

**Collaborative design and feasibility  
assessment of computational nutrient  
sensing for simulated food-intake tracking  
in a healthcare environment**

by

Kaylen J. Pfisterer

A thesis  
presented to the University of Waterloo  
in fulfillment of the  
thesis requirement for the degree of  
Doctor of Philosophy  
in  
Systems Design Engineering

Waterloo, Ontario, Canada, 2021

© Kaylen J. Pfisterer 2021

### **Examining Committee Membership**

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Alex Mihailidis, Professor  
Dept. of Occupational Science & Occupational Therapy,  
Institute of Biomedical Engineering  
University of Toronto

Supervisor(s): Alexander Wong, Professor  
Dept. of Systems Design Engineering,  
University of Waterloo

Internal Member: Jennifer Boger, Adjunct Assistant Professor  
Dept. of Systems Design Engineering,  
University of Waterloo

Internal Member: Maud Gorbet, Associate Professor  
Dept. of Systems Design Engineering,  
University of Waterloo

Internal-External Member: Heather H. Keller, Professor  
Dept. of Kinesiology and Health Sciences,  
University of Waterloo

### **Author's Declaration**

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Statement of Contributions

This thesis uses content from 2 published journal papers, 1 currently under review, an unpublished manuscript, and 3 conference papers all of which I was a either first or co-first author. I provided major contributions on the vision, design, development, evaluation and writing of the papers' material. Paper-specific contributions as they pertain to each chapter's content are included below.

**KJ Pfisterer**, J Boger, A Wong. Prototyping the Automated Food Imaging and Nutrient Intake Tracking (AFINI-T) system: A modified participatory iterative design sprint. *JMIR Human Factors* 2019;6(2):e13017. doi: <http://dx.doi.org/10.2196/13017>. **This paper** is incorporated in Chapters 3 and 5 of this thesis. On this paper, K.J.P. was the main contributor from project inception to planning, implementation, data collection, analysis, interpretation, and writing.

**KJ Pfisterer**, R Amelard, AG Chung, B Syrnyk, A MacLean, HH Keller, A Wong. When segmentation is not enough: Rectifying visual-volume discordance through multisensor depth-refined semantic segmentation for food-intake tracking in long-term care. *Submitted: Scientific Reports*. **This paper** is incorporated in Chapter 4 of this thesis. On this paper, K.J.P and R.A contributed equally to this work. K.J.P conceptualized the system, was the main contributor to experimental design, data acquisition protocols, data collection, interpretation, and writing of the manuscript, with additional contributions to the algorithmic design. R.A. was the main contributor to algorithmic design and technical implementation. K.J.P and R.A. conducted data analyses.

**KJ Pfisterer**, R Amelard, A Wong. Volume to nutrients. **This unpublished manuscript** is incorporated in Chapter 4 of this thesis. On this paper, K.J.P, was the main contributor to experimental design, data collection, data analysis planning, interpretation, and writing of the manuscript, with additional contributions to the algorithmic design. R.A. led algorithmic and technical implementation. K.J.P and R.A. conducted data analyses.

**KJ Pfisterer**, J Boger, A Wong. Food for thought: Ethical considerations of user trust in computer vision. *CVPR2019 Fairness Accountability Transparency and Ethics in Computer Vision Workshop*, Long Beach, United States. **This paper** is incorporated in Chapter 5 of this thesis. On this paper, K.J.P. was the main contributor from project inception to planning, implementation, data collection, analysis, interpretation, and writing.

**KJ Pfisterer**, R Amelard, AG Chung, A Wong. A new take on measuring relative nutritional density: The feasibility of using a deep neural network to assess commercially-prepared pureed

food concentrations. *Journal of Food Engineering* 2018; 220(223). **This paper** is incorporated in Chapter 6 of this thesis. On this paper, K.J.P was the main contributor to experimental design, data acquisition protocols, technical implementation, data analysis, interpretation, and writing of this manuscript.

**KJ Pfisterer**, R Amelard, A Wong. Differential color space analysis for investigating nutrient content in a pureed food dilution-flavor matrix: a step toward objective malnutrition risk assessment. *SPIE Photonics West BiOS 2018*, San Francisco, USA. **This paper** is incorporated in Chapter 6 of this thesis. This paper built on one aspect of the previous paper. On this paper, K.J.P was the main contributor to experimental design, data acquisition protocols, data analysis, interpretation, and writing of this manuscript. Technical implementation support was provided by R.A.

**KJ Pfisterer**, R Amelard, B Syrnyk, A Wong. Towards computer vision powered color-nutrient assessment of pureed food. *CVPR2019 Women in Computer Vision Workshop*, Long Beach, United States. **This paper** is incorporated in Chapter 6 of this thesis. This paper built further on the previous two papers. On this paper, K.J.P was the main contributor to experimental design, data acquisition protocols, data analysis, interpretation, and writing of this manuscript. Technical implementation support was provided by R.A.

## Abstract

One in four older adults (65 years and over) are living with some form of malnutrition. This increases their odds of hospitalization four-fold and is associated with decreased quality of life and increased mortality. In long-term care (LTC), residents have more complex care needs and the proportion affected is a staggering 54% primarily due to low intake. Tracking intake is important for monitoring whether residents are meeting their nutritional needs however current methods are time-consuming, subjective, and prone to large margins of error. This reduces the utility of tracked data and makes it challenging to identify individuals at-risk in a timely fashion.

While technologies exist for tracking food-intake, they have not been designed for use within the LTC context and require a large time burden by the user. Especially in light of the machine learning boom, there is great opportunity to harness learnings from this domain and apply it to the field of nutrition for enhanced food-intake tracking. Additionally, current approaches to monitoring food-intake tracking are limited by the nutritional database to which they are linked making generalizability a challenge.

Drawing inspiration from current methods, the desires of end-users (primary users: personal support workers, registered staff, dietitians), and machine learning approaches suitable for this context in which there is limited data available, we investigated novel methods for assessing needs in this environment and imagine an alternative approach. We leveraged image processing and machine learning to remove subjectivity while increasing accuracy and precision to support higher-quality food-intake tracking. This thesis presents the ideation, design, development and evaluation of a collaboratively designed, and feasibility assessment, of computational nutrient sensing for simulated food-intake tracking in the LTC environment.

We sought to remove potential barriers to uptake through collaborative design and ongoing end-user engagement for developing solution concepts for a novel Automated Food Imaging and Nutrient Intake Tracking (AFINI-T) system while implementing the technology in parallel. More specifically, we demonstrated the effectiveness of applying a modified participatory iterative design process modeled from the Google Sprint framework in the LTC context which identified priority areas and established functional criteria for usability and feasibility. Concurrently, we developed the novel AFINI-T system through the co-integration of image processing and machine learning and guided by the application of food-intake tracking in LTC to address three questions: (1) *where is there food?* (i.e., food segmentation), (2) *how much food was consumed?* (i.e., volume estimation) using a fully automatic imaging system for quantifying food-intake. We proposed a novel deep convolutional encoder-decoder food network with depth-refinement (EDFN-D) using an RGB-D camera for quantifying a plate's remaining food volume relative to reference portions in whole and modified texture foods. To determine (3) *what foods are present* (i.e., feature extraction and classification), we developed a convolutional autoencoder to learn

meaningful food-specific features and developed classifiers which leverage *a priori* information about when certain foods would be offered and the level of texture modification prescribed to apply real-world constraints of LTC. We sought to address real-world complexity by assessing a wide variety of food items through the construction of a simulated food-intake dataset emulating various degrees of food-intake and modified textures (regular, minced, puréed). To ensure feasibility-related barriers to uptake were mitigated, we employed a feasibility assessment using the collaboratively designed prototype. Finally, this thesis explores the feasibility of applying biophotonic principles to food as a first step to enhancing food database estimates. Motivated by a theoretical optical dilution model, a novel deep neural network (DNN) was evaluated for estimating relative nutrient density of commercially prepared purées. For deeper analysis we describe the link between color and two optically active nutrients, vitamin A, and anthocyanins, and suggest it may be feasible to utilize optical properties of foods to enhance nutritional estimation.

This research demonstrates a transdisciplinary approach to designing and implementing a novel food-intake tracking system which addresses several shortcomings of the current method. Upon translation, this system may provide additional insights for supporting more timely nutritional interventions through enhanced monitoring of nutritional intake status among LTC residents.

## Acknowledgements

I feel so overwhelmingly fortunate to have had such an incredibly positive graduate student experience. My love of research started in my undergraduate degree - thank you to Dr. Thorsten Dieckmann for taking me into your lab and providing me with the opportunity to get hands-on learning and for your sprinkled mentorship over the past 10+ years. My time in your lab helped me build the confidence I needed to continue my learning journey.

A big part of my graduate studies journey success is because Dr. Mike T. Sharratt always encouraged me and met my enthusiasm for learning and cultivated our mutual curiosity with such zeal. You taught me many valuable life lessons and I wish so deeply that I could share this with you. You are missed more than I can say. Without that encouragement, numerous pep talks, and advocating for me in every facet of life ... I'm not sure I would have been brave enough to have begun.

Starting with my work with Heather in my masters, this was such a wonderful way to continue that learning journey together. Heather, you are a fantastic mentor and every opportunity you have set me up enabled me to embark on this PhD journey. I am so pleased to have been able to share in this chapter with you as well and look forward to more collaborations too!!

Maud, thank you for mentoring me in a complementary way by empowering me to get teaching and leadership experience in the classroom. I never would have expected my comprehensive exam to be a job interview in disguise but teaching BME 361 with you and Nima was such a pleasure, and it was so rewarding to connect with students and apply my developing skills. You took my self-doubt about being qualified to TA an engineering course and helped me see my value and contributions. Much of what I learned through those experiences is reflected here. Thank you.

Jen, thank you for letting me design a unique reading course together and taking me on as a student in that capacity. That spring term was one of the most intense I had and, my goodness, it was a lot of fun to jam pack so much learning into such a short time. I'm so proud to be showcasing so much of the user-study work I did under your wing in this thesis.

On a related note, thank you so much to Susan Brown, Lora Bruyn-Martin and the Schlegel-University of Waterloo Research Institute for Aging for your help in setting that study up for success. A big thank you to Jill Estioko for your strong support from the Schlegel Villages side. I owe much of the progress showcased here to the opportunities you provided and connected me with. Thank you. To all the project advisors, thank you for your time and shared enthusiasm. Translational research is *hard* and I didn't get all the way there, but we now have the blueprints of what's needed to come next and I'm excited to see how that might transpire.

Alex... I don't even know where to start. I've told you often how much I appreciate the numerous opportunities with which you have provided me. Being able to present at national and





international conferences in several different fields was absolutely incredible. You support your research-kids in incredibly unique ways, and you took a risk on me. I'm so grateful you did. You helped me hone leadership skills through planning high-profile tours, trusted me to plan CVIS, and helped me find wonderful undergrads to help support the technical implementation. Thank you for giving me the courage not only to expand my career aspirations, but also for your intensely strong support of embarking on parenthood. This was something I thought I would need to sacrifice for career. With your support to buoy me up, I am overwhelmingly joyful to have brought those worlds together. To say you empowered me is an understatement. I'm so excited for what is to come and how we can continue to get creative about harnessing the power of machine learning and artificial intelligence for helping society.

Alex MacLean and Braeden Syrnyk, you two are incredible allies. I'm so proud of the work and learning we did together. Thank you for sharing your talents and expertise with me.

All my parents, grandparents, and siblings, I am very appreciative of your support - each in unique ways. You helped me lay the foundation to my learning journey and have been there since the start. Thank you for your love and care.

My dear friends: Janine Blair, and Victoria Tolton, you provided me with so much support behind the scenes in ways I still don't completely understand. Amy Matharu, I feel so fortunate to know you. Thank you for sharing your positive energy and unbridled enthusiasm with me. Thank you for making things happen; you are such a cool human. And again, Victoria, thank you for your many hours of both literal and metaphorical counter-pressure ... what a beautiful start for everything to come next. Audrey Chung, Brendan Chwyl and Charlie C<sup>2</sup>: Brendan, your humour and friendship helped keep me light and my heart full. Audrey, I am so thankful for the friendship we developed and how much you helped me learn and grow as a human. Between care packages, side-by-side coding, painting, clay making, popcorn adventures, you nurtured my sense of self and helped me emerge so much more confident and capable in many ways. Armin Hasanbegovic, Mo Musbah, and Eric Pisani, our friendship began in undergrad. Mo and Armin, it's been so neat to go from honorary classmate to learning and building in an adjacent field. With love, support, care, gaming, visits and geek-outs together, you three have been with me throughout this journey - thank you, dear ones.

Robert, my very best big friend, our ongoing collaborative learning is among my fondest of activities. It's SO MUCH FUN! I love you not only for who you are, but for who I am when I am with you. As I am your pillar, you are my rock in all aspects of life. Your support and advocacy is everything to me. Today, I am here with you .

Arienne, you have inoculated me with more ambition and sense of purpose than I could have known possible. You teach me so much every day and have grounded me so deeply that I can reach "up. UP. *UP!!*" without fear of falling over. Thank you little one. You are my light .

## **Dedication**

To my best friend, my favourite collaborator.  
For my daughter, forging a path of endless possibility.

this is for you

# Table of Contents

<b>List of Figures</b>	<b>xiv</b>
<b>List of Tables</b>	<b>xvi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Problem Statement . . . . .	3
1.3 Challenges . . . . .	3
1.4 Objectives . . . . .	4
1.5 Research Contributions . . . . .	5
1.6 Thesis Structure . . . . .	8
<b>2 Background</b>	<b>9</b>
2.1 Machine Learning for Food Intake Innovation . . . . .	11
2.1.1 Segmentation Accuracy Challenges . . . . .	12
2.1.2 Volume Estimation Challenges . . . . .	13
2.1.3 Classification Challenges for Accurate Nutritional Inference . . . . .	15
2.1.4 Training and Validation Dataset Selection . . . . .	16
2.1.5 Advances Towards a Holistic Solution . . . . .	18
2.2 Utility of Food Biophotonics for Enhancing Estimates . . . . .	20
<b>3 What We Know: Expert User Knowledge and Needs Assessment</b>	<b>26</b>
3.1 Overview of Design Strategy . . . . .	27
3.2 Methods . . . . .	28
3.2.1 Stage 1: Design Ideation . . . . .	28
3.2.2 Stage 2: Reflect and Storyboard . . . . .	29
3.2.3 Stage 3: Storyboard Critiques . . . . .	30
3.2.4 Stage 4: Design of the Goldilocks' Quality Horizontal Prototype . . . . .	30

3.3	Analyses . . . . .	30
3.4	Results . . . . .	34
3.4.1	Stage 1: Design Ideation . . . . .	34
3.4.2	Stage 2: Reflect and Storyboard . . . . .	37
3.4.3	Stage 3: Storyboard Critiques . . . . .	38
3.4.4	Stage 4: Design of the Goldilocks' Quality Horizontal Prototype . . . . .	38
3.5	Discussion . . . . .	40
3.6	Conclusion . . . . .	41
3.6.1	Key Contributions . . . . .	42
<b>4</b>	<b>Re-imagining Food Intake Tracking to “Know Better”</b>	<b>43</b>
4.1	Seeing food intake objectively . . . . .	44
4.1.1	Food Related Segmentation and Classification Progress . . . . .	45
4.1.2	Methods . . . . .	47
4.1.3	Results . . . . .	55
4.1.4	Discussion . . . . .	62
4.2	Inferring Nutritional Intake . . . . .	65
4.2.1	Methods . . . . .	66
4.2.2	Results . . . . .	78
4.2.3	Discussion . . . . .	82
4.3	Conclusion . . . . .	98
4.3.1	Key Contributions . . . . .	98
<b>5</b>	<b>A New System Prototype's Palatability</b>	<b>100</b>
5.1	Establishing System Acceptability, Perceived Workload, and User Perceptions of Trust in the System . . . . .	101
5.2	Methods . . . . .	101
5.2.1	Interview Walk-through of Prototypes . . . . .	101
5.2.2	Usability Assessment Tools . . . . .	102
5.2.3	Webinar Feedback . . . . .	103
5.3	Analyses . . . . .	103
5.4	Results . . . . .	103
5.4.1	Users' Subjective Usability Scores . . . . .	103
5.4.2	Technical Expert's Subjective Usability in Context . . . . .	104
5.4.3	Users' Subjective Workload (RTLX) . . . . .	105
5.4.4	Users' Perceptions of Trust . . . . .	107
5.4.5	Qualitative Feedback . . . . .	108
5.5	Discussion . . . . .	109

5.6	Conclusion	112
5.6.1	Key Contributions	113
<b>6</b>	<b>Food Biophotonics: Optical Imaging for Estimating Human-Observable Nutrient Properties</b>	<b>114</b>
6.1	A New Take on Measuring Relative Nutrient Density	116
6.1.1	Methods	117
6.1.2	Results	125
6.1.3	Discussion	134
6.2	Towards Computer Vision Powered Colour-Nutrient Assessment of Pureéd Food	136
6.2.1	Methods	136
6.2.2	Results & Discussion	138
6.3	Conclusions	141
6.3.1	Key Contributions	141
<b>7</b>	<b>Towards Translation: Current Limitations and Future Directions</b>	<b>143</b>
7.1	Current Limitations and Additional Considerations	143
7.1.1	Datasets	143
7.1.2	Low-Density Foods	144
7.1.3	Volume Estimation	144
7.1.4	Pushing Food Biophotonics Further	145
7.1.5	Practical Implications of Altering Workflow	145
7.2	Future Directions: Opportunities to Disrupt for Translational Impact	146
7.2.1	Further Automation for Decreased Time Requirements	146
7.2.2	Incorporation of Fluid Intake Tracking	147
7.2.3	Supporting Enhanced Personalisation	147
7.2.4	Broadening AFINI-T's Reach with the Mobile Era	149
	<b>References</b>	<b>151</b>

# List of Figures

1.1	Overview of objectives and output. . . . .	6
2.1	From a food image to nutritional inference. . . . .	11
2.2	Reflectance and transmittance imaging in terms of scattering events. . . . .	23
3.1	Stage 2 PSW user interface. . . . .	31
3.2	Stage 2 Registered Team user interface. . . . .	32
3.3	Stage 2 Registered Dietitian user interface. . . . .	33
3.4	Primary and secondary users of AFINI-T . . . . .	35
3.5	Current workflow of food service and food and fluid intake charting. . . . .	36
4.1	System diagram of the encoder-decoder food network (EDFN). . . . .	51
4.2	Graph cut annotation example. . . . .	55
4.3	Comparing EDFN and “applied ground truth” graph cut (GC) segmentation with- out and with depth-refinement (EDFN-D, GC-D) . . . . .	57
4.4	EDFN and EDFN-D intake accuracy. . . . .	61
4.5	Imaging set-up and weigh stations. . . . .	69
4.6	Effect of under-representation of green foods on decoder output. . . . .	73
4.7	Convolutional autoencoder network for learned feature representation and in the context of classification. . . . .	74
4.8	The toast occlusion conundrum. . . . .	81
4.9	Nutrients of interest correlation and agreement between mass and volume esti- mates. . . . .	87
5.1	Subjective usability scores. . . . .	104
5.2	Trust ratings of existing system and AFINI-T prototype. . . . .	108
6.1	Dilution series. . . . .	118
6.2	Deep neural network architecture for relative nutrient density classification. . . .	120

6.3	Overview of comparative classification methods. . . . .	122
6.4	Normalized absorbance spectra for (a) blueberry, and (b) strawberry. . . . .	124
6.5	Descriptive analysis plots of purée flavours based on colour. . . . .	126
6.6	Descriptive analysis plots of purée flavours based on saturation and texture (entropy) . . . . .	127
6.7	Sample patches for each purée flavour and dilution. . . . .	128
6.8	Vitamin A containing purée samples. . . . .	133
6.9	Anthocyanin containing purée samples. . . . .	135
6.10	Two pronged purée assessment for bulk nutrient estimation and single nutrient assessment. . . . .	137

# List of Tables

2.1	Food databases and the appropriateness for LTC. . . . .	25
3.1	Time required to complete food and fluid intake charting in one neighbourhood. . .	38
3.2	Key inspiration concepts from commercially available online healthcare tools. . .	39
4.1	List of foods included in the regular and modified texture food datasets. . . . .	49
4.2	Comparative analyses of system performance between our proposed method and graph cut. . . . .	58
4.3	Modified texture foods: Volume estimation accuracy. . . . .	58
4.4	Dataset characteristics. . . . .	70
4.5	Cumulative imaged foods list. . . . .	71
4.6	Conversion of % daily values to absolute. . . . .	78
4.7	Average segmentation and classification accuracies within and across datasets. . .	80
4.8	Bulk intake accuracy within and across datasets. . . . .	83
4.9	Macronutrient intake accuracies within and across datasets. . . . .	84
4.10	Micronutrient intake accuracies of elements within and across datasets. . . . .	85
4.11	Micronutrient intake accuracies of vitamins within and across datasets. . . . .	86
4.12	Theoretical completion time . . . . .	94
5.1	Ravden usability checklist results. . . . .	106
5.2	Perceived workload measures of the current system and AFINI-T prototype. . . .	107
6.1	Sensitivity, specificity, and classification accuracy comparison (autoencoder vs. random forest and SVM). . . . .	130
6.2	Correlation between colour features and relative nutrient density. . . . .	132
6.3	Vitamin A content in raw sweet potato and carrot. . . . .	139
6.4	Sweet potato dilution prediction network accuracy. . . . .	140



# Chapter 1

## Introduction

This thesis describes a novel approach to support enhanced food intake tracking in long-term care homes (LTC) by addressing several shortcomings of the current systems in place and the technological solutions designed without the lens of LTC. It explores the ideation, design, development and evaluation of a collaboratively designed and feasibility assessment of computational nutrient sensing for simulated food intake tracking in the LTC environment. As its focus was application-driven, our story begins with the magnitude of the need and health implications for when nutritional needs are not being met as described in Section 1.1. Following this, comes the problem statement in Section 1.2, as well as a challenges and objectives nested within current limitations in Sections 1.3 and 1.4, and a summary of research contributions in Section 1.5. This chapter concludes with an overview of the remaining thesis structure in Section 1.6.

### 1.1 Motivation

The link between poor nutritional status (e.g., malnourished or at risk for malnourishment) and disease is well established; malnutrition is associated with decreased quality of life [124], increased hospital stays and pressure ulcers [222], morbidity [184] and mortality [222, 184]. Furthermore, malnutrition-related costs the health care system \$10 billion per year in each the USA and UK [80, 202]. Older adults are at increased risk for nutritional deficiency due to physical and physiological changes (e.g., reduced lean muscle, less efficient gastrointestinal tracts, changes in sensory ability like smell or taste), in addition to having a higher degree of co-morbidity [33]. With one in four older adults (65 years or older) malnourished [116], an additional 15% at medium or high risk for malnutrition [151, 113], and over 4 times higher odds of hospitalization and \$21,892 more in total charges for those with malnutrition [133], it's clear

nutritional status has multidomain effects with both fiscal and clinical ramifications and should be monitored.

Canadian older adults living in long-term care (LTC) are particularly vulnerable; 54% of the LTC population is malnourished or at risk [121] primarily due to low food intake [122]. This is higher than global estimates ranging from 19% to 42% (37 studies, 17 countries) [19]. Additional independent risk factors for malnutrition are eating challenges, and increased cognitive impairment [122, 232] which describes between 47% to 90% of the Ontario LTC population [86, 35]! Thus, tracking and preventing poor food intake is paramount. However, we lack quality tracking methods for food and fluid intake, especially needed with multiple staff involved in the care of residents over the course of a day or week.

**\$15.5 B** annual malnutrition related costs

**4x** odds of hospitalization from malnutrition

**54%** LTC residents malnourished - OR AT RISK -

Best practice metrics for ongoing nutritional assessment include monitoring unintentional weight loss, usual low intake of food, or other quality indicators to prioritise referrals and monitor effectiveness of nutritional support systems [60]. Nursing assistants or personal support workers (PSWs) chart food and fluid intake of residents using either a paper-based or electronic form to capture intake across a meal at 25% incremental proportions of intake. However, while inadequate intake is manageable [214], present guidelines for a nutritional intervention stipulate a resident must consume less than 75% of a meal most of the time [7, 211, 212]. Half of these residents who would benefit from an intervention are missed [211, 212] because of difficulties assessing and charting food intake. More specifically, the accuracy and validity of these methods is known to be poor [154, 240] with incorrect estimates over 50% of the time [39, 23]. As a result, trust in these measurements is low, with limited utility in practice. But as we will see in Chapter 3, care providers would like to utilize this information if measurement reliability and trust in these measurements could be ensured [180].

Outside of LTC, while traditional methods to measure nutritional intake exist (e.g., food frequency questionnaires, food diaries, 24 hour recall, and food intake patterns [178], they rely on self-report methods which are subjective and suffer from the same two key limitations: poor accuracy and validity [154, 240]. Alternative methods exist (e.g., weighed food record, digital photography [240]) however, these alternatives are time-consuming, require trained personnel, and are therefore impractical for large-scale adoption in LTC. Automated tools may provide a time efficient, cost efficient, and objective alternative. However, they are not without their own challenges specifically around food classification (segmentation and recognition), portion size

estimation (scale inference), and food mixing (occlusions). While others have successfully implemented automatic food intake systems, they rely on images from multiple perspectives [128], require a single image with a fiducial marker (i.e., reference object for scale inference) [176], or require manual segmentation and labelling for each food each item [162], which involves personnel time and may impact accuracy.

## 1.2 Problem Statement

In a 2016 review, two explicit gaps in the literature were identified as persisting: the need for user adoption studies and the need to address more complex meal scenarios [189]. For the most impactful research, these challenges must be addressed to remove barriers for user-adoption which is the focus of my research described throughout this thesis. The overall aim of this research is to enhance food intake assessment by reducing subjectivity while improving accuracy and precision to support higher-quality food intake tracking. The technology developed and assessed for feasibility through this process may be used as a system to assess changes in nutritional status, to provide more reliable data for nutritional interventions, and is specifically tailored for healthcare settings (e.g., long-term care). Figure 1.1 provides an overview with challenges and objectives described in detail below. Specific contributions are outlined in Section 1.5. This project improves upon and innovates beyond previous work by addressing three challenges (C1-C3) unique to designing a system to support enhanced nutrition tracking in long-term care.

## 1.3 Challenges

**Challenge 1 (C1): Understand and remove potential barriers to uptake.** The first challenge relates to the general approach of building a solution. Traditionally, technological solutions to food tracking have been approached in a silo-ed manner. Here, we bridged the transdisciplinary gap through collaborative design by bringing together key stakeholders across the nutrition, long-term care sector, and engineering disciplines to ensure real-world relevance. While progress in the field has led to successful food recognition tasks and some useful applications for calorie tracking and weight-loss management (e.g., [128, 163, 176, 189]), there has been a disconnect between the powerful nutrition approach as part of holistic care planning or as part of a healthy lifestyle. As a result, these applications have had limited clinical applicability.

**Challenge 2 (C2): Address real-world complexity.** Classically, food tracking and food databases oversimplify foods either in the presentation (e.g., banana in peel) or in how they are imaged (e.g., one food per plate, only full portions). In the real-world there are fewer constraints imposed which must be considered in the design of a robust system to navigate complex (e.g., multi

foods, mixed/touching foods) meal scenarios. Additionally, in practice, food intake tracking assumes equal consumption across a plate (or across the mean “plate” of all possible options) and is recorded only at the plate level (bulk food intake). The desired output was the ability to track, with limited human input, simulated food intake at the food item and nutrient level. Additionally, traditional image-based approaches tend to combine segmentation with classification making it difficult to evaluate system error.

**Challenge 3 (C3): Explore food database enhancement.** To date, food tracking systems rely on food databases for nutrient content estimates however, foods included in the database may not be representative of specific food items on the plate due to different ingredients included, cook times, or preparation methods. Leveraging biophotonic principles for food may be possible, but feasibility of this approach is unknown.

## 1.4 Objectives

**Objectives to address C1.** To understand and remove potential barriers to uptake, we used novel avenues for end-user engagement and sought to understand of the problem space to inform the physical design of a vision-based imaging system for food intake tracking driven by user needs. Typically, collaborative design requires a large team, intensive time commitments, and has not often been applied within the aging context. To address this, we developed a novel infrastructure for end-user engagement enabling them to share knowledge in a timely, iterative manner that is time and resource inexpensive. The target end-users of our primary research goals were health care professionals (e.g., dietitians, nurses), primary care givers (e.g., personal support workers, food aides), and clinical researchers within the long-term care domain. **The objectives** of addressing this challenge were therefore:

1. To ensure the system is user-friendly by conducting focus-groups, interviews and surveys to assemble a list of needs and priority areas with the design in mind, and
2. To reduce barriers to adoption by ascertaining feasibility from user perceptions perspective by conducting human factors and user-centred evaluation. Specifically, the performance and workload measures of this system were compared against traditional food intake charting methods from the end-user perspective.

**Objectives to address C2.** To address real-world complexities, we considered more complex meal scenarios and facilitated user trust in the system by disentangling sources of error. **The objectives** of addressing this challenge were therefore:

1. To construct a simulated food intake dataset which emulates a real-world environment and considers various degrees of food intake (i.e., plate leftover images).

2. To assess the feasibility and accuracy of an image-based system leveraging volume estimates for estimating bulk food intake.
3. To leverage *a priori* information to constrain the food classification problem by considering the long-term care context around menu planning and the likelihood of offered foods based on set menus at certain times.

**Objective to address C3.** We borrowed biophotonic principles and applied them to food. **The objective** of addressing this challenge was to assess the feasibility of enhanced imaging as a foundational step to enhancing food database estimates.

## 1.5 Research Contributions

Based on the objectives outlined above, the research contributions outlined in this thesis can be broadly categorized into hardware/software contributions, empirical research contributions, and dataset contributions [241] as outlined below:

1. **Hardware and Software Contributions.** These outcomes resulted from successful completion of Challenges C1 and C2.
  - (a) Significant algorithm development to bridge disciplines and apply current machine learning trends to food intake tracking. At each stage of the pipeline, careful application-driven design decisions were made to consider the constraints and requirements of workflow in LTC which led to the infrastructure for:
    - i. A novel food intake tracking system powered by machine learning for assessing proportion of food ingested and food intake tracking at the nutrient level. Leveraging a depth-refined imaging system for bulk food and nutrient intake assessment informed by...
    - ii. Proof-of-concept mock-ups of nutrient-sensing system using end-to-end design mentality informed by end-user engagement.
2. **Empirical Research Contributions.** These contributions resulted from successful completion of Challenge C1.
  - (a) A novel framework for conducting collaborative design of technology in a healthcare setting through participatory iterative design.
  - (b) Specific design elements, functionality, priority areas, and constraints necessary to consider for reducing barriers for uptake by end-users informed by user-needs assessment.

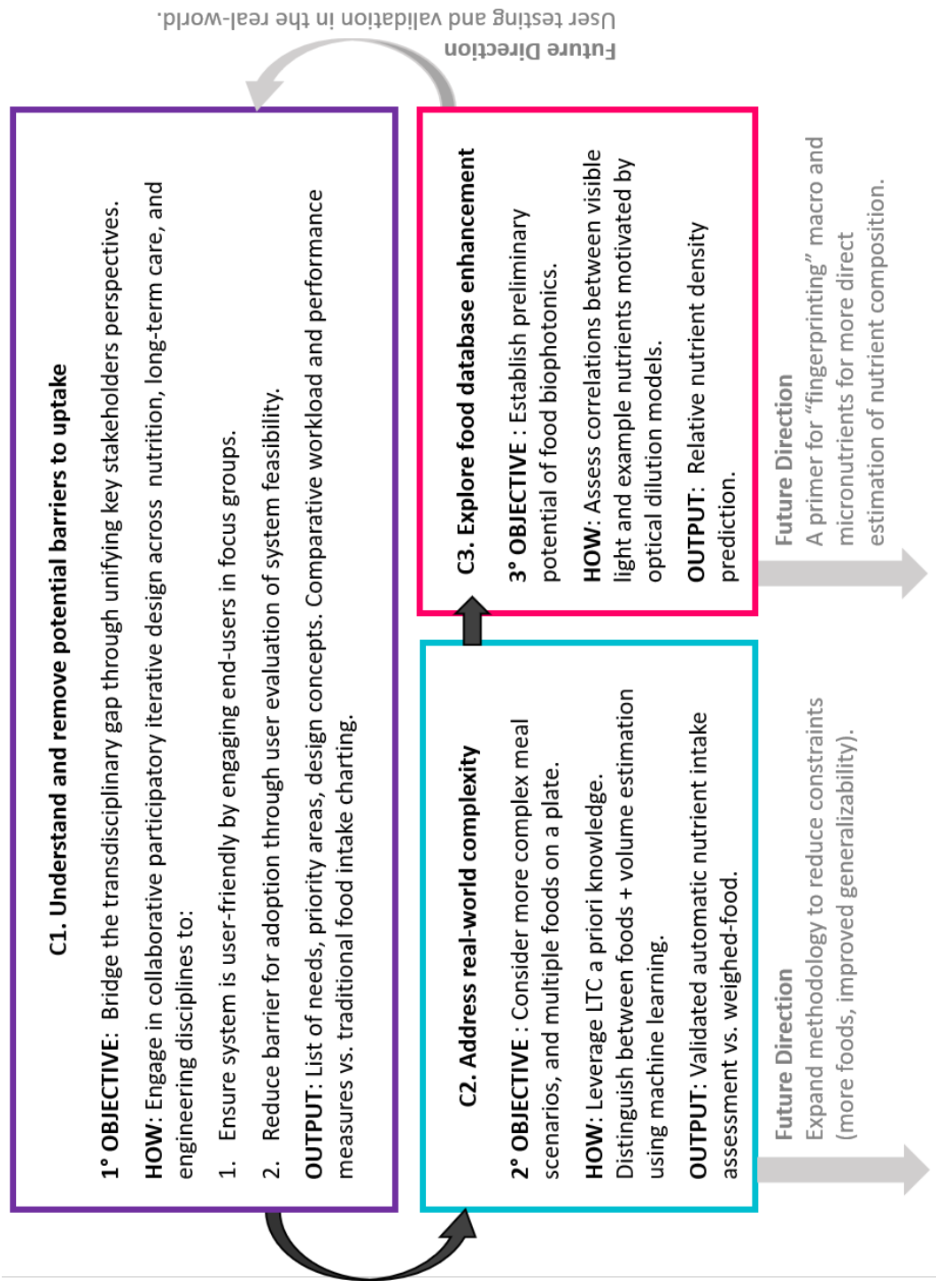


Figure 1.1: Overview of challenges and objectives scoped within this thesis and in the context of future directions.

- (c) Evaluation of the system prototype assessing feasibility, comparing perceived performance and workload measures of the novel system against traditional food intake charting methods, and user perceptions of trust.
  - (d) Proof of concept analysis of computational nutrient assessment theoretically motivated through food biophotonic optical dilution models.
3. **Dataset Contributions.** Since no datasets exist that represent LTC foods, I carefully and painstakingly curated three datasets to evaluate the algorithmic development and the accuracy of the system. This involved acquisition planning, building an imaging system and chassis for consistent imaging, as well as data collection. For the regular and modified texture long-term care food dataset, this also involved weighing each item separately followed by hand-segmentation and hand-labelling of every image. These datasets provided the background data upon which to tackle Challenges C2 and C3.
- (a) ***Regular and modified texture long-term care food dataset.*** This 1,039-image dataset is a fully labelled, high resolution dataset and consists of 47 foods representative of LTC and includes a variety of fruits, vegetables, pastas, soup, and meat dishes. This dataset is comprised of two subsets: a “regular” dataset, and a “modified texture foods dataset”. The **regular dataset** is comprised of 9 regular texture foods across three meals each containing up to three meal items and imaged at every 25% incremental amount relative to the full portion. It yielded 125 unique plates per meal and 375 unique plates across meals. The **modified texture foods dataset** is comprised 664 images across 93 classes of modified texture food samples representing 47 unique foods prepared by a LTC kitchen. All samples include hand-segmented and hand-labelled pixel-level segmentations. A 314-image subset of the modified texture foods dataset representing 63 food samples (56 unique) and 27 unique foods additionally includes full nutritional information provided by the LTC home. These foods were imaged one per plate and samples were imaged at different simulated intake levels by progressively removing some of the sample. Foods were imaged using the Intel RealSense to obtain RGB images and infrared depth images. For more detail, see Chapter 4.
  - (b) ***Dilutions: A puréed food dataset.*** This 3,540-image dataset is a fully-labeled, high resolution dataset collected to study the visible spectrum optical properties of commercially prepared puréed food flavours. This dataset is comprised of two subsets: “reflectance mode”, and “transmittance mode”. The **reflectance mode dataset** is comprised of six 5mL samples of each of thirteen types of commercially prepared purées at five discrete dilutions relative to initial concentration (20%, 40%, 60%, 80%, and 100% initial concentration) for 390 unique samples. Additionally, these

390 samples were imaged using three polarizations (i.e., unpolarized, and horizontal and vertical linearly polarizations) and were imaged using three exposures at each polarization for a total of 3,510 RGB (i.e., red, green, blue) images. The **transmittance mode dataset** provides an additional 30 full-field white normalized transmittance images of five calibration images and a five-tier dilution series consisting of five vitamin-A-rich commercially-prepared food samples imaged as 15 mL aliquots. Samples were imaged using a DSLR camera (Canon T4i) with a broadband tungsten-halogen light source and front glass fabric diffuser. For more detail, see Chapter 6.

## 1.6 Thesis Structure

The remainder of this thesis is organized as follows. Chapter 2, “Background” provides a literature review on nutrient sensing with special mention of relevant machine learning, optics, and imaging methods from three perspectives: (1) food segmentation, (2) food volume estimation, (3) food classification. Chapter 3 provides an overview of expert user knowledge and needs assessment and design of the Automated Food Imaging and Nutrient Intake Tracking (AFINIT) system, while Chapter 4 provides a technical solution to address these needs, and Chapter 5 describes and evaluates users’ receptivity to the system from the perspectives of usability, workload, and trust. Moving beyond the typical approach for an end-to-end food intake tracking system, Chapter 6 summarises my work on leveraging nutrient information embedded within the visible spectrum through the lens of food biophotonics. Finally, Chapter 7 describes the system’s current limitations and discusses possible future directions for further enhancement.



# Chapter 2

## Background

As we saw in Section 1.2, monitoring nutritional status in LTC is crucial, but difficult to do effectively. Time constraints are a pervasive barrier in the LTC sector. This is further compounded by frequent retrospective charting, which increases the probability of reporting errors [7]. While accuracy is important to ensure appropriate referrals of residents to a registered dietitian (RD) [211], the current method fails to differentiate between aspects of a meal; equal consumption across a plate is assumed. To address this, Andrews and Castellanos developed a food-type specific tool, however, consumption was still underestimated 25% of the time [7]. The challenge remains that comparisons either require time consuming methods or need to be completed by highly qualified personnel [211].

Technological innovations may provide a solution to remove subjectivity, enhance reproducibility, and inform higher levels of detail. In a 2010 NIH report, a plea for technological innovation for food intake was issued [223]. This need exists largely due to the aforementioned limitations surrounding accuracy and validity of traditional food intake methods [154, 240]. Thompson et al. discuss that prior advances were limited to making the recording process more effective (e.g., cell phone recording devices with voice recognition for interviews) [223]. For context, visual estimation of food intake versus actual weighed food records is known to have 50% error for food items and 20% for nutrients [23] which is further corroborated with [39] reporting incorrect portion estimation at least 56% of the time. With such a high degree of subjectivity and large margins of error, there is an opportunity for improved methods with greater utility. As we will see, the 2010 plea for technological innovation for food intake has been relatively unaddressed, especially in the LTC domain.

Regarding the general progress on automatic food intake tracking systems, several devices have been proposed for an individual to track and manage weight loss by recording intake using a

mobile device [128, 163, 176, 189]. While these on-the-go approaches could potentially be modified for appropriate use in LTC settings, in their current state, they are tailored for a different purpose, rely on self-monitoring, and do not adhere to related best practices for food and fluid intake. Additionally, they require a series of images from multiple perspectives [128], or depend on reference objects to infer scale (i.e., fiducial marker) [176]. In a time-constrained environment such as LTC and hospital settings, these requirements make these approaches infeasible. Consistent with this apparent gap, a 2016 review by Pouladzadeh and colleagues [189] summarize both traditional and newer (smart-phone vision-based) methods for calorie intake tracking in the context of weight loss and weight maintenance. They conclude there remains several challenges including: the explicit need for user acceptance studies of nutritional monitoring technology, consideration of more complex meal scenarios, and computational requirement consideration [189]. Within the LTC context, the closest technological solution was a comparison to estimate food waste of regular and modified texture diets either with the visual estimation method or using digital photographs afterwards [177]. The above highlights the opportunity for new innovations designed with the LTC context in mind. To achieve this, I embarked on constructing a vision-based system powered by machine learning.

It may be easy for us, as humans, to take our ability to recognize food or other objects as a relatively simple task. For example, when looking at an image of food on a plate we see a plate and can easily identify items on the plate whether they are mixed together or not. To a computer however, this is a much more complex yet crucial task to build upon for inference about how much food, and which types of food or nutrients are consumed. Regardless of methodology, the high-level sub-tasks to answer three main questions remain the same as illustrated in Figure 2.1:

1. *Where is there food?* This sub-task involves food detection or differentiation between food and plate (i.e., segmentation).
2. *What foods are present?* This sub-task involves comparing foods on a plate to one another (i.e., feature extraction), and deciding which type of food each item is (i.e., classification)
3. *How much food remains relative to the initial amount?* This sub-task involves estimation about how much food is there (i.e., volume estimation) and is essential for drawing inferences.

Since each stage relies upon all previous stages, an upstream error translates to a potentially amplified downstream error. The result is estimates of estimates and at each level with increasingly larger margins for error. This may be why in the literature, few researchers have concentrated on the entire process, and have instead focused on a subset of the process. To date, food classification has received the most attention with less attention given to segmentation, volume estimation and limited attention for a holistic food intake tracking system.

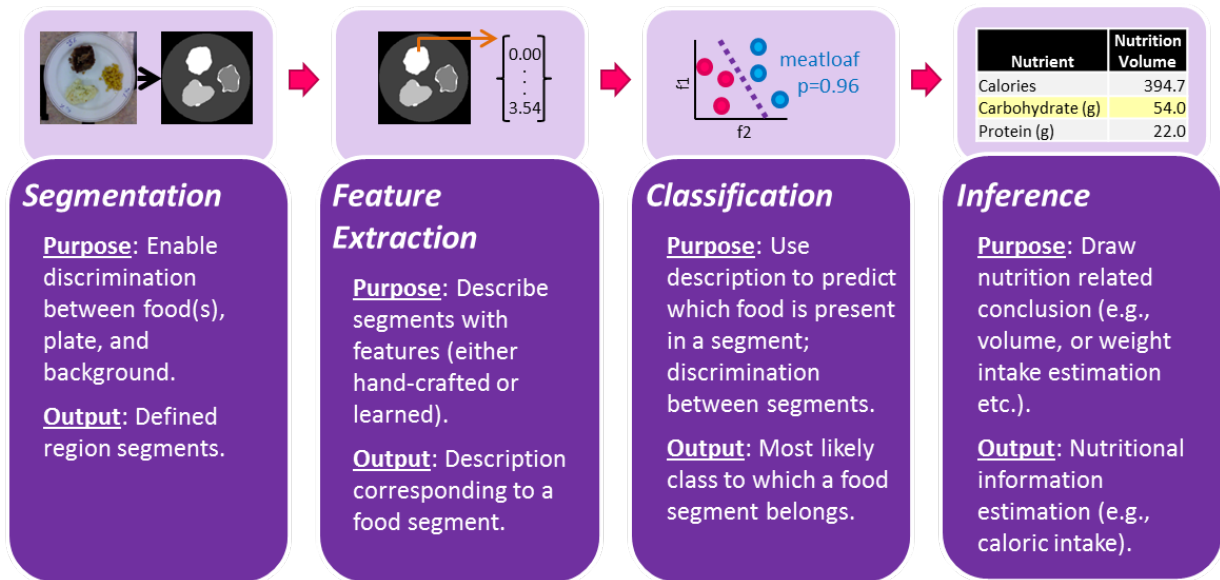


Figure 2.1: From a food image to nutritional inference.

The remainder of this chapter outlines the rationale for applying machine learning in the context of addressing three main questions to approach a machine-learning-powered food imaging system. While in-depth reviews for classification can be found in [144, 34, 61, 189, 218]), I have included a brief summary highlighting key classification progress and accuracy from segmentation to inference in the context of limitations and unsolved challenges.

## 2.1 Machine Learning for Food Intake Innovation

Machine learning methods provide a data-driven approach to remove objectivity and maintain consistency. Several examples of machine learning techniques have been published even twenty years ago where machines outperform humans, yet in the nutrition community these approaches have had minimal application. It's time to bridge these fields. For example, in a traffic sign recognition task, [50] presented a machine learning framework, specifically using convolutional neural network (CNN) with half the error rate of identification by humans. For object classification within a scene more broadly, the use of CNNs have attained surprisingly strong results on standardized image databases. For example, on an object recognition task with 1000 classes Krizhevsky et al. achieved an accuracy of 62.5% [130], whereas random chance would yield an accuracy of 0.1% (1/1000). Lawrence et al., applied CNNs for facial recognition and achieved 96.2% accuracy [134]. In more a complex task comparing human and machine performance,

an additional example from the gaming community was the win of Google’s AlphaGo against a grandmaster at the strategic game, Go [161, 74]. Intuitively, with advances in car automation and self-driving cars, leveraging machine learning and artificial intelligence is becoming common in our society. With promising accuracies attained by applying machine learning methods in other fields, leveraging these methods in the world of nutrition may provide an objective and reliable alternative to traditional food and nutrition intake methods especially when considering these novel methods need not be perfect to match or outperform humans given human error margins of 20%, 50%, and up to 400% [23, 39]; this is possible as discussed below.

### 2.1.1 Segmentation Accuracy Challenges

While food image segmentation progress has been made, error assessment in these systems tends not to be reported, or segmentation is coupled with either classification [186, 120, 93, 162, 207, 237, 246] or volume [89]. This makes sources of error difficult to disentangle and has practical implications. There is generally no way to systematically assess error propagation as part of the pipeline for predicting nutritional outcomes. This results in the system operating as a “black-box”, which may limit the uptake of these approaches in practice due to low perceived trustworthiness. Beyond the user and ethical perspectives, several researchers also describe the need for accurate segmentation methods for accurately predicting nutritional information (e.g., energy, macro-/micronutrient content) down-stream in the pipeline [162, 10, 237, 252].

One challenge comes from a lack of consistency in which segmentation accuracy is evaluated and quantified. For example, [160] reported a receiver operating characteristic curve of 0.9982, with precision and recall approaching 0.81 and 0.55, respectively [238]. Intersection over union (IOU) has been the most consistently reported accuracy metric [249, 48, 10, 4]. IOU has several advantages over more traditional precision/recall metrics as it considers the proportion of properly assigned pixels but also penalizes false positive predictions. That said, it may not necessarily capture an assessor’s perspective on what is relevant food (e.g., to include or not include crumbs) - more on that in Chapter 4.

Existing food intake tracking systems rely on images from multiple perspectives [129, 187], require a single image with a fiducial marker (i.e., reference object [176, 252]), or may not be suitable for real-time monitoring [162] because of the time required to build adequate models from scanners. Others are limited by requiring manual selection of bounding boxes [120], to predicting food areas with a bounding box [207] or require manual labelling for each food item [162]. These methods involve operator time and may impact accuracy. For example, two operators may segment food differently, foods may be incorrectly labelled, or labels may be missed in some cases. In the case of limiting predictions to within a bounding box, these methods cannot be used to accurately estimate intake as they include both food and background and are not assessed on

a per-pixel level. One semi-automatic method, interactive graph cut segmentation, has yielded strong accuracy in food segmentation [186, 89, 120]. It does not impart the same degree of burden as manual segmentation and we consider this as an “applied ground truth”. However, interactive annotation graph cut [28] requires user input to initialize the segmentation process (e.g., drawing areas to keep or discard). Even adding only a few seconds per resident per plate when scaled within the LTC environment makes it prohibitive within this context.

Others have reported methods such as adaptive k-means segmentation yielding an IOU of 0.64 for [249] (single image analysis). This method requires the number of items (k) per plate and forces each pixel to belong to one of the k classes and would require a separate model for each different value of k. Other methods have achieved higher IOU using deep convolutional neural networks (DCNN) when trained/tested on the same food dataset. Ciocca *et al.* achieved an IOU of 0.79 [48], Aslan *et al.* achieved an IOU of 0.931 using semantic segmentation (pixel-level labeling) [10], and Aguilar *et al.* achieved an IOU of up to 0.971 for spatially distinct (i.e., non-touching) food items [4]. All of these implementations were trained and tested on the same food dataset which may impact their generalizability to real-world scenarios because of inherently learned biases of each dataset. This is an important consideration with LTC in mind.

Assuming every other aspect of the process from segmentation through inference were 100% accurate, this implies food intake results would have a margin of error of up to 100-error%. While these may often fall within tolerance on this sub-task, improvements upon these methods for increased accuracy is desired. Improper segmentation will translate to either underestimates of food intake (i.e., the segments are smaller or do not capture all food before and after), or overestimates (i.e., the plate is captured and counted as food ingested). This ramification requires careful consideration when designing a system whose purpose is to achieve interpretable and understandable accuracy for nutrient intake estimation.

### 2.1.2 Volume Estimation Challenges

One consideration often overlooked is not considering portion size. This may not be as large of an issue in the context of weight loss management where the assumption is the entire imaged food is consumed. However, in the LTC setting where low food intake is commonplace, this assumption is violated frequently and must be accounted for any degree of utility within the LTC context. Food volume estimation systems estimating portion sizes of consumed food by subtracting the remainder of food from the original portion size. Several attempts have been made using template shape matching [93, 243, 40, 110, 175, 193, 96]. Three main drawbacks of these methods are the requirement of a shape library, difficulty with template matching with occlusions, and varying preparation methods of the same food. For example, if a food is prepared differently, it may not map onto the appropriate shape model (i.e., 3D banana in peel may be in the library but sliced

or diced banana may not). Similar to the segmentation problem, others have applied a multi-image perspective or stereo reconstruction for volume estimation [57, 190] or building a 3D representation through point-cloud representation [194]. The main drawback of these approaches is the time required to take the photos from different perspectives or gather enough sample points scanned for an accurate 3D representation; lack of time is a main concern when considering the LTC context. Another challenge of these approaches is accurate modelling/measuring of highly textured foods. This is a particularly salient issue for LTC where modified textures are often prescribed as part of a therapeutic diet [234] so very different foods can appear similar (e.g., minced or puréed foods).

Others have employed depth cameras or structured lighting to map the topology of the foods [71, 205, 162, 42, 139]. At a high level, structured lighting is when a pattern of light is projected upon the surface with a camera to image the result. Based on the degree of deformation between the known projected pattern and resultant image captured by the camera, the topography of the food surface can be inferred. There has been very little work on incorporating structured lighting. Only two papers were identified to have used structured lighting for volume estimation [205, 139], the latter of which reported an average error in weight of 2.56-5.01%. One potential reason for few papers leveraging a structured light approach is that until recently, these systems were difficult or expensive to build. However, with recent advances in the Kinect and Intel's RealSense RGB-D (i.e., red, green, blue plus depth) cameras, an off-the-shelf solution is now available at relatively low cost (i.e., under \$200) and in the current versions of iPhone, depth cameras are built-in. As a result, I think moving forward, there is great promise in solutions leveraging a structured light approach. Especially when considering the multiplicative nature of error propagating from each stage of the system, it is promising that the accuracy of a structured lighting approaches 97.44% [139]. One drawback of depth- and structured-light-only methods is highly reflective foods (e.g., gelatin, soup) which can throw off the readings, but they have an advantage at being more robust against illumination variations [139]. Taking a hybrid approach with additional refinement from RGB images, it may be possible to circumvent these drawbacks and capitalise on this illumination robustness advantage.

Despite the deep learning boom, acquiring adequately large and complete food datasets for training and testing has limited progress [250] of applying machine learning to volume estimation. The few forays have been fairly early-stage as they contain either large mean volume estimation errors (up to 400 mL error [162]), measure in terms of niche units (bread units tailored to diabetes [73]), or are limited to a small number of or synthetic food items in a highly controlled environment (e.g., [142, 143]).

The most common method of inferring portion size has been using fiducial markers (e.g., an item that is of known size and shape) to infer scale. Some have attempted using a thumb in screen [5, 186], while others use a checkerboard approach [93]. However, this approach may not

be practical in the LTC environment or may increase the complexity of the system to account for the fiducial marker moving from frame to frame. Typically, after the fiducial marker is used to infer scale, geometric shape matching is commonly applied to images to infer volume from a two-dimensional image. For example, given the shape in an image (e.g., unpeeled orange), look up the closest geometric model (e.g., sphere) and use that geometric model to infer volume.

This combined method has been shown to yield rather accurate results; [40] reported a percent error of 11% based on volume estimation and [244] reported a percent error of 18% for weight inference through volume estimation when compared to ground truth measurements. While this method far surpasses human error in portion estimation around 56% [39], the inherent limitation is that a geometric shape model is required for this calculation. In the case of foods served in LTC, particularly puréed foods, shape models may not exist due to high intra and inter-class variation for each food item. A more general approach that does not rely on prior shape or scale inference would provide a more robust solution within this context.

### **2.1.3 Classification Challenges for Accurate Nutritional Inference**

Considering the end goal of eventual nutritional estimation, classification is an important piece of the puzzle. From a segmentation perspective, this could be seen as a special case of classification in which a food image must be segmented into the food regions and the background (a two-class classification problem). As such, the implications of misclassification errors (i.e., improperly identifying food as background or background as food) provide relevant context. While extensive reviews of the literature can be found elsewhere (e.g., [144, 34, 61, 189, 218]), it is clear that automatically identifying food assigning it to a food-item category (i.e., food classification) has received the majority of attention. Indeed, this is an important step towards an automated system for food and nutritional intake analysis, however for malnutrition risk assessment, *how much* food was consumed is more clinically relevant.

In the 2020 review by Lo et al. [144], they summarize food classification systems ranging from a top-1 (i.e., the systems ‘best-guess’ for food category) accuracy of 50.1% on 256 classes [119], to 94.5% on 30 classes [188] using traditional approaches on food classification (i.e., one-vs-rest classifier [119], support vector machine classifier with handcrafted colour and texture features [188]), and from 54.7% on 256 classes [140] to an impressive 93.0% on 101 food classes [220] using deep learning approaches. But even when classification accuracies are high, misclassification errors are problematic. For example, misclassifying applesauce as mayonnaise or cream of wheat would yield wildly different nutritional compositions which limits utility in practice. Regardless of the number of classes a system may be able to automatically detect, if the automated classification is not always reliable, it violates at least one of the constraints required within the LTC sector. Practically, this means the system will lack trustworthiness, undermine

utility in practice, and may inhibit initial uptake [225]. For example, Yunus et al. [246] map words to estimate ingredients and generate nutritional info for an entire dish or item based on the top-1 classification (i.e., the best guess of what a food is). The strength of this approach is the ability to detect numerous dishes (100 classes of images) and has quite a high top-1 accuracy (up to 85%). However, this implies misclassification is occurring 15% of the time. Similarly, Wang et al. [237] make use of classification to refine segmentation. They are able to detect many different foods (42 items), provide a good segmentation feedback loop, account for personalized preferences but are limited in accuracy; the average daily classification accuracy was at best 69.81%. In the case of Wang et al. [237] while segmentation is the goal, it relies on feedback from the classification, so a classification error rate of 30% is problematic in this setting.

