Probing Sucrose Contents in Everyday Drinks Using Miniaturized Near-Infrared Spectroscopy Scanners

WEIWEI JIANG, The University of Melbourne, Australia
GABRIELE MARINI, The University of Melbourne, Australia
NIELS VAN BERKEL, University College London, United Kingdom
ZHANNA SARSENBAYEVA, The University of Melbourne, Australia
ZHEYU TAN, Japan Advanced Institute of Science and Technology, Japan
CHU LUO, The University of Melbourne, Australia
XIN HE, Anhui Normal University, China
TILMAN DINGLER, The University of Melbourne, Australia
JORGE GONCALVES, The University of Melbourne, Australia
YOSHIHIRO KAWAHARA, The University of Tokyo, Japan
VASSILIS KOSTAKOS, The University of Melbourne, Australia

Near-Infrared Spectroscopy (NIRS) is a non-invasive sensing technique which can be used to acquire information on an object's chemical composition. Although NIRS is conventionally used in dedicated laboratories, the recent introduction of miniaturized NIRS scanners has greatly expanded the use cases of this technology. Previous work from the UbiComp community shows that miniaturized NIRS can be successfully adapted to identify medical pills and alcohol concentration. In this paper, we further extend this technology to identify sugar (sucrose) contents in everyday drinks. We developed a standalone mobile device which includes *inter alia* a NIRS scanner and a 3D printed clamp. The clamp can be attached to a straw-like tube to sense a liquid's sucrose content. Through a series of studies, we show that our technique can accurately measure sucrose levels in both lab-made samples and commercially available drinks, as well as classify commercial drinks. Furthermore, we show that our method is robust to variations in the ambient temperature and lighting conditions. Overall, our system can estimate the concentration of sugar with ±0.29 g/100ml error in lab-made samples and < 2.0 g/100ml error in 18 commercial drinks, and can identify everyday drinks with > 99% accuracy. Furthermore, in our analysis, we are able to discern three characteristic wavelengths in the near-infrared region (1055 nm, 1235 nm and 1545 nm) with acute responses to sugar (sucrose). Our proposed protocol contributes to the development of everyday "food scanners" consumers.

CCS Concepts: • **Human-centered computing** → *Mobile devices*; Ubiquitous and mobile computing systems and tools; Ubiquitous and mobile computing design and evaluation methods.

Authors' addresses: Weiwei Jiang, The University of Melbourne, Melbourne, Australia, weiwei.jiang@student.unimelb.edu.au; Gabriele Marini, The University of Melbourne, Melbourne, Australia, marinig@student.unimelb.edu.au; Niels van Berkel, University College London, London, United Kingdom, n.vanberkel@ucl.ac.uk; Zhanna Sarsenbayeva, The University of Melbourne, Melbourne, Australia, zsarsenbayev@student.unimelb.edu.au; Zheyu Tan, Japan Advanced Institute of Science and Technology, Nomi, Japan, zheyut@gmail.com; Chu Luo, The University of Melbourne, Melbourne, Australia, chul3@student.unimelb.edu.au; Xin He, Anhui Normal University, Wuhu, China, xin.he@ahnu.edu.cn; Tilman Dingler, The University of Melbourne, Melbourne, Australia, tilman.dingler@unimelb.edu.au; Jorge Goncalves, The University of Melbourne, Melbourne, Australia, jorge.goncalves@unimelb.edu.au; Yoshihiro Kawahara, The University of Tokyo, Tokyo, Japan, kawahara@akg.t.u-tokyo.ac.jp; Vassilis Kostakos, The University of Melbourne, Melbourne, Australia, vassilis.kostakos@unimelb.edu.au.

2474-9567/2019/12-ART136 $15.00
https://doi.org/10.1145/3369834

Additional Key Words and Phrases: Near-Infrared spectroscopy, liquid sensing, food scanner, mobile sensing, machine learning

ACM Reference Format:
Weiwei Jiang, Gabriele Marini, Niels van Berkel, Zhanna Sarsenbayeva, Zheyu Tan, Chu Luo, Xin He, Tilman Dingler, Jorge Goncalves, Yoshihiro Kawahara, and Vassilis Kostakos. 2019. Probing Sucrose Contents in Everyday Drinks Using Miniaturized Near-Infrared Spectroscopy Scanners. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 136 (December 2019), 25 pages. https://doi.org/10.1145/3369834

1 INTRODUCTION

Advances in food scanning and sensing have enabled the emergence of mobile food scanners [39]. Previous work has shown great promise for a wide range of food and drinks scanning tasks, such as maturity estimation [6, 10], contamination detection [25, 27], and most importantly, nutrition content probing [16, 38]. In our work, we are especially interested in probing contents in food, and particularly, in everyday drinks.

Different methods have been proposed for probing everyday drinks. For instance, Nutrilyzer [38], Al-Light [28] and iTube [5] show great accuracy when using bespoke hardware to measure proteins, ethanol, and peanut allergen in milk, alcohols, and liquid food, respectively. However, it is difficult to generalize this approach to other types of liquids without substantial hardware customization. In this paper, we show a method that can measure sucrose contents accurately in everyday drinks, and can be generalized to other liquid sensing tasks that consider other contents.

Our method is based on Near-Infrared Spectroscopy (NIRS). NIRS is a rapid and non-invasive technology which is used for retrieving information of an object at a chemical level [3]. An increasing number of studies use NIRS to analyze food constituents, particularly in the agriculture industry [7]. However, conventionally, this technology was expensive [48] and designed for use exclusively in scientific laboratories [15, 35]. More recently, it has been possible to miniaturize this technology, which has opened up a wide range of opportunities for mobile devices. For example, prior work shows how handheld NIRS can be used to identify specific pharmaceuticals [21]. Our work contributes towards the growing body of research on leveraging NIRS to identify everyday objects. We show that NIRS can be used to accurately measure the amount of a particular content – sugar – in commercially available drinks. Due to the small size and mobility of our prototype, we argue that users with this technology can directly measure the sugar concentration of drinks they consume, and thus have an opportunity to have a more informed diet.

More crucially, however, we recognize that the research field is likely to become fragmented if, as a community, we continue to publish disjointed studies, each one showing that NIRS can detect yet another nutrient (*e.g.*, salt, gluten, alcohol, etc.). Therefore, in this paper, we develop and validate a protocol that can be followed when designing a NIRS scanner for any arbitrary content. Our work makes two contributions to the growing body of literature that utilizes NIRS for food sensing tasks.

- First, we design a handheld prototype device that uses NIRS to measure sucrose concentration in liquids, and test this with both controlled lab samples and commercial drinks. Our results show high accuracy, and robustness to substantial variations in ambient temperature and light.
- Second, we validate a calibration, sampling and analysis protocol that can be replicated, and which allows researchers to use NIR spectrometers to measure other food and drink qualities, such as alcohol, protein, and fat. A wider range of applications, such as detecting counterfeit or diluted perfume as we show, are also possible using our methodology and our open-sourced design and software[1].

Our work takes a step towards enabling our community to develop a robust "food scanner", whereby a consistent methodology is used to collect empirical evidence on the identification of a range of ingredients, nutrients, and allergens of interest to consumers in everyday settings.

[1] https://github.com/HighTemplar-wjiang/IMWUT-NIRS-Liquid

2 RELATED WORK

We first describe existing work related to the use of mobile devices in the context of Near-Infrared Spectroscopy (NIRS), liquid sensing, and food scanning. Generic liquid sensing methods are capable of sensing sucrose concentration in liquids. We are particularly interested in those applying a spectroscopic method since it is non-invasive and provides rapid and accurate sensing results. Alternative methods beyond spectroscopy have been utilized in other studies, but their implementations have several limitations, and thus are not suitable for our targeted task (*i.e.*, sensing sugar contents in drinks), or cannot be generalized to other tasks.

2.1 Mobile Near-Infrared Spectroscopy Methods

NIRS is one of the spectroscopic methods that utilize Near-Infrared (NIR) light. It is a non-destructive and rapid chemical analysis method [45] – which measures interactions between matter and electromagnetic radiation. NIRS is especially popular for sensing organic content, including in detecting various contents in liquids, due to the characteristics of NIR lights [32]. It leverages the phenomenon that different chemical bonds can absorb significantly more NIR lights with specific wavelengths. By analyzing the absorbance spectrum, the chemical composition information can be retrieved [3]. NIRS has been adopted for a number of applications as a non-invasive sensing method, such as human hemoglobin (Hb) detection [26], quality control in the pharmaceutical industry [40], and food maturity monitoring [2].

While most of the studies with NIRS are carried out in laboratories, recently available commercial off-the-shelf portable and miniaturized NIRS scanners have enabled additional applications. For instance, using a portable NIRS scanner, Taira *et al.* conducted an experiment to predict the sugar contents in sugar canes [47], while Yu *et al.* presented a handheld Visible-Near-Infrared (VIS-NIR) spectrometer to estimate sugar content in pears with standard error of prediction value of 0.46 °Bx. Furthermore, Ibanez *et al.* successfully predicted the taste of tomatoes by determining related compounds [16]. Because their method can be applied *in situ*, the analysis procedure has been significantly simplified and accelerated.

To make NIRS further accessible, Klakegg *et al.* studied the possibility of placing a miniaturized NIRS scanner into the hands of people without prior knowledge on NIRS [22, 23]. The study by Klakegg *et al.* suggested that with a well-designed system, non-experts can utilize the NIRS method for daily identification tasks. Furthermore, Klakegg *et al.* developed a NIRS system that supported nurses in pill identification, which could significantly reduce medicine management errors [21]. They envisioned that NIRS could be successfully applied in out-of-laboratory settings by non-experts. Adopting NIRS for sucrose content sensing in everyday drinks remains an under-explored area given the challenges associated with scanning liquids. Our work aims to identify these issues and provide solutions that will improve the ability to use NIRS for liquid sensing. We further envision that our work can be extended to applications beyond sensing sucrose contents in liquids such as allergen detection, counterfeit perfume detection, etc.

2.2 Liquid Sensing Methods

Besides NIRS, there are several other methods for liquid sensing in general, which have been, or potentially can be, used for sensing sucrose contents in beverages, but with different limitations. One of the most frequently used methods for liquid sensing is the use of a refractometer [11]. A refractometer is a mobile device that measures the solution's index of refraction that is related to the content's concentration. Refractometers have been widely used in agriculture for, *e.g.*, Brix sensing (sweetness sensing, including sucrose) and alcohol sensing [12]. There are, however, some limitations to the application of refractometers. In particular, it is not trivial to extend a specific refractometer to measure other contents, as a refractometer is mostly tuned to a designated wavelength that may not be able to detect other contents [31]. Besides, it is also difficult to measure opaque liquids which do not allow light to pass through (*e.g.*, milk, juices).

Alternative methods leverage different sensing principles with some disadvantages compared to NIRS. For instance, Jiang *et al.* presented a graphene test paper that successfully recognized several complex solutions, including four saline solutions and four polymer solutions [18]. A downside of this approach is that it requires a particular material (graphene-cellulose nanocomposite test paper) that is not yet commercially available and cannot be reused. There also exist certain low-cost liquid sensing technologies, but their application scenarios are limited. For example, Hassan *et al.* demonstrated an inkjet-printed radio frequency (RF) microstrip that recognizes water, ethanol, water/ethanol 50:50 mixture and synthetic engine oil, while wider-ranging applications for sensing other contents, including sucrose, remain challenging [13]. The main challenge was the technical limitations of RF microstrip sensing. Specifically, they utilized the phenomenon that resonance frequencies of the RF microstrip change with respect to dielectric constants. The dielectric constants, however, do not change drastically with either different concentrations of sucrose nor all materials, resulting in negligible resonance frequency shift that is difficult to be measured even by sophisticated laboratory equipment.

