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

Kaylen J. Pfisterer\textsuperscript{a,b,}\textsuperscript{*}, Robert Amelard\textsuperscript{a,b}, Audrey G. Chung\textsuperscript{a}, Alexander Wong\textsuperscript{a,b}

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

Abstract

Dysphagia affects 590 million people worldwide and increases risk for malnutrition. Puréeed 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 puréeed 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 puréeed food shows promise as a novel tool for nutrient quality assurance.

Keywords: nutrient sensing, deep learning, image processing, puréeed food

1. Introduction

Dysphagia (swallowing difficulty) affects approximately 590 million people worldwide \cite{1} and at least 15\% of American older adults \cite{2} increasing these individuals’ risk for malnutrition \cite{3,2}. Malnutrition impacts quality of life \cite{4} and accounts for significant annual burden to the health care system of approximately $15.5$ billion in the United States \cite{5} and 7.3 billion in the UK \cite{6}. Modified texture diets (e.g., puréeed food) have been used to allow safe ingestion of nutritional requirements in this population \cite{7}. However, based on differences in
preparation methods, nutrient composition can be highly variable [3]. 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 [3]. Thus, puréed food quality assurance is required [3].

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) [1]. 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 [2]. 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 healthcare 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 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 [8]. 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 [9]. Specifically, deep neural networks (DNNs) are biologically inspired by the visual cortex for decision making [10], and have been used with great success for specific complex tasks such as speech recognition [11, 12, 13], object recognition [14, 15, 16, 17], and natural language processing [18, 19]. 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. 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 flavors 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 flavors 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). Figure 3 provides a summary of color and texture across the samples. The images in Figure 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 Figure 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 interpolation. The three RGB channels were concatenated to yield 378, 25×50×3 (or 75×50) pixel patches for processing for a given dilution for a specific purée flavor and 1890 patches for a specific purée flavor. Therefore, the final set of images consisted of 13,230 auto-labeled patches.

2.5. Network architecture

Images were then passed into a deep neural network (DNN), consisting of two layers of pretrained stacked autoencoders and a final softmax activation layer. At a high level, five global, general networks were formed using randomly initiated weights and passing through all of the unlabeled patches (i.e., there was no flavor or dilution information provided to the system). These general networks were then fine-tuned using flavor-specific labeled data. Given a specific flavor, the system predicted the dilution class to which a patch from an image belonged.

The autoencoders were implemented as feed-forward fully-connected implementations which use a logistic sigmoid function as a transfer function for non-linearity. The input layer consisted of 3750 input neurons from image dimensions 50×75. Similar to the process described by Hinton et al. [20] the first autoencoder layer was pretrained to receive the high-dimensional data from input layer and convert it to lower-dimensional data with 100 outputs via a learned nonlinear mapping. This large reduction in dimensionality was desirable based on the relatively basic characteristics of the data (e.g., texture and color). The output from the first autoencoder was passed into the second autoencoder in which dimensionality was further reduced to 50 outputs via a second learned nonlinear mapping, providing a “distilled” feature set and back propagation was performed after training. Following the unsupervised learning from the
Figure 1: Deep neural network architecture. A subimage is decomposed into a series of patches which are then downscaled to half their original dimensions (100×50 –>50×25). RGB channels for each patch are concatenated to create one 50×75 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 flavor, the network output is one of five dilution classes (20%, 40%, 60%, 80%, and 100%).

stacked autoencoder, one final softmax layer was leveraged as a means to map the automatically learned “distilled” features onto one of five output classes (20%, 40%, 60%, 80%, or 100% initial concentration) as the top-1 hit (i.e., the one class with the strongest prediction) class label. Finally, the autoencoders and softmax layer were stacked together to form a general deep network where each autoencoder operates as a discrete feature extractor [20]. This forced the network to learn increasingly higher level features. Running a final iteration of backpropagation across the whole system (i.e., the fine-tuning phase), we make use of the retained features that differentiate between classes [20] (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 flavor separately. This final fine-tuning step is the only iteration which uses the labeled, flavor-specific data. As a result, no hand-crafted features were used for the purpose of distinguishing between classes; all features were automatically discovered through the use of the stacked autoencoders.

2.6. Validating our network: Pretraining and testing the network

For each flavor-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 backpropagation across the entire system for further fine-tuning of the weights. This was repeated five times for each flavor-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 flavor.

2.7. Comparative analyses

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 color (64 features) and texture (seven features) characteristics \[21\]. For further formulations of features, refer to Section 3.1.

Once features were extracted (either from each of five general networks second autoencoders or from hand crafted features), each feature set for every patch was passed into one of three methods for distinguishing between dilution classes and making a prediction about to which dilution class a patch belonged: softmax layer, random forest, support vector machine (both linear and radial basis function kernel). 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. For comparative purposes we provided the random forest and support vector machine approaches with hand-crafted features as well as the auto-generated features output from autoencoder 2. For further explanation of these machine learning methods, refer to Section 3.2.

2.8. Data analyses

Descriptive analyses were summarized based on accuracy at predicting concentration for a given purée flavor 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 \[22, 23\]. 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 \(I\), local texture was computed as a region-wise neighborhood entropy computation:

\[
H_i = - \sum_{x_j \in N_i} p_{N_i}(x_j) \cdot \log_2(p_{N_i}(x_j))
\]  

(1)

where \(H_i\) is the local entropy of the \(i^{th}\) pixel neighborhood \(N_i\), \(p_{N_i}\) is the intensity histogram (i.e., sample intensity probability distribution) of pixel neighborhood \(N_i\), and \(x_j \in N_i\) is the intensity value of pixel \(j\). According to this formulation, samples with smooth homogeneous texture contain little intensity variation over localized patches, resulting in a sparse \(p_{N_i}\) and thus low entropy. Conversely, samples with inhomogeneous or heterogeneous rough texture contains highly varying intensity values, resulting in a dense \(p_{N_i}\) and thus high entropy. We used \(9 \times 9\) pixel neighborhoods, resulting in 81 pixel intensities to populate a distribution containing 256 bins.