One crucial factor in determining the potential for successful classification is based on the features used to discriminate between classes of food item. If selected features do not represent the data correctly, their utility in distinguishing between different food types is limited. Typically, in the nutrition literature, handcrafted features have been used based on colour and texture. Food is typically very colourful, so this has been fairly successful in the past. Since 2010, eight articles [69, 8, 252, 245, 94, 120, 155, 90] were identified as relevant (i.e., pertaining to food classification) leveraging handcrafted features with classification accuracies ranging from 53.7% [155] to 90% [90] on distinguishing between six [8] and 100 types of food [120]. That said, a more data-driven approach using learned features may provide more generalizable and better performing models upon which to classify food types. While this data-driven approach may potentially yield less intuitive features, given food images are highly colourful, these features may reflect handcrafted features in addition to other nuances which may not be as obvious to humans. These learned features may provide more robust alternatives to hand-crafted. For example, more recently, two articles focussed on food classification leveraged learned features achieving classification accuracies of 60.47% on a 10-class [115] problem and 76.4% on a 100 class problem [140]. Conveniently, while there are few examples upon which to compare accuracy between handcrafted and learned features, the 100 food class dataset example is ideal as both papers [120, 140] used the same dataset (UEC-FOOD100) and had the same number of classes. The 76.4% [140] was achieved through the use of learned features which provides an example where learned features outperformed handcrafted classification accuracy at 59.6% [120]. This suggests learned features may be more powerful for distinguishing between food classes.

#### **2.1.4 Training and Validation Dataset Selection**

As we saw in 2.1.1, the notion of training and testing on the same dataset is an important concept when considering generalizability of results as it risks learning biases and nuances specific to that specific dataset and may limit the ability to adapt to new instances. As with all machine learning,



pre-existing examples are necessary to build, train, and validate a model. For this work, we were very intentional and deliberate about training and validating on a separate pre-existing dataset to which we could then test models developed on our novel datasets. When selecting a dataset for training and validation, careful consideration must be made when evaluating its appropriateness. Specifically with food intake assessment in LTC we must consider:

- **Colour:** Food comes in many colours. Part of supporting a healthy diet includes the mentality of “eat a rainbow” to ensure various micronutrient needs in addition to macronutrient needs are met [165]. As such, there should be a wide distribution of colours naturally found in foods captured in the training and validation datasets. In Chapter 4, we see and discuss limitations of our best available training dataset option because of limited representation of green as well as steps we took to ameliorate this issue.
- **Texture:** Similar to colour, foods also inherently come in a variety of textures. This aspect is particularly salient when considering the LTC population where 47% of residents receive modified texture diets (e.g., minced, puréed) as part of a strategy to address the high prevalence of swallowing difficulties pervasive in LTC [234].
- **Portion:** Given the application to food intake assessment, the ideal training dataset would have representative intake images that include both before and after images of meals (i.e., not just full portions of served food). The rationale is by having more representative examples of what foods can look like partially, or fully disassembled in the case of food mixing, in the training dataset, the network will be more likely to learn representative examples of foods apart from the original served context.
- **Orientation:** The occlusion conundrum where one food is in the way of another is very difficult to circumvent especially in the LTC environment when taking multiple images from many perspectives is infeasible due to time constraints. For the purpose of acquisition of LTC images “in the wild”, images taken from the above configuration is preferred to facilitate volume estimation and down-stream nutritional intake estimation. The issue of occlusion (e.g., seeing only the bun as part of an assembled hamburger) remains an issue, however, is reasonable when accepting the assumption that complex foods (e.g., foods with multiple components like a hamburger) are eaten in similar proportions. While this is undoubtedly a fallible assumption, errors in down-stream nutritional estimation are constrained to a specific food as opposed to influencing the entire plate. As such, the ideal training and validation database would be acquired in the top-view configuration.
- **Label level:** To facilitate volume estimation and down-stream nutrient estimation, image segmentation must be conducted pixel-wise as opposed to using bounding boxes. As such, the ideal training and validation database would be labelled at the pixel-level.
- **Style:** Multiple food items on a plate is common in LTC as the “family style” approach

to eating has been shown to enhance food intake [233]. While multiple plates may also be used in LTC (e.g., soup, side salad, dessert), the training and validation dataset would ideally show representation consistent with “family style” as opposed to solely “single item” plates.

- **Accessibility:** The ideal dataset needs to be readily available (i.e., non-proprietary).

Based on the above considerations, the most imperative are orientation and label level and the most suitable food database available for training and validation for LTC (i.e., top-view with pixel-wise labelling) is therefore UNIMIB2016 [49]. Table 2.1 below provides a comprehensive comparison of existing food databases along with characteristics that are relevant to food intake estimation and a summary of rationale for selecting UNIMIB2016 [49].

### 2.1.5 Advances Towards a Holistic Solution

Based on the above summary, a system to accurately identify food items and make accurate nutritional inferences is complex and despite the following advances, can be considered an unsolved problem, particularly when considering consistency in system error reporting. This section summarizes progress on end-to-end systems within the context of error reporting.

A variety of approaches of end-to-end systems are described below. In several cases, the method for segmentation was either not mentioned [42, 166, 67, 70, 45] or was explicitly stated as being beyond the scope of the present version of the system [166]. For the papers which do mention their segmentation methodology, GraphCut [186], semantic segmentation [162], minimum spanning trees [93] were applied. For feature extraction, as seen in the papers which focus explicitly on feature extraction, as part of the holistic system, we again see the majority employed hand-crafted features [166, 42, 93, 45]. However, in more recent years (i.e., since 2015, three quarters of the identified systems have instead leveraged learned features [162, 186, 67].

At the level of classification, particularly in the case of learned features, this implicates the methods employed for classification as well. In the case of learned features via neural networks, with a trained network, feature extraction and classification can be combined instead of considered as two sub-processes as in the case of [162, 186, 67] with reported classification accuracies of 100% (11 classes) [186], 82.5%(15 classes) [67]. Alternative methods employed for classification were AdaBoost [166], K Nearest Neighbours [93], and support vector machines [45, 42] with reported classification accuracies of 68.3% (50 classes) [42], 99.1%(6 classes) [45]. In these cases, comparison on relative performance for the ideal method for classification is difficult to establish due to inconsistencies of datasets and the number of classes; the higher number of classes, the more difficult the problem.

Finally at the inference level, few papers reported percent error in calorie estimation or weight

estimation. For calorie estimation, the reported values are: 0.09% [45], 0.25% [186], 30% [67], 35% [162]. [93] reports a 10%-11% weight estimation error depending on the method used (area vs. shape template). Using two models for comparing volumes (depth images vs. geometric models, [70] report respective average errors of 27.3%-56.0% and 14.0%-18.0%.

Liao et al. [139] approach the problem by estimating mass of food consumed using a depth camera and a specific gravity function to go from food density to estimated mass. One benefit of depth camera images is they are robust against illumination variations which typically plagues classification. They use an elegantly simple approach to determining where food is on a plate based on depth images and inpainting where specular reflection becomes an issue for foods with high reflectivity (e.g., soups, gelatin etc). This approach holds promise however, not all foods specific gravities are known which limits potential reach of this method. Additionally, the mass estimation accuracy was only tested on three food samples (egg, rice and tofu) and in its current form, it cannot address multiple foods on a plate as is standard in LTC.

Others still report a proxy of accuracy based on correlation coefficient compared to a ground truth method as in the case of [166](0.32), and [67] (0.78). It is important to note that a strong correlation does not necessarily imply accurate results. For example, two signals may be highly correlated, but one may be highly biased resulting in consistent over or underestimates.

While each of the above methods have their own quirks and limitations, the bottom line with respect to portion size or calorie estimation is that they at least achieve similar accuracies to average reported human error. As such, there is room for improvement.

This is where the power of leveraging *a priori* knowledge comes into affect. Instead of treating classification in isolation, we have several insights driven by how food is prepared and served within the context of long-term care. Firstly, menus are planned in advance - this provides crucial information in terms of limiting possible foods as well as providing an opportunity to leverage menus planned with Canada's Food Guide in mind for the purpose of providing a nutrient level baseline of what is in each portion of food. Secondly, certain foods are more likely to be served at certain times of the day. While there may be exceptions to this rule for honouring preferences, from a constraints perspective this enables us to limit the potential possibilities and simplify the classification problem. Instead of needing to ascertain if a sandwich was from any number of restaurants, we can assume it's most likely to be what was on the menu for a given day and again more likely for that sandwich to be served at lunch. With this approach to constraining the problem, we can make some good headway towards a system that works within a naturally constrained environment. While this approach may have blind spots for when residents leave the home for restaurant meals with family, this remains a constant challenge within the LTC system and are therefore treating this as outside the scope of this dissertation.

More generally, progress in this field of end-to-end systems for nutrition monitoring has been

outside the context of LTC with an emphasis of an individual tracking and managing their personal weight loss or health tracking using mobile devices [128, 162, 176, 189, 10, 129, 4]. While these approaches could be modified for use in LTC, in their current form, they target a different purpose (e.g., calorie tracking), still rely on self-monitoring, and do not consider the LTC context for food and fluid intake tracking best practices. Doulah et al. provide a review of technology-driven methodologies including a summary of work exploring wearable devices and sensor [61]. However, these wearable sensor approaches have typically been developed for individual use and require individuals to wear sensors like microphones or strain sensors. Within the LTC setting the wearables approach is inappropriate from a privacy perspective, when considering the degree of assistance needed during mealtime, as well as the financial implications of the number of devices required to track intake. As such, current approaches are infeasible for large-scale monitoring, especially in time-constrained and financially constrained environments such as LTC or hospital settings. Perhaps most relevant is the work of Astell and colleagues [12], who developed an effective electronic food record system for nutrient tracking system for community dwelling older adults. While promising (approximately 97% agreement for energy intake compared to food diaries), the comparison method of food diaries is similar to the monitoring already in place within LTC so true accuracy of this method remains unclear.

Corroborated by three recent reviews [189, 218, 61], further innovation is needed. These reviews summarize both traditional and newer (smartphone vision-based) methods for calorie intake tracking in the context of weight loss and weight maintenance. They conclude that several challenges remain, including: the explicit need for user acceptance studies of nutritional monitoring technology [189], computational requirement consideration [189], the need for a comprehensive food image dataset for benchmarking and better generalization of learned classifiers [218], consideration of more complex meal scenarios (i.e., beyond solid, separated foods, or synthetic foods) [189, 218], the need for proper comparisons against gold standard weighed food records as well as the need for adequate statistical analyses of methods [61]. Within the LTC context the closest technological analog in this domain was a comparison to estimate food waste of regular- and modified-texture diets either with the visual estimation method or by using digital photographs for retrospective analysis [177] both methods required significant operator time as it was a manual process. More broadly, further work is needed for developing an accurate, objective and cost-effective automated system [218, 61].

## 2.2 Utility of Food Biophotonics for Enhancing Estimates

Based on the current state of literature, the level of nutritional inference always relies upon some form of food database. This presents an additional challenge relating to data quality and/or sensitivity of the food database. A large portion of resources are dedicated to USDA food database,

but accuracy is limited to the foods tested and included in database. For example, not all cooking methods are considered, holding times for foods which may impact the nutritional composition is not accounted for, or specific variants on food items (e.g., my maternal grandmother's family bun recipe) may not be reflected.

In my mind, if we can build a system that can leverage the base nutritional content from menu planning and food databases while improving and extending beyond this technique, we will reach a more reliable, generalizable solution. For example, by augmenting or supplementing the nutrient estimates with measurements we take from each specific plate for nutrients of interest, the estimates will be closer to the ground truth. While this may not be necessary in all cases, keeping the system in mind from a research tool perspective in addition to its application in the LTC community, this may have added value and could potentially provide a far more cost and time-effective alternative to sending food samples for component analysis for proof-of-concept studies. For these reasons, Chapter 6 takes a first step into the utility of visible spectrum light and how it relates to nutrients using vitamin A and a group of antioxidants called anthocyanins which are optically active in the visible spectrum. This section provides additional rationale for these considerations of "beyond the database" as well as a primer on food biophotonics to aid in that assessment.

Further motivating the need to explore additional avenues is one condition which increases the risk for malnutrition, dysphagia (swallowing difficulty) [104, 219]. Dysphagia affects approximately 590 million people worldwide [47] and at least 15% of American older adults [219] increasing these individuals' risk for malnutrition. In LTC 47% of residents receive modified texture diets (e.g., minced, purées) [234] as they have been used to allow safe ingestion of nutritional requirements in this population [78]. However, based on differences in preparation methods, nutrient composition can be highly variable [104]. This has practical implications especially for older adults with a generally lower intake which decreases with age [122] further confounding the problem. More specific to LTC, when the menus were assessed in LTC, [232] reported nutrient inadequate provision of calcium for the regular texture menu choices as well as potassium, vitamins D and E, folate for both regular and puréed textures. This is in line with older research finding residents did not contain adequate quantities of vitamins and minerals to meet residents needs in a full portion of food when residents typically consumed only a half portion [239]. These findings suggest that not only are LTC residents at increased risk for malnutrition even if residents were able to consume a full portion, but they would also not meet their daily nutritional requirements for several nutrients. So, while monitoring nutritional status is crucial to ensure inadequacy does not translate to deficiency or to facilitate early detection and provide an opportunity for intervention, food must also be as nutritious as possible to ensure adequate nutrient consumption is attainable.

There is currently a lack of tools to quantitatively and objectively assess the nutritional density

of purées. To in part address this, international definitions for modified texture foods (including purée) were recently released by the International Dysphagia Diet Standardization Initiative (IDDSI) ([47]). However, implementation of these international definitions does not address nutrient density beyond purée consistency and adoption may be limited in practice. An automated imaging system may help reduce variance within or between human assessors due to differences in learning or experience; a seasoned purée cook has more intuition about what makes a safe and nutritious purée than a new cook ([104]). A system that can quantify the concentration of the purée could reduce cost and time while providing insight into nutrient density of a purée in health care settings. More specifically, optical imaging systems may provide a powerful solution to this problem. These systems use the same type of information (visible optics) as what is conducted by human assessor; however, computational models provide objective and repeatable predictions.

Recent advances in machine learning have been successfully applied to a vast range of fields from object recognition to pharmacy and genomics ([135]). Specifically, deep neural networks (DNNs) are biologically inspired by the visual cortex for decision making ([20]), and have been used with great success for specific complex tasks such as speech recognition ([97, 54, 85]), object recognition ([130, 91, 136, 213]), and natural language processing ([21, 51]). In image classification and other applications, however, there is often insufficient training data to properly train a conventional DNN due to the nature of supervised learning which require a large number of network parameters and an abundance of labeled training data. In the case of puréed food analysis, data insufficiency becomes a prominent concern due to the limited amount of available labeled data. Labeled data requires the acquisition of spectral and texture information of the puréed food via imaging, and the cumbersome manual labeling process of the images by trained personnel. Leveraging instead, unsupervised learning techniques may provide alternative solutions when paired with food biophotonics which I describe in more detail in Chapter 6. Borrowing from the field of biomedical optics, photon migration models have been used to estimate quantitative tissue properties such as blood oxygen saturation and hemoglobin concentration ([22]). Though primarily used in biomedical applications, these models provide a theoretical basis for quantitative nutritional assessment using optical imaging data. The remainder of this section provides a primer on some key fundamental biophotonic principles as follows.

Figure 2.2 provides some intuition some key biophotonic principles showcasing where higher concentration of puréed food means light coming in has more chances of bumping into food particulates and results in it appearing darker from less light bouncing back (reflectance mode), or less light goes through (transmittance mode). Conversely, for more diluted puréed foods, there is more water and fewer food particles for the light to interact with so samples appear lighter from more light bouncing back (reflectance mode), or more light goes through (transmittance mode).

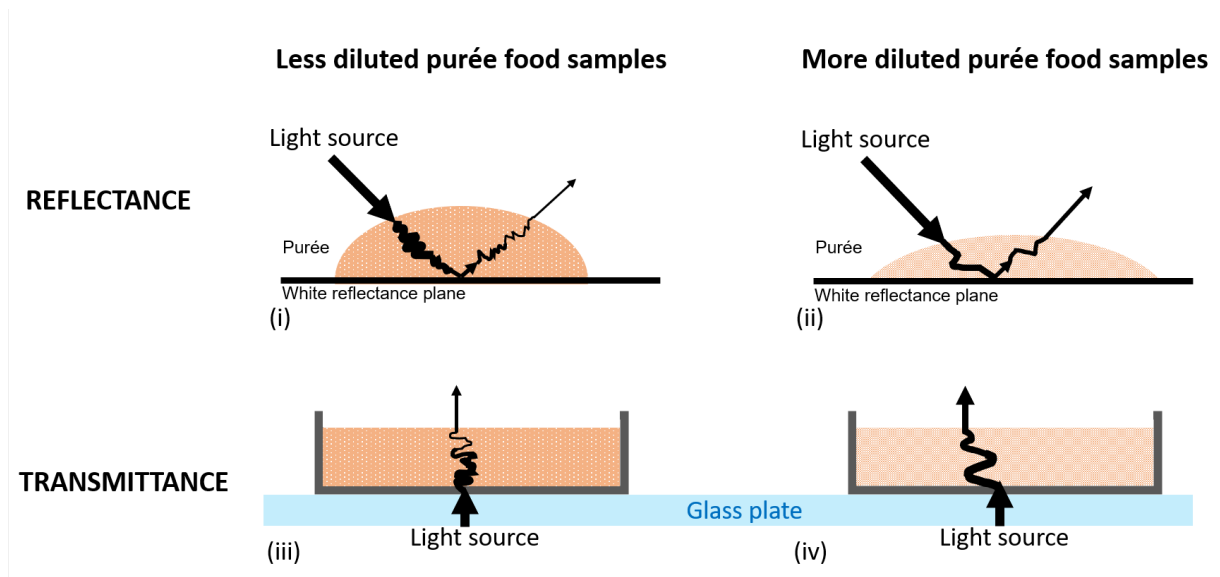


Figure 2.2: Comparing relative scattering events in reflectance (i, ii) and transmittance (iii, iv) modes in higher and lower concentrations of puréed food samples. (i) and (iii) Depict a relatively less diluted purée food sample. There is relatively less water compared to more diluted samples so there are more scattering and absorption events per unit length (i.e., higher  $\mu_s$  and  $\mu_c$ ); greater reduction in reflected photon intensity. (ii) and (iv) Depict a relatively more diluted purée food sample with relatively more water and therefore fewer scattering and absorption events per unit length (i.e., lower  $\mu_s$  and  $\mu_c$ ); less reduction in reflected photon intensity.

So there you have it, how bright or dark a food appears can tell us about how much light is reflected or transmitted, absorbed, or scattered. The more concentrated a sample, the more opportunities there are for light to be absorbed or scattered. And these principles can tell us about the underlying composition of foods. For example, fats are highly scattering substances and iron, and vitamin A absorb light at specific wavelengths [95, 253]. Colour can also play a role in the underlying composition as different types of absorbers appear to absorb certain wavelengths and reflect others to appear specific colours (e.g., vitamin A appears orange). Since we can use these visually observable features to distinguish between types of foods with our own eyes, the goal is to train a computer to do the same through deep learning.

#### UP NEXT...

We've now seen the general steps for going from a food image to drawing nutrient intake estimation (segmentation, feature extraction, classification, inference), as well as opportunity areas within these steps. We discussed the need for disentangling sources of error, as well as key considerations for selecting a training and validation dataset and discussed the theoretical potential for incorporating food biophotonics for enhanced composition analysis. With this fresh on our minds, we are ready to continue our journey towards further enhancing the field. To set our stage, we must first understand the key players specific to food intake tracking in LTC. Afterall, before we can “do better” we must first “know better”, which is where Chapter 3 begins.



Table 2.1: Summary of popular food databases and their appropriateness for use in LTC.

Dataset	Content	Portion	Orientation	Acquisition	Label Level	Accessibility	Ref.	LTC Requirements		
								Top	Pixel	Intake
PFID	USA Fast Foods	full	variable or angled	Restaurant + lab	image-level	open source	[41]	N	N	N
TADA	USA Foods	n/a	n/a	Lab	n/a	proprietary	[152]	N/A	N/A	N/A
Food85	Japanese Foods	n/a	n/a	Web + prev. dataset	n/a	proprietary	[99]	N/A	N/A	N/A
Chen	Chinese Foods	full	variable or angled	Web	image-level	open source	[42]	N	N	N
UEC Food-100	Japanese Foods	full	variable or angled	Digital camera	bounding box	open source	[156]	N	N	N
UNICT-FD889	Variety of Foods	full	variable or angled	Smart phone	image-level	open source	[72]	N	N	N
Food-101	USA Foods	full	variable or angled	Web	image-level	open source	[27]	N	N	N
UEC Food-256	Japanese & Other Foods	full	variable or angled	Digital camera	bounding box	open source	[118]	N	N	N
Food201-Segmented	USA Foods	full	variable or angled	Web	pixel-level	open source	[163]	N	Y	N
Menu-Match	Restaurant Foods (Asian, Italian, Soup)	full	variable or angled	Digital camera	image-level	open source	[18]	N	N	N
UNIMIB2015	USA Dining Hall	full+leftovers	top-view	Digital camera	image-level	open source	[49]	Y	N	Y
UNIMIB2016	USA Dining Hall	full	top-view	Digital camera	pixel-level	open source	[49]	Y	Y	N
VireoFood-172	Chinese Foods	full	variable or angled	Web	image-level	open source	[112]	N	N	N
ChineseFoodNet	Chinese Foods	full	variable or angled	Web	image-level	open source	[43]	N	N	N
ChinFood1000	Chinese Foods	full	variable or angled	Web	image-level	open source	[77]	N	N	N
Recipe1M+	USA Foods	full	variable or angled	Web	image-level	open source	[153]	N	N	N
FoodX-251	Variety	full	variable or angled	Web	image-level	open source	[117]	N	N	N
AFood	Variety	full	variable or angled	Web	image-level	open source	[137]	N	N	N
CROCUFID	Variety	full	variable or angled	Lab	image-level	open source	[224]	N	N	N
FoodDD	Entire Foods (e.g., whole apple)	full	variable or angled	Smartphone, Web	pixel-level	open source	[208]	N	Y	N
ISIA Food-500	Variety	full	variable or angled	Web	image-level	open source	[164]	N	N	N

## Chapter 3

# What We Know: Expert User Knowledge and Needs Assessment

In the last chapter we were primed from the literature with the opportunity for an easy to use, accurate, and comprehensive food intake system designed with LTC in mind. However, to ensure efforts can have real-world impact, the perspectives of end-users is essential. This chapter describes the end goal of collaborating with representative end users to design a novel prototype system for Automated Food Imaging and Nutrient Intake Tracking (AFINI-T), and how we sought to accelerate uptake of novel technological solutions through practice-informed research. In this chapter I seek to address my thesis C1.1 to ensure the system is user friendly. More specifically, two objectives to achieve this aspiration were:

- To **identify practice-relevant problems** through user-centered participatory design (O1).
- To **remove feasibility-related barriers** to uptake (O2).

To meet these objectives, we sought to:

- **Understand workflow** and the problem space (identifying primary and secondary users, the flow of food and fluid intake charting currently in place, including user perceptions of workload of the current system as described in Chapter 5 (A)).

### This chapter contains content previously published from...

**KJ Pfisterer**, J Boger, A Wong. Prototyping the Automated Food Imaging and Nutrient Intake Tracking (AFINI-T) system: A modified participatory iterative design sprint. *JMIR Human Factors* 2019;6(2):e13017. doi: <http://dx.doi.org/10.2196/13017>. On this paper, K.J.P. was the main contributor from project inception to planning, implementation, data collection, analysis, interpretation, and writing.

- **Conduct a needs assessment** within the problem space (B).
- **Establish functional criteria** for usability and feasibility including user interface requirements (C).

The remainder of this chapter describes the process used to conduct rapid prototyping through user-centred participatory iterative design adapted from the Google Sprint method, results including interface prototypes, and discussion and limitations [180].

### 3.1 Overview of Design Strategy

With the end goal of creating a *Goldilocks’ quality horizontal prototype*, we implemented an iterative participatory iterative design process modeled off the Google Sprint framework to develop and evaluate this prototype for monitoring food and fluid intake in LTC [126, 169]. Here, *Goldilocks’ quality* refers to having the “just right” amount of fidelity to elicit useful feedback from users without having to build an entirely functional prototype [126] and a *horizontal prototype* refers to a user interface-based design to allow user feedback on an early-stage conceptual walk-through of the process [171]. For the creation of the prototype, the stages conducted were as follows:



The purpose of Stage 1, Design Ideation, was to engage with end-users as collaborators to establish design directions. Specifically, we sought to understand current workflow, evaluate priorities, understand perceived workload of the current system, and identify potential project advisors. The output from this directly informed Reflect and Storyboard (Stage 2) and Usability Assessment (discussed in Chapter 5).



The purpose of Stage 2, Reflect and Storyboard, was to use storyboarding to generate solution concepts of the user interface and system output that reflect identified needs and priorities for project advisors’ critique in the next stage.



The purpose of Stage 3, Storyboard Critiques, was to assess the storyboards created through collaborating with expert users to establish design directions and to finalize solution concepts for incorporation into Stage 4’s design of the Goldilocks’ quality prototype.



The purpose of Stage 4, Design of the Goldilocks’ Quality Horizontal Prototype, was to create low-fidelity prototypes by incorporating the most promising solution concepts identified through the storyboard critiques in previous stage.

The design process was guided by my personal work experience at the Schlegel-UW Research Institute for Aging (RIA) as the Assistant Research Coordinator as well as several conceptual frameworks. More specifically, when I started my role at the RIA, the Schlegel Villages were undergoing a major shift towards changing the culture (and language) of aging. This was being conducted through participatory action research and over the course of a year, there were tangible, measurable differences towards resident centred care and putting living first [38]. I had logged this away as a clearly effective strategy across all roles within the LTC setting. But for my thesis, since I was developing technology to support enhanced quality of care, I needed a complementary perspective to support Age-Tech development. When I was introduced to the google sprint methodology, it seemed a natural fit with an opportunity for adaptation to my purposes including a modifications for a more rigorous evaluation using standardised tools (e.g., Subjective Usability Scale, Raw Task Load Index which are included in Chapter 5. While the sprint methodology has long-standing success in industry, to my knowledge, it hadn't been adapted for use in a health care setting. However, by combining sprint methodology with additional conceptual frameworks, I felt confident this approach would be an effective strategy for rapid prototyping in LTC. The following pertain to the considered conceptual frameworks: (1) conducting interdisciplinary research [24, 37]; (2) leveraging user-centered design and participatory design [185, 203]; (3) applying rapid prototyping methodology via a modified Sprint [126, 171]; and best practices for user interface design [171, 105, 14, 114, 170, 210]. Examples of information flow through each stage is shown in Figures 3.1 and 3.3.

## 3.2 Methods

### 3.2.1 Stage 1: Design Ideation

Stage 1 consisted of a 60-minute workshop in which three activities were completed: Activity 1: The “Ask the Experts” activity; Activity 2: Priority ranking survey completion, and Activity 3: “Vote with dots” exercise to keep participants engaged and reflect on priorities. Three research assistants plus the lead author took notes during this discussion and transcribed several comments verbatim. Following the workshop, three informal open-ended interviews were conducted to further inform the problem-space. The lead author took notes during these interviews; several comments were transcribed verbatim.

For the workshop, 21 participants representing 12 LTC and retirement homes were recruited through self-enrollment with following roles: Administrative Assistant, Chef, Dining Lead (similar to a dining room manager), Director of Recreation, Dietary Aides, Neighbourhood Coordinator, Recreation Assistant, Restorative Care, Senior Nurse Consultant, Directors and Assistant Directors of Food Services, Nurse, and Personal Support Workers (PSW). Activities were discussed

with the Schlegel-UW Research Institute for Aging's (RIA) Research Application Specialist for input on how to conduct this exercise successfully with front-line team members.

#### **3.2.1.1 Activity 1: The “Ask the Experts” Activity**

Workshop participants were asked about their experience with food and fluid intake. This aimed to build participants' confidence in the value of their experiences while probing current workflow and problem space.

#### **3.2.1.2 Activity 2: Priority Ranking Survey**

Participants independently completed a survey to evaluate priorities and needs to limit bias. This survey asked about the current charting process (e.g., when it is done, task completion time, barriers and facilitators to task completion). For evaluating priorities, 5-point Likert scales were used to rate 16 statements' importance from “Not Important” (i.e., 0) to “Very important” (i.e., 4) or “Not Applicable”.

#### **3.2.1.3 Activity 3: “Vote with Dots” Exercise**

Modeled from [126], participants transposed their individual Activity 2 responses into a group response by voting their preference using stickers on giant sticky notes to amalgamate opinions, keep participants engaged and to facilitate additional discussion.

### **3.2.2 Stage 2: Reflect and Storyboard**

The data from Stage 1 design ideation was combined with the heuristics outlined below to create a series of storyboard solution concepts. Each storyboard was designed using Balsamiq and included tailored concepts developed for three types of primary users identified in Stage 1: a PSW, registered team, and registered dietitian (RD). As the system is expected to run on iOS based software and hardware to mesh with the current charting practice on iPads, storyboards were loosely based on iOS Human Interface Guidelines [105]; general iOS expectations will need to be balanced with the current electronic health record system in place (i.e., PointClickCare).

In Stage 2, we explored usability from the designer's perspective by applying the heuristics outlined by Shneiderman's 8 golden rules [210] and Nielsen and Molich's 10 user interface design heuristics [170] as well as considering heuristics to support trust cues and credibility [52, 76] while adhering to best practices for user interface design [171, 105, 14, 114]. For example, as shown in Figures 2-3, buttons were designed in accordance with the affordance principle

where visual cues act as clues to suggest how an object might be used [171, 14] and informative labels [114] were included to reinforce affordances. Colour was used to make the buttons appear actionable [14, 114] and users were “rewarded with visual feedback” [114] in the form of confirmation pop-ups, as well as warning and success screens. Finally, inspiration was drawn from three healthcare record systems (e.g., Prognosis EHR, ChiroSpring, and Aprima EHR [106, 44, 68] noted as the top electronic medical records software from 2018 online reviews [36, 215]).

### **3.2.3 Stage 3: Storyboard Critiques**

Five participants self-selected as project advisors during Stage 1’s workshop from the perspectives of PSW, dining lead, LTC RD, food and nutrition consultant, and food/dietary aide. Similar to the sprint process described by [126], storyboard critiques were conducted with each participant. Feedback was gathered through in-person meetings or over a virtual screen sharing teleconference (a Zoom meeting) when it was infeasible to meet in person on areas of interest, utility, or needing improvement using a wire diagram prototype mockup developed in Stage 2. The first author transcribed feedback in real-time with on-going participant clarification and confirmation. The outputs from Stage 3 included spatial heatmaps on preferred design elements and qualitative feedback for additional consideration. These heatmaps provided feedback similar to the vote with dots exercise described in Stage 1 where more popular concepts received more votes.

### **3.2.4 Stage 4: Design of the Goldilocks’ Quality Horizontal Prototype**

Design decisions were informed by heuristics as in Stage 2 [210, 171, 170] and feedback received from the storyboard critiques in Stage 3. The following heuristics were emphasized: universal usability was considered by testing the prototypes with different types of users (e.g., academics, PSWs), providing informative feedback and error prevention, the output this stage (Stage 4) was a Goldilocks quality horizontal prototype. This included interfaces for each of the three levels of primary users currently involved in residents’ food and fluid intake charting (i.e., PSW, registered nursing team, and RD).

## **3.3 Analyses**

Given the nature and size of this pilot study, a preliminary thematic analysis was used for qualitative components (e.g., discussions, comments, verbal/written feedback) that was combined with descriptive statistics for quantitative information including the average ( $\mu$ ), standard deviation

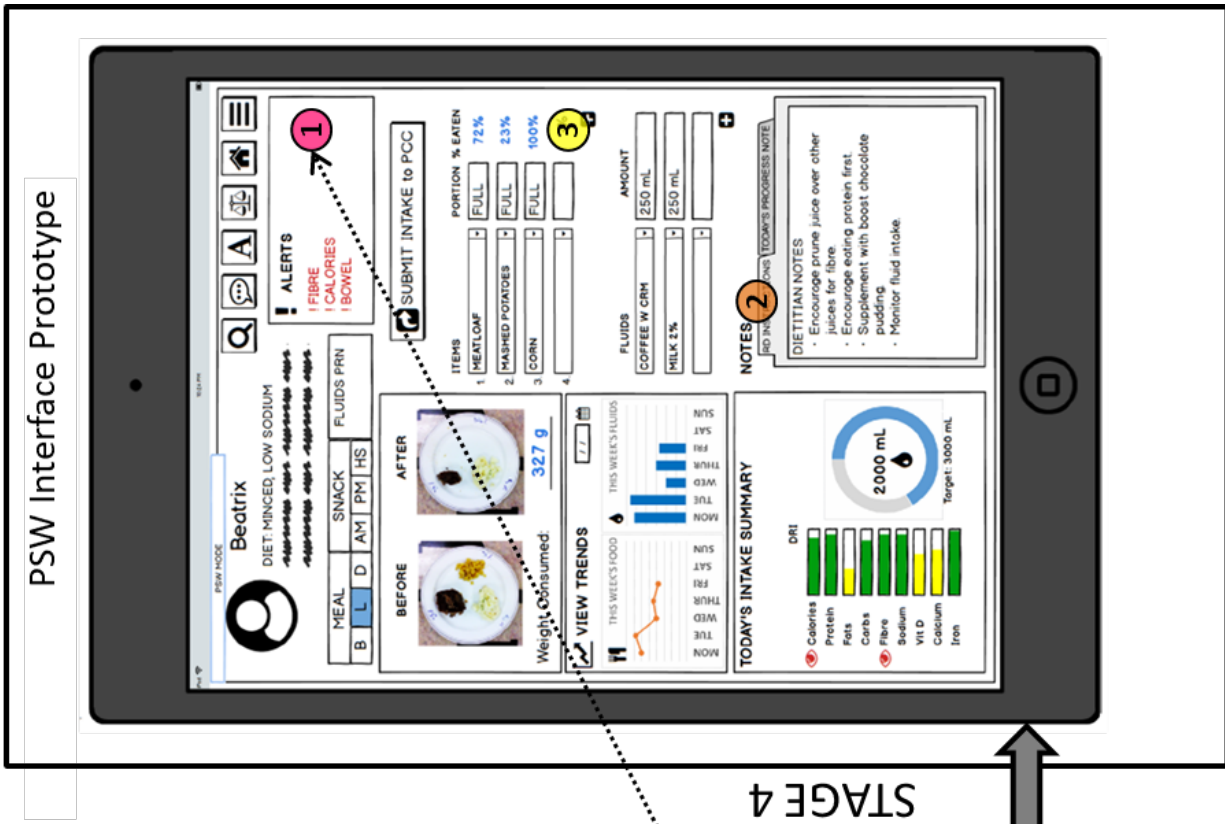
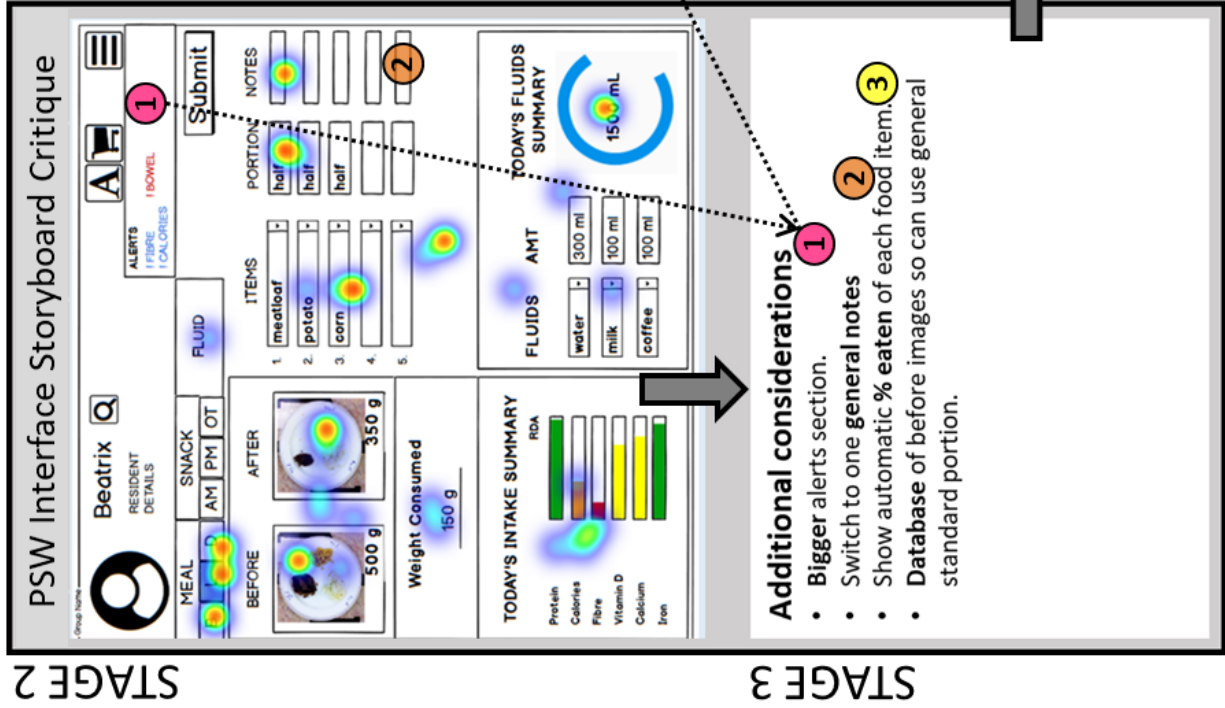


Figure 3.1: Stage 2 PSW user interface. Output from Stage 3 included a heatmap on the most promising aspects (red indicates more votes, n=5) with qualitative feedback highlights for additional considerations. The right pane illustrates an example of the prototype interface. Numbers correspond to the flow of information and adapted feedback from Stage 2 through to 3 and 4 using the first example (#1 in pink) to further illustrate flow with the dashed arrow.

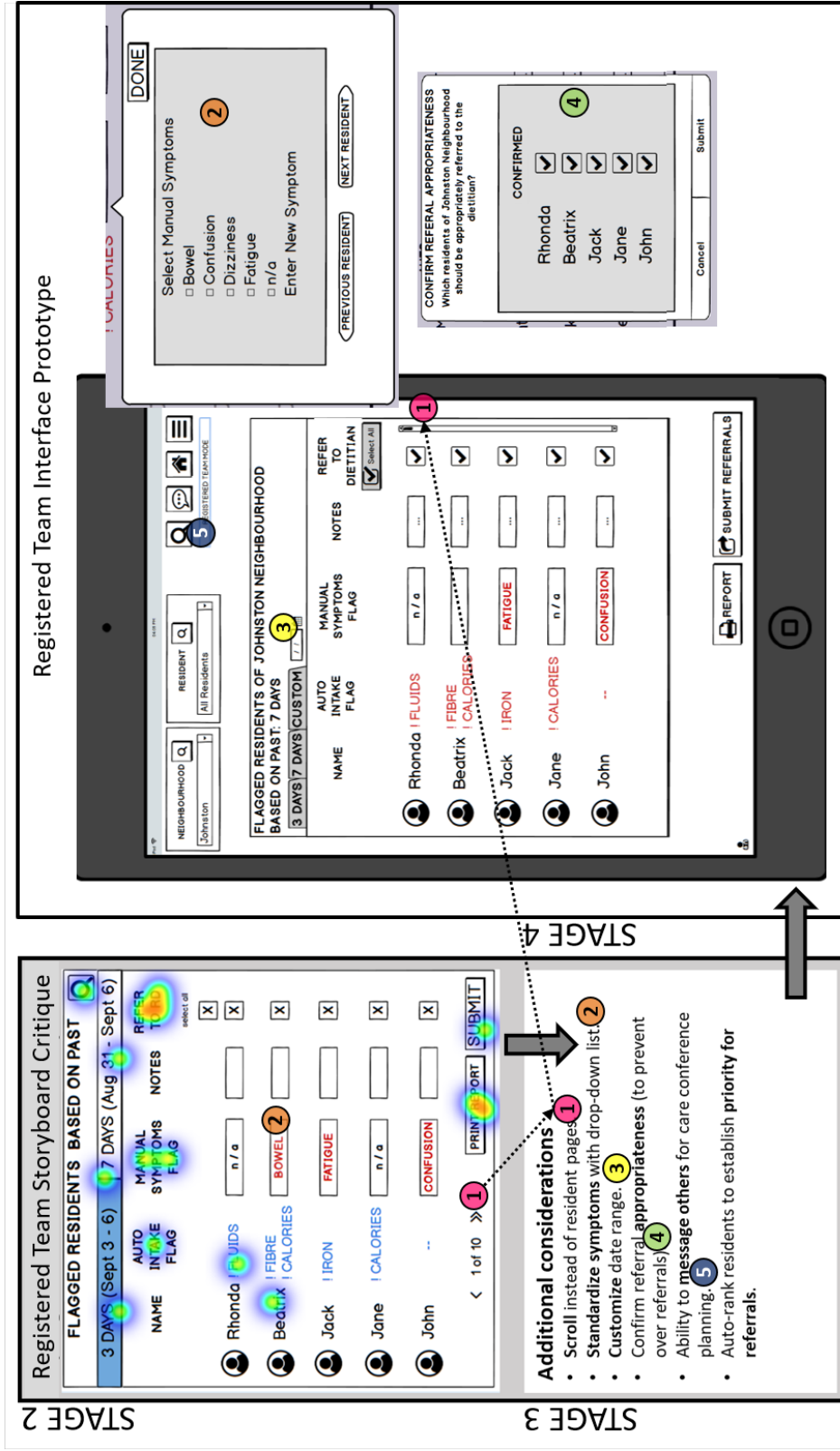


Figure 3.2: Stage 2 Registered Team user interface. Output from Stage 3 included a heatmap on the most promising aspects (red indicates more votes, n=5) with qualitative feedback highlights for additional considerations. The right pane illustrates an example of the prototype interface. Numbers correspond to the flow of information and adapted feedback from Stage 2 through to 3 and 4 using the first example (#1 in pink) to further illustrate flow with the dashed arrow.



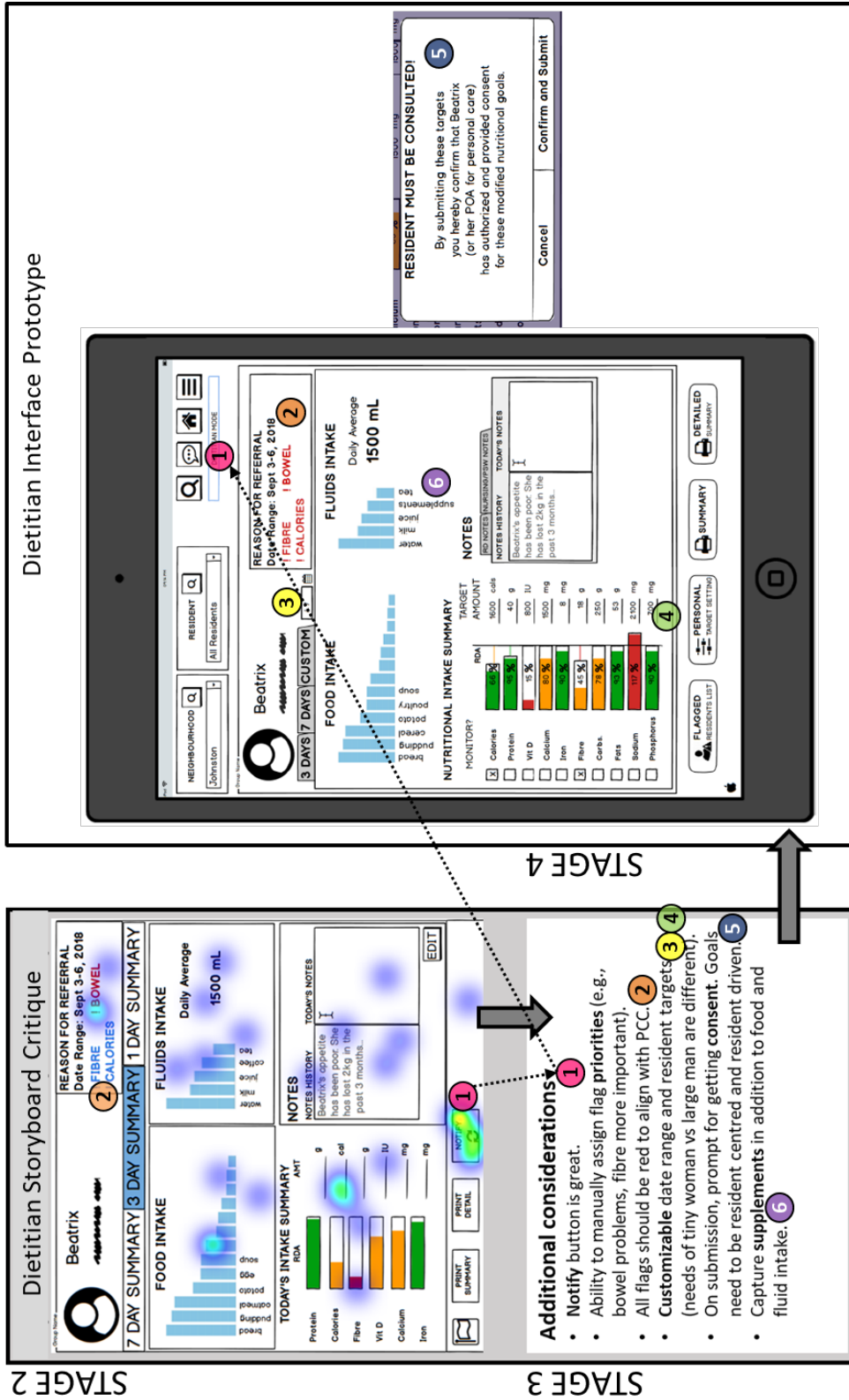


Figure 3.3: Stage 2 Registered Dietitian user interface. Output from Stage 3 included a heatmap on the most promising aspects (red indicates more votes, n=5) with qualitative feedback highlights for additional considerations. The right pane illustrates an example of the prototype interface with a sample pop-out box. The numbers correspond to the flow of information and feedback from Stage 2 through to 3 and 4 using the first example (#1 in pink) to further illustrate flow with the dashed arrow.

( $\sigma$ ), mode, and median scores [198]. A weighted average was used to analyse Likert survey questions, excluding “Not Applicable”, to yield a ranking of each statement.

## **3.4 Results**

### **3.4.1 Stage 1: Design Ideation**

Results from Stage 1 pertained to Objective 1: Address a practice-relevant problem through user-centered participatory design (Goals A, B) and Objective 2: Remove feasibility-related barriers to uptake and are as follows (Goal C):

#### **3.4.1.1 Goal A: Identify primary users**

Based on discussions, surveys and interviews in Stage 1, the following primary and secondary users were identified Figure 3.4. Based on best practices and policies for tracking and monitoring food intake tracking of residents, PSWs are responsible for charting. Nurses oversee based on how intake fits into other domains of healthfulness (e.g., constipation), and the dietitian oversees based on referrals for more closely monitoring residents at highest risk. That said, DFS were also excited about this from a food planning perspective. Saying they could use the feedback of what people are actually eating as well as to inform how to tweak recipes for enhancing nutrient density. Dietary Aides, Neighbourhood Coordinators (NC) and the Recreation Team may also interact with the AFINI-T system but to a lesser extent. From the policy side Administrators and Directors of Care (DoC) may leverage output. Family members often ask for reports but the Resident and family are not responsible for tracking/monitoring of food intake directly; the resident is not involved in the recording of food and fluid intake tracking.

#### **3.4.1.2 Goal A: Understand workflow and problem space**

Based on discussions, surveys and interviews in Stage 1, the current workflow infrastructure is illustrated in Figure 3.5. Typically, food is prepared in an industrial kitchen and held at a temperature until it is ready to be distributed to each neighbourhood. At each neighbourhood, dietary aides plate and serve the meals with support from PSWs. PSWs, registered nursing team, RDs are primary users and PSWs conduct charting of food and fluid intake on iPads. This charting is completed whenever primary users have time which could be during meal service or retrospectively, consistent with [7]. In a follow-up discussion with the organization-wide director of food services who is responsible for policy, she indicated that conducting food intake in real-time is mandated (as opposed to retrospectively), but from the workshop discussion, it is clear there is a gap between policy and practice. While the workflow of AFINI-T is congruent with

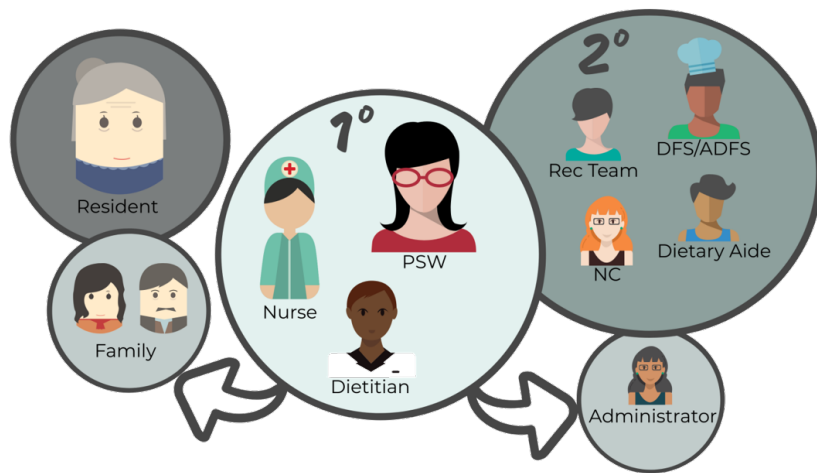


Figure 3.4: Primary and secondary users of the AFINI-T system. Primary users are Nurses, Personal Support Workers (PSWs) and the Dietitian. Secondary users are the Recreation Team (Rec Team), Directors and Assistant Directors of Food Services (DFS/ADFS), Dietary Aides, and Neighbourhood Coordinators (NC).

this mandate, a solution to support this mandate in practice may require policy modifications. For example, one person may need to be assigned to the sole task of tracking food and fluid intake during mealtime, which means they would be unavailable to help with residents’ care needs for the duration of the meal. Changing policy is outside the scope of the current AFINI-T project but having sensitivity to this issue provides helpful context and informs that this may be a potential barrier to uptake of the system in practice.

The main sources of nutrients are through food, fluids, and supplements. Participants felt supplements were well documented as they are tracked using the medcart system. That left fluids and foods, both of which are crucial to monitor. Feedback from the nutrition research expert informed that while fluids receive the majority of focus in the literature and is important for monitoring hydration status, the more clinically relevant area for nutritional status is food. With fluids, it is easy to measure, the question is more about “*will* people measure fluid intake” due to conflicting priorities. With food intake however, the question is more “*can* we measure food intake” as it is more challenging. While measuring food intake is a proxy for nutritional status, measuring food intake gives us a sense of *why* something might be going wrong (in combination with biomarkers). While the utility of current food intake measurements is limited, fundamentally, it raises awareness to some extent regardless of whether the measurements are inaccurate (e.g., food spills). Currently in LTC, this charting conducted is based on visual assessment of the plate after a meal potentially missing food spills and is accepted as an appropriate limitation.

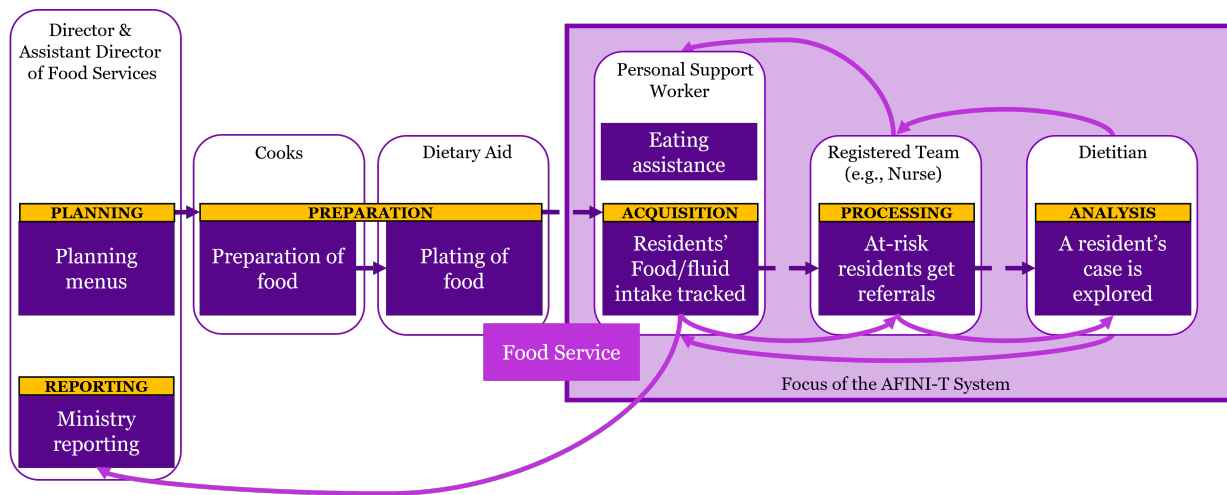


Figure 3.5: Current workflow of food service and food and fluid intake charting.

Better, more reliable measurements can enable more meaningful assessment of probing when, how, and why, something may be going wrong informing intervention opportunities. In line with this perspective, workshop participants' domain knowledge was helpful to further converge the focus on first monitoring food intake in a novel way.

Participants felt one facilitator to recording food and fluid intake was the PCC system enables fluid intake measurements at the mL level. Across two sources (the "Ask the Experts", Survey), they thought fluids were fairly well documented/were more confident in those values (more useful measurements) because there is less ambiguity around how much since it is collected to the mL. As a result, as the first step for enhancing nutrition monitoring, I chose to focus on a system to support food intake since this was the largest barrier with least utility.

More specifically, regarding the current system, respondents appreciated the ability to track fluids so they need not manually add and the output has units (mL). While the current system is dependable, substantial barriers and limitations were identified regarding the effectiveness and accuracy of the current system. One workshop participant shared, "What's being collected for solid food isn't useful. It's so high level and minimal can't make use of it. [We] can't infer anything regarding health or category of at-risk. [We] look at last 7 days, see 'they had 75% of a meal so they're eating well', but it doesn't say anything. [We] don't get a lot of info from the charts."

Insufficient time, data inaccuracy, unreliability and non-standardized measurements were identified as the largest barriers for task completion. Additionally, inability to differentiate between types of foods, and lack of relation to original serving size lead to data interpretation difficulties.

For example, some residents prefer half portions; if they eat half of their portion, this could be recorded as 50% (i.e., half of the serving they received), or it could be input as 25% (i.e., 1/4 relative to the full portion). There is no guarantee the proportion is input accurately or consistently. These themes were apparent through two sources, the “Ask the Experts” as well as on the survey. For more detail regarding the current system’s retrospective analysis of perceived user workload, see the sections of Table 3 pertaining to the “Current” system.

#### **3.4.1.3 Goal B: Conduct a needs assessment of problem space including priority areas**

Workshop participants were asked to rate need statements’ importance. The top three ranked priorities were tied between: (1) “ease of use” and “accuracy” ( $\mu=3.9$ , mode: “very important”, 15/16 votes), (2) “reliability” and “maintenance” ( $\mu=3.9$ , mode: “very important”, 14/16 votes), and (3) “The system should work well with PointClickCare.” ( $\mu=3.8$ , mode: “very important”, 12/16 votes).

The following five themes emerged as wishes for a novel system to extend beyond the current infrastructure: (1) being able to leverage weight of food as a ground truth instead of relying solely on subjective proportions, (2) having the ability to track trends over time, (3) being able to discriminate between types of food, (4) being able to include fluid intake as well to discriminate between types of fluids, and (5) operating the system in different modes to accommodate various use cases (i.e., in the dining room vs for in-room service). One additional, complementary theme relevant to priorities identified independently through three interviews was the need to support prioritising referrals that consider symptoms and risk flags severity. One project advisor articulated, “There is 1 Registered Dietitian for 300 residents. It’s impossible to track properly . . . People are often missed because nurses aren’t identifying properly. . . If charting were accurate, this would help with the referral process.”

