feature vector containing 128 elements known as SIFT keys were used as part of the training features to train the machine learning algorithms.

4.2.2.2 Image classification model (hyperparameter tuning, model selection, and model evaluation)

In order to identify the machine learning algorithm that is best suited for the dataset, seven different machine learning classification algorithms (Random Forrest, Logistic Regression, Linear Discriminate Analysis, K-Nearest Neighbor, Classification and Regression Tree, Gaussian Naïve Bayes and Support Vector Machine) were trained using the training set. Different combinations of the algorithm's hyperparameters were compared and tested for the purpose of selecting the top-performing ones.

Hyperparameter tuning and model selection

Different sets and combinations of hyperparameters were experimented on during the training of the seven algorithms considered and were used to produce a set of classification models. The best set of hyperparameters and the top-five-performing models were selected using K-fold Cross-validation and learning curves approaches. In addition to selecting hyperparameters and model, these approaches also helped to obtain the right balance between bias and variance that best yield an optimal model performance.

a. 10-fold Cross-validation

In this approach, the training data (80% of the dataset) was shuffled and randomly split into 10 groups. As illustrated in Figure 4.4, one group was used for validation and the rest for training. The process was repeated until all the ten groups have been used for validation once.

Figure 4.4: Illustration of 10-fold cross-validation

b. Learning curve:

In the learning curve, the performance of the models both on the training and validation set was plotted as a function of the training set size. The learning curve was used to observe the effect of the varying size of the training data on the generalization performance of developed classification models. This helped to further tune the parameters of the models and to select top performing models.

Model evaluation

The selected models were evaluated in order to estimate the generalization ability of the selected model on unseen data or on the held-out test set, i.e., how well the selected model performed on unseen data using specific evaluation metrics. Accuracy, precision, and recall computed using Equation 4.3, 4.4, and 4.5*)*, were the three (3) main model evaluation metrics used to report the performance of the model on the held-out test set. Accuracy described how often the models were

correct overall, precision described how precise or how often the models were able to correctly predict the actual or correct labels, and recall (also known as sensitivity) described how sensitive or how often the models were able to detect the correct or actual labels.

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}\ ; \tag{4.3}$$

$$\boldsymbol{Precision} = \frac{TP}{TP + FP}\ ; \tag{4.4}$$

$$\boldsymbol{Recall} = \frac{TP}{TP + FN} \tag{4.5}$$

Where *TP* indicates True Positives (foods correctly detected); *FP* indicates False Positives (foods incorrectly detected or misclassified foods); *TN* indicates True Negatives (food not detected).

Confusion matrix was further used to display the performance of the models on the test set by matching the predicted classes with the actual classes. The diagonal elements of the confusion matrix showed the fractions of food images that were correctly predicted while the off-diagonal elements represented misclassified food images.

4.2.2.3 **Model classification strategy**

Two different strategies of deploying classification models namely: Single and Ensembled classifier strategy, were used in this study.

Study A: Single classifier strategy

Using the feature extraction algorithms (e.g. Color Histogram, HOG, LBP), every image in the training set was represented by a collection of corresponding feature vector(s) and their associated class label. In the single classifier strategy, the feature vectors were used to independently train and build the classification models from the selected classification algorithms (K-Nearest Neighbors, Logistic Regression, Random Forest, Support Vector Machine, and Linear Discriminant Analysis). In addition, the single classifier strategy also involved training the models on different combinations of the feature vectors and then compare the results. The algorithm of some of the selected classifiers are described in *Appendix A*.

Study B: Multiple or ensemble classifier strategy

In the ensemble strategy illustrated in Figure 4.5, the 5 selected classification or learning algorithms were trained on the extracted feature vectors and then combined or ensembled to form

a single classification model. Furthermore, different combinations of the feature vectors were also used to train and build ensemble models in order to compare their performances. The ensemble method used called *Voting method* involved three steps. Firstly, the selected classifiers were trained independently on the training set. In the step that followed, each of the models returned for a testing instance, a probability vector for each predicted food class i.e. probabilities of predicting the food classes being considered. Finally, the final class label was selected by first computing the linear combination of the weights $\boldsymbol{w}$ and the predicted class probabilities $\boldsymbol{p}$ and then selecting the class with the highest probability as seen in Equation 4.6.

$$\hat{y} = \arg\max_{i} \frac{1}{n} \sum_{j=1}^{n} w_{ij} p_{ij} \tag{4.6}$$

Where, $\hat{y}$ = final predicted class, w_i = weight assigned to the jth classifier, p_i = predicted probability for the jth classifier

n = number of classifiers

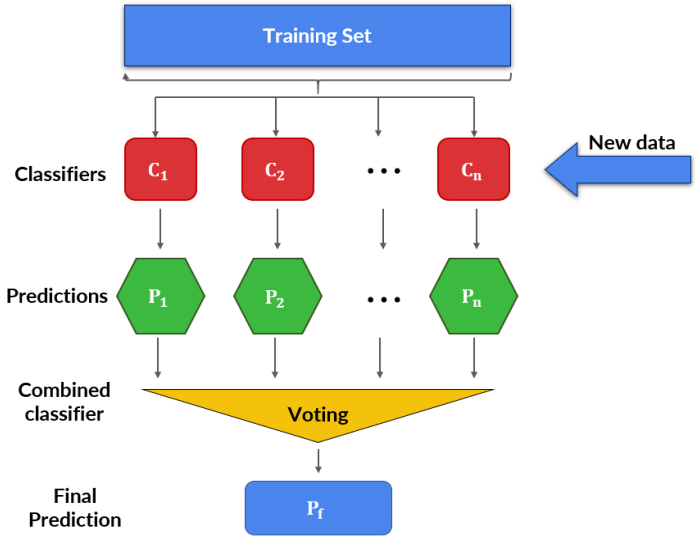


Figure 4.5: Architecture of the Ensemble method

4.2.3. *Weight and nutrient estimation*

The weights and nutrients composition of the food items considered in this study were obtained from the Food and Nutrient Database for Dietary Studies (FNDDS). The FNDDS also consist of

known serving size and weights of commonly consumed foods and beverages as reported in the What We Eat in America (WWEIA) database. As a typical illustration, as shown in Table 4.1, the portion weight of a single medium serving size of a regular croissant is 57 g as reported in the FNDDS database. The portion weight of the food can be varied to account for larger food portions or multiple serving based on the quantity consumed by multiplying the portion weight by a factor e.g. $1, 1.5, 2, 3$ etc. The weight and the nutrient values obtained for each of the food class were further used as part of the input data for the nutrient profiling.

Table 4.1: Nutrient information of the 10 foods retrieved from the FNDDS

	Portion description	Estimated portion weight	Energy	Protein	Fiber	Calcium, Ca	Iron, Fe	Ascorbic acid, Vit. C	Vitamin D	Sodium, Na	Saturated Fatty Acid	Added Sugar
				65	**25**	**900**	**12.5**	**110**	**5**	**3153**	**22**	**50**
Food item		**(g)**	**(kcal)**	**(g)**	**(g)**	**(mg)**	**(mg)**	**(mg)**	**(µg)**	**(mg)**	**(g)**	**(g)**
Avocado	1 regular size	201	321.6	4.02	13.467	24.12	1.1055	20.1	0	14.07	4.2733	1.3266
Egg, cheese and ham on bagel	2 servings	436	522	25.72	2	248	3.82	2.8	1.2	1508.00	8.1240	6.1800
Banana (raw)	1 medium (7" long)	118	105.02	1.2862	3.068	5.9	0.3068	10.266	0	1.18	0.1322	14.4314
Cheeseburger	2 regular (medium)	290	780.1	22.0545	1.74	259.55	1.479	0.435	0.145	856.95	10.1979	5.2200
Coconut	1 serving	80	283	2.664	7.2	11.2	1.944	2.64	0	16	23.758	4.984
Cooked beans	1 serving	250	210	10.725	6	106.25	6.25	20.625	0	306.25	0.6738	0.7750
Cooked rice	1 serving	153	253.98	7.8948	0.153	168.3	1.377	0.306	0.459	492.66	4.5365	0.1530
Croissant	3 medium size	171	694.26	14.022	4.446	63.27	3.4713	0.342	0	798.57	19.9369	19.2546
Pizza	2 slices (1/8 of the whole 12")	250	705	29.35	5.75	382.5	6.3	2.25	0	1712.50	12.7300	8.1500
Spaghetti	1 serving	260	218.4	14.612	9.256	59.8	2.808	28.6	0	899.60	3.5230	37.0240

4.2.4. *Nutrient profiling*

4.2.4.1 The SAIN, LIM model

Following the food recognition using the food recognition model, and its corresponding nutrient values retrieved from the FNDDS, the nutrient scoring and nutrient profiling steps were then carried out using the SAIN, LIM model (Equation 4.7 and 4.8 respectively). In this study, the SAIN, LIM model was used to categorize the recognized foods based on their degree of healthiness and unhealthiness and then distributed them into one of the four different classes (or quadrants) as follows: *i. Recommended for good health; ii. food with neutral or balanced health benefits; iii. recommended in less quantities or to be consumed occasionally; iv. consumption should be limited.*

$$\boldsymbol{SAIN\ Score} = \left(\frac{\left(\frac{\left(\frac{Protein}{65} + \frac{Fibre}{25} + \frac{Vit\ C}{110} + \frac{Ca}{900} + \frac{Fe}{12.5} + \frac{Vit\ D}{5} - \min\ term \right)}{5} \times 100 \right)}{E} \right) \times 100 \qquad (4.7)$$

Where:

Protein = Protein content in $g/100\ g$;

Fibre = Fibre content in $g/100\ g$;

Vit C = Vitamin C content in $mg/100\ g$;

Ca = Calcium content in $mg/100\ g$;

Fe = Iron content in $mg/100\ g$;

Vit D = vitamin D content in $\mu g/100\ g$

E = energy density in kcal/edible $100\ g$;

Minimum ratio (min ratio) = the lowest of the 6 [nutrient/DRV] ratios.

$$\boldsymbol{LIM\ Score} = \left(\frac{\frac{Na}{3153} + \frac{SFA}{22} + \frac{Added\ Sugar}{50}}{3} \right) \times 100 \qquad (4.8)$$

4.2.5. *Algorithm implementation*

All the analysis were carried out using Python programming *(Van Rossum, 2007)* within the Jupyter Notebook environment equipped with computer vision, machine learning, and data visualization libraries such as OpenCV *(Bradski, 2000)*, Scikit-learn *(Pedregosa et al., 2011)*, and Matplotlib *(Hunter, 2007),* respectively. The functions to execute the feature extraction and machine learning algorithms were pre-installed as part of the OpenCV and Scikit-learn respectively. The machine used was equipped with an Intel Core i7 7th generation CPU, Geforce GTX1050 NVIDIA graphic card and 16 GB of RAM.

4.3. RESULTS AND DISCUSSIONS

4.3.1. *Model selection and optimization results*

The result (accuracy and standard deviation) of the 10-folds cross-validation for the considered classifiers on the training set is displayed in Table 4.2. The result showed that out of the seven

classification models, the top 5 performing ones were Random Forest, Logistic Regression, Linear Discriminate Analysis, K-Nearest Neighbor, and Support Vector Machine.

Table 4.2: Result of 10-folds cross-validation of the classifier on the training set

Classifiers	Mean Accuracy (%)	Standard deviation
Random Forest	91.82	0.0089
Support Vector Machine	90.46	0.0113
Logistic Regression	88.82	0.0153
Linear Discriminant Analysis	87.30	0.0142
K-Nearest Neighbors	82.00	0.0187
Classification and Regression Tree	79.21	0.0204
Gaussian Naïve Bayes	72.51	0.0106

The learning curve was used to further analyze the performance of the models on the training set in terms of their tendency to overfit or underfit the data. The learning curve compared the performance of the classification models on the training and validation set as a function of increasing size of the training set. With the learning curves, as shown in Figure. 4.6, it was observed that the validation score was first gradually increasing with increasing training set size. However, the validation score started to decline at about 84% which was indicative of the fact that the model had started to pick up noise from the training data (overfitting) due to insufficient training data and this can result to failure of the model to generalize to unseen dataset. The overfitting was mitigated by regularizing the model using the data augmentation technique which was carried out by artificially increasing and balancing the size of the training set by applying computer vision image transformation technique such as rotation, translation, shifting, and scaling on the training data. As a result, the regularization step applied to the training data increased the size by 20% which in turn improved the validation score as shown in Figure. 4.7.

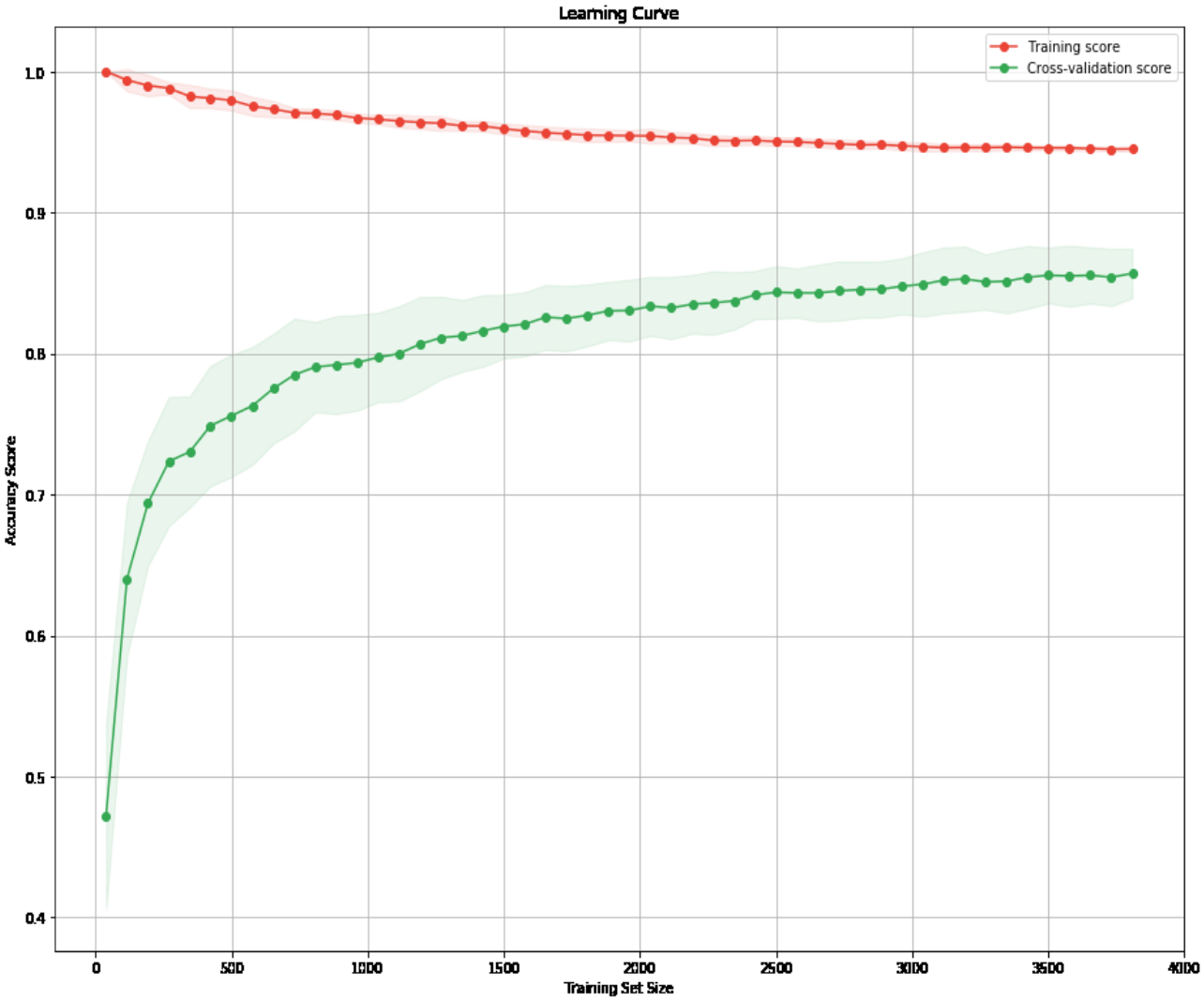


Figure. 4.6: Learning curve before data augmentation regularization step

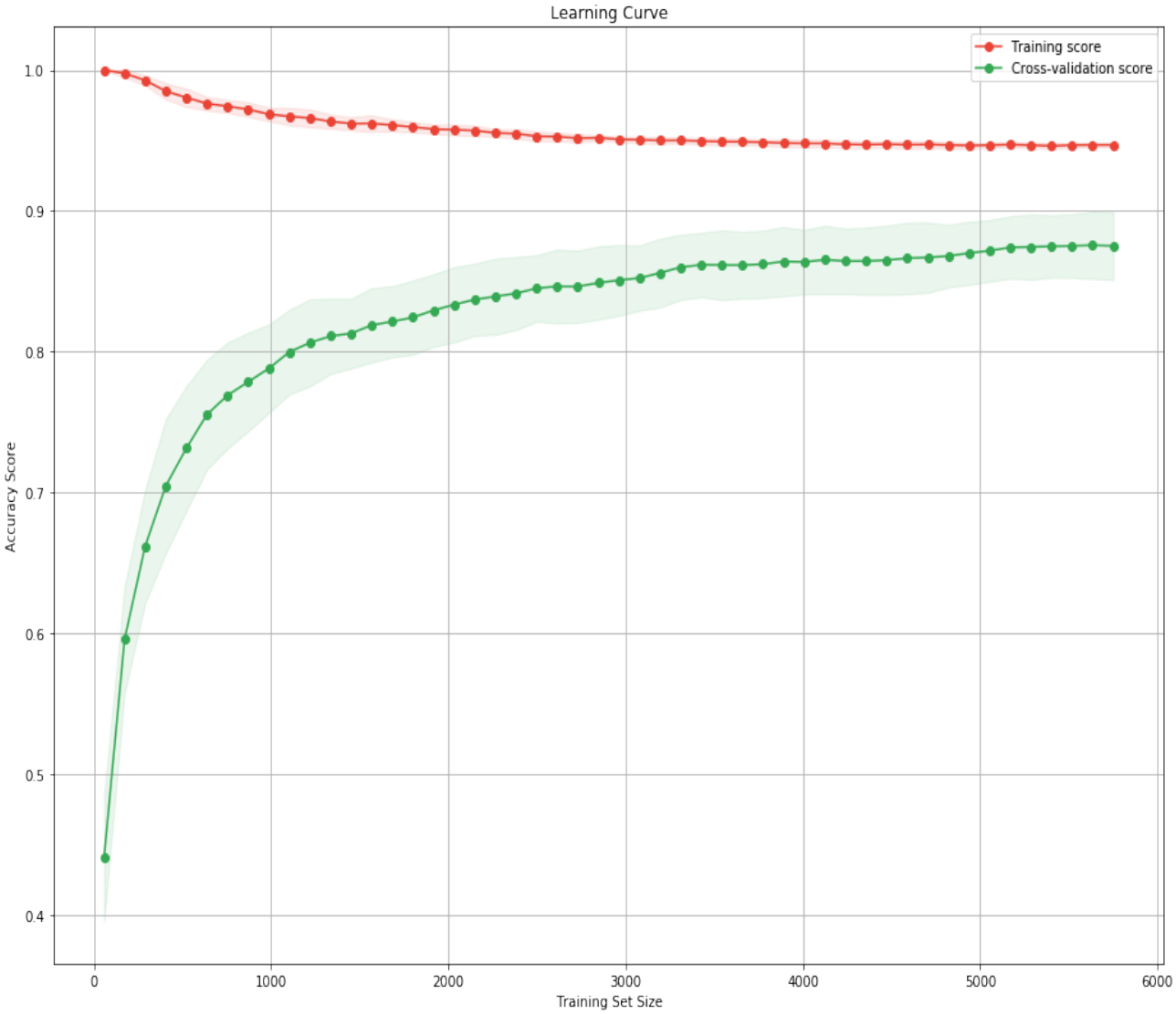


Figure. 4.7: Learning curve indicating improved model performance based on increased training dataset size

4.3.2. *Performance evaluation of the classification algorithms and features extraction algorithm on model generalization*

4.3.2.1 Study A: Results of comparing the performance of individual classification models on individual and combined feature descriptors

This section presents the performances of individual classification models (Random Forest, Logistic Regression, Linear Discriminate Analysis, K-Nearest Neighbor, and Support Vector Machine) built from using the HOG, SIFT, color histogram, and LBP feature descriptors when used individually and in combination as the input feature vectors for training the classification models.. Table 4.3 shows the comparative results of the performance of each of the classifiers on the individual feature descriptors (HOG, Color Histogram (CH), SIFT, and LBP). The results revealed low accuracy, precision and recall when the classifiers are trained on the individual feature descriptors. This implied that the image features generated using the individual features descriptor were insufficient to train the machine learning algorithms (underfitting) and thus, the model performed low on the validation set and is very likely to further fail to generalize to the test data. Table 4.4 presents the results of different combinations of the feature descriptors which included HOG + SIFT, HOG + LBP, HOG + CH, HOG + CH + SIFT, HOG + CH + LBP, and HOG + CH + SIFT + LBP used to train the selected classifiers. This showed better improvement over the individual feature descriptors in term of accuracy, precision and recall across the machine learning algorithms. This indicated that concatenating the feature vectors extracted using the individual machine learning algorithms into a single robust feature vector provided a better visual representation of the foods in the image dataset. This however implied that increasing the robustness of the training features by concatenating the feature vector has the potential to improve the performance of the classification models and also lead to better generalization to the test data. The results showed that the combination of HOG, Color Histogram, and LBP was best as the feature descriptors to consider for building the models which also aligns with the results presented in the work of *Pouladzadeh et al. (2014)* and *Christodoulidis et al. (2015a)* .

Table 4.3: Classification accuracy, precision, and recall (or sensitivity) of the classifiers on single feature descriptors

Classifiers	KNN			LDA			LR			RF			SVM		
Features descriptors	Acc. (%)	Pr. (%)	Rc. (%)	Acc. (%)	Pr. (%)	Rc. (%)	Acc. (%)	Pr. (%)	Rc. (%)	Acc. (%)	Pr. (%)	Rc. (%)	Acc. (%)	Pr. (%)	Rc. (%)
HOG	65.58	70.18	64.05	59.90	60.88	58.64	63.71	63.91	62.14	68.52	69.45	67.43	68.27	67.75	67.07
CH	75.02	74.32	73.32	64.46	66.93	62.51	67.27	67.44	65.17	84.70	84.19	83.71	75.01	74.36	73.15
SIFT	47.59	48.41	45.78	50.09	49.78	49.93	48.90	47.34	47.87	59.17	59.21	58.16	53.35	52.31	52.24
LBP	59.09	59.68	57.16	52.34	50.81	50.78	42.49	31.41	39.15	65.40	65.67	64.05	42.35	32.21	39.35

Table 4.4: Classification accuracy, precision, and recall (or sensitivity) of the classifiers on combined feature descriptors

Classifiers	KNN			LDA			LR			RF			SVM		
Features descriptors	Acc (%)	Pr. (%)	Rc (%)	Acc (%)	Pr. (%)	Rc (%)	Acc (%)	Pr. (%)	Rc (%)	Acc (%)	Pr. (%)	Rc (%)	Acc (%)	Pr. (%)	Rc (%)
HOG + SIFT	47.59	48.41	45.78	69.13	69.25	68.20	58.67	58.02	57.99	72.70	74.29	71.43	53.29	52.27	52.16
HOG + LBP	65.83	70.41	64.32	69.08	69.76	68.09	66.15	66.26	64.70	75.14	77.59	74.03	70.14	69.69	68.96
HOG + CH	73.34	77.16	72.01	78.33	79.52	77.56	81.20	80.75	80.33	86.63	86.48	85.73	83.64	83.09	82.82
HOG + CH + SIFT	47.59	48.41	45.78	82.41	82.84	81.83	62.93	62.28	52.12	87.54	87.33	86.68	53.29	52.27	52.16
HOG + CH + LBP	73.27	77.13	71.95	82.76	83.40	82.26	81.50	81.89	81.05	89.19	88.76	88.66	84.39	83.86	83.64
HOG + CH + SIFT + LBP	47.59	48.41	45.78	84.16	84.15	83.65	63.18	62.56	63.25	89.29	88.88	88.76	53.29	52.27	52.15

4.3.2.2 Study B: Comparison of individual classification models and ensemble model on selected combination of feature descriptors

In this study, the five classifiers (Random Forest, Logistic Regression, Linear Discriminant Analysis, K-Nearest Neighbor, and Support Vector Machine) were individually trained on the combined HOG, color histogram (CH) and LBP training feature vectors and then used to develop the ensemble classification model. The comparative result of the performance of the developed models on the validation set using the individual and the ensembled model is presented in Table 4.5. The result showed that the ensemble model achieved highest accuracy, precision and recall compared to using the classifiers individually. This pointed out the capability of the ensemble model to systematically harmonize the predictive capacity of each of the classification models as well as to mitigate their individual classification errors. The results further implied that ensemble model has higher potential than individual classification model in term of accuracy, precision and recall for developing accurate classification models to better represent the visual features of the food image dataset as well as to produce improved generalization to the test data. A similar ensemble technique was also employed in the work of *Pandey et al. (2017)*.

Table 4.5: Classification accuracy, precision, and recall of the individual vs ensemble model on HOG + CH + LBP feature descriptors

Classifiers [HOG + CH + LBP]	KNN	LDA	LR	RF	SVM	Ensemble model
Accuracy (%)	73.27	82.76	81.5	89.19	84.39	**90.32**
Precision (%)	77.13	83.4	81.89	88.76	83.86	**90.13**
Recall (%)	71.95	82.26	81.05	88.66	83.64	**89.95**

4.3.3. *Model performance comparison and evaluation*

Using classifiers and feature descriptors individually to develop the models led to underfitting and poor generalization performance to the unseen data (held-out test set) as shown in study A and B. As observed, the high bias and low training and validation accuracies were due to inadequacy of the model to learn enough features from the training data. However, the performance of the models increased when the feature descriptors were combined, and with the combination of *HOG, color histogram, and LBP* producing the highest across-board classification result while the combination of *HOG and SIFT* performed the lowest. Random forest (RF) and support vector machine (SVM) with global accuracies of 89.19% and 84.13% respectively, performed well when the classifiers were considered individually but the ensemble method displayed a much more significant performance with the combination of *HOG, color histogram, and LBP* with global accuracy of 90.3%. The results obtained were higher than that obtained in the work presented

by *Anthimopoulos et al. (2014); Yanai et al. (2015a)* and *Farinella et al. (2016)*. However, the results were in line with results obtained in the work of *Pouladzadeh et al. (2015), E Silva et al. (2018)* and *Ahmed et al. (2019).* This shows that the combination of HOG, color histogram, and LBP feature descriptors, which extracted gradient-based features, color space-orientation and textural features respectively were good visual descriptors that enabled the model to maintain a good balance of variance and bias and to also generalize well to the unseen dataset with a considerable level of accuracy, precision and recall as compared to other combinations of feature descriptors experimented on. The concatenation of these feature vectors into a single and robust feature vector gave rise to a good and rich feature descriptor with wide range of varying image data representation. The performance of selected feature descriptor combination was further compared on the basis of individual models as well as the ensemble model on each class of food as seen in Table 4.6. The table shows the results of comparing the classification performance of the individual classifiers and the ensemble method on the food classes. Figure 4.8 (a and b) shows the confusion matrices for the classification of the 10 unique foods in the test set using the ensemble model and the best performing model among the 5 single classification models. The diagonal elements of the confusion matrices indicated the fractions or percentages of food items that were correctly classified or predicted while the off-diagonal elements represent misclassified food items. The misclassifications, even though in infinitesimal percentages, were majorly a result of similarities in the appearance (color and texture) of the food items which can be mitigated by collecting more training images. Figure 4.9 shows samples of food that were misclassified due to similarities in their features. Figure 4.10 (a) and (b) presents a graphical representation of the comparative results.