Color 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.
3. Theory

3.1. Comparative analyses: hand crafted features

The color features were constructed using a discrete quantized color histogram. Color histograms are relatively invariant to rotation and translation, and coarse color quantization encourages perceptual similarities through enlarged bin sizes. Environmental consistency (e.g., exposure time, white correction, illuminant spectrum, etc.) is important for such color comparisons. A controlled optical setup was used to fix the relevant optical parameters, and is discussed further in Section 2.2. Specifically, 64 color features were extracted by quantizing each color channel into four bins, yielding $4 \times 4 \times 4 = 64$ features. Given 64 colors, the number of pixels within a patch pertaining to each of the 64 color 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 color bin. Mathematically, given image $I(x, y)$, the quantized values at each coordinate $q(x, y)$ was computed as:

$$q(x, y) = \arg \min_i \{ ||I(x, y) - c_i||_2 \}$$

(2)

where $c_i \in C$ is the set of uniformly spaced bins spanning the color space. The final quantized feature set was computed as:

$$f_c(i) = \sum_{(x,y) \in D} \delta(q(x, y), i)$$

(3)

where $D$ is the image domain, and $\delta$ is the Kronecker delta function:

$$\delta(i, j) = \begin{cases} 
1, & \text{if } i = j \\
0, & \text{if } i \neq j
\end{cases}$$

(4)

With regards to texture features, we used a set of texture descriptors based on differential translation histograms [21]. First, all patches were converted from color to grayscale, yielding $I_g(x, y) \in R$. Then, the translational sum and difference maps were computed:

$$s_\Delta(x, y) = I_g(x, y) + T_\Delta I_g(x, y)$$

(5)

$$d_\Delta(x, y) = I_g(x, y) - T_\Delta I_g(x, y)$$

(6)

where $T_\Delta$ is the translation operator by coordinates $\Delta$. Then, the normalized sum and difference translation histograms were computed as:

$$h_{s,\Delta}(i) = \frac{1}{N} \sum_{(x,y) \in D} \delta(s_\Delta(x, y), i)$$

(7)

$$h_{d,\Delta}(i) = \frac{1}{N} \sum_{(x,y) \in D} \delta(d_\Delta(x, y), i)$$

(8)
where $D$ is the image domain, $N$ is the number of pixels, and $\delta(\cdot)$ is the Kronecker delta function from Eq. (4). These histograms represent texture descriptors. For example, pixels in homogeneous regions which exhibit low texture will be approximately equal, resulting in $s_{\Delta}(x,y) \approx 2I_g(x,y)$ and $d_{\Delta}(x,y) \approx 0$ for all $(x,y)$ in the homogeneous region.

Using these fundamental computations, the following texture features were computed for each patch [21, 24]:

Mean: $\mu = \frac{1}{2} \sum_i i h_{s,\Delta}(i)$  \hspace{1cm} (9)

Contrast: $c_n = \sum_j j^2 h_{d,\Delta}(j)$  \hspace{1cm} (10)

Homogeneity: $h_g = \sum_j \frac{1}{1 + j^2} h_{d,\Delta}(j)$  \hspace{1cm} (11)

Energy: $e_n = \sum_i h_{s,\Delta}(i)^2 \sum_j h_{d,\Delta}(j)^2$  \hspace{1cm} (12)

Variance: $\sigma^2 = \frac{1}{2} \left( \sum_i (i - 2\mu)^2 h_{s,\Delta}(i) + \sum_j j^2 h_{d,\Delta}(j) \right)$  \hspace{1cm} (13)

Correlation: $c_r = \frac{1}{2} \left( \sum_i (i - 2\mu)^2 h_{s,\Delta}(i) - \sum_j j^2 h_{d,\Delta}(j) \right)$  \hspace{1cm} (14)

Entropy: $h_n = - \sum_i h_{s,\Delta}(i) \log(h_{s,\Delta}(i)) - \sum_j h_{d,\Delta}(j) \log(h_{d,\Delta}(j))$  \hspace{1cm} (15)

3.2. Comparative analyses: methods for distinguishing between dilution classes

As mentioned previously, three methods for distinguishing between dilution classes were used: a softmax layer, and for comparative purposes, random forests [25], and support vector machines (SVM) [26]. Here we describe a softmax layer. A 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.

3.3. 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 color. 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 color. Mathematically, expressing this using the Beer-Lambert law of light attenuation produces:

$$A = \log \left( \frac{I_0}{I} \right) = \epsilon_{H_2O} \cdot c_{H_2O} \cdot l_{H_2O} + \epsilon_p \cdot c_p \cdot l_p$$  \hspace{1cm} (16)$$

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:

$$A = -(\epsilon_{H_2O}(1 - c_p) + \epsilon_p c_p)l$$  \hspace{1cm} (17)$$

where $l = l_p = l_{H_2O}$. Representative perceptual image patches were derived from the mixture absorbance spectra by computing the perceived spectra color according to the CIE LMS cone responsivity curves:

$$I_v = \int_{400}^{700} Z_v(\lambda) \cdot \exp \left( -(\epsilon_{H_2O}(1 - c_p(\lambda)) + \epsilon_p c_p(\lambda))l \right) \, d\lambda$$  \hspace{1cm} (18)$$

where $I_v$ and $Z_v$ are the image and spectral cone response for color channel $v \in \{R, G, B\}$.

4. Results

To understand the performance of the deep neural network (DNN) for predicting purée sample concentration, results are organized as follows: (1) supporting evidence from the optical dilution model that dilution is quantifiable through perceptual data; (2) descriptive analyses of each image class in terms of color, texture and saturation to provide quantitative insights; (3) sample patches for each class across every purée flavor 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.

4.1. 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 2 demonstrates how the absorbance spectrum changes according to the puree concentration (i.e., dilution) according to (17) using published absorbance curves for water [28] and blueberry [27]. Blueberries contain anthocyanins, pH sensitive chromophores which shift from red to blue with increasing pH. For example pH = 1 appear red, pH = 4.5 they appear colorless and at pH 7-8 appear blue [29]. As the undiluted blueberry pur ee becomes more diluted, its pH increases causing a shift from red to blue. In its undiluted state the
Figure 2: (a) Normalized absorbance spectra for blueberry [27] and water [28]. (b) Effect of increasing dilution (decreasing relative purée concentration) on the water-blueberry mixture spectral curve. As purée dilution increases, the overall absorption decreases due to fewer photon absorption events during the photon migration path, leading to lighter observed images (bottom patch). These findings are valid for other absorbing purée chromophores without loss of generality.

4.2. Descriptive analyses

Figures A.1 and A.2 provide qualitative information about each class of images with respect to color (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 color, with the exceptions of banana and chicken the colors 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±0.008, 100% IC 0.988±0.015) and the least saturated samples were blueberry (saturation: 20% IC 0.538±0.071, 100% IC 0.128±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 3. In terms of texture (entropy) the more diluted samples were more similar in appearance to water and aside from their color, 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 have a tendency to form gels 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). For a full list of saturation, RGB, and entropy values, refer to Table A.3.

4.3. Sample patches

Figure 3 depicts sample patches for each class of purée flavor 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 2 which supports the hypothesis that quantifiable observational evidence can be used to estimate relative nutrient concentration. A color intensity gradient across the concentrations was observed. Several purée flavors, 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 as a result of the preparation of the dilution series. These gradients are visual indications which the network may be using to distinguish between different concentration classes.