#### **3.4.1.4 Goal C: Establish functional criteria for usability and feasibility**

The current system mode time to complete task defined the time completion target: 10-14 minutes maximum per neighbourhood (i.e., “ward”) of 16 residents. Of the 21 workshop attendees, 11 self-identified as being involved in charting resident food and fluid intake and were asked about the amount of time required to complete intake charting for each of food, fluid or snack. Survey responses are outlined in Table 3.1

### **3.4.2 Stage 2: Reflect and Storyboard**

Design decision inspirations were drawn from Prognosis EHR, ChiroSpring, and Aprima EHR [62–64] (Table 3.2). These inspirations led to a full, yet organized screen for each of the inter-

Table 3.1: Summary of length of time required to complete food and fluid intake charting for one neighbourhood comprised of 16 residents (Stage 1). \* n is the number of responses with the mode rating out of N, the total number of responses.

Charting Type	Mode Time (minutes)	(n/N responses)*	Time Range (minutes)
Food (meal)	10-14	30% (3/9)	<10 to 25+
Fluid	10-14	40% (4/10)	<10 to 25
Snack	<10	64% (5/9)	<10 to 19

faces (Figures 3.1 and 3.3). Many of the inspiration examples and informed design decisions described in Table 2 are illustrated as the output from Stage 2’s storyboard critiques ( 3.1 and 3.3). These figures also depict how Stage 2 output informed Stage 3 (including the heatmap overlay in the Stage 2 panes) and Stage 4 (discussed further in subsequent stages).

### 3.4.3 Stage 3: Storyboard Critiques

In Figures 3.1 and 3.3, the heatmap overlaid on the Stage 2 pane illustrates the most promising concepts voted by project advisors (e.g., drop-down items and meal-specific tabs). The lower left-hand pane depicts additional considerations captured through discussion. For example, on the PSW interface, building a database of pre-meal images to save time along with the domain knowledge that this solution would work for around 90% of the population (i.e., 10% may require special pre-meal images due to residents receiving non-standard portion sizes).

When design decisions were made, the advisor’s perspective was considered and weighed accordingly. As the team of project advisors was relatively small, it will be important for the final product to be tested with a larger sample of users to make sure concepts captured generalize to users’ needs more broadly. Generally, the project advisor feedback on storyboard solution concepts (Stage 3) were well received with project advisors actively engaged (e.g., “this system will give me confidence on neighbourhoods I don’t know.”). Regarding the solution concepts considered together, one advisor said, “This is a fantastic tool for dietitians to use. I can see it is needed. It will advance the profession and advance quality of care for residents. And it will happen in my lifetime”. There were no negative comments.

### 3.4.4 Stage 4: Design of the Goldilocks’ Quality Horizontal Prototype

Design heuristics were applied in the four ways and sample output from this stage is illustrated in the right pane of Figures 3.1 and 3.3. First, related to universal usability, mapping was considered

Table 3.2: Summary of key inspiration concepts from commercially available online healthcare tools. Numbers in brackets correspond to corresponding design decisions; many are highlighted in Figures 3.1 and 3.3.

Inspiration Example	AFINI-T Prototype Design Decision	
Click saving features (1)	Smart tabs opening based on time of day.	
Tap on name (2)	Tap on a name to open the profile for loading a resident profile.	
Prognosis EHR [106]	Clinical snapshot (3)	“Today’s Intake Summary” clinical snapshot pane.
	One check default clicks (4)	“select all” capability for registered nursing team referral.
	Input and edit notes (5)	Ability to add and edit notes
	Customizability (6)	Planned: panes to be moved, expanded, minimized.
	Solution from scheduling to billing to claims to task management (7)	Supports the process from intake tracking to referrals to further investigation.
Ability to skip questions (8)	Incomplete data can be entered and edited later or skipped (with a warning).	
Last and current visit notes visible(9)	RD pane (not shown) has “Notes History” right beside “Today’s Notes”	
Hand off from one person to the next (7)	Supports the process from intake tracking to referrals to further investigation.	
Adaptive learning capabilities with intelligent navigation (10)	Planned: smart food suggestion/selection based on learned preferences and already selected items. Automatically highlighting/changing focus after sub-tasks completed (e.g., progress note pane becomes in-focus after intake completed)	

through matching the system with users language and familiar concepts in reality (e.g., Figure 2 contains tab names for snacks such as “AM”, “PM” and “HS” refer to the morning, afternoon, and evening snacks respectively) [171, 170]. Second, informative feedback on a change of state was provided [210, 171] when users attempted to submit or track an action, a pop-up there is a pop-up banner at the bottom of the screen (not shown). Third, error prevention [210, 171, 169] was incorporated through limiting types of responses and providing feedback. For example, the PSW interface would prompt for a picture, or a progress note before submission with the ability to finish charting at a later point of the meal service. Fourth, efforts were made to reduce short-term memory load and enhance visibility/discoverability [210, 171, 169] by placing the workspace into panes with all information accessible on one screen. Other features included making “smart” suggestions when selecting items or filling out portion sizes. For example, notes entered from RD interface (not shown) would auto-populate on RD instructions tab in the PSW interface.

### 3.5 Discussion

Regarding timeliness in the time-constrained dementia care context, one substantial difference between previous work on developing technology for consumer-centred nutrient intake tracking (e.g., [128, 163, 176, 189]) and the work presented here is that the purpose of our technology is to support tracking in a regulated LTC environment. This means considerations regarding consumer uptake and use are different than with general consumer market. For example, the novelty does not arise from tracking food and fluid intake per se; this is something that is already mandated for at-risk residents. Instead, the novelty is in improving the method for tracking beyond the current system in place. Other work involving diet tracking apps tend to focus on weight loss and are meant for tracking of an individual’s food intake by the individual. Here, we seek to leverage LTC as an infrastructure already in place to conduct more efficient mandated multiperson monitoring.

The role of nutrition as part of a holistic care plan for individuals living with dementia is discussed in the 2015 European Society for Parenteral and Enteral Nutrition (ESPEN) guidelines. They indicate that malnutrition contributes to disease progression and increased caregiver burden and that “non-pharmacological strategies like nutritional interventions are of particular interest as part of disease management” [231]. There is evidence to suggest that adhering to a particular pattern of dietary intake (e.g., the Mediterranean diet) is associated with reduced cognitive decline [228]; however, these authors state “more conclusive evidence is needed to reach more targeted and detailed guidelines to prevent or postpone cognitive decline”. Leveraging the necessity to monitor at-risk residents living in LTC through a novel, objective approach to food intake tracking, may be beneficial for gaining new insights for defining guidelines.



Specifically considering the dementia care context and nutrition's role in the process, according to a 2016 systematic review [1], relatively few interventions have been conducted to explore the effect of food intake in mild cognitive impairment or dementia. They conclude that all 43 controlled interventions were at risk of bias and resulted in no consistent evidence either in support or against the effectiveness of nutrition focused interventions [1]. By providing an alternative method for tracking, we seek to improve upon how these allocated resources are used and aim to provide more informative data. One future direction of the AFINI-T system is to use artificial intelligence to learn food preferences.

In terms of the physical design requirements, additional discussion is required as the exact location to house the system remains unclear as do size restrictions. What was gleaned, however, is that the AFINI-T system must work on the iPad since this is what is currently in use. The acceptable level of accuracy target was not well defined with project advisors. That said, we can turn to the literature for some insight and important context. There is a tendency for frequent overestimation of food consumption [211, 39]; in terms of degree of inaccuracy estimates of food intake are typically over 50% for food items [39, 23] with reported over-estimation of food 22% of the time [211]. Furthermore, the source of error is said to be random [23] implying compensation is not possible with current methods. With the AFINI-T system, we should set our targets to be much more stringent because the automated image-based system removes subjectivity. Careful documentation and exploration of the conditions where the system does not perform optimally will be necessary. One challenging situation is plates where the food items get mixed up over the course of the meal. However, even more crude estimates, where we assume equal eating distributions across types of foods for a plate average, would still improve on the current system as it eliminates subjectivity, and reflects relative changes in mass and volume.

## 3.6 Conclusion

Using a multi-stage participatory iterative design sprint of a Goldilocks quality horizontal prototype for the Automated Food Imaging and Nutrient Intake Tracking (AFINI-T) system, a user interface prototype was collaboratively designed and primed for system evaluation. Specifically, practice relevant problems were identified (the desire for more accurate food and fluid intake tracking for enhanced clinical utility), and requirements were defined (e.g., need to integrate with existing electronic health record system). Functional criteria for usability and feasibility were additionally defined (e.g., ease of use, accuracy, reliability, and ease of maintenance), with reference time to complete current food and fluid intake tracking (e.g., 10-14 minutes each for meals and fluids for 16 residents). This provides context for the remainder of this thesis.

### 3.6.1 Key Contributions

- Novel application of a modified participatory iterative design sprint in the LTC setting.
- Primary and secondary users identified, with preference for level and flow of information and translated to requirements for an enhanced food and fluid intake system (emphasis on food intake tracking).
- User interface for primary users reflecting requirements designed.

#### UP NEXT...

We've seen how to effectively collaborate using a new infrastructure for collaborative design. We've learned pain points of the current system and re-imagined food intake data utility with needs translated to requirements. We've designed the interface guided by end-users. Now that we've learned how we'd like to "know better", we can embark on "doing better" - which is where [Chapter 4](#) takes us next.

## Chapter 4

# Re-imagining Food Intake Tracking to “Know Better”

Now that we’ve seen how users would prefer to interact with the system and have a sense for priorities, we can begin working on the “guts” of the system. Recent reviews corroborate that further work is needed for developing an accurate, automated systems [218, 61] and describe the need for considering more complex meal scenarios (i.e., beyond solid, separated foods, or synthetic foods) [189, 218] and require adequate statistical analyses of methods [61]. From the food imaging perspective, this involves four subcomponents which are expanded on in the remainder of this chapter - the first 2 in Section 4.1 and the latter two in Section 4.2:



Establishing *where* on a plate there is food (i.e., segmentation) for ...



Establishing *how much* food is present (i.e., volume estimation) so we can infer food intake relative to an initial portion.



Establishing *which foods* are present (i.e., classification) so we can tease out food intake at the food-item level and address the current shortcoming of assuming equal consumption across foods.



Computing *nutrient-level intake estimates* by linking food-item level intake with recipes.

### This chapter contains content previously published from...

**KJ Pfisterer**, R Amelard, AG Chung, B Stryk, A MacLean, HH Keller, A Wong. When segmentation is not enough: Rectifying visual-volume discordance through multisensor depth-refined semantic segmentation for food-intake tracking in long-term care. (Submitted: *Scientific Reports*). On this paper, K.J.P and R.A contributed equally to this work. K.J.P conceptualized the system, was the main contributor to experimental design, data acquisition protocols, data collection, interpretation, and writing of the manuscript, with additional contributions to the algorithmic design. R.A. was the main contributor to algorithmic design and technical implementation. K.J.P and R.A. conducted data analyses.

Section 4.2 also contains an unpublished manuscript. On this paper, K.J.P, was the main contributor to experimental design, data collection, data analysis planning, interpretation, and writing of the manuscript, with additional contributions to the algorithmic design. R.A. led algorithmic and technical implementation. K.J.P and R.A. conducted data analyses.

## 4.1 Seeing food intake objectively

Automated tools may provide a time efficient, cost effective, and objective alternative to current subjective human assessor estimates. We saw in Chapter 3, the LTC sector requires a system that is reliable, accurate, cost effective and time efficient for measuring food and resulting energy, macro and micronutrient intake [180]. In Chapter 2 we saw that for food intake tracking we must answer: *where* is there food (segmentation), *what* foods are present (classification), and *how much* food remains relative to the initial amount (volume estimation)?

To date, food classification has received the most attention; more-in-depth reviews for classification can be found elsewhere (e.g., [144, 34, 61, 189, 218]), however for completeness, I have included an overview of relevant works preceded with the focus of my work. In-

### Section Summary

This section outlines two colour and volume food intake datasets I collected to capture simulated intake of regular texture and modified texture foods against weighed food records. It further describes how I leveraged AI powered computer vision techniques for food intake tracking through segmentation and volume estimation. The general process was, from an image, determine *where* food was on the plate and take the associated depth map for volume. To answer *how much* food was eaten, for leftover plates, compare to the full portion reference image to establish % food intake both using segmentation maps alone as well as depth-refined segmentation for calculating 2D and 3D % intake error.

#### Key take-aways:

- Depth information improved accuracy.
- We explored “**visual-volume discordance**”, the disagreement between where food pixels are and where meaningful food resides, where segmentation alone necessitated capturing additional context (e.g., when to automatically omit sauce remnants).
- The system excelled at modified texture food estimates and struggled with low-density foods (e.g., salad).

stead, I focus more on food segmentation as a relatively unexplored domain to address food intake from imaging to estimation. For the purpose of this section, we focus on the *where* (segmentation), and the *how much* (volume estimation), as types of food items in LTC are well constrained through monthly menu-planning.

#### 4.1.1 Food Related Segmentation and Classification Progress

While automated tools may provide a time efficient, and objective alternative, little work has been done within the LTC context providing an opportunity for high impact in this field. The following section provides an overview of related work around food classification, volume estimation and nutritional intake estimation within the context of LTC’s requirements. It concludes with an overview of the focus of our proposed solution.

Certainly there has been progress made in food classification and nutrient intake, however, these previous efforts have tended to focus on classification in isolation ( e.g., [144, 34, 61, 189, 218]). Food segmentation is comparatively unexplored but has typically required fiducial markers [176, 252], multiple images [129, 187], manual labelling for *every* food item and each food image [162], or did not predict food areas at the pixel-level (e.g., bounding boxes[207, 120]). However, in LTC semantic segmentation is required when measuring food and nutrient *intake* across a plate as many residents do not consume their entire portion. For food volume estimation, template matching has been popular [93, 243, 40, 110, 175, 193, 96]. However, within the context of LTC, the same food can take on various shapes (e.g., banana in peel versus sliced banana, versus pureed banana) and 47% of the LTC population receives modified texture foods [234]) limiting the utility of template-matching in this context. Others have leveraged stereo reconstruction [57, 190] for volume estimation and 3D point-clouds [194]. However, taking multiple images and building 3D point clouds require additional time, an extremely limited commodity in the LTC context.

An alternative route which has been gaining popularity is depth imaging and structured lighting for mapping food topology [71, 205, 162, 42, 139]. While depth- and structured-light-only approaches are more robust to illumination variation, highly reflective foods (e.g., gelatin, soup) pose a challenge for accurate readings [139]. Leveraging structured light for measuring volume shows promise as a means to address previous shortcomings particularly on the required operator time with accuracy approaching 97.44% [139]. While structured light has been gaining popularity in the agrifood industry for measuring volume [145], monitoring fermentation in bread [108, 229], classifying fruits and vegetables [179] and apple quality analysis [147, 148], there has been very little work on incorporating structured lighting for monitoring food intake [205, 206, 139, 71]. Shang et al. imaged only 9 food replicas and did not report error [205, 206], Fang et al. reported error of 2% but did not disclose how many foods or what was used as ground

truth [71], but is consistent with Liao et al. reporting an average error in weight of 7.5% [139] which is a marked improvement over methods which rely on human assessors with accuracy less than 75% and incorrect measurements up to 66% of the time when assessed retrospectively [7, 39]. One potential reason for few papers leveraging a structured light approach is that until recently, these systems were difficult or expensive to build. However, with recent advances in the Kinect and Intel's RealSense RGB-D cameras, an off-the-shelf solution is now available at relatively low cost (i.e., under \$200) or comes standard in newer versions of iPhones and iPads.

An additional factor limiting progress in food *intake* estimation more broadly is the need for large, complete food datasets [250]. As a result, few attempts have been made to solve this problem, they have high error (up to 400 mL error [162]), or are limited to synthetic foods [142, 143] which do not accurately represent a real-world environment. While we can borrow inspiration from these approaches, they were not designed with the needs of LTC in mind. For example, weight loss was the desired outcome [129, 163, 176, 189, 10, 4] which is inappropriate in the LTC context and they were designed to be used by the user. In LTC we have proxy users where one staff monitors many residents. While these approaches could be modified for use in LTC, in their current form, they target a different purpose (e.g., calorie tracking), still rely on self-monitoring, and do not consider the LTC context for food and fluid intake tracking best practices.

Further consideration is needed to also disentangle sources of error as error assessment is typically not reported or segmentation is coupled with either classification [186, 120, 93, 162, 207, 237, 246] or volume estimation [89]. Others have also identified this as a limitation for accurately predicting down-stream nutritional information [162, 10, 237, 252]. These gaps have been corroborated by recent reviews identifying the need for automated systems [218, 61] which consider more representative foods [189, 218] with the need for more in-depth evaluation and statistical analyses [61].

Within the LTC context the closest technological analog in this domain was a comparison to estimate food waste of regular- and modified-texture diets either with the visual estimation method or by using digital photographs for retrospective analysis [177] both methods required significant operator time as it was a manual process. Borrowing from the hospital setting, a more recent pilot study by Ofie et al. used a combination of RFID scanner, built-in food scale and digital photography method [174]. They yielded promising results in terms of accuracy with relatively high agreement between trained and untrained assessors (ICC = 0.88,  $P < 0.01$ ) and excellent agreement for protein and energy intake between their method and the weighed food method (ICC = 0.99,  $P < 0.01$ ) [174]. These are encouraging results for an adjacent application in LTC. However, weighing each portion or manually imaging each plate before and after a meal to establish what was consumed is time-consuming, may require trained personnel for best accuracy, and is therefore impractical for large-scale adoption in our LTC setting. More broadly, further work is

needed for developing an accurate, objective and cost-effective automated system [218, 61]. Our work seeks to address this gap in the context of LTC.

There is a need to address higher quality tracking with decreased margins of error (e.g., errors up to: 400%, 24-hr recall; 50%, portion size [23], correct estimation of intake occurs only 44% for correct portion size estimation in LTC, but is low as 38% of the time with delayed recording [39]. In line with these opportunity areas and current human-computer interaction trajectories [2], we seek to improve trust and transparency by focusing on developing an explainable system which approximates human assessors but with enhanced objectivity, accuracy, and precision. Specifically, we describe and evaluate a novel fully automated depth-refined multisensor food intake tracking system. Here the depth-refined segmentation and volume estimation have been decoupled to disentangle and assess potential sources of error. Through this decoupling, we seek to enhance reliability for eventual integration with nutritional intake estimation, and to reduce potential barrier to uptake in practice. We designed the segmentation system to be used in clinical settings (such as LTC or hospitals) with acquisition consistent with LTC food and fluid intake visual assessment procedures. Our system is comprised of an RGB-D camera, a novel deep convolutional neural network encoder-decoder food network, called EDFN, with fused output from superpixel processing. It also incorporates depth information for enhanced segmentation and volume estimation. The use of a single RGB-D camera brings simplicity over a multi-camera or multi-perspective set-up, reducing processing and acquisition time while removing subjectivity in the assessment. We trained our EDFN on the UNIMIB2016 dataset [49] and test it on two novel datasets to reduce bias and enhance generalizability. The two novel datasets are 1) a regular texture foods dataset and, 2) a modified texture foods dataset (e.g., puréed, minced). To the best of our knowledge, this is the first modified texture foods dataset used for segmentation or volume estimation. We conducted analyses including IOU, 2D- and 3D- percent intake error, absolute intake error, and mean intake error bias. We use ground-truth hand segmentation and comparison against an “applied ground truth” through the graph cut semi-automated method. We supplement these analyses with volume disparities to illuminate how segmentation strategies impact accuracy and under what conditions IOU may be insufficient to assess true accuracy. Using this more holistic construct for assessment, we aim to enable trust in the system and document potential circumstances and limitations of the system relevant to the LTC domain for early malnutrition detection via plate-by-plate food consumption tracking.

## **4.1.2 Methods**

### **4.1.2.1 Data Collection**

As we saw in Chapter 3, in the Schlegel Villages, food is prepared in the central industrial kitchen where it is held until it is time to serve from the neighbourhood servery and intake is

assessed by personal support workers whenever they are able to after a resident has finished their meal. To align with this standard of practice, data were collected in an industrial research kitchen which conforms to LTC kitchen standards to enable the most realistic full-portion images collected and to collect data under similar lighting conditions as would be the case in practice. We constructed an image acquisition system that enabled top-down image capture. We imaged 36 foods representative of LTC where 9 were regular texture foods listed as options on a LTC menu comprising our novel “regular texture foods” dataset. A set of 63 modified texture food samples representing 27 unique foods were prepared by a LTC kitchen (The University Gates, Schlegel Villages) and either minced or puréed comprising our “modified texture foods” dataset. During image acquisition, the room temperature varied from 20.6°C to 22.5°C. Before and after imaging via an RGB-D camera (Intel RealSense F200), the scale was tared with a reference plate of the same mass to ensure consistency in portion size (via mass) relative to the matching reference portion. The reference weight for 0% eaten (i.e., full portion) was recorded as part of the weighed-food records for validation and used to calculate each quarter portion eaten. Images were saved to a computer for model training and evaluation. For a summary of the types and representation of foods imaged, see Table 4.1.

***Regular Texture Foods Acquisition:*** We imaged 9 unique food items across three representative meals each consisting of three food items (breakfast: oatmeal, toast, eggs; lunch: pasta, salad, cookie; and dinner: meatloaf, mashed potatoes, corn) were selected from an LTC menu and imaged as part of this data collection series. Each plate was assembled with up to three food items. One full serving of each food item was defined by the nutritional label serving size by mass. Plates were imaged at every permutation of 0%, 25%, 50%, 75%, 100% of each food item consumed. Here, 0% corresponds to the initial, largest mass portion (P1), and 100% corresponds to no amount of that food component remaining (P5). The largest mass portion, P1 was deemed a “full” portion with P2-P5 representing smaller and smaller masses. These 25% incremental bins were selected based on standard dietary intake record forms used in LTC [158, 31]. This yielded 125 unique plates per meal (375 unique plates). Foods were selected to be representative based on a LTC menu.

***Modified Texture Foods Acquisition:*** We imaged 63 food samples (56 unique + 7 duplicates) representing 27 unique food items. Each set of samples for a given unique food contained at least one example of a modified texture (i.e., minced, puréed or both) either imaged fresh, after being held at serving temperature, or both. Holding food at serving temperature is standard practice in LTC serveries as meal items are prepared in advance of meal service. Each food sample was imaged at 5 different portions (P1-P5), with one exception containing 4 portions, by progressively removing some of the sample with a spoon to simulate varying degrees of leftovers. The largest mass portion, P1, was deemed a “full” portion for intake purposes, with



Table 4.1: Imaged foods represented in each of the regular texture foods dataset and the modified texture foods dataset. Regular texture foods were derived from LTC menus and imaged with three foods per meal (breakfast: oatmeal, toast, scrambled eggs; lunch: tortellini, salad, cookie; dinner: meatloaf, potatoes, corn) at every 25% incremental amounts of each food item relative to the full portion. Modified texture foods were imaged separately with five incrementally smaller portions. Recipes were available for all modified texture foods included.

<b>Food Component</b>	<b>Regular Texture Foods</b>	<b>Modified Texture Foods</b>
<b><i>Grains</i></b>	Cheese tortellini /w tomato sauce	Bow tie pasta /w carbonara
	Oatmeal	Macaroni salad
	Whole wheat toast	Vegetable rotini
<b><i>Vegetables and fruits</i></b>	Corn	Asian vegetables
	Mashed potatoes	Baked polenta /w garlic
	Mixed greens salad	California vegetables
		Greek salad
		Mango & pineapple
		Red potato salad
		Sauteed spinach & kale
		Seasoned green peas
		Stewed rhubarb & berries
		Strawberries & bananas
		Sweet and sour cabbage
<b><i>Proteins</i></b>	Meatloaf	Baked basa
	Scrambled egg	Braised beef liver & onions
		Braised lamb shanks
		Hot dog wiener
		Orange ginger chicken
		Salisbury steak & gravy
		Teriyaki meatballs
		Tuna salad
<b><i>Mixed dishes</i></b>	Oatmeal cookie	Barley beef soup
		Blueberry coffee crumble cake
		Eggplant parmigiana
		English trifle
		Lemon chicken orzo soup

P2-P5 representing smaller and smaller masses. This yielded a total of 314 images. Foods were representative of a typical LTC menu as they were prepared by the LTC kitchen and represented a variety of fruits, vegetables, pastas, soup, and meat dishes.

#### 4.1.2.2 Food Image Volume Estimation System with an Encoder Decoder Food Network (EDFN)

The goal was to estimate volumetric food intake from RGB-D images of food on a plate. We developed a deep convolutional neural network (DNN) for generating food segmentation maps, which was refined using depth heuristics and combined with calibrated pixel-wise food heights to estimate food consumption (in mL). Figure 4.1 shows a visual representation of the system diagram.

More specifically, we were inspired by the success of encoder-decoder networks for semantic image segmentation [15]. We designed the macroarchitecture of the proposed food segmentation DNN as a multi-scale encoder-decoder network architecture tailored for downsampled, pixel-level semantic segmentation of food images. Figure 4.1 shows the network architecture, which consists of a residual encoder microarchitecture, a multi-scale hierarchical decoder microarchitecture, and a final high-resolution, per-pixel classification layer for producing a food segmentation map. The residual encoder microarchitecture is responsible for encoding RGB images into a set of feature maps describing the objects in the image. The encoder feature map outputs are then processed through the decoder microarchitecture which parses the scene at multiple spatial scales. These multi-scale representations were concatenated to the feature map outputs, and a  $1 \times 1$  convolutional layer was trained to output a two-class per-pixel segmentation map (food or no-food).

For the residual *encoder* microarchitecture, we leveraged a spliced ResNet101 architecture with pre-activation [92]. The ResNet101 architecture was chosen because of its powerful representational capability for learning discriminative feature representations from complex scenes. We leveraged the notion of transfer learning by beginning with a ResNet101 network architecture designed for classification, trained on the ImageNet dataset of natural scenes [58], and splicing off the deeper ResNet101 layers to create the final encoder microarchitecture. More specifically, we splice at the third unit of the first residual block [248], leading to the proposed residual encoding microarchitecture, which encodes  $120 \times 160$  RGB images into  $256 \ 15 \times 20$  feature maps. As such, the image was fed through a  $7 \times 7$  convolutional layer with 64 kernels and a stride of 2. Then, a  $3 \times 3$  max pool with stride of 2 was performed to downsample the image. These representations were fed through the first ResNet101 block, consisting of 64  $1 \times 1$  convolution, 64  $3 \times 3$  convolution, and 256  $1 \times 1$  convolution layers three times, with skip connections after every set of 3 layers. The last  $3 \times 3$  layer was downsampled using a stride of 2. Thus, the encoder



microarchitecture outputs 256 feature maps at 1/8 the input image size.

The *decoder* microarchitecture of the proposed food segmentation network was designed to decode the feature maps from the encoder microarchitecture into hierarchical global priors using a region binning scene parsing network architecture design. It is well known that multi-scale context aids pixel segmentation [141] which is particularly relevant within the context of food. As humans observing food, there are two main components: the colour and the texture of the food. Texture also varies across scales (i.e., food has a hierarchical visual nature to it). To account for the multi-scale context of food, we leveraged a pyramid scene parsing network (PSPNet) [248] which was connected to the feature outputs from the encoder microarchitecture. As such, the PSPNet decoder microarchitecture performs analysis across four spatial scales, which adds information representing the underlying feature representation and provides local-to-global context of the plate of food. The feature maps were fed into four parallel max-pool layers, with bin sizes of  $1\times 1$ ,  $2\times 2$ ,  $3\times 3$ , and  $6\times 6$ . The upscaled hierarchical global prior outputs were concatenated to the encoder feature maps and two class (food or no-food) pixel-level segmentation was performed using a  $1\times 1$  convolution layer. A circle Hough transform [13] was used to mask the plate from the table, eliminating detection of food outside the plate boundaries (e.g., on tables with complex patterns).

#### 4.1.2.3 Training and Validating the Network

We trained the proposed encoder-decoder food network (EDFN) on the UNIMIB2016 food dataset (1027 tray images, 73 categories), which contains per-pixel ground-truth segmentation [49]. The encoder weights were frozen to conserve deep computational feature extraction from large robust datasets, and only the decoder weights were optimized. The UNIMIB2016 dataset was chosen due to its food variety, overhead view, and pixel-level segmentation annotation. Additionally, since our method was driven by LTC application requirements with data collection in a specific manner, we needed a dataset that was similarly acquired (e.g., pixel-wise annotation, not bounding boxes) so training/fine-tuning could be accomplished without bias by our novel LTC test datasets. While our LTC test datasets comprise a representative sample albeit with relatively few food groupings, training on the UNIMIB2016 dataset with 76 food categories enhanced generalizability. Downsampling was conducted to align the spatial feature sizes of the encoder and decoder microarchitectures with the UNIMIB2016 dataset [49]. The UNIMIB2016 data were resized to match our image height/width which were at the same aspect ratio (4:3). This resizing to  $120\times 160$  images provided two key advantages: (1) computation reduction, (2) better scaled kernels for the image size. We empirically observed that there was not enough global context at the original resolution, resulting in the middle of foods getting misclassified. By downsampling our image, the network was able to identify primary *high*-level features instead of getting stuck in the texture of the food and could be successfully decoded by the pyramid scene parsing de-

coder microarchitecture. The UNIMIB2016 data were randomly split into training and validation subsets (80%/20%). Since all UNIMIB2016 plates were placed on the same tan colored tray, we found that the machine learning model inappropriately learned that the tray colour was always indicative of non-food. Thus, we performed data augmentation on the UNIMIB2016 data by randomly rotating the hue channel of each image’s background (non-food) pixels and adding it to the dataset, thus effectively doubling the training and validation datasets. The network was trained using batch size 32 using RMSProp optimizer with softmax cross-entropy loss, a learning rate of 0.0001 and a decay of 0.995. The network was trained over 200 epochs, and the best model according to the validation loss was kept.

#### 4.1.2.4 Testing the Network

We tested the network on our two custom LTC datasets consisting of 689 (375+314) plates representing 36 (9+27) different foods (as outlined in Table 4.1). Original images were downsampled from 480×640 to 120×160 to decrease the number of network parameters and improve computation time. The images were hand segmented to define ground truth segmentation masks of the food on the plates.

We compared our results to those generated by semi-automatic graph cut segmentation. Since user input is required for initialization, for consistency in the regular texture dataset, one line was used to denote each food item present on the plate and one squiggled background line was indicated around the top and right side of the image as shown in Figure 4.2. The modified texture dataset required substantially and inconsistently more user-defined seeding. The circle Hough transform plate masking used in our proposed system was used here too. The output from this method is a plate-level food segmentation mask.

#### 4.1.2.5 Segmentation Depth-Refinement

The generated food segmentation map from EDFN is based solely on visual information and is thus privy to visual-volume discordance. We therefore developed a heuristic for excluding labeled food areas that are irrelevant to food consumption (e.g., pasta sauce remnants). To do this, co-aligned depth maps were acquired synchronously with the RGB images for the plate under analysis as well as an empty calibration plate.

Ten depth maps were averaged for each acquisition to account for measurement noise. The calibration depth map  $d_C$  was registered to the plate depth map  $d_P$  to account for any changes in camera-plate orientation between calibration and plate acquisitions. Food height was used to refine the segmentation mask (“depth-refinement”) based on *a priori* knowledge that visual-volume discordance is observed when very shallow and inconsequential foods are visually apparent, but

are irrelevant to volumetric analysis. Specifically, the image was decomposed into 250 perceptually meaningful superpixels using simple linear iterative clustering [3] (compactness=20,  $\sigma=2$ ). For each superpixel  $\mathcal{S}_i$ , the constituent pixels were removed from the food map using statistical thresholding on the pixel height distribution:

$$Q_{h_{\mathcal{S}_i}}(p) < \tau \quad (4.1)$$

where  $Q_{h_{\mathcal{S}_i}}$  is the quantile function of the distribution of pixel heights in  $\mathcal{S}_i$ . We set  $p = 0.75$  and  $\tau = 2$  mm based on measurement error along a flat table.

### ***Food Volume Calculation***

Food height was determined in mm units, but to calculate food volume, pixel spacing needed to be calibrated to mm (a pixel-to-mm conversion). Using the known diameter of the plate  $d$  (259 mm), we used the detected plate radius  $\hat{r}$  from the circle Hough transform to compute the conversion:

$$\Delta x = \frac{d}{2\hat{r}} \quad (4.2)$$

Food volume in mL could then be computed by summing the per-pixel differential volumes within the food mask:

$$V = \sum_{i \in F} (\Delta x)^2 h_i \quad (4.3)$$

where  $F$  is the set of segmented food pixels. Volumetric food intake was computed by subtracting the plate volume from the full portion volume. Similarly, percent intake was calculated relative to the full portion volume.

#### **4.1.2.6 Data Analysis**

To compare quantitative performance between methods, we use the common performance measures of global accuracy (Equation 4.4) to describe the percentage of correctly classified pixels, food segmentation accuracy (Equation 4.5) to describe the percentage of correctly classified food pixels), as well as the intersection over union (IOU) (Equation 4.6) both within a meal (i.e., breakfast, lunch, dinner, modified texture single-imaged foods) and across meals. For this application, the IOU provides a more representative metric for how our segmentation system is performing as it captures accuracy within the context of the true bounded food areas since false positive predictions are penalized. The theoretical maximum value of IOU is 1.0 when the intersection maps perfectly over the union without deviation. We define the metrics described above as follows:

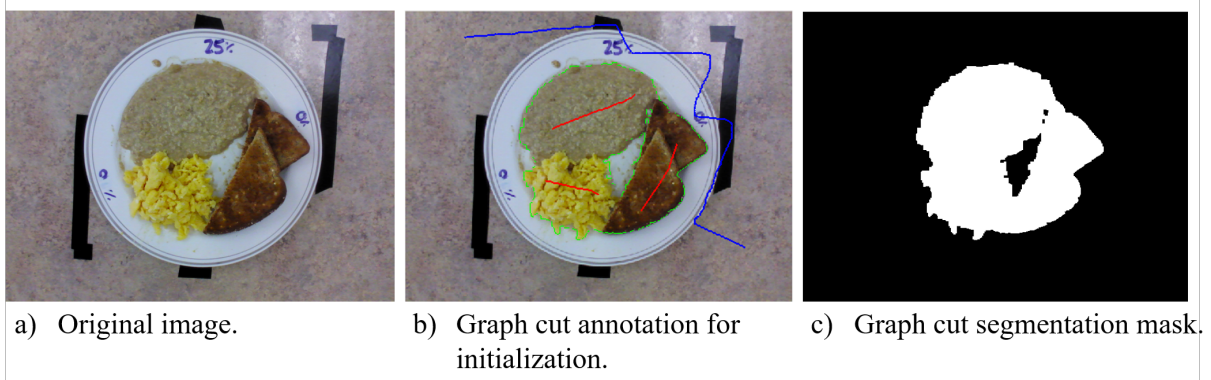


Figure 4.2: Sample graph cut annotation with one line per food item (red) and one background line (blue) and resulting segmentation mask.

$$\text{Global Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.4)$$

$$\text{Food Segmentation Accuracy} = \frac{TP}{TP + FN} \quad (4.5)$$

$$IOU = \frac{\text{target} \cap \text{prediction}}{\text{target} \cup \text{prediction}} = \frac{TP}{TP + FP + FN} \quad (4.6)$$

Volume estimation accuracy was assessed by computing mean absolute intake error (mL) as volume calculated using our proposed method or the “applied ground-truth” method with or without depth-refinement relative to the volume across the ground-truth hand segmented areas. Error (mL) was calculated similarly but preserves the direction of error. Volume intake error (mL) is the difference between the current portion relative to the full portion. Intake error was calculated for both segmentation (2D) and volume (3D) data relative to the full portion. All values are reported as mean  $\pm$  SD.

### 4.1.3 Results

When solely considering intake using segmentation, the form factor of food inherently assumes depth is uniform across the segmented portion. However, this approach fails to fully capture the context of food; we refer to this as the “**visual-volume discordance**”. For example, consider one tablespoon (15 mL) of tomato sauce, this sauce could be piled relatively high into a mound (i.e.,

representing relatively few plate pixels), or could be very thinly spread across the majority of the plate (i.e., representing many plate pixels). From a computer vision approach to segmentation, these two plates would yield extremely different segmented areas while the absolute volume of the sauce would be the same. A human assessor can note there is little sauce on the plate in either configuration. Depth-refinement, either as part of the segmentation pipeline or conducted through relative changes in volume for food volume intake assessment, circumvents this issue by providing context beyond the pixel count of a segment and brings assessment closer (but with higher precision and accuracy) to a human assessor. Now consider what is deemed “ground truth” from hand segmentation of an image. Here, human assessors indicate *where* on the plate there was food to generate the ground truth. However, considering segmentation accuracy in isolation misses important context about *how much* food is present, which is the more pertinent question for assessing food intake. As such, we cannot rely fully on classical segmentation accuracy for evaluating system performance since there can be a strong discordance between visual (RGB) and volume (RGB-D) assessments. Volume consideration is particularly essential for estimating food intake when accounting for the high prevalence of modified texture foods in LTC. While metrics pertaining to visual accuracy are most synonymous with traditional assessments of segmentation accuracy (e.g., IOU), metrics pertaining to volume accuracy may be more representative of true intake, especially for modified texture foods which have higher fluidity. Figure 4.3 provides a visual analyses of the results taking into account these considerations while Table 4.2 provides a numerical summary of results. Data are reported as (mean  $\pm$  SD).

#### 4.1.3.1 Regular Texture Foods

##### *Food Segmentation Accuracy*

Regarding the regular texture foods dataset, food segmentation accuracy was high for our proposed system without and with depth-refinement (EDFN:  $0.943 \pm 0.047$ , EDFN-D:  $0.918 \pm 0.048$ ). Our proposed, depth-refined system was comparable to the no-depth and depth-refined graph cut implementations (GC:  $0.955 \pm 0.055$ , GC-D:  $0.933 \pm 0.060$ ). Variance was similar but slightly lower for our proposed EDFN and EDFN-D than GC and GC-D.

##### *Segmentation Agreement*

Segmentation agreement (IOU) was also good, and further improved through depth-refinement for our proposed system (EDFN:  $0.885 \pm 0.069$ , EDFN-D:  $0.927 \pm 0.029$ ). The graph cut IOU both without and with depth-refinement outperformed our proposed methods owing to user-specified seed points (GC:  $0.938 \pm 0.036$ , GC-D:  $0.941 \pm 0.029$ ). We attribute the improvements of both our proposed system and our applied ground truth with depth-refinement to reduction in the visual-volume discordance due to low-profile yet visually distinct foods on a plate, such as pasta sauce.



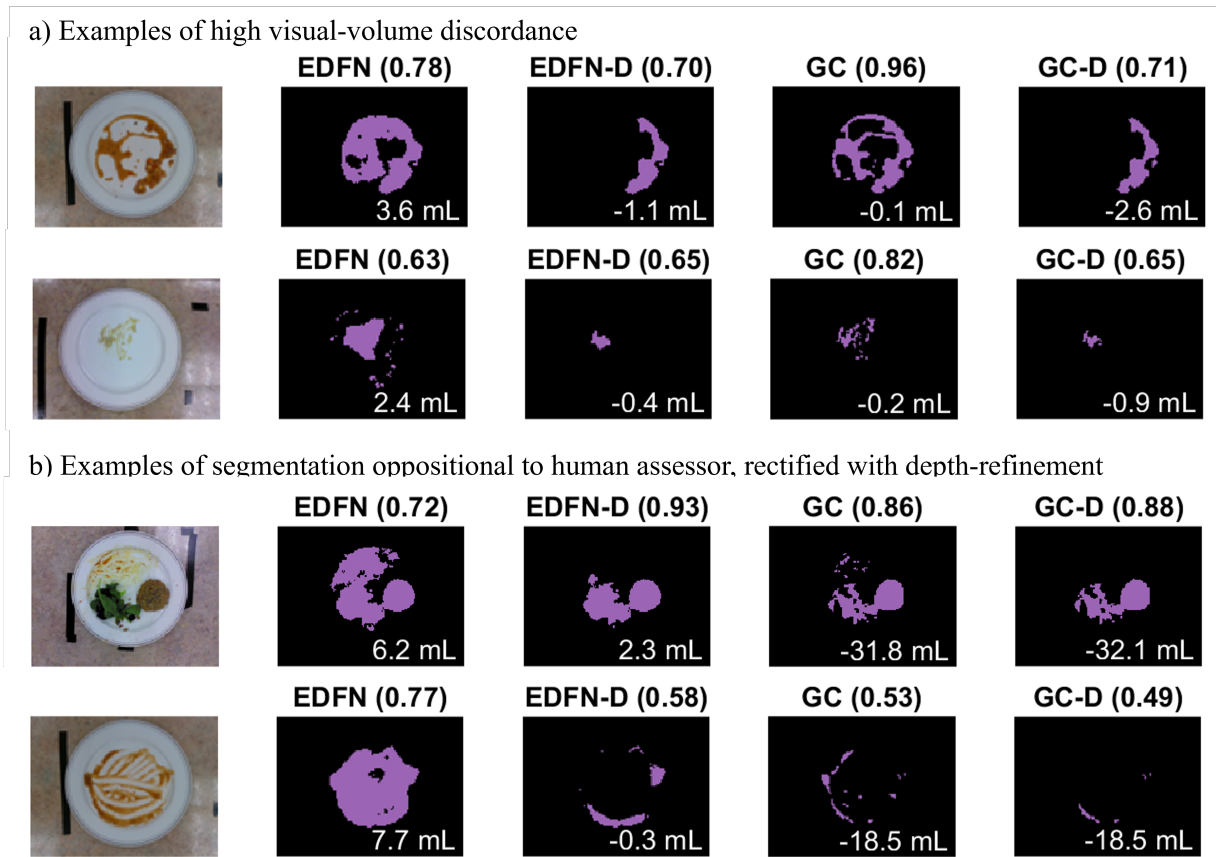


Figure 4.3: Visual comparison of proposed method (EDFN) and our “applied ground truth” (GC) both without and with depth-refinement (EDFN-D, GC-D). Examples span both regular and modified textures and illustrate a) plates with high visual-volume discordance (i.e., many pixels contain food but represent little volume), as well as b) examples where depth-refinement rectified oppositional segmentation compared to human assessor (i.e., a special case of EDFN examples of visual-volume discordance human assessors would segment differently and EDFN-D compensates for these differences). IOU is shown in black bold at the top of the frame while volume error (mL) is indicated in white text at the bottom of each frame; negative volume error implies under-segmentation (i.e., over-estimation of intake).

Table 4.2: Comparative analyses of system performance within and across LTC datasets between our proposed method (EDFN, EDFN-D) and the “applied ground truth” graph cut (GC, GC-D). % error intake refers to the portion of segmented pixels calculated using the predicted estimate minus the target ground-truth hand segmented regions; 2D: no-depth, 3D: with depth-refinement.

<i>Dataset Method</i>	Segmentation accuracy			Intake accuracy			Volume estimation accuracy (mL)		
	GSA	FSA	IOU	2D % intake error	3D % intake error	MAE	MEB	VIE	
<i>Regular texture foods</i>									
<i>EDFN</i>	0.973 (0.018)	0.943 (0.047)	0.885 (0.069)	-23.4 (21.2)	-9.1 (8.8)	17.1 (49.2)	-14.7 (50.0)	-129.2 (154.3)	
<i>EDFN-D</i>	0.984 (0.012)	0.918 (0.048)	0.927 (0.029)	-16.5 (17.6)	-9.0 (8.9)	18.0 (50.0)	-17.2 (50.3)	-130.2 (154.8)	
<i>GC</i>	0.987 (0.009)	0.955 (0.055)	0.938 (0.036)	-6.8 (13.5)	0.4 (1.3)	4.5 (5.5)	-0.0 (7.1)	1.8 (6.6)	
<i>GC-D</i>	0.988 (0.005)	0.933 (0.060)	0.941 (0.029)	-5.3 (12.7)	0.4 (1.3)	4.6 (5.4)	-1.8 (6.9)	0.2 (6.5)	
<i>Modified texture foods</i>									
<i>EDFN</i>	0.990 (0.011)	0.922 (0.165)	0.846 (0.114)	-44.0 (38.6)	-0.8 (5.2)	2.8 (3.1)	1.7 (3.8)	0.3 (3.6)	
<i>EDFN-D</i>	0.991 (0.014)	0.697 (0.348)	0.819 (0.166)	-16.5 (20.0)	1.4 (5.9)	2.3 (3.2)	-0.7 (3.9)	0.8 (3.6)	
<i>GC</i>	0.995 (0.006)	0.834 (0.157)	0.898 (0.082)	-27.4 (25.5)	0.7 (3.1)	2.2 (2.7)	-1.9 (2.9)	-0.9 (3.3)	
<i>GC-D</i>	0.991 (0.013)	0.656 (0.328)	0.819 (0.165)	-14.5 (18.1)	2.1 (4.6)	3.2 (3.4)	-3.1 (3.5)	-0.5 (3.4)	
<i>All foods</i>									
<i>EDFN</i>	0.981 (0.017)	0.934 (0.116)	0.867 (0.094)	-32.8 (32.0)	-5.3 (8.5)	10.8 (37.4)	-7.4 (38.2)	-71.4 (131.7)	
<i>EDFN-D</i>	0.987 (0.013)	0.819 (0.259)	0.879 (0.125)	-16.5 (18.7)	-4.2 (9.2)	11.0 (38.1)	-9.9 (38.4)	-71.8 (132.3)	
<i>GC</i>	0.991 (0.008)	0.901 (0.128)	0.920 (0.064)	-16.2 (22.3)	0.5 (2.3)	3.5 (4.6)	-0.8 (5.7)	0.6 (5.6)	
<i>GC-D</i>	0.990 (0.010)	0.809 (0.262)	0.887 (0.127)	-9.5 (16.1)	1.1 (3.3)	4.0 (4.7)	-2.4 (5.7)	-0.1 (5.3)	

Values are (mean  $\pm$  SD) GSA: Global segmentation accuracy, FSA: Food segmentation accuracy, IOU: intersection over union, MAE: Mean absolute error, MEB: Mean error bias, VIE: Volume intake error.

Table 4.3: Summary of volume estimation accuracy across portion sizes for the modified texture foods dataset across our proposed EDFN and EDFN-D.

<i>Dataset Portion</i>	Segmentation accuracy			Intake accuracy			Volume estimation accuracy (mL)		
	GSA	FSA	IOU	2D % intake error	3D % intake error	MAE	MEB	VIE	
<i>EDFN (no-depth-refinement)</i>									
<i>P1</i>	0.996 (0.003)	0.970 (0.066)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	3.3 (3.4)	2.0 (4.3)	0.0 (0.0)	
<i>P2</i>	0.994 (0.006)	0.965 (0.089)	-30.2 (17.2)	-0.8 (8.0)	-0.8 (8.0)	3.0 (3.1)	2.1 (3.8)	-0.1 (3.3)	
<i>P3</i>	0.992 (0.009)	0.941 (0.125)	-50.1 (23.5)	-0.9 (5.3)	-0.9 (5.3)	2.9 (2.9)	1.9 (3.6)	0.1 (3.7)	
<i>P4</i>	0.987 (0.013)	0.912 (0.176)	-68.0 (35.3)	-1.1 (5.6)	-1.1 (5.6)	3.0 (3.2)	1.4 (4.2)	0.6 (4.5)	
<i>P5</i>	0.982 (0.014)	0.821 (0.254)	-71.9 (43.4)	-1.0 (3.8)	-1.0 (3.8)	2.0 (2.6)	1.1 (3.1)	0.9 (4.6)	
<i>EDFN-D (depth-refined)</i>									
<i>P1</i>	0.996 (0.004)	0.945 (0.074)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	2.6 (3.3)	0.1 (4.3)	0.0 (0.0)	
<i>P2</i>	0.994 (0.010)	0.908 (0.132)	-22.8 (17.5)	0.3 (8.7)	0.3 (8.7)	1.9 (3.0)	0.0 (3.6)	0.1 (3.3)	
<i>P3</i>	0.992 (0.013)	0.824 (0.162)	-30.6 (17.5)	1.4 (6.3)	1.4 (6.3)	2.3 (3.5)	-0.5 (4.1)	0.7 (4.2)	
<i>P4</i>	0.987 (0.018)	0.640 (0.275)	-25.2 (20.1)	2.8 (6.6)	2.8 (6.6)	2.9 (3.8)	-1.7 (4.5)	1.8 (4.4)	
<i>P5</i>	0.986 (0.016)	0.157 (0.262)	-3.9 (16.2)	2.3 (3.5)	2.3 (3.5)	1.7 (2.2)	-1.5 (2.4)	1.6 (4.0)	

Values are (mean  $\pm$  SD) GSA: Global segmentation accuracy, FSA: Food segmentation accuracy, IOU: intersection over union, MAE: Mean absolute error, MEB: Mean error bias, VIE: Volume intake error.

**Percent Intake Error (2D and 3D)** Regarding 2D percent intake error (i.e., using segmentation alone), our non-depth-refined and depth-refined proposed systems were outperformed by the depth-refined graph cut implementation (EDFN<sub>2D</sub>:  $-23.4\% \pm 21.2$ , EDFN-D<sub>2D</sub>:  $-16.5\% \pm 17.6$ , GC<sub>2D</sub>:  $-6.8\% \pm 13.5$ , GC-D<sub>2D</sub>:  $-5.3\% \pm 12.7$ ). These negative values, which were improved with depth-refinement, imply a bias towards under-segmenting an image. Refer to the Discussion for clinical implications of this bias towards under-segmentation. Regarding the 3D percent intake error, the graph cut implementations outperformed our proposed system (EDFN<sub>3D</sub>:  $-9.1\% \pm 8.8$ , EDFN-D<sub>3D</sub>:  $-9.0\% \pm 8.9$ , GC<sub>3D</sub>:  $0.4\% \pm 1.3$ , GC-D<sub>3D</sub>:  $0.4\% \pm 1.3$ ).

**Volume Estimation Accuracy** Regarding the volume error (mL) on the regular texture dataset, while initially it appeared depth-refinement worsened performance across methods (EDFN:  $-14.7 \text{ mL} \pm 50.0$ , EDFN-D:  $-17.2 \text{ mL} \pm 50.3$ , GC:  $0.0 \text{ mL} \pm 7.1$ , GC-D:  $-1.8 \text{ mL} \pm 6.9$ ), the plate-level absolute volume error was improved with depth refinement and variance was much smaller (e.g., EDFN-D volume intake error  $-130.2 \text{ mL} \pm 154.8$ ; EDFN-D mean absolute error  $18.0 \text{ mL} \pm 50.0$ ). However, intake error from volume was high and both our proposed system and graph cut implementations (EDFN:  $-129.2 \text{ mL} \pm 154.3$ , EDFN-D:  $-130.2 \text{ mL} \pm 154.8$ , GC:  $1.8 \text{ mL} \pm 6.6$ , GC-D:  $0.2 \text{ mL} \pm 6.5$ ). Corroborated by the negative mean error bias for EDFN and EDFN-D, we empirically attribute this wide variance and high intake error paired with low plate-level absolute volume error to salad. As intake error is calculated with a plate relative to the full portion, variability in a food’s appearance could be high leading to high intake error, with accurate absolute volume. Salad, as part of the lunch plates had low-density with widely varying degrees of air pockets between leaves of lettuce at each plating. Depending on how “fluffy” the salad was put into position, it could take on differing volumes. This is discussed in detail later.

#### 4.1.3.2 Modified Texture Foods

##### **Food Segmentation Accuracy**

Regarding the modified texture foods dataset, food segmentation accuracy was high for our proposed system, however depth-refinement reduced the accuracy based on classical segmentation accuracy calculations (EDFN:  $0.922 \pm 0.165$ , EDFN-D:  $0.697 \pm 0.348$ ). In both cases however, our proposed system outperformed the graph cut analogs (GC:  $0.834 \pm 0.157$ , GC-D:  $0.656 \pm 0.328$ ). Depth-refinement brings the context of volume, whereas the initial ground-truth hand segmentation used for comparison was based solely on visual appearance of food on the plate. Modified texture foods by nature tend to spread out more due to increased fluidity and as such, are at greater risk for the visual-volume discordance.

##### **Segmentation Agreement (IOU)**

Compared to the regular texture food dataset, segmentation agreement (IOU) was adequate (over 0.80) and relatively unaffected by depth-refinement. Our proposed system was slightly outper-

formed by the graph cut implementation on IOU, with the graph cut variance smaller in non-depth-refined segmentation (EDFN:  $0.846 \pm 0.114$ , GC:  $0.898 \pm 0.082$ ). Our proposed method was comparable to graph cut for the depth-refined counterparts (EDFN-D:  $0.819 \pm 0.166$ , GC-D:  $0.819 \pm 0.165$ ).

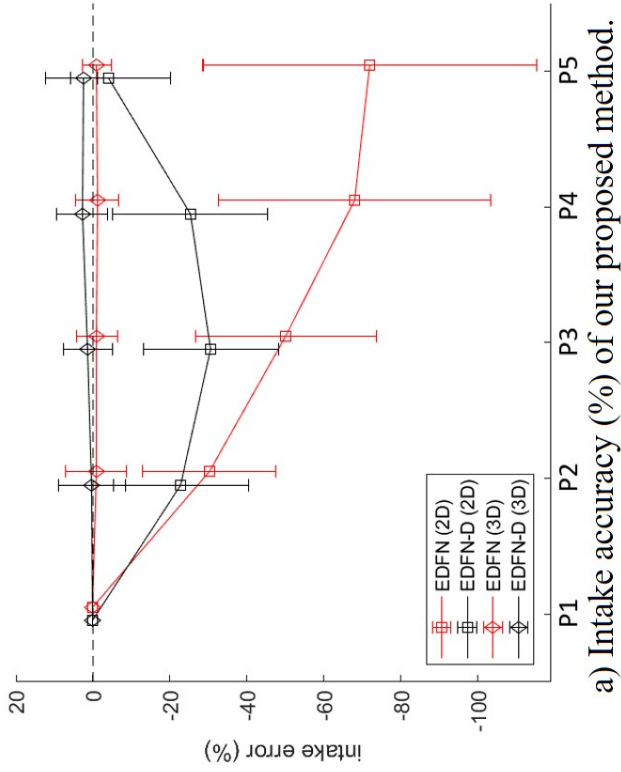
### ***Percent Intake Error (2D and 3D)***

Regarding the 2D percent error of intake, non-depth-refined implementations had unacceptably high percent error of intakes (EDFN<sub>2D</sub>:  $-44.0\% \pm 38.6$ , GC<sub>2D</sub>:  $-27.4\% \pm 25.5$ ) with unacceptably wide variances still present in depth-refined implementations (EDFN-D<sub>2D</sub>:  $-16.5\% \pm 20.0$ , GC-D<sub>2D</sub>:  $-14.5\% \pm 18.1$ ). This provides additional evidence to support the need to rectify the visual-volume discordance, especially in modified texture foods where visually salient food remnants are more likely to remain on the plate after consumption. For 3D percent error intake, depth-refinement had a minimal impact on percent intake error for all implementations which already yielded very low error in estimating volume intake error (EDFN<sub>3D</sub>:  $-0.8\% \pm 5.2$ , EDFN-D<sub>3D</sub>:  $1.4\% \pm 5.9$ , GC<sub>3D</sub>:  $0.7\% \pm 3.1$ , GC-D<sub>3D</sub>:  $2.1\% \pm 4.6$ ). Again, negative intake implications are addressed in the discussion.