Table 4.6: Classification precision and recall (sensitivity) of the individual classifiers and the Ensemble method (EM) on each food

Classifiers	**LR**		**RF**		**SVM**		**EM**	
Foods	Precision (%)	Recall (%)	Precision (%)	Recall (%)	Precision (%)	Recall (%)	Precision (%)	Recall (%)
avocado	96	97	99	99	96	99	99	98
bagel	74	72	78	83	75	76	78	87
banana	93	90	97	97	93	94	97	94
cheeseburger	85	83	85	93	83	85	90	91
coconut	100	89	99	92	97	94	100	93
cookedBeans	82	68	87	86	86	69	93	88
cookedrice	84	93	90	91	87	87	91	93
croissant	69	67	81	79	71	76	85	79
pizza	63	72	86	87	77	75	90	89
spaghetti	68	79	87	79	73	82	78	87

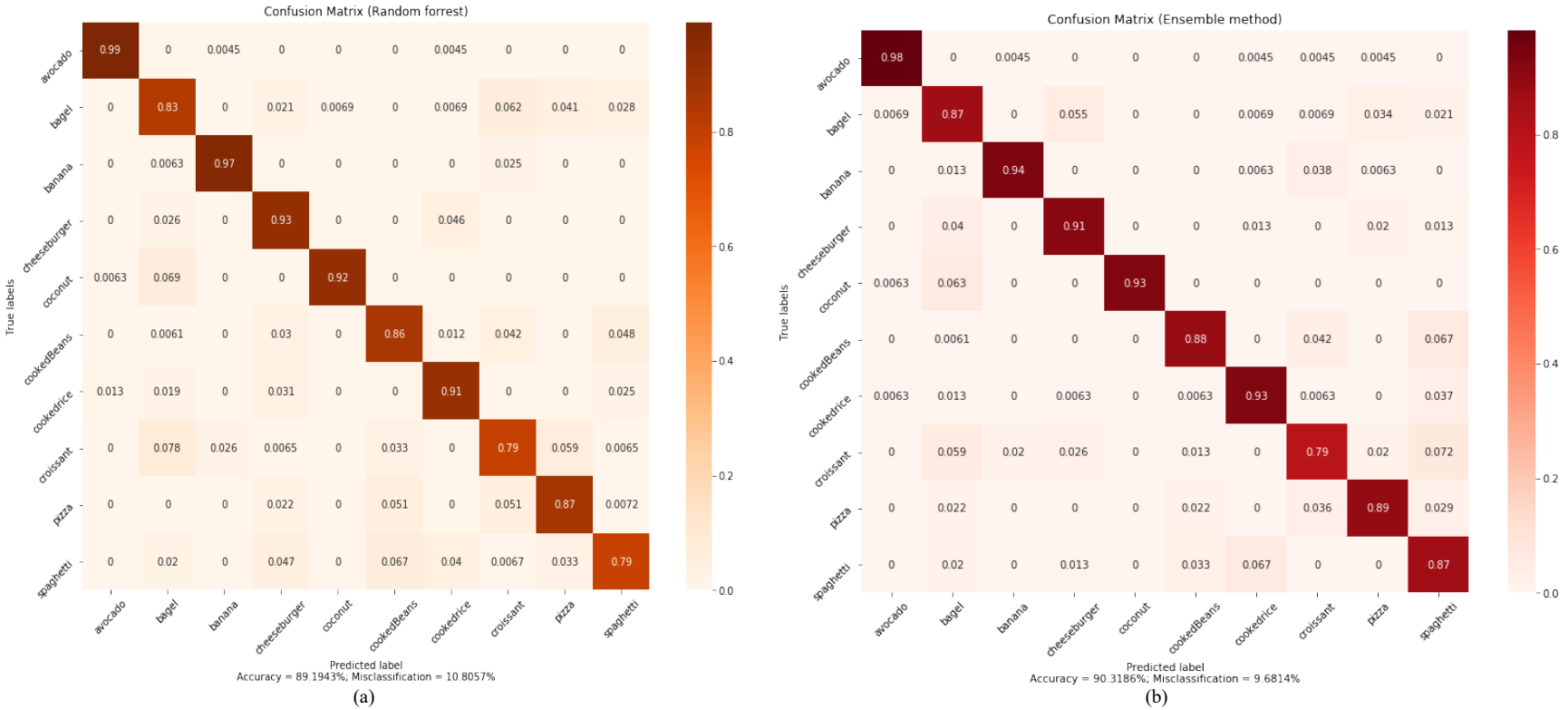
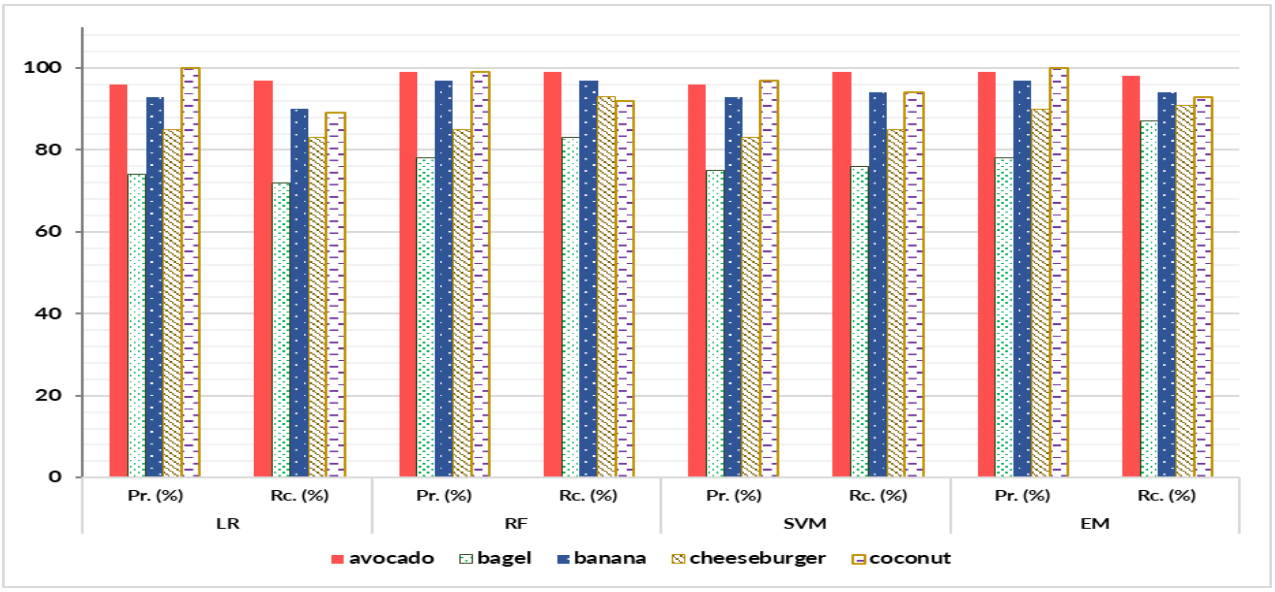


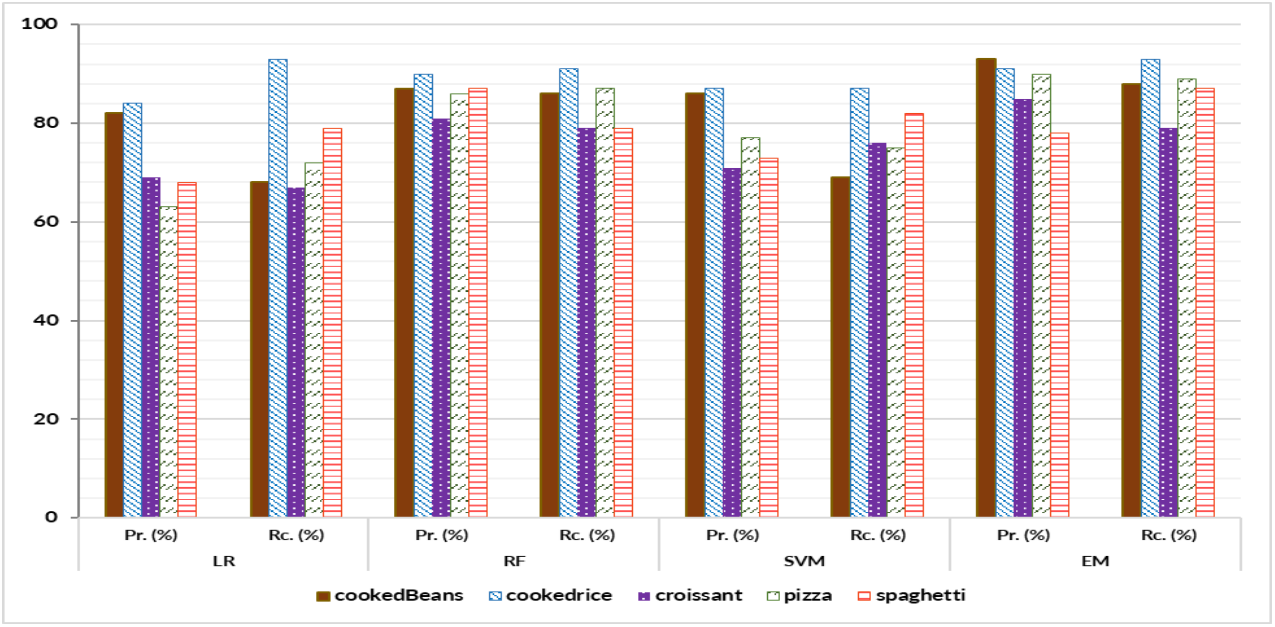
Figure 4.8: Confusion matrix of (a) Random forest model (b) Ensemble model for the classification of the 10 food products in the dataset

(a) (b) (c)

Figure 4.9: Examples of incorrectly classified food images (a) bagel classified as croissant (b) pizza classified as cheeseburger (c) cooked beans classified as spaghetti



(a)



(b)

Figure 4.10 (a and b): Graphical representation of the performance comparison between individual classifiers ensemble method on the foods

4.3.4. *Nutrient profiling*

4.3.5. *Evaluating and implementing the nutrient scoring model*

Table 4.7 shows the food items considered in this study, their portion description, respective serving or portion size weight and as well as their SAIN and LIM scores. As an illustration using Equation 4.7 and 4.8, Cooked beans with serving size of $\mathbf{250\ g}$, contains **Energy**: 210 kcal, **Protein**: 10.725 g, **Fibre**: 6 g, **Ca**: 106.25 mg, **Fe**: 6.25 mg, **Vit** C: 20.625 mg, **Vit** D: 0 μg, **Na**: 306.25 mg, **SFA**: 0.674 mg, **Added sugar**: 0.775 g. The SAIN and LIM scores were computed as follows:

$$\boldsymbol{SAIN\ Score} = \left(\frac{\left(\frac{\left(\frac{10.725}{65} + \frac{6}{25} + \frac{106.25}{900} + \frac{6.25}{12.5} + \frac{20.625}{110} + \frac{0}{5} - 0.013 \right)}{5} \times 100 \right)}{210} \right) \times 100$$

$$= \left(\frac{\left(\frac{(1.2105 - 0.118)}{5} \times 100 \right)}{210} \right) \times 100$$

$$= \left(\frac{21.85}{210} \right) \times 100$$

$$\boldsymbol{SAIN\ Score = 10.405}$$

$$\boldsymbol{LIM\ Score} = \left(\frac{\frac{306.25}{3153} + \frac{0.674}{22} + \frac{0.775}{50}}{3} \right) \times 100$$

$$= \left(\frac{0.1432}{3} \right) \times 100$$

$$\boldsymbol{LIM\ Score = 4.773}$$

From the above analysis, once cooked beans have been recognized using the ensemble classification model, the diet quality of the food was then analyzed using the SAIN-LIM nutrient profiling model. Cooked beans have a SAIN Score and LIM score computed as $\mathbf{10.405}$ and $\mathbf{4.773}$ respectively. This result shows that cooked bean has high SAIN score which is indicative of a relatively high amount of qualifying nutrients such as Protein, Calcium and Fiber while the low LIM score indicated that the cooked beans contained minimal amount of the disqualifying nutrients or nutrients to be consumed in less quantities such as Sodium, SFA, and Added Sugar.

The results however implied that the diet quality of consumed food (e.g. cooked beans) in terms of their healthiness significance can be modeled by estimating and analyzing the presence of certain nutrients present in food (qualifying and disqualifying nutrients). Similar findings were also deduced from the study carried out by *Darmon et al. (2009).*

Table 4.7: Computed SAIN, LIM scores for the 10 foods

	Portion description	Estimated portion weight	SAIN - LIM		
Food item		**(g)**	**Min ratio**	**SAIN**	**LIM**
Avocado	1 regular size	201	0.0268	5.4210	7.5038
Egg, cheese and ham on bagel	2 servings	436	0.0255	4.9688	32.3716
Banana (raw)	1 medium (7" long)	118	0.0066	4.9587	9.8337
Cheeseburger	2 regular (medium)	290	0.0040	2.1654	27.9909
Coconut	1 serving	80	0.0124	3.5937	39.4888
Cooked beans	1 serving	250	0.1181	10.4048	4.7752
Cooked rice	1 serving	153	0.0028	4.0676	12.1838
Croissant	3 medium size	171	0.0031	2.1363	51.4862
Pizza	2 slices (1/8 of the whole 12")	250	0.0205	4.5689	42.8257
Spaghetti	1 serving	260	0.0664	9.8872	39.5311

4.3.6. *Nutrient profile visualization*

In order to verify the ability of the food recognition and nutrient profiling system, the 10 classes of food considered in this study were presented as a food record plan indicating a set of foods consumed by a user during lunch over a period of ten days. The results obtained support the following discussion. Firstly, the image features were extracted and classified for the 10 foods into their respective classes with reliable accuracies. Secondly, the adoption of the SAIN, LIM nutrient profiling model enabled the classification of the foods based on *their healthy'' and ''unhealthy''* benefits to the diets of the user over the pre-defined food record duration using the calculated SAIN and LIM scores. The values obtained by computing the SAIN and LIM scores for all the 10 food items were also used to visualize the nutrient profile of the foods as shown in Figure 4.11. The figure illustrated the SAIN and LIM nutrient profile of the 10 foods considered in the 10-day food record (Avocado, Bagel, Banana, Cheeseburger, Coconut, Cooked beans, Cooked rice, Croissant, Pizza, and Spaghetti) on a logarithmic scale. The logarithmic scale was used both for the LIM and the SAIN scores, due to the ease of response of the scale to skewness towards large

values; i.e., in cases where one or a few values are much larger than the majority of other values of the two scores, especially the SAIN.

Figure 4.11: Graphic representation of the classification of selected foods with the SAIN,LIM (score of nutritional adequacy of the individual foods, and score of nutrients to be limited) model and their position within the 4 nutrient profile classes (on log scales)

Thirdly, the portion weight retrieved from the FNDDS database for Avocado and Cooked beans were 201 g and 250 g respectively and their associated SAIN scores (5.42 and 10.41 respectively) were greater than 5 but their LIM scores (7.50 and 4.78 respectively) were not more than 7.5. This implied that the foods have high nutrient density, i.e. rich in nutrients recommended for healthy living and low in nutrients that are considered to be unhelpful to the body. Spaghetti in the 3rd quadrant is high in both the SAIN and LIM scores (9.89 and 39.53 respectively) is considered to be less beneficial in terms of nutritional contribution and hence, recommended in less quantities or to be consumed occasionally. The SAIN scores for Bagel, Banana, Cheeseburger, Coconut, Cooked rice, Croissant, and Pizza ($4.97, 4.96, 2.17, 3.59, 4.07, 2.14$ *and* 4.57 respectively) in the 4th quadrant were less than 5 and their LIM scores ($32.37, 9.83, 27.99, 39.49, 12.18, 51.49$, and 42.83 respectively) were all greater than 7.5. This is indicative of the high constituents of the nutrients considered to be unhealthy namely: sugar, Sodium and saturated fat and are therefore, based on the quadrant they fall into, recommended to be consumed in very limited quantities. Hence, the results obtained implied that the approach has good practical value in terms of determining the quality of the dietary intake of users and hence potentially manage the nutritional value of their food intakes.

4.4. CONCLUDING REMARKS

In this study, a food recognition and image-based nutrient profiling system that is capable of predicting the class of a captured food image and then assessing the nutritional benefits of the food to the diets of the user or a population group under study has been presented. The approach first trained and validated a food recognition model developed for single foods using state-of-the-art hand-engineered computer vision feature extraction algorithms (HOG, color histogram, and LBP) and machine learning classification algorithms (Random Forrest, Logistic Regression, Linear Discriminant Analysis, K-Nearest Neighbor, and Support Vector Machine) by experimenting on different cross-sections of the dataset provided. The developed model was used to make reasonable prediction or generalize to new food image data or unseen data presented to it. Furthermore, the diet quality of the recognized food was then assessed using the SAIN, LIM nutrient profiling model. The proposed system has the potential of being implemented on smart mobile devices and as such becoming a reliable and convenient tool for daily self-administered dietary intake monitoring, recording, and nutritional status assessment. The experimental results proved the effectiveness of the system which produced satisfactory results for recognizing food with high precision and recall values ranging from $75 - 100\%$. The computer vision features extraction and machine learning strategy adopted for the development of the recognition model is very promising and is often beneficial in events where the dataset size is small and computational capacity is limited as is the case this study. The strategy was easy to deploy to extract visual image features which were classified by the mainstream machine learning algorithms.

As observed in this study, the greater the number of features extracted from images of food and the more distinct those features are, the better the performance of the recognition system. In future work, there are intentions to improve the food recognition system by integrating deep neural network-based approaches including transfer learning to improve the performance of the image recognition component of the system. Another possible direction for future work is to cover more food and meal instances in the database to support more mixed or composite foods.

Preface to chapter 5

An interesting lesson from the study in the previous chapter was that, the greater the features extracted from food images and the more distinct those features are, the better the performance of the recognition system. However, this is very tedious and time-consuming. So, this now begs the question, "*can an approach to generate or extract or learn unlimited abstract features from images, provide a quality visual representation of the object(s) in the image… and can this be done automatically?*"

Careful review of literature revealed that there have been promising advances in the application of Deep Learning for image classification and object recognition.

Chapter 4 explored Deep Learning techniques, specifically Deep Convolutional Neural Networks and established how they can be used for food recognition to better aid dietary intake monitoring and nutritional status evaluation. In particular, this chapter deals with evaluating the nutrients present in Composite or Mixed foods (from images) with complex and overlapping features.

5 Deep Learning Assisted Composite Food Recognition and Diet Quality Assessment System

ABSTRACT

Poor dietary intake is implicated in the occurrence of ill health and deaths around the world. Researchers attempts to address these challenges by measuring, analyzing and monitoring dietary intakes. However, the process comes with a lot of challenges. Conventional methods often fail due to the inherent complexity and inaccurate reporting by users. In the past decade, image-based diet monitoring and assessment approaches is fast becoming the most promising methods for managing diets and assessing diet-related diseases.

Computer vision and machine learning offer a cornucopia of useful ways to aid diet monitoring challenges that otherwise defy conventional approaches. This study presents a deep learning assisted diet quality assessment system capable of recognizing and evaluating the nutritional adequacy and healthiness benefits or level of recommendation of a food consumed by a user.

Concretely, a transfer learning approach was implemented to develop a food recognition system capable of recognizing a food from a captured image with very reliable accuracy while the SAIN-LIM model was used to assess the diet quality of the food in order to determine the nutritional adequacy and healthiness significance of the food. Experimental results showed a promising potential use of the system by dietary researcher to track and analyze the quality of a client's diet as well as by any user to self-administer a well-controlled and maintained healthy diet.

Index Terms: Dietary quality assessment, computer vision, deep learning, deep convolutional neural networks, transfer learning, artificial intelligence, image-based, nutrient profiling, nutritional status evaluation

5.1. INTRODUCTION

Conventionally, smart mobile-based dietary assessment tools attempt to solve the problem of food intake monitoring in four fundamental approaches namely: recognize or detect the food, classification of the type of food in image, manually compute volume or weight estimation, and then estimate or retrieve the nutrients information in the food from publicly available food and nutrient databases *(Dehais et al., 2017; Hamid et al., 2016; Heravi et al., 2015; Lo et al., 2018).* Previous studies *(Anthimopoulos et al., 2015; Zawbaa et al., 2014)* focused largely on the conventional food image recognition approaches which involve extraction of visual features from food images using several combinations of specialized traditional computer vision algorithms such as Gabor filters, Scale Invariant feature Transform (SIFT), Color Histograms, Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG), and several others. The feature vectors extracted were then used to train and validate classification models developed from classical machine learning algorithms such as Support Vector Machine (SVM), Random forest, and Logistic Regression, in order which can be used to classify new food images. In the work of *Chen et al. (2012),* accuracy of 46% and 53% were obtained when SIFT and LBP features extracted from food image dataset were used independently to train a multi-label SVM classifier. When the features were combined in addition to Gabor filter and color histogram features, the accuracy was improved to 68%. In a similar study by *Beijbom et al. (2015),* on the same published dataset, SIFT, LBP and color features were extracted in combination with other hand-engineered computer vision features extraction algorithm such as HOG and MR8 filter and the accuracy improved to 77.4%

Food items generally tend to show intra-class variation depending upon the method of preparation, as well as the ingredients used. This leads to complex variations in terms of shape, size, texture, and color. Traditional features perform considerably well on small to medium size homogeneous dataset with few variations in the feature space. However, the features tend to fail if there are no adequate training images to capture complex variations in the feature vector especially in mixed or composite food images *(Heravi et al., 2015; Subhi et al., 2019)*. The importance of automatically generating complex and meaningful visual feature from food images using much more robust feature engineering tools cannot be overemphasized, hence the need for deep learning.

In the last decade, computer hardware resources and computational power has grown tremendously, and researchers now use graphical processing units (GPUs) for parallel computations during the implementations of artificial neural networks *(Oh et al., 2004).* The inherent ability of GPUs to perform massive parallel computations made it possible to effectively handle the popular dense linear algebra

matrix-vector multiplication steps in neural networks *(Pierson et al., 2017)*. In a similar manner, data has never been more available, and progressively becoming the '***plethoric oil***'. Particularly, ease of generating image data from different kind of devices have witnessed tremendous growth in recent years which is indicative of the leap into the Big Data and Internet of Things (IoT) era of technology where data and computing capabilities are readily available and easily accessible. These recent breakthroughs have spurred a great deal of research toward the application of deep learning *(Goodfellow et al., 2016; LeCun et al., 2015; Schmidhuber, 2015)* to solve problems in various fields such as robotics navigation *(Pierson et al., 2017),* speech recognition *(Noda et al., 2015),* natural language processing *(Zhang et al., 2019),* self-driving cars *(Bojarski et al., 2016)*, agricultural production *(Kamilaris et al., 2018)*, remote sensing *(Cheng et al., 2016)*, medical imaging analysis *(Shen et al., 2017)*, and food detection and recognition *(Aguilar et al., 2017; Christodoulidis et al., 2015a)*. Deep learning implementations particularly Deep Convolutional Neural Network (CNN, or ConvNet) *(Krizhevsky et al., 2012)* have been seen to surpass human performance on the popular object recognition benchmark dataset, ImageNet *(He et al., 2015).* The idea behind ConvNets is that, it progressively passes input images through sequence of convolutional (feature extraction) and pooling layers, assign importance parameters (learnable weights and biases) to various aspects or objects in the image in order to perform classification. ImageNet gave rise to the development of several well-known open source CNN architectures including, AlexNet *(Krizhevsky et al., 2012)*. VGGNet *(Simonyan et al., 2014)*, GoogLeNet *(Szegedy et al., 2015)*, and ResNet *(He et al., 2016)*. The impressive classification accuracies derived from these architectures subsequent to being trained on huge dataset motivated the concept of Transfer Learning *(Pan et al., 2009)* which deals with repurposing the knowledge (learned parameters) acquired from training a CNN on huge dataset to a different but related task with smaller dataset.

In food images recognition tasks, the major challenges are the large variations in food due to shape, color, texture, volume, ingredients, composition as well as image background noise *(Zhou et al., 2019)*. The successes and popularity of CNN architectures as well as the increasingly publicly available food image datasets contributed to the growing success in the application of deep learning for food recognition (see *Appendix C*), detection, and to aid dietary intake assessment *(Min et al., 2019).* Furthermore, researchers can now (e.g. using deep learning API libraries (Tensorflow, Pytorch, Keras) implement transfer learning by fine-tuning and retraining open source pre-trained models (with inherent ability to extract features such as texture, color, high-level abstract representations, etc.) on their own food image dataset. This approach has proven to significantly reduce training time and increase accuracy *(Sahoo et al., 2019; Zhou et al., 2019)*.

Composite food (e.g. pasta dishes, sandwiches, salad, etc.) can be described as foodstuff intended for human consumption that contain a compound of different food products either from animal origin or plant based, consumed in processed, semi-processed or unprocessed/ raw form *(UK Food Standards Agency, 2017).* Composite foods contain several combinations and overlapping of foods and ingredients which give rise to numerous blends of visual features including color, shape, and texture. Images generated from composite foods contain complex variations of visual features which tends to be difficult for classical machine learning algorithm or simple ConvNets to adequately represent or classify. Researchers such as, *Heravi et al. (2015)*, and *Subhi (2018)* have presented recognition systems to aid different applications where they utilized dataset with food images that contain single food, or foods with distinctive shape, colors, and textures (often of uniform pixels), or datasets that contain images of very few composite food and several single food combinations. However, in recent time, there is a need for systems that can recognize composite foods since these days people sometimes tend to consume more of composite foods compared to single foods (Figure 5.1). However, this research area has received very minimal research attention due to complexity involved in manipulating the highly varying visual features. To this end, recognizing composite foods requires food recognition systems to adopt, not just convNets, but state-of-the-art deeper convolutional neural network architectures or through the implementation of transfer learning to extract robust high-level visual features that can be effectively used to recognize food from images.

(a) (b) (c) (d)

Figure 5.1: Comparison between Single food (a) & (b) and Composite or Mixed food (c) & (d)

In light of the current context, the objective of this study was to develop a deep learning assisted composite food recognition and diet quality assessment system.

5.2. MATERIALS AND METHODS

5.2.1. *Dataset protocol*

5.2.1.1 Image data acquisition for the food recognition system

A 15-class composite food images dataset was used in this study and was obtained from two sources, namely: (i) randomly selected composite food images from the Food101 *(Bossard et al., 2014),* (ii) composite food images collected by crawling image search engines like Google Images, Bing Images, and Pinterest, using an open source web crawling tool *(Fatkun, 2019)*.

5.2.1.2 Image pre-processing

The preprocessing steps employed in order to ensure the dataset is compatible and indeed suitable for the recognition model development are as follows.

Data formatting, cleaning, labelling and partitioning

The food images in the dataset (especially the ones in unsupported image formats) were first converted into Portable Network Graphic (PNG) formats because of the inherent characteristics of preserving the quality of the original images and preventing data loss which in turn makes the pre-processing easy. The images were then resized (i.e. to 300×300 with an intuitive foresight that, the models considered in this study require input images with size of either 244×244 or 299×299 which can be further resized and delivered in batches of image tensors to the models during the training phase), labelled and partitioned in the ratio [70%: 15%: 15%] into training, validation and test sets each containing 15 classes of food. Image Dataset Pre-Processing algorithm (see *Appendix A*) was used for this pre-processing step. This resulted into a well-balanced and distributed dataset with 13,030 images which contained 15 classes of randomly selected composite food images, and by popular nomenclature, the dataset was referred to as $\boldsymbol{Food15}$**.** In details, the $Food15$ dataset contained 9,030 training set ($602\ images\ per\ class$) and 2000 (133 images per class) validation and test sets belonging to 15 classes, namely*: Lasagna, Steak_with_mashed_potatoes, Spaghetti_beef_tomato-sauce, Macaroni_and_cheese, Fried_rice, Fish_and_chips, Chicken_curry, Hot_dog, Rice_and_beans, Beef_salad, Pizza, Egusi_Soup, Pad_thai, Waffles_and_fruits, Cheeseburger,* which cut across dishes from different regions including Asia, West Africa, and North America. Figure 5.2 shows examples of images in the Food15 dataset.

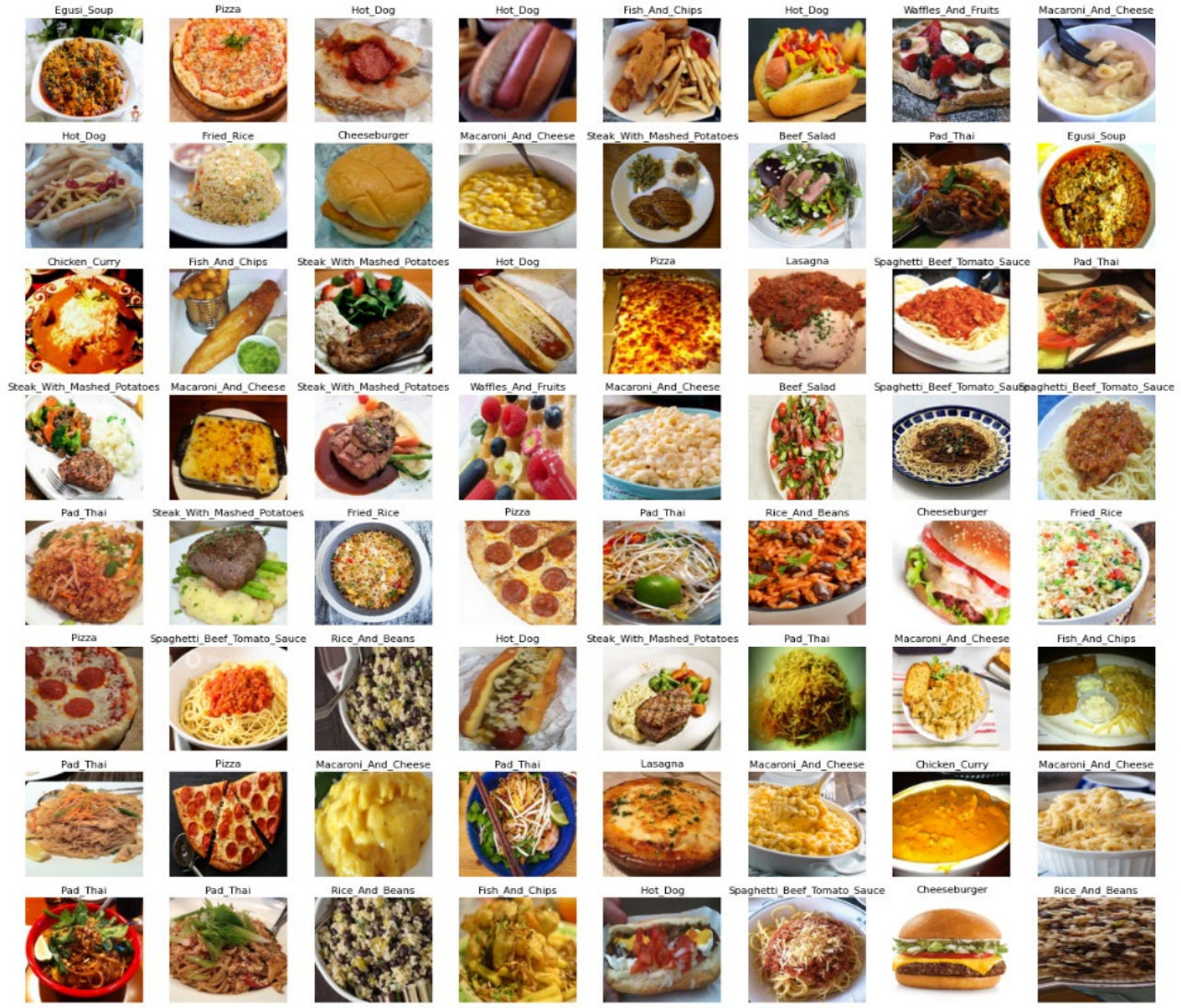


Figure 5.2: Examples of images in the Food15 dataset

Feature scaling

The images in the dataset were scaled to be homogeneous i.e. scaled to take small values in order to improve the numerical operations and performance of the optimization function during the training of the neural network. For the feature scaling technique called the *min-max scaling* or *normalization* (computed using Equation 5.1), the features (image pixels) were casted to *float32* and then scaled from range of $[0 - 255]$ to have a range of [0 and 1]. This was done by subtracting the min value and dividing by the difference between the max and min values. The image normalization was carried out at the point of feeding the images to the network algorithm.

$$z = \frac{x - min\ (x)}{max(x) - min\ (x)} \tag{5.2}$$

Where $z\ is\ the$ normalized value for pixel value $\boldsymbol{x}$, $min(x)\ and\ max(x)$ are the minimum and maximum values in $\boldsymbol{x}$ given its range.