4.4. Network accuracy

The observations noted provide both quantitative (Figures A.1 and A.2) and qualitative (Figure A.3) insight into performance. The method with the highest performance across flavors was our proposed DNN with an overall accuracy of 92.2%±4.1%, sensitivity of 83.0%±1.5%, and specificity of 95.0%±4.8%. This was closely followed by the handcrafted features paired with random forests for discrimination between dilutions. As illustrated in Tables A.1, A.2, the most consistently highest performing purée flavor was strawberry. Pertaining to the most highly performing method for discrimination was the stacked autoencoders.
and softmax layer approach, blueberry performed best. Across 10 trials, the mean accuracy for classifying blueberry dilutions was 99.6% ± 0.6% (sensitivity 98.9% ± 1.9%, specificity 99.7% ± 0.5%). These results are consistent with the descriptive analyses based on the high variance of color, 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 flavors' average accuracy ranged between 73.3% ± 7.8% (chicken) and 98.2% ± 1.2% (strawberry). Across all seven methods for discrimination between dilutions, chicken was the most difficult flavor 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 with no discernible features. This was consistent with the low entropy and high saturation seen in Table I.
Table 1: Summary of sensitivity, specificity and accuracy across all flavors using either self-generated features extracted from an autoencoder or color and texture based handcrafted features using four methods to discriminate between dilutions: softmax layer, random forest, SVM - linear kernel SVM, and SVM - radial basis kernel.

### SUMMARY OF PERFORMANCE ACROSS 13 FLAVORS

| Method (AutoFeat) | Sens  | Spec  | Acc   |
|------------------|-------|-------|-------|
|                  | $\mu_{Sens} \pm \sigma_{Sens}$ | $\mu_{Spec} \pm \sigma_{Spec}$ | $\mu_{Acc} \pm \sigma_{Acc}$ |
| SVM Linear       | 0.690 ± 0.187 | 0.900 ± 0.074 | 0.844 ± 0.061 |
| SVM Radial Basis | 0.609 ± 0.179 | 0.868 ± 0.082 | 0.794 ± 0.070 |
| Random Forest    | 0.790 ± 0.143 | 0.937 ± 0.046 | 0.903 ± 0.039 |
| Softmax          | 0.830 ± 0.150 | 0.950 ± 0.048 | 0.922 ± 0.041 |

| Method (HandFeat) | Sens  | Spec  | Acc   |
|------------------|-------|-------|-------|
|                  | $\mu_{Sens} \pm \sigma_{Sens}$ | $\mu_{Spec} \pm \sigma_{Spec}$ | $\mu_{Acc} \pm \sigma_{Acc}$ |
| SVM Linear       | 0.665 ± 0.188 | 0.890 ± 0.081 | 0.830 ± 0.057 |
| SVM Radial Basis | 0.577 ± 0.175 | 0.850 ± 0.078 | 0.770 ± 0.063 |
| Random Forest    | 0.826 ± 0.158 | 0.949 ± 0.047 | 0.920 ± 0.044 |

5. Discussion

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 generalizability is considered our DNN method using stacked autoencoders and a softmax layer provides several advantages. First, since autoencoders leverage unsupervised learning for the purpose of training, labels are unnecessary to create a global dilution model. Our initial global model contained all 13 flavors (flavors unlabeled) to learn the five dilutions. To fine tune this global model of dilution classes for a specific flavor a relatively small amount of labelled data for the new flavor would be required compared to forming a new flavor specific model as in the case of hand-crafted features with random forests for distinguishing between dilutions. Second, while it is unclear how either method would perform on more complex foods (e.g., regular texture, multiple food types on a plate), the deep learning approach may show more promise for extensibility since DNNs have historically be more flexible as evidenced by high accuracy across diverse applications (e.g., speech recognition [11, 12, 13], object recognition [14, 15, 16, 17], and natural language processing [18, 19]). A third advantage is in the case of introducing a new flavor. Specifically, for the hand-crafted feature and random forest approach, no classification can be made without training a new model. Instead, our approach is based upon a general global dilution model and may therefore be used a starting point for discriminating between classes; initial accuracy could then be improved by fine-tuning the global dilution model for a new-flavor specific model.

The subimages and patches used for this study reflected the $1/10$ s expo-
sure time. By re-running the same tests on the 1/20 s data, we may be able to improve performance. Since the color 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 color (and perhaps relative or normalized entropy) and composition of macronutrients (e.g., carbohydrates, protein, fat) or micronutrients (e.g., vitamin A, iron).

The system proposed here is the first step towards the end-goal of a “smart” imaging system to automatically detect the concentration and composition of commercially prepared puréed samples. As such, the output from the system offered only the best label. Future work will explore the confidence of a particular class instead of simply the output (i.e., label) from the softmax layer. In addition, classification was performed on a patch by patch basis; it may be more meaningful to instead classify based on the entire subimage using either a weighted average with mean squared error as a measure of accuracy or even simply taking the mode class of patches. The end goal would be to provide a means for a cook, dietitian or caregiver to run an image through the pre-trained system and have a relative nutrient density estimate to inform nutritional density and texture safety measurements. Future directions of output from the system should be based on input from end-users to ensure output from the system is most clinically meaningful and relevant. In addition, since our system generalized well to new data, we expect this approach will allow for robust extensions including combining classification of purée flavor as well as dilution, and to extend beyond puréed food to other modified textures (e.g., from regular to minced, to puréed, and to liquidized) consistent with the recent development of the International Dysphagia Diet Standardization Initiative’s dysphagia diet terminology [1].

6. Conclusions

In this paper, we demonstrated the feasibility of automatic nutritional density analysis using deep neural networks to predict the concentration of commercially prepared purées. Using multispectral imaging data at different polarizations, a stacked autoencoder attained promising accuracy across thirteen common puréed foods. Of the 13 purée flavors tested, the most promising methodology was using DNNs comprised of stacked autoencoders with a softmax layer to distinguish between dilutions. Over all flavors and across 10 trials the mean accuracy of this method was 92.2%±4.1% with mean sensitivity 83.0%±15.0%, and mean specificity 95.0%±4.8%. Dilution classification results were strongest for purée flavors with observable texture differences reflected in higher entropy, higher variation in color across dilution classes, and lower saturation. In contrast, purée flavors performed more poorly when there were fewer visual cues to discriminate between dilution classes which was further reflected in low color 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. If accuracy can be further improved and the system can achieve similar accuracy on a wider range of modified texture foods, this machine learning DNN imaging system for nutrient density analysis of purées show promise as a tool for purée nutrient quality assurance.

Acknowledgments

The authors would like to acknowledge Professor Heather Keller for her contribution to the overall vision of this developing system. This work was supported by the Natural Sciences and Engineering Research Council of Canada, the Canada Research Chairs Program, and Nvidia for the GPU hardware used in this study through the Nvidia Hardware Grant Program.

References

[1] J. A. Y. Cichero, P. Lam, C. M. Steele, B. Hanson, J. Chen, R. O. Dantas, J. Duivestein, J. Kayashita, C. Lecko, J. Murray, M. Pillay, L. Riquelme, S. Stanschus, Development of international terminology and definitions for texture-modified foods and thickened fluids used in dysphagia management: The IDDSI framework, Dysphagia (2016) 1–24 [doi: 10.1007/s00455-016-9758-y]. URL http://dx.doi.org/10.1007/s00455-016-9758-y

[2] L. Sura, A. Madhavan, G. Carnaby, M. A. Crary, Dysphagia in the elderly: Management and nutritional considerations, Clin Interv Aging 7 (287) (2012) 98.

[3] N. Ilhamto, K. Anciado, H. H. Keller, L. 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) (2014) 210–228.