2.3 Mobile Sensing for Food and Beverages

Broadly speaking, we are interested in mobile sensing methods that can be used with food and beverages in everyday scenarios. The most relevant study is the smart ice cube presented at UbiComp by Matsui *et al.* [28]. The smart ice cube utilizes a single frequency of NIR (1450 nm) assisted by RGB lights to predict alcohol concentration in 13 commercially available liquors with approximately 2% v/v errors. Our study does not only extend this work to sucrose content sensing, but also provides the basis for a more flexible food scanning method by using arbitrary frequencies.

Previous work that considers liquid sensing cannot estimate sucrose contents with great accuracy. For instance, Fan *et al.* developed a capacitive sensor that can detect the liquid level inside a cup, although it cannot identify the liquid itself [8]. Lester *et al.* presented a smart cup equipped with pH and conductivity sensors to classify 68 drinks with up to 79% accuracy [24]. Furthermore, Mezgec *et al.* adopted deep neural networks and trained with a food image database, including drinks and soups [30]. Though they achieved 86.72% accuracy in identifying food and drinks in the database, their system had 55% top-five accuracy in a field test, which reduces its applicability in real-life scenarios compared to NIRS.

Besides beverage sensing, in general, previous work has aimed to develop food scanning for mobile contexts [39]. Some of the most popular methods are, in particular, colorimetric-based methods, photoacoustic-based methods, and spectroscopic methods. The colorimetric method can be used to determine the concentration of chemical content in a solution with the help of a color reagent [33]. Mobile colorimetric methods for food scanning utilize the camera of a smartphone for image or signal processing. For instance, Coskun *et al.* developed a smartphone attachment (iTube) for allergen detection [5], and they successfully identified three different Mrs. Fields Cookies that did or did not include peanuts with $\geq$ 1 ppm concentrations. For beverages, Masawat *et al.* developed a photography lightbox for detecting Tetracycline[2] concentrations in milk [27]. The main disadvantage of the colorimetric method is that it requires specific color reagents for particular contents to be detected, as well as time-consuming sample preparation that requires professional training (*e.g.*, iTube requires 20-minute incubation after grinding the sample and mixing it with 50~60 °C hot water) [39]. In contrast, the other two methods, photoacoustic and spectroscopic methods, including NIRS, are more rapid (up to several seconds), non-invasive and *in situ* without requiring professional sample preparation.

The photoacoustic method utilizes the photoacoustic effect that is distinct from, but related to, NIRS. It utilizes the variations of light absorption to different material samples. By transmitting a modulated light wave, a sound wave can be formed and detected, after interacting with the material sample, by a microphone or a piezoelectric sensor [41]. For instance, Rahman *et al.* proposed a novel mobile sensing system (Nutrilyzer) that

[2]Tetracycline is an antibiotic used in agriculture, and its residues might affect the health of the consumers.

successfully estimated protein concentrations in four kinds of milk with RMSE=0.2468 [38]. Furthermore, the authors demonstrated that Nutrilyzer could also characterize adulterant in milks, including detergent, salt, starch, and water. The photoacoustic method, however, is particularly useful for detecting low concentration contents. As also shown by authors in [38], the accuracy of alcohol detection by Nutrilyzer is relatively lower with RMSE=15.23, compared to Rum:Water ratio detection with RMSE=2.84 (Rum mixed with water, with ratios of 100%, 80%, 60%, 40%, and 20% respectively).

Finally, there are also several mobile systems based on spectroscopic methods, not limited to NIRS, that show promising results in various food scanning tasks. For instance, Goel *et al.* presented a hyperspectral imaging system (HyperCam) that successfully detected the quality and the ripeness of avocados in 50 pairs with 94% accuracy, by expanding the spectrum from the visible region into NIR region (820 nm, 950 nm, and 990 nm) [10]. Furthermore, Das *et al.* proposed a miniaturized ultra-violet (UV) spectrometer that was particularly useful for testing fruit ripeness, based on the UV fluorescence of Chlorophyll peaked at 680 nm [6]. There are other spectroscopic methods using different wavelengths such as X-ray photoelectron spectroscopy (XPS) [42] and visible spectroscopy (VIS). However, UV and X-ray lights are well-known as containing harmful electromagnetic radiations to the human body [1, 36]. It is less preferable to adopt such technologies for everyday settings. In consequence, we are particularly interested in NIRS for its versatility and safety on detecting a wide range of contents in liquids.

3 APPARATUS DESIGN AND EXPERIMENTAL PROTOCOL

We investigate the use of miniaturized NIRS scanners for measuring sucrose contents in liquids. It is worth noting that NIRS is a non-selective sensing technology, *i.e.*, NIRS responds to changes in the experimental conditions. This is because NIRS utilizes sensors to detect the variation of NIR light after interacting with the sample. Conditional changes that can affect the NIR light, therefore, directly affect the scanning results. Most conditional changes, however, do not drastically impact the result, but some changes should be noted to avoid inconsistent results for the same sample. For instance, changing the surface shape or the orientation of the sample may greatly alter the light scattering path. To this end, it is important to keep constant those significant conditions to ensure meaningful results.

In particular, when considering scanning a liquid through a container (since liquids themselves are shapeless), the structure and material of the container itself will directly affect the scanning results. It is also challenging to probe a liquid directly on its surface: liquid fluctuations further add to the challenge as the scanner is sensitive to the relative position between the sample and the sensor due to the nature of NIR light.

3.1 Prototype and Experimental Setup

To address the challenges of sensing liquids, we designed a clamp to keep the relative position of the scanner and the liquid samples constant as shown in Figure 1. We designed and fabricated the clamp using a Makerbot Z18 3D printer. The body size was $H \times W \times D = 83mm \times 114mm \times 64mm$ to wrap a miniaturized NIRS scanner (Texas Instruments NIRScan Nano [17]). The sample liquid for analysis can be kept in a tube that is clamped inside the tube tunnel that is in front of the scanning window, the tube tunnel is wrapped by a tube holder module for fitting different sized tubes, as shown in Figure 1 (a), (d) and (e). We have also attached a Raspberry Pi A+ with a 2.8-inch screen (Adafruit PiTFT) for controlling the scanner. A Li-Po battery with a power management module (PowerBoost 1000C) is also included, meaning that the prototype runs on battery and is effectively a mobile clamp. The design resources, as well as the software, are open-sourced [3].

Previous work by Klakegg *et al.* [21] on identifying solid objects already addressed challenges related to user-induced errors and the potential negative impact of ambient light. Our design further addresses those challenges

[3] https://github.com/HighTemplar-wjiang/IMWUT-NIRS-Liquid

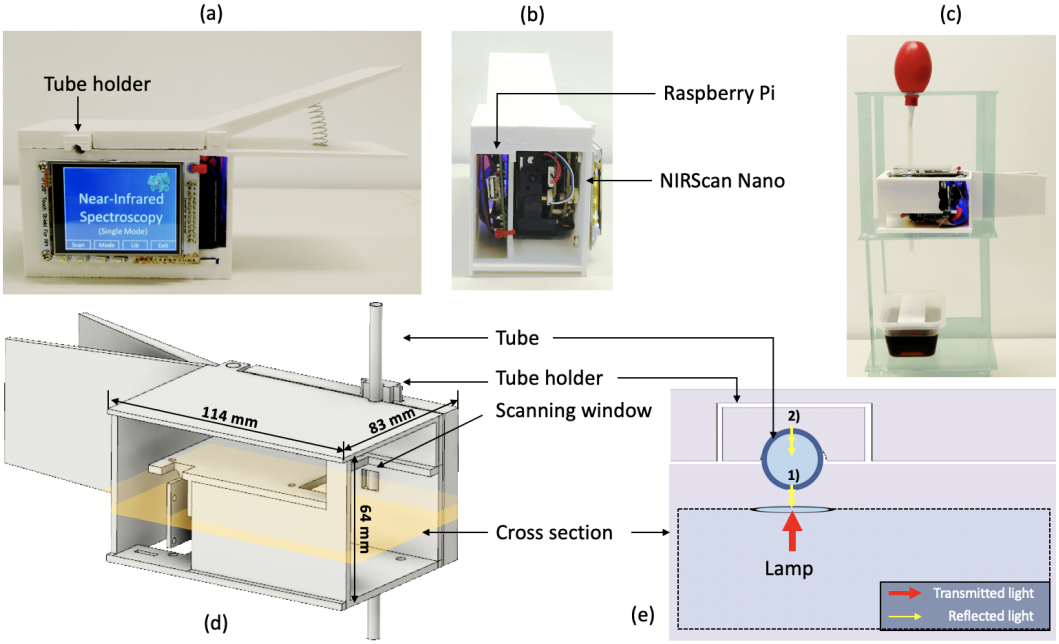


Fig. 1. Design, prototype, and experimental setup. (a) Front view of the prototype, including an Adafruit PiTFT 2.8-inch screen. (b) Side view of the prototype, including a NIRScan nano and a Raspberry Pi A+. (c) Experimental setup, including a cuvette containing a Coca-Cola Zero sample, a bulb blower attached with a transparent tube for pumping liquid, and our prototype to perform liquid scanning. (d) Bird's eye view of the clamp design. (e) Cross-section including scanning window and tube. Lights that are received by the scanner include: 1) lights reflected by the near-end air/tube interface and the tube-liquid interface, 2) lights reflected by the far-end tube-liquid interface and air-holder interface.

by using a tube following a protocol described in Subsection 3.2, thus reducing errors caused by shapeless and fluctuations that cause distortions of NIR spectra. Our design also differs from the one by Matsui *et al.* [28], as their prototype requires direct contact with the liquid by sinking the device into the sample, which introduces hygienic concerns in practice. Our prototype can take a sample from a tube – such as a straw for consuming drinks – without direct contact with the sample, thus reducing the risk of contamination and improving the practicality of our approach.

To evaluate our prototype, we built an experimental setup, as shown in Figure 1 (c). We fabricated a Perspex shelf using a laser cutter. Our prototype was placed in the middle of the shelf, with a transparent UL224 tube (6 mm diameter, but the design of the prototype allowed us to change the diameter easily) clamped. A cuvette was put on the bottom containing a liquid sample (Coca-Cola Zero in Figure 1). We used a bulb blower to pump the liquid inside the tube during the scanning.

It is worth noting that the scanner works in reflection mode (*i.e.*, detecting the light intensity reflected back by the sample), rather than transmission mode (*i.e.*, detecting the light intensity after passing through the sample). The latter is mostly used for liquid sensing. We chose reflection mode for compatibility of both opaque liquids and translucent liquids. In consequence, as shown in Figure 1 (e), the detected NIR spectra mainly include two parts: 1) The lights reflected by the near-end air/tube interface and the tube-liquid interface. 2) The lights reflected by the far-end tube-liquid interface and air-holder interface. For opaque liquids part 1) contributes much more

Fig. 2. Illustration of the sampling procedure.

Table 1. Configuration of NIRScan Nano during the experiments.