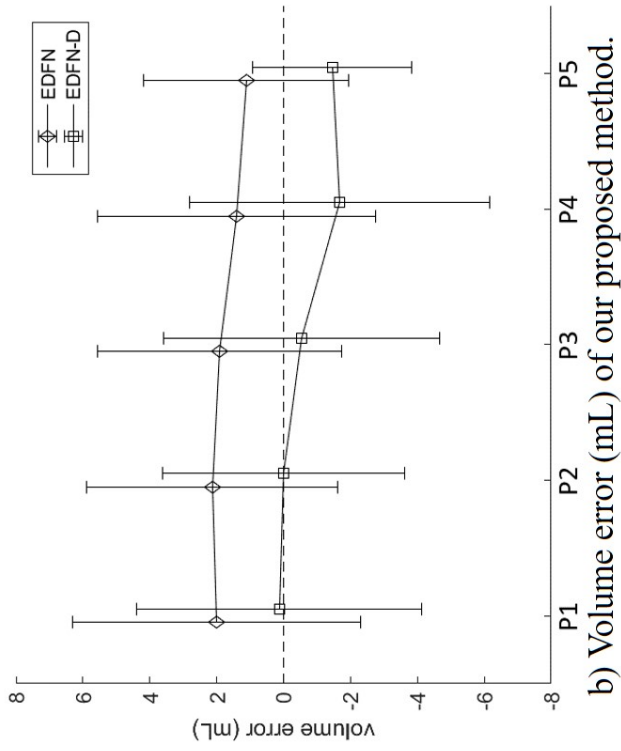
Figure 4.4(a) depicts these visual-volume discordance errors and illustrates how depth-refinement (black lines) both reduced the variance of the percent intake errors as well as reduced the relative intake error compared to the non-depth-refined counterparts (red) and using P1 as the reference “full-portion”. Across methods, error tended to increase as the remaining portion size diminished except for EDFN-D (2D) with error peaking at the third portion and receding across portions P4 and P5. We attribute this trend to depth-refinement compensating for higher degrees of visual-volume discordance on plates more likely to have smearing (i.e., plates with less on the plate relative to the initial portion size). This highlights why visual representation context of *where* food resides is inadequate and how additional depth context pertinent to *how much* food is present is required particularly for modified texture foods.

### ***Volume Estimation Accuracy Across Portion Sizes***

Typically, volume errors, as well as volume intake errors, were low (less than 4.0 mL) across EDFN and GC implementations as shown in Table 4.3. To supplement this using our proposed EDFN and EDFN-D, Figure 4.4(b) shows volume accuracy across each of the 5 portions (P1-P5) relative to the volume across the ground-truth hand segmented food regions. While our EDFN-D implementation was more accurate, it had similar precision to EDFN. A similar trend in reduced error in mL across smaller portions was observed for both EDFN and EDFN-D albeit with EDFN-D error shifted downwards. With EDFN-D, the three smallest portions (P3, P4, P5) had negative values indicating the depth-refinement omitted depth values across additional pixels relative to the initial segmentation and yielded a slight under-segmentation of food.



a) Intake accuracy (%) of our proposed method.



b) Volume error (mL) of our proposed method.

Figure 4.4: Accuracy of our proposed method without (EDFN) and with (EDFN-D) depth-refinement on the modified texture foods dataset. Here, P1 is the reference plate with P2-P5 representing simulated degrees of increasing intake (i.e., P5 most eaten). (a) 2D refers to two-dimensional % intake error in terms of the proportion of pixels estimated compared to the ground-truth hand segmentation of food areas; 3D refers to three-dimensional % intake error in terms of the relative volume across the estimated food segment compared to the volume across the ground-truth segmentation of food areas. (b) Proposed method is compared against volume across the ground-truth hand segmented food area.

#### 4.1.4 Discussion

Our proposed fully automatic system imparts a reduced processing burden compared to semi-automatic segmentation. Using graph cut as our “applied ground-truth”, task completion time represents an additional point for consideration, and our automated approach provides a key advantage in the food and nutrition tracking context. Empirically, user-defined seed initialization for graph cut implementation incurred approximately 5 seconds of manual annotation time per image. Assuming 192 residents across 6 neighbourhoods (units) in LTC, this implies 48 additional minutes during a meal-service simply to annotate the images. The average time for charting residents’ food intake for a day is already at least 270 minutes [180], which implies that annotation could impart an 18% time increase to complete food intake charting. This approach is infeasible and prohibitive within this context. Instead, compared to the graph cut method which tended to under-segment food, our proposed automatic segmentation method requires minimal additional time commitment from the user enhancing its potential for uptake in the field.

Segmentation errors (under- or over-segmentation) have clinical significance for identification of inadequate intake before low intake progresses to eventual malnutrition. Estimated intake could be incorrect due to either under-estimating food consumption (i.e., less food reported than actually consumed) or over-estimating food consumption (i.e., more food reported than actually consumed). Over-segmenting food areas (i.e., under-estimating intake) implies the predicted food area is over-estimated and translates to reporting that there is more food present (less food intake) than what is true. Conversely, under-segmenting food areas (i.e., over-estimating intake) implies the predicted food area is under-estimated and translates to reporting there is less food present (more food intake) than there is. Over-segmentation (under-estimating food consumption) may be the lesser of two evils for the majority of LTC residents; if food intake is better than what is reported, residents at-risk for malnutrition may be less likely to be missed. Clinically, residents receiving modified texture foods consume significantly fewer calories, have higher cognitive impairment, and require more assistance with activities of daily living than residents on regular texture diets [123] making these even higher-risk residents. However, under-segmentation (over-estimating food consumption), increases risk of missing residents with nutrient inadequacies through introducing false negatives and potentially missing residents with poor food intake that could result in malnutrition. While under-reporting may lead to increased referrals for malnutrition screening, most at-risk residents who eat very little would still be identified which might help to identify residents who could benefit from a dietary intervention.

It appears segmentation alone is inadequate and there is a need for context from volume. We observed that under-segmentation was an issue mostly with 2D % volume intake across implementations and datasets. We evaluated 2D % volume intake for consistency with existing methods, however, the system’s 3D intake estimation is preferred whenever possible. Addition-

ally, returning to Figure 4.4, this plot showcases the need for methods beyond segmentation for food intake tracking. While we observed a decrease in 2D % intake error with depth-refinement across the remaining plate portions, errors for P2, P3, and P4 were more than 20%. A system based on segmentation methods alone may be accurate when very little food is consumed or nearly all the food is consumed. Certainly, there is value in improved accuracy for these edge-cases [177]. However, were we to rely on a system without-depth refinement, this degree of error may be deemed inappropriately high and be a barrier to uptake. Depth-refined volume estimations yielded additional improvement in 3D percent error intake and may provide a palatable alternative within an acceptable error margin of less than 10% (where current practice is up to 62% error [39]; 50% for portion size [23]) with the added benefit of more fine-grained assessment (continuous measurement vs. 25% incremental bins). While we also observed negative values for volume intake error for our proposed system, EDFN-D, this error was largely due to salad. It appeared GC seemed to understand the visual representation of salad better than our proposed method. We suspect this could be improved with additional instances of salad included in the training and validation set, which we include in Section 4.2. Green vegetables were generally under-represented in UNIMIB2016 and provides an opportunity area for additional comprehensive food intake tracking databases. Whereas the volume intake error for regular texture foods was on the order of 130 mL, the modified texture foods circumvented the low-density foods problem with volume intake error of 0.8 mL. Arguably, accuracy on these modified texture foods is more clinically relevant for at-risk residents. Additionally, building in some depth redundancy may better approximate a human assessor and improve initial acceptability.

Perhaps more pertinent to uptake of the system and recurring use of the system is *how* the computer “sees” food on a plate and how this compares to the human experience for supporting trust in the automated system. For example, from the human assessor perspective, a plate containing sauce remnants would be ignored and treated as completely consumed food while the computer vision approach would observe each of the pixels containing sauce remnants and mark it as still containing food as in the case of Figure 4.3. This may lead to distrust in the system because the system makes decisions differently and in opposition to how a human assessor would classify the presence of food. Borrowing from clinicians’ perceptions of artificial intelligence tools, the alignment between the system output and what would be expected from a human’s interpretation is essential for continued use [225]. By incorporating depth-refinement, even though the IOU for modified texture foods decreased, it brings the assessment closer to how a human would interpret a plate with the added benefits of greater precision and objectivity. This area has been largely unexplored because, with one exception (UNIMIB2015 [49]), available food datasets only include full plates as shown in Table 2.1. For measuring food intake, additional application appropriate metrics such as intake error and volume estimation accuracy must be considered since assessing system performance solely from a segmentation accuracy perspective can be misleading.

Albeit challenges with low-density foods, our system attained substantially higher accuracy than current LTC methods. As aforementioned, current LTC home food intake accuracy shows correct estimation of intake occurring as low as 38% of the time [39] and when portion size is mis-estimated, has error up to 50% [23]. Part of the issue, in practice, may be that the granularity of these estimates is also wide since estimates are recorded as 25% incremental food intake bins [158, 30]. This may introduce further subjectivity between assessors. Hypothetically, one human assessor may estimate a plate to be 30% eaten so a value of 25% would be recorded; another assessor may estimate intake at 45% reflecting a record of 50% consumed. With depth-refinement, our proposed EDFN-D removes this subjectivity, operates on a continuous scale, and has a mean 3D % intake of estimation error of -4.2% across both datasets and a mean volume intake error of 0.8 mL on the modified texture foods dataset.



## 4.2 Inferring Nutritional Intake

While initial results on three regular texture foods demonstrated a high correlation and strong agreement between simulated intake from ground truth hand segmentation and weighed food ( $r \leq 0.97$ ,  $\sigma < 20$ ), it assumed each food item was hand segmented and hand-labelled (not published). Given LTC's extreme time pressures, to minimise potential barriers to uptake a level of automated segmentation (previous section) and classification is required. However, the challenge with even the most successful of generic classification methods is they fail to reach 100% accuracy

which may limit their utility in practice especially if classification does not include assessment of food (or nutrient) intake accuracy. Within nutrition tracking in long-term care, this can have very real consequences. If mayonnaise gets misclassified as low-fat greek yogurt for example, the nutritional composition recorded may be drastically affected and limit trustworthiness of the system. When considering the necessity for accurate nutrition tracking, we therefore need a solution that performs reasonably well and acknowledge some degree of human assessor intervention may be required if error-margins are inappropriately high. Human-in-the-loop may be essential, but we can limit the extent to which it is needed by:

- Incorporating *a priori* information about when to expect which foods (e.g., LTC menu).
- Leveraging prescribed diet for further refinement.

For example, Mrs. Brown takes her chicken minced. Let's use today's minced lunch classifier.

### Section Summary

This section concludes the AFINI-T system saga. It describes how I continued applying computer vision, machine learning, and prior knowledge of menus for identifying *which foods* are present (i.e., classification) and finally how the accuracy of simulated *nutrient-level intake estimates* using our EDFN-D segmentation method compared against the ground truth weighed food records.

#### Key take-aways:

- Mean classification accuracy by meal: 88.9% .
- Mean absolute error food volume:  $3.8 \pm 8.8$  mL.
- Mean bulk 3D % absolute intake error:  $9.9 \pm 9.7$ .
- For 13 nutrients of interest intake estimation relative to weighed food records:
  - High correlation:  $r^2 = 0.96 \pm 0.02$ .
  - High similarity (low bias):  $\sigma_{max} = -2.7$  kcal.
- Future work should address low-density foods (e.g., salad, stacked toast) which account for a high degree of variance.

This work builds on the first section of this chapter of leveraging a specialized food-segmentation method powered by deep learning for automated segmentation and moves from bulk food segmentation to nutritional estimation with a few additional steps which we modularized for systematic assessment of error through our system. Here, we focus on the characterization of changes in volume on a whole plate level for bulk intake estimation reporting degree of consumption (i.e., proportion of food consumed) as well as simulated nutritional intake estimation using a nutritional look-up table at the food-item and whole plate level. “Simulated” because the datasets collected were mock-ups of plates of leftovers not true leftover plates. When considering the application realities and needs surrounding use in LTC and specifically how classification might be applied, an additional point for consideration is limited computational resources. To meet the needs of this environment, we require:

- (a) A salient feature extractor that can be trained in advance and supports real-time use.
- (b) A classification method that is light-weight for mobile application use.
- (c) A daily, easily updatable classifier to account for *a priori* menu plans.

With these needs steering development, **the purpose** of this study is to explore nutritional composition estimation using an integrated RGB-D camera for image acquisition and to quantify the error of computing nutritional intake from relative changes in volume as compared to ground truth nutritional intake from gold standard weighed-food records. This system measures food intake, addresses automatic segmentation with integrated RGB-D assessments, and considers both regular and modified texture foods.

## 4.2.1 Methods

### 4.2.1.1 Climate and Context: Motivation from a Case Study

Insights motivating our technical approach here were gathered through interviews and workshop discussion from previously unreported results from the user study outlined in Chapters 3, and 5. Participants identified potential barriers to uptake including time, and whether the level of detail is desired or seen as valuable. Additional insights are provided from two interviews (a registered dietitian (RD) nutrition research expert, an RD working in LTC), and discussion with experts during a workshop. The workshop included 21 participants representing 12 LTC and retirement homes who were recruited through self-enrollment (for a list of roles, refer to Section 3.2.1).

#### *Focus areas*

Regarding the interviews, ultimately it was clear that the dietitian needs to be the gate-keeper to ensure any changes to nutritional management is holistically assessed and surface-level issues are not being masked as may be the case if dietary aides were to be automatically directed to

suggest or encourage eating alternative foods. More specifically, the interview with the LTC RD conveyed that, in Ontario, dietitians get one half-hour block per resident each month. This isn't much time. The ability to streamline the workload by automating manual tasks (e.g., manual calculations of individualised targets) may help to offset the time required and allow the RD to assess resident nutritional status more deeply. This RD saw the most potential for this technology to help make more appropriate referrals and to decrease the data overload of multiple, unranked referrals. They indicated that more accurate intake tracking would help especially considering there is no current discrimination between which foods were consumed. For example, if 25% of the main dish was consumed, there is no way to know if it was the protein or the vegetable; it assumes an even distribution across the foods offered on the plate. For this reason, our system must provide more precise information pertaining to the types of foods consumed, and track at a lower level (nutrient level) so key nutrients with clinical endpoints can be better monitored. This RD identified the following clinical endpoints: falls (calcium, vitamin D, protein), new or worsening pressure ulcers (protein), constipation (fibre, fluids), and weight loss (calories, protein). Similarly, zinc, vitamin K, and vitamin B6 are often poorly consumed in this population with clinical relevance pertaining to cognition, wound healing and immunity [103, 197, 123].

We took the union of nutrients of interest identified here along with the nutrients of interest identified through the design process (calories, carbohydrates, protein, fats, calcium, iron, sodium and vitamin C). This resulted in a final set of 13 nutrients of interest. **Macronutrients:** calories, carbohydrates, fats, fibre, and protein. **Micronutrients:** calcium, iron, sodium, vitamin B6, vitamin C, vitamin D, vitamin K, and zinc.

The interview with the RD nutrition research expert indicated that currently food intake is tracked by team members walking around the dining room implying any sort of food imaging will impart an increased time. Furthermore, in LTC the emphasis is more on quality of life and honouring resident preferences, additionally corroborated during the quarterly registered dietitian meeting. Typically, even menu planning is conducted at the macronutrient level, so this type of system has potential to meet resistance regardless of the analytical power without some form of incentivisation. Based on workshop discussion including perspectives of dietary aides, recreation staff, and directors and assistant directors of food services, they expressed the desire to more precisely record food and fluid intake to inform individual interventions and menu-planning more broadly. It seems there is a strong desire to provide enhanced care and make better use of tracking information than what is mandated, but that barrier to do so is the inaccuracy of the current system. This suggests, knowing more “useful” data could be collected may provide incentives inherently.

### ***Real-time use***

Regarding practicalities of real-time use, insight from the workshop motivated this work. One main emerging theme was the very clear need for the system needs to work even when internet

is down. They need to log into the system to track. If it was down, then currently they could not log anything forcing retrospective charting in these situations. There was much enthusiasm for compatibility offline and would prefer the option to save offline with the ability to sync when able both for when the internet is down or to support dining field trips to outside restaurants. For more than just privacy concerns and considerations, this further corroborates the need for an on-board system. This applies both to the computational requirements of the system as well as how classification models are updated.

### ***Leveraging a priori information***

Regarding the use of *a priori* menu-plan information, in this way, we are able to simplify the classification problem by applying real-world constraints. At a given time of day, certain foods are more likely to be served. To start, we can limit to the set of menu item options. One foreseen challenge with the modified texture foods is discrimination between the same food at different levels of modification. While they are the same foods, they are prepared differently (e.g., broth is added to purées) making it important to properly discriminate between sub-categories of the same food. Here, we make use of the additional prior of which level of texture is prescribed to each resident. As we know the AFINI-T system needs to be linked in with the current electronic health record software (e.g., through PointClickCare), access to this information would be feasible.

### **Informed Design Requirements**

***Nutrients of concern:*** 13 nutrients of concern were identified based on clinical endpoints and expert user feedback. **Macronutrients:** calories, carbohydrates, fats, fibre, and protein. **Micronutrients:** calcium, iron, sodium, vitamin B6, vitamin C, vitamin D, vitamin K, and zinc.

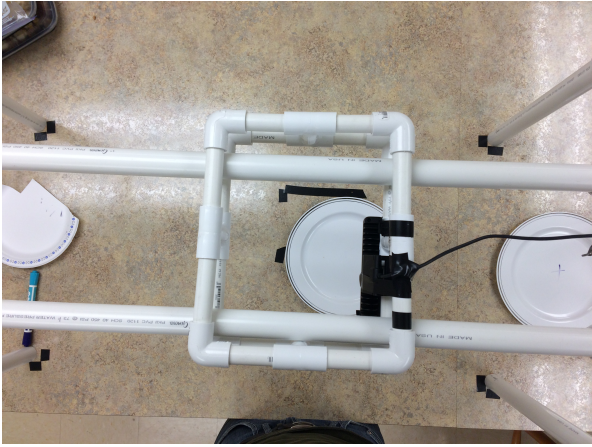
***Dietitian's role:*** The system must record accurate information which can then be leveraged by the dietitian to prioritize referrals and more easily manage nutritional interventions as well as the complex considerations for what level of information to be provided to which level of care (e.g., dietary aide versus dietitian). The dietitian must be the gatekeeper.

***Real-time use:*** An on-board system without constant internet access dependency is needed and informs how classification models are trained and updated.

***Simplifying classification:*** We leverage meal time and prescribed diet texture for constraining the classification problem.

### **4.2.1.2 Experimental Procedure**

As described previously, data were collected in an industrial research kitchen at the Schlegel-University of Waterloo Research Institute for Aging's Centre of Excellence for Innovation in Aging. This kitchen was modeled after industrial research kitchens found in LTC homes. An optical imaging cage was constructed that enabled top-down image capture (see Figure 4.5). The



(a) Physical imaging set-up from above.



(b) Physical imaging set-up from side.

Figure 4.5: Physical set-up of imaging and weigh stations.

camera was connected to a computer for data acquisition and plates were weighed at an adjacent weigh station.

#### 4.2.1.3 Regular and Modified Texture Food Datasets

We used our regular texture foods dataset and our modified texture foods dataset. Table 4.4 provides an overview of dataset characteristics and a summary of all food items imaged can be found in Table 4.5. For each food item, one full serving was defined by the nutritional label portion size (regular texture dataset) or the recipe-defined portion size received from the kitchen, was weighed to the nearest 1 gram using an Ohaus Valor Scale.

For the *regular texture foods dataset*, where a serving size was referenced using volume, that volume of food (e.g., corn) was weighed and the mass was used thereafter. Nutritional information for each food item was supplied by the manufacturer except for meatloaf and mashed potatoes. Here, the nutritional information supplied by the manufacturer was combined for both food items and not on the individual food item level so nutritional information was approximated using USDA's food database. Since manufacturers supply nutritional information for minerals as percent daily value (assuming a 2000 calorie diet), for the wholefoods dataset, minerals were reported similarly. For consistency mass in grams was used to define all serving sizes.

For the *modified texture foods dataset* here, we expanded our MTF dataset with additional examples (without recipes) for further segmentation and volume estimation analysis. The nutritional analysis was conducted on the subset of the 314 images. To approximate a true food consump-

Table 4.4: Overview of dataset characteristics. The regular texture food dataset (RTF) was comprised of 3 “meal” plates each consisting of 3 foods imaged at every permutation of 25% simulated intake. The modified texture food dataset (MTF) set consisted of 134 food samples representing 47 foods each consisting of a set of at least one purée and one minced texture food. Each sample was imaged 5 times by progressively removing food with the exception of 6 samples consisting of 4 each with one lost image.

Dataset overview	<b>RTF</b>	<b>MTF</b>	<b>Total</b>
# images	375	664	1039
# samples	3	134	137
# classes	9	93	102
# foods represented	9	47	56
# foods with recipes	9	27	36

tion more closely, the modified texture food dataset was less stringently controlled in terms of portion size. The portion size was defined as the amount recorded on the LTC kitchen’s recipes (defined in millilitres). Since nutritional information were provided according to a volumetric serving size, we needed to convert from mass to volume for using weighed food for nutrient intake validation to define the expected mass of a full portion in millilitres. To accomplish this, we calculated the food’s density to convert by using the full plate’s “true volume” (in mL) with its mass (in grams). This enabled us to scale nutritional information using the same pipeline as the regular texture foods dataset for validating our findings using mass; it was not required for the system to operate.

#### 4.2.1.4 Training Dataset

To train our convolutional autoencoder, we again used the UNIMIB2016 dataset. However, we were aware of key limitations that green foods were underrepresented which affected the autoencoder’s ability to differentiate between all colours and textures. Noting the scarcity of green foods in the original dataset, we augmented the training dataset by adding in 91 examples of lettuce, 91 examples of peas, and 89 examples of spinach from the FoodX-251 food dataset [117]. We refer to this as the UNIMIB+ dataset. Figure 4.6 shows the effect of this under-representation of green through an orange-ish hue across the autoencoder’s decoder output trained solely on the UNIMIB2016 dataset for validation examples. The autoencoder was able to converge to a lower validation loss on the UNIMIB+ dataset. Empirically, this resulted in greens appearing greener, reds appearing redder, yellows and whites appearing less murky as well in bottom UNIMIB+ example compared to the UNIMIB2016 in Figure 4.6. This suggests the addition of the green samples enabled the autoencoder to learn good representations and encode features more deeply

Table 4.5: Cumulative list of foods imaged.

<b>Food Component</b>	<b>Regular Texture foods</b>	<b>Modified Texture Foods (with recipes as in Section 4.1)</b>	<b>Additional Modified Texture Foods (with segmentations)</b>
<b><i>Grains</i></b>	Cheese tortellini /w tomato sauce Oatmeal Whole wheat toast	Bow tie pasta /w carbonara sauce Macaroni salad Vegetable rotini	Basmati Rice
<b><i>Vegetables and fruits</i></b>	Corn Mashed potatoes Mixed greens salad	Asian vegetables Baked polenta /w garlic California vegetables Greek salad Mango & pineapple Red potato salad Sauteed spinach & kale Seasoned green peas Stewed rhubarb & berries Strawberries & bananas Sweet and sour cabbage	Beet & onion salad Cantaloupe Chunks Green beans with pimento Grilled vegetable salad Roasted cauliflower
<b><i>Proteins</i></b>	Meatloaf Scrambled egg	Baked basa Braised beef liver & onions Braised lamb shanks Hot dog wiener Orange ginger chicken Salisbury steak & gravy Teriyaki meatballs Tuna salad	Bean & sausage strata Grilled lemon & garlic chicken Pork tortiere Roast beef with miracle whip
<b><i>Mixed dishes</i></b>	Oatmeal cookie	Barley beef soup Blueberry coffee crumble cake Eggplant parmigiana English trifle Lemon chicken orzo soup	Black bean soup Broken glass parfait Butternut squash soup Cranberry spice oatmeal cookie Lemon meringue pie Peach jello Pear crumble cake Roast beef with miracle whip on whole wheat Turkey burger on wheat bun

and were aligned more closely with how a human would perceive the foods - a crucial point for our LTC application.

#### 4.2.1.5 Computational Methods

The following describes how our segmentation strategy was refined compared to Section 4.1, our general classification approach, followed by system automation through the use of a convolutional autoencoder.

##### *Refined segmentation strategy*

To enhance our EDFN-D beyond Section 4.1, here we incorporated more representation of green in the UNIMIB+ dataset for training and validation, and introduced a more optimal stop-criteria for training for segmentation for supporting more salient feature extraction. To summarize, an integrated imaging system was developed that employed three main steps to yield food intake volume estimates:

1. RGB information was used for automatic segmentation using a refined version of our EDFN-D. Compared to section 4.1, for this iteration, we incorporated more representation of green in the UNIMIB+ dataset for training and validation, and introduced a more optimal stop-criteria for training.
2. Segmentation was mapped onto depth information and converted from arbitrary depth units to  $\text{cm}^3$  for volume estimation, and comparisons of volumes were made to infer volume of food ingested from the relative difference to the full serving reference portion (Section 4.1)
3. Volume consumed was mapped onto nutritional information for intake approximation. These nutrient level intake estimates were then validated against the ground truth nutritional information by mass in Section 4.2.1.6.

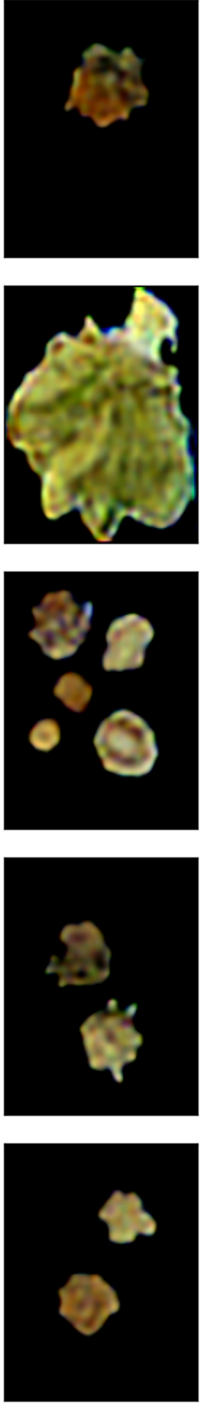
##### *General classification approach*

Here we use the UNIMIB+ data and splice off the autoencoder at the feature map as a latent feature extractor for classification (see Figure 4.7 for a system diagram and network architecture). As we will see in Chapter 6, we have previously taken a similar approach to classification on an similar task for predicting relative nutritional density of a dilution series of commercially prepared purées [181]. There we trained a global autoencoder across flavours and fine-tuned the model for flavour-specific modules and compared against hand-crafted features (64 colour; 7 texture) and the learned features. Given the similarity between this dilution series and modified texture foods which comprise 64% (664/1039) of our testing dataset and 47% of the LTC population receives modified texture foods [234]), taking a similar approach seems appropriate.





Input validation images



Decoder output from autoencoder trained on UNIMIB2016



Decoder output from autoencoder trained on UNIMIB+

Figure 4.6: Effect of under-representation of green foods on decoder output. The decoder output from the autoencoder trained on the UNIMIB+ dataset in the bottom appears less murky, more vibrant, and with truer perceived greens than the UNIMIB2016 counterpart in the middle.

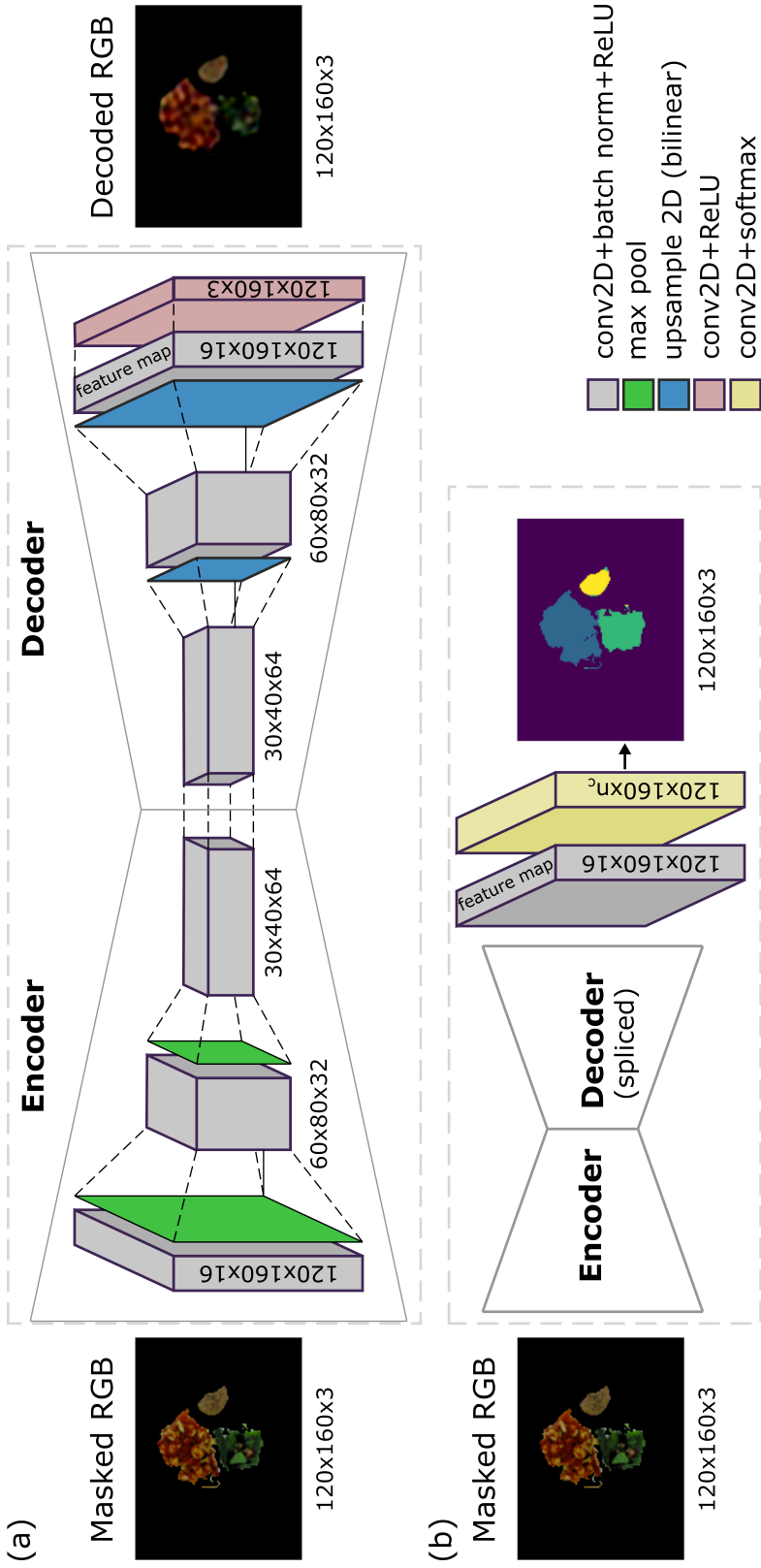


Figure 4.7: The top convolutional autoencoder network shows the architecture for learning feature representation: an input image is given and the output is a reconstruction of that image. Training minimized the error between input and output images; we used a MSE loss with Adam optimizer, a learning rate of 0.0001, and a batch size of 32. Our early stop criteria were a change of loss of  $< 0.0001$ , and a patience of 5 epochs. For the bottom, the autoencoder was spliced, weights frozen and only the last layer was retrained for classification. We used a categorical cross-entropy (ignoring background pixels) loss, with Adam optimizer, and a learning rate of 0.1. Our early stop criteria were a change of loss  $< 1 \times 10^{-5}$ , and a patience of 5 epochs. We used a 70%:30% train to validation split of augmented data. The data were augmented by generating 300 images from the full set of plates and applying random flips, rotations, and increased or decreased contrast.

### *Automation with a convolutional autoencoder*

For brevity, we report nutrient intake accuracy using the automated system (i.e., the automated classification case). A human-in-the-loop version where there is the opportunity to correct all misclassified regions (i.e., the best-case scenario) may improve results beyond what is reported here. By incorporating an initial pass of automation, the goal is to reduce the degree of intervention required by the user. For this automated approach, we leveraged a convolutional autoencoder for classification of foods which was roughly inspired by a highly successful convolutional neural network (CNN), the VGG network [213]. For a given meal or time of day, we fed the “masked” output from our EDFN-D (food/no-food detector) into our convolutional autoencoder. CNNs encode spatial information and given how food has differing degrees of cohesion, we felt the context of spatial information would be an asset. We chose a convolutional autoencoder approach because while we have fully labelled data for these datasets, in practice, this may be infeasible to collect - it will be far more likely that we have a large proportion of unlabelled data. Autoencoders can extract meaningful, generalizable features because their framework was built for compression, which tries to preserve as much information as possible to reconstruct the original input in an efficient manner. They’re comprised of an encoder which learns the key features in a compressed way, and a decoder, to get back the estimated signal and compare the output to the input based on a cost function; as a result, they do not require labelled training data. It also provides an advantage in interpretability because you can visualise the reconstruction and observe its behaviour as a sanity-check. As we saw in Section 4.6, this characteristic can be incredibly valuable for understanding inherent bias of training datasets. The “convolution” refers to learning where certain features exist over an image. A good analogy is using a blacklight flashlight at a crime scene to see blood. In this analogy the convolution is the shining flashlight across the scene and where blood illuminates, is the activation map for the learned blood-detector. In our case, the autoencoder learns what type of flashlight to use (i.e., a learned kernel) to see specific features which likely encodes some form of colour and texture information but are not easily intuitable. We trained an autoencoder to be a feature extractor using the UNIMIB+ dataset. The data were augmented by generating 300 images from the full set of plates and applying random flips, rotations, and increased or decreased contrast. We used one reference image (the full-portion image) to learn what each class looked like and then mapped subsequent instances onto these pre-labelled classes by grouping all the “full plates” of food for a given meal into the training set. We used a 70%:30% train to validation split of augmented data. We built a different classifier for each meal (described in more detail below); the idea is given the vast number of options, constrain the classification problem to each meal for fewer possible options. Given that there are many options and as new meals are planned, we will need something easily updateable given the realities of multiple food choices each day. This is an advantage of using our approach as only one labelled example is required. Using this method, our system requires

one full reference portion to which we can compare leftover plates to classify foods and infer intake. For nutritional intake estimation, we leveraged nutritional info per portion based on the Schlegel Villages menu planning software (or supplied by the manufacturer in the regular texture dataset) to link up proportional nutrient intake with the portion of food under the assumption that recipes were followed exactly.

More specifically, our autoencoder takes our input image which has three channels (R,G,B) and reduces to activations for each of 16 learned features as a global feature extractor. To go from the natural image output from the decoder to a feature extractor, we strip off the final layer which would output the decoded image (a lower resolution image similar to the input image) and take the output right after the 1x1x3 convolution (1x1 convolution across the 16-dimensional latent feature map, producing an RGB image) occurs to get activation maps of the learned features across each of the three channels. These learned feature activation maps are used to discriminate between foods using the reference full portion reference plates to fine-tune the network on a meal-by-meal basis. Learned weights are frozen from the global feature extractor and the final 1x1 (per-pixel) convolution layer is fine-tuned to map features to specific food classes for each meal. For example, in the full plate, the autoencoder learns  $n$  kernels that combine the features at each pixel into the target class. Then, for a new image, it goes to each pixel and applies the  $n$  kernels and says, "which one generated the strongest response?" and chooses that label for that pixel. We apply ground truth labels to the full portion plate so we can link up the proper proportional intake at the nutrient level and assess accuracy of the intake estimates including how this image-based system compares to the gold-standard weighed food approach. These classes *could* be left general for bulk intake estimates that are class-agnostic and would require no user input, it would assume equal consumption across a plate, similar (but less subjective) to what is currently being tracked in LTC. For example, lunch today contained 6 classes and this food blob belongs to class 2. While the network doesn't know which food comprises class 2, it knows they ate 79% of the plate that was served to them. Here, we instead chose to incorporate an additional level of detail (i.e., class 1 is "meatloaf") for more accurate nutrient intake assessment to create a system which allows for flexibility in the level of detail it provides. For example, there may be times when bulk intake estimates are preferred for their increased efficiency, while at others, dietitians would prefer a more fine-grained approach for a subset of at-risk referred residents. In this case our automated approach (with possible human-in-the-loop refinement) provides a automated (or semi-automated) alternative to support food-item specific intake without the assumption that foods are equally consumed.

#### **4.2.1.6 Nutrient Intake Association**

This step was comprised of three general stages: (1) determine the relative consumption of each food item compared to a full reference portion using food volume estimation from the depth

maps, (2) compare relative consumption to nutritional information to infer nutritional intake for each item, and (3) sum the inferred nutritional intake for each item across a plate for an estimation of total nutrition consumed during a meal (for the modified texture foods, this was across the plate of one food item). More specifically,

1. **Use relative changes in volume to estimate food intake:** The 0% eaten portion for each food item was considered as the full portion. Each subsequent portion (i.e., 25%, 50%, 75%, 100% consumed, or portions 2, 3, 4, 5 for the MDF dataset) for each food item was compared against the initial full portion to yield a relative volume change representing the intake of that specific food item. The relative change in volume was compared to the ground-truth relative change in mass for each food item.
2. **Use food intake to estimate nutrient intake for each food item:** The nutritional information for one serving was used as the reference full portion for each of the priority nutrients: calories, carbohydrates, fats, fibre, protein, calcium, iron, sodium, vitamin B6, vitamin C, vitamin D, vitamin K, and zinc. Given the proportion of consumed food for each food item, the nutritional information was scaled accordingly to estimate intake for each of the priority nutrients at the food item level.
3. **Use food intake to estimate overall nutrient intake:** Finally, the total nutrient intake consumed during the meal was estimated by summing the nutrient intake across all food items for each plate.

#### 4.2.1.7 Statistical Analyses

##### *System accuracy*

Segmentation accuracy was assessed using intersection over union as described in Section 4.1.2.6. Classification accuracy was described using top-1 accuracy and is summarized with per-meal classifiers. Bulk intake accuracy (i.e., class-agnostic, overall food volume intake) was assessed using mean absolute error (mL) and 3D% intake error also described in Section 4.1.2.6 where intake error was calculated for volume (3D) data relative to the full portion. All values are reported as mean  $\pm$  SD. Nutrient intake accuracy was assessed using the fully automated classification approach (i.e., without updating misclassified regions) to evaluate nutrition intake accuracy and is reported as mean  $\pm$  SD as well as % error.

##### *Validating nutrient intake estimation against weighed-food records*

All data were analyzed using Matlab version 2020b software. Linear regression was used to determine the goodness of fit through the degree of correlation with  $r^2$  to summarize the extent to

Table 4.6: Converting the regular texture dataset % daily values into absolute value to match the modified texture dataset. The Recommended Dietary Allowance (RDA) or Adequate Intake (AI) was used for individuals over 70 years of age and an average across the two sexes was assumed to represent 100% daily values.

Input Units	RDA/AI for > 70 Years			Output Units
	Males	Females	Assumed 100% Daily Value	
Fat (g)	n/a	n/a	n/a	g
Carbohydrates (g)	n/a	n/a	n/a	g
Fibre (g)	n/a	n/a	n/a	g
Protein (g)	n/a	n/a	n/a	g
Calcium%	1200 mg	1200 mg	1200 mg	mg
Iron%	8 mg	8 mg	8 mg	mg
Sodium (mg)	n/a	n/a	n/a	mg
Vitamin B6%	1.7 mg	1.5 mg	1.6 mg	mg
Vitamin C%	90 mg	75 mg	82.5 mg	mg
Vitamin D (IU)	ns	ns	ns	IU
Vitamin K (mcg)	ns	ns	ns	mcg
Zinc (%)	11 mg	8 mg	9.5 mg	mg

Where: g: grams, mg: milligrams, IU: international units, mcg: micrograms. While the AI is known for vitamins D and K, no foods in the regular texture foods dataset reported % daily values and are thus marked as not specified (ns)

which nutritional intake information from weighed-food mass is related to estimated nutritional information from food volume. The Bland-Altman method was used to describe the level of agreement between nutritional intake info from weighed-food mass compared to intake volume with mean agreement ( $\sigma$ ) and bias ( $\mu$ ) between methods [79]).

Several nutrients of concern in the regular texture foods dataset were reported in % daily value (i.e., calcium, iron, vitamin B6, vitamin C, and zinc). We converted these to absolute values to match the modified texture foods dataset using the 2005 Health Canada reference values for elements and vitamins. Where there was a difference across age, we used the >70 years old reference; where there was a difference in requirement by sex, we took the average value.

## 4.2.2 Results

### 4.2.2.1 Segmentation Accuracy

Based on the output summarized in Table 4.7, segmentation accuracy was good with an average IOU of 0.879 across datasets which was consistent with results in Table 4.2. Segmentation

accuracy ranged from 0.823 on the modified texture dataset at lunch, to 0.944 on the regular texture dataset for breakfast. From the perspective of IOU, the modified texture dataset was more poorly segmented by the EDFN-D, however, as we saw in Section 4.1.3, the degree of visual-volume discordance is higher and is discussed in Section 4.2.2.3.

#### 4.2.2.2 Classification Accuracy

As shown in Table 4.7, classification accuracy was higher for the regular texture dataset with top-1 ranging from 93.5% on breakfast and lunch to 95.1% on dinner. However, the regular texture dataset had only three classes per meal, so it was a less challenging classification problem, especially when considering the modified texture dataset had less texture variance. In contrast, the modified texture dataset ranged from 64.9% on Day 2 Dinner with 12 classes, to 89.0% on Day 1 Lunch with 15 classes. While these classification results are strong, there are still misclassifications of some segmented food regions which, as expected, may necessitate human-in-the-loop for rectifying segmented region errors. These misclassifications tended to occur near the edges of a food segment regardless of dataset which may be due to a less uniform representation near the edges either due to higher crumbliness (e.g., meatloaf crumbs), or due to the convolutional kernel extending into the "empty space" (i.e., the plate) making it easier to classify a pixel as food when there are food pixels surrounding it. However, as we will see in Section 4.2.2.5, nutrient intake estimation relative to ground truth weighed-food records indicate these errors in classification of some regions do not appear translate to large intake errors and this fully automated classification strategy may be deemed acceptable given the time-savings. Additional consideration for technology translation is warranted.

#### 4.2.2.3 Volume Estimation Accuracy

Similar to what we observed in the former part of this chapter, volume estimation was good with food volume error of  $2.5 \pm 9.2$  mL. However, this system continues to struggle with low-density foods. Our system previously struggled with salad and continues to here with the largest food volume error seen for RTD:L of  $-10.1 \pm 22.2$  mL. A similar issue of low-density foods is seen through the 3D % absolute error intake of  $14.4 \pm 13.1$  % which we suspect is due to the air pocket below some of the pieces of toast which sit on a tangential angle to the plate, or when two pieces are stacked with overhang as shown in Figure 4.8. This could be considered one of the classic examples of the "occlusion conundrum" with the imaging limitation of collection from an overhead view. This is an example of where segmentation could be done perfectly but would translate to volume estimation errors.

Table 4.7: Average segmentation and classification accuracies within and across datasets. There were no samples for Day 5 Dinner. RTD: Regular texture foods dataset; MTD: Modified texture foods dataset.

<b>Dataset</b>		<b>Segmentation accuracy</b>	<b>Classification Accuracy</b>
<b>Meal</b>	<b># of classes (# images)</b>	<b>IOU</b>	<b>Top-1</b>
RTD: Breakfast	3 (125)	$0.944 \pm 0.019$	93.5%
RTD: Lunch	3 (125)	$0.919 \pm 0.033$	93.5%
RTD: Dinner	3 (125)	$0.928 \pm 0.019$	95.1%
<i>RTF subtotal</i>	<i>9 (375)</i>	<i><math>0.929 \pm 0.027</math></i>	<i>93.9%</i>
MTD: Day 1 - Lunch	15 (90)	$0.841 \pm 0.123$	70.2%
MTD: Day 1 - Dinner	5 (25)	$0.823 \pm 0.099$	89.0%
MTD: Day 2 - Lunch	12 (74)	$0.863 \pm 0.118$	70.6%
MTD: Day 2 - Dinner	12 (91)	$0.840 \pm 0.122$	64.9%
MTD: Day 3 - Lunch	10 (85)	$0.834 \pm 0.132$	80.4%
MTD: Day 3 - Dinner	15 (109)	$0.859 \pm 0.100$	70.4%
MTD: Day 4 - Lunch	9 (60)	$0.871 \pm 0.113$	72.2%
MTD: Day 4 - Dinner	10 (90)	$0.837 \pm 0.107$	67.8%
MTD: Day 5 - Lunch	5 (41)	$0.881 \pm 0.117$	87.8%
<i>MTD subtotal</i>	<i>93 (665)</i>	<i><math>0.849 \pm 0.116</math></i>	<i>73.7%</i>
<b>TOTAL</b>	<b>104 (1040)</b>	<b><math>0.879 \pm 0.101</math></b>	<b>88.9%</b>



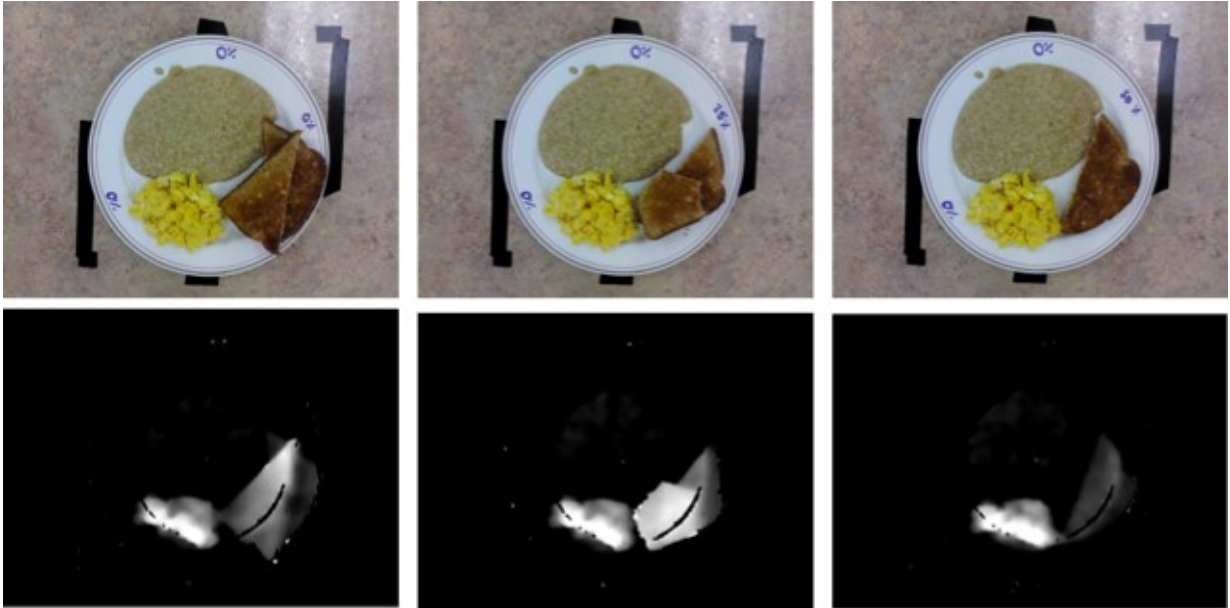


Figure 4.8: The toast occlusion conundrum. Where two pieces of toast are stacked with an overhang, the overhang is assumed to contain toast as a limitation to overhead imaging. Similarly, since toast is rigid, if it is placed on an angle up the side of the plate, there’s a similar issue. This is seen in the depth images where the brighter the pixel is the closer it is to the camera; brighter pixels are foods of greater height. Toast is a rigid plane and in the first example, we see a gradient from lower to higher near the tip with a similar, but less obvious trend in the third depth image. The visual colour range was adjusted to exemplify the toast height.

#### 4.2.2.4 Bulk Intake Accuracy

Table 4.8 summarizes the bulk intake accuracy within and across datasets. Compared to Section 4.1, for this iteration, we incorporated more representation of green in the UNIMIB+ dataset for training and validation, and introducing a more optimal stop-criteria for training for segmentation. In Table 4.2, we saw the mean absolute volume error was  $18.0 \pm 50.0$  mL for RTFs and  $2.3 \pm 3.2$  mL for MTFs and mean volume intake error of  $130.2 \pm 154.8$  mL and  $0.8 \pm 3.6$  mL for RTFs and MTFs, respectively. Here, our accuracy is higher with mean absolute food volume error of  $6.6 \pm 13.6$  mL for RTFs, and  $2.1 \pm 3.1$  for MTFs. Similarly, the bulk intake accuracy is higher with mean absolute intake error is greatly reduced for the RTFs with  $39.9 \pm 39.9$  mL but slightly higher for MTFs  $6.0 \pm 5.6$  mL. The higher degree of visual-volume discordance for MTFs compared to RTFs is again corroborated in Table 4.8 with the mean food volume error of  $3.8 \text{ mL} \pm 8.8$ , and higher mean volume error on the RTF dataset ( $6.6 \pm 13.6$  mL) than for the MTF dataset

( $2.1 \pm 3.1$  mL).

#### 4.2.2.5 Validating Nutrient Intake from Volume with Nutrient Intake from Mass

In Figure 4.9, the modified texture foods plates (blue) tended to be of lesser mass than the regular texture foods (red) so we observe a clustering effect for the macronutrients. We also observe a banding effect on fibre for the whole foods dataset due to how mass was controlled for matching 25% portion increments and given the relatively few foods that contained fibre in the regular texture food dataset. There is also much higher variance for the modified texture dataset for larger amounts of a nutrient (e.g., protein, fat, iron) with tighter variances observed on smaller portion sizes.

Based on the coefficients of determination for all nutrients of interest, nutrient estimates by volume are tightly linearly correlated with nutrient estimates from mass with  $r^2$  values ranging from 0.92 for fat to 0.99 for vitamins C, and K. Based on the Bland-Altman plots, not only are they tightly correlated but there is also good agreement between methods as evidenced by very little bias and zero contained within the limits of agreement. Ideally, the bias distributions would be centred around the y-intercept (i.e., a  $\mu$  of 0). This is typically the case with  $\mu$  ranging from a minimum of -0.01 for vitamin B6 (mg), zinc (mg), and fat (g), to a maximum of -2.7 for calories (kcal). This reflects the nutrient-level error having a degree of pixels misclassified and may be further improved with human-in-the-loop refinement. Taken together, these results suggest nutrient estimation using our AFINI-T system appears to be valid with the caveat that a wider range of portion sizes would assist in clarifying robustness.

### 4.2.3 Discussion

#### 4.2.3.1 Making sense of visual trends in Figure 4.9

There was a trend of the modified texture foods (blue) consistently appearing closer to the origin at lower values both by volume and mass. This was due to the size of the samples; for the regular texture foods, the mass of these portions were greater than that of the modified texture foods dataset. Some additional considerations to keep in mind are for the modified texture foods, these were collected on a continuous scale, whereas for the regular texture foods, the mass was strictly controlled to adhere to the 25% incremental bins. There was also a larger range in different samples and foods represented in the modified texture food dataset compared to the regular texture foods. As a result, we see a banding pattern among the regular texture foods in fibre and sodium which is an artefact of the data collection process. For the micronutrients, there was incomplete data available for the regular texture food items since these were prepared by a grocery store and a full nutritional breakdown is not mandated to be included on the nutritional label. I

Table 4.8: Bulk intake accuracy within and across datasets. There were no samples for Day 5 Dinner. “Food volume error” is equivalent to “Mean error bias”; “Error, intake” is equivalent to “Volume intake error”; and “3D % intake error” is the same as in Section 4.1. RTF: Regular Texture Foods Dataset; MTF: Modified Texture Foods Dataset; B: Breakfast, L: Lunch, D: Dinner; D#: Day number. For example, MTD: D1 - L is Modified Texture Foods Dataset: Day 1 - Lunch.

Meal	Dataset	# of classes (# images)	Food volume error			Bulk intake accuracy		
			Mean absolute food volume (mL)	Food volume error (mL)	Mean absolute error, intake (mL)	Error, intake (mL)	3D % absolute intake error	3D % intake error
RTD: B		3 (125)	3.0 ± 4.1	-2.2 ± 4.5	17.0 ± 14.3	-15.1 ± 16.3	14.4 ± 13.1	-12.0 ± 15.3
RTD: L		3 (125)	11.0 ± 21.7	-10.1 ± 22.2	76.1 ± 48.5	18.1 ± 88.7	13.7 ± 9.0	7.6 ± 14.6
RTD: D		3 (125)	6.0 ± 6.0	-5.8 ± 6.1	26.5 ± 14.4	-24.5 ± 17.7	11.2 ± 9.9	-2.9 ± 14.7
<i>RTF subtotal</i>		9 (375)	6.6 ± 13.6	-6.0 ± 13.9	39.9 ± 39.9	-7.2 ± 56.0	13.1 ± 10.9	-2.5 ± 16.8
MTD: D1 - L		15 (90)	1.0 ± 1.1	-0.7 ± 1.3	3.4 ± 3.3	-0.9 ± 4.7	5.0 ± 4.6	-0.3 ± 6.9
MTD: D1 - D		5 (25)	1.9 ± 2.9	-1.1 ± 3.3	4.1 ± 3.7	2.5 ± 5.0	7.4 ± 14.1	6.4 ± 14.6
MTD: D2- L		12 (74)	2.2 ± 3.3	0.0 ± 4.0	7.4 ± 7.3	6.1 ± 8.4	6.7 ± 5.5	5.1 ± 7.0
MTD: D2 - D		12 (91)	1.2 ± 1.0	0.1 ± 1.5	4.6 ± 4.3	2.9 ± 5.5	8.3 ± 7.7	5.5 ± 9.9
MTD: D3 - L		10 (85)	3.8 ± 5.1	-3.3 ± 5.5	7.6 ± 6.3	5.0 ± 8.5	11.5 ± 10.0	10.0 ± 11.5
MTD: D3 - D		15 (109)	1.9 ± 2.0	0.3 ± 2.7	5.5 ± 3.8	3.9 ± 5.4	6.7 ± 4.7	4.9 ± 6.6
MTD: D4 - L		9 (60)	1.5 ± 2.5	0.3 ± 2.9	5.6 ± 7.5	4.8 ± 8.0	6.3 ± 3.9	5.3 ± 5.2
MTD: D4 - D		10 (90)	2.1 ± 1.9	0.7 ± 2.8	6.5 ± 4.8	5.8 ± 5.6	6.0 ± 4.7	5.0 ± 5.8
MTD: D5 - L		5 (41)	3.4 ± 5.2	-1.5 ± 6.1	9.5 ± 6.7	7.8 ± 8.6	9.9 ± 5.9	7.7 ± 8.7
<i>MTD subtotal</i>		93 (665)	2.1 ± 3.1	-0.5 ± 3.8	6.0 ± 5.6	4.4 ± 6.9	7.6 ± 8.0	5.9 ± 9.4
<b>TOTAL</b>		<b>104 (1040)</b>	<b>3.8 ± 8.8</b>	<b>-2.5 ± 9.2</b>	<b>19.9 ± 30.8</b>	<b>-0.4 ± 36.7</b>	<b>9.9 ± 9.7</b>	<b>2.4 ± 13.6</b>

Table 4.9: Macronutrient intake accuracies within and across datasets. There were no samples for Day 5 Dinner and no recipes available for foods imaged on Day 2 Lunch. RTF: Regular Texture Foods Dataset; MTD: Modified Texture Foods Dataset; B: Breakfast, L: Lunch, D: Dinner; D#: Day number. For example, MTD: D1 - L is Modified Texture Foods Dataset: Day 1 - Lunch.