[4] H. H. Keller, T. Østbye, R. Goy, Nutritional risk predicts quality of life in elderly community-living Canadians, The Journals of Gerontology series A: Biological sciences and Medical sciences 59 (1) (2004) M68-M74.

[5] S. Goates, K. Du, C. A. Braunschweig, M. B. Arensberg, Economic burden of disease-associated malnutrition at the state level, PloS one 11 (9) (2016) e0161833.

[6] C. A. Russell, The impact of malnutrition on healthcare costs and economic considerations for the use of oral nutritional supplements, Clinical Nutrition Supplements 2 (1) (2007) 25–32.

[7] I. Germain, T. Dufresne, K. Gray-Donald, A novel dysphagia diet improves the nutrient intake of institutionalized elders, Journal of the American Dietetic Association 106 (10) (2006) 1614–1623.
[8] I. J. Bigio, S. Fantini, Quantitative Biomedical Optics: Theory, Methods, and Applications, Cambridge University Press, 2016.

[9] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[10] Y. Bengio, Learning deep architectures for AI, Foundations and Trends in Machine Learning 2 (1) (2009) 1–127.

[11] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. 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) (2012) 82–97.

[12] G. E. Dahl, D. Yu, L. Deng, A. Acero, Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition, IEEE Transactions on Audio, Speech, and Language Processing 20 (1) (2012) 30–42.

[13] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates, et al., Deep speech: Scaling up end-to-end speech recognition, arXiv preprint arXiv:1412.5567.

[14] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[15] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE transactions on pattern analysis and machine intelligence 37 (9) (2015) 1904–1916.

[16] Y. LeCun, F. J. Huang, L. 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, Vol. 2, IEEE, 2004, pp. II–104.

[17] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556.

[18] Y. Bengio, R. Ducharme, P. Vincent, C. Jauvin, A neural probabilistic language model, Journal of machine learning research 3 (Feb) (2003) 1137–1155.

[19] R. Collobert, J. Weston, A unified architecture for natural language processing: Deep neural networks with multitask learning, in: Proceedings of the 25th international conference on Machine learning, ACM, 2008, pp. 160–167.

[20] G. E. Hinton, R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[21] Y. Zhang, L. Wu, Classification of fruits using computer vision and a multiclass support vector machine, Sensors 12 (9) (2012) 12489–12505.

[22] M. Bosch, F. Zhu, N. Khanna, C. J. Boushey, E. J. Delp, Combining global and local features for food identification in dietary assessment, in: Image Processing (ICIP), 18th IEEE International Conference on, IEEE, 2011, pp. 1789–1792.

[23] F. Zhu, M. Bosch, N. Khanna, C. J. Boushey, E. J. Delp, Multilevel segmentation for food classification in dietary assessment, in: Image and Signal Processing and Analysis (ISPA), 2011 7th International Symposium on, IEEE, 2011, pp. 337–342.

[24] M. Unser, Sum and difference histograms for texture classification, IEEE Transactions on Pattern Analysis and Machine Intelligence (1) (1986) 118–125.

[25] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32.

[26] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (3) (1995) 273–297.

[27] F. Teoli, S. Lucioli, P. Nota, A. Frattarelli, F. Matteocci, A. Di Carlo, E. Caboni, 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 (2016) 24–30.

[28] M. P. Robin, S. F. Edward, Absorption spectrum of pure water (380–700 nm). II. integrating cavity measurements, Applied Optics 36 (33) (1997) 8710–8723.

[29] R. E. Wrolstad, et al., Color and pigment analyses in fruit products, Tech. rep., Corvallis, Or.: Agricultural Experiment Station. Oregon State University. (1993).

[30] M. D. Alvarez, W. Canet, Dynamic viscoelastic behavior of vegetable-based infant purees, Journal of Texture Studies 44 (3) (2013) 205–224.
Figure A.1: Descriptive analysis plots of purée flavors based on color. RGB values have been normalized. Typically color varied between the dilution classes of a purée flavor and were distinguishable between different purée flavors.
Figure A.2: Descriptive analysis plots of purée flavors based on saturation and texture (entropy). Saturation was normalized; entropy was used to describe texture. Typically saturation and texture, varied across a purée flavor’s dilution classes and were distinguishable between different purée flavors.
Figure A.3: Sample horizontal polarized subimages taken with ISO 100, EXP 1/20 s. Note the visible variations across dilutions for the best performing purée flavor, with the stacked autoencoders and softmax layer, highlighted in blue (average accuracy blueberry: 99.6%) and the indistinguishable color and lack of texture consistent across the poorest performing purée flavor highlighted in red (average accuracy chicken: 73.3%); the subimages at 60% and 80% appear nearly absent or completely saturated.
Table A.1: Summary of sensitivity, specificity and accuracy for each flavor using self-generated features extracted from an autoencoder and four methods for to discriminate between dilutions: softmax layer, random forest, SVM - linear kernel SVM, and SVM - radial basis kernel. Results summarized represent five randomly initiated networks with 6-fold cross-validation (i.e., leave one of each of the six imaged positions out for testing) for each flavor.

| Flavour  | Sens $\mu_{\text{Sens}} \pm \sigma_{\text{Sens}}$ | Spec $\mu_{\text{Spec}} \pm \sigma_{\text{Spec}}$ | Acc $\mu_{\text{Acc}} \pm \sigma_{\text{Acc}}$ |
|----------|---------------------------------|---------------------------------|---------------------------------|
| Apple    | 0.732 ± 0.211                   | 0.923 ± 0.065                   | 0.880 ± 0.043                   |
| Apricot  | 0.750 ± 0.112                   | 0.927 ± 0.036                   | 0.887 ± 0.026                   |
| Banana   | 0.607 ± 0.214                   | 0.859 ± 0.119                   | 0.792 ± 0.096                   |
| Beef     | 0.608 ± 0.252                   | 0.868 ± 0.096                   | 0.798 ± 0.070                   |
| Blueberry| 0.840 ± 0.101                   | 0.956 ± 0.031                   | 0.931 ± 0.030                   |
| Carrot   | 0.628 ± 0.171                   | 0.884 ± 0.068                   | 0.820 ± 0.043                   |
| Chicken  | 0.395 ± 0.322                   | 0.781 ± 0.186                   | 0.621 ± 0.156                   |
| Mango    | 0.668 ± 0.243                   | 0.894 ± 0.080                   | 0.841 ± 0.072                   |
| Parsnip  | 0.765 ± 0.198                   | 0.928 ± 0.071                   | 0.889 ± 0.085                   |
| Pea      | 0.657 ± 0.140                   | 0.892 ± 0.052                   | 0.834 ± 0.037                   |
| Squash   | 0.713 ± 0.195                   | 0.911 ± 0.067                   | 0.865 ± 0.057                   |
| Strawberry| 0.918 ± 0.063                  | 0.979 ± 0.016                   | 0.966 ± 0.013                   |
| Sweet Potato | 0.688 ± 0.202             | 0.903 ± 0.080                   | 0.850 ± 0.062                   |
| Across Flavours | 0.690 ± 0.187       | 0.900 ± 0.074                   | 0.844 ± 0.061                   |