5.2.2. *Experimental setup: system architecture and model development*

5.2.2.1 Model training and optimization experiments

The successful application of deep learning to computer vision task contributed to the development and popularity of the convolutional neural networks (CNNs). CNNs models have outperformed the conventional food image recognition approaches which involve extraction of visual features from food images using several combinations of specialized traditional computer vision algorithms such as Gabor filters, Scale Invariant feature Transform (SIFT), Color Histograms, Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG), and several others. With deeper and deeper architectures as well as through transfer learning approach, CNNs have achieved higher and reliable accuracies far better than the conventional approaches.

The process of building a deep neural network model (*See Appendix D)* involved first choosing an appropriate network architecture and then training the network on the available training set. In this study, deep convolutional neural network was employed and repurposed through the transfer learning approach.

The training process involved estimating the weight parameters of the network by solving a non-convex optimization algorithm problem. The purpose of the network optimization algorithms was to continuously update the weights of the network as a function of the errors made by the model during training until it attained convergence. In this study, three standard optimization algorithms namely: Stochastic Gradient Descent (SGD), RMSprop and Adam(*See Appendix C)* were used and compared to solve the optimization problem in order to come up with the model that best represents the ***Food15*** training set and as well generalize to never-before-seen test set.

Features extraction, model fine-tuning and model selection

Due to the *low-to-medium* size of the Food15 dataset and also to avoid undue demand for computing resources and training time required to train a ConvNet from scratch the popular state-of-the-art *Transfer Learning* approach was used to develop accurate deep convolutional neural network model in time-saving manner. Concretely, five different pre-trained networks namely: VGG16, VGG19, ResNet101V2, InceptionV3, and Xception (*see Appendix C*) were repurposed and trained on the 70% *Food15* training set using the *Fine-tuning* approach and their performance were evaluated on the 15% validation set.

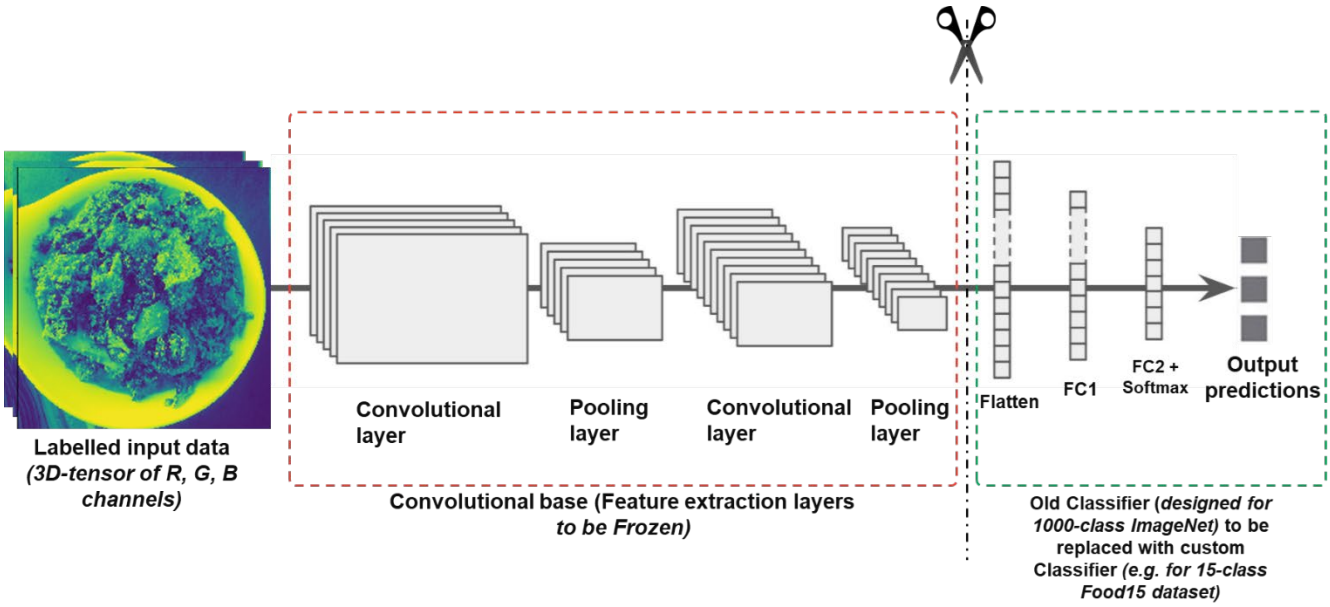


Figure 5.3: Transfer learning: Repurposing pretrained ConvNets

Each of the pre-trained models was made up of a convolutional base (feature extraction layer) and a customized densely connected classification layer trained to classify the 1000 classes in the ImageNet dataset. The convolutional bases as illustrated in Figure 5.3 above, consisted of several sequence of parameterized convolutional and pooling layers arranged in different structures based on the model's architecture. Earlier layers, i.e. layers deep down in the convolutional bases of the models have already been encoded with highly-generic and reusable features or ***feature maps*** such as visual edges, textures and colors, etc., whereas layers at the top of the convolutional bases were encoded with more-specialized features containing patches such as those of "mashed potatoes", "spaceship", "dog eye", etc. based on the ImageNet database on which the models were trained. Since *Food15* has only 15 classes of food items, the first step of the fine-tuning approach was to remove the classification layer on top of each of the five pre-trained models and then replaced them with a custom-built classifier consisting of a 512-hidden units dense or fully-connected layer with a **ReLU** activation function as illustrated in Figure 5.4, followed by a 50% ***dropout*** layer as well as the dense output layer with 15 hidden units (representing the number of classes) and a ***softmax*** activation function. Intuitively, it was more useful to consider beginning fine-tuning from the higher-up layers with the more-specialized features and not the deeper layers because they are flexible and could easily be repurposed for classifying the Food15 dataset. In the second step, the convolutional base was ***frozen*** for each network in order to train the randomly initialized weights in the top custom-built classifier that was newly added. This was done to prevent the process of training the fresh classifier from propagating large error gradients through the entire network which could destroy the previously learned or encoded feature representations in the layers of the convolutional base. In the steps that followed, the fresh classifier for each of the model was mildly and separately trained using the ***RMSprop***

$(learning\_rate: 1e-3)$ ***Adam*** $(learning\_rate: 1e-3)$ and ***SGD*** $(learning\_rate: 1e-2)$ and for **10** ***epochs***. After this, the top layer was considered to have been trained and harmless to the convolutional base. Next, some of the top layers in the convolutional base of the models that were encoded with features more specialized to the ImageNet dataset were strategically ***unfrozen*** in order for them to re-adjust to or learn features more relevant to the ***Food15*** dataset. It would be counter-productive to unfreeze the entire convolutional base of the pre-trained models because of the massive computing resources and long duration of time that would be required to train the loads of parameters of the network from scratch and the high risk of overfitting. Finally, the unfrozen layers and the added dense classifier were jointly re-trained on the training set for **30** more ***epochs*** and at very low $(learning\_rate: 1e-4\ to\ 1e-6)$ to further improve or ***fine-tune*** the performance of the model. A low learning rate was used in order to minimize the magnitude of the updates made to the features being adjusted. The entire process was repeated for each of the model using the three different optimization algorithms, their performances were evaluated on the validation set, and the top 3 performing models were selected for further analysis.

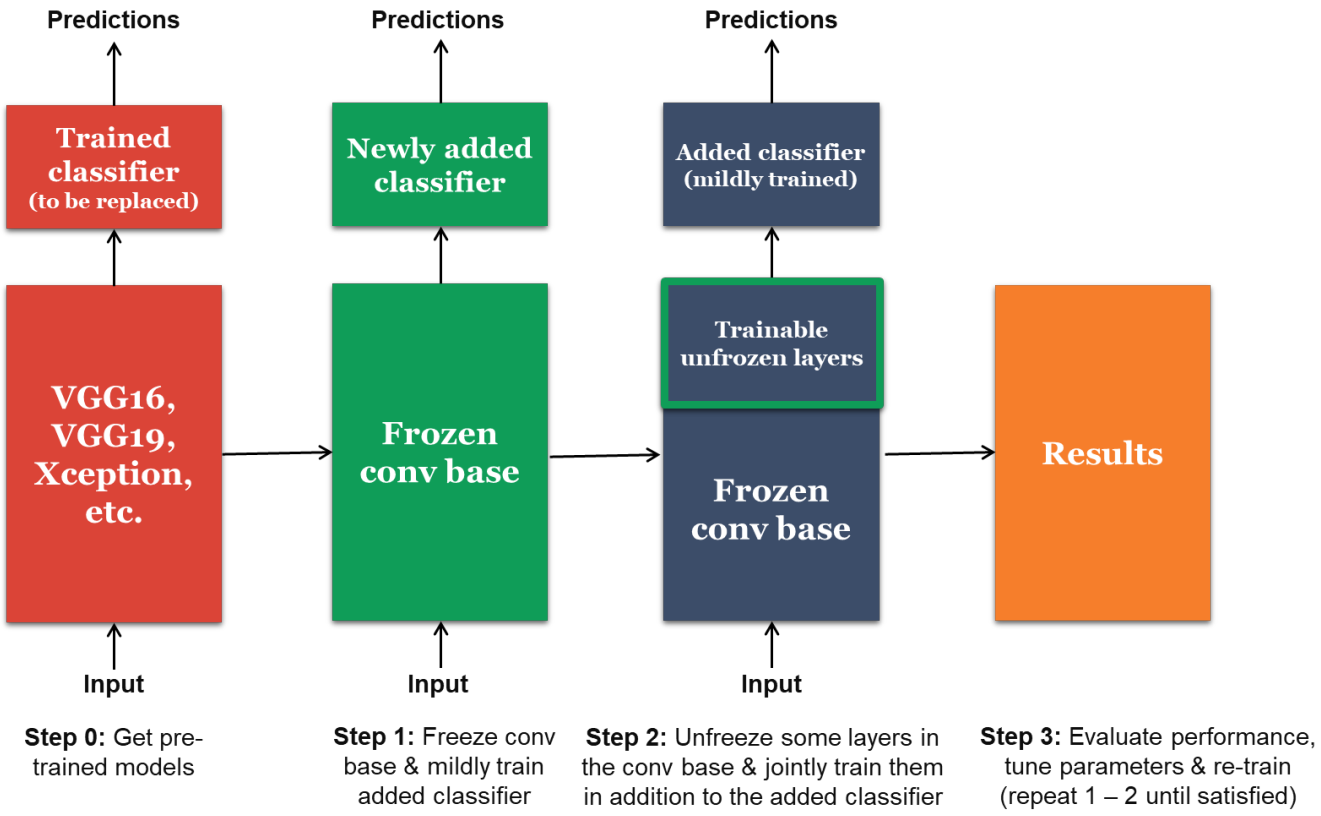


Figure 5.4: Model Fine-tuning

5.2.2.2 **Model regularization: reducing overfitting and maximizing generalization**

During the training phase, millions of parameters were generated by each of the five architectures. Learning such huge number of parameters introduced high risk of the overfitting as the performance of the model on the validation set began to decline at accuracy range of 88% which implied that the models had begun to learn patterns that could be misleading or irrelevant for generalization. The overfitting was mitigated by implementing the following Regularization techniques (*See Appendix C* for detailed description of the regularization techniques).

Data augmentation

Data augmentation is an easy and very popular method to mitigate overfitting. It involves artificially increasing the size of the training set in order to improve the performance of the model. Two distinct forms of data augmentation were employed, both of which helped transform the original training images with very minimal computation. The first form of data augmentation involved generating new images by performing geometric transformation methods on the training set as follows: i) *Image rotation in the range* $[-\mathbf{30}°, +\mathbf{30}°]$, and ii) *Zooming factor of* $\mathbf{0.2}$. The images generated were saved directly to disk thereby increasing the training set size by 44% (from 9000 to 16000), a value similar to that achieved in *Chen et al. (2019)* but less than what was achieved in *Krizhevsky et al. (2012)*. The second form of data augmentation involved further transforming the images during the model training process i.e. the images were transformed as they were being served in batches to the network. The transformation techniques adopted include: i) *Image rotation in the range* $[-\mathbf{40}°, +\mathbf{40}°]$, ii) W*idth and height shifting range* $(of\ \mathbf{0.1}\ of\ the\ total\ width\ and\ height\ of\ the\ image)$, iii) *Zooming factor of* $\mathbf{0.2}$, iv) *Random horizontal flipping*. The transformed images were generated and queued temporarily on the CPU while the Graphical Processing Unit (GPU) is training on the previous batch of images. The augmentation not only increased the size of the training set, it also further improved the robustness of the model by making it invariant to rotation, translation, and viewpoint.

Dropout and weight regularization

A *dropout rate* of range $[\mathbf{0.1 - 0.5}]$ and $\boldsymbol{\ell_2}$ **norm** (or weight decay) parameter constraint of $[\mathbf{0.1}, \mathbf{0.01}, \boldsymbol{and}\ \mathbf{0.001}]$ were applied to the added classifier and experimented on individually and in combination during the training process.

5.2.2.3 **Summary of learning: stochastic gradient descent with momentum**

Following the choice of Stochastic Gradient Descent as the main optimization algorithm, and after several experimentation and hyper-parameter tuning as described above, the summary of the training process is

represented by Equation 5.2 below. The parameters used included a batch size of $\mathbf{10}$ examples, momentum of $\mathbf{0.9}$, learning rate of $\mathbf{0.001}$ and weight decay of $\mathbf{0.001}$ which also served as a regularization term to help reduce training error. The momentum term as the name implied was added to the gradient descent algorithm in order to speedup or introduce ***velocity*** in the algorithm. The basic idea behind the role of the momentum was to compute an exponentially weighted average of the gradients at each iteration, $\boldsymbol{t}$, and then use the average to update the weights instead of updating it with the classical gradient descent weight update method at the end of each iteration. As shown in the contour plots in Figure 5.5, the black arrow indicated the large up and down oscillations of the classical gradients which tend to prolong the rate of convergence (or learning) while the red path (towards the center of the contour or global minimum) indicated the faster and quicker gradients steps (faster learning) as a result of the momentum. Hence, the gradient descent update rule (with momentum) for the weight $\boldsymbol{W}$ at $\boldsymbol{t}$ iteration and on the current mini-batch was:

$$
\begin{gathered}
v_{\partial W} := \beta v_{\partial W} + (1-\beta)\partial W \\
v_{\partial W} := 0.9 v_{\partial W} + (1-0.9)\partial W \\
v_{\partial W} := 0.9 v_{\partial W} + 0.1\partial W \\
W := W - \alpha v_{\partial W} \quad (instead\ of\ W := W - \alpha \partial W) \\
W := W - 0.001(0.9 v_{\partial W} + 0.1\partial W) \\
W := W - 0.009 v_{\partial W} - 0.0001\partial W
\end{gathered}
\tag{5.3}
$$

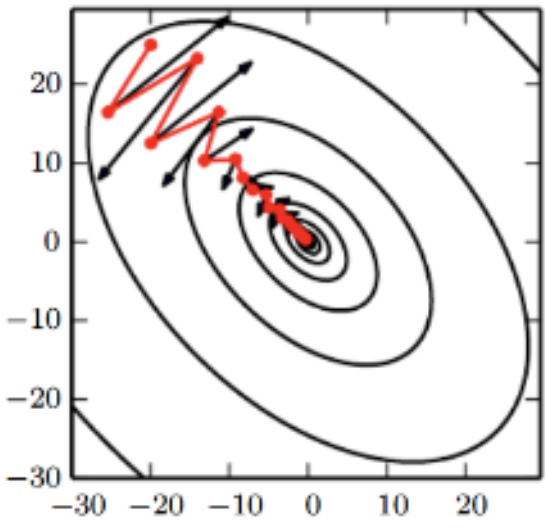


Figure 5.5: Contour plots of SGD with momentum

Where

$v_{\partial W}$ = momentum variable, $\boldsymbol{\alpha}$ is the learning rate, $\boldsymbol{\beta}$ is the momentum term, and $\boldsymbol{\partial W}$ is the average over the last iteration gradients.

The models were trained for 50 $epochs$ each over the **16 035** training images on **Google Colab** equipped with $Tesla\ K80\ GPU$ and 12 $GB\ of\ RAM$.

Figure 5.6 shows an illustration of the **64** activation maps of size **109 × 109**, the feature representations learned by the **4th convolutional layer** of the **Xception model** from the sample input image.

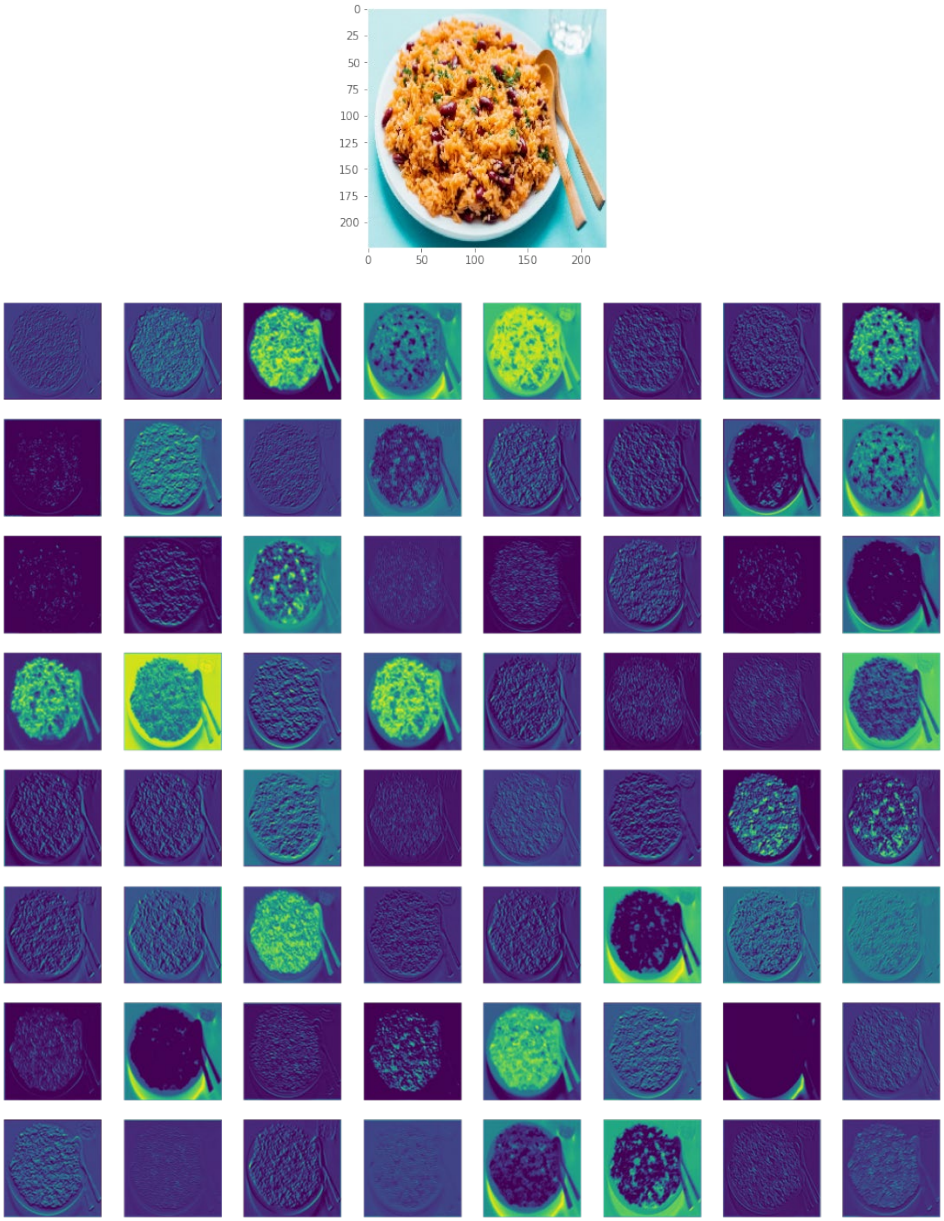


Figure 5.6: Activation maps (each image size, 109 ×109; total number, 64) of the 4th convolutional layer of the Xception model used as sample test images

5.2.2.4 Metrics for performance evaluation of recognition model

In this study, the performances of the models were measured by monitoring the rate at which the models correctly predicted the class of each food image in the test set using the following metrics.

Classification accuracy

This represented the number of correctly classified food images made as a ratio of all food images classification made. It was computed using Equation 5.3.

$$\frac{TP + TN}{TP + TN + FP + FN} \tag{5.4}$$

Where, TP (True positive) and TN (True negative) are the number of food images that the models correctly classified into their actual classes, FP (False positive) and FN (False Negative) represent the number of food images misclassified by the models.

A variant of classification accuracy metric called ***Top-K-Categorical-Accuracy*** was also used. Using the Top-K-Categorical-Accuracy, the model was considered to have correctly predicted a test food image if the predicted probability of that food image falls within the Top-K of all predicted probabilities for that class.

Confusion matrix and error matrix

The confusion matrix was used to plot the performance of the models on the test set by matching the classes of the predictions with their actual classes, making it easy to identify misclassified items. The diagonal elements of the confusion matrix showed the fractions of food images that were correctly predicted while the off-diagonal elements represented misclassified food images. In a similar context, the error matrix was used to indicate the error or uncertainty in a reported measurement.

5.2.3. *Food portion size and nutrient estimation*

The 15 composite foods (*Beef_salad', 'Cheeseburger', 'Chicken_curry', 'Egusi_Soup', 'Fish_and_chips', 'Fried_rice', 'Hot_dog', 'Lasagna', 'Macaroni_&_cheese', 'Pad_thai', 'Pizza', 'Rice_and_beans', 'Spaghetti_beef_tomato-sauce', 'Steak_with_mashed_potatoes'*, and *'Waffles_and_fruits'*) considered in this study cut-across cuisines from different regions of the world including ***Asia, West Africa, and North America.*** The serving size weights, and nutrients composition of the food items considered in this study were obtained from the Food and Nutrient Database for Dietary Studies (FNDDS). The database also consisted of known portion size and weights of commonly consumed foods and beverages as reported in the What We Eat in America (WWEIA) database.

The FNDDS contains over 8000 food and beverages, 65 nutrient components for each of the food and over 30000 typical portion weights. For instance, according to the database, a typical portion weight of the single serving size of a home-made beef salad is 182 g. The portion weight of the food can then be varied to account for larger food portions or multiple serving based on the quantity consumed by multiplying the portion weight by a factor e.g. 1.5, 2, 3 etc. The serving size weight and the nutrient values obtained for each of the food class (as seen in Table 5.1) were further used as part of the input data for the SAIN-LIM nutrient profiling model.

Table 5.1: Nutrient information of the 15 foods retrieved from the FNDDS

	Portion description	Estimated portion weight	Energy	Protein	Fibre	Calcium, Ca	Iron, Fe	Ascorbic acid, Vit. C	Vitamin D	Sodium, Na	Saturated Fatty Acid	Added Sugar
				65	**25**	**900**	**12.5**	**110**	**5**	**3153**	**22**	**50**
Food item		**(g)**	**(kcal)**	**(g)**	**(g)**	**(mg)**	**(mg)**	**(mg)**	**(µg)**	**(mg)**	**(g)**	**(g)**
Beef salad	1 serving	182	473.2	32.58	0.73	34.58	2.95	1.456	0.182	609.7	7.01	3.82
Cheeseburger	2 serving	400	1232	66.04	8	476	10.84	0.6	0.4	2060	25.89	19.52
Chicken curry	1 serving	233	191.06	13.35	3.50	74.56	1.72	20.97	0.47	0	1.56	3.43
Egusi soup	1 serving	130	543.4	6.29	0.98	137.45	1.53	0	0	19.071	3.80	0
Fish and chips	1 serving	250	1087.5	51.00	5.50	122.50	4.38	23.50	13	2145.00	26.36	5.23
Fried rice with shrimps	1 serving	130	215.57	8.91	1.17	31.12	0.79	4.15	0	448.84	0.69	0.62
Hot dog	2 serving	114	332.88	11.66	0	127.68	1.25	21.546	0.798	1112.64	9.10	3.31
Lasagna	1 serving	375	742.5	53.10	6	1083	4.13	3	0.75	1515	19.95	7.23
Mac & Cheese	1 serving	120	265.04	10.28	1.44	202.96	1.27	0	0.72	436.70	6.31	3.03
Pad Thai	1 serving	250	362.5	17.00	3.00	83	1.45	12.25	0.5	875	2.81	10.70
Pizza	3 standard slices	321	850.65	32.68	5.78	426.93	7.67	8.67	0	1335.36	19.87	17.05
Rice and beans	1 serving	150	185.77	6.04	5.40	51.03	2.42	5.10	0	265.48	0.82	1.56
Spaghetti with beef & tomato-sauce	1 serving	260	218.40	15.39	3.64	59.80	2.81	28.60	0	899.60	3.52	9.26
Steak with mashed potatoes	1 serving	320	491.35	44.85	2.01	47.08	5.50	0	0.27	1129.89	8.08	0.84
Waffles and fruits	1 serving	180	630	11.86	4.86	324	2.45	1	1.44	1112.40	7.45	15.77

5.2.4. *Nutrient profiling*

5.2.4.1 The SAIN, LIM model

The portion size and nutrient content values retrieved from the FNDDS for each recognized food were used by the SAIN, LIM model to perform nutrient scoring and profiling analysis. The SAIN and the LIM nutrient scoring models were used to classify the foods based on their degree of **healthiness** and **unhealthiness** and then categorized them into one of the four different classes as follows:

i. Foods recommended to be consumed and that supports good health ($SAIN > 5 \; and \; LIM < 7.5$);
ii. Food with neutral or balanced health benefits ($SAIN < 5 \; and \; LIM < 7.5$);
iii. Foods recommended in less quantities or to be consumed occasionally ($SAIN > 5 \; and \; LIM > 7.5$); and
iv. Foods that their consumption should be limited ($SAIN < 5 \; and \; LIM > 7.5$).

The SAIN score (computed using Equation 5.4), computed for 100 kcal of food, is an un-weighted arithmetic mean of the percentage adequacy for five qualifying nutrients (plus 1 optional nutrient) in the food composition tables and for which a daily recommended value (DRV) existed. The nutrients needed for computing SAIN score include Protein, Calcium, Iron, fiber, and Vitamin C, were obtained from the FNDDS.

$$\textbf{SAIN}\ Score = \left(\frac{\left(\frac{\left(\frac{Protein}{65} + \frac{Fibre}{25} + \frac{Vit\ C}{110} + \frac{\frac{Ca}{900}}{5} + \frac{Fe}{12.5} + \frac{Vit\ D}{5} - \min\ ratio \right)}{5} \times 100 \right)}{E} \right) \times 100 \qquad (5.5)$$

Where:

Protein = protein content in $g/100\ g$;

Fibre = fibre content in $g/100\ g$;

Vit C = vitamin C content in $mg/100\ g$;

Ca = calcium content in $mg/100\ g$;

Fe = Iron content in $mg/100\ g$;

Vit D = vitamin D content in $\mu g/100\ g$

E = energy density in kcal/edible $100\ g$;

Minimum ratio (min ratio) = the lowest of the 6 [nutrient/DRV] ratios.