Dataset	Meal	# of classes (# images)	Nutrient intake accuracy $\mu \pm \sigma$ (% error)					
			Calories (kcal)	Carbohydrates (g)	Fibre (g)	Fats (g)	Protein (g)	
RTD: B		3 (125)	18.13 $\pm$ 16.76 (18)	2.68 $\pm$ 2.36 (3)	0.50 $\pm$ 0.48 (1)	0.67 $\pm$ 0.74 (1)	1.08 $\pm$ 1.07 (1)	
RTD: L		3 (125)	23.43 $\pm$ 21.64 (23)	2.70 $\pm$ 2.99 (3)	0.47 $\pm$ 0.30 (0)	0.88 $\pm$ 0.82 (1)	0.82 $\pm$ 0.85 (1)	
RTD: D		3 (125)	17.04 $\pm$ 17.38 (17)	2.57 $\pm$ 2.37 (3)	0.22 $\pm$ 0.23 (0)	0.66 $\pm$ 0.80 (1)	0.59 $\pm$ 0.75 (1)	
<i>RTF subtotal</i>		9 (375)	<i>16.31 <math>\pm</math> 13.63 (16)</i>	<i>2.11 <math>\pm</math> 2.05 (2)</i>	<i>0.37 <math>\pm</math> 0.39 (0)</i>	<i>0.83 <math>\pm</math> 0.80 (1)</i>	<i>1.11 <math>\pm</math> 0.98 (1)</i>	
MTD: D1 - L		15 (90)	9.31 $\pm$ 13.27 (9)	1.38 $\pm$ 1.52 (1)	0.08 $\pm$ 0.07 (0)	0.28 $\pm$ 0.57 (0)	0.31 $\pm$ 0.52 (0)	
MTD: D1 - D		5 (25)	28.17 $\pm$ 25.01 (28)	1.33 $\pm$ 1.06 (1)	0.09 $\pm$ 0.07 (0)	1.19 $\pm$ 1.09 (1)	2.93 $\pm$ 2.69 (3)	
MTD: D2- L		12 (74)	n/a	n/a	n/a	n/a	n/a	
MTD: D2 - D		12 (91)	37.25 $\pm$ 38.66 (37)	3.26 $\pm$ 4.10 (3)	0.11 $\pm$ 0.14 (0)	1.92 $\pm$ 1.83 (2)	1.77 $\pm$ 1.57 (2)	
MTD: D3 - L		10 (85)	12.46 $\pm$ 9.61 (12)	1.41 $\pm$ 1.15 (1)	0.13 $\pm$ 0.09 (0)	0.43 $\pm$ 0.33 (0)	0.72 $\pm$ 0.52 (1)	
MTD: D3 - D		15 (109)	5.06 $\pm$ 3.53 (5)	1.30 $\pm$ 0.90 (1)	0.10 $\pm$ 0.07 (0)	0.02 $\pm$ 0.01 (0)	0.05 $\pm$ 0.03 (0)	
MTD: D4 - L		9 (60)	5.90 $\pm$ 3.55 (6)	0.35 $\pm$ 0.16 (0)	0.08 $\pm$ 0.06 (0)	0.46 $\pm$ 0.32 (0)	0.14 $\pm$ 0.10 (0)	
MTD: D4 - D		10 (90)	5.06 $\pm$ 5.41 (5)	0.81 $\pm$ 0.86 (1)	0.23 $\pm$ 0.25 (0)	0.18 $\pm$ 0.20 (0)	0.19 $\pm$ 0.20 (0)	
MTD: D5 - L		5 (41)	14.98 $\pm$ 8.52 (15)	2.11 $\pm$ 1.21 (2)	0.19 $\pm$ 0.11 (0)	0.44 $\pm$ 0.26 (0)	0.65 $\pm$ 0.36 (1)	
<i>MTD subtotal</i>		93 (665)	<i>18.44 <math>\pm</math> 23.82 (18)</i>	<i>1.72 <math>\pm</math> 2.17 (2)</i>	<i>0.13 <math>\pm</math> 0.13 (0)</i>	<i>0.78 <math>\pm</math> 1.14 (1)</i>	<i>1.14 <math>\pm</math> 1.70 (1)</i>	
<b>TOTAL</b>		<b>104 (1040)</b>	<b>16.85 <math>\pm</math> 16.83 (17)</b>	<b>2.01 <math>\pm</math> 2.09 (2)</b>	<b>0.31 <math>\pm</math> 0.36 (0)</b>	<b>0.82 <math>\pm</math> 0.90 (1)</b>	<b>1.12 <math>\pm</math> 1.20 (1)</b>	

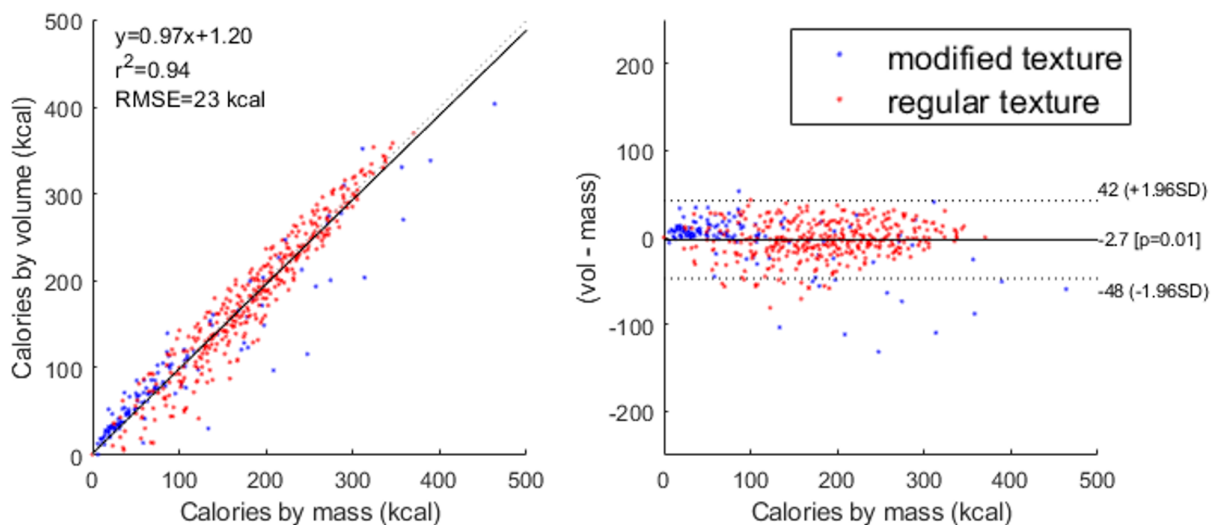
Table 4.10: Micronutrient intake accuracies of elements within and across datasets. There were no samples for Day 5 Dinner and no recipes available for foods imaged on Day 2 Lunch. RTD: Regular texture foods dataset; MTD: Modified texture foods dataset; B: Breakfast, L: Lunch, D: Dinner; D#: Day number. For example, MTD: D1 - L is Modified Texture Foods Dataset: Day 1 - Lunch.

Dataset	# of classes (# images)	Nutrient intake accuracy $\mu \pm \sigma$ (% error)			
		Calcium (mg)	Iron (mg)	Sodium (mg)	Zinc (mg)
RTD: B	3 (125)	10.66 $\pm$ 10.09 (11)	0.26 $\pm$ 0.24 (0)	20.77 $\pm$ 18.60 (21)	0.04 $\pm$ 0.03 (0)
RTD: L	3 (125)	11.96 $\pm$ 12.34 (12)	0.11 $\pm$ 0.07 (0)	32.20 $\pm$ 34.33 (32)	0.00 $\pm$ 0.00 (0)
RTD: D	3 (125)	2.95 $\pm$ 3.72 (3)	0.08 $\pm$ 0.09 (0)	45.71 $\pm$ 46.31 (46)	0.00 $\pm$ 0.00 (0)
<i>RTF subtotal</i>	9 (375)	9.94 $\pm$ 8.48 (10)	0.17 $\pm$ 0.18 (0)	18.50 $\pm$ 15.58 (18)	0.05 $\pm$ 0.06 (0)
MTD: D1 - L	15 (90)	8.45 $\pm$ 14.01 (8)	0.05 $\pm$ 0.05 (0)	16.43 $\pm$ 20.69 (16)	0.02 $\pm$ 0.02 (0)
MTD: D1 - D	5 (25)	1.86 $\pm$ 1.64 (2)	0.62 $\pm$ 0.57 (1)	17.05 $\pm$ 15.11 (17)	0.50 $\pm$ 0.46 (1)
MTD: D2 - L	12 (74)	n/a	n/a	n/a	n/a
MTD: D2 - D	12 (91)	6.04 $\pm$ 7.03 (6)	0.17 $\pm$ 0.19 (0)	11.16 $\pm$ 9.22 (11)	0.10 $\pm$ 0.12 (0)
MTD: D3 - L	10 (85)	2.53 $\pm$ 2.03 (3)	0.09 $\pm$ 0.08 (0)	13.87 $\pm$ 11.42 (14)	0.05 $\pm$ 0.04 (0)
MTD: D3 - D	15 (109)	1.08 $\pm$ 0.75 (1)	0.02 $\pm$ 0.01 (0)	0.08 $\pm$ 0.06 (0)	0.01 $\pm$ 0.01 (0)
MTD: D4 - L	9 (60)	1.34 $\pm$ 0.93 (1)	0.03 $\pm$ 0.02 (0)	12.11 $\pm$ 8.05 (12)	0.01 $\pm$ 0.01 (0)
MTD: D4 - D	10 (90)	2.98 $\pm$ 3.22 (3)	0.07 $\pm$ 0.08 (0)	3.84 $\pm$ 3.32 (4)	0.04 $\pm$ 0.04 (0)
MTD: D5 - L	5 (41)	7.12 $\pm$ 4.54 (7)	0.12 $\pm$ 0.06 (0)	18.93 $\pm$ 10.33 (19)	0.10 $\pm$ 0.06 (0)
<i>MTD subtotal</i>	93 (665)	4.37 $\pm$ 6.19 (4)	0.19 $\pm$ 0.34 (0)	12.92 $\pm$ 12.73 (13)	0.14 $\pm$ 0.27 (0)
<b>TOTAL</b>	<b>104 (1040)</b>	<b>8.51 <math>\pm</math> 8.31 (9)</b>	<b>0.18 <math>\pm</math> 0.23 (0)</b>	<b>17.07 <math>\pm</math> 15.09 (17)</b>	<b>0.07 <math>\pm</math> 0.15 (0)</b>

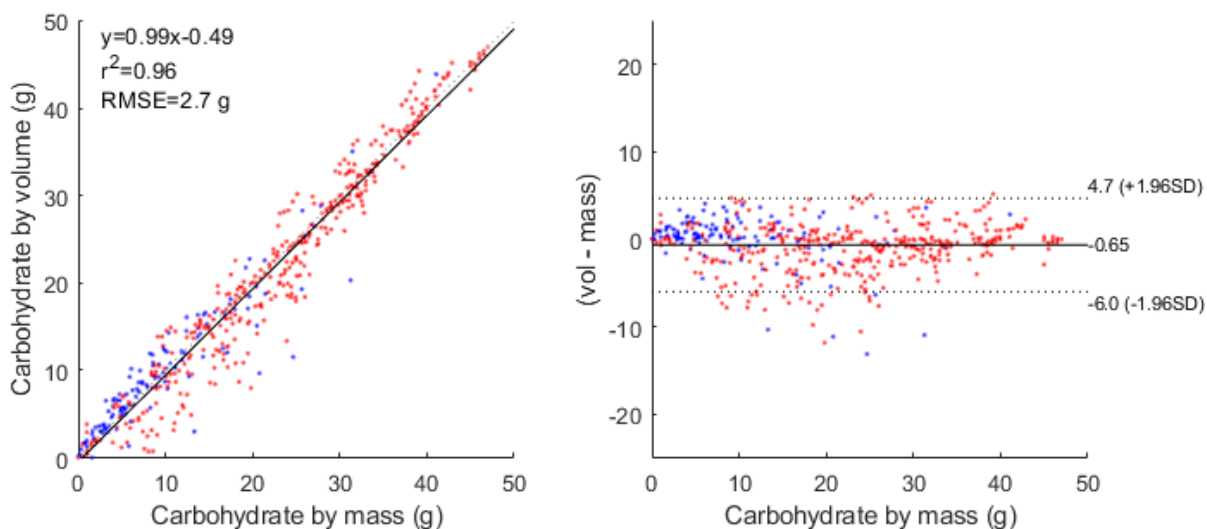
Table 4.11: Micronutrient intake accuracies of vitamins within and across datasets. There were no samples for Day 5 Dinner and no recipes available for foods imaged on Day 2 Lunch. RTF: Regular Texture Foods Dataset; MTD: Modified Texture Foods Dataset; B: Breakfast, L: Lunch, D: Dinner; D#: Day number. For example, MTD: D1 - L is Modified Texture Foods Dataset: Day 1 - Lunch.

Dataset	# of classes (# images)	Nutrient intake accuracy			
		Vitamin B6 (mg)	Vitamin C (mg)	Vitamin D (IU)	Vitamin K (mcg)
Meal		$\mu \pm \sigma$ (% error)			
RTD: B	3 (125)	0.02 ± 0.02 (0)	0.00 ± 0.00 (0)	n/s	n/s
RTD: L	3 (125)	0.00 ± 0.00 (0)	5.70 ± 4.05 (6)	n/s	n/s
RTD: D	3 (125)	0.00 ± 0.00 (0)	0.78 ± 0.96 (1)	n/s	n/s
<i>RTF subtotal</i>	<i>9 (375)</i>	<i>0.01 ± 0.02 (0)</i>	<i>0.00 ± 0.00 (0)</i>	<i>n/s</i>	<i>n/s</i>
MTD: D1 - L	15 (90)	0.01 ± 0.01 (0)	1.87 ± 1.59 (2)	0.00 ± 0.01 (0)	2.02 ± 2.29 (2)
MTD: D1 - D	5 (25)	0.10 ± 0.09 (0)	0.71 ± 0.50 (1)	4.38 ± 4.03 (4)	0.32 ± 0.29 (0)
MTD: D2 - L	12 (74)	n/a	n/a	n/a	n/a
MTD: D2 - D	12 (91)	0.02 ± 0.03 (0)	0.18 ± 0.19 (0)	0.06 ± 0.09 (0)	0.01 ± 0.01 (0)
MTD: D3 - L	10 (85)	0.01 ± 0.00 (0)	0.31 ± 0.21 (0)	0.00 ± 0.00 (0)	1.69 ± 1.18 (2)
MTD: D3 - D	15 (109)	0.01 ± 0.01 (0)	1.70 ± 1.18 (2)	0.00 ± 0.00 (0)	0.16 ± 0.11 (0)
MTD: D4 - L	9 (60)	0.00 ± 0.00 (0)	0.66 ± 0.46 (1)	0.00 ± 0.00 (0)	2.33 ± 1.63 (2)
MTD: D4 - D	10 (90)	0.01 ± 0.01 (0)	2.80 ± 3.04 (3)	0.03 ± 0.03 (0)	3.55 ± 3.86 (4)
MTD: D5 - L	5 (41)	0.01 ± 0.01 (0)	0.96 ± 0.70 (1)	0.16 ± 0.10 (0)	0.66 ± 0.36 (1)
<i>MTD subtotal</i>	<i>93 (665)</i>	<i>0.03 ± 0.05 (0)</i>	<i>1.05 ± 1.50 (1)</i>	<i>0.87 ± 2.41 (1)</i>	<i>1.13 ± 1.98 (1)</i>
<b>TOTAL</b>	<b>104 (1040)</b>	<b>0.02 ± 0.03 (0)</b>	<b>0.27 ± 0.88 (0)</b>	<b>0.87 ± 2.41 (1)</b>	<b>1.13 ± 1.98 (1)</b>

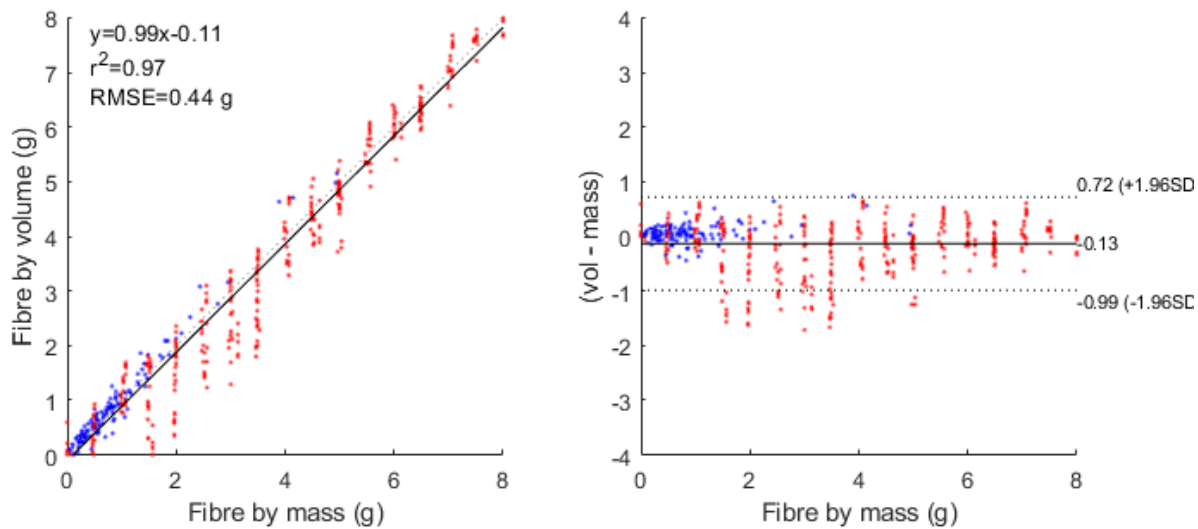
Figure 4.9: Correlation and agreement between mass and volume estimates for determining nutritional intake at the whole plate level across all imaged samples. Left depicts the goodness of fit with linear regression and coefficient of determination ( $r^2$ ), right depicts the degree of agreement between measures and bias from the Bland-Altman method.



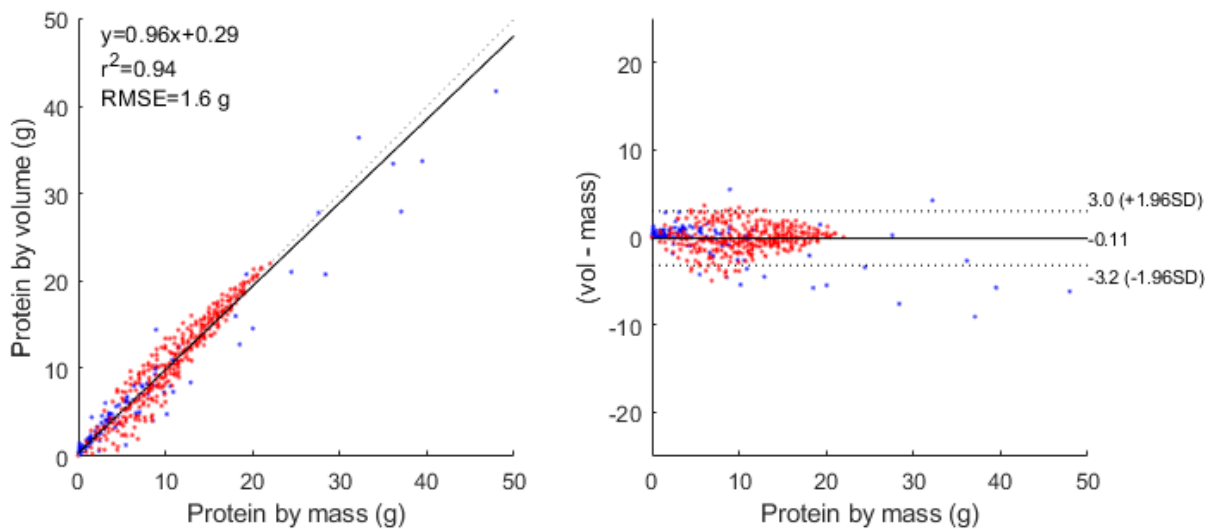
(a) Correlation and agreement between mass and volume estimates calories.



(b) Correlation and agreement between mass and volume estimates of carbohydrates.

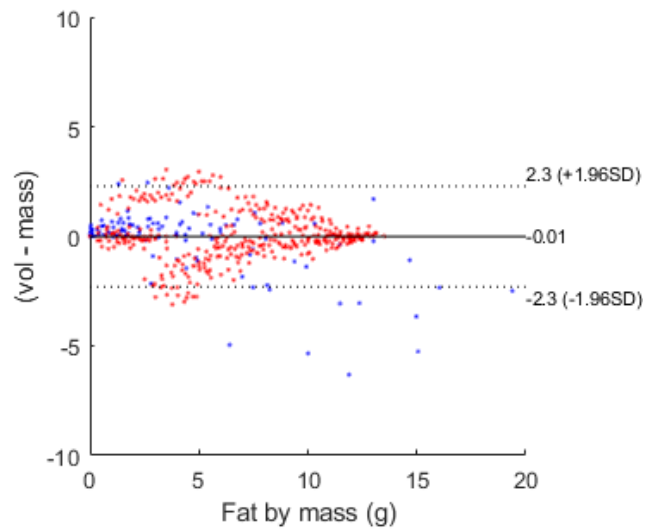
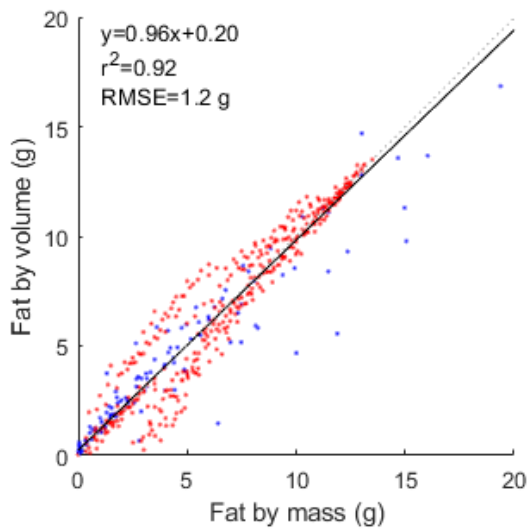


(c) Correlation and agreement between mass and volume estimates of fibre.

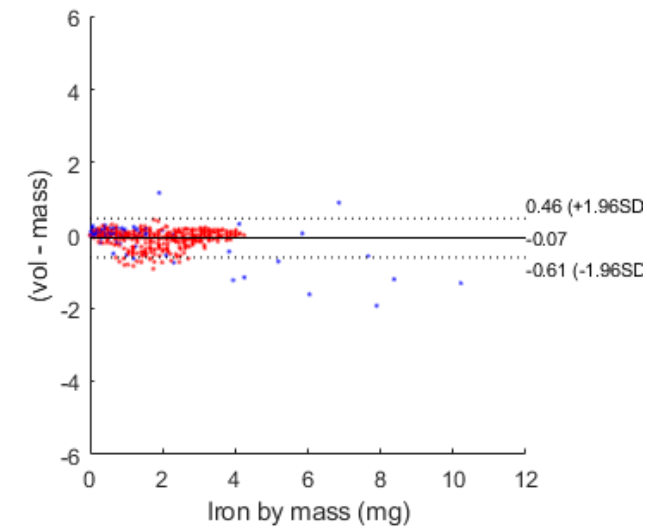
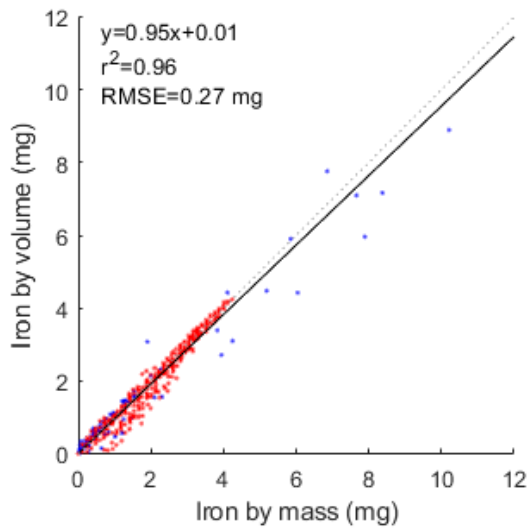


(d) Correlation and agreement between mass and volume estimates of protein.

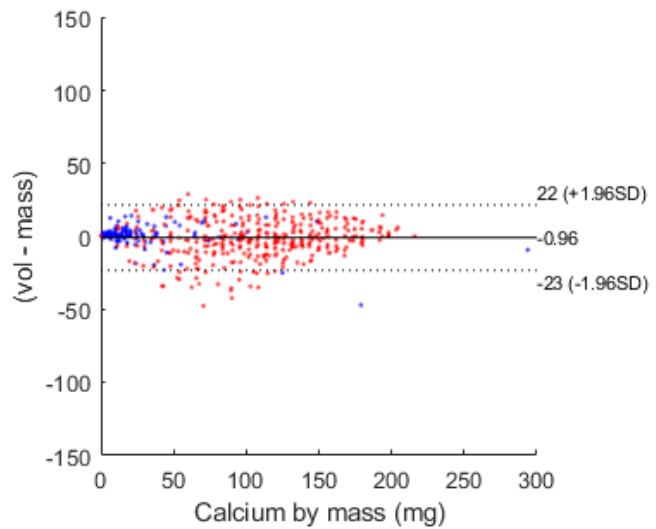
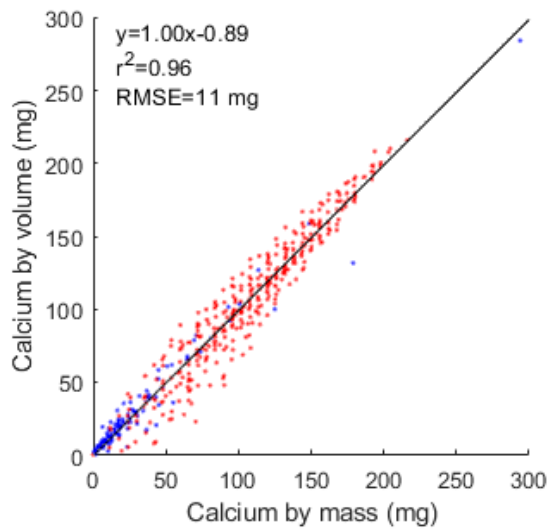




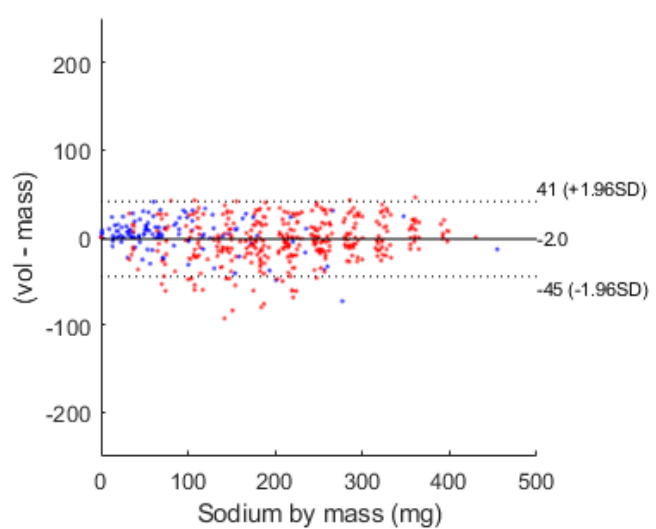
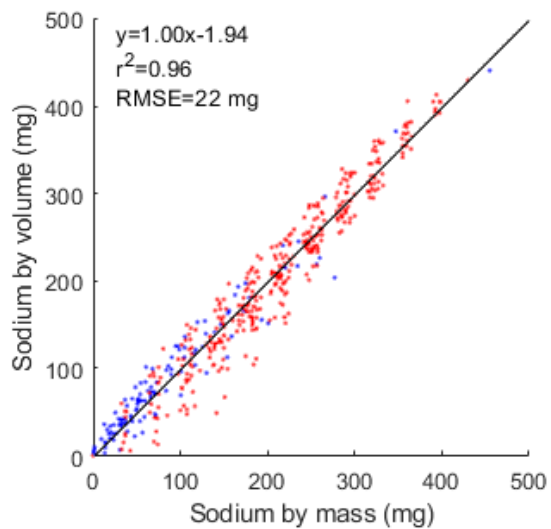
(e) Correlation and agreement between mass and volume estimates of fat.



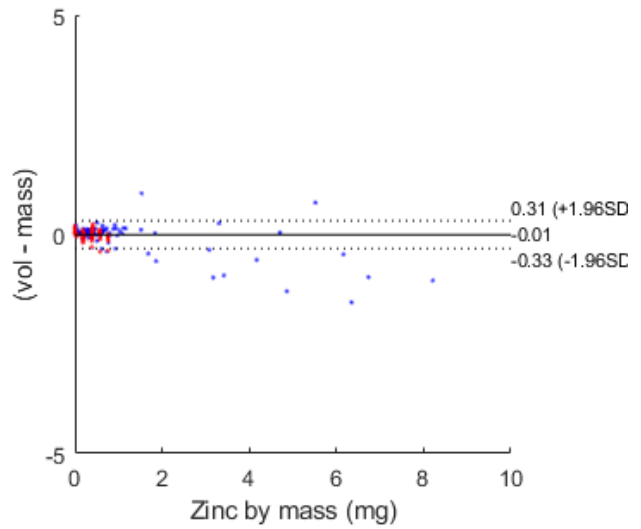
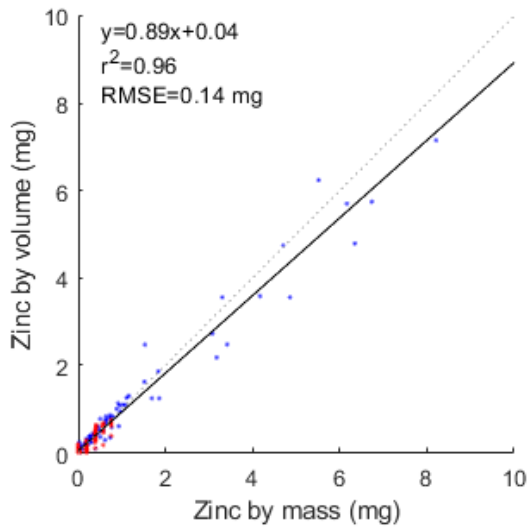
(f) Correlation and agreement between mass and volume estimates of iron.



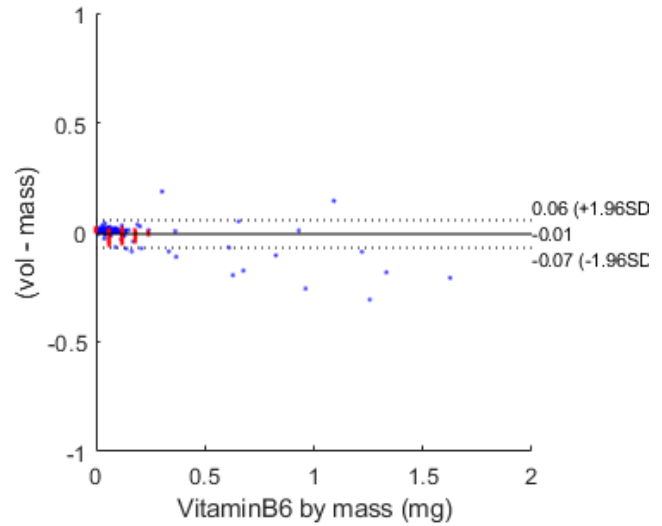
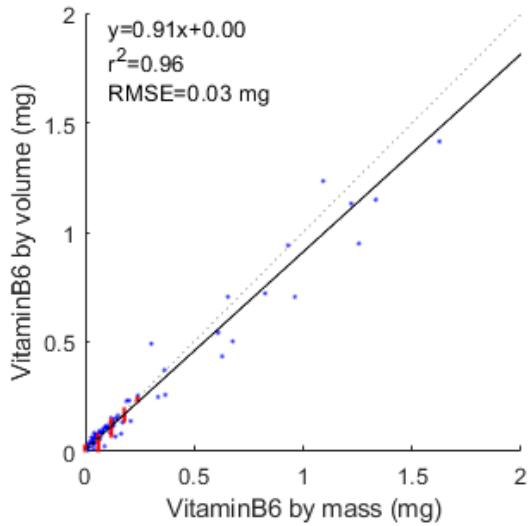
(g) Correlation and agreement between mass and volume estimates of calcium.



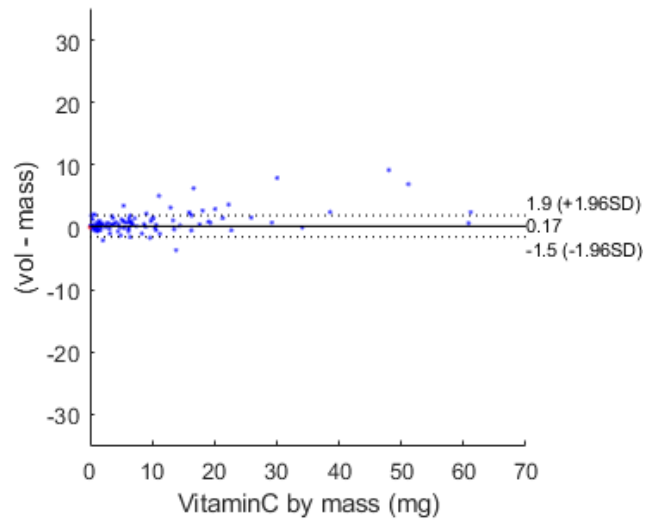
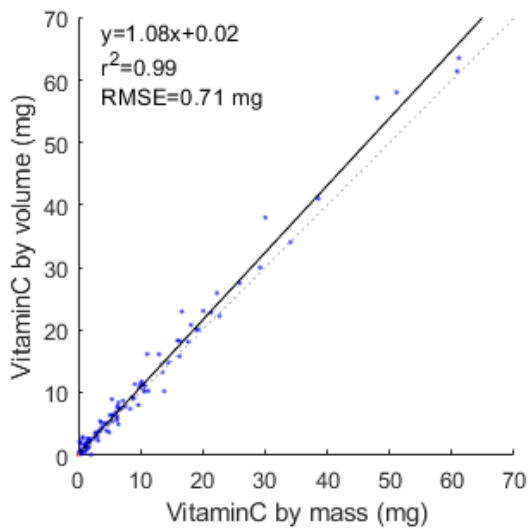
(h) Correlation and agreement between mass and volume estimates of sodium.



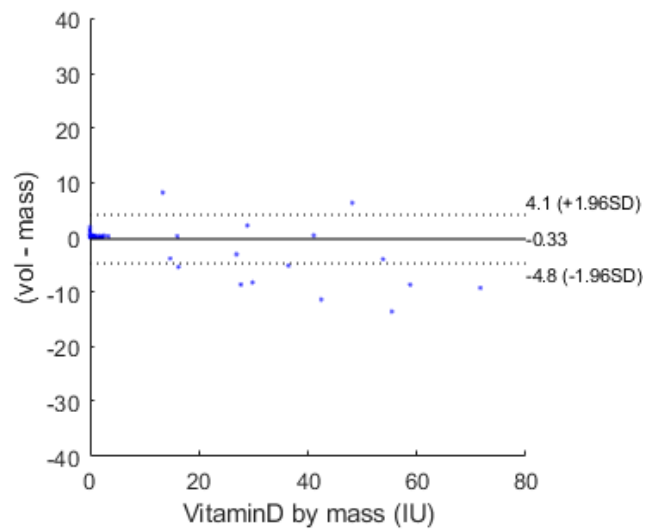
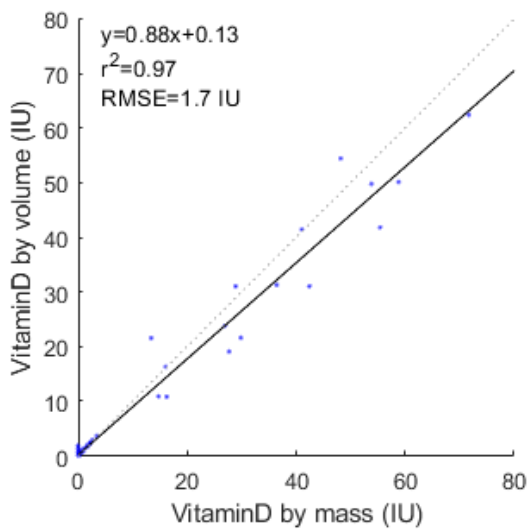
(i) Correlation and agreement between mass and volume estimates of zinc.



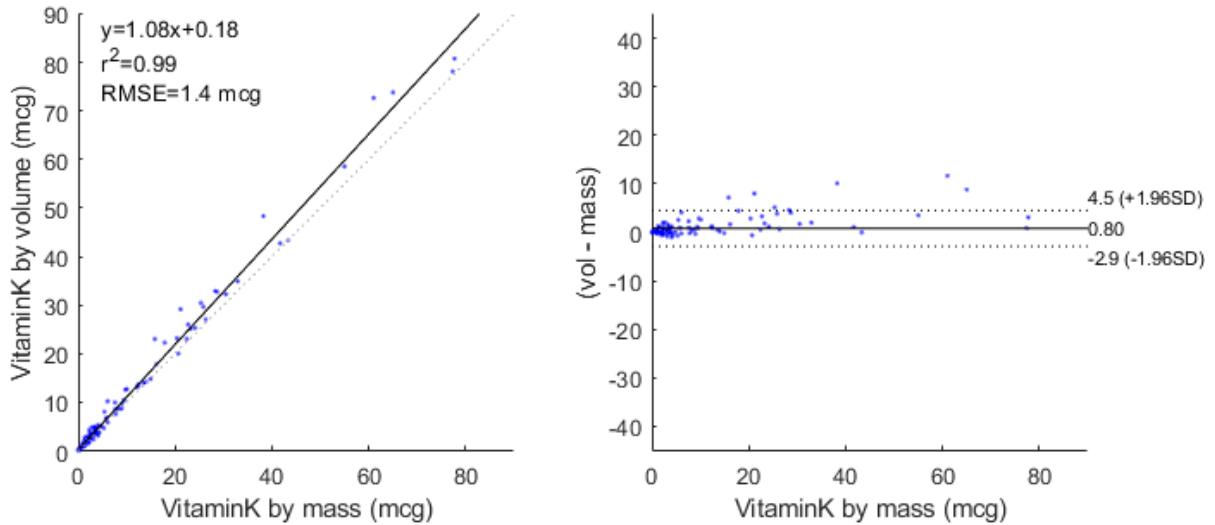
(j) Correlation and agreement between mass and volume estimates of vitamin B6.



(k) Correlation and agreement between mass and volume estimates of vitamin C.



(l) Correlation and agreement between mass and volume estimates of vitamin D.



(m) Correlation and agreement between mass and volume estimates of vitamin K.

suspect that with an increased dataset, much of the trends observed would be dampened as the true distribution emerges, especially for higher mass samples.

#### 4.2.3.2 Benchmarking the AFINI-T approach with current practice and requirements

With the end-to-end AFINI-T system in place, we can compare against the current workflow. One requirement identified in Chapter 3, was for the system to run on an iPad. By design, methodology and models were selected to support portability. More specifically, for segmentation, the EDFN network is 300 MB, and for classification the autoencoder is 1 MB with 0.5 MB for each meal-specific classifier.

The second benchmark is with respect to theoretical task completion time. In terms of benchmarking theoretical task completion time, we can compare to results from Chapter 3. When assuming a very conservative estimate including food handling of 10 seconds per image for acquisition (results from Chapter 3 did not include food handling time), the time for preprocessing (e.g., plate finding) takes approximately 2.5 seconds per image, with segmentation taking 0.7 seconds per image and classification of 0.05 seconds per image. As shown in Table 4.12, even based on these conservative estimates, the theoretical completion time using AFINI-T meets the low end of task completion times (9 minutes 45 seconds vs a mode rating of 10-14 minutes of completion time for charting one meal). Here, I've assumed separate imaging for each of the appetizer, main and dessert for each resident. If instead we consider acquisition as only acquir-

Table 4.12: Summary of length of time required to complete food and fluid intake charting for one neighbourhood comprised of 16 residents (Stage 1) compared to theoretical AFINI-T processing. \* n is the number of responses with the mode rating out of N, the total number of responses.

Type	Mode Time (n/N responses)	Time Range	AFINI-T Estimate (1 sec acquisition)	AFINI-T Estimate (10 sec acquisition)
Food (per meal)	10 - 14 mins (3/9)	<10 to 25+ mins	2 min 34 sec	9 mins 45 sec
Fluid (per meal)	10-14 mins (4/10)	<10 to 25 mins	N/A	N/A
Snack (per snack)	<10 mins (5/9)	<10 to 19 mins	52 sec	3 mins 15 sec

ing the image (estimated 1 second), this drops to 2 minutes 34 seconds. The true completion time will likely take between these upper and lower bounds, but the key take-away is AFINI-T is platformed to take less time than the current methodology, and with the added benefit of being objective and capture data at a resident-centric level. Instead of a resident’s intake being binned into the 25% bin across the average foods served that day, AFINI-T captures details to the mL, and tracks personalized items ordered on a resident-by-resident basis.

#### 4.2.3.3 Comparison to the literature

In terms of how this classification accuracy stacks up next to the literature, it is a bit challenging to assess since there are no food intake datasets to benchmark against. Additional considerations affecting the ability to compare include the number of classes included, inconsistencies with “accuracy” reporting (e.g., top-1 vs top-4 accuracy) and the complexity of classification problem (e.g., whole raw foods vs prepared meals vs modified texture versions of those prepared foods). That said, for classification methods based on handcrafted features: Zhang et al. report 88.2% accuracy for whole foods (entire pineapple) consisting of 18 classes [247], Bolle et al. achieved 95% using top-4 accuracy for vegetable identification in a supermarket [25], Rocha et al. achieved 99% using top-2 accuracy for some fruits and vegetables by fusing three types of features (including Unser’s features) [201], Arivazhagan et al achieved 85% accuracy on 15 types of produce using a minimum distance classifier [9], and Chowdury et al. attained 96.55% accuracy on 10 vegetables using colour and texture features and a neural network classifier [46].

Regarding a trend for learned features, a deep learning approach has had a comparatively slow adoption in the field of food imaging. As we saw in Chapter 2, accuracy around segmentation and classification tends to be either not mentioned [42, 166, 67, 70, 45] or was explicitly stated as being beyond the scope of the present version of their system [166]. In recent years, systems have begun to leverage learned features [162, 186, 67], an important consideration for both classification in particular. Oftentimes, feature extraction and classification are combined,

making sources of error difficult to disentangle. This is further confounded when segmentation and classification accuracies are combined instead of as two sub-processes. As we saw in Chapter 2, classification accuracies of 100% (11 classes) [186], 82.5%(15 classes) [67]. Alternative methods employed for classification were AdaBoost [166], K Nearest Neighbours [93], and support vector machines [45, 42] with reported classification accuracies of 68.3% (50 classes) [42], 99.1%(6 classes) [45]. While direct comparison between the AFINI-T system and LTC intake data isn't possible since the AFINI-T system is the first to measure food intake and considers modified texture foods, it would seem that based on our results, our DNN approach is on among the highest performing approaches with a top-1 accuracy of 88.9%. Furthermore, the type of data represented in our MTF and RTF datasets for LTC contain more complex food scenarios as they are prepared foods (RTF: 93.9% accuracy; MTF: 73.7%), and the accuracy we report is top-1 suggesting the AFINI-T approach may outperform the others.

At the inference level, few papers report percent error at the nutrient level and tend to focus on calorie estimation. The few that have reported calorie estimation error of: 0.09% (mean absolute error) on 6 categories using random forests and support vector machines [45], and 0.25% (mean standard error) on 11 categories of entire foods (e.g., green pepper) using a CNN [186]. Additionally, [67] report 80% of calorie estimates falling within 40% error (35% within 20% error) on 15 classes using a multi-task CNN with a maximum correlation coefficient of 0.81 ( $r^2$  0.64 equivalent), and top-1 accuracy of 82.48% [67] and they report in this manner for comparison to [166] with 79% of calorie estimates falling within 40% error (35% within 20% error) using hand-crafted features with a correlation coefficient of 0.32 ( $r^2$  0.10 equivalent) [166]. Our AFINI-T system had an % error of 2.4% across 13 nutrients on the 56 categories (102 classes) of food with an minimum  $r^2$  value of 0.92 (0.94 for calories). Our average top-1 accuracy was 88.9%, ranging between 95.1% on 3 classes (RTD:Dinner) to 70.4% and 89.0% on 15 class meals (MTD: Day 3 Dinner, MTD: Day 1 - Lunch). Based on these comparisons, our work performs among the best despite accounting for more complex meal scenarios and across 13 nutrients. While there has been relatively little work done in this area, our results highlight the significance of our contribution both from the technical implementation as well as from the applied perspective working towards translational research.

While each of the above methods have their own quirks and limitations, the bottom line with respect to portion size or calorie estimation is that they at least achieve similar accuracies to average reported human error. With the AFINI-T system, not only are we able to measure a specific resident's intake (as opposed to the proportion consumed across the average of all foods offered), it removes subjectivity, and can be tracked to the nutrient level in competitive time. The result: higher quality data that can be used to inform resident preferences, streamline referrals to registered dietitians along with a data-driven approach for monitoring and evaluating nutritional interventions.

#### 4.2.3.4 Additional considerations

While we were able to provide complete total automation for segmentation, classification, and food and nutrient intake estimation for simulated intake plates, the current system relies on one fully hand-labelled and hand segmented image for a full reference portion to provide specific semantic segmentation and classification labels for each meal-specific classifier. If this would be deemed as too cumbersome, the next level of automation would be incorporating a semi-automatic method (e.g., graph cut) on the backend to aid the user in hand segmenting the reference images. In that case the dietitian or user needn't hand segment anything - they would simply assign a label to each of the pre-segmented types of foods present with a few lines to indicate where each food type resides (as opposed to hand segmenting an outline for each segment containing every type of food and its associated food-name label). This latter example is in line with the co-designed user interface and workflow we saw in Chapter 3.

Our results suggest that our method for estimating food intake is in strong agreement and tightly correlated with true intake. Especially in the case of larger intake portions, our methodology yielded accuracy of nutrient content with less than 5% error. Interpretation of the acceptability of the precision and accuracy of the system requires further input from users. If improvement is warranted, it will require a degree of human input or expanded models. This may be in the form showing output classification masks so misclassified segments could be reclassified as appropriate. Alternatively it could be to seed regions from menu to select item so it's tightly constrained and applying region growing; this approach is in line with what was integrated into the collaborative co-design prototype development outlined in Chapter 3.

While evaluation of accuracy “in the wild” is prudent, in the current approach segmentation of only one reference image is required and we have shown even when some pixels are misclassified (typically around the edges where a food's variance is higher), there is reasonable nutrient intake accuracy robustness. One other consideration is that not all foods have nutrient values for every nutrient of interest often rely on complex imputations for estimates [107]. Nutrients like vitamin D are often incomplete [102], and this is also reflected in the few datapoints on vitamin D plot in Figure 4.91. Because there has been so little work around more accurate nutrient intake tracking within the LTC domain, the error targets may require further thought. Additional discussions with end-users and nutrition experts is warranted to evaluate the utility and appropriateness of reporting these values, what margin of error is deemed acceptable for supporting trust in the system, and other considerations given the quality of data included in the underlying nutritional databases.

With smaller intake amounts and therefore smaller relative portion differences, the error was larger. This was ultimately why there was clustering near the origin for several of the nutrients, especially those from the MTF dataset (shown in red). To improve on this in the future, we must



both collect supplementary data to expand the dataset as well as consider from where this error arose. We hypothesize this error is primarily due to segmentation, rounding, and potentially depth map variance.

One reason we chose to keep segmentation separate from classification is because when combined, it is difficult to tease out where error has been introduced into the system and to evaluate the extent of the error. This is particularly relevant within the long-term care context where some residents may consume very few bites of food. Monitoring these residents who are at highest risk for malnutrition is imperative so “hidden” errors which may not be accounted for would propagate to a decreased overall accuracy of food intake. In the case of under-estimation this is less of a problem if dietary requirements are being met; however, over-estimation of food intake has a large impact on quality of care and the ability to detect which residents are at risk (and their level of risk).

With respect to rounding, recall that relative volume differences were compared between the reference portion and the leftovers. For smaller absolute intake (or smaller portion sizes as was the case with the modified texture foods samples), this error is then propagated forward likely yielding a less precise nutritional intake estimate.

Regarding depth-map variance results indicate that nutritional intake estimates, had greater variation at the lower levels of intake (larger spread at lower intake levels). This may correspond to the amount of variation in estimation at smaller levels of intake/larger food left on the plate. We speculate this is because of a compounding of small discrepancies in depth maps which gets propagated to volume and then to nutritional intake. Future work will address this by incorporating depth-map variance as a feature to describe the food item. For example, for a green salad, we would expect a higher variance in depth map because it is a non-dense food item and in contrast, meatloaf or slab cake would have very low depth-map variance across the food item as these items are more block-like. Supporting this hypothesis was the observation that most of the classification errors arose from the peripheral edges of a sample, where variability was higher. Recall that the training data contained only full-reference portions. I expect that with additional data representing a wider variety of “intake” data, that model refinement could improve in these higher variance settings.

To support a “glass box” level of interpretability for classification we had initially considered using support vector machines, random forests and logistic regression for classification. We first employed colour-based features (e.g., Unser’s features, previously used regarding food [247, 227, 64, 63, 181, 201]). However, it appeared colour-based and texture features do not encode deeply enough the nuances between food types which is an especially important consideration for modified texture foods (e.g., minced or puréed foods) where there is less variability. Preliminary analysis yielded classification accuracy below 70% with these methods. And while more easily

intuitive features based on colour and texture may bring the decision process more closely to resemble a human’s discrimination, given they do not appear to be descriptive enough in this case, the goal here is to sacrifice a degree of intuitability to distill a feature set from training a convolutional auto-encoder.

More broadly, future directions will be to add an additional stage for automatic food-type classification as specific foods rather than arbitrary classes with associated nutritional values (i.e., mashed potatoes are classified as mashed potatoes after the initial segmentation step). In addition, we wish to improve upon the algorithms to handle more complex food types (e.g., salads and/or soups in which the food is comprised of multiple components) as well as more complex plates of food to address food mixing as seen with mashed potatoes or more generally to address plates “in the wild”.

## 4.3 Conclusion

In summary, we proposed an application-driven design for a novel fully automatic multisensor segmentation system which leverages depth-refinement for improved accuracy. We assessed our system on two representative LTC food intake datasets which included simulated intake plates since current datasets contain only full-plate portions or do not contain pixel-level segmentations. For further advancing the field, additional food intake datasets with pixel-level segmentation are needed. A system such as the one presented here which approximates a human assessor but is objective, more consistent, and can more accurately quantify food intake measurements may provide a valuable step towards automated tracking of food and fluid intake within LTC.

### 4.3.1 Key Contributions

The ability to image to nutritional intake in literature previously has not been as depth, did not include modified texture foods, and was not designed for LTC or healthcare setting (the focus has been individual-centric and typically for for weight loss). The AFINI-T system provides a very different application in a way that has not been done elsewhere.

- The first simulated regular and modified texture food intake dataset for LTC.
- A novel DNN for automated depth-refined food segmentation, and with that the ability to “see” bulk food intake objectively. This has demonstrated the novel ability to segment foods, including modified texture foods, and it works well.
- A novel DNN for LTC food classification incorporating domain-specific priors, and with that the ability to infer macro and micronutrient with an end-to-end solution.

#### UP NEXT...

Based on these strong results, it would appear our general framework for a novel proof-of-concept automated system to measure food intake and nutritional composition estimation using an integrated RGB-D camera for image acquisition is both feasible and valid. Specifically, our nutritional intake from relative changes in volume were strongly correlated with intake from weighed-food records ( $r^2 \geq 0.92$ ) with excellent agreement on intake estimates for all thirteen nutrients ( $\sigma \leq 2.0$ ). But regardless of how well the developed technology performs, if the approach and methodology is not deemed acceptable by users, it will not be used. In Chapter 5, these potential barriers to uptake are assessed to probe AFINI-T's palatability including acceptability, perceived workload and user perceptions of trust.

# Chapter 5

## A New System Prototype's Palatability

Chapter 4 introduced the “brains” of the AFINI-T system answering the questions of where food is on a plate, what food is there, and how much was consumed using simulated intake. But the question remains, how might acceptability of this technology be met from the user perspective? To set this system up for minimizing barriers to uptake user perspectives of system acceptability, perceived workload and user perceptions of trust in the system are essential. This is the focus of this chapter where I seek to address reducing barriers to adoption by ascertaining feasibility from the user's perspective. More specifically, the objectives of this chapter are to describe continued efforts to remove feasibility-related barriers to uptake and facilitate confidence in design decisions for user-centered technology development. Here, we evaluate a user-driven, practice relevant early-stage prototype to inform future directions including user perceptions of workload, usability, and receptivity of the AFINI-T system prototype.

### **This chapter contains content previously published from...**

**KJ Pfisterer**, J Boger, A Wong. Prototyping the Automated Food Imaging and Nutrient Intake Tracking (AFINI-T) system: A modified participatory iterative design sprint. *JMIR Human Factors* 2019;6(2):e13017. doi: <http://dx.doi.org/10.2196/13017>. On this paper, K.J.P. was the main contributor from project inception to planning, implementation, data collection, analysis, interpretation, and writing.

## **5.1 Establishing System Acceptability, Perceived Workload, and User Perceptions of Trust in the System**

Returning to the modified participatory google sprint mentality we conducted a usability assessment to elucidate preliminary feasibility with end-users early on through the evaluation of prototypes through pilot testing. Output from this stage informed how the prototypes could be improved for development of a working system in the future. We added one additional stage as well for final prototype validation, to receive additional feedback from a new group of RDs, directors and assistant directors of food services not previously involved in the prototype design to provide a fresh perspective to minimize bias.

## **5.2 Methods**

The system acceptability assessment built off Chapter 4 and was guided by several conceptual frameworks: (1) conducting interdisciplinary research [24, 37]; (2) leveraging user-centered design and participatory design [185, 203]; and (3) evaluating usability [32, 196] and perceived workload [88]. The prototypes developed in Stage 4 of Chapter 4 were used for pilot evaluation.

Five participants self-selected as project advisors during Stage 1's workshop from the perspectives of PSW, dining lead, LTC RD, food and nutrition consultant, and food/dietary aide. By word of mouth, two new project advisors (Director and Assistant Director of Food Services) requested inclusion as observers for a total of seven advisors. Two technical experts with backgrounds in systems design engineering and limited exposure to the users' perspectives were recruited through word of mouth for completing the Ravden Checklist. An additional 13 RDs/directors/assistant directors of food services were recruited to provide usability data who were not previously involved with the design process as described in Section 5.2.3.

### **5.2.1 Interview Walk-through of Prototypes**

All testing was completed in-person though one-on-one sessions. Testing sessions were audio-recorded and relevant quotes were transcribed verbatim. Testing began with an interview walk-through of the prototypes based on script adapted from [126] to ascertain usability and feasibility barriers. A novel pre-defined strict set of tasks was completed by each advisor. I completed a checklist to capture the degree of success to which each task was completed (i.e., success, required prompting, or failed).