**AUTOENCODER 2 FEATURES**

**SVM Linear Kernel**

| Flavour  | Sens $\mu_{\text{Sens}} \pm \sigma_{\text{Sens}}$ | Spec $\mu_{\text{Spec}} \pm \sigma_{\text{Spec}}$ | Acc $\mu_{\text{Acc}} \pm \sigma_{\text{Acc}}$ |
|----------|---------------------------------|---------------------------------|---------------------------------|
| Apple    | 0.635 ± 0.188                   | 0.895 ± 0.063                   | 0.828 ± 0.061                   |
| Apricot  | 0.702 ± 0.134                   | 0.907 ± 0.052                   | 0.859 ± 0.046                   |
| Banana   | 0.509 ± 0.159                   | 0.807 ± 0.129                   | 0.717 ± 0.105                   |
| Beef     | 0.446 ± 0.267                   | 0.795 ± 0.108                   | 0.679 ± 0.082                   |
| Blueberry| 0.736 ± 0.096                   | 0.923 ± 0.035                   | 0.880 ± 0.028                   |
| Carrot   | 0.573 ± 0.131                   | 0.865 ± 0.064                   | 0.786 ± 0.049                   |
| Chicken  | 0.385 ± 0.317                   | 0.776 ± 0.190                   | 0.609 ± 0.162                   |
| Mango    | 0.655 ± 0.240                   | 0.889 ± 0.082                   | 0.833 ± 0.078                   |
| Parsnip  | 0.627 ± 0.232                   | 0.880 ± 0.091                   | 0.813 ± 0.107                   |
| Pea      | 0.578 ± 0.128                   | 0.851 ± 0.061                   | 0.777 ± 0.043                   |
| Squash   | 0.645 ± 0.193                   | 0.887 ± 0.081                   | 0.827 ± 0.060                   |
| Strawberry| 0.852 ± 0.063                  | 0.958 ± 0.016                   | 0.935 ± 0.019                   |
| Sweet Potato | 0.567 ± 0.181             | 0.854 ± 0.094                   | 0.773 ± 0.076                   |
| Across Flavours | 0.609 ± 0.179       | 0.868 ± 0.082                   | 0.794 ± 0.070                   |

**SVM Radial Basis Kernel**
Table A.1 continued:

| Flavour  | Sens                        | Spec                        | Acc                         |
|----------|-----------------------------|-----------------------------|-----------------------------|
|          | $\mu_{Sens} \pm \sigma_{Sens}$ | $\mu_{Spec} \pm \sigma_{Spec}$ | $\mu_{Acc} \pm \sigma_{Acc}$ |
| Apple    | 0.824 ± 0.140               | 0.952 ± 0.039               | 0.925 ± 0.036               |
| Apricot  | 0.859 ± 0.097               | 0.962 ± 0.030               | 0.940 ± 0.029               |
| Banana   | 0.841 ± 0.160               | 0.956 ± 0.043               | 0.931 ± 0.050               |
| Beef     | 0.708 ± 0.204               | 0.910 ± 0.064               | 0.862 ± 0.046               |
| Blueberry| 0.906 ± 0.077               | 0.975 ± 0.026               | 0.961 ± 0.027               |
| Carrot   | 0.709 ± 0.144               | 0.914 ± 0.062               | 0.866 ± 0.033               |
| Chicken  | 0.522 ± 0.206               | 0.826 ± 0.085               | 0.744 ± 0.073               |
| Mango    | 0.759 ± 0.195               | 0.928 ± 0.060               | 0.890 ± 0.053               |
| Parsnip  | 0.867 ± 0.144               | 0.963 ± 0.039               | 0.942 ± 0.046               |
| Pea      | 0.728 ± 0.119               | 0.920 ± 0.040               | 0.876 ± 0.032               |
| Squash   | 0.793 ± 0.175               | 0.941 ± 0.048               | 0.908 ± 0.041               |
| Strawberry| 0.934 ± 0.066              | 0.983 ± 0.017               | 0.973 ± 0.014               |
| Sweet Potato | 0.813 ± 0.126         | 0.948 ± 0.038               | 0.918 ± 0.032               |
| Across Flavours | 0.790 ± 0.143           | 0.937 ± 0.046               | 0.903 ± 0.039               |

**Softmax Layer**

| Flavour  | Sens                        | Spec                        | Acc                         |
|----------|-----------------------------|-----------------------------|-----------------------------|
|          | $\mu_{Sens} \pm \sigma_{Sens}$ | $\mu_{Spec} \pm \sigma_{Spec}$ | $\mu_{Acc} \pm \sigma_{Acc}$ |
| Apple    | 0.871 ± 0.154               | 0.966 ± 0.041               | 0.947 ± 0.033               |
| Apricot  | 0.920 ± 0.073               | 0.979 ± 0.017               | 0.967 ± 0.018               |
| Banana   | 0.874 ± 0.188               | 0.964 ± 0.054               | 0.944 ± 0.057               |
| Beef     | 0.709 ± 0.249               | 0.914 ± 0.076               | 0.865 ± 0.063               |
| Blueberry| 0.989 ± 0.019               | 0.997 ± 0.005               | 0.996 ± 0.006               |
| Carrot   | 0.769 ± 0.194               | 0.935 ± 0.074               | 0.897 ± 0.059               |
| Chicken  | 0.501 ± 0.286               | 0.824 ± 0.124               | 0.733 ± 0.078               |
| Mango    | 0.748 ± 0.232               | 0.923 ± 0.074               | 0.883 ± 0.068               |
| Parsnip  | 0.923 ± 0.118               | 0.980 ± 0.033               | 0.968 ± 0.038               |
| Pea      | 0.765 ± 0.113               | 0.935 ± 0.032               | 0.898 ± 0.024               |
| Squash   | 0.887 ± 0.138               | 0.969 ± 0.039               | 0.952 ± 0.043               |
| Strawberry| 0.955 ± 0.048              | 0.988 ± 0.013               | 0.982 ± 0.012               |
| Sweet Potato | 0.884 ± 0.133        | 0.969 ± 0.039               | 0.951 ± 0.035               |
| Across Flavours | 0.830 ± 0.150           | 0.950 ± 0.048               | 0.922 ± 0.041               |
Table A.2: Summary of sensitivity, specificity and accuracy for each flavor using 71 hand-crafted features pertaining to color and texture and three methods to discriminate between dilutions: random forest, SVM - linear kernel SVM, and SVM - radial basis kernel. Results summarized represent the average from five random forests each containing 10 trees, or one run of SVM for each of linear and radial basis kernels. All methods used 6-fold cross-validation (i.e., leave one of each of the six imaged positions out for testing) for each flavor.