Method	Start nm	End nm	Width nm	Dig. Resolution	Exposure ms	Num. of scans by device to average	Num. of scans by user per sample
Hadamard	900	1700	7.03	228	0.635	6	6

in the received lights than 2), while for translucent liquids part 2) contributes much more than the other. In consequence, spectra with significant magnitude for both opaque liquids and translucent liquids can be detected.

3.2 Experimental Protocol

We conducted two studies: the first one focused on estimating sucrose concentrations on pre-defined solutions, while the second study focused on estimating sucrose contents as well as classifying everyday drinks. Both studies used the same sampling protocol.

3.2.1 Procedure. For analyzing samples using NIRS, it is recommended by literature to have a minimum of 12 spectra for each raw material (two lots per material, two samples per lot, three scans per sample) [32]. In our experiment, for each liquid type, we used a single lot, took three samples[4] per lot, and for each sample we conducted six consecutive scans, resulting in 18 spectra per liquid type as shown in Figure 2. We decided to take more than three scans as miniaturized NIRS scanners tend to be noisier than benchtop ones. In addition, we have not taken multiple lots from the same liquid (material), as we did not aim to monitor production procedures or conduct a quality test. Between two consecutive scans, a one-second interval was set to minimize interference. Before scanning a new liquid, the tube was flushed with pure water and dried to prevent contamination. The tube remained fixed during the whole experiment. Following this procedure, scanning a single sample took approximately half a minute.

3.2.2 Configuration. We configured the NIRS scanner as shown in Table 1, based on the device's guidebook [17] and prior work [21]. Typical considerations related to these parameters include increasing the signal-to-noise ratio (SNR) using a specific scan method, a trade-off between scanning duration and noise, wavelengths that can better respond to the targeted contents, and choosing a resolution that can give higher accuracy.

We adopted the Hadamard scan method, which was preferred to the alternative Column method in most scenarios, as it could provide greater SNR by spreading the noise to the whole wavelength span. The start and end wavelengths were chosen for the maximal possible spectral span, *i.e.*, from 900 nm to 1700 nm. This is due to the phenomenon that the chemical bonds disperse NIRS signals across a wide range of wavelengths, and hence, it is prudent to capture multiple wavelengths for acquiring reliable results [32]. The width of the generated light pattern was set to 7.03 *nm* for an improved distinction of the incident light. In consequence, the digital

[4]A sample means carrying out an action of pumping liquid from the cuvette.

resolution was set to 228, resulting in double the oversampling rate necessary to fulfill the Nyquist-Shannon sampling theorem [14]. The exposure time was set to 0.635 ms to prevent potential attenuation induced by possible movement during a scan. Six repeated back-to-back scans were performed and averaged to further reduce the noise[5]. This results in a trade-off between signal noise and scanning time.

All results were obtained using Python's SciPy [19] and scikit-learn [37] packages. We wrapped the communication protocol library for the NIRScan Nano into a Python package for both Version 2 and 3, which is made public via open-sourcing. Thus one can easily build proprietary applications based on our prototype. We have verified that the library could work on Raspberry Pi and laptops running Windows or Linux, and consistent results were obtained across these platforms.

4 STUDY 1 - QUANTITATIVE EXPERIMENT OF SUCROSE CONTENT SENSING

In this section, we describe the evaluation using our prototype for sucrose contented liquids, following the experimental protocol presented in Subsection 3.2. The purpose of the study is to predict sucrose concentrations in 25 predefined sucrose-water solutions (referred to as sucrose solutions in this paper). Although a number of studies have reported sensing sucrose contents using conventional NIRS hardware [20, 44], little work has focused on running such an experiment with a miniaturized NIRS scanner [21, 22]. The main outcome of this study is a linear regression model, and an evaluation of all the collected data for selecting the most explainable wavelengths (*aka.*, characteristic wavelengths). We also show that machine-learning based methods can give reliable results with low errors.

4.1 Data Collection

We prepared a total of 25 sucrose solutions with different concentrations ranging between 0 and 20.0 g/100ml. Higher concentrations are out of scope since it is rare to have a liquid with sucrose content beyond 20.0 g/100ml (*e.g.*, highly sucrose-contented energy drinks such as Monster Green contains 11.4 g/100ml sucrose). We first prepared 21 solutions at 1.0 g/100ml intervals (*i.e.*, 0.0, 1.0, 2.0, etc.). To investigate the capability of the prototype in more extreme cases, we prepared 4 additional solutions with consecutive halving concentrations from 0.5 g/100ml diluted from a 1.0 g/100ml sample (*i.e.*, 0.5, 0.25, 0.125 and 0.0625 g/100ml). The experiment was conducted on a cloudy afternoon, with the room temperature between 20 and 25 °C, with humidity between 20% and 30%. The indoor light condition was in line with normal office conditions.

The sucrose we used was ordinary, off-the-shelf, white sucrose (cane sugar). A digital kitchen scale ($\pm 0.01\ g$ precision) and a plastic graduated cylinder ($\pm 2\ ml$ precision in 20 °C) were used to measure the weight and liquid volume respectively. It should be noted that mixing a specific amount of sucrose with 100 *ml* of water did not necessarily result in a 100 *ml* solution, and the volume of the mixed solution should be measured for higher accuracy.

4.2 Preprocessing

Following our sampling protocol, we collected 450 spectra (18 spectra × 25 solutions). Before analyzing, preprocessing is required as partially documented in the literature [32]. The preprocessing method includes two steps: calibration and smoothing. Additional preprocessing methods reported in the literature [21, 32] are not necessary since our design controlled the conditions. The original hardware (scanner) focuses on retrieving consistent NIRS spectrum given the same sample in the sample ambient conditions. However, it does not address the issue

[5]In all other parts of this paper. We define a scan as a user initialized scan command and acquires one spectrum as a result (*e.g.*, push the scan button on the scanner, or send a command to the scanner via a program.). Internally by the device, six scans are performed and averaged into one spectrum, but users are unaware of.

that the scanning results may vary with respect to ambient conditions due to its sensitivity issue. Hence, proper preprocessing is required to improve the performance across different scenarios.

4.2.1 Calibration. The raw NIRS spectra were calibrated to address the sensitivity issue. We calibrated the collected data by defining the relative absorbance. Besides calibration, the relative absorbance works as a normalization for the non-standard scanning method we use for our design as a trade-off between usability and performance. In our design, we adopted reflection mode for liquid sensing, while conventionally, NIRS with the alternative transmission mode is preferred for quantitative analysis [32]. Consequently, when analyzing a translucent liquid with our experimental setup, the majority of the detected NIR intensity is obtained from the reflected lights after passing through the liquid, which works as a mixed-mode of the two standard modes.

To correlate the detected light intensity and the solution concentration, we derive the relative absorbance based on the Beer-Lambert Law [46] which can be expressed as

$$I = I_0 e^{-\sigma l N}, \tag{1}$$

where I_0 is the intensity of incident IR light, σ is the absorption cross section, l is the optical path length, N is the concentration of particles. To predict N, Equation 1 can be transformed to

$$N = -\frac{1}{\sigma l} \cdot \log \frac{I}{I_0} \tag{2}$$

$$= \frac{1}{\sigma l \cdot log_{10} e} \cdot A \tag{3}$$

where log and e are the natural logarithm and its base respectively, $A = log_{10}(I_0/I)$ is defined as the absorbance. Furthermore, the particle concentration is defined by $N = N_v/m$, where N_v is the mass concentration (*e.g.*. with unit g/100ml) and m is the molecular mass with unit g/mol. Therefore, we can express the mass concentration as

$$N_v = \frac{m}{\sigma l \cdot log_{10} e} \cdot A. \tag{4}$$

Equation 3 shows a linear correlation between the absorbance and concentration, since m, σ and l are constants in our experimental setup. However, the incident light intensity I_0, and the output intensity I are difficult to measure in our prototype. Therefore, we substitute them as I_0' and I' respectively, where I_0' is the measured intensity of a reference liquid (*e.g.*, for pure water), and I' is the detected light intensity for the liquid sample. Since the difference between I_0' and I' is primarily caused by the difference in concentration of the liquid solution, we can then define relative absorbance as a substitution in Equation 3, resulting

$$A' = \log_{10} \frac{I_0'}{I'} \tag{5}$$

$$N_v = C_1 \cdot \frac{m}{\sigma l \cdot log_{10} e} \cdot A' + C_2, \tag{6}$$

where C_1 and C_2 are constants caused by possible scaling of the substitution. It should be noted that this substitution process also deals with the spectrum shift caused by ambient condition changes. Together with applying the Beer-Lambert law, the spectra are calibrated against both our non-standard scanning method and the ambient conditional changes.

In practice, we first record 18 spectra of pure water and average them to establish our reference spectrum. Then, for all spectra collected at various concentration levels, we divide the obtained raw spectrum (light intensity) by the reference spectrum and apply the logarithm to the spectrum. The resulting spectrum is what we use in the rest of the analysis. This process works as calibration. Thus it should be noted that the pure water spectra (reference spectrum) should be re-collected when the ambient conditions are substantially changed (*e.g.*, temperature, brightness, humidity, etc.).

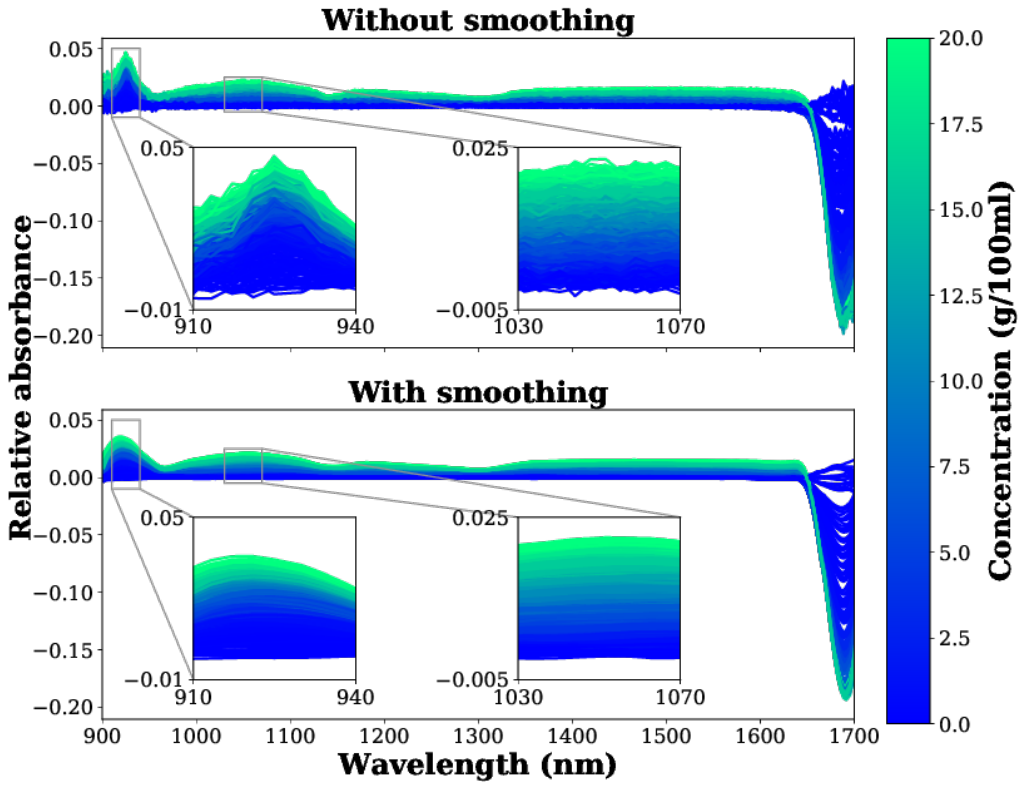


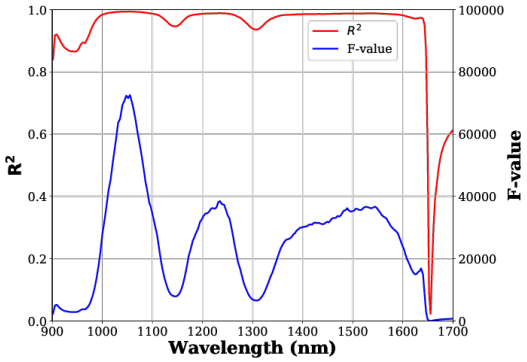
Fig. 3. Preprocessed 450 NIRS spectra across 25 sucrose concentrations ranging between 0 and 20 g/100ml. Top: spectra without smoothing, bottom: spectra with smoothing.