## 5.2.2 Usability Assessment Tools

Prototypes were evaluated by comparing perceptions of the AFINI-T prototype to the system currently in place regarding usability (user and technical expert), workload, and trust as follows:

1. **Users' Subjective usability Scale (SUS)** [32]: The SUS was selected over other usability questionnaires for its ease of use, minimal training requirements, and low application time [216, 217]. By word of mouth, two new project advisors (Director and Assistant Director of Food Services) requested inclusion as observers for a total of seven advisors.
2. **Technical Expert's Subjective Usability (Ravden Checklist)**: For evaluating usability more formally, an adapted Ravden checklist [196] was used by two technical experts with backgrounds in systems design engineering and limited exposure to the users' perspectives. The Ravden checklist was selected for its low-cost and ease of use to assess the interface with good inter-rater reliability and predictive validity [216, 217]. Items pertaining to help, including all of section 9, were removed as this was beyond the scope of the Goldilocks quality horizontal prototype.
3. **Users' Subjective workload (RTLX)**: The RTLX [88, 87] was administered to enable comparison of perceived workload of the current method in place with the AFINI-T system prototype (Table 3). Perceived workload of the current system was retrospectively evaluated with the Raw Task Load Index (RTLX) [88, 87] for its application simplicity and comparability to the NASA-TLX [87, 216, 149, 230]. For evaluating perceived workload of the AFINI-T prototype, four project advisors from Stage 4 were tester participants (PSW, Dining Lead, Dietary Aide, and Nutrition Research Expert). By word of mouth, two new project advisors (Director and Assistant Director of Food Services) requested inclusion as observers for a total of six advisors.
4. **Users' Perceptions of Trust** (adapted from [111]): Ethical considerations around users' perceptions of trust in technology is a crucial but often neglected aspect when conducting computer vision research in areas that can have a major impact on societal well-being. A subset of Jian et al. [111]'s tool was used to capture perceptions related to deception, wariness, confidence, dependability, reliability, trust and familiarity with the system [100]. Statements were comprised of 7-point Likert scales ranging from not at all (1) to extremely (7). Responses were re-categorized from a 7-point Likert scale to "No", "Neutral", and "Yes" to summarize trends; the original 7-point Likert scale ratings were used for calculating a two-tailed t-test assuming unequal variances [75, 172] to compare the existing electronic paper-based system and the AFINI-T prototype.

### **5.2.3 Webinar Feedback**

After testing was completed, on a separate occasion new RDs, directors and assistant directors of food services from across the Schlegel Villages were invited to participate in a webinar outlining the progress to date along with tandem survey completion for assessing perceived usability and workload. 13 people participated in the webinar (43% participation rate), which is consistent with typical attendance of quarterly dietitian meetings at Schlegel Villages due to scheduling complexities.

## **5.3 Analyses**

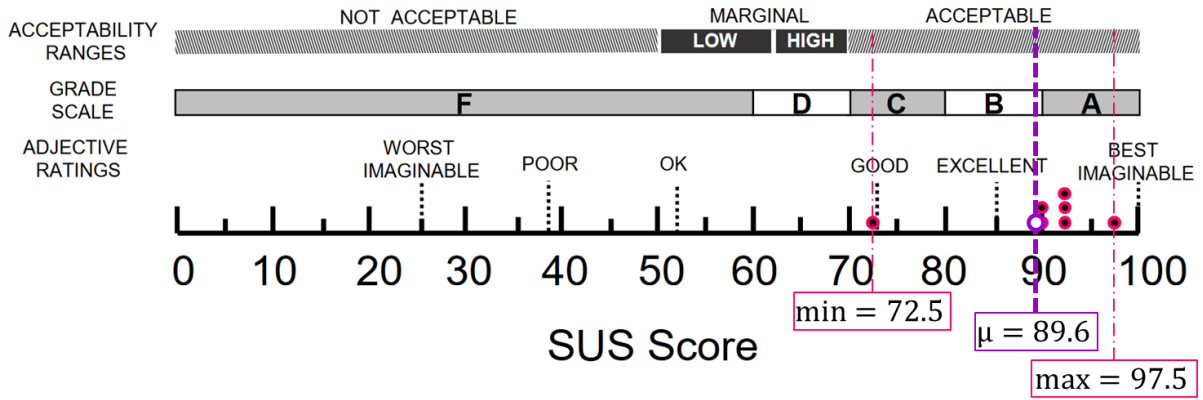
For scales with five or more categories (e.g., RTLX),  $\mu(\sigma)$  are used; the mode was used for categorical data with fewer than five categories (e.g., Ravden Checklist). A two-tailed t-test assuming unequal variances [75, 172], was conducted to compare the current system and the AFINI-T system for users' perceived workload for the RTLX. Quantitative data were analysed using descriptive statistics, with highlights from qualitative data as described in Stage 1.

## **5.4 Results**

### **5.4.1 Users' Subjective Usability Scores**

Subjective usability was rated as “acceptable” with average SUS scores of 89.2 and 78.2 translating to a B+ or C+ on the grade scales for project advisors and webinar participants, respectively. Mapping these scores onto the adjective ratings as described by [16, 17], the majority of usability scores therefore fall between “excellent” and “best imaginable” for project advisors (5/6) and “okay” to “good” for webinar participants (5/9) as shown in Figure 5.1.

**a) Project Advisors (n=7)**



**b) Webinar RD Participants (n=9)**

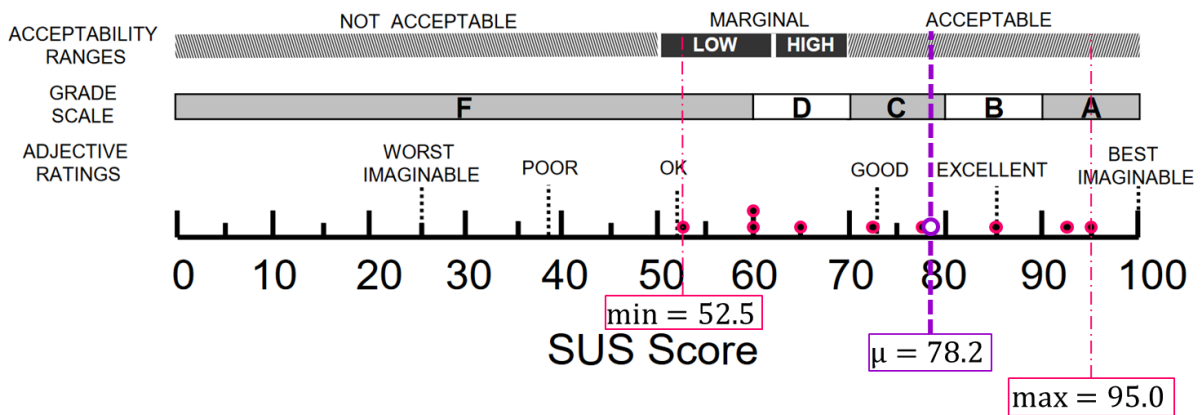


Figure 5.1: Overview of subjective usability ratings by project advisors (a), and webinar participants (b).

In line with these quantitative results, project advisor users commented that: “It’s quite intuitive, the key things were easily found”, “It’s a lot but it’s easy to learn and it’s colourful”, “I’m not technologically inclined, but most things I was able to do intuitively”, “I think someone could use this if they were just thrown onto the floor with it.”

**5.4.2 Technical Expert’s Subjective Usability in Context**

Two technical experts completed a modified Ravden usability checklist evaluation with favourable ratings (Table 5.1). Ratings across both raters for sections 1-8 were very satisfactory (7/8 sec-



tions) or split between “satisfactory” and “very satisfactory” (1/8 sections) and mode for section 10 on system usability of “no problems”. Consistent with comments from user testing, the main suggested area for improvement was to increase customisability options (e.g., sort resident list in multiple ways, allow more flexibility in the order of operations such as allow charting before a picture is taken).

An additional comment received pertained to “the order of operations” causing minor problems. The AFINI-T prototype was operating on the assumption that a before picture would be taken before selecting food items. This may not be the case, and because of this feedback, we can incorporate this additional flexibility moving forwards. Another example of reported minor problems was that the “system response times were too quick for you to understand what is going on”, which is consistent with the user testing as there were several instances in which a prompt was needed to observe a change in state (e.g., input fluids, and not only does the fluid intake progress bar update, so too does the nutrient intake summary). Both project advisors and technical experts suggested highlighting these changes with colour; this should be incorporated moving forwards.

### **5.4.3 Users’ Subjective Workload (RTLX)**

As highlighted in Table 5.2, performance was rated comparably with average score of 16.8 and 15.2 for the AFINI-T and current systems, respectively. In the case of mental demand, time demand, effort and frustration, subjective workload ratings were significantly lower for the AFINI-T system than the current system ( $p < 0.05$ ). This suggests the AFINI-T system is perceived to require less effort and lower overall workload than the current system and is consistent with comments from the participants including: “*[This would take a] huge burden off me as a clinician. This is hugely better than paper... there are no guesstimates... I don’t have to do work.*”, and “*It makes life so much easier*”.

Since the focus for performance was more on workflow and process, the emphasis for either system was on task completion, not accuracy of task completion. For the AFINI-T system, when accuracy was considered (i.e., the number of correct task completions without any prompting were tallied and divided by the number of valid tasks), the average accuracy was 88%. A rating of 16.8/20 equates to around 84% accuracy which indicates a very close mapping between perceived performance and task completion accuracy; however, this may be spurious. These results are presented only as a discussion piece as assessing accuracy of data entry is beyond the scope of this study that aimed to address high-level interface design. Entry errors for data accuracy would need to be assessed with the fully functional system as part of future work.

Table 5.1: A summary of the Ravden usability checklist evaluation conducted by two technical experts; section 9 was removed as it was not applicable to this version of the prototype.

Section	Mode Rating	Expert 1 % of valid “Always” ratings (n/N)	Expert 1 % of valid “Most of the time” ratings (n/N)	Expert 2 % of valid “Always” ratings (n/N)	Expert 2 % of valid “Most of the time” ratings (n/N)
Section 1: VISUAL CLARITY	Very Satisfactory	73% (11/15)	27% (4/15)	50% (7/14)	50% (7/14)
Section 2: CONSISTENCY	Very Satisfactory	91% (10/11)	5% (1/11)	73% (8/11)	27% (3/11)
Section 3: COMPATIBILITY	Very Satisfactory	79% (11/14)	21% (3/14)	64% (9/14)	36% (5/14)
Section 4: INFORMATIVE FEEDBACK	Very Satisfactory	75% (9/12)	25% (3/12)	69% (9/13)	31% (4/13)
Section 5: EXPLICITNESS	Very Satisfactory	91% (10/11)	9% (1/11)	83% (10/12)	17% (2/12)
Section 6: APPROPRIATE FUNCTIONALITY	Very Satisfactory	100% (8/8)	0% (0/8)	88% (7/8)	13% (1/8)
Section 7: FLEXIBILITY AND CONTROL	Satisfactory/ Very Satisfactory	56% (5/9)	22% (2/9)	89% (8/9)	11% (1/9)
Section 8: ERROR PREVENTION AND CORRECTION	Very Satisfactory	89% (8/9)	11% (1/9)	100% (7/7)	0% (0/7)
Section 10: SYSTEM USABILITY PROBLEMS*	No Problems	71% (15/21)	29% (6/21)	81% (17/21)	19% (4/21)

\* Section 10 was reverse coded. Instead of % of valid “always” and “most of the time”, these columns refer to “% of valid no problems” and “% of valid minor problems”.

Table 5.2: Comparing retrospective perceived users' workload measures of current food/fluid intake system from Stage 1 to the AFINI-T prototype results from Stage 5. Values could take on a range from 0 to 20; 0 implies no workload, 20 implies highest imaginable workload except in the case of performance which is reverse coded.

Workload Measure	System	Mean	Mode(s)	Min	Max	Responses	t-Test
			n	n	n	(N)	t, df (P-value)
Mental demand	Current	10.2	6	4	19	10	2.56, df = 13.8 (P=0.023)
	AFINI-T	4.4	3	1	10	6	
Physical demand	Current	6.4	2	1	15	9	1.41, df = 12.5 (P=0.183)
	AFINI-T	3.5	1	1	6	6	
Time demand	Current	16.7	19	5	20	10	4.89, df = 10.8 (P<0.001)
	AFINI-T	5.5	3	1	12	6	
Performance	Current	15.2	18, 20	3	20	10	0.722, df = 13.7 (P=0.722)
	AFINI-T	16.8	20	11	20	6	
Effort	Current	13.2	6	6	20	10	5.55, df = 13.5 (P<0.001)
	AFINI-T	3.7	3	1	7	6	
Frustration	Current	11.5	15	1	20	10	3.80, df = 13.0 (P=0.002)
	AFINI-T	3	2	1	8	6	

#### 5.4.4 Users' Perceptions of Trust

Consistent with our previously reported high usability (SUS score of 89.2), and significantly higher perceived performance with the AFINI-T prototype than the existing system ( $P < .05$ ) in [180], Figure 5.2 indicates there is low trust in the existing system (55% of respondents do not trust the system,  $n=11$ ), the AFINI-T system was perceived to be more trustworthy, and generally, AFINI-T system trust ratings were opposite and more positive compared to the existing system for food and fluid intake charting. For example, advisors rated the AFINI-T system as less deceptive (deceptive "yes": 17% AFINI-T, 45% existing system), less wary of the system (wariness "yes": 0% AFINI-T, 50% existing system), and more confident in the AFINI-T system (confident "yes": 83% AFINI-T, 18% existing system). As shown in Figure 5.2 using a two-tailed t-test assuming unequal variances [75, 172] indicate that these three scores of deceptiveness, wariness, and confidence were significantly different between systems all with  $p$ -values  $< 0.05$ . While not statistically significant, there were additionally higher ratings for dependability, reliability, and familiarity with the AFINI-T system prototype compared to the existing system. The statement regarding "I can trust" the system was higher for AFINI-T system prototype than to the existing system and this difference approached significance ( $p=0.08$ ).

<b>THE SYSTEM IS DECEPTIVE. ('NO').</b>	<b>83%</b> AFINI-T	<b>VS</b>	<b>9%</b> EXISTING	t = 3.45, df = 11.2 p=.005
<b>I AM WARY OF THE SYSTEM ('YES').</b>	<b>0%</b> AFINI-T	<b>VS</b>	<b>50%</b> EXISTING	t = 3.34, df = 12.5 p=.006
<b>I AM CONFIDENT IN THE SYSTEM ('YES').</b>	<b>83%</b> AFINI-T	<b>VS</b>	<b>18%</b> EXISTING	t = 2.42, df = 8.03 p=.042
<b>THE SYSTEM IS DEPENDABLE ('YES').</b>	<b>83%</b> AFINI-T	<b>VS</b>	<b>30%</b> EXISTING	t = 1.55, df = 6.80 p=.166
<b>THE SYSTEM IS RELIABLE ('YES').</b>	<b>67%</b> AFINI-T	<b>VS</b>	<b>30%</b> EXISTING	t = 1.07, df = 7.83 p=.319
<b>I CAN TRUST THE SYSTEM ('YES').</b>	<b>83%</b> AFINI-T	<b>VS</b>	<b>18%</b> EXISTING	t = 2.00, df = 8.29 p=.080
<b>I AM FAMILIAR WITH THE SYSTEM ('YES').</b>	<b>55%</b> AFINI-T	<b>VS</b>	<b>100%</b> EXISTING	t = 1.31, df = 14.9 p=.210

Figure 5.2: Advisors’ perceived trust of existing food/fluid intake system (Existing) to the AFINI-T prototype (AFINI-T). 7-point Likert scale ratings were condensed to “No” for Likert ratings 1-3, “neutral” for a rating of 4, and “yes” for ratings 5-7.

### 5.4.5 Qualitative Feedback

Project advisors’ receptivity of the AFINI-T system prototype was positive with several areas identified for improvement. For example, regarding the general concept for the dietitian interface: “[It] would be good to personalize these specific needs and set it so the flags sent to nursing/PSW for these items based on what dietitian enters . . . This would save a lot of time especially if individualized.”, “Capturing [supplement intake] would enable dietitians to monitor intervention adherence . . . If it shows up that they never have it, then great feedback to change the intervention.”

Receptivity of webinar participants was moderate. The main reservation pertained to how the system would integrate with the current method and PointClickCare (corroborated also in Chapter 3), and workflow more generally. For example, three webinar participant’s direct messages were as follows: (1) “I love the idea of this system, we are concerned about workload, as well as if the systems (AFINI-T and PCC) talk to each other.”, (2) “Would this be a separate system that

would be linked to PCC?”, and (3) “I hope a PCC progress note is generated from any notes [a registered dietitian] adds”.

Additionally, webinar participants expressed reservations regarding the proposed AFINI-T system. One dietitian expressed concern about overemphasizing the importance of nutrition “*in a population that should have the main focus of just making sure [residents] are enjoying the food we are serving*”. There was also concern over how this will translate to Ontario Ministry of Health and Long-Term Care (MOHLTC) inspectors’ inspections and the perception that using a system like this will take more time. Additionally, it was stated that there was no perceived value to having access to more detailed nutrient data in the LTC population as, to them, the largest issue contributing to malnutrition is the impact dementia has on calories consumed. However, they did suggest that if there was an ability to screen for residents to focus on only those at greater risk for malnutrition that the AFINI-T system would be helpful while still meeting the MOHLTC standards since only those at risk for malnutrition are mandated to track food and fluid intake. This provides an interesting complementary perspective and warrants further probing and discussion.

## **5.5 Discussion**

The overall purpose of this work was to continue to remove feasibility-related barriers to uptake and empower user-centred technology development through usability assessment and prototype validation.

While usability assessment and prototype evaluation informed the design process, one specific opportunity for further enhancement was the opportunity for collecting webinar feedback. We conducted a hybrid webinar survey to connect during a quarterly dietitian meeting. The concept of the AFINI-T system was completely new to most participants which made it difficult to build rapport with this group. However, we believe at this stage of the design process this was a strength; this may have helped participants to provide candid, objective feedback. That said, there were several examples of difficulty in keeping webinar participants engaged. For example, the webinar was run with a brief adjournment for completion of a survey that was then used to encourage group discussion. The ability to take a poll during the webinar may have been more effective at keeping engagement. In addition, the method by which participants attended was inconsistent across locations. For example, most participants joined individually, however at venues where multiple participants joined from one location (e.g., RD, director, and assistant director of food services), they filled out the corresponding survey together as well. This may have resulted in bias in some of the feedback collected but also enabled conversation and collaborative thought. Given the exploratory, qualitative nature of the feedback received during this stage, it does not undermine the results of previous stages, and for Stage 6, may have resulted in

more critical appraisal from potential group discussion.

In terms of time requirements and concerns raised by webinar participants, this is valid and is a next step. When the fully functional prototype is developed, it will be important to evaluate task completion time. Even if the AFINI-T system requires a comparable amount of time, it will yield a trove of powerful nutritional insights so direct comparison of approaches may be more complex than a simple timed trial. Through this approach, the AFINI-T system may support care givers' efforts in promoting enjoyment of food consumed for residents with communication changes as part of living the dementia journey. Within the scientific community context, additionally, the proposed AFINI-T system may enable knowledge discovery through a thorough automated approach to understanding dietary patterns in the LTC context and beyond.

We were particularly interested in trust as it pertains to the ability to introduce a new level of automation in this field by leveraging computer vision techniques. Several factors described by others that influence trust relevant to developing computer-vision-based applications are: the type of system, how complex the system is, how the system will be used, and the cost-benefit of using the system (e.g., high risk or low risk) [100]. However, in the case of the AFINI-T system, we must be careful since this is a tool to support care in a vulnerable setting (older adults). As such, it must be clear that decision making rests with humans and reinforce that the AFINI-T system is a decision-making aid. Specifically, when automation comes with great benefit and fewer risks, people tend to increase reliance on automation; even under high-risk scenarios if the level of automation is low [100, 150] and humans tend to view automated tools as more accurate than humans [150]. This contributes to over-trust [138] and may be more important to consider than initially thought given readiness to accept the prototype-level technology and high trust ratings for even for the prototype that is not fully developed.

From an uptake of technology perspective, the landscape for AFINI-T appears to be promising when reflecting on trust ratings and usability ratings, and several design considerations that were incorporated that otherwise would have been missed if not for the diverse set of collaborators. It is also encouraging that project advisors were anecdotally engaged with frequent comments of, *"I'm so excited!"* and *"I just want to play!"* Regarding three wishes to improve, project advisors commented: *"I just want to use it now"* and *"[I would use food-specific info to] identify trends with people. Try different plates, see if it changed the intake pattern. What do individual residents like and what don't they like. If they eat a lot, is it because they like the food, or was it a one-off? Having data accessible opens the door to how can be analyzed."* When considered together, output from these various phases suggest that project advisors are keen to try a technology like this and that we're on the right track for developing the AFINI-T system; users seem to be engaging with it, enjoying it, and see it as value-added.

Particularly in computer vision research building intelligent systems with end-users engaged,

we should be aware of how the perception of the technology changes through this process as suggested through this work. Specifically, we must be wary of potential over-trust in the system, which is when “trust exceeds system capabilities” [138]. Of interest in this application, the risk for over-trust is higher when building tools that will be used in high stress environments due to time constraints and have potentially large benefits over the existing method (e.g., time saving, improved quality of care) [100]. Regarding the receptivity to the AFINI-T system, advisors commented *“This would save a lot of time especially if individualized. [The % daily value of nutrients is a] proportional calculation [that is currently] a manual process.”* In contrast, perhaps contributing to relatively low trust in the existing system, several users articulated that the quality of the existing method for data collection is not helpful at prioritizing resident referrals to dietitians. When considered in the LTC environment where being short-staffed is the norm, this may in part explain why trust ratings in the AFINI-T prototype were high and highlights the importance of these design considerations.

This participatory iterative design methodology provided a more holistic approach to developing a solution. We took this approach to work directly with representative end-users to understand and incorporate their perspective and concerns to consider and appropriately support trust from multiple angles. By doing so, we aimed to support trust cues, credibility [52, 76], and to adhere to best practices for user interface design [105, 14, 114, 170, 171, 210]. This translated to incorporating ease of navigation (reinforced by click-saving features), use of good visual design elements by on-screen chunking of relevant information, aiming for an overall professional look, supporting search ability as well as smart guidance through transactions and overall, working directly with users to provide appropriate and useful content. This is corroborated with qualitative feedback as well. In speaking with one project advisor, they were impressed with what was captured. Specifically, the consideration to have the reminder that dietitians need consent from residents to make any modifications to their nutritional goals. They asked, *“How did you know to include that?”* When it was explained that it came from discussion with other project advisors they said, *“You really did your homework.”* This is one example of the powerful synergy of incorporating multidisciplinary perspectives when users are included as collaborators. Instead of being limited to domain knowledge in one field, this transdisciplinary approach allows for a more informed and feasible solution.

Between workshop participants and project advisors (Chapter 3), 27 unique collaborators representing 15 different roles were engaged in this participatory iterative design process. This sample size is consistent with recent analogous health-care-related user-centered design as well as usability and feasibility studies [50–58] [59, 82, 101, 109, 125, 131, 167, 195, 199] with sample sizes ranging from five as in [125] to 32 as in [199]. Between 11 and 13 additional participants were involved in the webinar exercise and contributed to nine survey responses (several individuals filled out a response together) and are described in Chapter 5. Therefore, the total sample size ranged

between 35-40, however, not all collaborators contributed to every aspect of the process (e.g., user testing in Stage 5 was comprised of a subsample of 6 individuals). While this sample size is consistent with early pilot project prototyping [37, 59, 82, 101, 109, 125, 131, 167, 195, 199], generalizability remains unclear. As the team of project advisors was relatively small and from the same organization, it will be important for the final product to be tested with a larger sample of users to make sure concepts captured generalize well to users' needs more broadly.

Our data collection strategy was grounded in theory, guided by several conceptual frameworks, and grounded expertise to complement the interdisciplinary and complexity of the problem space (e.g., [24, 37]). We also borrowed from transdisciplinary research that “explicitly recognizes the value of partnerships and the different stakeholders along with their roles in facilitating and supporting innovation” [24]. In the AFINI-T design this was reflected through recruiting diverse and multidisciplinary project advisors. That said, no demographic information was collected; this should be considered moving forward especially when recruiting for a larger sample for user testing. A larger sample size for the final prototype will help deepen our understanding of usability. Finally, given the stage of this research, qualitative analyses were limited to extracting overarching themes across sources; an additional avenue for future work, pending completion of a high-fidelity prototype is to conduct a more thorough qualitative analysis vetted in an evaluation framework (e.g., grounded theory or narrative content analysis) alongside prototype testing and evaluation.

## 5.6 Conclusion

While fast-paced and time intensive, the results gathered in this study indicate it *is* possible to design AI/ML/computer vision applications with end-users and can result in high receptivity and trust in these technologies. Our experience corroborates that engaging with end-users throughout the process as collaborators enhanced a “more comprehensive understanding of the problem space” [24]. Moving forward, we must re-assess trust in the fully functional system and probe more deeply into factors contributing to trustworthiness. More generally, as designers and researchers developing accountable computer vision systems, we must be the first line of defense and should consider our moral obligation for ensuring accuracy, system reliability, and understanding the psychological effects of using trust cues to enhance usability. Furthermore, output from these various stages suggest that while careful consideration for integration with the PointClickCare system is needed. Additional consideration for policy expectations are required but project advisors have interest to try a technology like this if these concerns are addressed. Advisors seem to be engaging with the AFINI-T prototype, receptive to the idea, and enjoying it. This modified participatory iterative design sprint was effective at understanding the problem space, making informed design decisions (Chapter 3), and evaluating receptivity to a novel pro-



prototype all within a compressed period of time (i.e., 6 weeks). Future directions for the AFINI-T system should include incorporation of learnings from this process, and the development of a fully working prototype for further user testing and evaluation with an emphasis on timed trials for assessing feasibility in the LTC environment.

### 5.6.1 Key Contributions

- Documented user acceptability.
- Reinforcement that the AFINI-T user interface is usable at right level of difficulty.
- Evidence supporting the AFINI-T system is primed for success based on trust cues and usability.

#### UP NEXT...

We have now seen priorities from end-users in LTC in Chapter 3, re-imagined what food intake might look like, evaluated our proposed solution in AFINI-T in Chapter 4, and showed high acceptability of this approach in this chapter. But we still haven't taken into consideration the error-prone nature of nutritional databases. In many cases, they are the best we can do without sending samples away to labs which is resource-intensive in terms of both time and money. My dream would be to be able to make some sort of at-the-plate inferences based on food biophotonic principles to "nudge" estimates closer to the truth especially for the end-application of conducting nutritional intervention research beyond the LTC perspective. While this is largely outside the scope of this thesis, Chapter 6, provides some encouraging results suggesting this may be within reach as a future direction.

## Chapter 6

# Food Biophotonics: Optical Imaging for Estimating Human-Observable Nutrient Properties

So far, we've seen how linking food images, volume estimation and nutritional information can yield a powerful system for objectively estimating food intake with accuracy and precision that outperforms current LTC standards. We assessed system accuracy by disentangling segmentation/volume estimation from the classification and combining it with nutritional information through planned menu recipes (which rely on databases). But we also know that food databases are notoriously missing information and there is also the issue of inter-chef variability in preparing foods. What if we could account for some of these shortcomings or differences by leveraging computational visual cues to nudge estimates towards a more accurate solution? In this chapter I seek to address my thesis C3, to explore feasibility of food database enhancement. While validation for a system like this is well outside the scope of this thesis, I conducted foundational investigations to probe the utility of computational relative nutritional density and the role of colour in investigating nutrient content for the nutrient families of vitamin A and antioxidants (anthocyanins). To explore this, I chose to use commercially prepared puréed foods for two reasons: (1) from an application perspective, the prevalence of modified texture diets and puréed foods in LTC is very high, and (2) commercially prepared puréed foods are well documented and have more uniform texture consistency to simplify the food scenario.

The remainder of this chapter outlines my work (1) describing a feasibility assessment of an image-based computational nutritional density analysis system [181], (2) assessing of the correlation between 13 commercially prepared puréed food samples' visible spectrum absorption profiles and nutritional composition across a flavour-stratified 5-tiered, 6-fold dilution series [183],

**This chapter contains content previously published from...**

**KJ Pfisterer**, J Boger, A Wong. Prototyping the Automated Food Imaging and Nutrient Intake Tracking (AFINI-T) system: A modified participatory iterative design sprint. *JMIR Human Factors* 2019;6(2):e13017. doi: <http://dx.doi.org/10.2196/13017>. On this paper, K.J.P. was the main contributor from project inception to planning, implementation, data collection, analysis, interpretation, and writing.

**KJ Pfisterer**, J Boger, A Wong. Food for thought: Ethical considerations of user trust in computer vision. *CVPR2019 Fairness Accountability Transparency and Ethics in Computer Vision Workshop*, Long Beach, United States. On this paper, K.J.P. was the main contributor from project inception to planning, implementation, data collection, analysis, interpretation, and writing.

**KJ Pfisterer**, R Amelard, AG Chung, A Wong. A new take on measuring relative nutritional density: The feasibility of using a deep neural network to assess commercially-prepared pureed food concentrations. *Journal of Food Engineering* 2018; 220(223). <https://doi.org/10.1016/j.jfoodeng.2017.10.016> On this paper, K.J.P. was the main contributor to experimental design, data acquisition protocols, technical implementation, data analysis, interpretation, and writing of this manuscript.

**KJ Pfisterer**, R Amelard, A Wong. "Differential color space analysis for investigating nutrient content in a pureed food dilution-flavor matrix: A step toward objective malnutrition risk assessment". Proc. SPIE Volume 10501, Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics; 105010J (2018) <https://doi.org/10.1117/12.2289028> This work, which built on one aspect of the previous paper. On this paper, K.J.P. was the main contributor to experimental design, data acquisition protocols, data analysis, interpretation, and writing of this manuscript. Technical implementation support was provided by R.A.

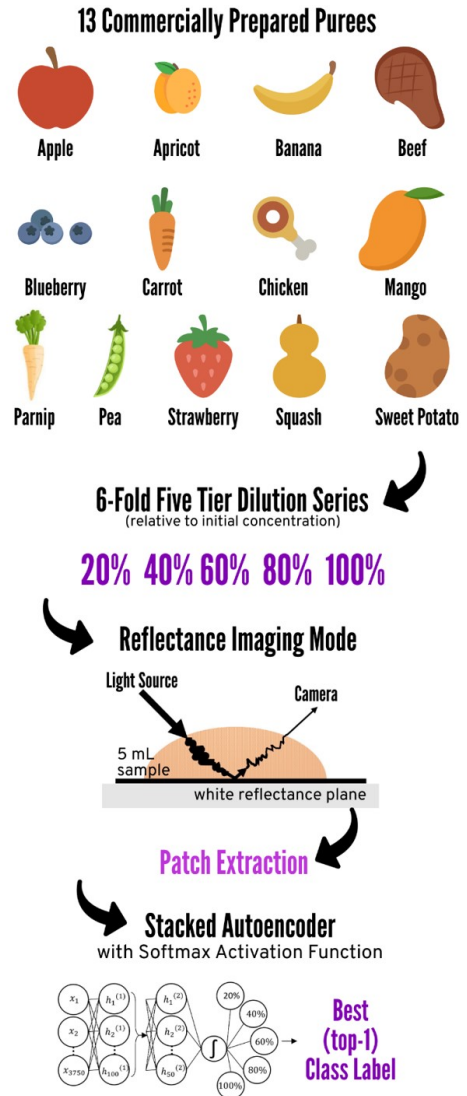
**K. Pfisterer**, R. Amelard, B. Syrnyk and A. Wong, "Towards Computer Vision Powered Color-Nutrient Assessment of Puréed Food," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 2019 pp. 490-492. doi: 10.1109/CVPRW.2019.00068 On this paper, K.J.P. was the main contributor to experimental design, data acquisition protocols, data analysis, interpretation, and writing of this manuscript. Technical implementation support was provided by R.A.

and (3) exploring sweet potato and carrot as proof-of-concept for optical links to vitamin A [182], a family of chromophores including carotenoids that are optically active in the visible spectrum [159].

## 6.1 A New Take on Measuring Relative Nutrient Density

This dilutions study is motivated by the end-goal of nutrient density assessment. Using relative water concentration to initial concentration (i.e., pure commercially prepared product), we prepared a dilution series to observe the effect of relative increased water content on optical properties (colour information, texture information, saturation etc.) for the purpose of determining the feasibility of using an optical imaging technique for discrimination. We used stacked autoencoders with a final softmax layer for dilution classification (i.e., discriminating between 20%, 40%, 60%, 80%, 100% initial concentration) as depicted in Figure 6.1.

To accomplish the task of learning what each flavour’s dilution looks like, I leveraged autoencoders. Autoencoders are DNNs that use unsupervised learning to provide a robust solution that is generalizable and extensible without compromising performance to complete a specific task and therefore circumvents the unlabelled-data problem. This study assesses the feasibility of using machine learning (i.e., DNNs) to automatically predict the concentration (as a proxy for nutrient density) of commercially-prepared purées. Furthermore, the use of DNNs for this task is motivated by the results of a theoretical optical dilution model. Since neural networks are biologically inspired machine learning methods and since in practice, food and food quality are often visually assessed, a theoretical optical validation of perceptually quantifiable nutrition composition can provide strong support for using machine learning. For example, passing input, such as a hypothetical concentration into a theoretical model, would yield an ideal output similar to the perception of the human eye. This present study, involving visible spectrum imaging data at different polarizations, provides a novel application of image classification to analyze thirteen types of commercially-prepared purées across three food categories (fruit, meat, vegetables) at five dilutions relative to initial concentration [181].



## **6.1.1 Methods**

### **6.1.1.1 Sample Preparation**

A six-fold five-tier dilution series for each of thirteen commercially-prepared purée flavours were prepared relative to initial concentration (20%, 40%, 60%, 80%, 100% initial concentration). The specific purée flavours selected were: apple, apricot, banana, beef, blueberry, carrot, chicken, mango, parsnip, pea, strawberry, squash (butternut), sweet potato. While the intended application is LTC, infant purées were selected to provide single food samples to avoid confounding of food blends' complex absorption profiles. At each of the relative dilutions, six 5 mL replicates of the flavour sample were loaded at an approximate height of 1cm from a standardized transparency sheet grid placed over a white reflectance plane and were imaged immediately. This yielded a total of 390 samples.

### **6.1.1.2 Data Acquisition**

Same-side reflectance was used (i.e., the light source and camera were positioned at the same location). The samples were illuminated using a broadband tungsten-halogen source with a front glass fabric diffuser for even illumination. A DSLR camera (Canon T4i) was used for high resolution image capture in the visible spectrum with consistent white balancing, aperture, and exposure settings. Vertically polarized images were acquired to maximize the subsurface characteristics, since purée samples were approximately horizontally planar. The samples were loaded onto a white reflectance plane to maximize reflectance. The room temperature varied during imaging from 21.9°C to 23.9°C.

### **6.1.1.3 Sample Subimages**

Since neural networks are biologically inspired and food consistency is presently visually inspected, it may be helpful to describe the data in terms of tangible features such as colour and texture. It is important to note that colour and texture are meant only to provide intuition into the data collected and, for our proposed system, were not used as hand-crafted features; features used for distinguishing between classes (classification) were automatically learned given no priors through the DNN (see Section 6.1.1.5 for more details). Figure 6.7 provides a summary of colour and texture across the samples. The images in Figure 6.1 were acquired from the sixth sample location on the sheet. To minimize glare the horizontal polarization of entire sample subimages were selected to provide further context with an ISO 100 and exposure 1/20 s.

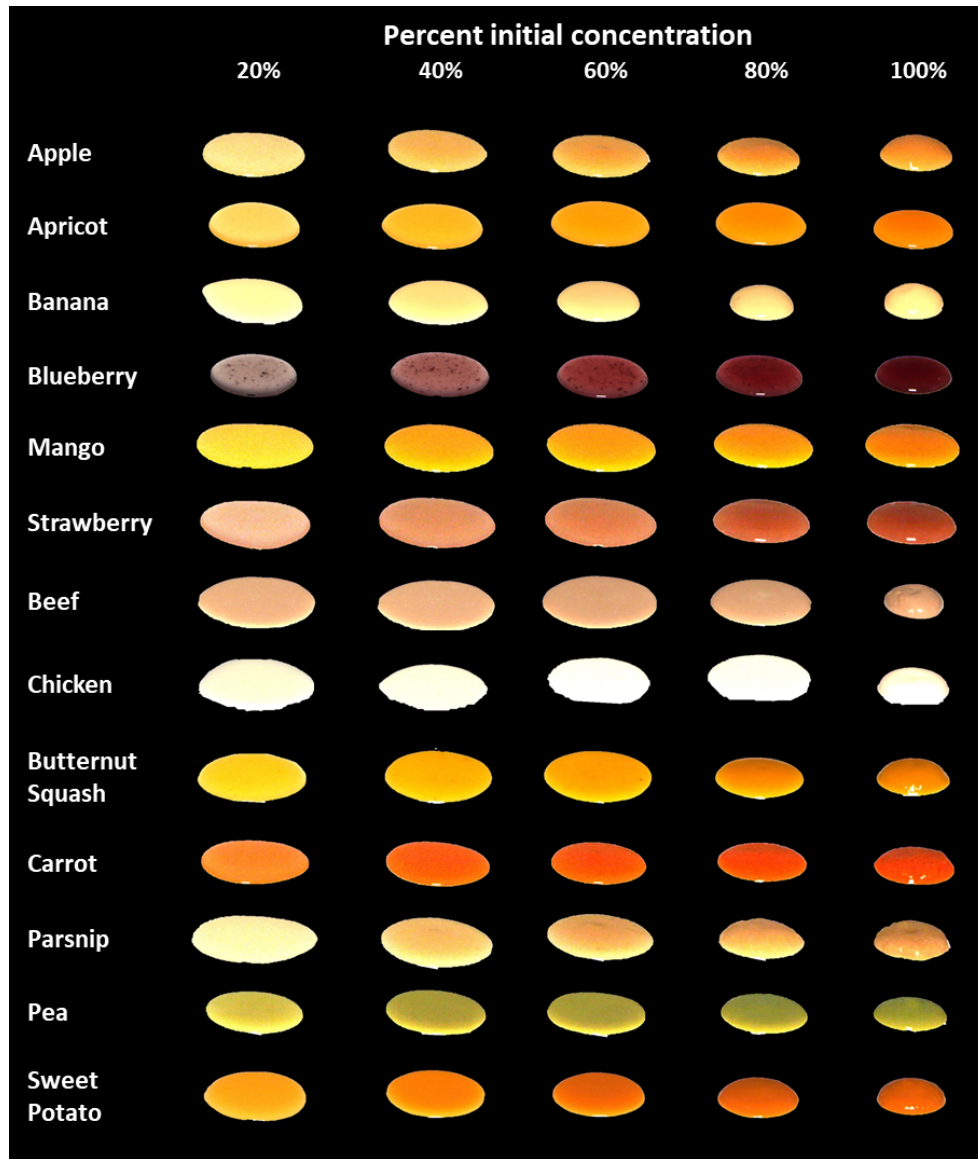


Figure 6.1: Dilution series example showing sample horizontal polarized subimages taken with ISO 100, EXP 1/20 s. Note the visible variations across dilutions for the best performing purée flavour, blueberry, with the stacked autoencoders and softmax layer achieved average accuracy of 99.6%; and poorest performing purée flavour, chicken, achieved an average accuracy of 73.3%. Here, chicken has an indistinguishable colour and lack of texture consistent across the subimages at 60% and 80% as they appear nearly absent or completely saturated.

#### 6.1.1.4 Training Data Set-up

Images were processed and data were analyzed using Mathworks' MATLAB version R2016b. Each image was white normalized by selecting a reference white rectangle from an in-frame white reflectance target. All images were labeled and deconstructed into six,  $100 \times 200$  pixel subimages (one for each sample on the sheet). As indicated in Figure 6.2, each three channel (i.e., RGB) subimage was decomposed into fifty-four patches using half overlapping windows of  $50 \times 100$  pixels. Rectangular patches were selected to improve the variance observed within a patch. These patches were downscaled to 50% of their original size ( $25 \times 50$ ) using bicubic interpolation. The three RGB channels were concatenated to yield 378,  $25 \times 50 \times 3$  (or  $75 \times 50$ ) pixel patches for processing for a given dilution for a specific purée flavour and 1890 patches for a specific purée flavour. Therefore, the final set of images consisted of 13,230 auto-labeled patches.

#### 6.1.1.5 Network Architecture

Images were then passed into a deep neural network (DNN), consisting of two layers of pre-trained stacked autoencoders and a final softmax activation layer. Unlike in Section 4.2.1.5, we did not use a convolutional neural network because the data we collected was uniform across a food blob so position provides little context. Instead, here, at a high level, five global, general networks were formed using randomly initiated weights and passing through all the unlabeled patches (i.e., there was no flavour or dilution information provided to the system). These general networks were then fine-tuned using flavour-specific labeled data. Given a specific flavour, the system predicted the dilution class to which a patch from an image belonged as shown in Figure 6.2. Here, we train the autoencoder, and then strip off the decoder so we can use the learned features from the encoder. We then take these features and use them as input into another autoencoder to further distill the feature set. After stripping off the second decoder, we pass the distilled features into an activation function for classification into one of five output classes (20%, 40%, 60%, 80%, or 100% initial concentration). By stacking these together, we form a general deep network where each autoencoder operates as a discrete feature extractor [98] and forces the network to learn higher level features. Then we fine-tuned the network by running a final iteration of backpropagation across the whole system. The result, a network that uses retained features that is capable of differentiating between classes [98] (i.e., inherent features that represent a concept such as the blueberry-ness of a specific concentration). From this global, general network, an additional step of fine-tuning was deployed for each flavour separately. This final fine-tuning step is the only iteration which uses the labeled, flavour-specific data. As a result, no hand-crafted features were used for the purpose of distinguishing between classes; all features were automatically discovered using the stacked autoencoders.

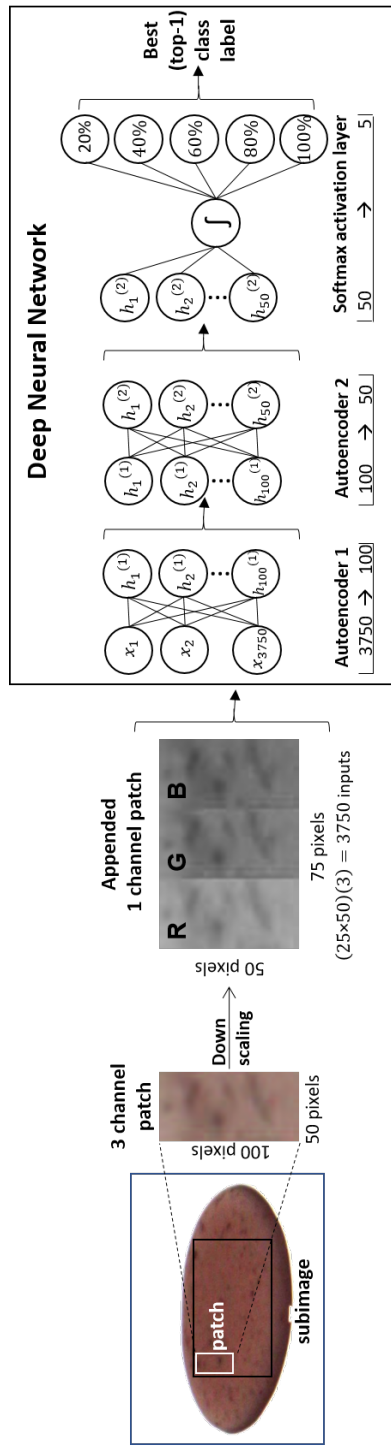


Figure 6.2: Deep neural network architecture. A subimage is decomposed into a series of patches which are then downscaled to half their original dimensions ( $100 \times 50 \rightarrow 50 \times 25$ ). RGB channels for each patch are concatenated to create one  $50 \times 75$  image per patch. Each patch is passed through two stacked autoencoders and a softmax layer to distinguish between classes (i.e., softmax classification layer). For a given flavour, the network output is one of five dilution classes (20%, 40%, 60%, 80%, and 100%).



#### 6.1.1.6 Validating our Network: Pretraining and Testing the Network

For each flavour-specific network,  $k$ -fold cross-validation was used by reserving one of the six positions for testing and completing training with the remaining five positions and conducting one final iteration of back propagation across the entire system for further fine-tuning of the weights. This was repeated five times for each flavour-specific network; specificity and sensitivity measures were averaged across each of the left-out positions. Accuracy of the network was assessed using confusion matrices whereby labels assigned by the network (i.e., observed class) were compared to the ground truth labels (i.e., expected class) and summarized by sensitivity and specificity for each class and across all classes for a given purée flavour.

#### 6.1.1.7 Feature Extraction

For comparative purposes, two methods were applied for feature extraction: 1) automatic extraction and learning of features by the second autoencoder, and 2) evaluation of hand-crafted features based on colour (64 features) and texture (seven features) characteristics [247]. The colour features were constructed using a discrete quantized colour histogram. Colour histograms are relatively invariant to rotation and translation, and coarse colour quantization encourages perceptual similarities through enlarged bin sizes. Environmental consistency (e.g., exposure time, white correction, illuminant spectrum, etc.) is important for such colour comparisons. A controlled optical setup was used to fix the relevant optical parameters, and is discussed further in Section 6.1.1.2. Specifically, 64 colour features were extracted by quantizing each colour channel into four bins, yielding  $4 \times 4 \times 4 = 64$  features. Given 64 colours, the number of pixels within a patch pertaining to each of the 64 colour bins was counted. Normalized histograms were used such that the value for each bin in the histogram represented the percent of pixels belonging to that colour bin. With regards to texture features, we used a set of texture descriptors based on differential translation histograms [247] after first converting colour images to grayscale. These histograms represent texture descriptors including mean, contrast, homogeneity, energy, variance, correlation, and entropy.

#### 6.1.1.8 Classification Methods

Features were extracted (either from each of five general networks' second autoencoders or from hand crafted features) and passed into one of three classification methods as shown in Figure 6.3.

More specifically, the classification methods were used to distinguish between dilution classes and making a prediction about to which dilution class a patch belonged. The three methods used were a softmax layer (tacked onto the end of our stacked autoencoder), and for comparative purposes, random forests [29], and support vector machines (SVM)(both linear and radial basis

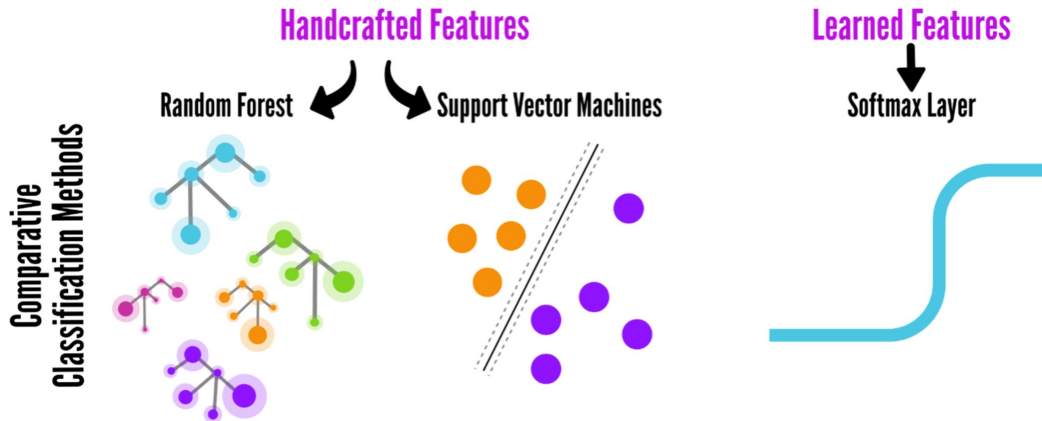


Figure 6.3: Overview of comparative classification methods.

function kernel) [53].

The softmax layer used the output features from the second autoencoder with the dilution labels to output the top-1 hit (i.e., the one class with the strongest prediction) class label. This method was applied to the features output from autoencoder 2 for each of the five general networks for the autoencoder based features only (i.e., hand crafted features were not fed into the softmax layer).  $k$ -fold cross-validation was applied to each iteration and results were averaged across the five networks. The main difference in implementing these methods is that the random forest and support vector machine approach require features to be provided, while the softmax approach relies on automatically learned and generated features.

### 6.1.1.9 Data Analyses

Descriptive analyses were summarized based on accuracy at predicting concentration for a given purée flavour from confusion matrices. Texture was summarized using entropy. Entropy is a rotation-invariant statistical measure of disorder, and thus was used to quantify texture variation similar to other food classification studies [26, 251]. In particular, local neighborhood entropy was used to assess the variation (or heterogeneity) of discrete image patches. Then, the spatial local neighborhood entropy distribution was used to summarize the texture of the entire image. Specifically, given a grayscale image, local texture was computed as a region-wise neighborhood entropy computation. The formulation we used resulted in low entropy for samples with smooth homogeneous texture containing little intensity variation over localized patches, and high en-

entropy for samples with inhomogeneous or heterogeneous rough texture contains highly varying intensity values. We used  $9 \times 9$  pixel neighborhoods, resulting in 81 pixel intensities to populate a distribution containing 256 bins.

Colour was summarized using the mean and standard deviation of red, green and blue values. Finally, saturation was summarized as a value between 0 and 1 where 1 represented totally saturated (white); saturation served as an indicator whether we could expect the system to work. If entropy was low and saturation was high, the data would represent pure white and may not contain discernible features upon which to correctly classify an image.

### 6.1.1.10 Optical Dilution Model

An optical dilution model was developed to motivate the use of deep neural networks for dilution classification. As a photon traverses through the purée sample, it undergoes a series of scattering and absorption events according to the constituent chromophores, resulting in the perceived colour. As the purée becomes diluted, the relative concentration of water increases while the photon path length stays relatively constant, leading to decreased overall absorption and thus changes in perceived colour. Representative perceptual image patches were derived from the mixture absorbance spectra by computing the perceived spectra colour according to CIE LMS cone responsivity curves [181].

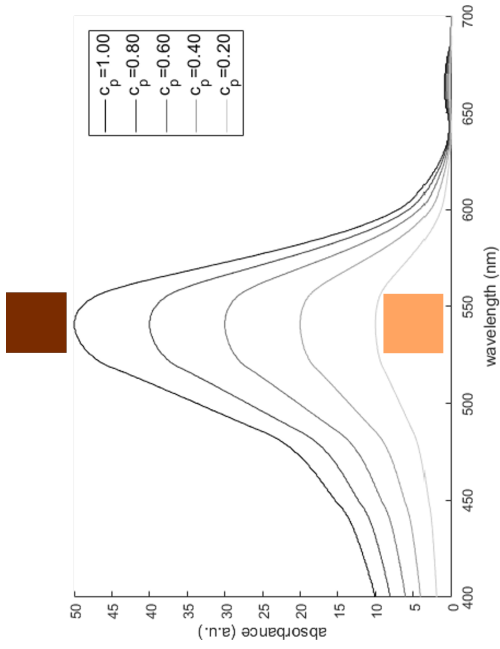
Mathematically, we can express this using the Beer-Lambert law of light attenuation to produce:

$$A = \log\left(\frac{I_0}{I}\right) = \epsilon_{H_2O} \cdot c_{H_2O} \cdot l_{H_2O} + \epsilon_p \cdot c_p \cdot l_p \quad (6.1)$$

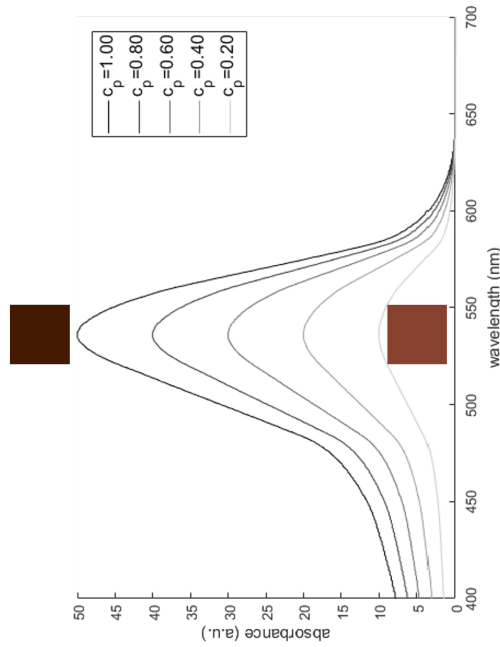
where  $A$  is absorbance,  $I_0$  and  $I$  are the incident and reflected illumination respectively,  $\epsilon_{H_2O}$  and  $\epsilon_p$  are the chromophore extinction coefficients for water and purée,  $c$  is the chromophore concentration, and  $l$  is the mean photon path length through the absorbing medium. Assuming a homogeneous dilution mixture ( $l_{H_2O} \approx l_p$ ), normalized incident illumination ( $I_0 = 1$ ), and normalized relative concentration ( $c_{H_2O} + c_p = 1$ ), this formulation simplifies to the following where  $l = l_p = l_{H_2O}$ :

$$A = -(\epsilon_{H_2O}(1 - c_p) + \epsilon_p c_p)l \quad (6.2)$$

Figure 6.4 illustrates what is mathematically described showing how relative changes in dilution impact on the spectral absorption response using two examples of purée samples (strawberry, and blueberry) based on published absorbance curves for blueberry and strawberry [221]. When samples are diluted with water, there are fewer absorption events and the spectral curve becomes more similar to pure water. Thus, a perceived increase in lightness of sample is observed as illustrated in the theoretical image patches in Figure 6.4.



(a) Mixture absorption of blueberry.



(b) Mixture absorption of strawberry.

Figure 6.4: Normalized absorbance spectra for (a) blueberry [221], and (a) strawberry [221]. Both (a) and (b) demonstrate the effect of increasing dilution (decreasing relative purée concentration) on purée samples’ spectral curves. Fewer photon absorption events are present in lower relative concentration resulting in a higher degree of reflection and lighter observed images (bottom patch) compared to the 100% initial concentration patches (top patch). Note that the optical dilution model models the degree of perceived “lightness” and “darkness”. An additional consideration is the effect of pH; were this to be reflected in the optical dilution model, we would also expect to observe a right shift of peak absorbance with lower, more acidic pH and a left shift of peak absorbance with higher, more basic pH [55].

## 6.1.2 Results

To understand the performance of the DNN for predicting purée sample concentration, results are organized as follows: (1) descriptive analyses of each image class in terms of colour, texture and saturation to provide context; (2) supporting evidence from the optical dilution model that dilution is quantifiable through perceptual data; (3) sample patches for each class across every purée flavour as a means to visualize and understand the underlying data; (4) an amalgamation of observations taken from confusion matrices to support accuracy of the system.

### 6.1.2.1 Descriptive Analyses

Figures 6.5 and 6.6 provide information about each class of images with respect to colour (mean R, G, B), texture (based on entropy, a statistical measure of variation) and saturation (where 0 is black and 1 is white). In terms of colour, with the exceptions of banana and chicken the colours appeared more vibrant as percent initial concentration increased. This is intuitive since the lower percent initial concentration (i.e., more diluted) samples contained more water than their higher initial concentration counterparts, resulting in texture and surface tension more similar to water than the pure purée. While the samples were all imaged using the same lighting conditions, camera settings, and were white corrected, there was a large range of saturation, texture (entropy), and RGB values. The samples most at risk for oversaturation were the 20% of initial concentration (IC) and the least at risk for oversaturation were the 100% IC. At both the 20% and 100% dilutions, the most saturated samples were chicken (saturation: 20% IC  $0.992 \pm 0.008$ , 100% IC  $0.988 \pm 0.015$ ) and the least saturated samples were blueberry (saturation: 20% IC  $0.538 \pm 0.071$ , 100% IC  $0.128 \pm 0.035$ ). With respect to texture (entropy), note the specks of blueberry seeds in blueberry, the smooth shininess of banana, beef and chicken, the more granular surface texture in the butternut squash, and the consistent, fine granularity across the sweet potato classes as shown in Figure 6.7. In terms of texture (entropy) the more diluted samples were more similar in appearance to water and aside from their colour, looked similar. Samples of lower dilution (more highly concentrated) tended to have higher entropy, however the most cohesive of samples (e.g., banana, beef, chicken, sweet potato) exhibited extremely smooth surface textures (i.e., lower entropy) across classes. This observation can be explained given that starches and proteins tend to form gels [6] as these were the products with the highest starch contents (sweet potato: 5 g/128 mL, banana: 3 g/128 mL) or protein contents (beef: 12 g/100 mL, chicken: 16 g/100 mL).