**HANDCRAFTED FEATURES**

**SVM Linear Kernel**

| Flavour   | Sens            | Spec            | Acc             |
|-----------|-----------------|-----------------|-----------------|
|           | $\mu_{Sens} \pm \sigma_{Sens}$ | $\mu_{Spec} \pm \sigma_{Spec}$ | $\mu_{Acc} \pm \sigma_{Acc}$ |
| Apple     | 0.695 ± 0.150   | 0.908 ± 0.062   | 0.859 ± 0.040   |
| Apricot   | 0.731 ± 0.108   | 0.919 ± 0.036   | 0.876 ± 0.015   |
| Banana    | 0.638 ± 0.270   | 0.876 ± 0.098   | 0.813 ± 0.098   |
| Beef      | 0.496 ± 0.317   | 0.817 ± 0.178   | 0.720 ± 0.116   |
| Blueberry | 0.760 ± 0.098   | 0.927 ± 0.030   | 0.888 ± 0.030   |
| Carrot    | 0.644 ± 0.183   | 0.889 ± 0.074   | 0.829 ± 0.053   |
| Chicken   | 0.398 ± 0.216   | 0.764 ± 0.150   | 0.632 ± 0.068   |
| Mango     | 0.696 ± 0.172   | 0.907 ± 0.054   | 0.858 ± 0.040   |
| Parsnip   | 0.850 ± 0.151   | 0.958 ± 0.047   | 0.933 ± 0.055   |
| Pea       | 0.556 ± 0.220   | 0.858 ± 0.111   | 0.774 ± 0.064   |
| Squash    | 0.741 ± 0.197   | 0.921 ± 0.055   | 0.879 ± 0.070   |
| Strawberry| 0.864 ± 0.157   | 0.962 ± 0.049   | 0.941 ± 0.036   |
| Sweet Potato | 0.580 ± 0.200 | 0.861 ± 0.112   | 0.787 ± 0.053   |

*Across Flavours* 0.665 ± 0.188 0.890 ± 0.081 0.830 ± 0.057

**SVM Radial Basis Kernel**

| Flavour   | Sens            | Spec            | Acc             |
|-----------|-----------------|-----------------|-----------------|
|           | $\mu_{Sens} \pm \sigma_{Sens}$ | $\mu_{Spec} \pm \sigma_{Spec}$ | $\mu_{Acc} \pm \sigma_{Acc}$ |
| Apple     | 0.613 ± 0.225   | 0.870 ± 0.086   | 0.806 ± 0.082   |
| Apricot   | 0.729 ± 0.066   | 0.917 ± 0.033   | 0.874 ± 0.025   |
| Banana    | 0.441 ± 0.219   | 0.780 ± 0.118   | 0.662 ± 0.116   |
| Beef      | 0.381 ± 0.310   | 0.745 ± 0.151   | 0.612 ± 0.086   |
| Blueberry | 0.548 ± 0.095   | 0.839 ± 0.059   | 0.756 ± 0.066   |
| Carrot    | 0.585 ± 0.118   | 0.865 ± 0.061   | 0.794 ± 0.039   |
| Chicken   | 0.332 ± 0.177   | 0.730 ± 0.104   | 0.566 ± 0.066   |
| Mango     | 0.701 ± 0.125   | 0.911 ± 0.062   | 0.862 ± 0.046   |
| Parsnip   | 0.580 ± 0.180   | 0.859 ± 0.054   | 0.783 ± 0.066   |
| Pea       | 0.515 ± 0.208   | 0.837 ± 0.111   | 0.746 ± 0.057   |
| Squash    | 0.731 ± 0.185   | 0.918 ± 0.054   | 0.873 ± 0.070   |
| Strawberry| 0.806 ± 0.186   | 0.943 ± 0.051   | 0.912 ± 0.049   |
| Sweet Potato | 0.539 ± 0.183 | 0.838 ± 0.075   | 0.758 ± 0.049   |

*Across Flavours* 0.577 ± 0.175 0.850 ± 0.078 0.770 ± 0.063
Table A.2 continued:

| Flavour      | Sens  | Spec  | Acc   |
|--------------|-------|-------|-------|
|              | $\mu_{\text{Sens}} \pm \sigma_{\text{Sens}}$ | $\mu_{\text{Spec}} \pm \sigma_{\text{Spec}}$ | $\mu_{\text{Acc}} \pm \sigma_{\text{Acc}}$ |
| Apple        | 0.871 ± 0.170 | 0.966 ± 0.051 | 0.946 ± 0.048 |
| Apricot      | 0.923 ± 0.068 | 0.980 ± 0.020 | 0.968 ± 0.017 |
| Banana       | 0.855 ± 0.177 | 0.960 ± 0.050 | 0.936 ± 0.053 |
| Beef         | 0.775 ± 0.155 | 0.935 ± 0.043 | 0.900 ± 0.039 |
| Blueberry    | 0.929 ± 0.090 | 0.981 ± 0.022 | 0.971 ± 0.028 |
| Carrot       | 0.769 ± 0.173 | 0.935 ± 0.056 | 0.897 ± 0.043 |
| Chicken      | 0.532 ± 0.257 | 0.833 ± 0.103 | 0.748 ± 0.087 |
| Mango        | 0.742 ± 0.241 | 0.923 ± 0.069 | 0.881 ± 0.061 |
| Parsnip      | 0.922 ± 0.141 | 0.979 ± 0.036 | 0.967 ± 0.040 |
| Pea          | 0.811 ± 0.138 | 0.948 ± 0.047 | 0.917 ± 0.040 |
| Squash       | 0.882 ± 0.165 | 0.969 ± 0.041 | 0.951 ± 0.047 |
| Strawberry   | 0.908 ± 0.139 | 0.976 ± 0.036 | 0.962 ± 0.035 |
| Sweet Potato | 0.812 ± 0.134 | 0.948 ± 0.043 | 0.919 ± 0.032 |
| Across Flavours | **0.826 ± 0.158** | **0.949 ± 0.047** | **0.920 ± 0.044** |
Table A.3: Descriptive statistics: Mean ($\mu \pm \sigma$) red (R), green (G), blue (B), texture (T), and saturation (S) values for each flavour across 20% (most diluted), 40%, 60%, 80%, and 100% (undiluted) concentrations.