4.2.2 Smoothing. After calibration, we applied a Savitzky-Golay filter for smoothing collected spectra data individually [43]. The Savitzky-Golay filter has been widely used in chemometrics to increase the SNR without greatly distorting the signal [32]. It utilizes a convolution process involving successive adjacent data points. The Savitzky-Golay filter equation is expressed as

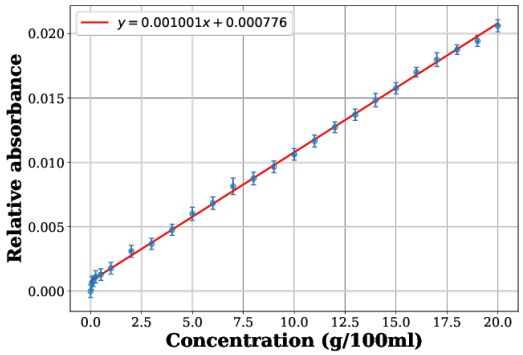
$$y_j^* = \sum_{i=-\frac{m-1}{2}}^{\frac{m-1}{2}} C_i y_{j+i}, \tag{7}$$

where y_j^* is the filtered jth data point, m is the convolution window size and C_i is the pre-defined polynomial coefficients according to the window size and degree of the polynomial. We chose $m = 21$ with third degree, resulting in a 140.6 nm window width, considering that the miniaturized NIRS scanner can give more noisy results due to its susceptibility to environmental changes.

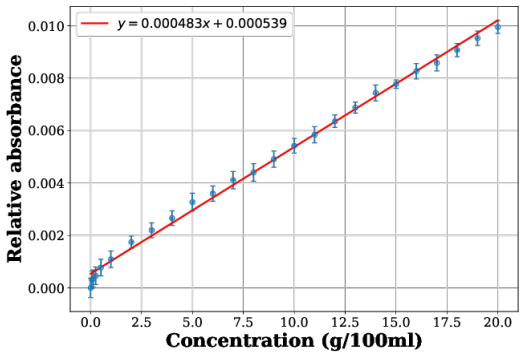
4.2.3 Results of Preprocessed Data. Figure 3 shows the preprocessed results for the collected spectra data. Mean values of all 18 spectra of water were used as reference data I_0' in Equation 5. The relative absorbance increases with respect to the sucrose concentration, except at the end of the spectrum that is an over-sensitive part of the



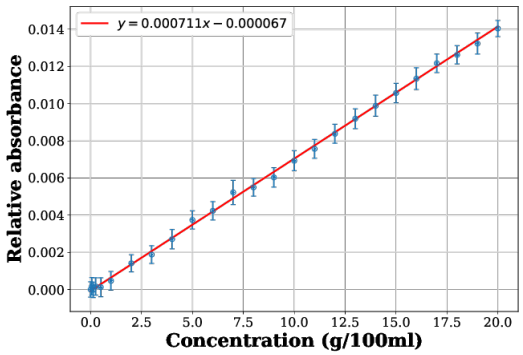
(a) Wavelength evaluation



(b) 1055 nm, $R^2 = 0.9939$, F-score=72550.64



(c) 1235 nm, $R^2 = 0.9885$, F-score=38486.06



d) 1545 nm, $R^2 = 0.9879$, F-score=36726.13

Fig. 4. Wavelength evaluation results in three overtone regions. (a) R^2 and F values with respect to wavelengths. (b) Peak wavelength (1055 nm) in the third overtone region. (c) Peak wavelength (1235 nm) in the second overtone region. (d) Peak wavelength (1545 nm) in the first overtone region.

NIRS scanner. We can also observe that after smoothing, samples between two consecutive concentrations are better bounded, with averaged SNR increasing from 4.88 to 6.78.

Since we adopted water as the reference liquid that has the lowest absorbance (*i.e.*, highest received intensity), the spectra data were automatically normalized between 0 to 0.05 with the logarithm operation (note that the reference spectra is the numerator to avoid negative number). This normalization can be helpful for the following regression models.

4.3 Identifying the Characteristic Wavelengths

The NIR absorption spectra are mostly complex and spread in broad overlapping NIR bands, with several characteristic wavelengths that are more responsive to specific chemical bonds than others [20, 32, 44]. To find those characteristic wavelengths and estimate the concentration, we adopted the simplest single-linear regression model using the ordinary least square (OLS) method for every possible wavelength [9]. As shown in Figure 4 (a), R^2 values for most of the wavelengths are above 0.9, with three NIRS overtone regions distributed in the spectra.

The peak wavelengths with the highest F values are 1055 nm, 1235 nm and 1545 nm, with formulae as follows, respectively,

$$A'_{1055nm} = 0.001001 \times N_v + 0.000776, \tag{8}$$

$$A'_{1235nm} = 0.000483 \times N_v + 0.000539, \tag{9}$$

$$A'_{1545nm} = 0.000711 \times N_v - 0.000067, \tag{10}$$

where A' is the relative absorbance and N_v is the labeled concentration in g/100ml.

4.4 Regression Models

We finally compared the different regression models, including linear regression methods and some common machine learning methods using the Python scikit-learn library [37]. We used these models to quantify the performance of different methods. Our configuration parameters are also described below, with the remaining parameters kept as default.

- Single linear regression (SLR): Single linear regression with OLS method is adopted for 1055 nm, 1235 nm and 1545 nm, respectively. SLR is the simplest model but most explainable.
- Multiple linear regression (MLR): Multiple linear regression with OLS method is adopted for the three selected wavelengths (1055 + 1235 + 1545 nm) and all wavelengths respectively. MLR is the expansion of SLR to include more correlated features.
- Support vector machine (SVM): RBF kernel is adopted with all wavelengths involved. SVM is suitable for high-dimensional as in our case (228 features in total).
- Random forest (RF): Mean squared error (MSE) criterion is adopted with all wavelengths involved. RF is also suitable for high-dimensional data.
- Multi-layer perceptron (MLP): A three-layer structure with 228 units in the hidden layer is adopted with all wavelengths involved, with *identity* activation function, with *lbfgs* solver, with adaptive training rate. This model works as a second-order polynomial regression, assuming the regression model is not necessarily linear.

We used two training/test data splitting schemes for evaluation. First was the skip-one-as-test scheme, where we adopted every other consecutive concentration as the training set, to evaluate the models when estimating unseen concentrations. The second scheme split a ratio of data for each concentration as test data (split-ratio-as-test). For a fair comparison, we set the training/test ratio as 50%/50%. Figure 5 shows an example of the SLR regression result with wavelength = 1055 nm.

4.5 Results

Table 2 shows the results for the regression models mentioned above. For the SLR models using one of the selected wavelengths, all three models had similar performance, for both training and test sets, for both skip-one-as-test and split-ratio-as-test schemes, topped by the wavelength with highest F-value (1055 nm). For higher accuracy, involving multiple wavelengths is necessary. On the one hand, MLR with all wavelength involved showed the lowest root-mean-square-error (RMSE) values for the training set, but with highest RMSE values for the test set in all methods with lower R^2 values ($R^2 < 0.98$). This indicates overfitting for the MLR model, suggesting that wavelength selection should be performed for linear regression models. On the other hand, MLR model with three selected wavelengths (1055 + 1235 + 1545 nm) shows an improved performance without overfitting.

For the other machine learning-based regression models, different behaviors are observed. SVM and MLP are relatively stable models showing close RMSE values for both training and test datasets. MLP gives the best RMSE performance in test datasets with reasonable R^2 values in both schemes ($R^2 > 0.99$), while also being the most complex model we used. Thus, the less complex RF model has a trade-off between performance complexity.

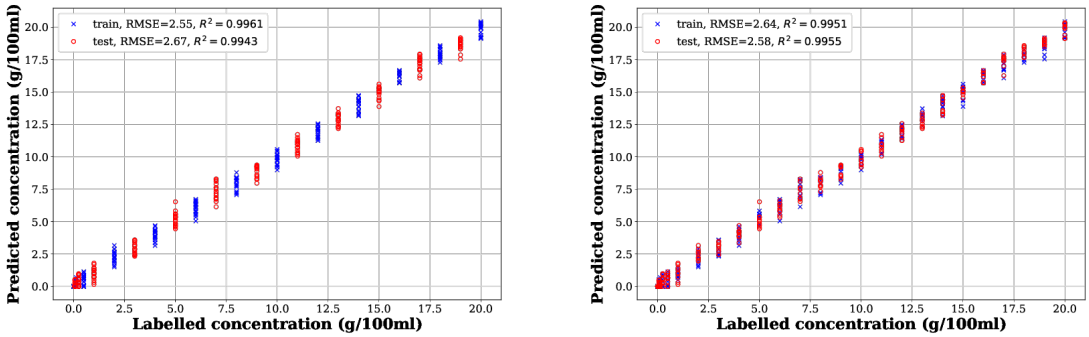


Fig. 5. Predictions with single-linear OLS regression model in different schemes (wavelength=1055 nm). Left: skip-one-as-test, right: split-ratio-as-test.

Table 2. Comparison of regression models, (SLR = single linear regression, MLR = multiple linear regression, SVM = support vector machine, RF = random forest, MLP = multi-layer perceptron). Reported number is error (g/100ml). Numbers in parentheses are R^2 values.

Method	**Skip-one-as-test**		**Split-ratio-as-test**	
	Train	*Test*	*Train*	*Test*
SLR (1055 nm)	2.55 (0.9961)	2.67 (0.9943)	2.64 (0.9951)	2.58 (0.9955)
SLR (1235 nm)	2.97 (0.9927)	3.01 (0.9907)	3.07 (0.9909)	2.91 (0.9927)
SLR (1545 nm)	3.15 (0.9908)	3.29 (0.9868)	3.29 (0.9880)	3.15 (0.9900)
MLR (1055 + 1235 + 1545 nm)	0.45 (0.9956)	0.53 (0.9932)	0.45 (0.9954)	0.48 (0.9948)
MLR (all wavelengths)	**0.05 (0.9999)**	1.07 (0.9720)	**0.05 (1.0000)**	1.68 (0.9349)
SVM	0.52 (0.9941)	0.57 (0.9922)	0.58 (0.9924)	0.50 (0.9944)
RF	0.17 (0.9994)	0.65 (0.9897)	0.21 (0.9990)	0.46 (0.9951)
MLP	0.25 (0.9986)	**0.29 (0.9979)**	0.27 (0.9983)	**0.26 (0.9984)**

However, RF did not provide much better performance than the MLR method for the selected wavelengths. In conclusion, MLR with selected wavelengths is preferred in this task.