The LIM score (computed using Equation 5.5) is the mean of the percentages by which a particular food exceeds the recommended nutritional value for each of the nutrients present in the food, namely: Sodium, added sugars, and saturated fatty acids (SFA) and it is expressed per 100g of cooked or rehydrated food.

$$\boldsymbol{LIM\ Score} = \left(\frac{\frac{Na}{3153} + \frac{SFA}{22} + \frac{Added\ Sugar}{50}}{3} \right) \times 100 \tag{5.6}$$

For analysis, tested foods were considered to have healthy or unhealthy profile based on their SAIN and LIM score. The higher the SAIN score in relation to the LIM score indicated that the food contained more of qualifying nutrients (Protein, Calcium, Iron, fiber, and Vitamin C) than the disqualifying ones and hence, the healthier the food is. Conversely, the higher the LIM score relative to the SAIN score indicated that the food consisted of more disqualifying nutrients (e.g. Saturated fatty acid, etc.) and hence the food was termed as threatening to the health of the consumer and its consumption should be reduced. In addition, ***Nutrient Score,*** a ratio of the SAIN score to the LIM score was further used to convey the healthiness and unhealthiness relationship of a food**.** The nutrient score (Equation 5.6) held comprehensible information about the healthiness of a food based on the computed SAIN and LIM scores. The higher the nutrient score, the healthier the food becomes and vice versa.

$$\boldsymbol{Nutrient\ Score} = \frac{\boldsymbol{SAIN\ Score}}{\boldsymbol{LIM\ Score}} \tag{5.7}$$

5.2.5. *Codes and experimental environment*

The analysis in this study were conducted using Tensorflow 2.1.0 *(Abadi et al., 2016)* deep learning framework which comes with Keras 2.3.0 *(Chollet, 2015)* high-level API integration and deployed in Python 3.6.9 *(Van Rossum, 2007)* programming language. The robustness of the framework made it easy to process the images, construct deep convolutional neural networks, acquire pretrained model's weight, as well as evaluate the developed models used in this research. Majority of the computation was done on Google Colab *(Bisong, 2019)* equipped with $Tesla\ K80\ GPU$ and 12 $GB\ of\ RAM$ and with easy access to the above frameworks. Some of the computations were also done within Jupyter Lab *(Pérez, 2014)* running on an Anaconda *(Anaconda, 2018)* virtual environment and on a Windows 10 OS MSI machine equipped with Intel Core i7 7th generation CPU, Geforce GTX1050 NVIDIA graphic card and 16 GB of RAM as well as the above frameworks. PyQt5 designer *(Willman, 2020)* was used to design and develop an interactive graphical user interface (GUI) application for the system to facilitate easy understanding and usage of the system. The application as illustrated in Figure 5.15 has four display windows as follows: *(i)* Query or test image window which displays a preview of the test image, *(ii)* Prediction image which displays the Top-5 predictions of the test image, *(iii)* Food nutrients analysis and scoring window which displays the weight, nutrients composition, SAIN, LIM, and nutrient scores of the food, and *(iv)* Nutrient distribution window which displays the distribution or profile of the nutrients composition by their individual percentage.

5.3. RESULTS AND DISCUSSIONS

5.3.1. *Model training, selection and optimization results*

In this study, the best performing optimization function and deep learning model were investigated. Three optimization functions (Stochastic Gradient Descent (SGD), RMSprop and Adam) were employed to train five pre-trained networks (VGG16, VGG19, ResNet101V2, InceptionV3, and Xception). Altogether, 15 deep learning models with different combinations of hyperparameters were investigated in order to select the best performing model, optimizer, and set of hyperparameters. All the five models were trained using the three optimization algorithms. Generally, the three optimizers performed well for the training process. However, with a more critical focus on evaluating the models in terms of the level of convergence and generalization, it was discovered that the Stochastic gradient descent (SGD: $(learning\_rate: 1e-5, momentum: 0.9)$) optimizer performed best for both the training and validation followed by the Adam ($learning\_rate: 1e-5$) optimizer as evidenced in their respective accuracies during the process of training the five models considered (see Table 5.2). In particular, the stochastic gradient descent (SGD) optimizer did a very neat job at fitting the data to the models and with the least validation error which resulted in Xception being the best performing model. Although the RMSprop optimizer had similar performance, its performance was not as good as the other optimizers regardless of the model it was used to train. Table 5.2 presents the empirical training and validation accuracies of the three algorithms on the five models experimented on. In addition, the trends of the training and validation losses over the **40** ***epochs*** for the best performing model (Xception) were also monitored as shown in Figure 5.7 (a – *d*). The results obtained and trends of the losses showed that SGD generally converged better to global optimum as well as have better ability to generalize regardless of the model it was used to train. This was also evident in the study carried out by *Krizhevsky et al. (2012)* and *Wu et al. (2019).*

Table 5.2: Empirical training and validation accuracies of the three algorithms on the five models

Optimizers	**Adam**		**RMSprop**		**SGD**	
Pre-trained models	**Train Acc (%)**	**Val Acc (%)**	**Train Acc (%)**	**Val Acc (%)**	**Train Acc (%)**	**Val Acc (%)**
VGG16	99.38	94.35	97.90	92.95	97.72	93.85
VGG19	99.48	94.35	98.26	92.50	98.95	93.950
Xception	95.71	90.25	88.05	81.50	99.07	94.40
InceptionV3	88.58	84.15	90.83	83.80	98.57	88.05
ResNet101V2	92.98	88.10	93.85	87.80	96.74	88.15

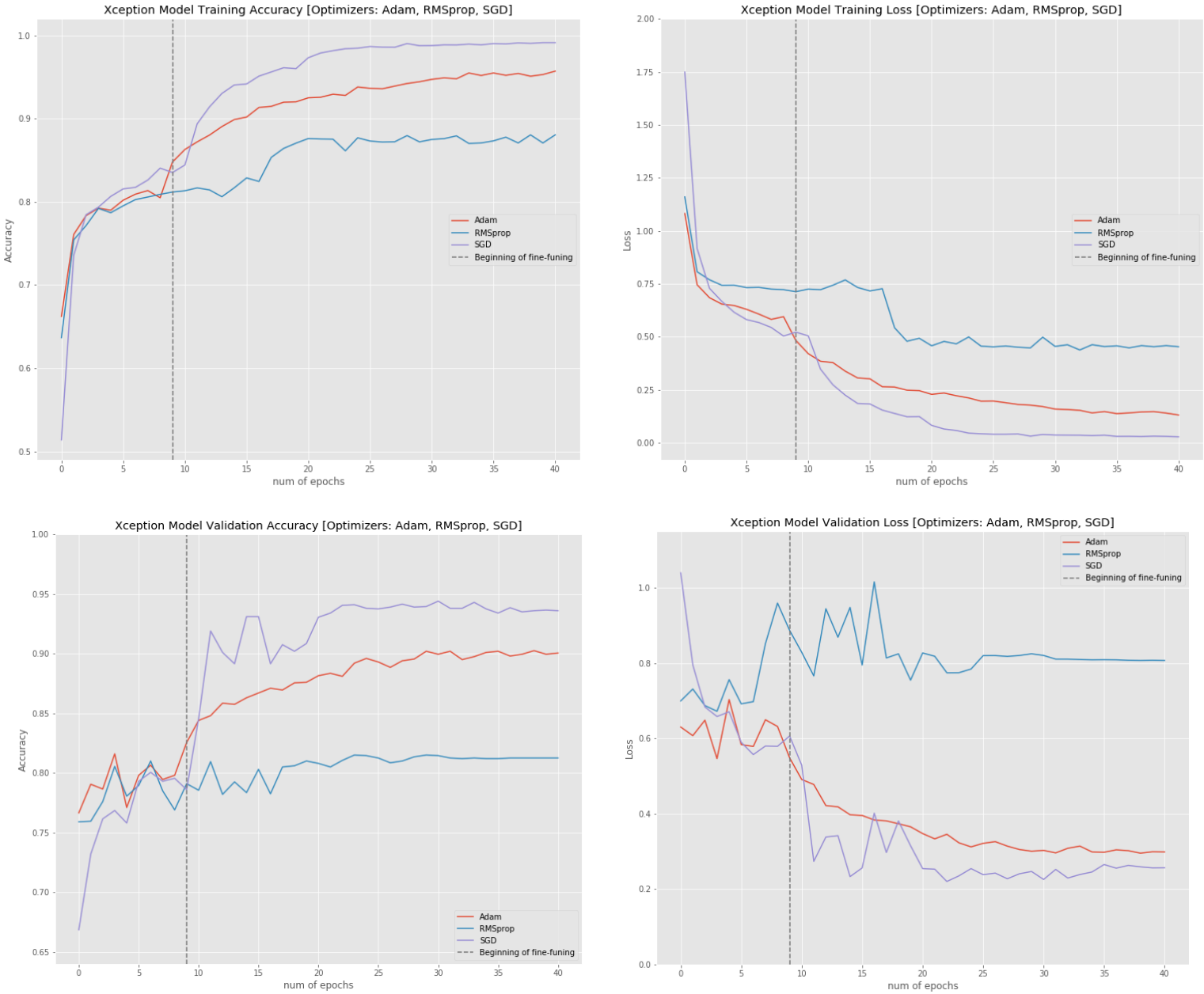


Figure 5.7: Trends of the training and validation losses of the Xception model over the three Optimization algorithms

Further investigation into the training process of the five models showed that the models began to overfit the training data at about $\mathbf{20-25}$ **epochs**. This implied that the models' performance on validation set started declining or getting worse compared to their performance on the training set which consistently improved as training progressed. This could have great effect on the generalization performance of the models on the *never-before-seen data* (test set). The result of the steps carried out towards mitigating overfitting and maximizing generalization are presented next.

5.3.2. *Model regularization result: reducing overfitting and maximizing generalization*

The results of the regularization techniques carried out on the model are given below.

5.3.2.1 Data augmentation

The augmentation strategy adopted resulted in an increase in the size of the training set as presented in Table 5.3.

Table 5.3: Reducing overfitting through Data Augmentation

Data	Train set	Validation set	Test set	Total
Before Augmentation [70: 15: 15]	~ 9000	2000	2000	13000
After Augmentation [80: 10: 10]	~ 16000	2000	2000	20000

5.3.2.2 Dropout and weight regularization

It was observed that applying both the dropout and the $\boldsymbol{\ell_2}$ **norm** (weight decay) individually did not lead to any significant improvement in the result, however, when both techniques are used in combination, it led to a further increase in the validation accuracies.

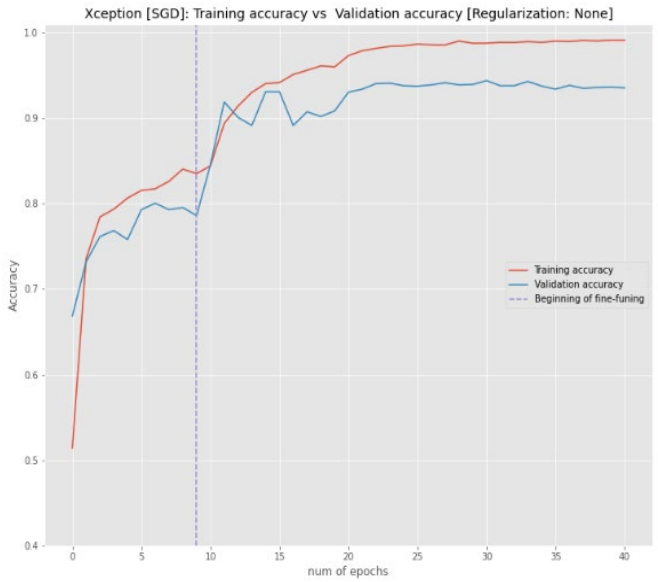
Several combinations and parameter tuning results, underlining the best strategy to reduce overfitting the data, showed that data augmentation, a dropout rate of $\mathbf{0.2}$ and weight decay of $\mathbf{0.001}$ were the best supporting training parameters for the SGD optimizer and a dropout rate of $\mathbf{0.5}$ and weight decay of $\mathbf{0.001}$ for the Adam optimizer. These parameters were used to further re-train and validate the top three models repeatedly for five times in order to get reliable statistics.

Table 5.4 presented below shows the results of comparison between non-regularized top three performing models namely: VGG16, VGG19, and Xception with their corresponding regularized artifacts in term of data augmentation, dropout, and weight decaying, over the Adam and the Stochastic Gradient Descent optimization algorithms. Figure 5.8 (a - f) shows graphical representation of the trends of how the

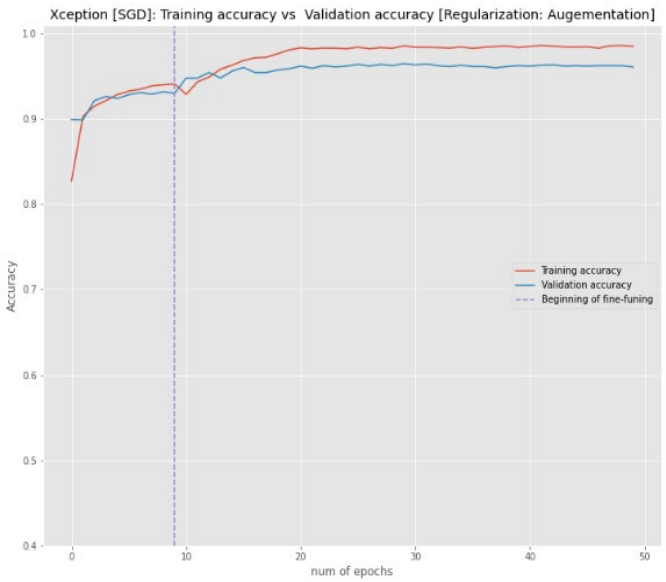
regularization strategies improved the performance of the Xception model when trained with the SGD algorithm. The results implied that the model regularization strategies were effective in improving the performance of the classification models.

Table 5.4: Results of Regularization Strategies on the models

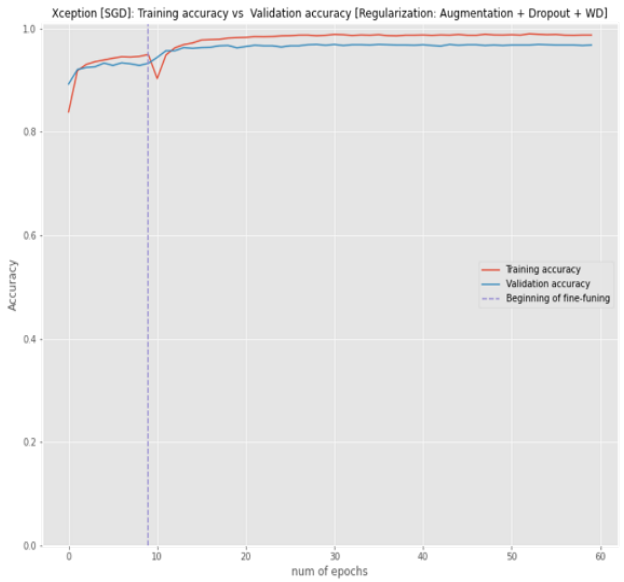
Optimizers	**Adam**						**SGD**					
Pre-trained models	**No regularization**		**Augmentation**		**Aug + Dropout + Weight Decay (WD)**		**No regularization**		**Augmentation**		**Aug + Dropout + Weight Decay (WD)**	
	Train Acc (%):	**Val Acc (%):**	**Train Acc (%):**	**Val Acc (%):**	**Train Acc (%):**	**Val Acc (%):**	**Train Acc (%):**	**Val Acc (%):**	**Train Acc (%):**	**Val Acc (%):**	**Train Acc (%):**	**Val Acc (%):**
VGG16	99.38	94.35	99.42	95.25	99.62	95.00	97.72	93.85	97.82	93.80	97.15	94.45
VGG19	99.48	94.35	99.15	94.30	99.26	93.90	98.95	93.95	99.53	94.95	96.61	94.40
Xception	95.71	90.25	99.37	95.65	99.84	97.70	99.07	94.40	98.52	96.35	98.55	96.65



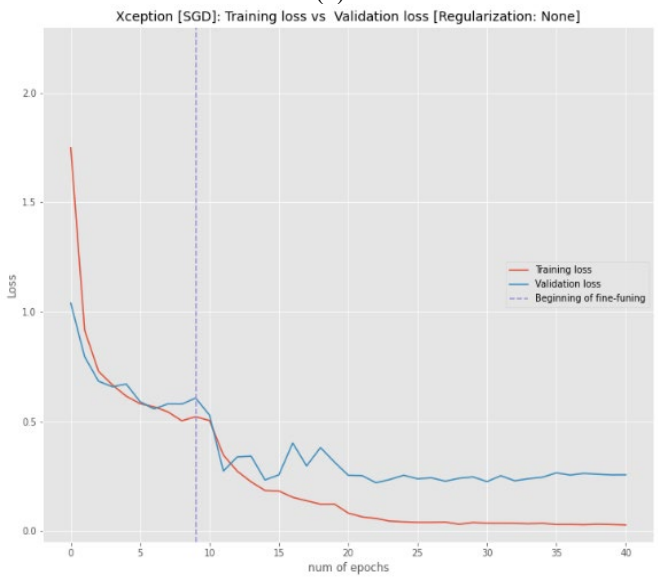
(a)



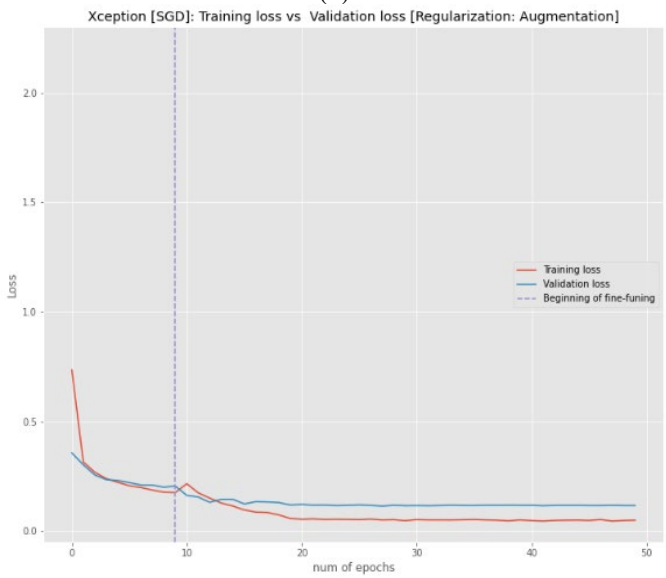
(b)



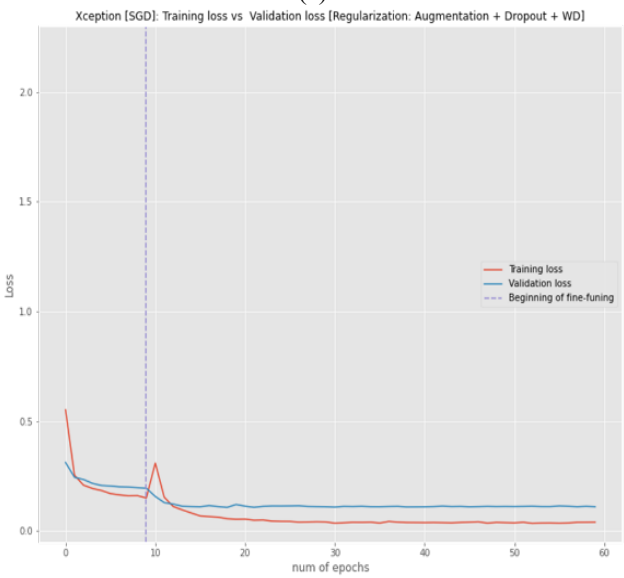
(c)



(d)



(e)



(f)

Figure 5.8: ***(a – f)****:* Empirical illustrations of how regularization strategies improved the performance of the Xception model using SGD

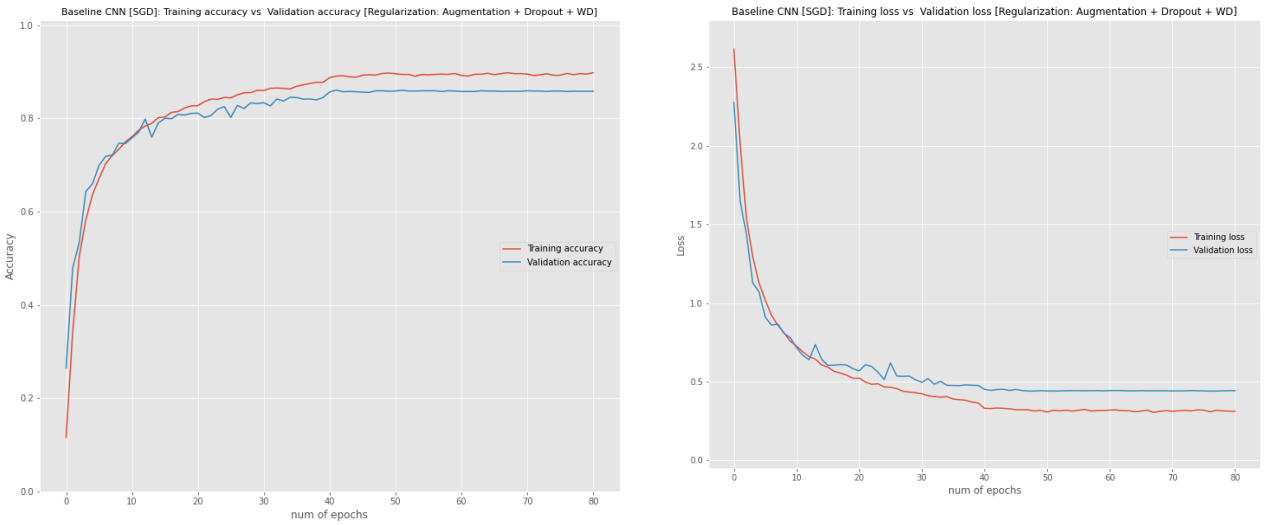
5.3.3. *Measure of generalization*

5.3.3.1 Classification accuracies

Table 5.5 represents the average validation and test accuracies obtained over the five repetitions. It compares the average performances in term of Top-1 and Top-3 validation and test accuracies of the three models alongside a baseline Vanilla CNN on the validation and test set. It was observed that the Top-1 and Top-3 of the three models are well above 90% with Xception model performing impressively on the test set with 98.10% and 99.80% on the Top-1 and Top-3 test accuracies respectively. Similar result were also obtained in other works *(Attokaren et al., 2017; Bootkrajang et al., 2020)*. These authors also observed that deep convolutional neural networks are more appropriate for image classification compared to traditional features extraction and classification algorithms. The 5-layer CNN baseline model trained from scratch, was validated and tested on the Food15 dataset for about 80 epochs. As shown in Figure 5.9, it took about **12 hours** for the 5-layer network to achieve a Top-1 accuracy of 85.90% using the SGD with momentum optimization algorithm before the training was stopped due to overfitting, compared to the pre-trained models which took an average of $3-4$ hours training time. To achieve accuracies similar to those of the pre-trained models, the baseline model would need huge (hundreds of thousands or millions) of images and much more layers, which however will require huge computation power and longer training time (days).

Table 5.5: Comparison of performances in term of Top-1 & Top-3 validation and test accuracies of the 3 models alongside a baseline CNN

Optimizers	**Adam**				**SGD**			
Pre-trained models	**Validation acc (%)**		**Test acc (%)**		**Validation acc (%)**		**Test acc (%)**	
	Top-1	**Top-3**	**Top-1**	**Top-3**	**Top-1**	**Top-3**	**Top-1**	**Top-3**
VGG16	95	99	97.70	99.69	94.45	99.35	96.75	99.90
VGG19	93.90	99.35	97	99.70	94.40	99.15	97.20	99.85
Xception	97.35	99.85	98.10	99.85	96.90	99.65	98.10	99.80
Baseline (Vanilla) CNN	81.35	94.80	92.55	98.20	85.90	96.95	93.85	98.60



(a) (b)

Figure 5.9: (a) Baseline CNN Training and Validation accuracy and, (b) Baseline CNN Training and Validation loss

5.3.3.2 Confusion matrix and error matrix

Figure 5.10 represents a confusion matrix of the performance of the Xception model on the test set. From the matrix, it can be deduced that the Xception model was able to correctly predict a large portion of the test set. However, few challenges exist where the model misclassified some food items such as Rice and beans as Fried rice, or Pad Thai as Beef salad as shown in Figure 5.11. This was as a result of the similar visual properties such as color and texture associated with the two foods as the food items can almost equally be misclassified by human observation. Figure 5.12 shows an Error matrix of all misclassified food items.

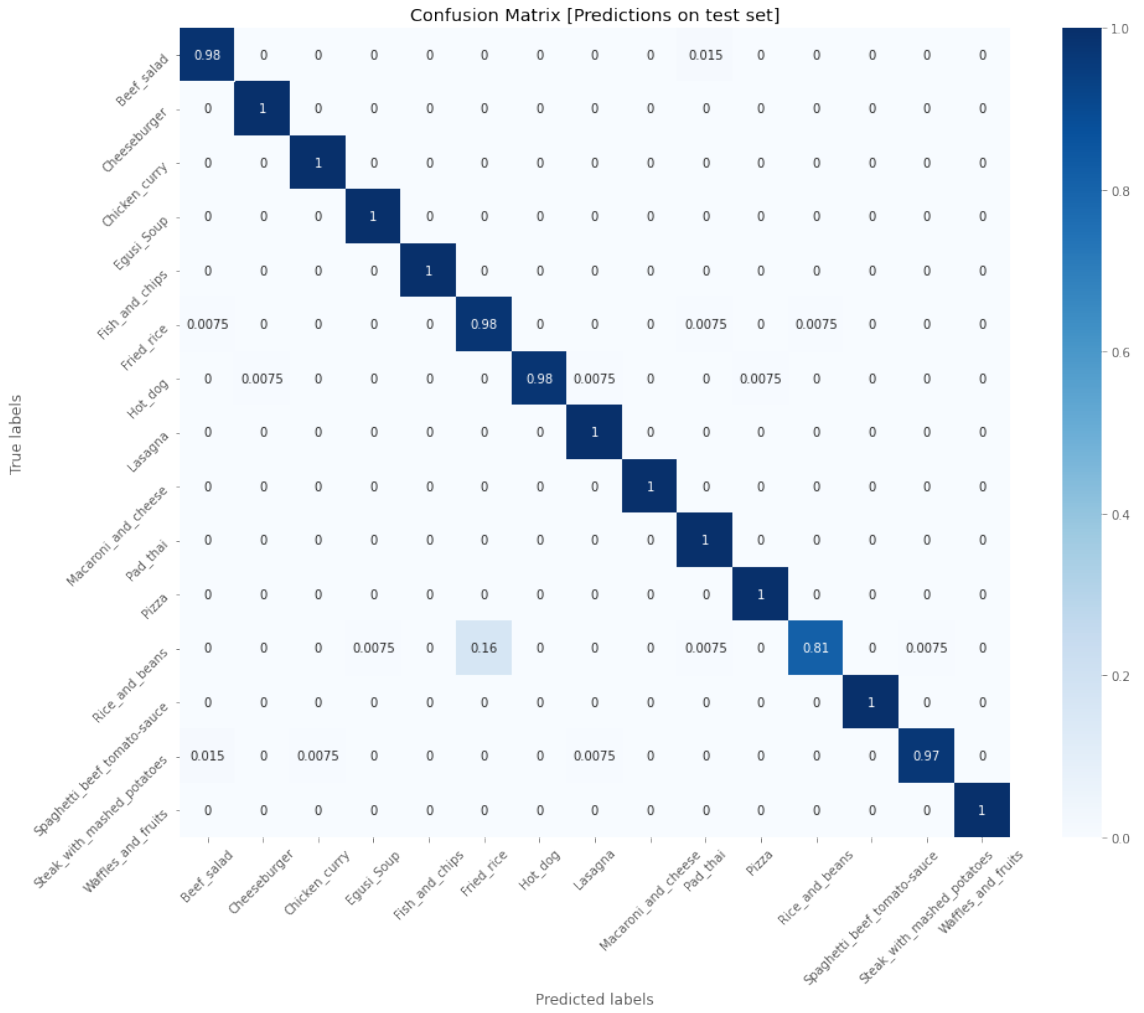


Figure 5.10: Confusion matrix of the performance of the Xception model on the test set.

(a) **Top:** Fried rice, **Bottom**: Pad Thai (b) **Top:** Rice and beans, **Bottom**: Beef salad

Figure 5.11: Samples of misclassified food items (Fried rice misclassified as Rice & beans; Pad Thai misclassified as Beef Salad)

Figure 5.12: Error plot of misclassified food items

In summary, the results of several combinations and experimentations involving fine-tuning of the parameters and hyperparameters of the five pre-trained models and the baseline convolutional neural networks are presented. It can be inferred that out of the five models (VGG16, VGG19, Xception, InceptionV3, and ResNet101V2) considered in this study, VGG16, VGG19, and Xception performed generally well in resulting a good representation of the dataset (Table 5.4). Furthermore, Adam and Stochastic Gradient Descent (SGD) with momentum out of the three optimization algorithms considered were the best two suited for efficiently training the networks. Based on the results, InceptionV3, and ResNet101V2 as well as the third optimizer, RMSprop did not hold satisfactory performance on the dataset. The strategy to reduce overfitting of the model using the combination of data augmentation, dropout, and weight decay when training the models using the SGD optimizer with momentum is considered to have performed best overall. The Xception model achieved a Top-1 validation and test accuracy of 96.90% and 98.10% respectively as well as a Top-3 validation and test accuracy of 99.65% and 99.80% respectively with the SGD while training the Xception model with Adam optimizer also achieved high accuracy with a slight Top-1 validation accuracy improvement of 0.45% over the SGD on the validation and test set. However, SGD was preferred due to is ease of interpretability and speed of generalization.

5.3.4. *SAIN-LIM model for nutrients profiling and scoring*

Table 5.6 shows the 15 classes of food items considered in this study including their portion description, individual weight per portion size, as well as their SAIN and LIM scores. The SAIN score and the LIM score, were computed using Equation 5.4 (described above) respectively while the nutrient score was computed using Equation 5.5. As an illustration, Beef salad with serving size of **182 g**, and **Energy**: 473.2 $kcal$, **Protein**: 32.58 g, **Fibre**: 0.73 g, **Ca**: 34.58 mg, **Fe**: 2.95 mg, **Vit** C: 1.46 mg, **Vit** D: 0.18 μg, **Na**: 609.7 mg, **SFA**: 7.01 g, **Added sugar**: 3.82 g, contains:

$$\boldsymbol{SAIN\ Score} = \left(\frac{\left(\frac{\left(\frac{32.58}{65} + \frac{0.73}{25} + \frac{34.58}{900} + \frac{2.95}{12.5} + \frac{1.46}{110} + \frac{0.18}{5} - 0.013 \right)}{5} \times 100 \right)}{473.2} \right) \times 100$$

$$= \left(\frac{\left(\frac{(0.8539 - 0.013)}{5} \times 100 \right)}{473.2} \right) \times 100$$

$$= \left(\frac{16.818}{473.2} \right) \times 100$$

$$\boldsymbol{SAIN\ Score = 3.55}$$

$$\boldsymbol{LIM\ Score} = \left(\frac{\frac{609.7}{3153} + \frac{7.01}{22} + \frac{3.82}{50}}{3} \right) \times 100; = \left(\frac{0.58837}{3} \right) \times 100$$

$$\boldsymbol{LIM\ Score = 19.61}$$

$$\boldsymbol{Nutrient\ Score} = \frac{\boldsymbol{SAIN\ Score}}{\boldsymbol{LIM\ Score}}$$

$$= \frac{\mathbf{3.55}}{\mathbf{19.61}}$$

$$\boldsymbol{Nutrient\ Score = 0.18}$$

From the above analysis, once Beef salad has been recognized using the deep learning classification model, the diet quality was further analyzed using the SAIN-LIM nutrient profiling model. The result indicated that the consumed Beef Salad had a low Nutrient Score of $\mathbf{0.18}$ which was majorly as a result of high amount of calorie as well as high disqualifying nutrients such as Sodium, Saturated fat and Added sugar with respect to low amount of the qualifying nutrients such as Protein, Calcium, etc. Hence, with this, based on the quantity

consumed in term of nutrients, relative to the *daily recommended value (DRV)*, the results implied that a recommendation for Beef salad to be consumed in limited quantity was made in order to maintain a healthy diet.

