Journal of Food Engineering xxx (2017) 1–16

Contents lists available at ScienceDirect

Journal of Food Engineering

journal homepage: www.elsevier.com/locate/jfoodeng

A new take on measuring relative nutritional density: The feasibility of using a deep neural network to assess commercially-prepared puréed food concentrations

Kaylen J. Pfisterer [a, b, *], Robert Amelard [a, b], Audrey G. Chung [a], Alexander Wong [a, b]

[a] University of Waterloo, Department of Systems Design Engineering, N2L 3G1, Canada
[b] Schlegel-UW Research Institute for Aging, Waterloo, N2J 0E2, Canada

A R T I C L E I N F O

Article history:
Received 15 June 2017
Received in revised form
18 October 2017
Accepted 22 October 2017
Available online xxx

Keywords:
Nutrient sensing
Deep learning
Image processing
Puréed food

A B S T R A C T

Dysphagia affects 590 million people worldwide and increases risk for malnutrition. Puréed food may reduce choking, however preparation differences impact nutrient density making quality assurance necessary. This paper is the first study to investigate the feasibility of computational puréed food nutritional density analysis using an imaging system. Motivated by a theoretical optical dilution model, a novel deep neural network (DNN) was evaluated using 390 samples from thirteen types of commercially prepared purées at five dilutions. The DNN predicted relative concentration of the purée sample (20%, 40%, 60%, 80%, 100% initial concentration). Data were captured using same-side reflectance of multispectral imaging data at different polarizations at three exposures. Experimental results yielded an average top-1 prediction accuracy of 92.2% ± 0.41% with sensitivity and specificity of 83.0% ± 15.0% and 95.0% ± 4.8%, respectively. This DNN imaging system for nutrient density analysis of puréed food shows promise as a novel tool for nutrient quality assurance.

1. Introduction

Dysphagia (swallowing difficulty) affects approximately 590 million people worldwide (Cichero et al., 2016) and at least 15% of American older adults (Sura et al., 2012) increasing these individuals' risk for malnutrition (Ilhamto et al., 2014; Sura et al., 2012). Malnutrition impacts quality of life (Keller et al., 2004) and accounts for significant annual burden to the health care system of approximately $15.5 billion in the United States (Goates et al., 2016) and 7.3 billion in the UK (Russell, 2007). Modified texture diets (e.g., puréed food) have been used to allow safe ingestion of nutritional requirements in this population (Germain et al., 2006). However, based on differences in preparation methods, nutrient composition can be highly variable (Ilhamto et al., 2014). This has practical implications especially for older adults with a generally lower intake; food must be as nutritious as possible to ensure adequate nutrient consumption. Additionally, purée thickness has safety implications; too thin a purée may cause choking (Ilhamto et al., 2014). Thus, puréed food quality assurance is required (Ilhamto et al., 2014).

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) (Cichero et al., 2016). 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 (Ilhamto et al., 2014). 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.

Optical imaging systems provide a powerful solution to this problem. Specifically, these systems use the same type of information (visible optics); however, computational models provide

* Corresponding author.
E-mail address: kpfisterer@uwaterloo.ca (K.J. Pfisterer).

https://doi.org/10.1016/j.jfoodeng.2017.10.016
0260-8774/

objective and repeatable predictions. 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 (Bigio and Fantini, 2016). Though primarily used in biomedical applications, these models provide a theoretical basis for quantitative nutritional assessment using optical imaging data. Additionally, recent advances in machine learning have been successfully applied to a vast range of fields from object recognition to pharmacy and genomics (LeCun et al., 2015). Specifically, deep neural networks (DNNs) are biologically inspired by the visual cortex for decision making (Bengio, 2009), and have been used with great success for specific complex tasks such as speech recognition (Hinton et al., 2012; Dahl et al., 2012; Hannun et al., 2014), object recognition (Krizhevsky et al., 2012; He et al., 2015; LeCun et al., 2004; Simonyan and Zisserman, 2014), and natural language processing (Bengio et al., 2003; Collobert and Weston, 2008). In image classification and other applications, however, there is often insufficient training data to properly train a conventional DNN due to 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.

In this paper, we assess the feasibility of computational nutritional density analysis using an imaging system to provide feedback without the need for human assessor input. This preliminary 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 prepare a dilution series to observe the effect of relative increased water content on optical properties (color information, texture information, satruation etc.) for the purpose of determining the feasibility of using an optical imaging techniques for discrimination. Instead of traditional supervised learning, we use stacked autoencoders with a final softmax layer for dilution classification (i.e., discriminating between 20%, 40%, 60%, 80%, 100% initial concentration). Autoencoders are DNNs that leverage unsupervised learning to provide a robust solution that is generalizable and extensible without compromising performance to complete a specific task. Specifically, this is the first study to our knowledge to assess the feasibility of using machine learning (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. In particular, 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 multispectral 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.

2. Material and methods

2.1. Sample preparation

Thirteen commercially-prepared purée flavours across three food categories were selected for this study: fruit (apple, apricot, banana, blueberry, mango, strawberry), meat (beef, chicken), and vegetables (carrot, butternut squash, parsnip, pea, sweet potato). Purée flavours were selected to maximize variations in texture and color. For each purée, a five tier dilution series was prepared relative to initial concentration: 20% (most diluted), 40%, 60%, 80%, and 100% (not diluted). For each dilution in the series, six 5 mL samples were systematically loaded onto a standardized transparency sheet grid from approximately one centimeter above the sheet at room temperature and imaged immediately, yielding a total of 390 samples.

2.2. Data acquisition

Same-side reflectance was used (i.e., the light source and camera were positioned at the same location). A DSLR camera (Canon T4i) was used for high resolution image capture in the visible spectrum with consistent white balancing, aperture, and exposure settings. Both unpolarized and linearly polarized data were acquired by positioning an oriented linear polarizer in front of the camera lens. The use of polarization provided higher variability of a purée's appearance by focusing on surface-level texture (horizontal polarization) and color (vertical polarization) information. To simulate various lighting conditions, three exposures were acquired (1/20 s, 1/10 s, and 1/5 s) for each polarization. These variations enable the system to learn more robust concepts about the purées. Over the course of imaging, the room temperature varied from 21.9 °C to 23.9 °C.

2.3. 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 color and texture. It is important to note that color and texture are meant only to provide intuition into the data collected and were not used as hand-crafted features; features used for distinguishing between classes (classification) were automatically learned given no priors through the deep neural network (see Section 2.5 for more details). Fig. 3 provides a summary of color and texture across the samples. The images in Fig. A.3 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.

2.4. 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, 100×200 pixel subimages (one for each sample on the sheet). As indicated in Fig. 1, each three channel (i.e., RGB) subimage was decomposed into fifty-four patches using half overlapping windows of 50×100 pixels. Rectangular patches were selected to improve the variance observed within a patch. These patches were downscaled to 50% of their original size (25×50) using bicubic