5 STUDY 2: PROBING EVERYDAY DRINKS

Next, we demonstrate how our prototype can be used to probe everyday drinks, in particular, estimating their sugar concentration. In this study, we consider 18 drinks which are commercially available, including 14 translucent drinks and 4 opaque drinks. Since NIRS cannot penetrate opaque objects as well as translucent ones, the regression model we trained in Section 4 cannot work well for non-translucent liquids. To address this issue, we show that the classification model is a reliable alternative approach without changing the experimental protocol. Also, we show that the performance does not change drastically across nine ambient conditions.

Our study is different to previous ones in literature that utilized NIRS for identifying everyday drinks, as we are the first to utilize a miniaturized NIRS scanner for such tasks (*e.g.*, Chen *et al.* successfully identified 12 types of teas with > 90% accuracy, but an expensive benchtop NIRS scanner was used that is not practical to move [4]).

[5]Here *translucent* means the light can penetrate the liquid, either partially or completely.

Table 3. Nutrition statistics for everyday drinks.

Soft drink	Sugar (g/100ml)	Fat (g/100ml)	Protein (g/100ml)	Color	Carbonated?	Translucent?
Coca Cola (classical)	10.6	0	0	Brown	✓	✓
Coca Cola (no sugar)	0.0	0	0.05	Brown	✓	✓
Coconut Water	3.9	<1	<1			✓
Fruit Drink Apple	6.0	<1	<1	Yellow		✓
Fruit Drink Blackcurrant	5.9	<1	<1	Red		✓
Fruit Drink Tropical	6.0	<1	0.1	Orange		✓
Lipton Lemon	4.5	<0.1	<0.1	Brown		✓
Lipton Original	4.6	<0.1	<0.1	Green		✓
Lipton Peach	4.5	<0.1	<0.1	Brown		✓
Milk Lite	4.9	1.3	3.4	White		
Milk Original	4.4	3.4	3.4	White		
Monster Energy (green)	11.4	0	0	Brown	✓	✓
Monster Energy (red)	4.2	0	0	Light brown	✓	✓
Monster Energy Original	6.3	0	0	Brown	✓	✓
Qoo	11.3	0	0	Light green		✓
Soy Milk Lite	2.1	1.8	3.0	White		
Soy Milk Original	2.1	3.0	3.0	White		
Water	0.0	0	0			✓

5.1 Data Collection

We selected 18 everyday drinks with varying nutritional content (*i.e.*, protein, fat, sugar, etc.) and color. Different products for the same category of drink were chosen to evaluate our prototype on similar liquids. Table 3 shows the properties of the drinks used in our study. We adopted the same experimental protocol, as stated in Subsection 3.2. For the carbonated drinks, we stirred and left the liquid in the sample cuvette for one day. Thus the effect of bubbles on the scanned spectra was suppressed.

For testing the robustness of our prototype, we conducted the experiment in nine ambient conditions: three temperature levels (16 °C, 26 °C and 32 °C) and three light levels (bright ∼ 495 lx, dim ∼ 253 lx and dark ∼ 136 lx), as illustrated in Figure 6. The temperature level was controlled by an air conditioner and measured by an independent digital thermometer close to the NIRS scanner until the targeted temperature was achieved. The light level was controlled by turning on/off of the lamps and measured by a smartphone light sensor. The humidity was stable at around 30%.

In total, 2916 spectra were collected throughout the experiment: 18 drinks × 9 ambient conditions × 18 spectra. We applied the same preprocessing method, as described in Subsection 4.2. Figure 7 shows representative (randomly chosen) spectra for each drink. Two distinct groups of spectra can be observed. The milk spectra appear in the relatively far and close negative region. In fact, milk is considered as emulsions with immiscible liquids [29]. Emulsions are not purely translucent with large particles in the solvent. Thus more light is reflected by the particles in the tube-liquid interface. Since the tube-liquid interface is closer to the scanning window, more light can be detected by the scanner. In consequence, this results in larger detected light intensity values as compared to water (used as the reference liquid in Equation 5), rendering negative absorbance values. The other group of spectra is the translucent liquids. It is also noticeable that the spectra are also similar to the sucrose solutions in Figure 3, as those drinks are mostly composed of sucrose and water.

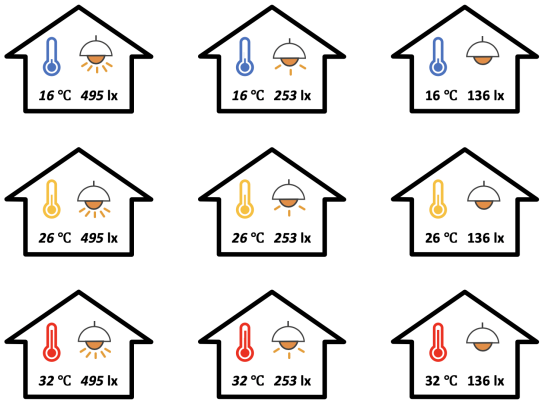


Fig. 6. Nine indoor scenarios for measuring the effect of ambient conditional changes, including three temperature conditions (16 °C, 26 °C and 32 °C) and three light conditions (bright ~ 495 lx, dim ~ 253 lx and dark ~ 136 lx), respectively.

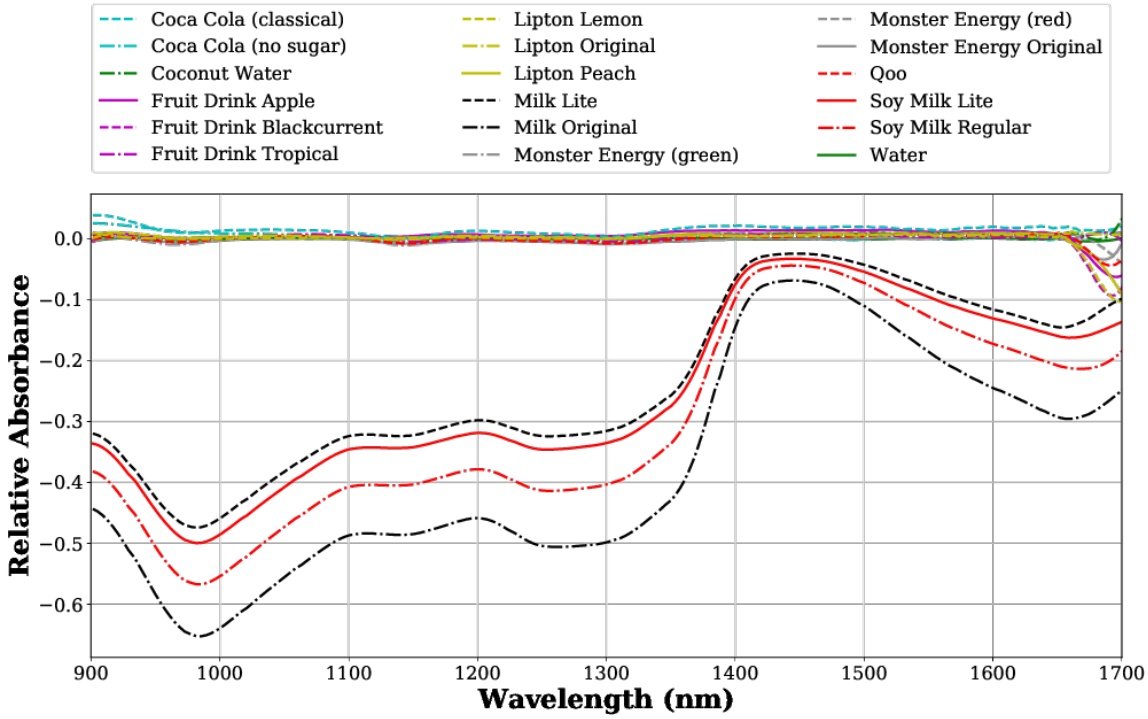


Fig. 7. Representative processed spectra for everyday drinks (colorful print-out recommended).

5.2 Estimating Sucrose Concentration

We first adopted the regression models from Section 4 to predict sucrose concentrations in everyday drinks, trained only with the sampled data from Study 1. The multiple linear regression model achieved the best performance

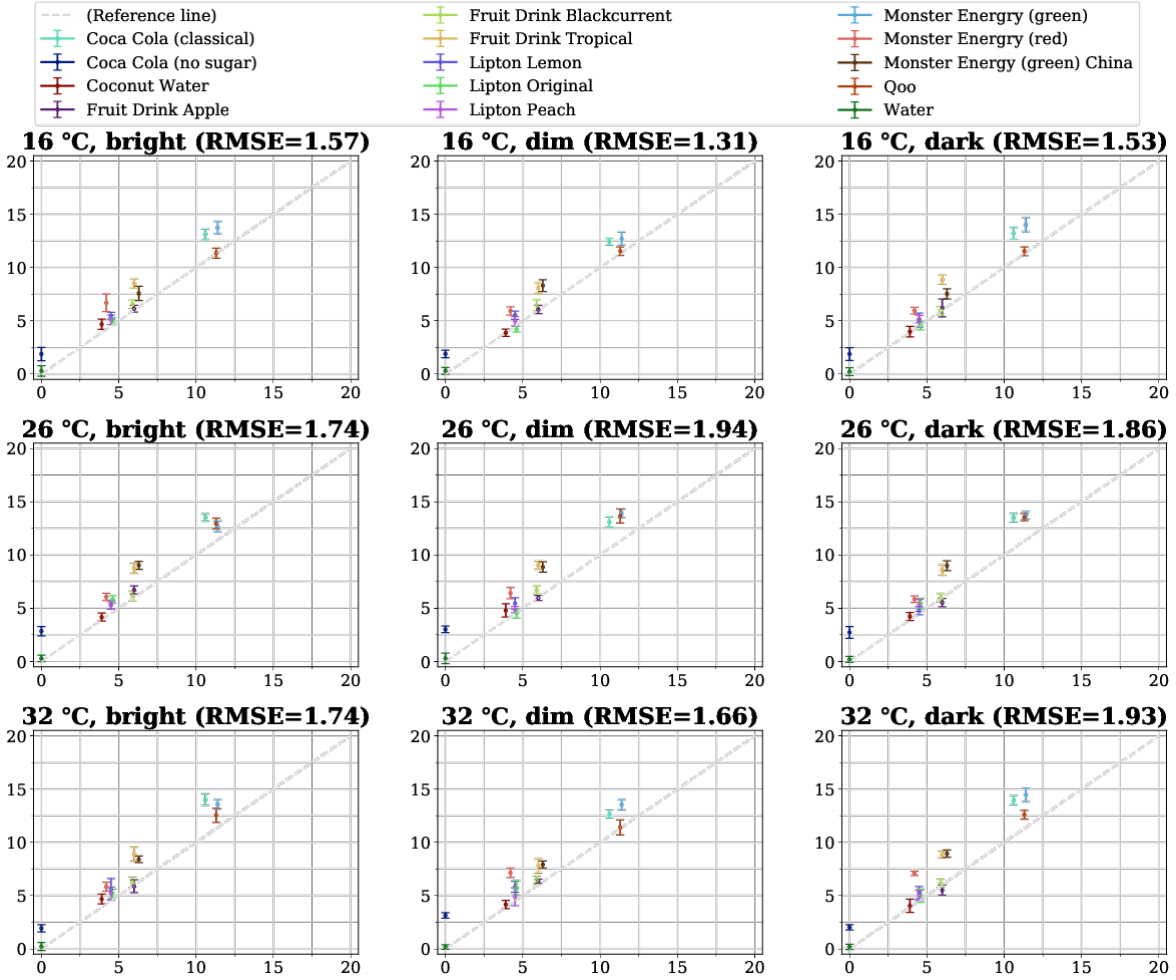


Fig. 8. Using a model from Study 1 to estimate sucrose concentrations in commercial drinks under different environmental conditions. The reference line shows the stated concentration = predicted concentration. (x-axis: ground truth; y-axis: estimation)

(RMSE=1.71), while the simple single linear regression model using 1055 nm wavelength achieved a much worse result (RMSE=8.32). The results for the multiple linear regression across the nine ambient conditions are shown in Figure 8. The RMSE performance varies across the ambient conditions, but the differences are less than 0.63 (RMSE=1.31 ~ 1.94). This variation is mainly caused by the variation of the miniaturized optical engine in the NIRS scanner, which is a research problem that is still being studied [32]. It is worth noting that our regression model was trained in a specific ambient condition (~25 °C room temperature, 30% relative humidity) but worked across all nine ambient conditions without significant performance change. The key is the calibration process by using a water spectrum as the reference for normalization, as we describe in Subsection 4.2.1. In consequence, ambient condition-dependent training as done by [28] is unnecessary for our method.