The results further implied that the diet quality of consumed foods (e.g. Beef salad) in terms of their healthiness benefits can be modeled by estimating and analyzing the presence of qualifying and disqualifying nutrients present in food which was also evident in the conclusions drawn from the study presented by *Darmon et al. (2009)*.

Table 5.6: Weight, nutrient composition, and SAIN, LIM (score of nutritional adequacy of individual foods, and score of nutrients to be limited) of individual foods

	Portion description	Estimated portion weight	SAIN - LIM			
						Weighted by 10
Food item		**(g)**	**Min ratio**	**SAIN**	**LIM**	**Nutrient score**
Beef salad	1 serving	182	0.013	3.555	19.610	1.81
Cheeseburger	2 serving	400	0.005	4.565	74.022	0.62
Chicken curry	1 serving	233	0.083	8.029	4.645	17.28
Egusi soup	1 serving	130	0	1.514	5.953	2.54
Fish and chips	1 serving	250	0.136	7.482	66.096	1.13
Fried rice with shrimps	1 serving	130	0.035	2.644	6.212	4.26
Hot dog	2 serving	114	0	4.669	27.754	1.68
Lasagna	1 serving	375	0.027	7.381	51.064	1.45
Mac & Cheese	1 serving	120	0	5.185	16.199	3.20
Pad Thai	1 serving	250	0.092	3.911	20.645	1.89
Pizza	3 standard slices	321	0.079	4.284	55.587	0.77
Rice and beans	1 serving	150	0.046	6.017	5.085	11.83
Spaghetti with beef & tomato-sauce	1 serving	260	0.066	7.940	21.019	3.78
Steak with mashed potatoes	1 serving	320	0	5.362	24.755	2.17
Waffles and fruits	1 serving	180	0	3.875	33.558	1.15

5.3.5. *Nutrient profile visualization*

Figure 5.13 illustrated a 4-quadrant-based graphical representation of the health benefits and implications of the 15 classes of food based on their respective SAIN and LIM scores on a logarithmic scale. A logarithmic scale was chosen both for the SAIN and the LIM scores, in order to ensure easy response of the scale to skewness towards large values. The serving size or portion weight for *Chicken curry and Rice and beans* were $233\,g$ and $150\,g$ respectively and their associated SAIN score (8.03 and 6.02 respectively) were greater than 5 but their LIM scores (4.65 and 5.09 respectively) were less than 7.5. This is indicative of the fact that the foods have high nutrient density, i.e. rich in nutrients recommended for healthy living and low in nutrients considered to be **unhealthy,** hence, their consumption supports good health. *Egusi soup (portion weight:* $130\,g$*, SAIN score:* 1.51*, LIM score:* 5.94*)* and *Fried rice & shrimps (portion weight:* $130\,g$*, SAIN score:* 2.64*, LIM score:* 6.21*)* in the $2nd$ quadrant has low SAIN and low LIM scores are considered to be **neutral** or balanced in term of their contribution to health benefits. *Spaghetti with beef & tomato-sauce (portion weight:* $260\,g$*, SAIN score:* 7.94*, LIM score:* 21.02*)*, *Macaroni & cheese (portion weight:* $120\,g$*, SAIN score:* 5.19*, LIM score:* 16.20*), Fish & chips (portion weight*: $250\,g$*, SAIN score:* 7.48*, LIM score*: 66.10*), Lasagna (portion weight:* $375\,g$*, SAIN score:* 7.38*, LIM score:* 51.06*), and Steak with mashed potatoes (portion weight:* $320\,g$*, SAIN score:* 5.36*, LIM score:* 24.76*)* in the $3rd$ quadrant were high in both SAIN and LIM scores and hence should be consumed in limited quantity or occasionally. *Pad Thai (portion weight:* $250\,g$*, SAIN score:* 3.91*, LIM score:* 20.65*), Pizza (portion weight:* $321\,g$*, SAIN score:* 4.28*, LIM score:* 55.59*), Waffles & fruits (portion weight:* $180\,g$*, SAIN score:* 3.88*, LIM score:* 33.56*), beef salad (portion weight*: $180\,g$*, SAIN score:* 3.56*, LIM score:* 19.61*), Cheeseburger (portion weight:* $400\,g$*, SAIN score:* 4.57*, LIM score:* 74.02*), and Hot dog (portion weight:* $114\,g$*, SAIN score:* 4.67*, LIM score:* 27.75*)* in the $4th$ quadrant all have their SAIN score less than 5 and their LIM scores are greater than 7.5 which implied that consuming these foods in their respective quantities could lead to chronic dietary related health problems such as cancer, diabetes, and heart disease. Therefore, it is recommended that the foods should be consumed in very limited quantities to ensure healthy living. Figure 5.14 shows the general overview of the proposed system including all components.

These results further implied that the deep learning assisted dietary assessment approach has significant value for determining the quality of the dietary intake of a user.

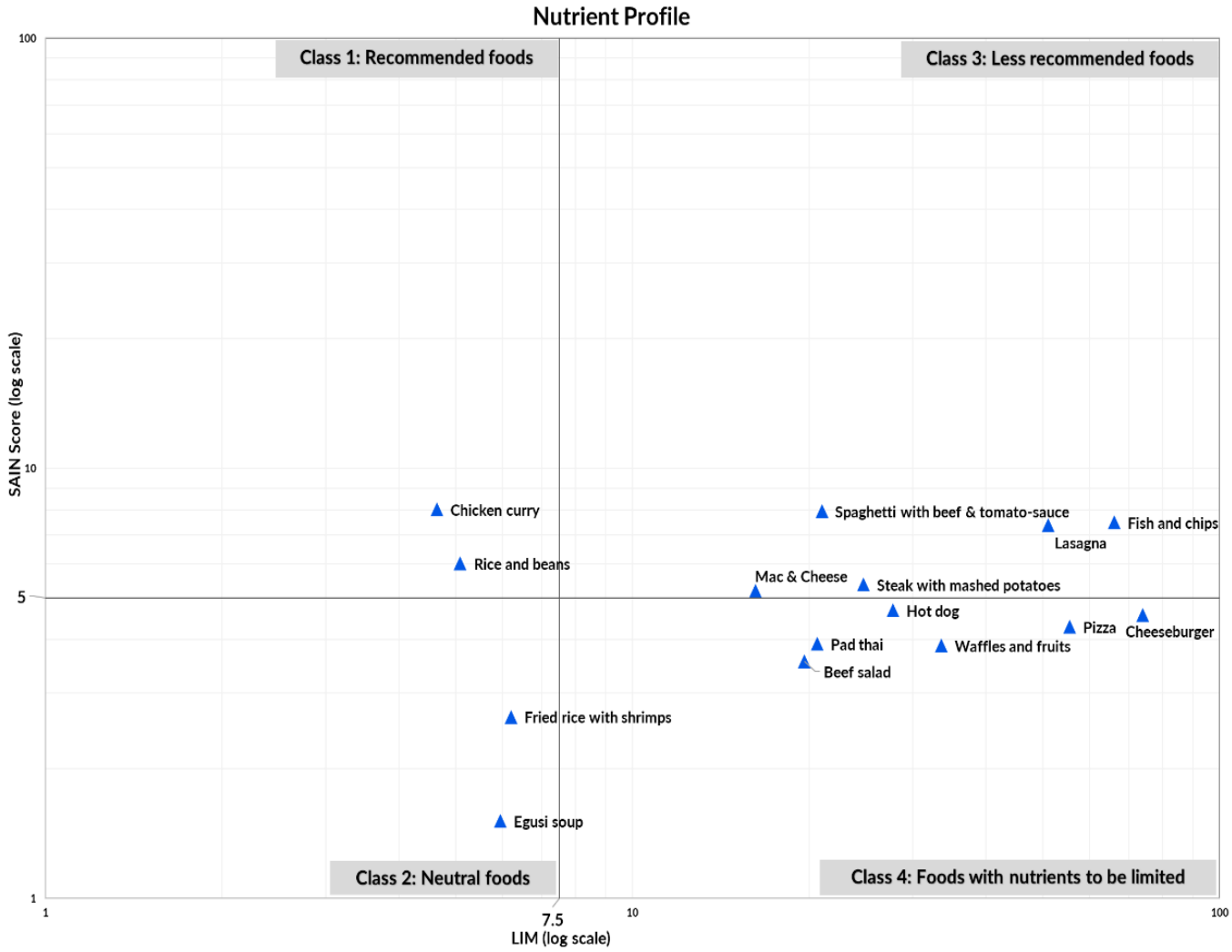


Figure 5.13: Graphic representation of the classification of selected foods with the SAIN,LIM (score of nutritional adequacy of the individual foods, and score of nutrients to be limited) model and their position within the 4 nutrient profile classes (on log scales)

Figure 5.15 showed the interactive computer-based application designed for the proposed food recognition and nutrient scoring system. In its operation, the application takes in captured "query or test food image", predicts the type of food present in the image, analyzes the diet quality of the food using the SAIN-LIM model and then displays the results of the analysis (estimated weight, nutrients composition, SAIN, LIM, and nutrient scores) as well as the distribution of the nutrients present in the food. The application can serve as a valuable computer-based user-friendly tool for self-assessment of diet quality of food consumed by a user. For more convenient self-administered diet quality assessment, the system can be alternatively be integrated into mobile app (Android, iOS, etc.) or web application, and deployed using smart mobile devices.

5.3.6. *Overview of system pipeline*

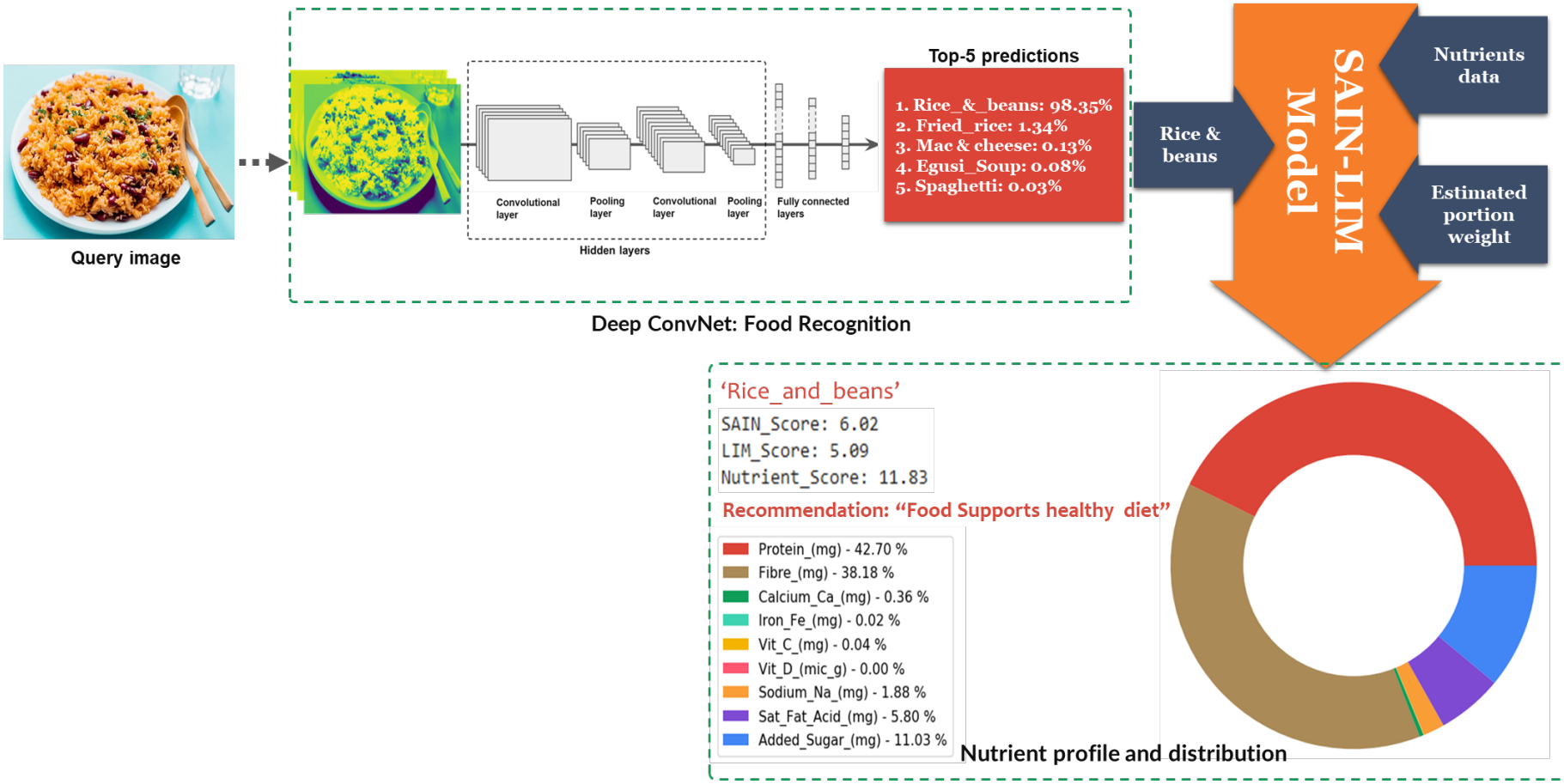


Figure 5.14: General Overview of the Deep Learning Assisted Composite Food Recognition and Diet Quality Assessment System

5.3.7. *An interactive computer-based application designed for the proposed system*

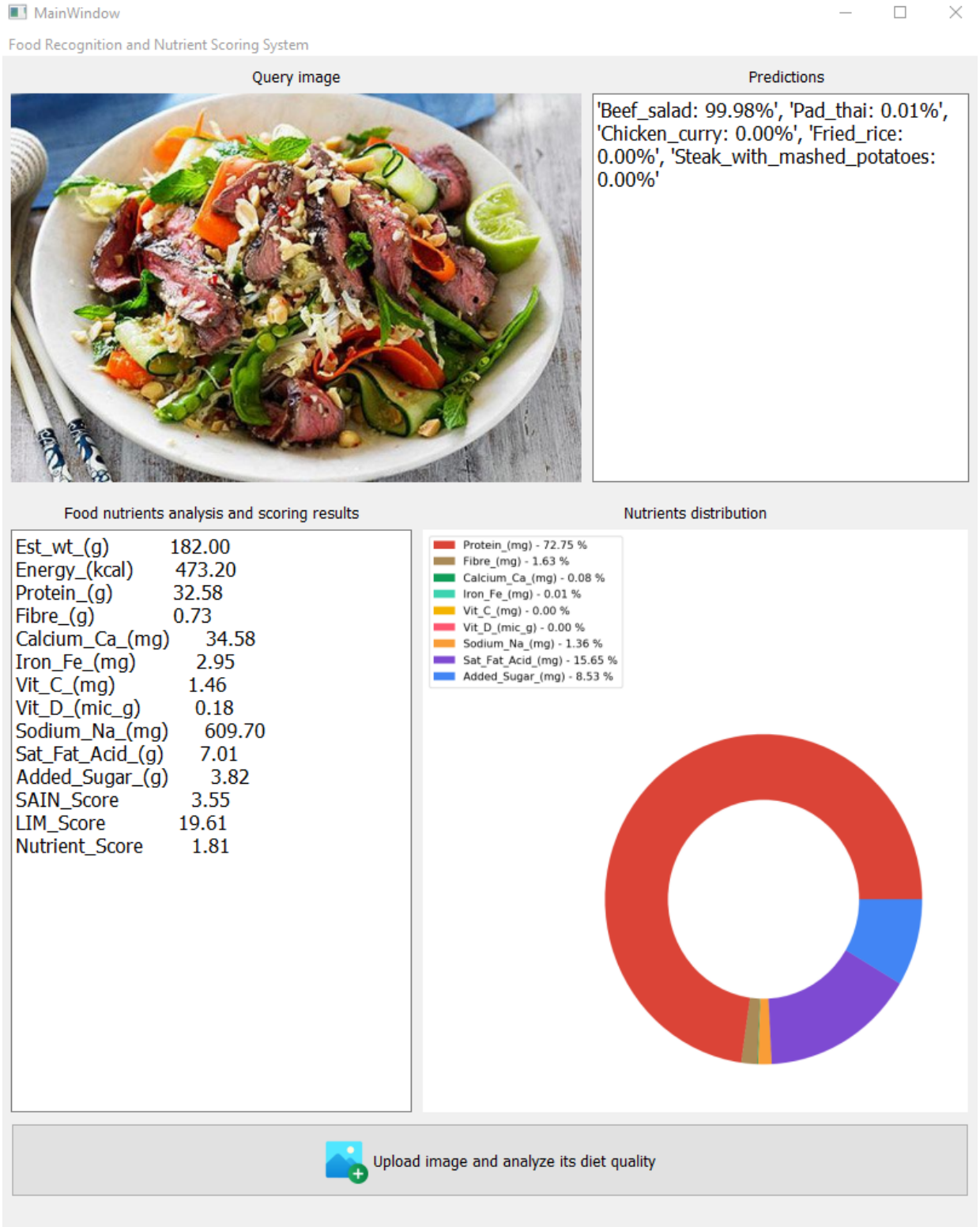


Figure 5.15: The Graphical User Interface (GUI) of the interactive computer-based application designed for the proposed System

5.3.8. *Conclusion and future work*

In this study, an images-based food recognition system based on deep convolutional neural networks, along with a diet quality assessment system was presented as a non-destructive tool for classifying and evaluating the healthiness benefits of foods consumed by a user. In the implementation of the recognition system, the deep convolutional neural network model was deployed through transfer a learning approach to develop an automatic, reliable and accurate Xception model to classify food images before the nutrient contents is being assessed using the SAIN, LIM nutrient profiling model.

Based on the results obtained, the system can promptly and accurately recognize composite or mixed dishes (which is often a challenging food recognition task) when carrying out dietary assessment tasks such as food record. In particular, out of the five pre-trained CNN models repurposed and an additional vanilla baseline CNN model trained from scratch for this study, the Xception model fine-tuned on the ***Food15*** training set generalized best both with ***Adam*** and ***Stochastic Gradient Descent*** optimization algorithm with test accuracies 98.10% and 98.10% (Table 5.5) respectively.

With an accurate food recognition component of the system already developed, the deep learning assisted diet quality assessment system can further effectively evaluate and analyze the nutrient contents present in the food consumed by its user thus yielding a reliable ***nutrient score*** for the food. The score can sufficiently serve as a ***rubric*** for evaluating the nutrient quality and contribution of the food to the user's diet. In its application, the system presented in this study can be effectively used as a tool to self-administer daily dietary intake monitoring, recording, and quality assessment of dietary intake as well as keeping up with nutrition goals of the user. The system is capable of outperforming traditional pen-and-paper based approaches for administering 24-hour dietary recall and food record for dietary assessment purposes.

For future studies, there are plans to expand the size of food database to cover more classes of food and this will bring about the need for fine-tuning deeper layers of the pre-trained models. In addition, further integration of more mainstream diet quality assessment and dietary diversity indicators is also envisioned in parallel in order to better assess the nutritional status and the diet quality of the food consumed by a user. Deploying the system presented in this study on cross-platform mobile device applications in order to enhance self-administering of the tool is also a key area of consideration.

6 Summary, General Conclusion and Recommendations for Future Studies

6.1. Summary and General Conclusion

Dietary intake remains a major subject of concern as part of the leading causes of nutrition-related illnesses such as chronic heart diseases, diabetes, vascular syndromes, and death in the world today. Measuring, analyzing and monitoring dietary intakes in order to investigate the diet quality has become a major area of study among researchers.

This study demonstrated the potential of using image-based technology-assisted techniques to monitor and analyze dietary intake of consumed foods. The first objective was met by utilizing mainstream computer vision image analysis and machine learning techniques to develop an automatic food recognition and nutrient profiling system for single foods such as Avocado, Bagel, and Croissant. The recognition model developed produced satisfactory results for recognizing food with high precision and recall values ranging from $75 - 100\%$. After this, the diet quality of the recognized food was then obtained using the SAIN-LIM nutrient profiling model.

The second objective deployed deep convolutional neural network assisted techniques to extend the food recognition and nutrient profiling system to detect and classify more complex composite foods such as beef salad, rice & beans, and lasagna. Five pre-trained CNN models were repurposed and an additional vanilla baseline CNN model (trained from scratch) were evaluated for this study. The Xception model fine-tuned on the ***Food15*** training set generalized best both with ***Adam*** and ***Stochastic Gradient Descent*** optimization algorithm with test accuracies 98.10% and 98.10% respectively. This served as a better food recognition model for the nutrient system and also indicative of the fact that deep learning assisted image-

based techniques performed better and is more reliable than the conventional computer vision image features analysis technique deployed in chapter 3. This was made possible due to the availability of the expensive and highly efficient computation capacity to deploy the deep convolutional neural network recognition models used in chapter 4.

Generally, the results obtained were indicative of the fact that both techniques are promising methods to adopt for the development of image-based dietary assessment and nutritional status evaluation.

6.2. RECOMMENDATION FOR FUTURE STUDIES

1. **Large-scale Standard and Global Food-image Dataset**: Like the huge ImageNet dataset which consists of several classes of general everyday objects mostly applicable in the computer vision domain, there is need for a large-scale ImageNet-type of dataset for food images which covers cuisines from all around the world. This can serve as a major resource for the development of advanced applications such as food-image search engines, robust classification and food image scene understanding, ingredients recognition system, as well as high-level image-based nutritional assessment systems.
2. **Large-scale Pre-trained Networks from Food Images**: In recent times, deep learning architectures such as Xception, VGG networks, etc. have been developed and have played significant roles in the field of computer vision and several others. It will be of great benefits to have pre-trained networks designed for food images from which representational knowledge can be acquired and transferred to similar tasks with small dataset. There are large RGB-D depth estimation datasets which are being repurposed to estimate volume of food from images. Results obtained from majority of these studies are very unreliable. Pretrained models developed from actual food volumes as ground-truth data would be very useful in better estimating volume of food from images for dietary assessment and nutritional status evaluation purposes.
3. **Robust and Unified Dataset of Nutrient Composition for Developing Countries**: It has become pertinent for developing countries e.g. as present in Africa to focus more attention on developing indigenous unified and reliable database of all commonly consumed food and inherent nutrients information in order to better account for an individual dietary intake and that of the population at large.

Appendices

7.1. APPENDIX A: MACHINE LEARNING CLASSIFICATION ALGORITHMS

Source: *Lindholm et al. (2019)*

Algorithm **1: Logistic Regression (LR)**

Given: Training data $\{\mathbf{x_i}, \mathbf{y_i}\}^{\mathbf{n}}_{\mathbf{i=1}}$ (with output classes $\mathbf{y = 0, 1}$) and test input $\mathbf{x_{new}}$

Result: Predicted test output $\hat{\mathbf{y}}$

Learning:

1. Compute : $\boldsymbol{\ell}(\mathbf{w}) = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{y_i} = \mathbf{ln}(\boldsymbol{\sigma}(\mathbf{w^T x_i})) + (\mathbf{1} - \mathbf{y_i})\,\mathbf{ln}(\mathbf{1} - \boldsymbol{\sigma}(\mathbf{w^T x_i}))$

Prediction:

2. Compute loglikelihood $\mathbf{p(y = 1 \mid x_{new})} = \frac{\mathbf{1}}{\mathbf{1+ e^{w^T x}}}$; $\mathbf{p(y = 0 \mid x_{new})} = \mathbf{1} - \frac{\mathbf{1}}{\mathbf{1+ e^{w^T x}}}$;
3. **If** $\mathbf{p(y = 1 \mid x_{new}) > p(y = 0 \mid x_{new})}$, **set** $\hat{\mathbf{y}} \leftarrow 1$, otherwise set **set** $\hat{\mathbf{y}} \leftarrow 0$

Algorithm **2: k-nearest neighbor, (k-NN)**

Given: Training data $\{\mathbf{x_i}, \mathbf{y_i}\}^{\mathbf{n}}_{\mathbf{i=1}}$ (with output classes $\mathbf{1}, \ldots \mathbf{K}$) and test input $\mathbf{x_{new}}$

Result: Predicted test output $\hat{\mathbf{y}}$

Learning: Nothing to do! (Just store the data).

Prediction:

1. Find the $\mathbf{k}$ training data point(s) $\mathbf{x_i}$ which has the shortest Euclidian distance $\|\mathbf{x_i} - \mathbf{x_{new}}\|$ **to** $\mathbf{x_{new}}$
2. Decide $\hat{\mathbf{y}}$ with a majority vote among those $\mathbf{k}$ nearest neighbors

Algorithm **3: Linear Discriminant Analysis (LDA)**

Given: Training data $\{\mathbf{x_i}, \mathbf{y_i}\}^{\mathbf{n}}_{\mathbf{i=1}}$ (with output classes $\mathbf{1}, \ldots \mathbf{K}$) and test input $\mathbf{x_{new}}$

Result: Predicted test output $\hat{\mathbf{y}}$

Learning:

1. **for** $\mathbf{k = 1}, \ldots \mathbf{K}$ **do:**
2. | **Compute :** $\hat{\boldsymbol{\pi}}_{\mathbf{k}} = \frac{\mathbf{n_k}}{\mathbf{n}}$**; and** $\hat{\boldsymbol{\mu}}_{\mathbf{k}} = \frac{\mathbf{1}}{\mathbf{n_k}} \sum_{\mathbf{i}:\mathbf{y_i}=\mathbf{k}} \mathbf{x_i}$
3. **end**
4. **Compute:**

$$\hat{\boldsymbol{\Sigma}} = \frac{\mathbf{1}}{\mathbf{n} - \mathbf{K}} \sum_{\mathbf{k=1}}^{\mathbf{K}} \sum_{\mathbf{i}:\mathbf{y_i}=\mathbf{k}} (\mathbf{x_i} - \hat{\boldsymbol{\mu}}_{\mathbf{i}})(\mathbf{x_i} - \hat{\boldsymbol{\mu}}_{\mathbf{i}})^{\mathsf{T}}$$

Prediction:

5. **for** $\mathbf{k = 1}, \ldots \mathbf{K}$ **do:**
6. | **Compute:**

$$\mathbf{p}(\mathbf{y = k} \mid \mathbf{x_{new}}) = \frac{\mathbf{e}^{-\frac{\mathbf{1}}{\mathbf{2}} (\mathbf{x_i} - \hat{\boldsymbol{\mu}}_{\mathbf{i}})^{\mathsf{T}} \boldsymbol{\Sigma}^{-\mathbf{1}} (\mathbf{x_i} - \hat{\boldsymbol{\mu}}_{\mathbf{i}})}}{(\mathbf{2}\hat{\boldsymbol{\pi}})^{\frac{\mathbf{m}}{\mathbf{2}}} \quad |\boldsymbol{\Sigma}|^{\frac{\mathbf{1}}{\mathbf{2}}}}$$

7. **end**
8. **Find largest** $\mathbf{p}(\mathbf{y = k} \mid \mathbf{x_{new}})$ **and set** $\hat{\mathbf{y}}_{\mathbf{new}}$ **to that** ***k***

Where: $\hat{\boldsymbol{\pi}}_{\mathbf{k}}$ and $\hat{\boldsymbol{\mu}}_{\mathbf{k}}$ = relative occurrence of class k in the training data and mean feature vector; $\boldsymbol{\Sigma}$ = shared covariance matrix for the entire dataset

7.2. APPENDIX B: IMAGE DATASET PRE-PROCESSING TOOL

Algorithm 4**: Image Dataset Pre-Processing Tool**

Input: Entire image dataset, D $\{\boldsymbol{x_i}\}^{\mathbf{n}}_{\mathbf{i=1}}$

Output: 3 sets of datasets belonging to *m* order of classes (e.g. 15) each with *n* number of x images and contain:

1. Training set: $\{\boldsymbol{x_i, y_i}\}^{\mathbf{n}}_{\mathbf{i=1}}$
2. Validation set: $\{\boldsymbol{x_i, y_i}\}^{\mathbf{n}}_{\mathbf{i=1}}$`
3. Test set: $\{\boldsymbol{x_i}\}^{\mathbf{n}}_{\mathbf{i=1}}$

Dataset preprocessing outline

Procedure:

1. for each image $\boldsymbol{x_i}$in D, do:
2. **if** $\boldsymbol{x_i}$ (is a) supported image format:
3. $\boldsymbol{x_i}\ (.\boldsymbol{png}) \leftarrow \boldsymbol{x_i}\ (.\boldsymbol{xyz})$
4. **else** delete $\boldsymbol{x_i}\ (.\boldsymbol{xyz})$
5. **end**

Covert each image to .png format and delete if unsupported

6. **for** each new image $\boldsymbol{x_i}$:
7. $\boldsymbol{x_i}\ (\mathbf{p} \times \mathbf{q}) \leftarrow \boldsymbol{x_i}\ (\mathbf{r} \times \mathbf{s})$
8. Append resized images as new dataset, D
9. **end**

Resize from $(r \times s)$ *width and height to* $(p \times q)$

10. Randomly shuffle and divide new dataset, D into three sets in ratio: $\mathbf{70}\%:\mathbf{15}\%:\mathbf{15}\%$ each belonging to *m* order of classes.
11. Set the $\mathbf{70}\%$ to be the training set and explicitly map each image $\boldsymbol{x_i}$ to label $\boldsymbol{y_i}$:such that Training set $= \{\boldsymbol{x_i, y_i}\}^{\mathbf{n}}_{\mathbf{i=1}}$
12. Set the a $\mathbf{15}\%$ to be the validation set: $\{\boldsymbol{x_i, y_i}\}^{\mathbf{n}}_{\mathbf{i=1}}$.
13. Set the last $\mathbf{15}\%$ to be the test set with no label mapping: $\{\boldsymbol{x_i}\}^{\mathbf{n}}_{\mathbf{i=1}}$
14. End

Dataset splitting and annotation

15. **Before feeding data to CNN**

Feature scaling

7.3. APPENDIX C: PERFORMANCE OF BENCHMARK DEEP LEARNING ARCHITECTURES ON POPULAR FOOD IMAGE DATASETS

Table 7.1: Top Performances of Deep learning models on Publicly available Dataset

Dataset	Dataset Description and Source	Research work ref	DL Architecture	Top-1% & Top-5% performance
UECFood-256: (Yanai et al., 2015b)	• Popular foods in Japan and other countries • 256/ 31397 images each with a bounding box indicating the location of the food item in the image. • http://foodcam.mobi/dataset256.html	*Martinel et al. (2018)*	WISeR	83.15 | 95.45
Food-101: (Bossard et al., 2014)	• Popular food in USA • 101/ 101,000 images • https://data.vision.ee.ethz.ch/cvl/datasets_extra/food-101/	*Martinel et al. (2018)*	WISeR	90.27 | 98.71
UECFood-100: (Matsuda et al., 2012)	• Popular Japanese food • 100/9060 images each with a bounding box indicating the location of the food item in the image. • http://foodcam.mobi/dataset100.html	*Martinel et al. (2018)*	WISeR	89.58 | 99.23
UECFood-100: (Matsuda et al., 2012)	• Popular Japanese food • 100/9060 images each with a bounding box indicating the location of the food item in the image. • http://foodcam.mobi/dataset100.html	*H. Hassannejad et al. (2016)*	Inception v3	81.50 | 97.30
Food-524: (G. Ciocca et al., 2017)	• Popular food in USA, China and Japan • 524/ 247,636 • http://www.ivl.disco.unimib.it/activities/food524db/	*G. Ciocca et al. (2017)*	ResNet-50	81.34 | 95.45
Food-475: (G. Ciocca et al., 2017)	• Popular food in USA, China and Japan • 475/ 247,636 • http://www.ivl.disco.unimib.it/activities/food475db/	*Ciocca et al. (2018)*	ResNet-50	81.59 | 95.50
VIREO Food172: (Chen et al., 2016)	• Popular Chinese dishes • 172/ 110241 • http://vireo.cs.cityu.edu.hk/VireoFood172/	*Ciocca et al. (2018)*	ResNet-50	85.86 | 97.32

7.4. APPENDIX D: OVERVIEW OF DEEP NEURAL NETWORKS

A Neural network architecture attempt to solve a supervised machine learning task in which, given a set of training examples (training set) of the form $\{(x_i, y_i); i = 1, 2, \dots, n\}$, such that x_i is the feature vector of i^{th} example also called the input variable and y_i is the desired output or target variable. The goal is, given the training set, to learn a function $f: X \rightarrow Y$ which maps the input variables to the target so that $f(x)$ also called the ***Hypothesis*** is considered to be a good predictor for the corresponding value of y.