### 6.1.2.2 Optical Dilution Model

The optical dilution model was evaluated with a candidate purée, blueberry, to motivate the use of neural networks as purée concentration estimators. Figure 6.4 demonstrates how the absorbance spectrum changes according to the purée concentration (i.e., dilution) using published

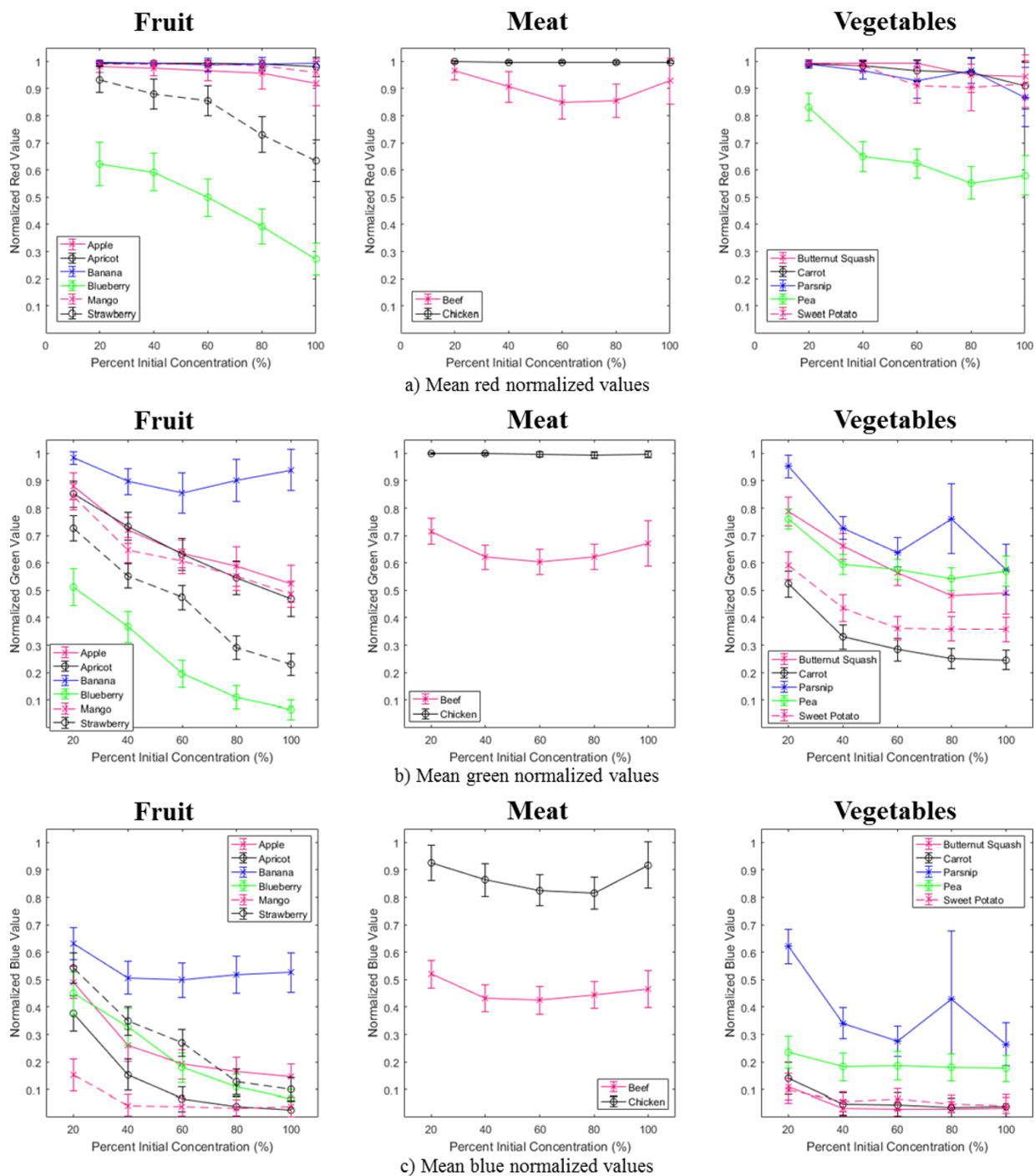
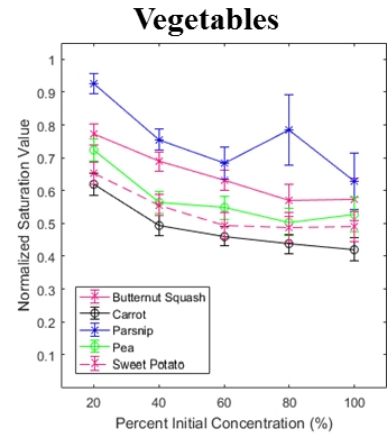
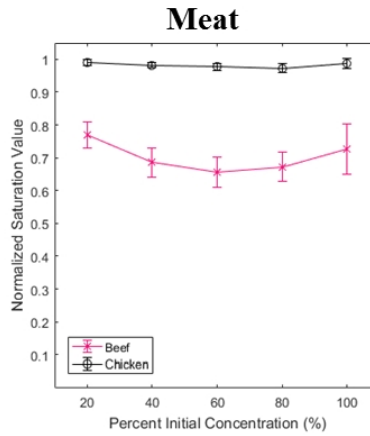
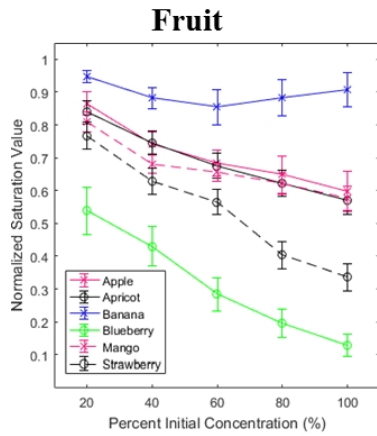
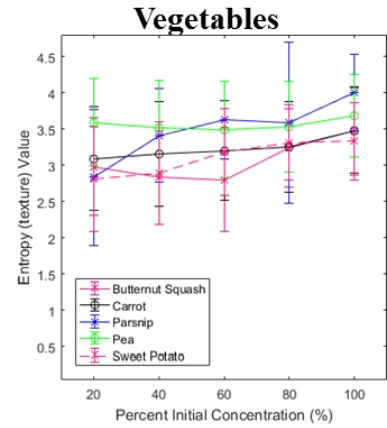
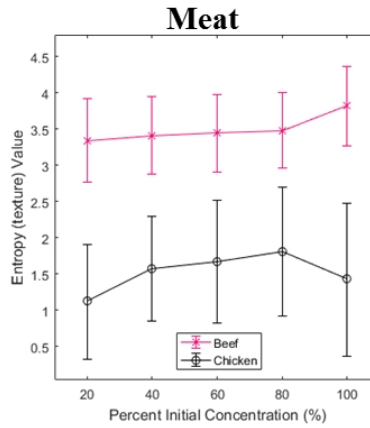
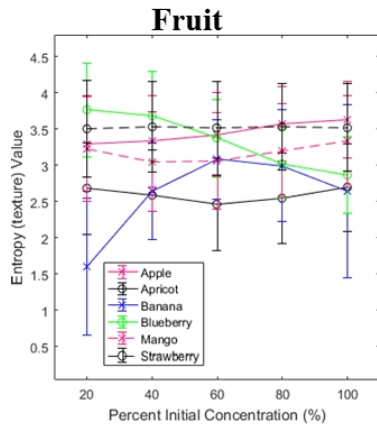


Figure 6.5: Descriptive analysis plots of purée flavours based on normalized RGB colour. RGB values have been normalized. Typically colour varied between the dilution classes of a purée flavour and were distinguishable between different purée flavours.



a) Mean saturation normalized values



b) Mean entropy (texture) values

Figure 6.6: Descriptive analysis plots of purée flavours based on saturation and texture (entropy). Saturation was normalized; entropy was used to describe texture. Typically, saturation and texture, varied across a purée flavour's dilution classes and were distinguishable between different purée flavours.

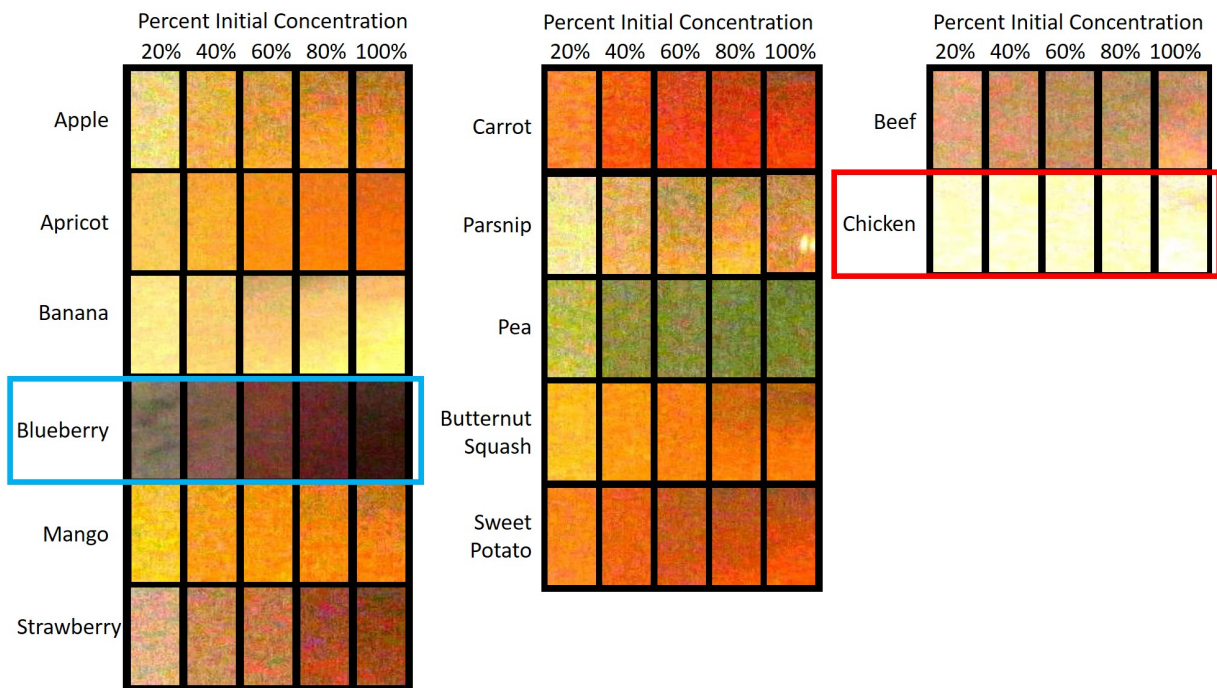


Figure 6.7: Sample patches for each purée flavour and dilution. Note the visible colour and texture variation across dilution classes and the indistinguishable nature of the poorest performing purée flavour highlighted in red. The sample blueberry patches outlined in blue had the best accuracy ( $99.6\% \pm 0.6\%$ ) with the autoencoders and softmax approach; these patches match the theoretical patches generated with the optical dilution model well (see Figure 6.4).



absorbance curves for water [200] and blueberry [221]. Blueberries contain anthocyanins which are pH sensitive chromophores which shift from red to blue with increasing pH. For example, pH = 1 appear red, while at a pH = 4.5 they appear colourless and at pH 7-8 they appear blue [242]. As the undiluted blueberry purée becomes more diluted, its pH increases causing a shift from red to blue. In its undiluted state the spectral curve matches that of blueberry. As the purée becomes diluted with water, which has weak visible absorption, the mixture's absorbance decreases, resulting in an observable difference in spectral composition. These phenomena are reflected in the generated theoretical image patches in Figure 6.4b. These findings support the hypothesis that purée concentration can be quantifiably estimated using a perceptual machine learning framework; there is consistency between what is visually observed and what can be quantifiably described by the optical dilution model using intensity. Thus, DNNs, which leverage visually observable information and model complex non-linear relationships, seem to be a good model for predicting purée concentrations since they are biologically inspired and modeled after the human visual cortex for decision making [20].

### 6.1.2.3 Sample Patches

Figure 6.7 depicts sample patches for each class of purée flavour taken from the eighth patch generated from the first subimage. The sample blueberry patches match the theoretical patches generated using the optical dilution model in Figure 6.4 which supports the hypothesis that quantifiable observational evidence can be used to estimate relative nutrient concentration. A colour intensity gradient across the concentrations was observed. Several purée flavours, most prominently in banana and beef, also exhibited a gradient across an image class most notably in the higher concentrations. For example, from bottom to top, the 100% beef samples darken. This was due to the highly cohesive nature of these samples; much more of the 5 mL sample loaded onto the sheet vertically rather than spreading horizontally. Specifically, this was due to the properties of the initial viscosity of the samples (i.e., there was variance in the viscosity of the 100% initial concentration) not because of the preparation of the dilution series. These gradients are visual indications which the network may be using to distinguish between different concentration classes.

### 6.1.2.4 Network Accuracy

The observations noted provide both quantitative (Figures 6.5 and 6.6) and qualitative (Figure 6.1) insight into performance. The method with the highest performance across flavours was our proposed DNN with an overall accuracy of  $92.2\% \pm 4.1\%$ , sensitivity of  $83.0\% \pm 1.5\%$ , and specificity of  $95.0\% \pm 4.8\%$ . This was closely followed by the handcrafted features paired with random forests for discrimination between dilutions. The most consistently highest performing

purée flavour was strawberry. However, the stacked autoencoder and softmax layer approach performed best on blueberry. Across 10 trials, the mean accuracy for classifying blueberry dilutions was  $99.6\% \pm 0.6\%$  (sensitivity  $98.9\% \pm 1.9\%$ , specificity  $99.7\% \pm 0.5\%$ ). These results are consistent with the descriptive analyses based on the high variance of colour, entropy (texture) observed between classes of blueberry dilutions in addition to less image saturation across dilution classes. For example, the lowest concentrations appeared more grey-blue compared to a more red-purple of the high concentration sample. Additionally, the blueberry samples also contained flecks of blueberry seeds or peels more visible in the lower concentrations than higher concentrations. These intuitive observations are congruent with the optical dilution model and the quantitative descriptive analyses with consistent and high accuracy. All other purée flavours’ average accuracy ranged between  $73.3\% \pm 7.8\%$  (chicken) and  $98.2\% \pm 1.2\%$  (strawberry). Across all seven methods for discrimination between dilutions, chicken was the most difficult flavour which was reflected in the single poorest accuracy, sensitivity, and specificity for every method. This was unsurprising since chicken samples were relatively indistinguishable to the human eye for the first several concentrations as they all simply looked white (i.e., high scattering, low absorption) with no discernible features. This was consistent with the low entropy and high saturation seen in Figure 6.6. While the softmax classification method using features from the stacked autoencoder produced the highest accuracy across the 13 flavours, handcrafted features using random forests for classification was a close second (accuracy: 0.922 vs 0.920).

Table 6.1: Summary of sensitivity, specificity and accuracy across all flavours using either self-generated features extracted from an autoencoder (softmax) or colour and texture based handcrafted features for using random forest, SVM - linear kernel SVM, and SVM - radial basis kernel implementations.

<b>SUMMARY OF PERFORMANCE ACROSS 13 FLAVOURS</b>			
<b>Method</b>	<b>Sens</b>	<b>Spec</b>	<b>Acc</b>
	$(\mu_{Sens} \pm \sigma_{Sens})$	$(\mu_{Spec} \pm \sigma_{Spec})$	$(\mu_{Acc} \pm \sigma_{Acc})$
<b>Softmax</b>	<b><math>0.830 \pm 0.150</math></b>	<b><math>0.950 \pm 0.048</math></b>	<b><math>0.922 \pm 0.041</math></b>
SVM Linear	$0.665 \pm 0.188$	$0.890 \pm 0.081$	$0.830 \pm 0.057$
SVM Radial Basis	$0.577 \pm 0.175$	$0.850 \pm 0.078$	$0.770 \pm 0.063$
Random Forest	$0.826 \pm 0.158$	$0.949 \pm 0.047$	$0.920 \pm 0.044$

### 6.1.2.5 Within-flavour Dilution Absorption Profile

The  $r^2$  values using an inverse exponential model were computed to compare the correlation and goodness of fit between the nine colour features (three colour spaces) and nutritional informa-

tion. To accomplish this, for each flavour at every dilution, the median was taken across the six instances and colour information plotted against the nutritional information obtained from the nutrition label (scaled accordingly by relative dilution). A summary of correlation coefficients for each colour feature and nutrients across all dilutions and flavours for images can be found in Table 6.2. Based on the trend across dilutions within a flavour, there was a clear increase in perceived lightness (increased reflectance) with decreasing relative concentration. For example, lower relative (e.g., 20% IC) concentrations appeared lighter than their higher relative concentration counterparts (e.g., 100% IC). Looking to the optical dilution model, this is intuitive in that lower relative concentrations would have fewer absorption events and a higher degree of reflection from the underlying white surface upon which the images were taken. This is consistent with changes and correlations with all three colour spaces. R,G,B, most consistently had the highest correlations between colour features and flavours which all contain both chroma and intensity information. To separate chroma and intensity information, HSV and CIELAB (Lab) colour spaces were computed. In Lab colour space, the highest correlation was found in the L channel comprised of the intensity information. In the case of the flavour “chicken”, consistent with the observation that all images of this flavour were oversaturated, the correlation between Lab channel  $r^2$  value were also the lowest with correlations of 0.11, 0.04, and 0.08, respectively. This corroborates that the current system and analysis failed for this flavour.

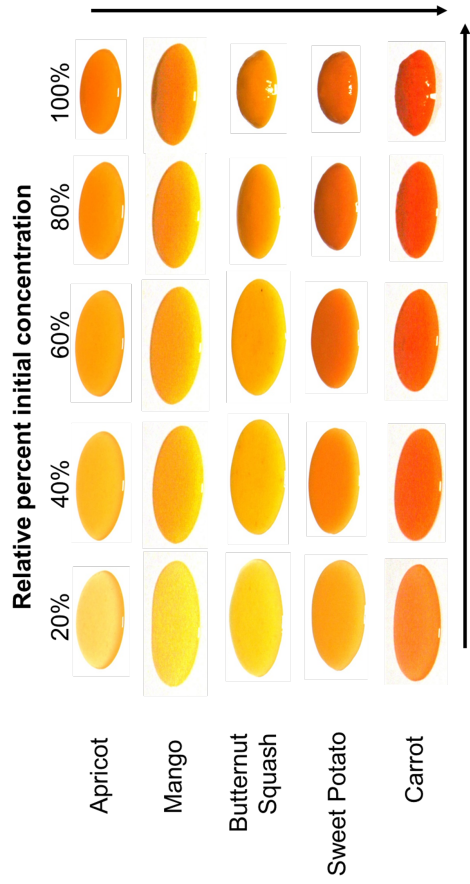
Two additional trends were observed when exploring the correlation between colour features and nutrients including: (1) the correlation between colour features and vitamin A (including beta-carotene), which contain known chromophores in the visible spectrum, specifically with absorption maxima around 328-575nm [253, 66], (2) the special case of blueberry in that there was an observable colour change shift due to the presence of anthocyanins in blueberry [221]; while the colour shift was not visibly perceptible in strawberry, they too contain anthocyanins and were analysed for a similar trend [221]. For consistency and comparison in other food applications [157, 192, 132, 204, 159], the Lab colour space will be discussed further.

#### 6.1.2.6 Vitamin A Rich Purée Samples

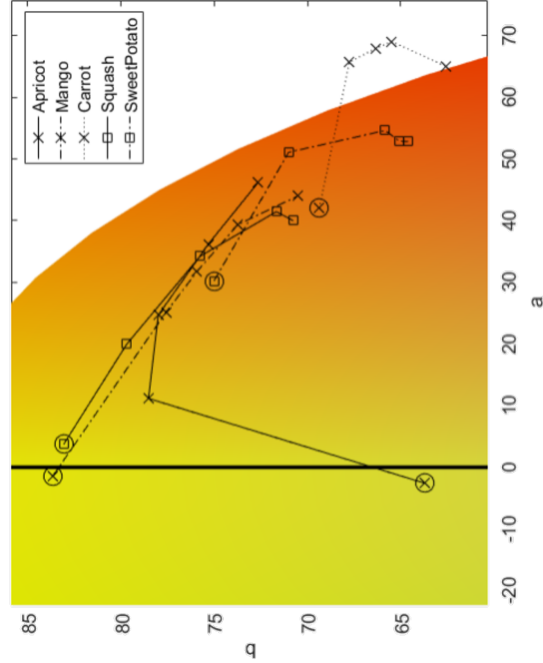
In the case of purée samples containing high amounts of vitamin A there is a similar trend moving from highest relative initial concentration on the right with a slight up and leftward direction with increased relative dilution. Specifically, in a 100mL sample of 100% IC samples the following flavours contain (% recommended daily intake): carrot (440%), sweet potato (200%), squash (160%), mango (107%), apricot (21%). Based on Figure 6.8, this translates to a spectral shift from a redder orange to a yellow and is consistent with the sample images as seen in Figure 6.8. Lower relative concentration samples appeared more yellow, while higher relative concentration samples appear more red. This was observed both qualitatively and quantitatively in Figures 6.8(a) and 6.8(b) respectively.

Table 6.2:  $r^2$  values demonstrating the correlation between different colour features and relative nutritional density across flavours; the average and median  $r^2$  values over all flavours are also presented.

<b>FLAVOUR</b>	<b>R</b>	<b>G</b>	<b>B</b>	<b>H</b>	<b>S</b>	<b>V</b>	<b>L</b>	<b>a</b>	<b>b</b>
Apple	0.8	0.97	0.79	0.93	0.7	0.8	0.97	0.93	0.13
Apricot	0.54	1	0.92	1	0.72	0.54	0.99	0.95	0.16
Banana	0.79	0.2	0.33	0.18	0.33	0.68	0.21	0.08	0.39
Beef	0.28	0.16	0.11	0.01	0.04	0.28	0.2	0	0.12
Blueberry	0.96	0.99	0.99	0.87	0.94	0.97	1	0.46	0.09
Carrot	0.82	0.73	0.34	0.69	0.37	0.82	0.82	0.48	0.96
Chicken	0.04	0.19	0.08	0.45	0.08	0.17	0.11	0.04	0.08
Mango	0.53	0.94	0.6	0.9	0.55	0.53	0.92	0.66	0.95
Parsnip	0.67	0.6	0.25	0.66	0.22	0.67	0.62	0.05	0
Pea	0.79	0.61	0.8	0.75	0.01	0.76	0.68	0.64	0.61
Squash	0.78	0.93	0.34	0.9	0.39	0.78	0.94	0.77	0.97
Strawberry	0.94	0.98	0.93	0.91	0.88	0.94	0.99	0.79	0.28
Sweet Potato	0.88	0.73	0.17	0.67	0.18	0.88	0.81	0.53	0.87
<i>Average</i>	0.68	0.69	0.51	0.69	0.42	0.68	0.71	0.49	0.43
<i>Median</i>	0.79	0.73	0.34	0.75	0.37	0.76	0.82	0.53	0.28



(a) Sample images of flavors containing vitamin A. Arrows indicate direction of increasing vitamin A content.



(b) Flavors containing high vitamin A plotted in L normalized Lab color space.

Figure 6.8: Purée samples containing high amounts of in vitamin A; open circle denotes lowest relative concentration in the dilution series. (a) Representative sample of the purée food sample dilution series from left (20% relative to initial concentration) to right (100% relative to initial concentration). Arrows denote increasing relative vitamin A content. Relative vitamin A content increases from left to right moving from yellow (lower vitamin A content) to red (higher vitamin A content). (b) L-normalized a,b plots of purée samples across dilutions. Comparing across dilutions between flavours, the same qualitative trends can be observed quantitatively in Lab colour space.

### 6.1.2.7 Anthocyanin-containing Purée Samples

Both blueberries and strawberries contain the phytonutrient anthocyanin which are pH sensitive chromophores which shift from red to blue with increasing pH. For example, at a pH = 1, anthocyanins appear red, while at a pH = 4.5 they appear colourless and at pH 7-8 they appear blue [242]. As the undiluted purée becomes more diluted, its pH increases towards water causing a shift from red to blue. Especially in the case of blueberry, this was observed both qualitatively and quantitatively in Figures 6.9(a) and 6.9(b) respectively.

### 6.1.3 Discussion

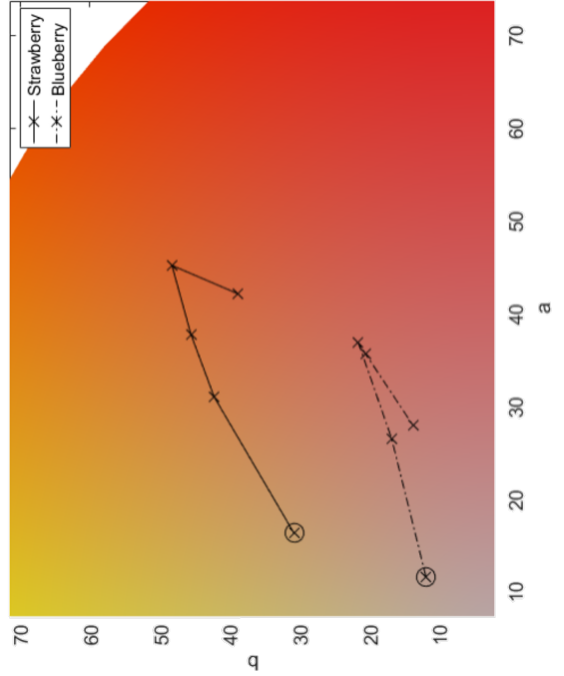
While many nutrient related signals fall outside the visible spectrum and instead in the UV, IR, or NIR, it may be possible to leverage visible spectrum colours to probe the presence or absence of specific nutrients such as vitamin A and anthocyanins. That said, if both blueberry and strawberry contain anthocyanins, why did the drastic colour shift only qualitatively observed in blueberry? One potential reason for this may be that the pH of strawberry purée sample in its undiluted form was above a pH of 4.5 (colourless anthocyanins). This may be plausible given the purée samples contain water in the ingredients and that initially strawberries are of higher pH than blueberries ( $\text{pH}_{\text{blueberry}} = 3.64$ ;  $\text{pH}_{\text{strawberry}} = 3.73$  [56]). It may also be due to other pigment molecules present in the strawberry which confound the anthocyanin colour shift. While qualitatively it may not be clear, quantitatively the strawberry a,b colour features did follow the same trend as that of blueberry.

While the highest performing method for discrimination between classes was the DNN, arguably, the combination of hand-crafted features with random forests for discrimination performed comparably. However, when

Since the colour of the same sample compared between exposures was not equal it appears there is a fundamental difference with how the light interacts with the samples. It would be interesting to test this hypothesis by investigating whether there is a correlation between the degree of difference and composition and to explore whether there are correlations between colour (and perhaps relative or normalized entropy) and composition of macronutrients (e.g., carbohydrates, protein, fat) or micronutrients (e.g., vitamin A, iron). Additionally, as part of future work, this computational nutritional density analysis should be validated with traditional rheology methods. For future extension, additional testing should be conducted using nutrient specific manipulations in which food components to determine whether the optical dilution model holds true for changes in substances extending beyond water content.



(a) Sample images of anthocyanin containing flavors. Arrow indicates increasing pH resulting from increasing relative dilution.



(b) Blueberry and strawberry in Lab color space.

Figure 6.9: Purée samples containing anthocyanins; open circle denotes lowest relative concentration in the dilution series. (a) Representative sample of the purée food sample dilution series from left (20% relative to initial concentration) to right (100% relative to initial concentration). Relative anthocyanin content increases from left to right while pH increases from right to left accounting for the observed colour shift from red (lower pH) to blue (higher pH). (b) L-normalized a,b plots of purée samples across dilutions. Comparing across dilutions between flavours, the same qualitative trends can be observed quantitatively in Lab colour space.

## 6.2 Towards Computer Vision Powered Colour-Nutrient Assessment of Pureéd Food

Leveraging the work earlier in this chapter, we developed a second way of assessing nutrient composition that, when paired with the autoencoder, can be used to estimate nutrient density for optically active nutrients in the future. This second approach co-integrates machine learning and computer vision with biophotonic analysis. Figure 6.10 shows a graphical representation of the system architecture.

### 6.2.1 Methods

Fine-grained, single nutrient assessment is accomplished through biophotonic analysis and yields an ellipse representing the Lab Gaussian distribution for each flavour and dilution, and a % transmittance map comparing highest (sweet potato) and lowest (carrot) vitamin A content. For this work, we used a transmittance-mode setup using petri dish samples to maintain strict volume consistency between samples and reduce the effects of specular reflectance and maximize diffuse reflectance. A five-tier dilution series was collected, this time in 15 mL aliquots in a petri-dish. We selected five commercially-prepared pureéd foods containing vitamin A: butternut squash, carrot, mango, and 6- and 8-month sweet potato. Thirty full-field white normalized transmittance images were acquired with a broadband tungsten-halogen light source and front glass fabric diffuser under a glass loading plate. Pixelwise spectral transmittance was computed using white and dark normalization.



To avoid photon boundary artifacts, a 35×35 mm region in the center of the sample was used to analyze the distribution of pixel transmittance spectra. Colour values were converted to Lab colour space, which has been shown to accurately capture the underlying chemical structure (i.e., conjugated double bonds) of carotenoid variants [159]. Images were processed using a photon migration model in a homogeneous pureéd mixture and Beer-Lambert exponential decay of light



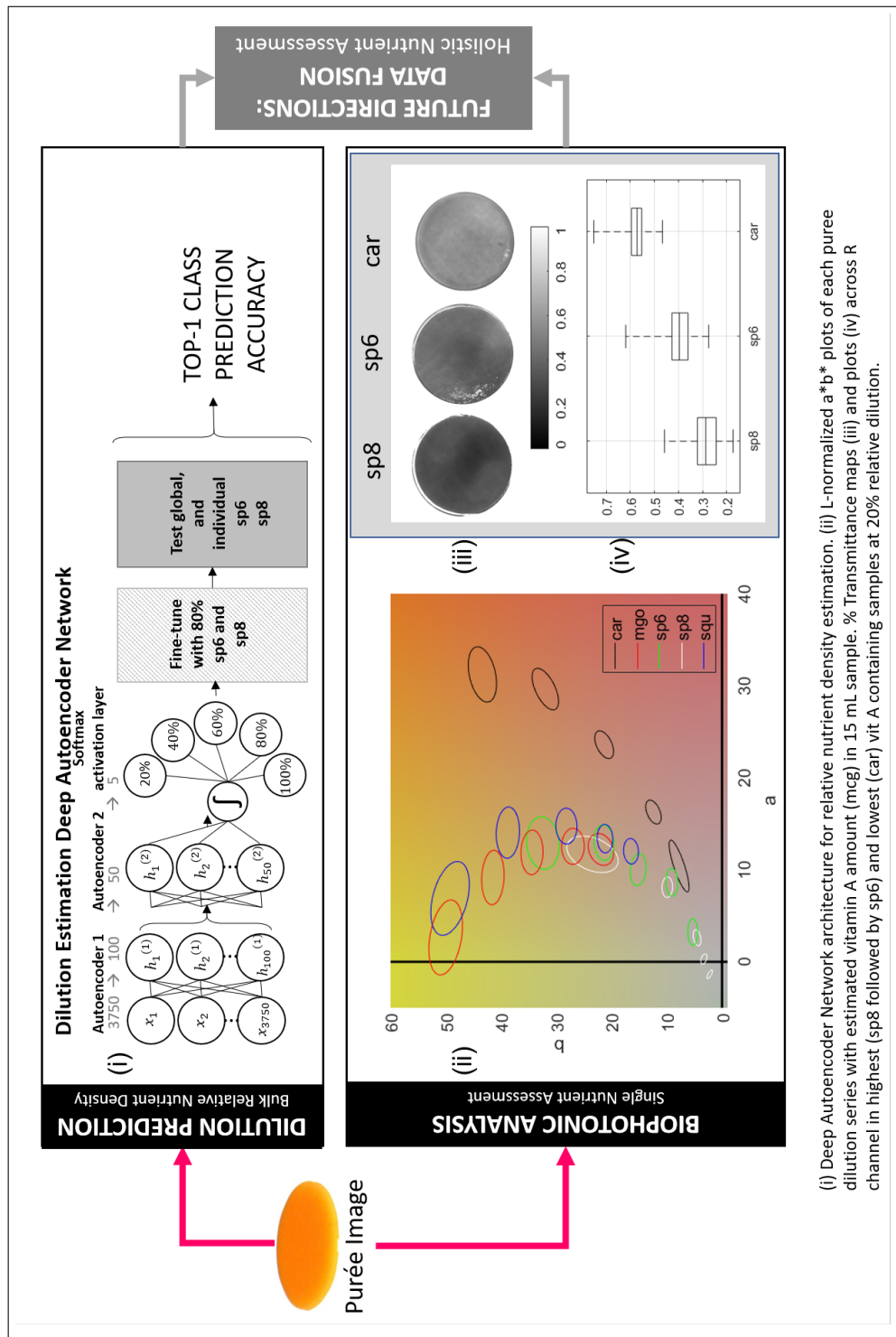


Figure 6.10: Two pronged purée nutrient composition assessment that, when paired with the autoencoder, can be used to estimate bulk nutrient density for optically active nutrients in the future system analysis comprised of coarse-grained bulk nutrient estimation (top) and single nutrient assessment through biophotonic analysis for fine-grained nutritional assessment (bottom). (i) Deep Autoencoder Network architecture for relative nutrient density estimation. (ii) L-normalized  $a*b$  plots of each puree dilution series in 15 mL samples. % transmittance maps (iii) and plots (iv) across R channel in highest (sp8 and sp6) and lowest (car) vit A containing samples at 20% relative dilution.

attenuation [183]:

$$T = \frac{I}{I_0} = \exp(\varepsilon_{H_2O} \cdot c_{H_2O} \cdot l_{H_2O} + \varepsilon_{vitA} \cdot c_{vitA} \cdot l_p) \quad (6.3)$$

where  $T$  is transmittance,  $I_0$  and  $I$  are the incident and transmitted illumination respectively. In the selected purées, we assume that the dominant absorber is vitamin A. For  $H_2O$  and  $vitA$ ,  $\varepsilon$  are the chromophore extinction coefficients,  $c$  are the concentrations, and  $l$  are the mean photon path length through each of water and the purée sample.

Coarse-grained bulk nutrient estimation was accomplished through predicting relative nutritional density using the deep relative nutrient density autoencoder network described and trained above. In this study, 400 RGB samples of size  $25 \times 50$  were extracted from the transmittance image acquisitions of the purée samples. To account for our imaging modality switch, using the transmittance data, we fine-tuned our reflectance-mode pretrained sweet potato network from [181] to a maximum of 400 epochs with a combined random 80% ( $n=320$ ) of transmittance samples of 6- and 8-month sweet potato to form a general fine-tuned “sweet potato” relative nutrient density network. To evaluate the accuracy of the proposed deep autoencoder network, we leveraged the remaining 20% of transmittance samples and compute accuracy for each of: global sweet potato (e.g., 6-month and 8-month combined), and individually for 6- and 8-month sweet potato. Five separate runs were performed.

## 6.2.2 Results & Discussion

Figure 6.10(ii) depicts L-normalized a,b values for all tested flavours at each of five concentrations relative to initial concentration. There is a trend observed with higher values of Vitamin A nearer to the origin with lower values arching upward and to the right. This is congruent with the visual appearance of the purées. For each flavour, the more highly concentrated purée samples were darker and between flavours, there was an observable colour difference. Except for carrots, the redder or more orange purées had higher %Daily Intake amounts of vitamin A. One interesting observation is that while sweet potato had the highest % Daily Value of vitamin A ( $\%DV_{vitA}$ ), carrot, with the second lowest ( $\%DV_{vitA}$ ) appeared much darker. We discuss and rationalize this empirical observation.

To interpret these data and make sense of the carrot anomaly, a few inferences were required. First, we had to look up the  $\%DV_{vitA}$  and scale it according to age range to obtain a specific quantity of vitamin A. We used the adequate intake (AI) for vitamin A measured in micrograms (mcg) of retinol activity equivalents (RAE) which compensates for the bioavailabilities of different forms of vitamin A (e.g.,  $1 \text{ mcg}_{retinol} = 12 \text{ mcg}_{\beta-carotene} = 24 \text{ mcg}_{\alpha-carotene}$ ). For infants 0-6 months and 7-12 months, the AI for vitamin A is 400 mcg, and 500 mcg for RAE [168]. When

Table 6.3: Summary of the vitamin A related nutritional information from raw samples of sweet potato and carrot [226, 173]

Vitamin A Component	Value in mcg per 1g	
	Sweet Potato	Carrot
Vitamin A, RAE	7.09	8.35
Retinol	0	0
$\beta$ -carotene	85.09	82.85
$\alpha$ -carotene	.07	34.77

considered together with the retinol activity equivalent, this implies that if considering a single source of vitamin A, double the amount of  $\alpha$ -carotene than  $\beta$ -carotene is required to achieve 100% adequate intake. This has some interesting optical implications. When considering the chromophores contributing to vitamin A content, let's focus on comparing the vitamin A components in sweet potato (highest vitamin A) to carrot (second lowest vitamin A). While there are similar amounts of  $\beta$ -carotene, in both sweet potato and carrot, carrot contains a substantial amount of  $\alpha$ -carotene (Table 6.3) [226, 173]. Assuming a homogeneous purée, normalized incident illumination and relative concentrations, and breaking vitamin A down further into the contributions from  $\beta$ - and  $\alpha$ -carotene ( $p_\beta, p_\alpha$ ) Equation 6.2 becomes:

$$A = -(\varepsilon_{H_2O}(1 - c_{vitA}) + \varepsilon_{vitA}(c_{p_\beta} + c_{p_\alpha}))l \quad (6.4)$$

where  $A = -\log T$  is optical absorbance. Now individually considering sweet potato (swp) with negligible  $\alpha$ -carotene, and carrot (car) in an undiluted pure sample we can further simplify by assuming the contribution due to water is consistent across flavours with the same extinction coefficient and pathlength which we represent with  $k$ , and expressing the concentration in mcg/g we get:

$$A_{swp} = -k(85.09) \quad (6.5)$$

$$A_{car} = -k(82.85 + 34.77) = -k(117.62) \quad (6.6)$$

This implies carrot has a  $(A_{car} - A_{swp}) \times 100\% / A_{swp} = 38\%$  higher absorbance than sweet potato. Only considering vitamin A content as contributing to photon absorption events is likely over-simplified especially when considering additional nutritional constituents which may affect absorption (e.g., chlorophyll, iron) and scattering (e.g., fat, starch) of incident photons.

Visual and biophotonic trends observed can reinforce our discussion and interpretation of our proposed fine-tuned deep autoencoder network. Regarding the former, analyzing the Lab space

Table 6.4: Summary of sweet potato dilution prediction network accuracy.

<b>Network Run</b>	<b>sp6+sp8</b>	<b>sp6 only</b>	<b>sp8 only</b>
1503191536	79%	68%	92%
1503191538	80%	79%	81%
1503191540	75%	69%	81%
1503191542	79%	74%	83%
1503191544	78%	76%	79%
<b>Average (<math>\mu \pm \sigma</math>):</b>	$78 \pm 2\%$	$73 \pm 5\%$	$83 \pm 5\%$
<b>Max:</b>	80%	79%	92%

of the biophotonics transmittance data, we observed a trend with higher values of Vitamin A nearer to the origin with lower values arching upward and to the right (Figure 6.10 (ii), Biophotonic Analysis left pane). This is congruent with the visual appearance of the purées; within a flavour the more highly concentrated purée samples were darker and between flavours, there was an observable colour difference. Except for carrots, the redder or more orange purées had higher % Daily Value (%DV) amounts of vitamin A. While sweet potato had the highest %DV of vitamin A ( $\%DV_{vitA}$ ), carrot, with the lowest  $\%DV_{vitA}$  appeared more red. When  $\beta$ - and  $\alpha$ -carotene are considered together, it would seem carrot has more carotenoid absorbers present. Since carotenoids absorb in the blue-green range, perhaps a larger relative amount of red is getting transmitted, which may account for the higher % transmittance in carrot compared to sweet potato as depicted in Figure 6.10 (iii). An additional observation was that while the nutritional composition for beginner baby food (sp6) and intermediate baby food (sp8) were similar, visually they appeared different; sp6 was slightly redder and lighter than its sp8 counterpart and is shown in Figure 6.10 (ii) and (iii) implying there were some nutritional differences perhaps not accounted for within the nutritional label.

Given there were visible differences between flavours, we wanted to explore whether it was possible to develop a generalizable deep autoencoder network for predicting relative nutritional density related to vitamin A concentration. Table 6.4 shows results using a fine tuned sweet potato deep network. For the combined testing on sp6 and sp8, we achieved a maximum top-1 prediction accuracy of 80%, with each the accuracy of sp6 and sp8 of 79%, and 92%, respectively. One potential contributing factor to errors is the overlap between classes based on visual and Lab similarity. For example, in Figure 6.10 (ii) there are three white ellipses near the intercept which correspond to  $sp8_{100\%}$ ,  $sp8_{80\%}$ , and  $sp8_{60\%}$  and these three overlap with the lower most green ellipse which belongs to  $sp6_{100\%}$ . With more data for fine-tuning, this approach can be expanded to additional flavours as well given that all except for carrot followed a similar trend but shifted up the arch.

Again, only considering vitamin A content as contributing to photon absorption events is likely an over-simplification. However, it provides a toy-case which opens the door of possibility for using light to assess nutrient amounts for additional optically active nutrients. Additionally, actual values of vitamin A in the purée may differ from its raw counterpart due to vitamin A's thermosensitivity and oxidation susceptibility during processing [127]. Exploring computer vision-based techniques on how to distinguish between contributions from biophotonic absorbers and scatterers may further enhance our ability to interpret nutritional quantity and quality. Next steps include the integration and data fusion from these two co-processes and expansion to additional nutrients and across a larger sample of food items.

## 6.3 Conclusions

We demonstrated the feasibility of automatic nutritional density analysis using deep neural networks to predict the concentration of commercially prepared purées. Dilution classification results were strongest for purée flavours with observable texture differences reflected in higher entropy, higher variation in colour across dilution classes, and lower saturation. In contrast, purée flavours performed more poorly when there were fewer visual cues to discriminate between dilution classes which was further reflected in low colour variation, low entropy across classes, and high saturation. These findings begin to clarify the constraints of working towards classification with naturalistic images taken in the field. Based on the promising preliminary findings of quantitatively observing differences in vitamin A and anthocyanin content, optical techniques leveraging visible spectrum information may be feasible for cost-effective nutrient density assessment in commercially puréed food samples. Additional investigations must be conducted to discern to which nutrients this approach may be reliably applied. This work provides evidence in support of an optically motivated objective malnutrition risk assessment tool.

### 6.3.1 Key Contributions

- Novel way for assessing relative nutrient density through different optical windows (shown here in the visible spectrum) to predict relative nutrient density of foods.
- This can inform new technology for supporting quality assessment of puréed foods.
- The application of deep learning in this field of food biophotonics is a novel contribution; to date, machine learning approaches typically consider hand crafted features and do not focus on predicting nutrient density.

#### UP NEXT...

This chapter concludes one story of my learning journey since 2016. As with all research, as one question is answered, another five are asked and our rear-view mirror can inform how we might continue to improve, grow, and build beyond. Chapter 7 provides some insight into my reflections on this learning journey, opportunities for further enhancement, and the next important steps for translation to the LTC environment.

# Chapter 7

## Towards Translation: Current Limitations and Future Directions

This thesis brings together my work across several different fields including user-centered co-design, nutrition, machine learning, and biophotonics. As a system, we have demonstrated that AFINI-T shows promise for more accurately tracking resident food intake. This section brings to a close a summary of current limitations and additional considerations followed by future directions and opportunities to disrupt for translational impact.

### 7.1 Current Limitations and Additional Considerations

As with all science, there are different ways of approaching a problem with their own unique set of advantages, and limitations. This section outlines opportunities for further enhancing my LTC food dataset, the issue of low-density foods, additional considerations for assessing volume estimation error, considerations for pushing food biophotonics further, and finally practical implications of altering workflow in LTC with the AFINI-T system.

#### 7.1.1 Datasets

The datasets collected as part of this thesis were the first representative of LTC foods, with pixel-wise segmentation and depth map information, and at multiple simulated intake levels. These provide a foundation upon which to grow. One initial way would be to prepare computationally simulated plates for modified texture foods on a plate by combining modified texture single food items from the modified texture food dataset into computationally simulated multi-item plates.

This may provide additional insights into performance and accuracy more closely resembling real-world data. Furthermore, additional expansion to test the system with a wider variety of foods with fewer constraints may enhance generalizability and encourage uptake of this technology. Moving forward, further data acquisition for additional food items and food plates would be prudent for further enhancing generalizability and evaluating this technology in an increasingly realistic real-world setting including piloting true intake acquisition (as opposed to simulated levels of intake). The level of plate mixing in a real-world setting may not be similar to what was tested within the scope of this thesis so should be considered as an avenue to further advance and validate this research.

### **7.1.2 Low-Density Foods**

Using the current system, low-density fluffy foods and rigid (e.g., salad, toast), highly absorptive foods (e.g., blueberry) continue to be challenging. To further improve reliability and consistency of the system, negative volume errors and the issue of low-density foods should be considered. For the proposed EDFN-D, the volume errors for plates after the highest intake (P4 and P5) were negative because the depth-refinement omitted volumes across pixels that were initially segmented as food. As part of future work, it would be interesting to explore how necessary initial colour-based segmentation is for establishing *how much* food is present or whether placing greater weight on depth maps could improve accuracy for volume estimation accuracy. Given the issue of low-density foods impacting volume regardless of segmentation method, this must be considered a limitation of over-head food intake systems and these types of foods (e.g., potato chips, salad) may need to be treated differently and separately to other foods. Perhaps repeat imaging of these foods separately, flattening the food before imaging, or applying a general food density score to estimate the range of volume values in these foods could address this limitation.

### **7.1.3 Volume Estimation**

While we collected ground truth weighed food records, we did not account for ground truth volume. As a result, we were working under the assumption that our volume estimation was accurate. Volume validation against gold-standard ground truth (e.g., water displacement) is needed to corroborate the accuracy (when in actuality, some evidence suggests there's less than 3% error [139]). This is an important consideration for more thoroughly quantifying error at each stage. Given the state of the literature in how error is typically reported (if it is), this thesis provides a step towards more transparent technology for supporting trust in the system.



### 7.1.4 Pushing Food Biophotonics Further

If continuing to apply food biophotonics, careful consideration about which wavelengths would be useful for capturing specific nutrients is needed. In the case of chicken, we saw that the visible spectrum was inadequate as all images, regardless of ISO, were oversaturated. Additionally, since chicken contains fat and protein, this increases the number of scattering events making it more difficult to make nutritional inferences. Were this approach deemed valuable to study further, there may be more promise in applying more involved methods such as spatial frequency domain imaging to quantify  $\mu_a$  and  $\mu_s$  at nutrient-specific wavelengths. To leverage visible spectrum information for the purpose of developing an optical imaging malnutrition risk assessment tool capable of determining relative nutritional density, several additional steps are required. First, a nutritional density prediction model must be developed based on the optical dilution models of major food chromophores to predict: the concentration of chromophores in a sample and the dilution of the sample based on the mode chromophore prediction. We can then estimate purée depth based on: relative absorption ( $A$ ), our known theoretical relative concentration ( $c$ ), and constant extinction coefficient ( $\epsilon$ ) for each puréed food flavour and potentially augment the system with a depth camera to yield higher accuracy of pathlength estimates. From here, we may disentangle  $\mu_s$  from  $\mu_a$  to make nutrient quantification inferences where  $\mu_s$  relates to structure and particle size whereas  $\mu_a$  relates to composition [191]. Additionally, we could optically estimate particle size for food quality safety; too large of particles may impact safe consumption of the food. Currently particle size is measured with physical deformation between the tines of a fork [47]. An optical approach leveraging pixel depth compensation for scale inference may improve measurement reliability, objectivity, and save personnel time. While technically challenging and interesting, the utility of this added information within the LTC context must be weighed against the added cost and more specialized equipment required for further development especially when considering the nutrients of interest in this population which tend not to have optical activity in the visible spectrum.

### 7.1.5 Practical Implications of Altering Workflow

Since this technology would alter the workflow within the LTC context, to evaluate and validate this work including feasibility, for translation of this work, a new model of care should be trialled. While outside the scope of this thesis, this should include an assessment for required time necessary to track residents as part of a cost-benefit analysis to a system like this, user evaluation of an operational and integrated system within the iOS platform, and partnership with PointClickCare for integration with their electronic health record system. When thinking towards translation to LTC, additional considerations must include the computational power to run the analysis. Ideally, the solution will be a self-contained fully integrated (perhaps battery powered) system, and

function in real-time. There would need to be privacy considerations addressed with the need for the technology to work both on- and offline. This is a current limitation of the system identified through interviews with end-users. The current system requires wi-fi connectivity and practically, the wi-fi is quite intermittent within the Villages. This is also infeasible whenever residents leave the Village setting for a dinner at a family member, friend, or restaurant. Leveraging iPads and other edge devices may provide a unique solution and may also address potential privacy concerns of updating resident food intake records. On the flip side, it may be prudent to consider whether a missed meal necessitates more accurate tracking. Given the accepted approach of collecting 3-day food intake records, primed with higher quality data, perhaps instead, the focus should emphasize resident enjoyment of foods and operating within a sliding window to assess 3-day food intake. Additional consideration for approximating intake from these settings and linking to the USDA or Canadian Nutrient File may be beneficial for extending charting accuracy beyond foods offered within the home.

## **7.2 Future Directions: Opportunities to Disrupt for Translational Impact**

### **7.2.1 Further Automation for Decreased Time Requirements**

One opportunity I see is to provide additional automation for decreased time requirements. Regarding the degree of human input for collecting data, hypothetically, if it was deemed too cumbersome to track on the food item level, we could do a bulk intake assuming equal consumption across foods. This is how food intake is currently tracked but using subjective visual assessment. Additionally, since food photography is inherently part of the workflow of the AFINI-T system we would be able incorporate *a posteriori* assessments for dietitians to apply on a case-by-case basis for residents requiring more in-depth assessment. This would help refine the referral process. Currently there is no way for dietitians to know what was consumed to cross-reference without them directly observing a mealtime experience. For the residents of interest, a dietitian could assign specific labels for only the most at-risk individuals to get a more accurate individual assessment to go from coarse-grain to fine-grain where needed. This is distinctly different from just using a general segmentation strategy because it has the infrastructure built-in to support fine-grain assessment by one-time application of labels to each class on the reference, full-portion image. While in this current implementation, one reference full-portion image needs to be hand-labelled at the pixel-level, unlike a semi-automatic method like graph cut, this reference image can be used for *ALL* residents who ordered the same meal items at the same serving size as opposed to defining it resident-by-resident image-by-image as would be required using graph cut alone. This allows flexibility for the level of human involvement depending on the desired

level of accuracy. By going this route even without assigning classes on the reference portion, the system provides enhanced accuracy and removes subjectivity around bulk food intake estimation in the coarse-estimation and empowers the dietitian to delve deeper for specific residents on a further enhanced fine-grain assessment leveraging objective portion estimation enhanced food photography. Hopefully even in this iteration instead of requiring each image to be hand-segmented or segmented using graph cut, our approach of combining output from our EDFN-D with the convolutional autoencoder will already remove a degree of burden and minimize barrier to uptake. To confirm, further testing would need to be conducted to ensure these hypothetical assumptions are valid in the real world, or the extent to which they hold true.

### **7.2.2 Incorporation of Fluid Intake Tracking**

A second opportunity is to incorporate fluid intake tracking. While the focus here was on nutritional intake through food, a system which additionally addresses fluid intake may be highly desirable. Dehydration happens rapidly, is avoidable, and increases potentially avoidable hospital admissions [236]. In some of my unpublished work that didn't make it into this thesis but was presented at the Canadian Nutrition Society 2019 Annual Conference, I explored food intake patterns in Ontario LTC homes. This was part of secondary data analysis of the Making the Most of Mealtimes data involving over 600 residents across 32 LTC homes in 4 provinces [123]. Through these analyses I've seen that proportionately large amounts of total caloric intake come from fluids including juice (11% of calories), milk or soymilk (9%), and oral nutritional supplements (3%). This is currently a blind spot in the AFINI-T system but accounts for over 20% of average daily calories. While this is beyond the scope of this thesis, template matching for vessel paired with a machine learning fluid module for classifying beverages based on observable optical properties (e.g., colour, transparency) may be a valuable place to start and provides a complementary and more holistic approach to food intake tracking. This may be where applying food biophotonics could have the greatest impact from the LTC home perspective.

### **7.2.3 Supporting Enhanced Personalisation**

A third opportunity is to support enhanced personalisation. While in the design phase, we learned that personalisation would be valuable especially for flagging risk factors for residents, this was not incorporated in the current form of the system. Two ways in which this could be addressed in future iterations are (1) providing population-level food intake insights, and (2) at the individual level. To apply population-level food intake insights, food databases and recipes could be linked with food healthfulness scores such as the Healthy Eating Index [84] or Nutrient Rich Foods [62]. While generally in-line with the new Canada's Food Guide recommendations of eating more of vegetables and fruits, nuts, seeds, and plant-based protein sources [81], the Healthy Eating Index

provides a scoring metric which encourages nutrient-dense foods and penalizes foods that are more processed or high in saturated fat or salt. Similarly, the Nutrient Rich Food score does the same but sums across nutrients to encourage and nutrients to limit. One other scoring system which may be of relevance to this population is the Dietary Inflammatory Potential scale [209] which was developed using associations between specific nutrients and blood biomarkers indicating increased inflammation. As inflammation plays a role in both cardiovascular disease and diabetes [146], two prevalent chronic conditions in this population (cardiovascular disease: 75%, diabetes: 30% [35]), designing foods to reduce inflammatory potential may provide a complementary perspective to address chronic disease management more holistically. Incorporation of these types of metrics may help aid menu-planning while optimising resident preferences and choices.

#### **7.2.3.1 AFINI-T data-driven approach to inform menu planning.**

Related to the third opportunity, AFINI-T is platformed to provide actionable insights for dietitians and director of food services based on data-driven insights. For example, it could be used to develop recipes that are more nutrient dense. Creating nutrient dense meals while minimizing cost is a contention in LTC as there is a fixed allocation of food per resident. In 2017 the raw food allocation in Ontario was \$7.30–12.50 [235] and up to \$9.54 in 2020 [11]. Until recently, there was a disconnect between requirement to serve full portion to meet nutritional requirements but because of budget, the foods were relatively inexpensively made and the quantity required was unsuitable; served portions were much too large and there was a high degree of food waste [65, 83]. AFINI-T can help provide tangible insights while switching toward more nutrient dense foods with the ability to assess the cost benefit. Output on which foods are consumed can inform how to design recipes to be smarter, more expensive but also more nutrient dense, with the expectation of less waste and more of portion consumed, especially when paired with software such as Food Processor for designing recipes. AFINI-T could close the loop on what is actually consumed to provide quantitative tangible numbers to support effectiveness of intervention trials as part of cost-benefit analysis.

It could also be used as a tool for developing more nutrient-dense recipes in which certain ingredients could be replaced with other. For example, replacing half of the ground beef in a chili recipe for lentils to decrease saturated fat, cholesterol while increasing fibre. When considering the individual-level, there may utility in learning resident preferences to support individual preferences even with rotating team member support. AFINI-T can provide insights when assessing the overall dining experience to help inform whether residents aren't consuming a food because they don't like the taste. A complementary future direction would be a user study to leverage information about what was consumed to learn whether a resident is consuming food because it is what's fed to them versus foods they actually enjoy. This could impact resident lives who

are living with dementia. For example for residents who cannot communicate verbally, a valuable extension would be to tap into a specific resident's intake trends over time to learn their resident preferences through this data-driven approach to support enjoyment factor. Having a system to provide these types of insights to new or different team members may support more resident-directed care. Additionally, analysing individual-level food intake patterns may provide a powerful tool to flag changes in eating habits. This could be used as a marker to potentially flag residents' worsening (or improving) condition. It may also serve as a means to flag residents who may be getting sick before symptoms present as a piece of infection control risk and management.