Overall, we find the regression model overestimates the labeled sucrose concentration of everyday drinks. This is because NIRS is a non-selective sensing method, *i.e.*, it responds to other ingredients in the solution besides the sucrose content. In consequence, the relative absorbance values increase, resulting in higher predicted values. This is a fundamental limitation of the NIRS method [3]. However, as most ingredients other than sucrose and water do not constitute a significant part of those liquids, the performance is still reasonable. Furthermore, we observe that the estimation results can be classified into three groups as follows.

- *Drinks with accurate estimations (RMSE* ≤ 1.0*):* Coconut Water ($mean = 4.29$, $RMSE = 0.50$, $std = 0.32$), Fruit Drink Apple ($mean = 6.04$, $RMSE = 0.36$, $std = 0.36$), Fruit Drink Blackcurrent ($mean = 6.30$, $RMSE = 0.42$, $std = 0.12$), Lipton Lemon ($mean = 5.38$, $RMSE = 0.89$, $std = 0.15$), Lipton Original ($mean = 5.05$, $RMSE = 0.52$, $std = 0.26$) and Lipton Peach ($mean = 5.09$, $RMSE = 0.60$, $std = 0.15$).
- *Drinks with less accurate absolute value estimations but accurate differential estimations:* There are two series of drinks for which the estimations are less accurate ($RMSE > 1.0$) but the differences between two estimations (*i.e.*, differential estimations) are accurate on average. One is the Monster series: Compared to Monster Energy (red) ($mean = 6.35$, $RMSE = 2.24$, $std = 0.65$), the differential estimations of Monster Energy Original ($mean = 8.45$, $RMSE = 2.25$, $std = 0.68$) and Monster Energy (green) ($mean = 13.54$, $RMSE = 2.28$, $std = 0.77$) are 2.10 and 7.19 on average, while the ground truths are 2.10 and 7.20, respectively. The other is the Coca Cola series: Compared to Coca Cola (no sugar) ($mean = 2.39$, $RMSE = 2.47$, $std = 0.65$), the Coca Cola (classical) ($mean = 13.31$, $RMSE = 2.78$, $std = 0.62$) was estimated as $10.92\ g/100ml$ higher than Coca Cola (no sugar), while the ground truth is 1.60.
- *Drinks with less accurate estimations:* The estimations for Lipton Lemon ($mean = 5.39$, $RMSE = 1.09$, $std = 0.63$), Fruit Drink Tropical ($mean = 8.59$, $RMSE = 2.66$, $std = 0.60$) and Qoo ($mean = 12.32$, $RMSE = 1.42$, $std = 0.99$) show neither precise estimation results nor the correct differential estimations. This may be caused by the extra contents and/or the physical properties of the drinks (*e.g.*, colors, carbon dioxide, etc.).

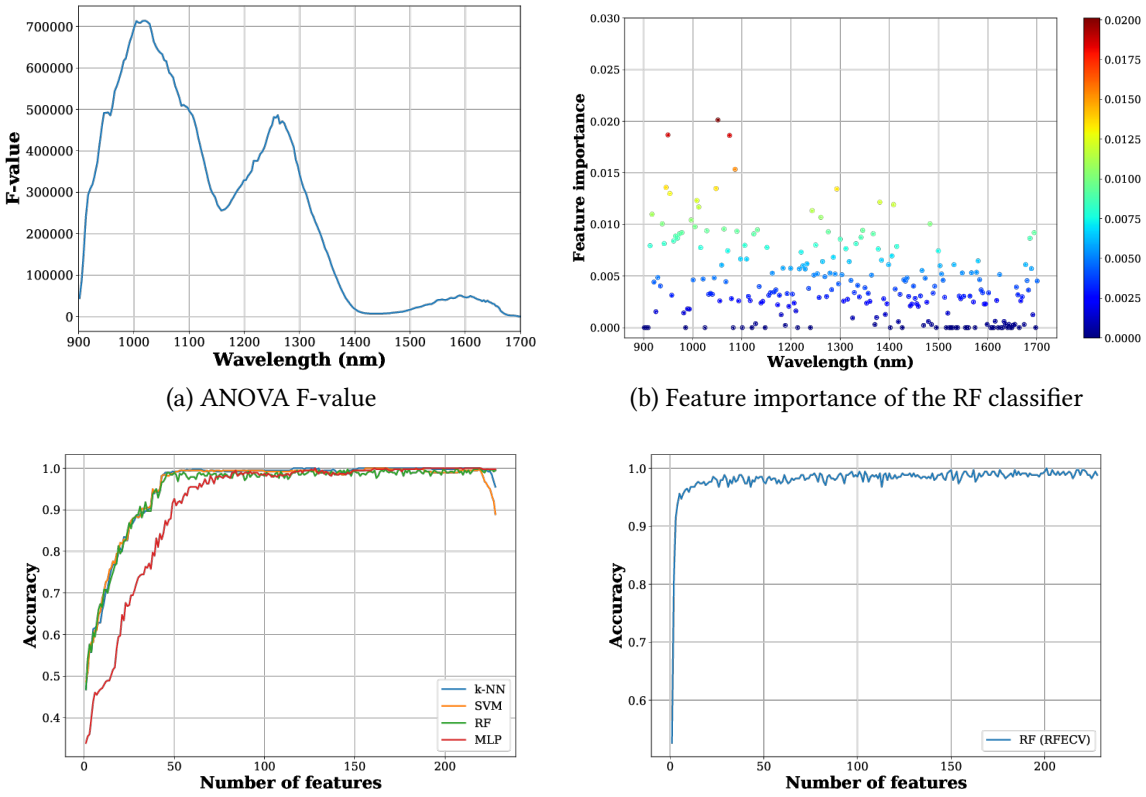
A similar phenomenon was observed in the study by Matsui *et al.* for estimating alcohol content in liquids [28]. To address this, the authors utilized RGB LEDs and sensors as side information to improve accuracy. In our study, alternatively, we show an indirect but more accurate method: by classifying the drinks without requiring additional sensing data.

5.3 Probing Sucrose Content by Classifying Everyday Drinks

We then investigate the performance of discriminating everyday drinks using classification models. If it is possible to identify commercially available drinks, then the system can automatically lookup pre-stored labeled data for the drink to acquire the sucrose concentration level as well as other nutrition statistical information. The method itself is also robust to counterfeit drinks since those drinks do not result in the same NIRS spectra as the real ones, and thus can be easily identified by classification. Since our results show that ambient conditions do not drastically impact the regression model, we aggregate data collected in all nine conditions. The methods we use are: k-nearest neighbors (k-NN), SVM, RF, and MLP. All models are imported from the scikit-learn Python library [37] with default parameters (version=0.20.2).

Furthermore, we perform principal component analysis (PCA) and feature selection for dimensionality reduction. This process can improve the training and estimating the efficiency of the machine learning models to reduce computational resource requirement, as well as help better understand the principle of the NIRS spectra for each material [32].

5.3.1 Feature Evaluation. We first performed a feature evaluation to analyze the importance of each wavelength. In Study 1 (Section 4) we observed that the sucrose content absorbed more NIR light in certain wavelengths than others. Hence, we can also classify those drinks by looking at specific wavelengths in which more NIR light



(a) ANOVA F-value

(b) Feature importance of the RF classifier

(c) Feature evaluation based on F-value

(d) Feature evaluation using RFECV

Fig. 9. Feature evaluations using ANOVA test and RFECV.

is absorbed by corresponding chemical components. We computed the analysis of variance (ANOVA) F-value for each wavelength and considered the feature importance with the Gini criterion attribute in the RF classifier. To evaluate how the features impact the classification accuracy, 10-fold cross-validation was conducted. We also applied recursive feature elimination and cross-validated (RFECV) selection to exclude the least important feature evaluated by cross-validation recursively. Since RFECV is only possible for classifiers that assess feature importance, we only conducted this process for the RF classifier.

The results are shown in Figure 9. We observe that the F-value shows three peak regions approximately 950 ∼ 1100 nm, 1200 ∼ 1300 nm and 1550 ∼ 1650 nm, which correspond to the third, second and first overtone regions respectively in Figure 9. The feature importance evaluation of the RF classifier also shows similar results. The first overtone region that has longer wavelengths is considered much less important than the other two overtone regions, which might be caused by the device we used, as it was more sensitive in this region [17]. We observe that in this region the sucrose solution spectra are very different in Figure 3 compared to the spectra in Figure 7. Thus, using this sensitive region might render unstable classification results.

Furthermore, results of feature evaluation based on F-value show that all four machine learning models achieve > 99% accuracy without all wavelengths involved. In fact, the accuracy of both k-NN and SVM drop with the least important features included (> 1650 nm), since more high-level noise is introduced (*i.e.*, the values in the

Table 4. Comparison of classification models (10-fold cross validation). Numbers in the parentheses are standard deviation values (k-NN = k-nearest neighbors (k=3), SVM = support vector machine, RF = random forest, MLP = multi-layer perceptron).

Metric (%)	**Without preprocess**				**With preprocess**			
	k-NN	*SVM*	*RF*	*MLP*	*k-NN*	*SVM*	*RF*	*MLP*
Accuracy	98.68 (0.0212)	98.68 (0.0270)	91.05 (0.0268)	98.42 (0.0175)	100.00 (0.0000)	100.00 (0.0000)	100.00 (0.0000)	99.74 (0.0079)
Precision	98.60 (0.0270)	99.04 (0.0202)	90.96 (0.0391)	98.51 (0.0229)	100.00 (0.0000)	100.00 (0.0000)	100.00 (0.0000)	99.82 (0.0053)
Recall	98.68 (0.0212)	98.68 (0.0270)	91.05 (0.0268)	98.42 (0.0175)	100.00 (0.0000)	100.00 (0.0000)	100.00 (0.0000)	99.74 (0.0079)
F1-score	98.42 (0.0262)	98.61 (0.0283)	90.00 (0.0317)	98.18 (0.0218)	100.00 (0.0000)	100.00 (0.0000)	100.00 (0.0000)	99.72 (0.0084)

noisy region are higher than normal values). The other two models (RF and MLP) are not affected by this noise. It is worth noting that with fixed hyper-parameters, both k-NN and SVM are more sensitive to the range of input values than RF and MLP. Hence, truncating high-level noisy values is necessary for the k-NN and SVM models.