In understanding how neural networks work, it is a common approach to resort to first principle where it was originally inspired by the goal of modeling biological neurons found in the human brain. A biological neuron is the basic computational unit of the brain. It is composed of a cell body containing the nucleus and most of the cell's complex components, many branching extensions called ***Dendrites*** as well as a very long extension called the ***Axon***.

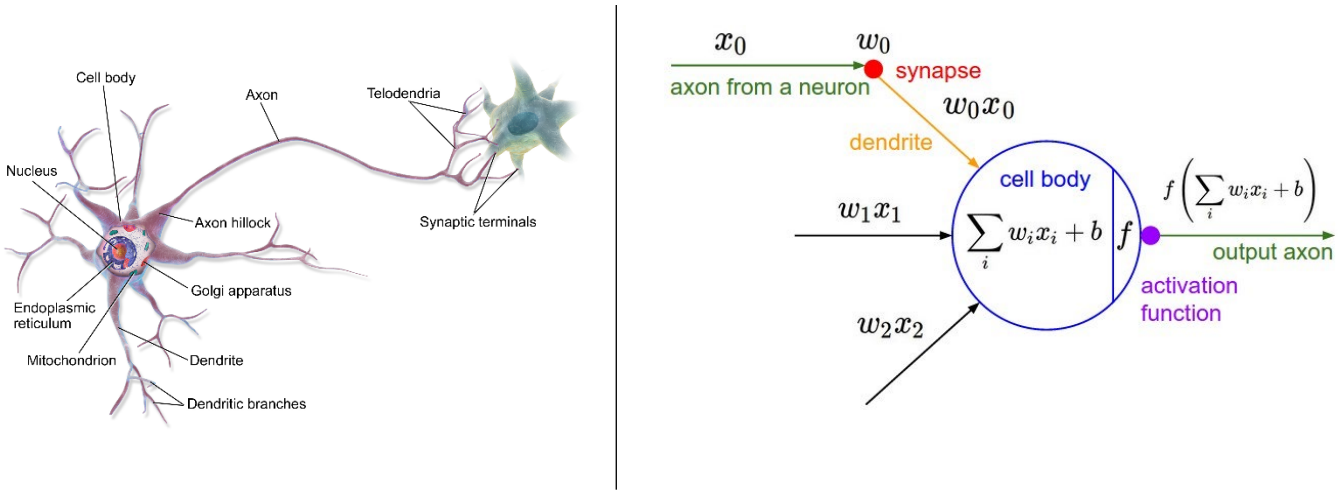


Figure 7.1: Mathematical model representation of Biological neuron

Approximately 86 billion neurons can be found in the human nervous system and they are connected with approximately $10^{14} - 10^{15}$ synapses. Figure 7.1 shows an image if a biological neuron (left) and a mathematical model representation (right). Each neuron receives input signals from its *dendrites* and gives an output signal along its axon. An artificial neuron is a mathematical function based on a model of biological neurons. In modeling the biological neuron, the signals that travel along the axons (aka input features $\boldsymbol{x_i}$) interact in a multiplicative manner (e.g. $\boldsymbol{w_i x_i}$) with dendrites of other neurons based on the synaptic strength at that synapse (aka $\boldsymbol{w_i}$). In the basic model, the dendrites carry the weighted signal to the *cell body* where they all get *summed*. If the final sum is above a particular threshold, the neuron can fire, sending an impulse along its axon. The firing rate of the neuron can be modeled with an ***activation function f(.),*** which represents the frequency of the impulse along the axon.

7.4.1. *The perceptron*

Perceptron (also called a Single-layer neural network) is one of the simplest artificial neurons or artificial neural network (ANN) architecture. As illustrated in Figure 7.2, it is a type of binary classification algorithm that makes its predictions based on a linear combination of a set of real-valued ***weights, bias*** and corresponding ***input feature vector***, i.e. it computes a weighted sum of its inputs ($z = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n = w^T x$) along with the bias term ***b*** and then passed the sum through a threshold or ***step function or activation function*** and obtain the results: $h_w(x) = step(z)$, where $z = w^T x$. The idea behind the step function given in Equation 7.1 is that, if the summed input gets above certain threshold (for example 0.5), then the neuron outputs a value of 1.0, otherwise it would output 0.0. To put it in more precise algebraic terms:

$$output = \begin{cases} 0, & if\ w \cdot x + b \leq 0 \\ 1, & if\ w\ .x + b > 0 \end{cases} \tag{7.1}$$

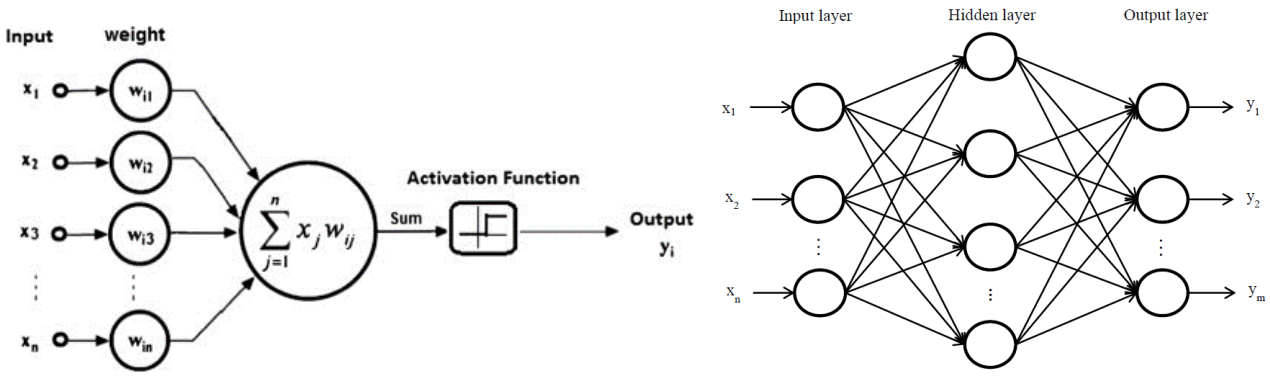


Figure 7.2: (a) Perceptron (b) Feed-forward Neural Network

A big draw-back with single-layer perceptron is that it can only classify linearly separable sets of vectors. Hence, if the vectors are not linearly separable, classification may never be achieved. An improvement on the single-layer perceptron, called a ***multi-layer perceptron (MLP***) or *multi-layer neural network* or ***feed-forward neural network (FFNN)*** is one which has the capacity to learn non-linear functions. A feedforward neural network is a structure composed of one *input layer*, several *hidden layers* of neurons with outputs serving as the inputs of the neuron of the next layer, as well as an *output layer*. When all the neurons in a layer are connected to every neuron in the previous layer (i.e. the input neurons), the layer is referred to as ***fully connected layer*** or *dense layer.* For deeper knowledge of feedforward neural network, see *(Goodfellow et al., 2016; LeCun et al., 2015).*

7.4.2. *Building a feedforward neural network*

Building a feedforward neural network involve four major steps namely: selecting a network architecture, determining the choice of hyperparameters, training the neural network, and regularizing the network.

7.4.2.1 Selecting a network architecture

This involve choosing a layout for the neural network which largely depends on the task at hand. A popular variant of feedforward neural network is the convolutional neural network which can be used for classification of input data that have grid-like topology e.g. for image classification and audio signal classification.

7.4.2.2 The choice of parameters and hyperparameters

This involve setting parameters such as the weights and biases needed for the model to learn as well as the hyperparameters which are needed to train the model including the number of input neurons/ units (a function of the dimension of the input feature vector), number of hidden units per (per hidden layers), and the number of output units (a function of the number of classes). Other major hyperparameters are discussed as follows:

Choice of Non-linear Activation functions

In selecting an activation functions for a feedforward neural networks, non-linearity is needed because its aim is to produce a nonlinear decision boundary through non-linear combinations of the weight and inputs. There are several activation functions in existence as illustrated in Figure 7.3, the three most common are:

a. **Sigmoid ($\sigma(z)$)**: The key idea underlying sigmoid is that, it takes a real-valued number and "squeezes" it into an interval [0; 1] using Equation 7.2. It has a property that, when $z = w \cdot x + b$ is large and positive, the output from the sigmoid neuron is approximately equal to 1 ($e^{-z} \approx 0 \; and \; \sigma(z) \approx 1$). In the same manner, when $z = w \cdot x + b$ is very negative, the output tends to infinity ($e^{-z} \approx \infty \; and \; \sigma(z) \approx 0$). The sigmoid function is affine close to $z = 0$ and saturates at 0 and 1 as z decreases or increases. This is a major limitation that leads to the problem of *vanishing gradients* and *slow convergence* which have restricted sigmoid's popularity from usage on hidden layers to only on output layer.

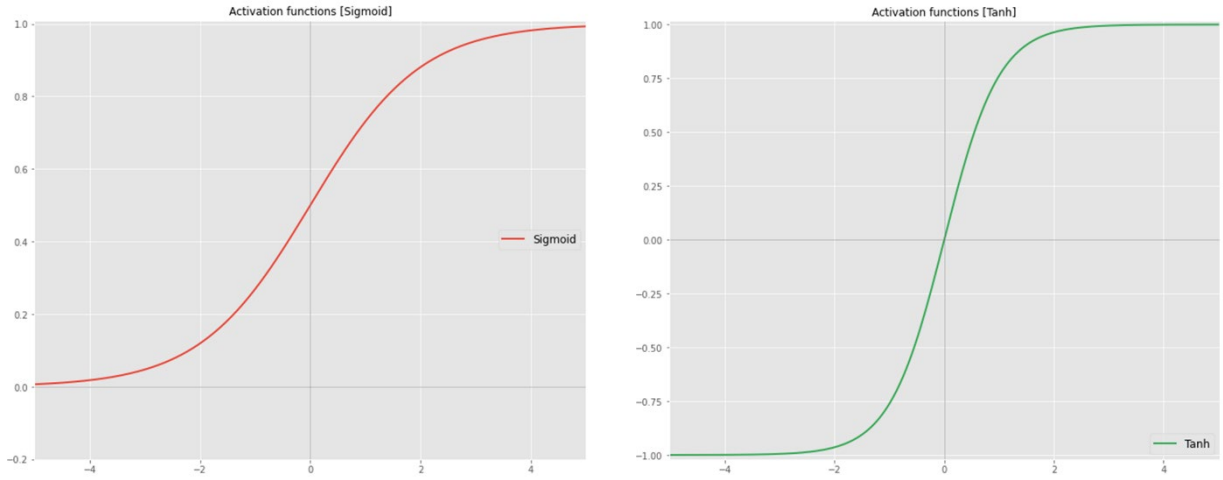
$$\sigma(z) = \frac{1}{(1 + e^{-z})} \tag{7.2}$$

b. **Hyperbolic tangent (*tanh(z)*)**: tanh(z) squashes a real-valued number into an interval [-1, 1] using Equation 7.3. Unlike sigmoid, it has a zero-centered output which makes learning even faster. However, like the sigmoid function, it still suffers from the problem of vanishing gradients.

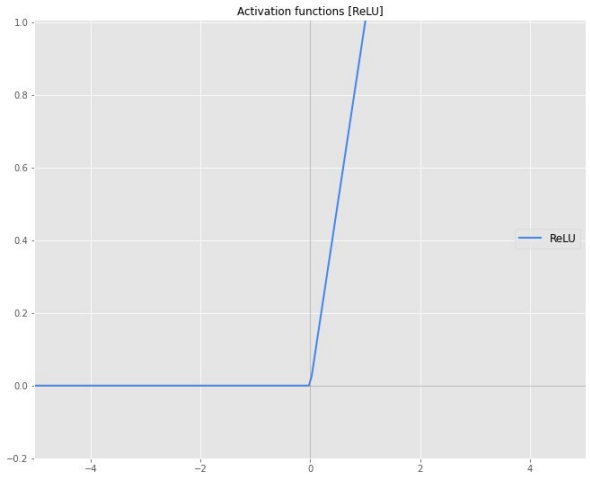
$$tanh(z) = \frac{e^{z} + e^{-z}}{e^{z} + e^{-z}} \tag{7.3}$$

c. **Rectified Linear Units (ReLU)**: ReLU has become very popular in recent years and now the standard choice in most neural network models. It has a linear non-saturating form and has been proven to have 6 times improvement in convergence over ***tanh*** function. It has a very simple and efficient mathematical form given as Equation 7.4;

$$ReLu(z) = max\,(0, z) \tag{7.4}$$



(a) (b)



(c)

Figure 7.3: Activation functions [(a) Sigmoid (b) Tanh (c) ReLU]

The usage of any particular activation function depends on the task at hand. This research work comprised of a multiclassification task, hence the use of the ReLU activation function in all hidden layers and ***Softmax function*** in the output layer. The Softmax function is a categorical probability distribution function that is applied to the output scores of the neurons before the loss computation. It attempts to squeeze the output of each output neuron such that the total sum is equal to 1. For a given class, the Softmax function is computed using Equation 7.5:

$$\boldsymbol{P(k \mid x_i) = f(s(x))_k = \frac{e^{s_k(x)}}{\sum_{j=1}^{K} e^{s_j(x)}}} \tag{7.5}$$

Where:

$\boldsymbol{K}$ = total number of classes;

$\boldsymbol{s(x)}$ = a vector of all individual scores $\boldsymbol{s_k}$ for each class $\boldsymbol{k}$ for each instance *x;*

$\boldsymbol{s_j}$ = scores inferred by the network for each class *k* in *K*

$P(k \mid x_i)$ = estimated class probability that an example $\boldsymbol{x}$ belongs to class $\boldsymbol{k}$, given the scores for each class for that example.

Learning rate (lr)

This is a tuning hyperparameter (required in an optimization algorithm during network training process) which must be set manually to control how much the network weights are being adjusted with respect to the loss gradient. It determines the length of the gradient step (at every iteration) or *how far* to move while moving towards the global optimum of a loss function for a batch or mini-batch of the training inputs.

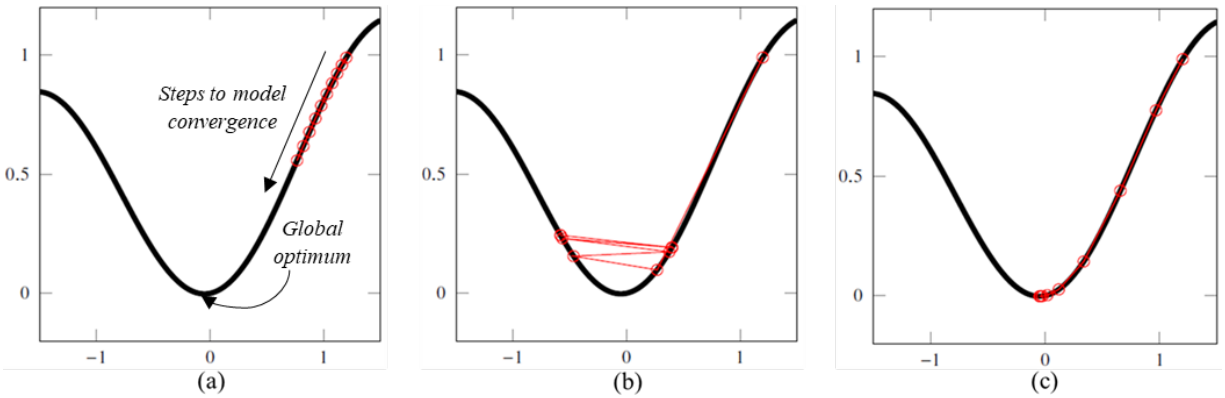


Figure 7.4: Plot of one-dimensional optimization problem using gradient descent of a cost function. (a) too low learning rate, (b) too high learning rate, (c) good learning rate

As illustrated in Figure 7.4, a learning rate which is too low leads to slow learning or convergence of the model to global optimum, and one which is too big may make the learning process overshoot (diverge) the global optimum (or bounce around indefinitely) and never converge since the step is too long.

A good strategy to find a good learning rate is to reduce (decay) it when the error keeps getting worse or increase it if the error becomes fairly constant or too slow. An illustration is given in Figure 7.4 above.

7.4.2.3 Training (learning) a feedforward neural network

After selecting a network architecture, including the number of inputs, hidden and output neurons, the process of training and making predictions using neural network involve solving a numerical optimization problem to find a good set of weights $\widehat{\boldsymbol{w}}$ that minimizes the ***error*** (cost function) at the output of the network usually of the form given in Equation 7.6:

$$\widehat{w} = \underset{w}{arg\,min}\, \boldsymbol{J(w)} \qquad \text{where,}$$

$$\boldsymbol{J(w) = log\, P(y_i \mid x_i; w) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{K} y_{ik}\, log\, P(k \mid x_i; w)} \tag{7.6}$$

Where:

$\boldsymbol{y_i}$ = K-dimensional one-hot encoding vector for i^{th} target belonging to K classes

$\boldsymbol{y_{ik}}$ = elements of the one-hot encoding vectors (the target (ground-truth) probability that the i^{th} example or instance belongs to class $\boldsymbol{k}$)

$\boldsymbol{K}$ = number of classes

$\boldsymbol{P(k \mid x_i; w)}$ = Output of the *Softmax function* = predicted probability that the example $\boldsymbol{x}$ belongs to class $\boldsymbol{k}$ given the scores of each class for that example (parameterized by $\boldsymbol{w}$)

Equation 6.6 is known as the ***Cross-Entropy Cost Function*** **(**often used for error computation in a ***Multiclassification task*** in machine learning).

The system of operation of any numerical optimization algorithm, require the parameters of such algorithm to be updated in an iterative manner. In deep learning, the most popular and effective optimization algorithm and as well the ones

experimented on in this research work include: *Stochastic Gradient Descent (SGD), Root Mean Square Propagation* (*RMSprop), and Adaptive Momentum Estimation (Adam).*

Stochastic Gradient Descent (SGD)

A summary of the implementation or pseudocode for the SGD Algorithm is given below:

Algorithm 5**: Stochastic gradient descent**

1. Initialize all weights $\mathbf{w_j}$ to small random numbers — *Weight initialization*

2. Repeat until convergence: — *Forward pass*
   a. Pick a training example, x at random from the dataset
   b. Feed example through network to compute output y for the given x
   c. For the output unit, compute the correction/ loss function:

$$\frac{\partial \mathbf{J}}{\partial \mathbf{w_{out}}} = \partial_{\mathbf{out}}\mathbf{x}$$

   — *Backpropagation (backward pass)*

   d. For each hidden unit j, compute its share of the correction:

$$\frac{\partial J}{\partial w_j} = \frac{1}{m}\sum_{i=1}^{m}(log\ P(k \mid x_i) - y_{ik})\, x_i$$

   e. Use gradient checking to confirm if backpropagation worked. Then disable gradient checking. — *Gradient checking*
   f. Update each network weight: — *Gradient descent*

$$\mathbf{w_j} := \mathbf{w_j} - \alpha\frac{\partial \mathbf{J}}{\partial \mathbf{w_j}}\ \forall\ \mathbf{j}, \qquad \mathbf{w_{out}} := \mathbf{w_{out}} - \alpha\frac{\partial \mathbf{J}}{\partial \mathbf{w_{out}}}$$

   - $\partial \mathbf{J}, \partial \mathbf{w_{out}}, \partial_{\mathbf{out}}$ = derivatives of loss, weight of the output neuron and error signal at the output neuron respectively.
   - $\mathbf{w_j}, \mathbf{w_{out}}$ = weight of the neuron at the jth layer and output layer respectively
   - $\alpha = \mathbf{learning\ rate}$

Stochastic gradient descent is a FFNN optimization algorithm that computes error on a single training example for every completed pass through the dataset (***one epoch***). Two other alternative methods exist, namely: Batch gradient descent and Mini-batch gradient descent. In *batch gradient descent*, the error is computed on all training examples per epoch. It loops through the training data, accumulate the weight changes and then update all the weights. In *mini-batch gradient descent*, the error is computed on randomly selected small subset (mini-batch), of the dataset before weights update.

The idea behind the ***Backpropagation algorithm*** is that, after a forward pass of the training example through the network (when making predictions), it goes from the output layer to the input layer, propagating the error gradient on the way. Once the algorithm has computed the gradient of the cost function in relation to each parameter in the network, it then performs a gradient descent step to update each parameter with the computed gradients.

Going into mathematical details of gradient descent and backpropagation is beyond the scope of this research work, for deeper contents, please see *(Bishop, 2006; Goodfellow et al., 2016; LeCun et al., 2015)*.

Root Mean Square Propagation (RMSprop)

RMSprop *(Tieleman et al., 2012)* is an unpublished, yet one of the most widely known gradient descent optimization algorithm for deep learning. It was developed to address the problem of aggressive, monotonically decreasing learning rates as well as the problem of large increments or decrements in the magnitude of the gradient of successive mini batches of training data. The RMSProp algorithm *(Géron, 2017)* is given below:

Algorithm 6: **RMSProp**

3. $\mathbf{s} \leftarrow \boldsymbol{\beta}\mathbf{s} + (\mathbf{1} - \boldsymbol{\beta})\, \nabla_{\mathbf{w}}\mathbf{J}(\mathbf{w}) \otimes \nabla_{\mathbf{w}}\mathbf{J}(\mathbf{w})$

4. $\mathbf{w} \leftarrow \mathbf{w} - \propto \nabla_{\mathbf{w}}\mathbf{J}(\mathbf{w}) \oslash \sqrt{\mathbf{s} + \boldsymbol{\varepsilon}}$

$\mathbf{J}(\mathbf{w})$ = cost function, $\nabla_{\mathbf{w}}\mathbf{J}(\mathbf{w})$ = gradient of cost function as a function of weight, $\propto$ = learning rate, $\mathbf{s}$ = gradient vector, $\boldsymbol{\varepsilon}$ = smoothing term, $\otimes$ and $\oslash$ = element-wise multiplication and division respectively

Adaptive Momentum Estimation (Adam)

Adam is an optimization algorithm that leverages the power of adaptive learning rates methods to find individual learning rates for each parameter of the model. The Adam algorithm *(Géron, 2017)* is given below:

Algorithm 7: **Adam**

1. $\mathbf{m} \leftarrow \boldsymbol{\beta}_1\mathbf{m} + (\mathbf{1} - \boldsymbol{\beta}_1)\, \nabla_{\mathbf{w}}\mathbf{J}(\mathbf{w})$
2. $\mathbf{s} \leftarrow \boldsymbol{\beta}_2\mathbf{s} + (\mathbf{1} - \boldsymbol{\beta}_2)\, \nabla_{\mathbf{w}}\mathbf{J}(\mathbf{w}) \otimes \nabla_{\mathbf{w}}\mathbf{J}(\mathbf{w})$
3. $\hat{\mathbf{m}} \leftarrow \frac{\mathbf{m}}{\mathbf{1} - \boldsymbol{\beta}_1^{\mathrm{T}}}$
4. $\hat{\mathbf{s}} \leftarrow \frac{\mathbf{s}}{\mathbf{1} - \boldsymbol{\beta}_2^{\mathrm{T}}}$
5. $\mathbf{w} \leftarrow \mathbf{w} + \propto \hat{\mathbf{m}} \oslash \sqrt{\hat{\mathbf{s}} + \boldsymbol{\varepsilon}}$

$\boldsymbol{\beta}$ = **momentum**, $\mathbf{m}$ = momentum vector, $\mathbf{s}$ = gradient vector, $\boldsymbol{\varepsilon}$ = smoothing term, $\oslash$ = element-wise multiplication and division respectively

7.4.2.4 Model regularization (handling overfitting)

The flexible structure of deep neural network makes it possible to generate tens or hundreds of thousands or even millions of parameters. These excess parameters can sometimes result to the model memorizing the training set which then, however, reduces the capacity of the model to generalize to new dataset. This phenomenon is referred to as Overfitting (the training set). A technique to mitigate overfitting is called Regularization. Several regularization techniques exist, the popular and best performing technique implemented in this research include batch normalization, early stopping, and dropout.

Batch normalization

This is technique used to improve the speed, performance, and stability of a FFNN (also addressing the problem of *vanishing and exploding gradients (Ioffe et al., 2015)*). It does this by adding a zero-centering and normalizing as well as a scaling and shifting operation before or after the activation function of each hidden layer. This operation gives the model the capability to learn the optimal scale and mean required for each input at the hidden layer.

Early Stopping

Early stopping *(Ruder, 2016)* is most commonly used form of regularization in deep learning. The process of implementing early stopping involve training the model for a fixed number of epochs that is sufficient enough for the model to reach convergence, while periodically checkpointing to record the validation error or accuracy and to optionally save a copy of the model parameters to file at the end of certain specified number of epochs. The model is said to start overfitting when the parameters of the model has stopped improving over the best saved ones or when the validation error stops decreasing or start increasing as seen in Figure 7.5. At that point, the training must be stopped, indicating the *point of optimal model performance*.