| Flavour (FRUIT) | 20%          | 40%          | 60%          | 80%          | 100%         |
|-----------------|--------------|--------------|--------------|--------------|--------------|
| Apple           | R ($\mu_R \pm \sigma_R$) | 0.981 ± 0.0228 | 0.976 ± 0.0279 | 0.966 ± 0.0385 | 0.957 ± 0.0578 | 0.919 ± 0.0816 |
|                 | G ($\mu_G \pm \sigma_G$) | 0.879 ± 0.0497 | 0.719 ± 0.0482 | 0.633 ± 0.0518 | 0.587 ± 0.0726 | 0.523 ± 0.0678 |
|                 | B ($\mu_B \pm \sigma_B$) | 0.489 ± 0.0581 | 0.261 ± 0.0577 | 0.192 ± 0.054  | 0.164 ± 0.0521 | 0.147 ± 0.0471 |
|                 | T ($\mu_T \pm \sigma_T$) | 3.3 ± 0.666   | 3.33 ± 0.631  | 3.42 ± 0.576   | 3.57 ± 0.509   | 3.63 ± 0.528   |
|                 | S ($\mu_S \pm \sigma_S$) | 0.865 ± 0.0358 | 0.743 ± 0.035  | 0.683 ± 0.0393 | 0.649 ± 0.0568 | 0.598 ± 0.0614 |
| Apricot         | R ($\mu_R \pm \sigma_R$) | 0.993 ± 0.00945 | 0.994 ± 0.00754 | 0.994 ± 0.00958 | 0.991 ± 0.015  | 0.98 ± 0.0355  |
|                 | G ($\mu_G \pm \sigma_G$) | 0.851 ± 0.0483 | 0.733 ± 0.0508 | 0.631 ± 0.0581 | 0.546 ± 0.0618 | 0.469 ± 0.0632 |
|                 | B ($\mu_B \pm \sigma_B$) | 0.377 ± 0.0643 | 0.154 ± 0.057  | 0.055 ± 0.0455 | 0.040 ± 0.0346 | 0.026 ± 0.0315 |
|                 | T ($\mu_T \pm \sigma_T$) | 2.68 ± 0.632  | 2.59 ± 0.617  | 2.46 ± 0.635  | 2.54 ± 0.625  | 2.69 ± 0.603  |
|                 | S ($\mu_S \pm \sigma_S$) | 0.839 ± 0.0347 | 0.745 ± 0.0351 | 0.675 ± 0.038  | 0.622 ± 0.0396 | 0.571 ± 0.0428 |
| Banana          | R ($\mu_R \pm \sigma_R$) | 0.997 ± 0.00461 | 0.994 ± 0.00705 | 0.987 ± 0.0247 | 0.991 ± 0.0231 | 0.994 ± 0.0155 |
|                 | G ($\mu_G \pm \sigma_G$) | 0.983 ± 0.023  | 0.897 ± 0.047  | 0.855 ± 0.0748 | 0.901 ± 0.0775 | 0.939 ± 0.0759 |
|                 | B ($\mu_B \pm \sigma_B$) | 0.631 ± 0.058  | 0.507 ± 0.0585 | 0.499 ± 0.0631 | 0.518 ± 0.0662 | 0.526 ± 0.0722 |
|                 | T ($\mu_T \pm \sigma_T$) | 1.6 ± 0.942   | 2.64 ± 0.663  | 3.08 ± 0.549  | 2.99 ± 0.773  | 2.64 ± 1.2    |
|                 | S ($\mu_S \pm \sigma_S$) | 0.947 ± 0.0187 | 0.882 ± 0.033  | 0.854 ± 0.0534 | 0.884 ± 0.0549 | 0.908 ± 0.0528 |
| Blueberry       | R ($\mu_R \pm \sigma_R$) | 0.622 ± 0.0796 | 0.592 ± 0.0694 | 0.498 ± 0.068  | 0.392 ± 0.0642 | 0.273 ± 0.0583 |
|                 | G ($\mu_G \pm \sigma_G$) | 0.512 ± 0.0686 | 0.367 ± 0.0572 | 0.196 ± 0.0481 | 0.111 ± 0.0422 | 0.0668 ± 0.0352 |
|                 | B ($\mu_B \pm \sigma_B$) | 0.449 ± 0.0779 | 0.328 ± 0.0665 | 0.179 ± 0.0527 | 0.11 ± 0.0461 | 0.0656 ± 0.0362 |
|                 | T ($\mu_T \pm \sigma_T$) | 3.76 ± 0.651  | 3.68 ± 0.618  | 3.37 ± 0.535  | 3.02 ± 0.482  | 2.86 ± 0.527  |
|                 | S ($\mu_S \pm \sigma_S$) | 0.538 ± 0.0709 | 0.43 ± 0.0592 | 0.284 ± 0.0505 | 0.195 ± 0.0431 | 0.128 ± 0.0345 |
Table A.3 continued:

| Flavour (FRUIT) | 20%     | 40%     | 60%     | 80%     | 100%    |
|-----------------|---------|---------|---------|---------|---------|
| Mango R (µR ± σR) | 0.99 ± 0.0103 | 0.991 ± 0.00995 | 0.99 ± 0.0111 | 0.984 ± 0.0215 | 0.96 ± 0.0492 |
| G (µG ± σG)     | 0.844 ± 0.0491 | 0.647 ± 0.0455 | 0.606 ± 0.0447 | 0.551 ± 0.0517 | 0.486 ± 0.0475 |
| B (µB ± σB)     | 0.153 ± 0.0587 | 0.0429 ± 0.0381 | 0.0383 ± 0.0328 | 0.0337 ± 0.0317 | 0.038 ± 0.0355 |
| T (µT ± σT)     | 3.22 ± 0.721 | 3.05 ± 0.693 | 3.06 ± 0.666 | 3.2 ± 0.649 | 3.33 ± 0.631 |
| S (µS ± σS)     | 0.809 ± 0.0317 | 0.681 ± 0.028 | 0.656 ± 0.0277 | 0.621 ± 0.0327 | 0.577 ± 0.0365 |
| Strawberry R (µR ± σR) | 0.933 ± 0.0466 | 0.879 ± 0.0554 | 0.855 ± 0.0565 | 0.73 ± 0.0663 | 0.634 ± 0.0775 |
| G (µG ± σG)     | 0.726 ± 0.0454 | 0.552 ± 0.0446 | 0.474 ± 0.0444 | 0.291 ± 0.0425 | 0.23 ± 0.0396 |
| B (µB ± σB)     | 0.542 ± 0.0551 | 0.348 ± 0.0519 | 0.269 ± 0.0496 | 0.127 ± 0.0458 | 0.102 ± 0.0412 |
| T (µT ± σT)     | 3.5 ± 0.662 | 3.53 ± 0.621 | 3.51 ± 0.647 | 3.53 ± 0.6 | 3.52 ± 0.607 |
| S (µS ± σS)     | 0.767 ± 0.0415 | 0.627 ± 0.0399 | 0.565 ± 0.0394 | 0.403 ± 0.0417 | 0.336 ± 0.0412 |

| Flavour (MEAT) | 20%     | 40%     | 60%     | 80%     | 100%    |
|-----------------|---------|---------|---------|---------|---------|
| Beef R (µR ± σR) | 0.966 ± 0.0345 | 0.907 ± 0.0564 | 0.849 ± 0.0607 | 0.855 ± 0.0613 | 0.928 ± 0.0842 |
| G (µG ± σG)     | 0.715 ± 0.0482 | 0.621 ± 0.0451 | 0.603 ± 0.0456 | 0.622 ± 0.0451 | 0.672 ± 0.0832 |
| B (µB ± σB)     | 0.52 ± 0.0503 | 0.432 ± 0.0484 | 0.425 ± 0.0499 | 0.444 ± 0.0486 | 0.466 ± 0.0684 |
| T (µT ± σT)     | 3.34 ± 0.575 | 3.41 ± 0.537 | 3.44 ± 0.532 | 3.48 ± 0.524 | 3.82 ± 0.549 |
| S (µS ± σS)     | 0.768 ± 0.0398 | 0.685 ± 0.0433 | 0.656 ± 0.0456 | 0.672 ± 0.045 | 0.725 ± 0.0766 |
| Chicken R (µR ± σR) | 0.998 ± 0.00329 | 0.997 ± 0.00401 | 0.997 ± 0.0045 | 0.996 ± 0.00506 | 0.997 ± 0.00417 |
| G (µG ± σG)     | 0.999 ± 0.00273 | 0.998 ± 0.00431 | 0.995 ± 0.00904 | 0.993 ± 0.0134 | 0.997 ± 0.0123 |
| B (µB ± σB)     | 0.925 ± 0.0646 | 0.863 ± 0.0602 | 0.826 ± 0.0575 | 0.815 ± 0.0579 | 0.918 ± 0.0844 |
| T (µT ± σT)     | 1.12 ± 0.784 | 1.57 ± 0.728 | 1.67 ± 0.848 | 1.81 ± 0.887 | 1.43 ± 1.05 |
| S (µS ± σS)     | 0.99 ± 0.00826 | 0.982 ± 0.00813 | 0.977 ± 0.0106 | 0.973 ± 0.0132 | 0.988 ± 0.0153 |
Table A.3 continued:

**PERCENT INITIAL CONCENTRATION**

| Flavour (VEGETABLE) | 20%                | 40%                | 60%                | 80%                | 100%               |
|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Carrot              | R ($\mu_R \pm \sigma_R$) | 0.991 ± 0.00947    | 0.983 ± 0.0196     | 0.966 ± 0.0405     | 0.958 ± 0.0549     | 0.911 ± 0.0863     |
|                     | G ($\mu_G \pm \sigma_G$) | 0.523 ± 0.0473     | 0.33 ± 0.0452      | 0.284 ± 0.0411     | 0.25 ± 0.0373      | 0.245 ± 0.0353     |
|                     | B ($\mu_B \pm \sigma_B$) | 0.14 ± 0.0576      | 0.0485 ± 0.0386    | 0.0455 ± 0.0415    | 0.0362 ± 0.0341    | 0.0389 ± 0.0352     |
|                     | T ($\mu_T \pm \sigma_T$) | 3.09 ± 0.718       | 3.16 ± 0.722       | 3.2 ± 0.69        | 3.25 ± 0.628       | 3.48 ± 0.584       |
|                     | S ($\mu_S \pm \sigma_S$) | 0.619 ± 0.0326     | 0.493 ± 0.0296     | 0.461 ± 0.0292     | 0.437 ± 0.0289     | 0.421 ± 0.0362     |
| Parsnip             | R ($\mu_R \pm \sigma_R$) | 0.99 ± 0.0144      | 0.967 ± 0.0329     | 0.929 ± 0.0631     | 0.967 ± 0.0487     | 0.869 ± 0.108      |
|                     | G ($\mu_G \pm \sigma_G$) | 0.952 ± 0.0425     | 0.727 ± 0.0412     | 0.638 ± 0.0541     | 0.761 ± 0.127      | 0.576 ± 0.0908     |
|                     | B ($\mu_B \pm \sigma_B$) | 0.621 ± 0.0636     | 0.341 ± 0.0562     | 0.276 ± 0.0545     | 0.429 ± 0.248      | 0.264 ± 0.0782     |
|                     | T ($\mu_T \pm \sigma_T$) | 2.83 ± 0.94        | 3.41 ± 0.651       | 3.62 ± 0.535       | 3.58 ± 1.11        | 4 ± 0.523          |
|                     | S ($\mu_S \pm \sigma_S$) | 0.926 ± 0.0316     | 0.755 ± 0.0325     | 0.684 ± 0.0487     | 0.785 ± 0.107      | 0.628 ± 0.0871     |
| Pea                 | R ($\mu_R \pm \sigma_R$) | 0.832 ± 0.0514     | 0.65 ± 0.0541      | 0.624 ± 0.0533     | 0.553 ± 0.0587     | 0.58 ± 0.072       |
|                     | G ($\mu_G \pm \sigma_G$) | 0.761 ± 0.0372     | 0.594 ± 0.0364     | 0.577 ± 0.0371     | 0.541 ± 0.0409     | 0.569 ± 0.0547     |
|                     | B ($\mu_B \pm \sigma_B$) | 0.236 ± 0.058      | 0.182 ± 0.0509     | 0.187 ± 0.0513     | 0.181 ± 0.0495     | 0.176 ± 0.0468     |
|                     | T ($\mu_T \pm \sigma_T$) | 3.59 ± 0.609       | 3.52 ± 0.644       | 3.49 ± 0.666       | 3.53 ± 0.625       | 3.68 ± 0.567       |
|                     | S ($\mu_S \pm \sigma_S$) | 0.722 ± 0.0336     | 0.564 ± 0.0347     | 0.547 ± 0.0355     | 0.504 ± 0.0401     | 0.528 ± 0.0523     |
| Squash              | R ($\mu_R \pm \sigma_R$) | 0.992 ± 0.00784    | 0.993 ± 0.00709    | 0.992 ± 0.001      | 0.95 ± 0.0629      | 0.943 ± 0.0796     |
|                     | G ($\mu_G \pm \sigma_G$) | 0.787 ± 0.0521     | 0.661 ± 0.0474     | 0.565 ± 0.0484     | 0.481 ± 0.0614     | 0.491 ± 0.0783     |
|                     | B ($\mu_B \pm \sigma_B$) | 0.111 ± 0.0481     | 0.0342 ± 0.0282    | 0.0289 ± 0.0241    | 0.0291 ± 0.028     | 0.032 ± 0.0508     |
|                     | T ($\mu_T \pm \sigma_T$) | 2.98 ± 0.672       | 2.83 ± 0.644       | 2.8 ± 0.712        | 3.24 ± 0.54        | 3.47 ± 0.614       |
|                     | S ($\mu_S \pm \sigma_S$) | 0.771 ± 0.0326     | 0.689 ± 0.0292     | 0.631 ± 0.03      | 0.57 ± 0.0499      | 0.574 ± 0.0644     |
| Sweet Potato        | R ($\mu_R \pm \sigma_R$) | 0.992 ± 0.00826    | 0.984 ± 0.02       | 0.911 ± 0.0643     | 0.904 ± 0.0851     | 0.917 ± 0.0866     |
|                     | G ($\mu_G \pm \sigma_G$) | 0.59 ± 0.0514      | 0.435 ± 0.0481     | 0.362 ± 0.0427     | 0.359 ± 0.0447     | 0.358 ± 0.0448     |
|                     | B ($\mu_B \pm \sigma_B$) | 0.0966 ± 0.0451    | 0.0563 ± 0.0363    | 0.0658 ± 0.0388    | 0.0498 ± 0.0305    | 0.0439 ± 0.0304    |
|                     | T ($\mu_T \pm \sigma_T$) | 2.81 ± 0.722       | 2.89 ± 0.712       | 3.18 ± 0.599       | 3.31 ± 0.522       | 3.33 ± 0.535       |
|                     | S ($\mu_S \pm \sigma_S$) | 0.654 ± 0.0337     | 0.556 ± 0.0329     | 0.493 ± 0.039      | 0.487 ± 0.0456     | 0.489 ± 0.046      |