Finally, using the RF classifier, the cross-validation accuracy results show that $> 95\%$ accuracy is achieved even with four selected features by the RFECV algorithm. With the ANOVA F-value criterion, $> 99\%$ accuracy is achieved by k-NN, SVM and RF classifiers with more than 60 features, while more than 100 features are required for MLP to reach similar accuracy. Given this observation, we selected 167 features by the F-value criterion, since the best accuracy performance can be achieved by all four classification models with the least number of features.

5.3.2 Classification Results. Using the 167 selected features, we tested four machine learning methods for classification using 10-fold cross-validation. As shown in Table 4, k-NN ($k = 3$) achieved the best overall performance with ~100% or 98.68% accuracy, with or without preprocessing, respectively, even though it is a relatively naive approach. This is to be expected since the optical path was well controlled by our design even without complex procedures involved, resulting in a simple classification problem where complex solutions are not necessary. Other more sophisticated methods struggled to reach 100% accuracy without preprocessing, suggesting that an elaborated tuning process might benefit, though not be necessary, for those classifiers.

In general, classifying everyday drinks with our prototype appears to be a simple task with high accuracy. This result indicates that our prototype successfully controls external interference, as we have already shown in Subsection 5.2. It is worth noting that even for drinks with very similar contents (*e.g.*, Lipton Lemon, Lipton Original and Lipton Peach), the model can still successfully classify the drink correctly using the NIRS spectra.

5.3.3 Principle Component Analysis. We ran PCA to further study the NIRS spectra as well as perform a dimensionality reduction. The 10-fold cross-validation was performed with k-NN ($k = 3$) to evaluate the feature extraction performed by PCA (*i.e.*, the training data set was used for searching for the optimal transformation, then the test data set was transformed accordingly, and finally a supervised machine learning algorithm was applied).

Figure 10 shows the result of PCA with k-NN ($k = 3$) classification. We observe that the NIRS spectra for translucent drinks and opaque ones are clearly separated by the first component (PC1, explained variance ratio = 0.9977), which is identical to the relative absorbance spectra in Figure 7. The second components (PC2, explained variance ratio = 0.0021) provides discrimination of NIRS spectra at a more detailed level. We observe that using the k-NN ($k = 3$) model with PCA, 0.986 accuracy is achieved by only the first two principal components, while 1.0 accuracy can be achieved with minimal five principal components. We note that PCA transforms the data

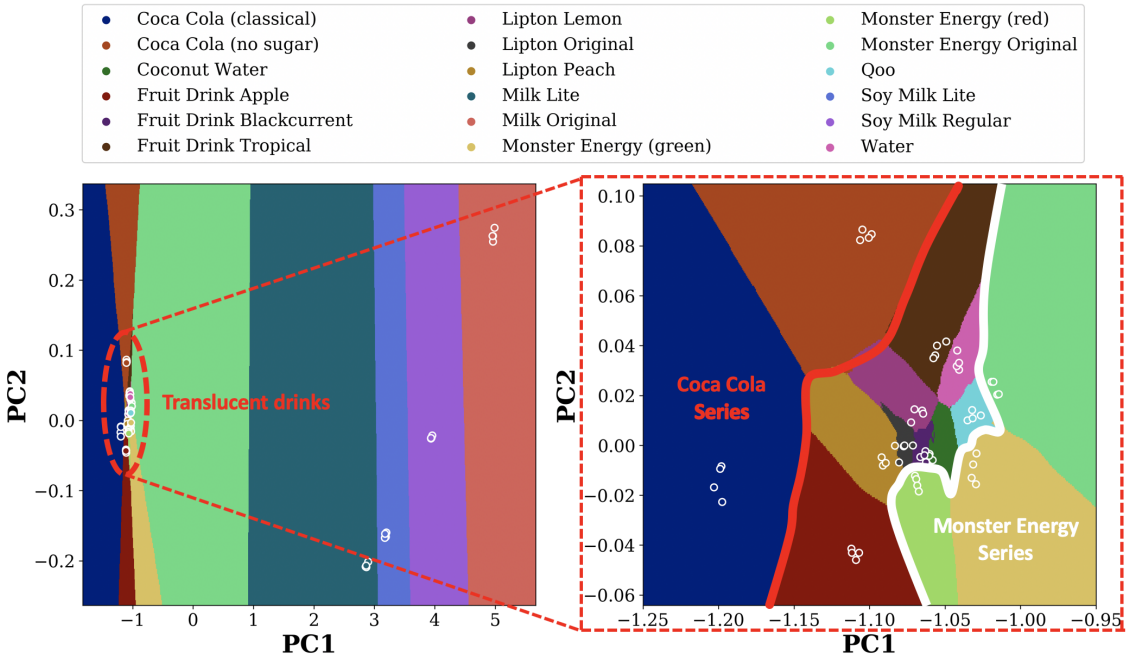


Fig. 10. PCA with k-NN classifier ($k = 3$), white-outlined points are test points. Ratios of variance explained are $PC1 = 0.9977$ and $PC2 = 0.0021$ respectively. Left: PC1 + PC2 for all drinks. Right: PC1 + PC2 for translucent drinks (colorful print-out recommended).

while feature selection does not. Therefore, we cannot compare the efficiency of the algorithms by the number of features directly.

Furthermore, we find that two brands of drinks, Monster and Coca Cola, have less estimation accuracy but precise differential estimations as we show in Subsection 5.2, and they are in the marginal PCA regions and grouped respectively as illustrated in Figure 10. This result means that for those drinks, additional content might be added, but not listed in detail on the labels, to adjust the color (*e.g.*, dark brown for Coca Cola), flavor, functions (*e.g.*, refreshing function for Monster), etc.

6 DISCUSSION

Our experiments yield promising results in terms of utilizing a miniaturized NIRS scanner for sucrose contented liquid sensing tasks, including both estimating concentrations and classifying everyday drinks. We present a prototype capable of estimating concentrations of sucrose solutions with ±0.29 g/100ml error with a complex MLP model, or ±2.67 g/100ml with a simple single linear regression model with one selected wavelength (1055 nm). This result is rather promising given the preparation of solutions in a less controlled setting rendering inevitable labeling errors in the ground truth references.

Furthermore, when estimating sucrose content in everyday drinks, we find that applying the regression models pre-trained by the sucrose solutions directly might not yield as precise results. Similar findings were also reported by Matsui *et al.* , and they suggest that color could have negative effects on NIR light measurements, given that they utilized RGB LED and digital color sensors as a compensation [28]. We did not adopt this method, since we can adopt an alternative classification method without modifying the experimental settings. In addition, the

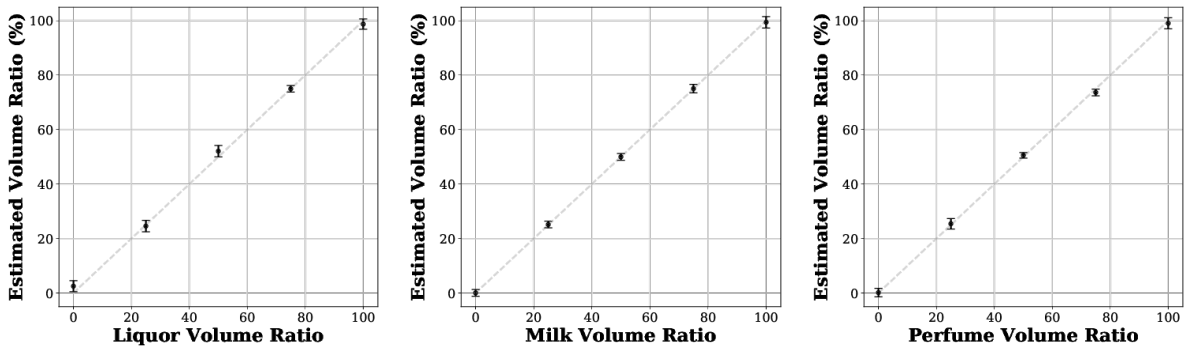


Fig. 11. Detecting diluted liquids. Classification models show 100% accuracy for all three contents. Left: diluted liquor (RMSE=2.09, R^2=0.99). Middle: diluted milk (RMSE=0.50, R^2=0.99). Right: diluted perfume (RMSE=1.19, R^2=0.99).

generalization of the RGB compensation method can be constrained because a specific design is required for different content sensing. Our method does not require bespoke hardware design and instead uses an off-the-shelf NIRS scanner, which can be more convenient to update for additional tasks, including sensing alcohol or classifying more drinks.

In terms of classification, our method also shows promising results in identifying 18 everyday drinks, with the most basic k-NN method outperforming other machine learning methods. This suggests that discriminating spectra were successfully acquired by our device, resulting in a non-complex classification problem. PCA analysis and feature evaluations also confirm this finding. By further analyzing the PCA results, we find that the first and second principal components for those over-estimated drinks are in the marginal regions, indicating the possible existence of other content with significant effect to the NIRS spectra. In fact, most organic molecules have their NIRS signatures, due to the absorbance of C-H chemical bonds, in the wavelength range of the NIRS scanner we used (900 nm~1700 nm). Therefore, it is possible to detect whether extra contents are added to the drink given a pre-scanned reference spectrum for that drink.

Overall, our findings show that our design enables both sucrose concentration estimation and classification of liquids. This is a task previously conducted in laboratories by trained experts. Having this capability in a small handheld device can be valuable for monitoring sucrose intake for decreasing risks of diabetes and weight-watching [49], as well as for users with chronic conditions that require regular monitoring (*e.g.*, sucrose intolerance) [34].

6.1 Towards food scanners: A standardized protocol

A motivation for our work is to develop a standardized calibration, sampling, and analysis protocol for conducting NIRS studies with miniaturized hardware. To demonstrate this, we have re-used the calibration process and MLR model in Subsection 4.4 to estimate concentrations of two further types of sugars (beyond sucrose): glucose and fructose, both of which are found in everyday food. We prepared three concentration values (5 g/100ml, 10 g/100ml and 15 g/100ml) of three categories (glucose, fructose, glucose:fructose 50:50 mixture) respectively. We tested the performance of the MLR model using the sampling protocol in Subsection 3.2. The results achieved overall RMSE=1.55, with RMSE $\leq$ 2.29 g/100ml in all nine ambient conditions, as show in Table 5. We note that no additional model training was carried out for this test: we reused the MLR model from Study 1.

Furthermore, the application of our work can be extended beyond sugar sensing. This is possible because our device uses a wide-ranging spectrometer. Therefore, unlike bespoke systems that respond to pre-determined

Table 5. RMSE for glucose and fructose content estimation using multi-linear regression model, unit: $g/100ml$).