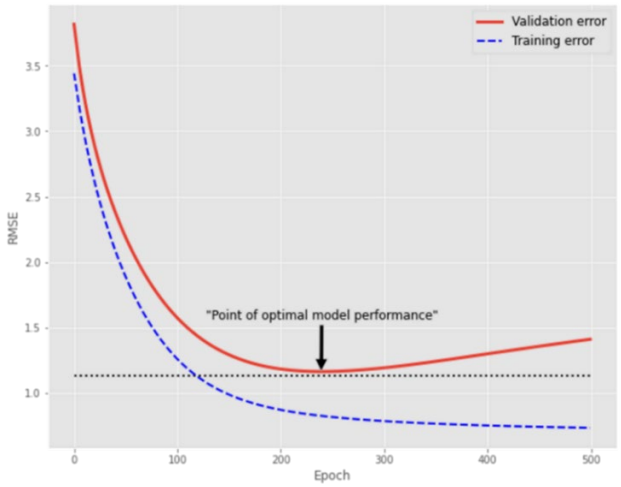


Figure 7.5: Early Stopping indication point of optimal model performance

Weight Regularization

This is a simple method to mitigate overfitting by penalizing or placing constraints on the parameters (i.e. weights) of the network such that they are forced to take smaller values and hence resulting in a simpler model that is less likely to overfit. The $\boldsymbol{\ell_2}$ **norm** *(Neyshabur et al., 2014)* and **weight decay** (similar to $\boldsymbol{\ell_2}$ **norm** but can be re-parameterized to become identical based on learning rate and purpose of implementation *(Loshchilov et al., 2017)*) are the most common regulation terms often added to the loss function during cost computation in order to regularize the weights. An $\boldsymbol{\ell_2}$ **norm** regularized cost function is given as Equation 7.7:

$$J_{reg} = \underbrace{J(w)}_{\text{Loss function}} + \underbrace{\lambda \sum_{i=1}^{N} \left\| W^{(i)} \right\|^2}_{\text{Regularization term}} \tag{7.7}$$

Dropout: To a first approximation, dropout *(Srivastava et al., 2014)* is a simple and powerful method of reducing overfitting or to reduce the problem of high variance and improve model generalization beyond the training set. The intuition behind dropout is that, at every training step, a random subset of every neuron (excluding the output neurons) in the network are either temporarily dropped or ignored with a probability or *drop rate* of $\boldsymbol{p}$ or kept with a probability of $\boldsymbol{1-p}$ after training as illustrated in Figure 7.6. The hyperparameter $\boldsymbol{p}$ is usually set between $\boldsymbol{10\% - 50\%}$ and mostly $30\% - 50\%$ for FFNN and its variants. The technique has proven to reduce overfitting in a variety of problems involving image classification, image segmentation, etc.

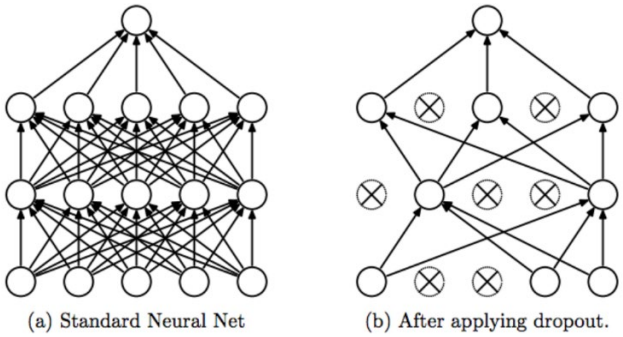


Figure 7.6: (a) A standard FFNN with two hidden layers and, (b) The same FFNN with dropped units. The units marked "×" have been dropped or turned off *(Srivastava et al., 2014)*

7.4.2.5 Convolutional Neural Networks (CNN) Architecture

Convolutional Neural Networks (CNN) are special case of feedforward neural networks designed for problems where the input data has a known grid-like topology, e.g. audio waveform and time-series data (1D-topology), image data (2D-topology), volumetric data e.g. CT scans or video data (3D-topology). A Convolutional network is simply feedforward neural networks that uses convolution instead of general matrix multiplication in at least one of its layers.

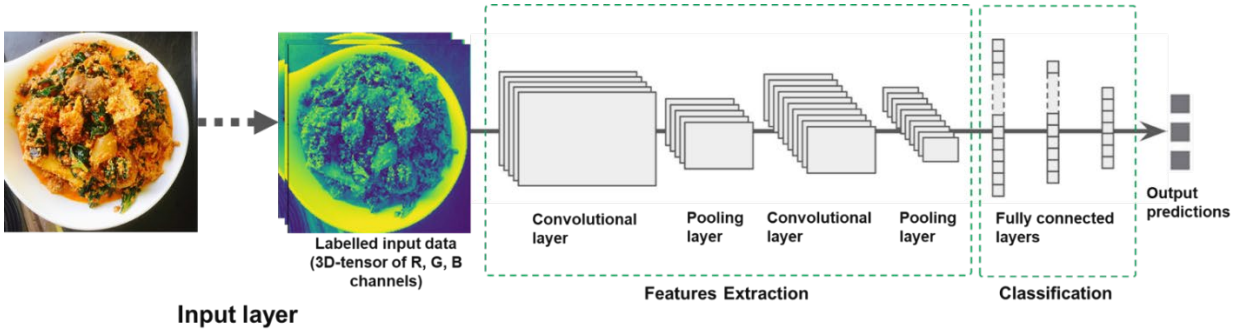


Figure 7.7: A typical Convolutional Neural Networks (CNN) Architecture

As illustrated in Figure 7.7, the CNN architecture used for building the food image classification model was made up of four major building blocks or layers, namely: input layer, convolutional layer, activation layer, and pooling layer.

a. **Input layer**: The input layer of the CNN architecture was a 3D tensor $(2D\ image \times 3\ channels)$ or simply an input image volume (a matrix of pixel values) with dimensions: $[width \times height \times depth]$ e.g. $[32 \times 32 \times 3]$, where the depth represents the RGB channels of the image. Unlike FFNN, the input variables

(pixel values) were not vectorized. If this was done, a lot of information in the input image would be lost. Conversely, the CNN preserves key information by representing input image as a tensor.

b. **Convolutional layer**: The convolutional layer consisted of a set of small size grid-like filters (or kernels) whose parameters need to be learned. Each filter was convolved with the input tensor to generate an activation map or feature map made of neurons. This was achieved by sliding the filter starting from the top-left corner horizontally and vertically across the input and then compute the dot products between the filter and the input tensor at every spatial position. The output tensor of the convolutional layer was then obtained by stacking the activation maps of all the filters along the depth axis. The convolution layer leveraged two key ideas namely: (i) Local connectivity, and (ii) Parameter sharing.

i. **Local connectivity**: Also referred to as ***Sparse interaction***. As illustrated in Figure 7.8, is a property that each hidden unit or neuron in the output activation map is sparely or scarcely connected only to a small local patch of the input tensor called ***receptive field*** (unlike in feedforward neural network where all the neurons are fully or densely connected), which is usually of same size as the filter. Local connectivity helped to reduce the number of parameters for the model and helped improve its statistical efficiency. The position of each neuron in the activation map is directly related to the position of the local patch. This implies that, if a neuron is moved by one or two pixels (called ***Stride***), the corresponding patch also moves to the right by one or two strides. However, for neurons on the border, the corresponding patch is partly located outside the input image and hence, results in shrinkage in the size of the output. This case where the size of the output is less than that of the input is referred to as ***valid convolution***. If the network can contain as many convolutional layers as the available computing resources can support, a technique called ***zero-padding*** is applied to fill the missing pixel on the border to keep the size of the output equal to the size of the input. The is referred to as ***same convolution***.

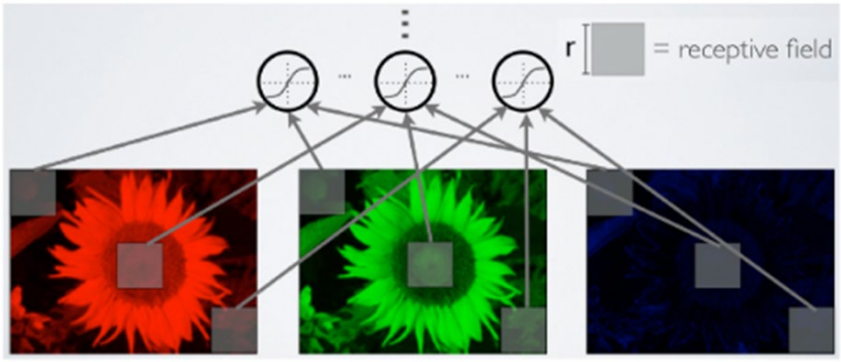


Figure 7.8: Illustration of Local connectivity, a property that each hidden unit or neuron in the output activation map is sparely connected only to a small local patch of the input tensor *(William L., 2019).*

ii. **Parameter sharing**: Unlike a feedforward neural networks where all the neurons are fully connected, in convNets, the parameters (weights) within a filter can be reused or shared across the neurons in the same feature map as illustrated in Figure 7.9. In other words, this implies that the weights of the connections between certain hidden units in a given hidden layer and the receptive field input from the

previous layer will be the same. The intuition here is that, if for instance, the same pattern is present, say at the top-left and at bottom-left receptive fields of an input image, then it makes sense to apply the same weight (filter) on such receptive fields to compute their corresponding activations. Parameter sharing helped to reduce more parameter space and hence, memory requirement. It also resulted in a nice translation property called ***equivariance*** which helped the model to generalize edge, texture, shape detection in different locations.

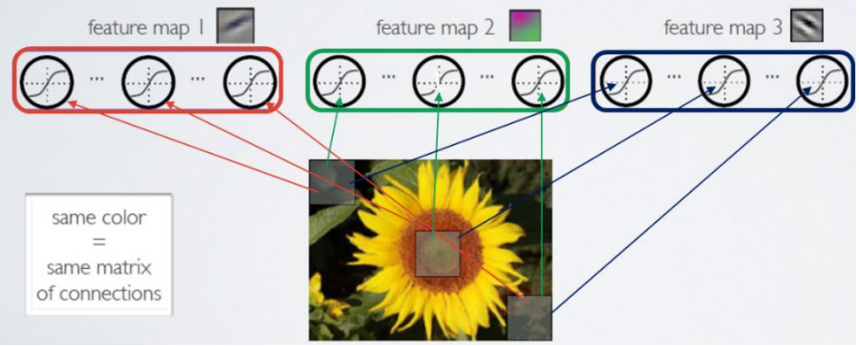
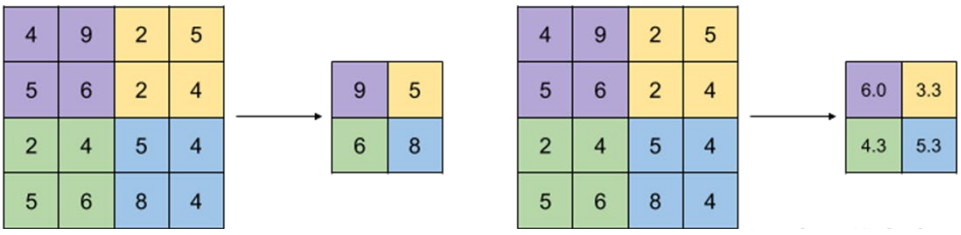


Figure 7.9: Parameter sharing illustrating that parameters (weights) within a filter can be reused or shared across the neurons in the same feature map *(William L., 2019)*

c. **Activation layer**: After each convolution operation in the convolutional layer, the activation of the feature map is computed in order to introduce non-linearity (e.g. tanh, ReLU)*.*. The ReLU function used in this study returns $\boldsymbol{x}$ for all values of $\boldsymbol{x} > \mathbf{0}$, and returns 0 for all values of $\boldsymbol{x} \leq \mathbf{0}$.

d. **Pooling layer**: In the pooling layer illustrated in Figure 7.10, the output of the convolutional layer is further modified. The pooling operation is performed to progressively reduce the spatial dimensions of the feature map, while still preserving the most critical feature information. This in turn helped reduced the number of parameters as well as mitigated overfitting. There are two variants of pooling function namely: maximum pooling (max pooling) or average pooling. Max pooling (used in this study) computes the maximum, or largest value in each patch of each feature map while average pooling computes the average value.



(a) Max pooling (b) Avg pooling

Figure 7.10: Variations of Pooling Function

CNN architectures fundamentally consist of several stacks of the four building blocks e.g. [Convolutional layer (+ReLU) + Pooling layer + Convolutional layer (+ReLU) + Pooling layer, + … + Fully connected layers (+ReLU) + Prediction layer (Softmax)]. In recent years, several sophisticated variants of this fundamental architecture have evolved and has brought about countless cutting-edge achievements in many fields. The popularity of GPU computing and birth of the

open source ImageNet dataset have inspired continuous progress in deep learning researches and its applications. In this study, five of the popular architectures namely: VGG16, VGG19, ResNet101V2, InceptionV3, and Xception were explored for building the classification model utilized in the food recognition system.

7.4.2.6 Pre-trained models and transfer learning for deep CNNs

It is currently a no-brainer that data is the new oil. The huge success of deep learning and convolutional neural networks today is considered to have been fueled by the massive availability of training data and the technological advancement in the computing resources for processing and generating data. However, this success is accompanied by expensive and tedious challenges such as data collection strategy, data cleaning, data labelling, and high cost computing resources especially in tasks like object recognition using deep convolutional neural network architectures. Hence, this begs the question: "*In order to minimize these efforts, how much data is enough for training a deep neural network?".* One state-of-the-art approach that has successfully provided a ubiquitous answer is ***Transfer Learning*** *(Pan et al., 2009)***.** To this end, deep neural networks need not to be trained from the scratch.

It is now a common practice to reuse an existing deep neural network (CNNs) that accomplishes a similar task to the target task and has been pre-trained on large dataset, e.g. *ImageNet, which contained 1.2 million images with 1000 categories,* (either as an initialization or a fixed feature extractor) for the task of interest (often with smaller dataset and less computational power requirement). In this study, five pre-trained models namely: VGG16, VGG19, ResNet101V2, InceptionV3, and Xception were experimented on in order to select the best model to be utilized as the classification model in the food recognition system.

There are two major approaches to implementing transfer learning as follows:

a. **Fixed feature extractor**: This involve removing the last fully-connected (FC) layer of a desired pre-trained CNN, i.e. the output layer as illustrated in Figure 7.7 above and the rest of the layers can be trained to serve as a fixed feature extractor for new database.

b. **Fine-tuning**: This approach involves first removing the fully-connected layer and replacing it with a new fully-connected layer(s) that matches the number of classes in the target dataset as illustrated in Figure 7.7. Next, the weights of the new fully-connected layer is initialized randomly while the rest of the layers of the pre-trained model is being frozen (making the weights non-trainable). The fully-connected layer can then be trained mildly (warmed up) while the performance of the model is being monitored. Based on the size of the available training set, one or two top (or all other) hidden layers can further be unfrozen and the training continued to give room for backpropagation to adjust their weights to improve the performance of the model. It is often a useful step to reduce the learning rate when unfreezing the hidden layers in order to improve model convergence.

7.4.2.7 CNN architectures

1. **VGGNet** *(Simonyan et al. (2014))*

Developed in the Visual Geometry Group (VGG) research lab at Oxford University, VGGNet was the runner-up in the ILSVRC 2014 ImageNet challenge. It had a very simple yet efficient structure with multiple sets of 2 and 3 block of

convolutional and poling layers amounting to 16 or 19 convolutional layers depending on the version of the network (VGG16 or VGG19). The model achieved a top-5 error rate of 7.3% on the ImageNet dataset.

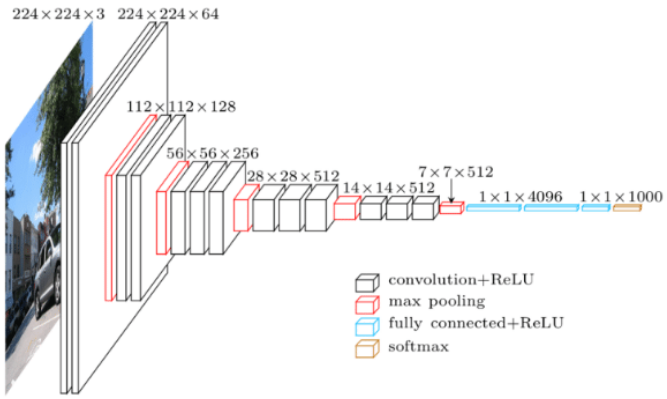


Figure 7.11: VGGNet Architecture

2. **ResNet** *(He et al. (2016))*

Kaiming He et al of Microsoft won the 2015 ILSVRC ImageNet competition with an amazing top-5 error rate of 3.57% (which was considered better than human classification accuracy *(Russakovsky et al., 2015))* with their Residual Network (ResNet). The basic building block for ResNets are the conv and identity blocks. The novelty in their network is the use of batch normalization and *Skip* or *shortcut connections* to bypass the input to the next layer when training deeper architectures. Each residual unit consist of 2 convolutional layers (without pooling), with batch normalization and ReLU activation, using 3 × 3 filters and preserving spatial dimensions (stride 1, "same" padding).

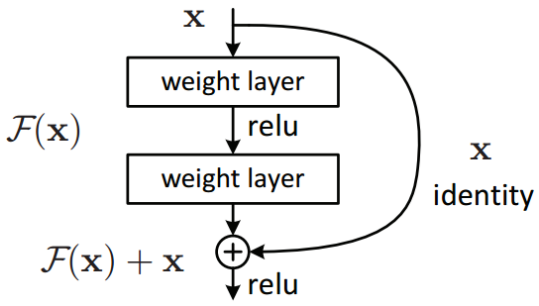


Figure 7.12: A building block of ResNet Architecture

3. **GoogLeNet (Inception)** *Szegedy et al. (2015)*

The inception architecture (as illustrated in Figure 7.13) was developed by researchers at Google and it won the 2014 ILSVRC ImageNet competition with a top-5 error rate of 7%. The model is comprised of a basic unit called *Inception module* where series of convolution are performed, and their results aggregated. This helps to drastically reduce the number of parameters and hence, reduce computing requirement. In order to reduce computational demand, 1 × 1 convolutions are used to reduce the depth of the input channel. For each of the inception module, a set of 1 × 1, 3 × 3, and 5 × 5 convolutional filters are learned which has the capacity to extract features at different scales from the input.

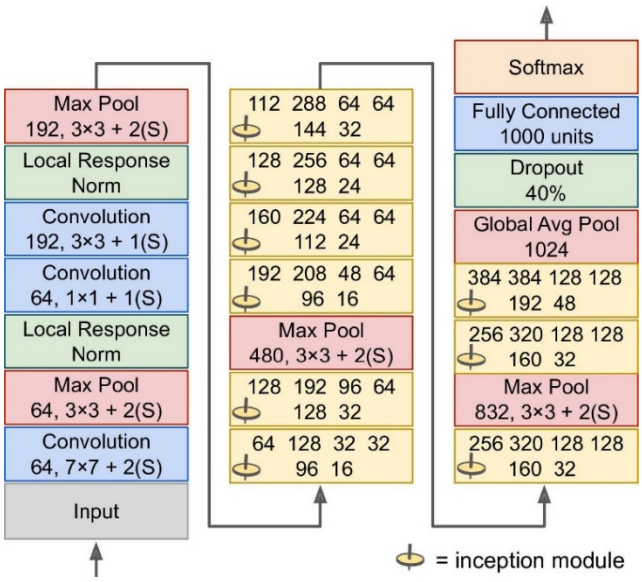


Figure 7.13: GoogleNet Architecture *(Géron, 2019)*

4. **Xception** *(Chollet, 2017)*

An extension of the inception architecture called Xception (Extreme Inception) was proposed by François Chollet, the author of the Keras library. Xception combines the ideas of GoogLeNet and ResNet but it was designed to replace the standard inception module in the inception architecture with depth-wise separable convolution layer. Regular convolutional layer uses filters that simultaneously capture spatial patterns and cross-channel patterns; however, a separable convolutional layer assumes that spatial patterns and cross-channel patterns can be modeled distinctly.

7.5. APPENDIX E: REPRODUCIBILITY AND RESOURCE AVAILABILITY

The results presented in this thesis has high degree of reproducibility and replicability given the hypothesis, original data, codes and computational resources. In addition, similar or consistent results can be obtained looking at the same scientific question but with different data. The image dataset that supported the findings in this study are available for public download via the following cloud storage API:

- https://food-image-dataset.s3.amazonaws.com/Composite_food_dataset.zip

All Python scripts used in this study for the deep learning model development (selection, training, fine-tuning, validation, and testing) and diet quality assessment of the 15 food categories considered are available on file via the following cloud storage APIs:

(i) Model development script: https://food-image-dataset.s3.amazonaws.com/Model+development+script.py

(ii) Food image recognition and diet quality assessment script: https://food-image-dataset.s3.amazonaws.com/Food+recognition+and+diet+quality+assessment+script.py

6 References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., . . . Isard, M. (2016). *Tensorflow: A system for large-scale machine learning.* Paper presented at the 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16).

Abramovitch, S. L., Reddigan, J. I., Hamadeh, M. J., Jamnik, V. K., Rowan, C. P., & Kuk, J. L. (2012). Underestimating a serving size may lead to increased food consumption when using Canada's Food Guide. *Applied Physiology, Nutrition, and Metabolism, 37*(5), 923-930.

Acharya, S. D., Elci, O. U., Sereika, S. M., Styn, M. A., & Burke, L. E. (2011). Using a Personal Digital Assistant for Self-monitoring Influences Diet Quality in Comparison to a Standard Paper Record among Overweight/obese Adults. *Journal of the American Dietetic Association, 111*(4), 583-588. doi:10.1016/j.jada.2011.01.009

Afshin, A., Sur, P. J., Fay, K. A., Cornaby, L., Ferrara, G., Salama, J. S., . . . Murray, C. J. L. (2019). Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The lancet, 393*(10184), 1958-1972. doi:https://doi.org/10.1016/S0140-6736(19)30041-8

AFSSA. (2008). Setting of Nutrient Profiles for Accessing Nutrition and Health Claims: Proposals and Arguments. Retrieved from https://www.anses.fr/en/system/files/NUT-Ra-ProfilsEN.pdf

Aguilar, E., Bolaños, M., & Radeva, P. (2017). *Food recognition using fusion of classifiers based on CNNs.* Paper presented at the International Conference on Image Analysis and Processing.

Ahmed, K. T., Ummesafi, S., & Iqbal, A. (2019). Content based image retrieval using image features information fusion. *Information Fusion, 51*, 76-99. doi:https://doi.org/10.1016/j.inffus.2018.11.004

Ahn, D., Choi, J.-Y., Kim, H.-C., Cho, J.-S., Moon, K.-D., & Park, T. (2019). Estimating the Composition of Food Nutrients from Hyperspectral Signals Based on Deep Neural Networks. *Sensors (Basel, Switzerland), 19*(7), 1560. doi:10.3390/s19071560

Ainaa, F., Poh, B., Nik Shanita, S., & Wong, J. (2018). Feasibility of reviewing digital food images for dietary assessment among nutrition professionals. *Nutrients, 10*(8), 984.

Allman-Farinelli, M., & Gemming, L. (2017). Technology Interventions to Manage Food Intake: Where Are We Now? *Current Diabetes Reports, 17*(11). doi:10.1007/s11892-017-0937-5

Almaghrabi, R., Villalobos, G., Pouladzadeh, P., & Shirmohammadi, S. (2012). *A novel method for measuring nutrition intake based on food image* (Vol. null).

Ambrosini, G. L., Hurworth, M., Giglia, R., Trapp, G., & Strauss, P. (2018). Feasibility of a commercial smartphone application for dietary assessment in epidemiological research and comparison with 24-h dietary recalls. *Nutrition Journal, 17*(1), 5. doi:10.1186/s12937-018-0315-4

Amft, O., Kusserow, M., & Troster, G. (2009). Bite Weight Prediction From Acoustic Recognition of Chewing. *IEEE Transactions on Biomedical Engineering, 56*(6), 1663-1672. doi:10.1109/TBME.2009.2015873

Amoutzopoulos, B., Steer, T., Roberts, C., Cade, J. E., Boushey, C. J., Collins, C. E., . . . Page, P. (2018). Traditional methods v. new technologies - dilemmas for dietary assessment in large-scale nutrition surveys and studies: a report following an international panel discussion at the 9th International Conference on Diet and Activity Methods (ICDAM9), Brisbane, 3 September 2015. *Journal of nutritional science, 7*, e11-e11. doi:10.1017/jns.2018.4

Anaconda, I. (2018). Conda documentation. In.

Andreoli, A., De Lorenzo, A., Cadeddu, F., Iacopino, L., & Grande, M. (2011). New trends in nutritional status assessment of cancer patients. *Eur Rev Med Pharmacol Sci, 15*(5), 469-480.

Anthimopoulos, M., Dehais, J., Shevchik, S., Ransford, B. H., Duke, D., Diem, P., & Mougiakakou, S. (2015). Computer Vision-Based Carbohydrate Estimation for Type 1 Patients With Diabetes Using Smartphones. *Journal of Diabetes Science and Technology, 9*(3), 507-515. doi:10.1177/1932296815580159

Anthimopoulos, M., Gianola, L., Scarnato, L., Diem, P., & Mougiakakou, S. (2014). *A Food Recognition System for Diabetic Patients Based on an Optimized Bag-of-Features Model* (Vol. 18).

Arne, H. (2019). Number of smartphone users worldwide from 2016 to 2021. Retrieved from https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

Assessment, D. (2018). A resource guide to method selection and application in low resource settings. *FAO: Rome, Italy*.

Attokaren, D. J., Fernandes, I. G., Sriram, A., Murthy, Y. S., & Koolagudi, S. G. (2017). *Food classification from images using convolutional neural networks.* Paper presented at the TENCON 2017-2017 IEEE Region 10 Conference.

Ball, J. W., Dains, J. E., Flynn, J. A., Solomon, B. S., & Stewart, R. W. (2014). *Seidel's Physical Examination Handbook-E-Book*: Elsevier Health Sciences.

Baranowski, T., Islam, N., Baranowski, J., Cullen, K. W., Myres, D., & Marsh, T. (2002). The food intake recording software system is valid among fourth-grade children. *Journal of the American Dietetic Association, 102*(3), 380-385.

Baranowski, T., Islam, N., Douglass, D., Dadabhoy, H., Beltran, A., Baranowski, J., . . . Subar, A. F. (2014). Food Intake Recording Software System, version 4 (FIRSSt4): a self-completed 24-h dietary recall for children. *Journal of Human Nutrition and Dietetics, 27*, 66-71. doi:10.1111/j.1365-277X.2012.01251.x

Beijbom, O., Joshi, N., Morris, D., Saponas, S., & Khullar, S. (2015). *Menu-match: Restaurant-specific food logging from images.* Paper presented at the 2015 IEEE Winter Conference on Applications of Computer Vision.

Bishop, C. M. (2006). *Pattern recognition and machine learning*: springer.

Bisong, E. (2019). Google Colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform* (pp. 59-64): Springer.

Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., . . . Zhang, J. (2016). End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.

Bolaños, M., Ferrà, A., & Radeva, P. (2017). *Food ingredients recognition through multi-label learning.* Paper presented at the International Conference on Image Analysis and Processing.

Bootkrajang, J., Chawachat, J., & Trakulsanguan, E. (2020) Deep-based openset classification technique and its application in novel food categories recognition. In*: Vol. 977. Advances in Intelligent Systems and Computing* (pp. 235-245): Springer Verlag.

Bossard, L., Guillaumin, M., & Van Gool, L. (2014). *Food-101-mining discriminative components with random forests* (Vol. null).

Boushey, C., Spoden, M., Zhu, F., Delp, E., & Kerr, D. (2017). New mobile methods for dietary assessment: Review of image-assisted and image-based dietary assessment methods. *Proceedings of the Nutrition Society, 76*(3), 283-294.

Boushey, C. J., Harray, A. J., Kerr, D. A., Schap, T. E., Paterson, S., Aflague, T., . . . Delp, E. J. (2015). How willing are adolescents to record their dietary intake? The mobile food record. *JMIR mHealth and uHealth, 3*(2), e47.

Bradski, G. (2000). The opencv library. *Dr Dobb's J. Software Tools, 25*, 120-125.

Burke, B. S. (1947). The dietary history as a tool in research. *Journal of the American Dietetic Association, 23*, 1041-1046.

Burrows, T. L., & Rollo, M. E. (2019). Advancement in dietary assessment and self-monitoring using technology. In: Multidisciplinary Digital Publishing Institute.

Burrows, T. L., Rollo, M. E., Williams, R., Wood, L. G., Garg, M. L., Jensen, M., & Collins, C. E. (2017). A systematic review of technology-based dietary intake assessment validation studies that include carotenoid biomarkers. *Nutrients, 9*(2), 140.

Cade, J. E. (2017). Measuring diet in the 21st century: use of new technologies. *Proceedings of the Nutrition Society, 76*(3), 276-282. doi:10.1017/s0029665116002883

Carbs & Cals. (2018). Carbs & Cals. Retrieved from https://www.carbsandcals.com/app/app

CDC. (2006). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Questionnaire (or examination protocol, or laboratory protocol). *http://www. cdc. gov/nchs/nhanes. htm*.

Chen, J., L., A., B., R., H., & M., A. F. (2017). The use of smartphone health apps and other mobile health (mHealth) technologies in dietetic practice: a three country study. *Journal of Human Nutrition and Dietetics, 30*(4), 439-452. doi:doi:10.1111/jhn.12446

Chen, J., & Ngo, C.-W. (2016). *Deep-based ingredient recognition for cooking recipe retrieval.* Paper presented at the Proceedings of the 24th ACM international conference on Multimedia.

Chen, M.-Y., Yang, Y.-H., Ho, C.-J., Wang, S.-H., Liu, S.-M., Chang, E., . . . Ouhyoung, M. (2012). *Automatic chinese food identification and quantity estimation.* Paper presented at the SIGGRAPH Asia 2012 Technical Briefs.

Chen, P. Y., Blutinger, J. D., Meijers, Y., Zheng, C., Grinspun, E., & Lipson, H. (2019). Visual modeling of laser-induced dough browning. *Journal of Food Engineering, 243*, 9-21.

Cheng, G., & Han, J. (2016). A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing, 117*, 11-28.

Chollet, F. (2015). Keras: Deep learning library for theano and tensorflow.(2015). *There is no corresponding record for this reference*.

Chollet, F. (2017). *Xception: Deep learning with depthwise separable convolutions.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Choras, R. (2007). Image feature extraction techniques and their applications for CBIR and biometrics systems. *1*(1), 6-16.

Christodoulidis, Anthimopoulos, M., & Mougiakakou, S. (2015a). *Food recognition for dietary assessment using deep convolutional neural networks.* Paper presented at the International Conference on Image Analysis and Processing.

Christodoulidis, Anthimopoulos, M., & Mougiakakou, S. (2015b). *Food recognition for dietary assessment using deep convolutional neural networks* (Vol. null).