#### **7.2.4 Broadening AFINI-T's Reach with the Mobile Era**

The fourth opportunity relates to expanding AFINI-T both within LTC as well as beyond. Currently, menu planning and reporting to ministry is completed by the director and assistant director of food services (DFS/ADFS). Cooks prepare the foods and dietary aides plate the food. It's the personal support workers who track and monitor resident intake with the registered team to process information and dietitian overseeing with neighbourhood coordinator. The resident is not currently in the loop. As part of the AFINI-T development, we did not include the include the resident in the loop. For LTC not appropriate given degree of memory impairment and focus needs to be on enjoyment not adding additional tasks. However, when considering translation to additional settings such as retirement, independent living, or community living, it would be interesting to have a system that would have the individual at centre of model of service (nurse-led or dietitian-led) to get feedback on the individual's specific nutrient targets. While this strategy may not be appropriate for LTC within retirement, this may be different. Exploring acceptability beyond LTC with the resident/patient at the centre of the process would provide valuable insights as part of future work. In these settings, leveraging smart phones depth sensors would be ideal as AFINI-T branches into the mobile realm. We already know that AFINI-T needs to work on the iPad since this is how it's currently being tracked in LTC; leveraging built-in depth sensors already present in newer technology would enable AFINI-T to be brought into the mobile era and provide added value especially as it broadens from LTC to the general public as an interesting future avenue to explore.

Of course, at this stage, these hypotheses still require testing to assess their utility in practice. While these strategies were not part of this thesis, these would be the next steps I would take to expand the research program outlined here as I see these as providing additional supports with great potential to impact and disrupt the way in which we assess nutrition management and beyond.

## FINAL THOUGHTS

As you can see, the work outlined in this thesis on its journey towards translation is far from complete but has provided me with a wealth of experiences that has helped me grow both as a human and as a scientist. My hope is that this work may lay the foundation for its translation through its blueprints and lessons learned. My dream would be to see this come to fruition where after further validating and testing this technology, it could be handed off to end-users to start assessing and making real-world impact in the daily lives of LTC residents and beyond. And with that,

*The adventure continues . . .*

# References

- [1] Asmaa Abdelhamid, Diane Bunn, Maddie Copley, Vicky Cowap, Angela Dickinson, Lucy Gray, Amanda Howe, Anne Killett, Jin Lee, Francesca Li, et al. Effectiveness of interventions to directly support food and drink intake in people with dementia: systematic review and meta-analysis. *BMC Geriatrics*, 16(1):1–18, 2016.
- [2] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y Lim, and Mohan Kankanhalli. Trends and trajectories for explainable, accountable and intelligible systems: an HCI research agenda. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, page 582, 2018.
- [3] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.
- [4] Eduardo Aguilar, Beatriz Remeseiro, Marc Bolaños, and Petia Radeva. Grab, pay, and eat: Semantic food detection for smart restaurants. *IEEE Transactions on Multimedia*, 20(12):3266–3275, 2018.
- [5] Rana Almaghrabi, Gregorio Villalobos, Parisa Pouladzadeh, and Shervin Shirmohammadi. A novel method for measuring nutrition intake based on food image. In *Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International*, pages 366–370. IEEE, 2012.
- [6] María Dolores Alvarez and Wenceslao Canet. Dynamic viscoelastic behavior of vegetable-based infant purees. *Journal of Texture Studies*, 44(3):205–224, 2013.
- [7] Yvette N Andrews and Victoria Hammer Castellanos. Development of a method for estimation of food and fluid intakes by nursing assistants in long-term care facilities: a pilot study. *Journal of the American Dietetic Association*, 103(7):873–877, 2003.
- [8] Marios Anthimopoulos, Joachim Dehais, Peter Diem, and Stavroula Mougiakakou. Segmentation and recognition of multi-food meal images for carbohydrate counting. In

*Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*, pages 1–4. IEEE, 2013.

- [9] Shebiah Arivazhagan, R Newlin Shebiah, S Selva Nidhyanandhan, and L Ganesan. Fruit recognition using color and texture features. *Journal of Emerging Trends in Computing and Information Sciences*, 1(2):90–94, 2010.
- [10] Sinem Aslan, Gianluigi Ciocca, and Raimondo Schettini. Semantic food segmentation for automatic dietary monitoring. In *Proceedings of the IEEE International Conference on Consumer Electronics-Berlin*, pages 1–6, 2018.
- [11] Ontario Long-Term Care Association. The role of long-term care, Apr 2020.
- [12] Arlene J Astell, Faustina Hwang, LJE Brown, C Timon, LM Maclean, Thomas Smith, T Adlam, H Khadra, and EA Williams. Validation of the nana (novel assessment of nutrition and ageing) touch screen system for use at home by older adults. *Experimental Gerontology*, 60:100–107, 2014.
- [13] Tim J Atherton and Darren J Kerbyson. Size invariant circle detection. *Image and Vision Computing*, 17(11):795–803, 1999.
- [14] Nick Babich. UX Design: Best Practices, Types and States. <https://uxplanet.org/button-ux-design-best-practices-types-and-states-647cf4ae0fc6>, 2016.
- [15] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017.
- [16] Aaron Bangor, Philip Kortum, and James Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3):114–123, 2009.
- [17] Aaron Bangor, Philip T Kortum, and James T Miller. An empirical evaluation of the system usability scale. *Intl. Journal of Human-Computer Interaction*, 24(6):574–594, 2008.
- [18] Oscar Beijbom, Neel Joshi, Dan Morris, Scott Saponas, and Siddharth Khullar. Menu-match: Restaurant-specific food logging from images. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 844–851. IEEE, 2015.
- [19] Christina L Bell, Bruce K Tamura, Kamal H Masaki, and Elaine J Amella. Prevalence and measures of nutritional compromise among nursing home patients: weight loss, low body mass index, malnutrition, and feeding dependency, a systematic review of the literature. *Journal of the American Medical Directors Association*, 14(2):94–100, 2013.



- [20] Yoshua Bengio. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009.
- [21] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. A neural probabilistic language model. *Journal of Machine Learning Research*, 3(Feb):1137–1155, 2003.
- [22] Irving J Bigio and Sergio Fantini. *Quantitative Biomedical Optics: Theory, Methods, and Applications*. Cambridge University Press, 2016.
- [23] Sheila A Bingham. Limitations of the various methods for collecting dietary intake data. *Annals of Nutrition and Metabolism*, 35(3):117–127, 1991.
- [24] Jennifer Boger, Piper Jackson, Maurice Mulvenna, Judith Sixsmith, Andrew Sixsmith, Alex Mihailidis, Pia Kontos, Janice Miller Polgar, Alisa Grigorovich, and Suzanne Martin. Principles for fostering the transdisciplinary development of assistive technologies. *Disability and Rehabilitation: Assistive Technology*, 12(5):480–490, 2017.
- [25] Ruud M Bolle, Jonathan H Connell, Norman Haas, Rakesh Mohan, and Gabriel Taubin. Veggievision: A produce recognition system. In *Proceedings Third IEEE Workshop on Applications of Computer Vision. WACV’96*, pages 244–251. IEEE, 1996.
- [26] Marc Bosch, Fengqing Zhu, Nitin Khanna, Carol J Boushey, and Edward J Delp. Combining global and local features for food identification in dietary assessment. In *2011 18th IEEE International Conference on Image Processing*, pages 1789–1792. IEEE, 2011.
- [27] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101—mining discriminative components with random forests. In *European Conference on Computer Vision*, pages 446–461. Springer, 2014.
- [28] Yuri Y Boykov and M-P Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 1, pages 105–112, 2001.
- [29] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [30] Briggs Healthcare. Food intake record form top-punch. <https://shop.briggscorp.com/e2wItemMain.aspx?parentID=IT00001135&parentLink=2100000709:3100002940:3100003791:3100004249:3100004284>, (accessed: 26.11.2018).
- [31] BRiGGS Healthcare. Dietary intake form. <https://shop.briggscorp.com/pdf/1182P.pdf>, (accessed: 29.06.2017).
- [32] John Brooke. SUS: a quick and dirty usability scale. *Usability Evaluation in Industry*, 189, 1996.

- [33] Sonya Brownie. Why are elderly individuals at risk of nutritional deficiency? *International Journal of Nursing Practice*, 12(2):110–118, 2006.
- [34] Vieira Bruno and Cui Juan Silva Resende. A survey on automated food monitoring and dietary management systems. *Journal of Health & Medical Informatics*, 8(3), 2017.
- [35] Canadian Institute for Health Information. 2018.
- [36] Capterra. Best Electronic Medical Records (EMR) Software. <https://www.webcitation.org/query?url=https%3A%2F%2Fwww.capterra.com%2Felectronic-medical-records-software&date=2018-11-26>, (accessed: 26.11.2018).
- [37] Eloise CJ Carr, Julie N Babione, and Deborah Marshall. Translating research into practice through user-centered design: an application for osteoarthritis healthcare planning. *International Journal of Medical Informatics*, 104:31–37, 2017.
- [38] Jennifer Carson. Working together to put living first: A culture change process in a long-term care and retirement living organization guided by critical participatory action research. 2015.
- [39] Victoria Hammer Castellanos and Yvette N Andrews. Inherent flaws in a method of estimating meal intake commonly used in long-term-care facilities. *Journal of the American Dietetic Association*, 102(6):826–830, 2002.
- [40] Junghoon Chae, Insoo Woo, SungYe Kim, Ross Maciejewski, Fengging Zhu, Edward J Delp, Carol J Boushey, and David S Ebert. Volume estimation using food specific shape templates in mobile image-based dietary assessment. In *Proceedings of SPIE*, volume 7873, page 78730K. NIH Public Access, 2011.
- [41] Mei Chen, Kapil Dhingra, Wen Wu, Lei Yang, Rahul Sukthankar, and Jie Yang. PFID: Pittsburgh fast-food image dataset. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 289–292. IEEE, 2009.
- [42] Mei-Yun Chen, Yung-Hsiang Yang, Chia-Ju Ho, Shih-Han Wang, Shane-Ming Liu, Eugene Chang, Che-Hua Yeh, and Ming Ouhyoung. Automatic chinese food identification and quantity estimation. In *SIGGRAPH Asia 2012 Technical Briefs*, pages 1–4. 2012.
- [43] Xin Chen, Hua Zhou, Yu Zhu, and Liang Diao. Chinesefoodnet: A large-scale image dataset for chinese food recognition. *arXiv preprint arXiv:1705.02743*, 2017.
- [44] ChiroSpringVideos. ChiroSpring - a cloud-based chiropractic practice management software. <https://www.youtube.com/watch?v=3Z6pAXhAH0c>, (accessed: 15.02.2015).

- [45] Manal Chokr and Shady Elbassuoni. Calories prediction from food images. In *Twenty-Ninth IAAI Conference*, 2017.
- [46] Md Towhid Chowdhury, Md Shariful Alam, Muhammad Asiful Hasan, and Md Imran Khan. Vegetables detection from the glossary shop for the blind. *IOSR Journal of Electrical and Electronics Engineering*, 8(3):43–53, 2013.
- [47] Julie A. Y. Cichero, Peter Lam, Catriona M. Steele, Ben Hanson, Jianshe Chen, Roberto O. Dantas, Janice Duivesteyn, Jun Kayashita, Caroline Lecko, Joseph Murray, Mershen Pillay, Luis Riquelme, and Soenke Stanschus. Development of international terminology and definitions for texture-modified foods and thickened fluids used in dysphagia management: The IDDSI framework. *Dysphagia*, pages 1–22, 2016.
- [48] Gianluigi Ciocca, Davide Mazzini, and Raimondo Schettini. Evaluating CNN-based semantic food segmentation across illuminants. In *Proceedings of the International Workshop on Computational Color Imaging*, pages 247–259, 2019.
- [49] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. Food recognition: a new dataset, experiments and results. *IEEE Journal of Biomedical and Health Informatics*, 21(3):588–598, 2017.
- [50] Dan Ciregan, Ueli Meier, and Jürgen Schmidhuber. Multi-column deep neural networks for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3642–3649. IEEE, 2012.
- [51] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167. ACM, 2008.
- [52] Cynthia L Corritore, Beverly Kracher, and Susan Wiedenbeck. On-line trust: concepts, evolving themes, a model. *International Journal of Human-Computer Studies*, 58(6):737–758, 2003.
- [53] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.
- [54] George E Dahl, Dong Yu, Li Deng, and Alex Acero. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):30–42, 2012.
- [55] Olivier Dangles and Julie-Anne Fenger. The chemical reactivity of anthocyanins and its consequences in food science and nutrition. *Molecules*, 23(8):1970, 2018.
- [56] Vanessa Rios de Souza, Patrícia Aparecida Pimenta Pereira, Thais Lomônaco Teodoro da Silva, Luiz Carlos de Oliveira Lima, Rafael Pio, and Fabiana Queiroz. Determination

- of the bioactive compounds, antioxidant activity and chemical composition of brazilian blackberry, red raspberry, strawberry, blueberry and sweet cherry fruits. *Food Chemistry*, 156:362–368, 2014.
- [57] Joachim Dehais, Marios Anthimopoulos, Sergey Shevchik, and Stavroula Mougiakakou. Two-view 3d Reconstruction for Food Volume Estimation. *IEEE Transactions on Multimedia*, 19(5):1090–1099, 2017.
- [58] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.
- [59] Clarissa J Diamantidis, Jennifer S Ginsberg, Marni Yoffe, Lisa Lucas, Divya Prakash, Saurabh Aggarwal, Wanda Fink, Stefan Becker, and Jeffrey C Fink. Remote usability testing and satisfaction with a mobile health medication inquiry system in CKD. *Clinical Journal of the American Society of Nephrology*, 10(8):1364–1370, 2015.
- [60] Dietitians of Canada. Dietitians of Canada - resource library: Best practices for nutrition, food service and dining in long term care homes. <https://www.dietitians.ca/Downloads/Public/2013-Best-Practices-for-Nutrition,-Food-Service-an.aspx>, 2019.
- [61] Abul Doulah, Megan A Mccrory, Janine A Higgins, and Edward Sazonov. A systematic review of technology-driven methodologies for estimation of energy intake. *IEEE Access*, 7:49653–49668, 2019.
- [62] Adam Drewnowski. Defining nutrient density: development and validation of the nutrient rich foods index. *Journal of the American College of Nutrition*, 28(4):421S–426S, 2009.
- [63] Shiv Ram Dubey and Anand Singh Jalal. Fruit disease recognition using improved sum and difference histogram from images. *International Journal of Applied Pattern Recognition*, 1(2):199–220, 2014.
- [64] Shiv Ram Dubey and Anand Singh Jalal. Application of image processing in fruit and vegetable analysis: a review. *Journal of Intelligent Systems*, 24(4):405–424, 2015.
- [65] Lisa M. Duizer and Heather H. Keller. Planning micronutrient-dense menus in ontario long-term care homes: Strategies and challenges. *Canadian Journal of Dietetic Practice and Research*, 81(4):198–203, 2020. PMID: 32495638.
- [66] Joseph Ratcliffe Edisbury, Albert Edward Gillam, Isidor Morris Heilbron, and Richard Alan Morton. Absorption spectra of substances derived from vitamin A. *Biochemical Journal*, 26(4):1164, 1932.

- [67] Takumi Ege and Keiji Yanai. Simultaneous estimation of food categories and calories with multi-task cnn. In *Machine Vision Applications (MVA), 2017 Fifteenth IAPR International Conference on*, pages 198–201. IEEE, 2017.
- [68] Aprima EHR. Aprima overview, 2016.
- [69] Yulia Eskin and Alex Mihailidis. An intelligent nutritional assessment system. In *AAAI Fall Symposium: Artificial Intelligence for Gerontechnology*, 2012.
- [70] Shaobo Fang, Chang Liu, Fengqing Zhu, Edward J Delp, and Carol J Boushey. Single-view food portion estimation based on geometric models. In *Multimedia (ISM), 2015 IEEE International Symposium on*, pages 385–390. IEEE, 2015.
- [71] Shaobo Fang, Fengqing Zhu, Chufan Jiang, Song Zhang, Carol J Boushey, and Edward J Delp. A comparison of food portion size estimation using geometric models and depth images. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 26–30. IEEE, 2016.
- [72] Giovanni Maria Farinella, Dario Allegra, and Filippo Stanco. A benchmark dataset to study the representation of food images. In *European Conference on Computer Vision*, pages 584–599. Springer, 2014.
- [73] Patrick Ferdinand Christ, Sebastian Schlecht, Florian Ettlinger, Felix Grun, Christoph Heinle, Sunil Tatavatry, Seyed-Ahmad Ahmadi, Klaus Diepold, and Bjoern H Menze. Diabetes60-inferring bread units from food images using fully convolutional neural networks. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 1526–1535, 2017.
- [74] Peter Fernandez. “Through the looking glass: envisioning new library technologies” understanding artificial intelligence. *Library Hi Tech News*, 33(3):20–23, 2016.
- [75] Andy P. Field. *Discovering Statistics using IBM SPSS*. Sage Publications, 2018.
- [76] BJ Fogg and Hsiang Tseng. The elements of computer credibility. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 80–87. ACM, 1999.
- [77] Zhihui Fu, Dan Chen, and Hongyu Li. Chinfood1000: A large benchmark dataset for chinese food recognition. In *International Conference on Intelligent Computing*, pages 273–281. Springer, 2017.
- [78] Isabelle Germain, Thérèse Dufresne, and Katherine Gray-Donald. A novel dysphagia diet improves the nutrient intake of institutionalized elders. *Journal of the American Dietetic Association*, 106(10):1614–1623, 2006.
- [79] Davide Giavarina. Understanding Bland Altman analysis. *Biochemia Medica*, 25(2):141–151, 2015.

- [80] Scott Goates, Kristy Du, Carol A Braunschweig, and Mary Beth Arensberg. Economic burden of disease-associated malnutrition at the state level. *PLOS ONE*, 11(9):e0161833, 2016.
- [81] Government of Canada. Canada’s Food Guide. <https://food-guide.canada.ca/en/>, (accessed: 24.08.2021).
- [82] Carolyn Steele Gray, Ashlinder Gill, Anum Irfan Khan, Parminder Kaur Hans, Kerry Kuluski, and Cheryl Cott. The electronic patient reported outcome tool: testing usability and feasibility of a mobile app and portal to support care for patients with complex chronic disease and disability in primary care settings. *JMIR mHealth and uHealth*, 4(2):e58, 2016.
- [83] JA Grieger and CA Nowson. Nutrient intake and plate waste from an australian residential care facility. *European journal of clinical nutrition*, 61(5):655–663, 2007.
- [84] Patricia M Guenther, Kellie O Casavale, Jill Reedy, Sharon I Kirkpatrick, Hazel AB Hiza, Kevin J Kuczynski, Lisa L Kahle, and Susan M Krebs-Smith. Update of the healthy eating index: Hei-2010. *Journal of the Academy of Nutrition and Dietetics*, 113(4):569–580, 2013.
- [85] Awni Hannun, Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, et al. Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567*, 2014.
- [86] L. Harris-Kojetin, M. Sengupta, E. Park-Lee, and R. Valverde. Long-term care services in the united states: 2013 overview. *Vital & Health Statistics*, 3(37), 2013.
- [87] Sandra G Hart. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 50, pages 904–908. Sage publications Sage CA: Los Angeles, CA, 2006.
- [88] Sandra G Hart and Lowell E Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*, volume 52, pages 139–183. Elsevier, 1988.
- [89] Hamid Hassannejad, Guido Matrella, Paolo Ciampolini, Ilaria Munari, Monica Mordonini, and Stefano Cagnoni. A new approach to image-based estimation of food volume. *Algorithms*, 10(2):66, 2017.
- [90] Hongsheng He, Fanyu Kong, and Jindong Tan. Dietcam: multiview food recognition using a multikernel svm. *IEEE Journal of Biomedical and Health Informatics*, 20(3):848–855, 2016.

- [91] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9):1904–1916, 2015.
- [92] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *Proceedings of the European Conference on Computer Vision*, pages 630–645, 2016.
- [93] Ye He, Chang Xu, Nitin Khanna, Carol J Boushey, and Edward J Delp. Food image analysis: segmentation, identification and weight estimation. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, pages 1–6, 2013.
- [94] Ye He, Chang Xu, Nitin Khanna, Carol J Boushey, and Edward J Delp. Analysis of food images: Features and classification. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 2744–2748. IEEE, 2014.
- [95] Georg Hennig, Christian Homann, Ilknur Teksan, Uwe Hasbargen, Stephan Hasmüller, Lesca M Holdt, Nadia Khaled, Ronald Sroka, Thomas Stauch, Herbert Stepp, et al. Non-invasive detection of iron deficiency by fluorescence measurement of erythrocyte zinc protoporphyrin in the lip. *Nature Communications*, 7(1):1–8, 2016.
- [96] David Herzig, Christos T Nakas, Janine Stalder, Christophe Kosinski, Céline Laesser, Joachim Dehais, Raphael Jaeggi, Alexander Benedikt Leichtle, Fried-Michael Dahlweid, Christoph Stettler, et al. Volumetric food quantification using computer vision on a depth-sensing smartphone: Preclinical study. *JMIR mHealth and uHealth*, 8(3):e15294, 2020.
- [97] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.
- [98] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.
- [99] Hajime Hoashi, Taichi Joutou, and Keiji Yanai. Image recognition of 85 food categories by feature fusion. In *2010 IEEE International Symposium on Multimedia*, pages 296–301. IEEE, 2010.
- [100] Kevin Anthony Hoff and Masooda Bashir. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3):407–434, 2015.
- [101] Jess Hohenstein, Dakota O’Dell, Elizabeth L Murnane, Zhengda Lu, David Erickson, and Geri Gay. Enhancing the usability of an optical reader system to support point-of-care rapid diagnostic testing: an iterative design approach. *JMIR Human Factors*, 4(4):e8621, 2017.

- [102] Joanne M Holden and Linda E Lemar. Assessing vitamin d contents in foods and supplements: challenges and needs. *The American Journal of Clinical Nutrition*, 88(2):551S–553S, 2008.
- [103] E Huskisson, S Maggini, and M Ruf. The influence of micronutrients on cognitive function and performance. *Journal of International Medical Research*, 35(1):1–19, 2007.
- [104] Nila Ilhamto, Katrina Anciado, Heather H Keller, and Lisa M Duizer. In-house pureed food production in long-term care: Perspectives of dietary staff and implications for improvement. *Journal of Nutrition in Gerontology and Geriatrics*, 33(3):210–228, 2014.
- [105] Apple Inc. Navigation bars - bars - ios - human interface guidelines - apple developer. <https://developer.apple.com/design/human-interface-guidelines/ios/bars/navigation-bars/>, (accessed: 26.11.2018).
- [106] Bizmatics Inc. Prognosis EHR Software Demo Video. <https://www.youtube.com/watch?v=phzFiyq6d8Q>, (accessed: 26.11.2018).
- [107] Gordana Ispirova, Tome Eftimov, and Barbara Koroušić Seljak. Evaluating missing value imputation methods for food composition databases. *Food and Chemical Toxicology*, 141:111368, 2020.
- [108] Eugenio Ivorra, Samuel Verdú Amat, Antonio J Sánchez, José M Barat, and Raúl Grau. Continuous monitoring of bread dough fermentation using a 3d vision structured light technique. *Journal of Food Engineering*, 130:8–13, 2014.
- [109] Michelle A Jahn, Brian W Porter, Himalaya Patel, Alan J Zillich, Steven R Simon, and Alissa L Russ. Usability assessment of secure messaging for clinical document sharing between health care providers and patients. *Applied Clinical Informatics*, 9(2):467, 2018.
- [110] Wenyan Jia, Hsin-Chen Chen, Yaofeng Yue, Zhaoxin Li, John Fernstrom, Yicheng Bai, Chengliu Li, and Mingui Sun. Accuracy of food portion size estimation from digital pictures acquired by a chest-worn camera. *Public Health Nutrition*, 17(8):1671–1681, 2014.
- [111] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1):53–71, 2000.
- [112] Chong-wah NGO Jing-jing Chen. Deep-based ingredient recognition for cooking recipe retrieval. *ACM Multimedia*, 2016.



- [113] Yvonne Johansson, Margareta Bachrach-Lindström, John Carstensen, and Anna-Christina Ek. Malnutrition in a home-living older population: prevalence, incidence and risk factors. a prospective study. *Journal of Clinical Nursing*, 18(9):1354–1364, 2009.
- [114] Justinmind. 7 rules for mobile ui button design. <https://uxplanet.org/7-rules-for-mobile-ui-button-design-e9cf2ea54556>, (accessed: 26.11.2018).
- [115] Hokuto Kagaya, Kiyoharu Aizawa, and Makoto Ogawa. Food detection and recognition using convolutional neural network. In *Proceedings of the 22nd ACM International Conference on Multimedia*, pages 1085–1088. ACM, 2014.
- [116] Matthias J Kaiser, Jürgen M Bauer, Christiane Rämisch, Wolfgang Uter, Yves Guigoz, Tommy Cederholm, David R Thomas, Patricia S Anthony, Karen E Charlton, Marcello Maggio, et al. Frequency of malnutrition in older adults: a multinational perspective using the mini nutritional assessment. *Journal of the American Geriatrics Society*, 58(9):1734–1738, 2010.
- [117] Parneet Kaur, , Karan Sikka, Weijun Wang, serge Belongie, and Ajay Divakaran. Foodx-251: A dataset for fine-grained food classification. *arXiv preprint arXiv:1907.06167*, 2019.
- [118] Yoshiyuki Kawano and Keiji Yanai. Automatic expansion of a food image dataset leveraging existing categories with domain adaptation. In *European Conference on Computer Vision*, pages 3–17. Springer, 2014.
- [119] Yoshiyuki Kawano and Keiji Yanai. Foodcam-256: A large-scale real-time mobile food recognition system employing high-dimensional features and compression of classifier weights. In *Proceedings of the 22nd ACM International Conference on Multimedia*, pages 761–762, 2014.
- [120] Yoshiyuki Kawano and Keiji Yanai. Foodcam: A real-time food recognition system on a smartphone. *Multimedia Tools and Applications*, 74(14):5263–5287, 2015.
- [121] Heather Keller, Vanessa Vucea, Susan E Slaughter, Harriët Jager-Wittenaar, Christina Lengyel, Faith D Ottery, and Natalie Carrier. Prevalence of malnutrition or risk in residents in long term care: comparison of four tools. *Journal of Nutrition in Gerontology and Geriatrics*, 38(4):329–344, 2019.
- [122] Heather H Keller, Natalie Carrier, Susan E Slaughter, Christina Lengyel, Catriona M Steele, Lisa Duizer, Jill Morrison, K Stephen Brown, Habib Chaudhury, Minn N Yoon, et al. Prevalence and determinants of poor food intake of residents living in long-term care. *Journal of the American Medical Directors Association*, 18(11):941–947, 2017.

- [123] Heather H Keller, Christina Lengyel, Natalie Carrier, Susan E Slaughter, Jill Morrison, Alison M Duncan, Catriona M Steele, Lisa Duizer, K Stephen Brown, Habib Chaudhury, et al. Prevalence of inadequate micronutrient intakes of canadian long-term care residents. *British Journal of Nutrition*, 119(9):1047–1056, 2018.
- [124] Heather H Keller, Truls Østbye, and Richard Goy. Nutritional risk predicts quality of life in elderly community-living Canadians. *The Journals of Gerontology: Series A*, 59(1):M68–M74, 2004.
- [125] Sundas Khan, Lauren McCullagh, Anne Press, Manish Kharche, Andy Schachter, Salvatore Pardo, and Thomas McGinn. Formative assessment and design of a complex clinical decision support tool for pulmonary embolism. *BMJ Evidence-Based Medicine*, 21(1):7–13, 2016.
- [126] Jake Knapp, John Zeratsky, and Braden Kowitz. *Sprint: How to solve big problems and test new ideas in just five days*. Simon and Schuster, 2016.
- [127] Griet Knockaert, Sudheer K Pulissery, Lien Lemmens, Sandy Van Buggenhout, Marc Hendrickx, and Ann Van Loey. Carrot  $\beta$ -carotene degradation and isomerization kinetics during thermal processing in the presence of oil. *Journal of Agricultural and Food Chemistry*, 60(41):10312–10319, 2012.
- [128] Fanyu Kong. *Automatic Food Intake Assessment Using Camera Phones*. PhD thesis, Michigan Technological University, 2012.
- [129] Fanyu Kong, Hongsheng He, Hollie A Raynor, and Jindong Tan. DietCam: multi-view regular shape food recognition with a camera phone. *Pervasive and Mobile Computing*, 19:108–121, 2015.
- [130] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.
- [131] Andre Kushniruk, Yalini Senathirajah, and Elizabeth Borycki. Towards a usability and error “safety net”: A multi-phased multi-method approach to ensuring system usability and safety. In *MEDINFO 2017: Precision Healthcare through Informatics*, pages 763–767. IOS Press, 2017.
- [132] Jane E Lancaster, Carolyn E Lister, Peter F Reay, and Christopher M Triggs. Influence of pigment composition on skin color in a wide range of fruit and vegetables. *Journal of the American Society for Horticultural Science*, 122(4):594–598, 1997.
- [133] David P Lanctin, Francheska Merced-Nieves, Renee M Mallett, Mary Beth Arensberg, Peggi Guenter, Suela Sulo, and Timothy F Platts-Mills. Prevalence and economic burden

- of malnutrition diagnosis among patients presenting to united states emergency departments. *Academic Emergency Medicine*, 28(3):325–335, 2021.
- [134] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew D Back. Face recognition: A convolutional neural-network approach. *IEEE Transactions on Neural Networks*, 8(1):98–113, 1997.
- [135] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [136] Yann LeCun, Fu Jie Huang, and Leon Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–104. IEEE, 2004.
- [137] G. G. Lee, C. Huang, J. Chen, S. Chen, and H. Chen. Aifood: A large scale food images dataset for ingredient recognition. In *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, pages 802–805, 2019.
- [138] John D Lee and Katrina A See. Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1):50–80, 2004.
- [139] Hsien-Chou Liao, Zi-Yi Lim, and Hua-Wei Lin. Food intake estimation method using short-range depth camera. In *2016 IEEE International Conference on Signal and Image Processing (ICSIP)*, pages 198–204. IEEE, 2016.
- [140] Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, and Yunsheng Ma. Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment. In *International Conference on Smart Homes and Health Telematics*, pages 37–48. Springer, 2016.
- [141] Chenxi Liu, Liang-Chieh Chen, Florian Schroff, Hartwig Adam, Wei Hua, Alan L Yuille, and Li Fei-Fei. Auto-deeplab: hierarchical neural architecture search for semantic image segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 82–92, 2019.
- [142] Frank P-W Lo, Yingnan Sun, Jianing Qiu, and Benny Lo. Food volume estimation based on deep learning view synthesis from a single depth map. *Nutrients*, 10(12):2005, 2018.
- [143] Frank P-W Lo, Yingnan Sun, Jianing Qiu, and Benny PL Lo. Point2volume: A vision-based dietary assessment approach using view synthesis. *IEEE Transactions on Industrial Informatics*, 16(1):577–586, 2019.

- [144] Frank Po Wen Lo, Yingnan Sun, Jianing Qiu, and Benny Lo. Image-based food classification and volume estimation for dietary assessment: A review. *IEEE Journal of Biomedical and Health Informatics*, 24(7):1926–1939, 2020.
- [145] Yaowei Long, Yuhao Wang, Ziming Zhai, Li Wu, Minzan Li, Hong Sun, and Qinghua Su. Potato volume measurement based on rgb-d camera. *IFAC-PapersOnLine*, 51(17):515–520, 2018.
- [146] Angel Lopez-Candales, Paula M Hernández Burgos, Dagmar F Hernandez-Suarez, and David Harris. Linking chronic inflammation with cardiovascular disease: from normal aging to the metabolic syndrome. *Journal of Nature and Science*, 3(4), 2017.
- [147] Yuzhen Lu and Renfu Lu. Using composite sinusoidal patterns in structured-illumination reflectance imaging (siri) for enhanced detection of apple bruise. *Journal of Food Engineering*, 199:54–64, 2017.
- [148] Yuzhen Lu and Renfu Lu. Structured-illumination reflectance imaging coupled with phase analysis techniques for surface profiling of apples. *Journal of Food Engineering*, 232:11–20, 2018.
- [149] Ameersing Luximon and Ravindra S Goonetilleke. Simplified subjective workload assessment technique. *Ergonomics*, 44(3):229–243, 2001.
- [150] Joseph B Lyons and Charlene K Stokes. Human–human reliance in the context of automation. *Human Factors*, 54(1):112–121, 2012.
- [151] BM Margetts, RL Thompson, M Elia, and AA Jackson. Prevalence of risk of undernutrition is associated with poor health status in older people in the uk. *European Journal of Clinical Nutrition*, 57(1):69–74, 2003.
- [152] Anand Mariappan, Marc Bosch, Fengqing Zhu, Carol J Boushey, Deborah A Kerr, David S Ebert, and Edward J Delp. Personal dietary assessment using mobile devices. In *Computational Imaging VII*, volume 7246, page 72460Z. International Society for Optics and Photonics, 2009.
- [153] Javier Marin, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, and Antonio Torralba. Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, 2019.
- [154] Corby K Martin, Hongmei Han, Sandra M Coulon, H Raymond Allen, Catherine M Champagne, and Stephen D Anton. A novel method to remotely measure food intake of free-living individuals in real time: the remote food photography method. *British Journal of Nutrition*, 101(3):446–456, 2008.

- [155] Niki Martinel, Claudio Piciarelli, and Christian Micheloni. A supervised extreme learning committee for food recognition. *Computer Vision and Image Understanding*, 148:67–86, 2016.
- [156] Yuji Matsuda, Hajime Hoashi, and Keiji Yanai. Recognition of multiple-food images by detecting candidate regions. In *2012 IEEE International Conference on Multimedia and Expo*, pages 25–30. IEEE, 2012.
- [157] Raymond G McGuire. Reporting of objective color measurements. *HortScience*, 27(12):1254–1255, 1992.
- [158] MedPass. Dietary intake form. [http://www.med-pass.com/media/pdf/CP1717\\_sp.pdf](http://www.med-pass.com/media/pdf/CP1717_sp.pdf), (accessed: 29.06.2017).
- [159] Antonio J Meléndez-Martínez, George Britton, Isabel M Vicario, and Francisco J Heredia. Relationship between the colour and the chemical structure of carotenoid pigments. *Food Chemistry*, 101(3):1145–1150, 2007.
- [160] Domingo Mery and Franco Pedreschi. Segmentation of colour food images using a robust algorithm. *Journal of Food Engineering*, 66(3):353–360, 2005.
- [161] Cade Metz. In major AI breakthrough, Google system secretly beats top player at the ancient game of go. <https://www.wired.com/2016/01/in-a-huge-breakthrough-googles-ai-beats-a-top-player-at-the-game-of-g> (accessed: 06.04.2016).
- [162] Austin Meyers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin P Murphy. Im2calories: towards an automated mobile vision food diary. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1233–1241, 2015.
- [163] Austin Meyers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin P Murphy. Im2calories: towards an automated mobile vision food diary. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1233–1241, 2015.
- [164] Weiqing Min, Linhu Liu, Zhiling Wang, Zhengdong Luo, Xiaoming Wei, Xiaolin Wei, and Shuqiang Jiang. *ISIA Food-500: A Dataset for Large-Scale Food Recognition via Stacked Global-Local Attention Network*, page 393–401. Association for Computing Machinery, New York, NY, USA, 2020.
- [165] Deanna M Minich. A review of the science of colorful, plant-based food and practical strategies for “eating the rainbow”. *Journal of Nutrition and Metabolism*, 2019, 2019.

- [166] Tatsuya Miyazaki, Gamhewage C de Silva, and Kiyoharu Aizawa. Image-based calorie content estimation for dietary assessment. In *Multimedia (ISM), 2011 IEEE International Symposium on*, pages 363–368. IEEE, 2011.
- [167] ZJ Nagykaldi, M Jordan, J Quitariano, CA Ciro, and JW Mold. User-centered design and usability testing of an innovative health-related quality of life module. *Applied Clinical Informatics*, 5(4):958, 2014.
- [168] National Institutes of Health. Vitamin A Fact Sheet for Health Professionals. <https://ods.od.nih.gov/factsheets/VitaminA-HealthProfessional/>, (accessed: 05.03.2019).
- [169] Jakob Nielsen. *Usability engineering*. Morgan Kaufmann, 1994.
- [170] Jakob Nielsen. 10 usability heuristics for user interface design. <https://www.nngroup.com/articles/ten-usability-heuristics/>, (accessed: 24.08.2021).
- [171] Don Norman. *The design of everyday things: Revised and expanded edition*. Basic books, 2013.
- [172] Geoffrey R Norman and David L Streiner. *Biostatistics: The bare essentials*. PMPH-USA, 2008.
- [173] United States Department of Agriculture Agricultural Research Service. National nutrient database: 11124, carrots, raw. <https://ndb.nal.usda.gov/ndb/foods/show/11124>, (accessed: 05.03.2019).
- [174] Kwabena T Ofei, Bent E Mikkelsen, and Rudolf A Scheller. Validation of a novel image-weighted technique for monitoring food intake and estimation of portion size in hospital settings: a pilot study. *Public Health Nutrition*, pages 1–6, 2018.
- [175] Kwabena T Ofei, Bent E Mikkelsen, and Rudolf A Scheller. Validation of a novel image-weighted technique for monitoring food intake and estimation of portion size in hospital settings: a pilot study. *Public Health Nutrition*, 22(7):1203–1208, 2019.
- [176] Koichi Okamoto and Keiji Yanai. An automatic calorie estimation system of food images on a smartphone. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, pages 63–70. ACM, 2016.
- [177] Maxim Parent, Helen Niezgodna, Heather H Keller, Larry W Chambers, and Shauna Daly. Comparison of visual estimation methods for regular and modified textures: real-time vs digital imaging. *Journal of the Academy of Nutrition and Dietetics*, 112(10):1636–1641, 2012.

- [178] Linda Penn, Heiner Boeing, Carol J Boushey, Lars Ove Dragsted, Jim Kaput, Augustin Scalbert, Ailsa A Welch, and John C Mathers. Assessment of dietary intake: NuGO symposium report. *Genes & Nutrition*, 5(3):205, 2010.
- [179] Rodrigo Méndez Perez, Fernando Auat Cheein, and Joan R Rosell-Polo. Flexible system of multiple rgb-d sensors for measuring and classifying fruits in agri-food industry. *Computers and Electronics in Agriculture*, 139:231–242, 2017.
- [180] Kaylen Pfisterer, Jennifer Boger, and Alexander Wong. Prototyping the automated food imaging and nutrient intake tracking (AFINI-T) system: A modified participatory iterative design sprint. *JMIR Human Factors*, 6(2):e13017, May 2019.
- [181] Kaylen J Pfisterer, Robert Amelard, Audrey G Chung, and Alexander Wong. A new take on measuring relative nutritional density: The feasibility of using a deep neural network to assess commercially-prepared puréed food concentrations. *Journal of Food Engineering*, 223:220–235, 2018.
- [182] Kaylen J. Pfisterer, Robert Amelard, Braeden Syrnyk, and Alexander Wong. Towards computer vision powered color-nutrient assessment of pureed food. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.
- [183] Kaylen J Pfisterer, Robert Amelard, and Alexander Wong. Differential color space analysis for investigating nutrient content in a puréed food dilution-flavor matrix: a step toward objective malnutrition risk assessment. In *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics*, volume 10501, page 105010J. International Society for Optics and Photonics, 2018.
- [184] Matthias Pirlich and Herbert Lochs. Nutrition in the elderly. *Best Practice & Research Clinical Gastroenterology*, 15(6):869–884, 2001.
- [185] Carolien E Postma, Elly Zwartkruis-Pelgrim, Elke Daemen, and Jia Du. Challenges of doing empathic design: Experiences from industry. *International Journal of Design*, 6(1), 2012.
- [186] Parisa Pouladzadeh, Pallavi Kuhad, Sri Vijay Bharat Peddi, Abdulsalam Yassine, and Shervin Shirmohammadi. Food calorie measurement using deep learning neural network. In *Proceedings of the IEEE International Instrumentation and Measurement Technology*, pages 1–6, 2016.
- [187] Parisa Pouladzadeh, Shervin Shirmohammadi, and Rana Al-Maghrabi. Measuring calorie and nutrition from food image. *IEEE Transactions on Instrumentation and Measurement*, 63(8):1947–1956, 2014.

- [188] Parisa Pouladzadeh, Shervin Shirmohammadi, Aslan Bakirov, Ahmet Bulut, and Abdulsalam Yassine. Cloud-based svm for food categorization. *Multimedia Tools and Applications*, 74(14):5243–5260, 2015.
- [189] Parisa Pouladzadeh, Shervin Shirmohammadi, and Abdulsalam Yassine. You are what you eat: So measure what you eat! *IEEE Instrumentation & Measurement Magazine*, 19(1):9–15, 2016.
- [190] Manika Puri, Zhiwei Zhu, Qian Yu, Ajay Divakaran, and Harpreet Sawhney. Recognition and volume estimation of food intake using a mobile device. In *2009 Workshop on Applications of Computer Vision (WACV)*, pages 1–8. IEEE, 2009.
- [191] Jianwei Qin and Renfu Lu. Measurement of the optical properties of fruits and vegetables using spatially resolved hyperspectral diffuse reflectance imaging technique. *Postharvest Biology and Technology*, 49(3):355–365, 2008.
- [192] Taha M Rababah, Khalil I Ereifej, and L Howard. Effect of ascorbic acid and dehydration on concentrations of total phenolics, antioxidant capacity, anthocyanins, and color in fruits. *Journal of Agricultural and Food Chemistry*, 53(11):4444–4447, 2005.
- [193] Laavanya Rachakonda, Saraju P Mohanty, and Elias Kougianos. ilog: an intelligent device for automatic food intake monitoring and stress detection in the iomt. *IEEE Transactions on Consumer Electronics*, 66(2):115–124, 2020.
- [194] Md Hafizur Rahman, Qiang Li, Mark Pickering, Michael Frater, Deborah Kerr, Carol Bouchev, and Edward Delp. Food volume estimation in a mobile phone based dietary assessment system. In *2012 Eighth International Conference on Signal Image Technology and Internet Based Systems*, pages 988–995. IEEE, 2012.
- [195] Jayant V Rajan, Juliana Moura, Gato Gourley, Karina Kiso, Alexandre Sizilio, Ana Maria Cortez, Lee W Riley, Maria Amelia Veras, and Urmimala Sarkar. Understanding the barriers to successful adoption and use of a mobile health information system in a community health center in são paulo, brazil: a cohort study. *BMC Medical Informatics and Decision Making*, 16(1):1–11, 2016.
- [196] Susannah Ravden and Graham Johnson. *Evaluating usability of human-computer interfaces: a practical method*. Halsted Press, 1989.
- [197] Christina Reginaldo, Hathairat Sawaengsri, Tammy Scott, Irwin Rosenberg, Jacob Selhub, and Ligi Paul. The association between vitamin B6 and cognitive decline is modified by inflammatory state. *The FASEB Journal*, 28(1\_supplement):LB425, 2014.
- [198] Mijke Rhemtulla, Patricia É Brosseau-Liard, and Victoria Savalei. When can categorical variables be treated as continuous? a comparison of robust continuous and categorical



- sem estimation methods under suboptimal conditions. *Psychological methods*, 17(3):354, 2012.
- [199] S Roberts, AP Marshall, R Gonzalez, and W Chaboyer. Technology to engage hospitalised patients in their nutrition care: a qualitative study of usability and patient perceptions of an electronic foodservice system. *Journal of Human Nutrition and Dietetics*, 30(5):563–573, 2017.
- [200] M Pope Robin and S Fry Edward. Absorption spectrum of pure water (380–700 nm). ii. integrating cavity measurements. *Applied Optics*, 36(33):8710–8723, 1997.
- [201] Anderson Rocha, Daniel C Hauage, Jacques Wainer, and Siome Goldenstein. Automatic fruit and vegetable classification from images. *Computers and Electronics in Agriculture*, 70(1):96–104, 2010.
- [202] Christine A Russell. The impact of malnutrition on healthcare costs and economic considerations for the use of oral nutritional supplements. *Clinical Nutrition Supplements*, 2(1):25–32, 2007.
- [203] Liz Sanders. On modeling an evolving map of design practice and design research. *Interactions*, 15(6):13–17, 2008.
- [204] A Sass-Kiss, J Kiss, P Milotay, MM Kerek, and M Toth-Markus. Differences in anthocyanin and carotenoid content of fruits and vegetables. *Food Research International*, 38(8):1023–1029, 2005.
- [205] Junqing Shang, Michael Duong, Eric Pepin, Xing Zhang, Kishore Sandara-Rajan, Alexander Mamishev, and Alan Kristal. A mobile structured light system for food volume estimation. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 100–101. IEEE, 2011.
- [206] Junqing Shang, Eric Pepin, Eric Johnson, David Hazel, Ankur Teredesai, Alan Kristal, and Alexander Mamishev. Dietary intake assessment using integrated sensors and software. In *Multimedia on Mobile Devices 2012; and Multimedia Content Access: Algorithms and Systems VI*, volume 8304, page 830403. International Society for Optics and Photonics, 2012.
- [207] Wataru Shimoda and Keiji Yanai. CNN-based food image segmentation without pixel-wise annotation. In *Proceedings of the International Conference on Image Analysis and Processing*, pages 449–457, 2015.
- [208] Parisa Pouladzadeh; Abdulsalam Yassine; Shervin Shirmohammadi. Foodd: Food detection dataset for calorie measurement using food images. <https://ieee-dataport.org/open-access/>

[foodd-food-detection-dataset-calorie-measurement-using-food-images#files](#), 2020.

- [209] Nitin Shivappa, Susan E Steck, Thomas G Hurley, James R Hussey, and James R Hébert. Designing and developing a literature-derived, population-based dietary inflammatory index. *Public Health Nutrition*, 17(8):1689–1696, 2014.
- [210] Ben Shneiderman and Catherine Plaisant. *Designing the user interface: Strategies for effective human-computer interaction*. Pearson Education India, 2010.
- [211] Sandra F Simmons and David Reuben. Nutritional intake monitoring for nursing home residents: a comparison of staff documentation, direct observation, and photography methods. *Journal of the American Geriatrics Society*, 48(2):209–213, 2000.
- [212] Sandra F Simmons and John F Schnelle. Feeding assistance needs of long-stay nursing home residents and staff time to provide care. *Journal of the American Geriatrics Society*, 54(6):919–924, 2006.
- [213] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [214] Philip D Sloane, Jena Ivey, Margaret Helton, Ann Louise Barrick, and Ana Cerna. Nutritional issues in long-term care. *Journal of the American Medical Directors Association*, 9(7):476–485, 2008.
- [215] Software Advice. Top electronic medical records software - 2018 reviews. <http://www.webcitation.org/query?url=https%3A%2F%2Fwww.softwareadvice.com%2Fca%2Fmedical%2Felectronic-medical-record-software-comparison&date=2018-11-26>, (accessed: 26.11.2018).
- [216] Neville A Stanton, Paul M Salmon, Guy H Walker, Chris Baber, and Daniel P Jenkins. *Human factors methods: a practical guide for engineering and design*. CRC Press, 2017.
- [217] Neville A Stanton, Mark S Young, and Catherine Harvey. *Guide to methodology in ergonomics: Designing for human use*. CRC Press, 2014.
- [218] Mohammed Ahmed Subhi, Sawal Hamid Ali, and Mohammed Abulameer Mohammed. Vision-based approaches for automatic food recognition and dietary assessment: A survey. *IEEE Access*, 7:35370–35381, 2019.
- [219] Livia Sura, Aarthi Madhavan, Giselle Carnaby, and Michael A Crary. Dysphagia in the elderly: Management and nutritional considerations. *Clinical Interventions in Aging*, 7(287):98, 2012.

- [220] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.
- [221] F Teoli, S Luciola, P Nota, A Frattarelli, F Matteocci, A Di Carlo, E Caboni, and C Forni. Role of ph and pigment concentration for natural dye-sensitized solar cells treated with anthocyanin extracts of common fruits. *Journal of Photochemistry and Photobiology A: Chemistry*, 316:24–30, 2016.
- [222] David R Thomas, Wendy Ashmen, John E Morley, and William J Evans. Nutritional management in long-term care: development of a clinical guideline. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 55(12):M725–M734, 2000.
- [223] Frances E Thompson, Amy F Subar, Catherine M Loria, Jill L Reedy, and Tom Baranowski. Need for technological innovation in dietary assessment. *Journal of the American Dietetic Association*, 110(1):48, 2010.
- [224] Alexander Toet, Daisuke Kaneko, Inge de Kruijf, Shota Ushiyama, Martin G van Schaik, Anne-Marie Brouwer, Victor Kallen, and Jan BF Van Erp. Crocufid: a cross-cultural food image database for research on food elicited affective responses. *Frontiers in Psychology*, 10:58, 2019.
- [225] Sana Tonekaboni, Shalmali Joshi, Melissa D McCradden, and Anna Goldenberg. What clinicians want: contextualizing explainable machine learning for clinical end use. In *Machine Learning for Healthcare Conference*, pages 359–380. PMLR, 2019.
- [226] United States Department of Agriculture Agricultural Research Service. National nutrient database: 11507, sweet potato, raw, unprepared. <https://ndb.nal.usda.gov/ndb/foods/show/11507>, (accessed: 05.03.2019).
- [227] Michael Unser. Sum and difference histograms for texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1):118–125, 1986.
- [228] Ondine van de Rest, Agnes AM Berendsen, Annemien Haveman-Nies, and Lisette CPGM de Groot. Dietary patterns, cognitive decline, and dementia: a systematic review. *Advances in Nutrition*, 6(2):154–168, 2015.
- [229] Samuel Verdú, Eugenio Ivorra, Antonio J Sánchez, Jose M Barat, and Raúl Grau. Relationship between fermentation behavior, measured with a 3d vision structured light technique, and the internal structure of bread. *Journal of Food Engineering*, 146:227–233, 2015.
- [230] MA Vidulich and PS Tsang. Techniques of subjective workload assessment: a comparison of swat and the nasa-bipolar methods. *Ergonomics*, 29(11):1385–1398, 1986.

- [231] Dorothee Volkert, Michael Chourdakis, Gerd Faxen-Irving, Thomas Frühwald, Francesco Landi, Merja H Suominen, Maurits Vandewoude, Rainer Wirth, and Stéphane M Schneider. Espen guidelines on nutrition in dementia. *Clinical nutrition*, 34(6):1052–1073, 2015.
- [232] Vanessa Vucea. Modified texture diet and long term care: A secondary data analysis of making the most of mealtimes (m3) project. Master’s thesis, University of Waterloo, 2017.
- [233] Vanessa Vucea, Heather H Keller, and Kate Ducak. Interventions for improving meal-time experiences in long-term care. *Journal of Nutrition in Gerontology and Geriatrics*, 33(4):249–324, 2014.
- [234] Vanessa Vucea, Heather H Keller, Jill M Morrison, Lisa M Duizer, Alison M Duncan, and Catriona M Steele. Prevalence and characteristics associated with modified texture food use in long term care: An analysis of making the most of mealtimes (m3) project. *Canadian Journal of Dietetic Practice and Research*, 80(3):104–110, 2019.
- [235] Vanessa Vucea, Heather H Keller, Jill M Morrison, Alison M Duncan, Lisa M Duizer, Natalie Carrier, Christina O Lengyel, and Susan E Slaughter. Nutritional quality of regular and pureed menus in canadian long term care homes: an analysis of the making the most of mealtimes (m3) project. *BMC nutrition*, 3(1):1–11, 2017.
- [236] Edith G Walsh, Joshua M Wiener, Susan Haber, Arnold Bragg, Marc Freiman, and Joseph G Ouslander. Potentially avoidable hospitalizations of dually eligible medicare and medicaid beneficiaries from nursing facility and home-and community-based services waiver programs. *Journal of the American Geriatrics Society*, 60(5):821–829, 2012.
- [237] Yu Wang, Ye He, Carol J Boushey, Fengqing Zhu, and Edward J Delp. Context based image analysis with application in dietary assessment and evaluation. *Multimedia tools and applications*, 77(15):19769–19794, 2018.
- [238] Yu Wang, Fengqing Zhu, Carol J Boushey, and Edward J Delp. Weakly supervised food image segmentation using class activation maps. In *Proceedings of the IEEE International Conference on Image Processing*, pages 1277–1281, 2017.
- [239] Barbara E. Wendland. Malnutrition in Institutionalized Seniors: The Iatrogenic Component. *Journal of the American Geriatrics Society*, 51(1):85–90, 2003.
- [240] Donald A Williamson, H Raymond Allen, Pamela Davis Martin, Anthony J Alfonso, Bonnie Gerald, and Alice Hunt. Comparison of digital photography to weighed and visual estimation of portion sizes. *Journal of the American Dietetic Association*, 103(9):1139–1145, 2003.
- [241] Jacob O Wobbrock and Julie A Kientz. Research contributions in human-computer interaction. *Interactions*, 23(3):38–44, 2016.

- [242] Ronald E Wrolstad et al. Color and pigment analyses in fruit products. Technical report, Corvallis, Or.: Agricultural Experiment Station. Oregon State University., 1993.
- [243] Chang Xu, Ye He, Nitin Khanna, Carol J. Boushey, and Edward J. Delp. Model-based food volume estimation using 3d pose. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 2534–2538. IEEE, 2013.
- [244] Chang Xu, Ye He, Nitin Khanna, Albert Parra, Carol Boushey, and Edward Delp. Image-based food volume estimation. In *Proceedings of the 5th international workshop on Multimedia for cooking & eating activities*, pages 75–80. ACM, 2013.
- [245] Shulin Yang, Mei Chen, Dean Pomerleau, and Rahul Sukthankar. Food recognition using statistics of pairwise local features. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2249–2256. IEEE, 2010.
- [246] Raza Yunus, Omar Arif, Hammad Afzal, Muhammad Faisal Amjad, Haider Abbas, Hira Noor Bokhari, Syeda Tazeen Haider, Nauman Zafar, and Raheel Nawaz. A framework to estimate the nutritional value of food in real time using deep learning techniques. *IEEE Access*, 7:2643–2652, 2018.
- [247] Yudong Zhang and Lenan Wu. Classification of fruits using computer vision and a multi-class support vector machine. *Sensors*, 12(9):12489–12505, 2012.
- [248] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [249] Xin Zheng, Qinyi Lei, Run Yao, Yifei Gong, and Qian Yin. Image segmentation based on adaptive k-means algorithm. *EURASIP Journal on Image and Video Processing*, 2018(1):68, 2018.
- [250] Lei Zhou, Chu Zhang, Fei Liu, Zhengjun Qiu, and Yong He. Application of deep learning in food: a review. *Comprehensive Reviews in Food Science and Food Safety*, 18(6):1793–1811, 2019.
- [251] Fengqing Zhu, Marc Bosch, Nitin Khanna, Carol J Boushey, and Edward J Delp. Multi-level segmentation for food classification in dietary assessment. In *2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 337–342. IEEE, 2011.
- [252] Fengqing Zhu, Marc Bosch, Nitin Khanna, Carol J Boushey, and Edward J Delp. Multiple hypotheses image segmentation and classification with application to dietary assessment. *IEEE Journal of Biomedical and Health Informatics*, 19(1):377–388, 2015.

[253] FP Zscheile, Jonathan W White Jr, BW Beadle, and JR Roach. The preparation and absorption spectra of five pure carotenoid pigments. *Plant Physiology*, 17(3):331, 1942.