Ciocca, Napoletano, P., & Schettini, R. (2017). Food Recognition: A New Dataset, Experiments, and Results. *IEEE Journal of Biomedical and Health Informatics, 21*(3), 588-598. doi:10.1109/JBHI.2016.2636441

Ciocca, G., Napoletano, P., & Schettini, R. (2017). *Learning CNN-based features for retrieval of food images.* Paper presented at the International Conference on Image Analysis and Processing.

Ciocca, G., Napoletano, P., & Schettini, R. (2018). CNN-based features for retrieval and classification of food images. *Computer Vision and Image Understanding, 176-177*, 70-77. doi:https://doi.org/10.1016/j.cviu.2018.09.001

Comrie, F., Masson, L. F., & McNeill, G. (2009). A novel online Food Recall Checklist for use in an undergraduate student population: a comparison with diet diaries. *Nutrition Journal, 8*(1), 13.

Dalal, N., & Triggs, B. (2005). *Histograms of oriented gradients for human detection.* Paper presented at the Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.

Darmon, N., Maillot, M., Darmon, M., & Drewnowski, A. (2005). A nutrient density standard for vegetables and fruits: nutrients per calorie and nutrients per unit cost. *Journal of the American Dietetic Association, 105*(12), 1881-1887.

Darmon, N., Vieux, F., Maillot, M., Volatier, J.-L., & Martin, A. (2009). Nutrient profiles discriminate between foods according to their contribution to nutritionally adequate diets: a validation study using linear programming and the SAIN, LIM system. *The American Journal of Clinical Nutrition, 89*(4), 1227-1236.

Dehais, J., Anthimopoulos, M., Shevchik, S., & Mougiakakou, S. J. I. t. o. m. (2017). Two-view 3d reconstruction for food volume estimation. *19*(5), 1090-1099.

Domhardt, M., Tiefengrabner, M., Dinic, R., Fötschl, U., Oostingh, G. J., Stütz, T., . . . Ginzinger, S. W. (2015). Training of Carbohydrate Estimation for People with Diabetes Using Mobile Augmented Reality. *9*(3), 516-524. doi:10.1177/1932296815578880

Drewnowski, & Fulgoni, V. (2008). Nutrient profiling of foods: creating a nutrient-rich food index. *Nutrition Reviews, 66*(1), 23-39.

E Silva, B. V. R., Rad, M. G., Cui, J., McCabe, M., & Pan, K. (2018). A Mobile-Based Diet Monitoring System for Obesity Management. *Journal of health & medical informatics, 9*(2), 307. doi:10.4172/2157-7420.1000307

Easy Diet Diary. (2018). Easy Diet Diary. Retrieved from https://easydietdiary.com/

Eldridge, A. L., Piernas, C., Illner, A.-K., Gibney, M. J., Gurinović, M. A., De Vries, J. H., & Cade, J. E. (2019). Evaluation of new technology-based tools for dietary intake assessment—An ilsi europe dietary intake and exposure task force evaluation. *Nutrients, 11*(1), 55.

Farinella, G. M., Allegra, D., Moltisanti, M., Stanco, F., & Battiato, S. (2016). Retrieval and classification of food images. *CBM Computers in Biology and Medicine, 77*, 23-39.

Fatkun. (2019). Fatkun Batch Download Image–Pro. Retrieved from https://chrome.google.com/webstore/detail/fatkun-batch-download-ima/nnjjahlikiabnchcpehcpkdeckfgnohf

Finucane, M. M., Stevens, G. A., Cowan, M. J., Danaei, G., Lin, J. K., Paciorek, C. J., . . . Bahalim, A. N. (2011). National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9· 1 million participants. *The lancet, 377*(9765), 557-567.

Forster, H., Fallaize, R., Gallagher, C., O?Donovan, B. C., Woolhead, C., Walsh, C. M., . . . Gibney, R. E. (2014). Online Dietary Intake Estimation: The Food4Me Food Frequency Questionnaire. *J Med Internet Res, 16*(6), e150. doi:10.2196/jmir.3105

Forster, H., Walsh, M. C., Gibney, M. J., Brennan, L., & Gibney, E. R. (2016). Personalised nutrition: the role of new dietary assessment methods. *Proceedings of the Nutrition Society, 75*(1), 96-105. doi:10.1017/s0029665115002086

Fowles, E. R., & Gentry, B. (2008). The feasibility of personal digital assistants (PDAs) to collect dietary intake data in low-income pregnant women. *Journal of nutrition education and behavior, 40*(6), 374-377.

Fulgoni, V., Keast, D. R., & Drewnowski, A. (2009). Development and validation of the nutrient-rich foods index: a tool to measure nutritional quality of foods. *The Journal of nutrition, 139*(8), 1549-1554.

Gazibarich, B., & Ricci, P. (1998). Towards better food choices: the nutritious food index. *Australian Journal of Nutrition and Dietetics*.

Gemming, Rush, E., Maddison, R., Doherty, A., Gant, N., Utter, J., & Mhurchu, C. N. (2015). Wearable cameras can reduce dietary under-reporting: doubly labelled water validation of a camera-assisted 24 h recall. *British Journal of Nutrition, 113*(2), 284-291.

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*: O'Reilly Media.

Géron, A. l. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems* [1 online resource (xx, 547 pages) : illustrations](First edition. ed.).

Gibson, R. S. (2005). *Principles of Nutrition Assessment* (Vol. null).

Gilhooly, C. H. (2017). Are Calorie Counting Apps Ready to Replace Traditional Dietary Assessment Methods? *Nutrition Today, 52*(1), 10-18. doi:10.1097/NT.0000000000000188

Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing*.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*: MIT press.

Hamid, H., Guido, M., Paolo, C., Ilaria, D. M., Monica, M., & Stefano, C. (2016). *Food Image Recognition Using Very Deep Convolutional Networks*. Paper presented at the Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, Amsterdam, The Netherlands. https://doi.org/10.1145/2986035.2986042

Hanley-Cook, G. T., Tung, J. Y. A., Sattamini, I. F., Marinda, P. A., Thong, K., Zerfu, D., . . . Lachat, C. K. (2020). Minimum Dietary Diversity for Women of Reproductive Age (MDD-W) Data Collection: Validity of the List-Based and Open Recall Methods as Compared to Weighed Food Record. *Nutrients, 12*(7), 2039.

Hassannejad, Matrella, G., Ciampolini, P. D., Munari, I., Mordonini, M., & Cagnoni, S. (2016). *Food image recognition using very deep convolutional networks* (Vol. null).

Hassannejad, H., Matrella, G., Ciampolini, P., Munari, I. D., Mordonini, M., & Cagnoni, S. (2016). *Food Image Recognition Using Very Deep Convolutional Networks*. Paper presented at the Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, Amsterdam, The Netherlands.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.* Paper presented at the Proceedings of the IEEE international conference on computer vision.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

He, Y., Xu, C., Khanna, N., Boushey, C. J., & Delp, E. J. (2013). FOOD IMAGE ANALYSIS: SEGMENTATION, IDENTIFICATION AND WEIGHT ESTIMATION. *Proceedings. IEEE International Conference on Multimedia and Expo, 2013*, 10.1109/ICME.2013.6607548. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5448794/

Health Canada. (2015, Feb, 2018). Canadian Nutrient File (CNF). *Canadian Nutrient File (CNF) - Search by food.* Retrieved from https://food-nutrition.canada.ca/cnf-fce/index-eng.jsp

HealthWatch 360. (2018). HealthWatch 360. Retrieved from http://www.gbhealthwatch.com/healthwatch360-app

Hemalatha, R., Kumari, S., Muralidharan, V., Gigoo, K., & Thakare, B. S. (2020) Literature Survey—Food Recognition and Calorie Measurement Using Image Processing and Machine Learning Techniques. In*: Vol. 570. 2nd International Conference on Communications and Cyber-Physical Engineering, ICCCE 2019* (pp. 23-37): Springer Verlag.

Heravi, E. J., Aghdam, H. H., & Puig, D. (2015). *A deep convolutional neural network for recognizing foods.* Paper presented at the Eighth International Conference on Machine Vision (ICMV 2015).

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in science & engineering, 9*(3), 90.

Hutchesson, M. J., Rollo, M. E., Callister, R., & Collins, C. E. (2015). Self-monitoring of dietary intake by young women: online food records completed on computer or smartphone are as accurate as paper-based food records but more acceptable. *Journal of the Academy of Nutrition and Dietetics, 115*(1), 87-94.

IBM Research Lab. (2018). Quantum Computing. Retrieved from http://research.ibm.com/5-in-5/quantum-computing/

Illner, A. K., Freisling, H., Boeing, H., Huybrechts, I., Crispim, S. P., & Slimani, N. (2012). Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. *International Journal of Epidemiology, 41*(4), 1187-1203. doi:10.1093/ije/dys105

Indumathi, D., & KP, D. (2017). NUTRITION: RISK FREE ANALYSIS TOOL. *International Journal of Curren*.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Jospe, M. R., Fairbairn, K. A., Green, P., & Perry, T. L. (2015). Diet App Use by Sports Dietitians: A Survey in Five Countries. *JMIR mHealth and uHealth, 3*(1), e7. doi:10.2196/mhealth.3345

Kagaya, H., Aizawa, K., & Ogawa, M. (2014). *Food Detection and Recognition Using Convolutional Neural Network.* Paper presented at the Proceedings of the 22nd ACM international conference on Multimedia, Orlando, Florida, USA. https://doi.org/10.1145/2647868.2654970

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70-90.

Kawano, Y., & Yanai, K. (2014). *Automatic expansion of a food image dataset leveraging existing categories with domain adaptation* (Vol. null).

Kennedy, E., Racsa, P., Dallal, G., Lichtenstein, A. H., Goldberg, J., Jacques, P., & Hyatt, R. (2008). Alternative approaches to the calculation of nutrient density. *Nutrition Reviews, 66*(12), 703-709.

Khanna, N., Boushey, C. J., Kerr, D., Okos, M., Ebert, D. S., & Delp, E. J. (2010). An Overview of The Technology Assisted Dietary Assessment Project at Purdue University. *ISM ... : ... IEEE International Symposium on Multimedia ... : proceedings. IEEE International Symposium on Multimedia*, 290-295. doi:10.1109/ISM.2010.50

Klurfeld, D. M., Hekler, E. B., Nebeker, C., Patrick, K., & Khoo, C. S. H. (2018). Technology Innovations in Dietary Intake and Physical Activity Assessment: Challenges and Recommendations for Future Directions. *American journal of preventive medicine, 55*(4), e117-e122.

Koegel, K. L., Kuczynski, K. J., & Britten, P. (2013). Addendum to the MyPyramid Equivalents Database 2.0. *Procedia Food Science, 2*, 75-80. doi:https://doi.org/10.1016/j.profoo.2013.04.012

Koen, N., Blaauw, R., & Wentzel-Viljoen, E. (2016). Food and nutrition labelling: the past, present and the way forward. *South African Journal of Clinical Nutrition, 29*(1), 13-21. doi:10.1080/16070658.2016.1215876

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet classification with deep convolutional neural networks* (Vol. null).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.

Lee, R. D., & Nieman, D. C. (1996). Nutritional assessment.

Lee, R. D., & Nieman, D. C. (2013). *Nutritional assessment*. New York, NY: McGraw-Hill.

Liang, H., Gao, Y., Sun, Y., Sun, X. J. I. J. o. C., & Applications. (2018). CEP: calories estimation from food photos. 1-9.

Liang, Y., & Li, J. (2017). Deep Learning-Based Food Calorie Estimation Method in Dietary Assessment. *arXiv preprint arXiv:1706.04062*.

Lindholm, A., Wahlström, N., Lindsten, F., & Schön, T. B. (2019). Supervised Machine Learning.

Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., & Ma, Y. (2016). *Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment.* Paper presented at the International Conference on Smart Homes and Health Telematics.

Lo, F., Sun, Y., Qiu, J., & Lo, B. J. N. (2018). Food Volume Estimation Based on Deep Learning View Synthesis from a Single Depth Map. *10*(12), 2005.

LoseIt. (2018). LoseIt. Retrieved from https://www.loseit.com

Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Lowe, D. G. (1999). *Object recognition from local scale-invariant features.* Paper presented at the iccv.

Lowe, D. G. (2001). *Local feature view clustering for 3D object recognition.* Paper presented at the Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001.

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision, 60*(2), 91-110. doi:10.1023/B:VISI.0000029664.99615.94

Lu, Y., Allegra, D., Anthimopoulos, M., Stanco, F., Farinella, G. M., & Mougiakakou, S. (2018). *A multi-task learning approach for meal assessment.* Paper presented at the Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management.

Ma, Y., Olendzki, B. C., Pagoto, S. L., Hurley, T. G., Magner, R. P., Ockene, I. S., . . . Hébert, J. R. (2009). Number of 24-Hour Diet Recalls Needed to Estimate Energy Intake. *Annals of Epidemiology, 19*(8), 553-559. doi:10.1016/j.annepidem.2009.04.010

Maillot, M., Darmon, N., Darmon, M., Lafay, L., & Drewnowski, A. (2007). Nutrient-Dense Food Groups Have High Energy Costs: An Econometric Approach to Nutrient Profiling. *The Journal of nutrition, 137*(7), 1815-1820. doi:10.1093/jn/137.7.1815

Mandal, B., Puhan, N., & Verma, A. (2018). Deep Convolutional Generative Adversarial Network Based Food Recognition Using Partially Labeled Data. *IEEE Sensors Letters, PP*, 1-1. doi:10.1109/LSENS.2018.2886427

Martin, C. K., Correa, J. B., Han, H., Allen, H. R., Rood, J. C., Champagne, C. M., . . . Bray, G. A. (2012). Validity of the Remote Food Photography Method (RFPM) for estimating energy and nutrient intake in near real-time. *Obesity, 20*(4), 891-899.

Martinel, N., Foresti, G. L., & Micheloni, C. (2018). *Wide-slice residual networks for food recognition.* Paper presented at the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV).

Masset, G. (2012). *Predictive validity of WXYfm and SAIN, LIM food nutrient profiling models in the Whitehall II cohort.* UCL (University College London),

Matsuda, Y., Hoashi, H., & Yanai, K. (2012, 9-13 July 2012). *Recognition of Multiple-Food Images by Detecting Candidate Regions.* Paper presented at the 2012 IEEE International Conference on Multimedia and Expo.

Meyers, A., Johnston, N., Rathod, V., Korattikara, A., Gorban, A., Silberman, N., . . . Murphy, K. P. (2015). *Im2Calories: towards an automated mobile vision food diary.* Paper presented at the Proceedings of the IEEE International Conference on Computer Vision.

Mezgec, S., & Koroušić Seljak, B. (2017a). NutriNet: A Deep Learning Food and Drink Image Recognition System for Dietary Assessment. *Nutrients, 9*(7). doi:10.3390/nu9070657

Mezgec, S., & Koroušić Seljak, B. (2017b). NutriNet: A Deep Learning Food and Drink Image Recognition System for Dietary Assessment. *Nutrients, 9*(7), 657. doi:10.3390/nu9070657

Min, W., Jiang, S., Liu, L., Rui, Y., & Jain, R. (2019). A survey on food computing. *ACM Computing Surveys (CSUR), 52*(5), 1-36.

Montville, J. B., Ahuja, J. K., Martin, C. L., Heendeniya, K. Y., Omolewa-Tomobi, G., Steinfeldt, L. C., . . . Moshfegh, A. (2013). USDA food and nutrient database for dietary studies (FNDDS), 5.0. *Procedia Food Science, 2*, 99-112.

MyFitnessPal. (2018). MyFitnessPal. Retrieved from https://www.myfitnesspal.com/

MyPlate. (2018). MyPlate. Retrieved from http://www.livestrong.com/myplate/

National Cancer Institute, D. o. C. C. a. P. S. (2018a). ASA24-Kids. Retrieved from https://epi.grants.cancer.gov/asa24/respondent/childrens.html

National Cancer Institute, D. o. C. C. a. P. S. (2018b, 2018, Feb 06). Automated Self-Administered 24-Hour (ASA24®) Dietary Assessment Tool. Retrieved from https://epi.grants.cancer.gov/asa24/

National Cancer Institute, D. o. C. C. a. P. S. (2018c). Diet History Questionnaire (DHQ). Retrieved from https://epi.grants.cancer.gov/dhq3/

Neyshabur, B., Tomioka, R., & Srebro, N. (2014). In search of the real inductive bias: On the role of implicit regularization in deep learning. *arXiv preprint arXiv:1412.6614*.

Nguyen, D. T., Zong, Z., Ogunbona, P. O., Probst, Y., & Li, W. (2014). Food image classification using local appearance and global structural information. *Neurocomputing, 140*, 242-251. doi:https://doi.org/10.1016/j.neucom.2014.03.017

Nishimura, J., & Tadahiro, K. (2008, 7-9 May 2008). *Eating habits monitoring using wireless wearable in-ear microphone.* Paper presented at the 2008 3rd International Symposium on Wireless Pervasive Computing.

Noda, K., Yamaguchi, Y., Nakadai, K., Okuno, H. G., & Ogata, T. (2015). Audio-visual speech recognition using deep learning. *Applied Intelligence, 42*(4), 722-737.

O'Loughlin, G., Cullen, S. J., McGoldrick, A., O'Connor, S., Blain, R., O'Malley, S., & Warrington, G. D. (2013). Using a wearable camera to increase the accuracy of dietary analysis. *American journal of preventive medicine, 44*(3), 297-301.

Oh, K.-S., & Jung, K. (2004). GPU implementation of neural networks. *Pattern Recognition, 37*(6), 1311-1314.

Ojala, T., Pietikainen, M., Maenpaa, T. J. I. T. o. p. a., & intelligence, m. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *24*(7), 971-987.

Oliver, E., Baños, R. M., Cebolla, A., Lurbe, E., Alvarez-Pitti, J., & Botella, C. (2013). An electronic system (PDA) to record dietary and physical activity in obese adolescents; data about efficiency and feasibility. *Nutricion hospitalaria, 28*(6).

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering, 22*(10), 1345-1359.

Pandey, P., Deepthi, A., Mandal, B., & Puhan, N. B. (2017). FoodNet: Recognizing foods using ensemble of deep networks. *IEEE Signal Processing Letters, 24*(12), 1758-1762.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research, 12*(Oct), 2825-2830.

Pérez, F. (2014). Project jupyter, 2015. In.

Pierson, H. A., & Gashler, M. S. (2017). Deep learning in robotics: a review of recent research. *Advanced Robotics, 31*(16), 821-835.

Pouladzadeh, Shirmohammadi, & Yassine. (2014). *Using graph cut segmentation for food calorie measurement.* Paper presented at the 2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA).

Pouladzadeh, P., Shirmohammadi, S., Al-Maghrabi, R. J. I. T. o. I., & Measurement. (2014). Measuring calorie and nutrition from food image. *63*(8), 1947-1956.

Pouladzadeh, P., Shirmohammadi, S., Bakirov, A., Bulut, A., & Yassine, A. (2015). Cloud-based SVM for food categorization. *Multimedia Tools and Applications, 74*(14), 5243-5260.

Qualcomm. (2018). Qualcomm 5G. Retrieved from https://www.qualcomm.com/invention/5g/what-is-5g

Recio-Rodríguez, J. I., Martín-Cantera, C., González-Viejo, N., Gómez-Arranz, A., Arietaleanizbeascoa, M. S., Schmolling-Guinovart, Y., . . . Gómez-Marcos, M. A. (2014). Effectiveness of a smartphone application for improving healthy lifestyles, a randomized clinical trial (EVIDENT II): study protocol. *BMC Public Health, 14*(1), 254.

Rhodes, D. G., Adler, M. E., Clemens, J. C., & Moshfegh, A. J. (2017). What we eat in America food categories and changes between survey cycles. *Journal of Food Composition and Analysis, 64*, 107-111. doi:https://doi.org/10.1016/j.jfca.2017.07.018

Rouse, M., & Pawliw, B. (2010). Quantum Computing. Retrieved from https://whatis.techtarget.com/definition/quantum-computing

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

Rusin, M., Årsand, E., & Hartvigsen, G. (2013). Functionalities and input methods for recording food intake: a systematic review. *International journal of medical informatics, 82*(8), 653-664.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., . . . Bernstein, M. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision, 115*(3), 211-252.

Sahoo, D., Hao, W., Ke, S., Xiongwei, W., Le, H., Achananuparp, P., . . . Hoi, S. C. (2019). *FoodAI: Food Image Recognition via Deep Learning for Smart Food Logging.* Paper presented at the Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

Scarborough, P., Rayner, M., Stockley, L., & Black, A. (2007). Nutrition professionals' perception of the 'healthiness' of individual foods. *Public Health Nutrition, 10*(4), 346-353.

Scheidt, D. M., & Daniel, E. (2004). Composite index for aggregating nutrient density using food labels: ratio of recommended to restricted food components. *Journal of nutrition education and behavior, 36*(1), 35-39.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85-117. doi:https://doi.org/10.1016/j.neunet.2014.09.003

Seebregts, C. J., Zwarenstein, M., Mathews, C., Fairall, L., Flisher, A. J., Seebregts, C., . . . Klepp, K.-I. (2009). Handheld computers for survey and trial data collection in resource-poor settings: Development and evaluation of PDACT, a Palm™ Pilot interviewing system. *International journal of medical informatics, 78*(11), 721-731.

Sharp, D. B., & Allman-Farinelli, M. (2014). Feasibility and validity of mobile phones to assess dietary intake. *Nutrition, 30*(11), 1257-1266.

Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. *Annual review of biomedical engineering, 19*, 221-248.

Shim, J.-S., Oh, K., & Kim, H. C. (2014). Dietary assessment methods in epidemiologic studies. *Epidemiology and health, 36*.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Simpson, E., Bradley, J., Poliakov, I., Jackson, D., Olivier, P., Adamson, A. J., & Foster, E. (2017). Iterative Development of an Online Dietary Recall Tool: INTAKE24. *Nutrients, 9*(2), 118.

Singha, M., & Hemachandran, K. (2011). Performance analysis of color spaces in image retrieval. *Assam University Journal of Science and Technology, 7*(2), 94-104.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929-1958.

Stumbo, P. J. (2013). New technology in dietary assessment: a review of digital methods in improving food record accuracy. *Proceedings of the Nutrition Society, 72*(1), 70-76. doi:10.1017/S0029665112002911

Stütz, T., Dinic, R., Domhardt, M., & Ginzinger, S. (2014). *Can mobile augmented reality systems assist in portion estimation? A user study.* Paper presented at the Mixed and Augmented Reality-Media, Art, Social Science, Humanities and Design (ISMAR-MASH'D), 2014 IEEE International Symposium on.

Su, C.-H., Chiu, H.-S., & Hsieh, T.-M. (2011). *An efficient image retrieval based on HSV color space.* Paper presented at the 2011 International Conference on Electrical and Control Engineering.

Subar, Kirkpatrick, S. I., Mittl, B., Zimmerman, T. P., Thompson, F. E., Bingley, C., . . . Potischman, N. (2012). The Automated Self-Administered 24-Hour Dietary Recall (ASA24): A Resource for Researchers, Clinicians, and Educators from the National Cancer Institute. *Journal of the Academy of Nutrition and Dietetics, 112*(8), 1134-1137. doi:10.1016/j.jand.2012.04.016

Subhi. (2018, 3-6 Dec. 2018). *A Deep Convolutional Neural Network for Food Detection and Recognition.* Paper presented at the 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES).

Subhi, Ali, S. H. M., Ismail, A. G., & Othman, M. (2018). Food volume estimation based on stereo image analysis. *21*(6), 36-43.

Subhi, M. A., Ali, S. H., & Mohammed, M. A. (2019). Vision-Based Approaches for Automatic Food Recognition and Dietary Assessment: A Survey. *IEEE Access, 7*, 35370-35381.

Sun, M., Burke, L. E., Baranowski, T., Fernstrom, J. D., Zhang, H., Chen, H.-C., . . . Yue, Y. (2015). An exploratory study on a chest-worn computer for evaluation of diet, physical activity and lifestyle. *Journal of healthcare engineering, 6*(1), 1-22.

Sun, M., Burke, L. E., Mao, Z.-H., Chen, Y., Chen, H.-C., Bai, Y., . . . Jia, W. (2014). *eButton: a wearable computer for health monitoring and personal assistance.* Paper presented at the Design Automation Conference (DAC), 2014 51st ACM/EDAC/IEEE.

Sun, M., Liu, Q., Schmidt, K., Jie, Y., Yao, N., Fernstrom, J. D., . . . Sclabassi, R. J. (2008, 20-25 Aug. 2008). *Determination of food portion size by image processing.* Paper presented at the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). *Going deeper with convolutions* (Vol. null).

Techradar. (2018). 5G Technology. Retrieved from https://www.techradar.com/news/what-is-5g-everything-you-need-to-know

Temple, N. J., & Gladwin, K. K. (2003). Fruit, vegetables, and the prevention of cancer: research challenges. *Nutrition, 19*(5), 467-470.

Tharrey, M., Maillot, M., Azaïs-Braesco, V., & Darmon, N. (2017). From the SAIN,LIM system to the SENS algorithm: a review of a French approach of nutrient profiling. *Proceedings of the Nutrition Society, 76*(3), 237-246. doi:10.1017/S0029665117000817

Thompson, F. E., & Byers, T. (1994). Dietary Assessment Resource Manual. *The Journal of nutrition, 124*(suppl_11), 2245s-2317s. doi:10.1093/jn/124.suppl_11.2245s

Tieleman, T., & Hinton, G. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning, 4*(2), 26-31.

Timon, C. M., van den Barg, R., Blain, R. J., Kehoe, L., Evans, K., Walton, J., . . . Gibney, E. R. (2016). A review of the design and validation of web- and computer-based 24-h dietary recall tools. *Nutrition Research Reviews, 29*(2), 268-280. doi:10.1017/S0954422416000172

UK Food Standards Agency. (2017). Composite Products. Retrieved from https://www.food.gov.uk/sites/default/files/media/document/compositeproductsqanda.pdf

USDA-ARS. (2010). USDA food and nutrient database for dietary studies, 4.1. *Beltsville, MD*.

USDA. (2010, April, 2018). USDA National nutrient database for standard reference, release 28. Retrieved from https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/nutrient-data-laboratory/docs/sr28-download-files/

Van Rossum, G. (2007). *Python Programming Language.* Paper presented at the USENIX annual technical conference.

Vila-Real, C., Pimenta-Martins, A., Gomes, A. M., Pinto, E., & Maina, N. H. (2016). How dietary intake has been assessed in African countries? A systematic review. *Critical Reviews in Food Science and Nutrition*, 1-21. doi:10.1080/10408398.2016.1236778

Weiss, R., Stumbo, P. J., & Divakaran, A. (2010). Automatic Food Documentation and Volume Computation Using Digital Imaging and Electronic Transmission. *Journal of the American Dietetic Association, 110*(1), 42-44. doi:https://doi.org/10.1016/j.jada.2009.10.011

WHO. (2010). Nutrient profiling: report of a WHO/IASO technical meeting. *London: WHO*.

WHO. (2018). Obesity and Overweight. Retrieved from https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

Willett, W. (2012). *Nutritional epidemiology*: Oxford University Press.

William L., H. (2019). COMP 551 -Applied Machine Learning: Lecture 15 - Convolutional Neural Nets. Retrieved from https://cs.mcgill.ca/~wlh/comp551/slides/15-conv_nets.pdf

Willman, J. M. (2020). Creating GUIs with Qt Designer. In *Beginning PyQt* (pp. 165-203): Springer.

Wirfält, E. (1998). Cognitive aspects of dietary assessment. *Näringsforskning, 42*(1), 56-59. doi:10.3402/fnr.v42i0.1762

Wu, L., Liu, Z., Bera, T., Ding, H., Langley, D. A., Jenkins-Barnes, A., . . . Xu, J. (2019). A deep learning model to recognize food contaminating beetle species based on elytra fragments. *Computers and Electronics in Agriculture, 166*, 105002.

Yanai, K., & Kawano, Y. (2015a). *Food image recognition using deep convolutional network with pre-training and fine-tuning* (Vol. null).

Yanai, K., & Kawano, Y. (2015b, 29 June-3 July 2015). *Food image recognition using deep convolutional network with pre-training and fine-tuning.* Paper presented at the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW).

Yang, Y., Jia, W., Bucher, T., Zhang, H., & Sun, M. (2018). Image-based food portion size estimation using a smartphone without a fiducial marker. *Public Health Nutrition*, 1-13. doi:10.1017/S136898001800054X

Zawbaa, H. M., Abbass, M., Hazman, M., & Hassenian, A. E. (2014, 2014//). *Automatic Fruit Image Recognition System Based on Shape and Color Features.* Paper presented at the Advanced Machine Learning Technologies and Applications, Cham.

Zelman, K., & Kennedy, E. (2005). Naturally nutrient rich… putting more power on Americans' plates. *Nutrition Today, 40*(2), 60-68.

Zhang, X., Bellolio, M. F., Medrano-Gracia, P., Werys, K., Yang, S., & Mahajan, P. (2019). Use of natural language processing to improve predictive models for imaging utilization in children presenting to the emergency department. *BMC Medical Informatics and Decision Making, 19*(1). doi:10.1186/s12911-019-1006-6

Zhou, L., Zhang, C., Liu, F., Qiu, Z., & He, Y. (2019). Application of Deep Learning in Food: A Review. *Comprehensive Reviews in Food Science and Food Safety, 18*(6), 1793-1811. doi:10.1111/1541-4337.12492

Zhu, F., Bosch, M., Woo, I., Kim, S., Boushey, C. J., Ebert, D. S., & Delp, E. J. (2010). The use of mobile devices in aiding dietary assessment and evaluation. *IEEE journal of selected topics in signal processing, 4*(4), 756-766.