Sugar Solution	16°C			26°C			32°C		
	Bright	Dim	Dark	Bright	Dim	Dark	Bright	Dim	Dark
Glucose	1.34	1.03	2.07	0.89	1.10	1.23	1.44	1.27	1.77
Fructose	1.46	0.92	2.29	1.27	1.80	1.12	1.66	2.12	1.54
Glucose:Fructose	1.08	1.42	1.62	1.70	1.33	1.47	1.90	2.00	1.73

frequencies, our prototype can be reprogrammed to detect other substances (*e.g.*, fat, proteins, etc.), simply by updating the classifier software on the device. To demonstrate this, we used our calibration and sampling protocol to collect training data to create new regression models and classification models that detect diluted alcohol (liquor), milk, and perfume. The classification models show $\sim$ 100% accuracy for liquor, milk, and perfume dilution. The regression models, as shown in Figure 11, perform quite similar to those presented in literature with other mobile (albeit bespoke) methods for both liquor (RMSE=1.45, R_2=0.99) [38] and milk (RMSE=0.50, R_2=0.99) [28]. We observe that our system performs best for milk due to its higher SNR since milk is more reflective, as we mentioned for Figure 7. Other possible extensions of our work include, but are not limited, to allergen detection in liquids, liquid authentication to avoid taking counterfeit liquids (*e.g.*, diluted perfumes as shown in Figure 11), assisting customizing drinks (*e.g.*, sweetness adjustment), etc.

Broadly speaking, our work demonstrates an important methodological step towards the development of a handheld food scanner that can be used in a range of everyday scenarios. We describe a calibration protocol and a sampling protocol that can be used to train models or classifiers that can be robust to environmental variations. Having a handheld food scanner means that users can have immediate feedback regarding the properties of the food or beverage they are about to consume. This can be valuable in everyday settings (where one is following a strict diet), in travel settings (where one may not be able to read the nutrition labels of a product), or in cases where the product does not have a nutrition label (such as vegetables at a food market or pastries at a local bakery).

Furthermore, since our prototype only uses off-the-shelf components, and the 3D model and software are all open-sourced[6], we encourage others to replicate and extend our work. Our work can be easily adapted to perform other tasks beyond liquid sensing, including food class estimation (*e.g.*, ranking beef), food poison detection (*e.g.*, detecting poisonous potatoes), food maturity prediction (*e.g.*, estimating the maturity of avocado as presented in [10]), etc. The main challenges are, in particular, to design a proper optical path for the NIRS scanner as we clarified in Section 3, and to choose the right models to process the spectra as we have shown in Section 4 and 5.

6.2 Hardware Miniaturization

We note that there exists potential to further miniaturize our prototype. There are two main components that could be further miniaturized: one is the Raspberry Pi A+ ($H \times W \times D = 12mm \times 65mm \times 56mm$) that we use for onboard computational resource. Alternatively, the Raspberry Pi Zero model is much smaller ($H \times W \times D = 12mm \times 65mm \times 30mm$), resulting in 46% size reduction. Furthermore, it is also feasible to design a customized computational board with a BCM283x chip that the Raspberry Pi uses or other System-on-Chip (SoC) parts.

The other component that takes space is the NIRS scanner itself ($H \times W \times D = 36mm \times 62mm \times 58mm$). It should be noted that the scanner we currently use is still an Evaluation Module (EVM), which will be revised in the future. Engineering wise, the main challenge to further reduce its size is to shrink the optical engine for light path control and detection. Currently, the optical engine is designed as a balance between cost and performance (SNR and resolution in particular), while the main cost of the NIRS scanner is by the Digital Micromirror Device (DMD)

[6]https://github.com/HighTemplar-wjiang/IMWUT-NIRS-Liquid

chip (~ 300 US dollars per chip) and the InGaAs optical sensor (~ 200 US dollars per chip). Using alternative designs (*e.g.*, with an optical sensor array) may greatly shrink the size with much higher cost. We can expect, however, the prices will go done drastically as the chip technology advances in the future.

6.3 Limitations

We note several limitations in our studies: most importantly, it is the physical principle for the NIRS method. NIRS is a non-selective method, which can be considered a double-edged sword. On the one hand, it provides rich information for the chemical composition as well as physical properties of the sample. On the other hand, the information that we are interested in may be overshadowed by noise. In this work, we utilized pure water as the reference sample to alleviate this issue. However, when significant motion occurs (*e.g.*, re-clamp the tube), or significant ambient conditional change happens (*e.g.*, the temperature changed from 16 °C to 26 °C), during sampling, one should re-sample the reference liquid for subsequent liquid sampling tasks. Alternatively, it is also possible to store a series of pre-scanned reference signals, and select one has been scanned in the closest condition, with the help of corresponding sensors.

We have tested the effects of ambient conditional changes during the same task, including three temperatures and three light conditions, respectively, which did not cover all possible combinations. We did not test the effect with humidity changes as humidity was not feasible to control. However, we have shown that our calibration process could well address this issue under the nine conditions we tested. Also, the datasheet of the scanner we used suggests that calibration by replacing the reference signal can help keep the performance when humidity changes [17], indicating the usefulness of our calibration process under humidity changes. Besides, as the main factor that affects a NIRS spectrum, under the same ambient condition, is the ingredient itself, it is reasonable to assume that our calibration process should also work in other ambient conditions. It should be noted, however, the device itself can only work under a specific environmental range that should cover most daily scenarios but may not include extreme environmental conditions.

Finally, in our study, we tested 25 sucrose solutions and 18 everyday drinks, respectively. For a specialized task, more samples should be taken to build the corresponding NIRS libraries, such as more fine-grained concentrations or more categories of everyday drinks. In addition, in the case of product quality monitoring, samples from different material lots should be used as suggested by conventional NIRS studies [32]. Furthermore, the bubble problems should be addressed for samples of carbonated liquids, for which we stirred the carbonated drinks for suppressing this effect.

7 CONCLUSION AND FUTURE WORK

We present a prototype utilizing a miniaturized commercial, off-the-shelf NIRS scanner for sucrose contented liquid sensing tasks, along with a replicable sampling and analysis protocol. We conducted two studies to evaluate the performance of our prototype systematically. Our experiments show ±0.29 g/100ml accuracy in concentration estimation in 25 sucrose solutions and > 99% accuracy in identifying 18 everyday drinks. These results suggest that our prototype can be used with reasonable performance on everyday drinks. We further provide examples where our design is used in other applications, including alcohol, milk, and perfume. We argue that our work is the first step towards a handheld food scanner that can be used to perform even more food-related tasks including food quality estimation, food poison/allergen detection, food maturity prediction, etc. We also discussed in detail, in future work, how the prototype could be miniaturized by customizing the hardware, allowing it to be appropriate in scenarios which require smaller-sized hardware. In conclusion, our method shows it is viable and promising to estimate sucrose concentrations in everyday drinks using a commercially off-the-shelf miniaturized near-infrared spectroscopy scanner.

ACKNOWLEDGMENTS

This work is partially funded by a Samsung Global Research Outreach grant, the ARC Discovery Project DP190102627 and JST ERATO Grant Number JPMJER1501, Japan.

REFERENCES

[1] Burkhard Beckhoff, Birgit Kanngießer, Norbert Langhoff, Reiner Wedell, and Helmut Wolff. 2007. *Handbook of practical X-ray fluorescence analysis*. Springer Science & Business Media.

[2] Robert J Blakey. 2016. Evaluation of avocado fruit maturity with a portable near-infrared spectrometer. *Postharvest Biology and Technology* 121 (2016), 101–105.

[3] Donald A Burns and Emil W Ciurczak. 2007. *Handbook of near-infrared analysis*. CRC press.

[4] Quansheng Chen, Jiewen Zhao, CH Fang, and Dongmei Wang. 2007. Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 66, 3 (2007), 568–574.

[5] Ahmet F Coskun, Justin Wong, Delaram Khodadadi, Richie Nagi, Andrew Tey, and Aydogan Ozcan. 2013. A personalized food allergen testing platform on a cellphone. *Lab on a chip* 13, 4 (02 2013), 636–640. https://doi.org/10.1039/c2lc41152k

[6] Anshuman J. Das, Akshat Wahi, Ishan Kothari, and Ramesh Raskar. 2016. Ultra-portable, wireless smartphone spectrometer for rapid, non-destructive testing of fruit ripeness. *Scientific Reports* 6 (08 09 2016), 32504 EP –. https://doi.org/10.1038/srep32504

[7] Claudia A Teixeira Dos Santos, Miguel Lopo, Ricardo NMJ Páscoa, and João A Lopes. 2013. A review on the applications of portable near-infrared spectrometers in the agro-food industry. *Applied spectroscopy* 67, 11 (2013), 1215–1233.

[8] Mingming Fan, Khai N Truong, and Abhishek Ranjan. 2016. *Exploring the Use of Capacitive Sensing to Externally Measure Liquid in Fluid Containers*. Technical Report.

[9] Francis Galton. 1886. Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland* 15 (1886), 246–263.

[10] Mayank Goel, Eric Whitmire, Alex Mariakakis, T. Scott Saponas, Neel Joshi, Dan Morris, Brian Guenter, Marcel Gavriliu, Gaetano Borriello, and Shwetak N. Patel. 2015. HyperCam: Hyperspectral Imaging for Ubiquitous Computing Applications. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 145–156. https://doi.org/10.1145/2750858.2804282

[11] Inc. Hanna Instruments. n.d.. Digital Refractometer for Sugar Analysis in Wine, Must and Juice - HI96812 - Hanna Instruments. https://goo.gl/y4jbrY. (Accessed on 07/12/2018).

[12] Rex Harrill. 1998. Using a refractometer to test the quality of fruits and vegetables. *PUBLISHING, Éd* (1998).

[13] Arshad Hassan, KiBae Lee, Jinho Bae, and Chong Hyun Lee. 2017. An inkjet-printed microstrip patch sensor for liquid identification. *Sensors and Actuators A: Physical* 268 (2017), 141–147.

[14] Simon Haykin and Barry Van Veen. 2007. *Signals and systems*. John Wiley & Sons.

[15] Haibo Huang, Haiyan Yu, Huirong Xu, and Yibin Ying. 2008. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *Journal of food engineering* 87, 3 (2008), 303–313.

[16] Ginés Ibáñez, Jaime Cebolla-Cornejo, Raúl Martí, Salvador Roselló, and Mercedes Valcárcel. 2019. Non-destructive determination of taste-related compounds in tomato using NIR spectra. *Journal of Food Engineering* 263 (2019), 237 – 242. https://doi.org/10.1016/j.jfoodeng.2019.07.004

[17] Texas Instruments Incorporated. n.d.. DLP NIRscan Nano Evaluation Module. http://www.ti.com/tool/DLPNIRNANOEVM. (Accessed on 07/11/2018).

[18] Xin Jiang, Tingting Yang, Changli Li, Rujing Zhang, Li Zhang, Xuanliang Zhao, and Hongwei Zhu. 2017. Rapid Liquid Recognition and Quality Inspection with Graphene Test Papers. *Global Challenges* 1, 6 (2017), 1700037.

[19] Eric Jones, Travis Oliphant, Pearu Peterson, et al. 2001–. SciPy: Open source scientific tools for Python. http://www.scipy.org/ [Online; accessed on 07/11/2018].

[20] Youngeui Jung and Jungseek Hwang. 2013. Near-infrared studies of glucose and sucrose in aqueous solutions: water displacement effect and red shift in water absorption from water-solute interaction. *Applied spectroscopy* 67, 2 (2013), 171–180.

[21] Simon Klakegg, Jorge Goncalves, Chu Luo, Aku Visuri, Alexey Popov, Niels van Berkel, Zhanna Sarsenbayeva, Vassilis Kostakos, Simo Hosio, Scott Savage, et al. 2018. Assisted Medication Management in Elderly Care Using Miniaturised Near-Infrared Spectroscopy. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 69. https://doi.org/10.1145/3214272

[22] Simon Klakegg, Jorge Goncalves, Niels van Berkel, Chu Luo, Simo Hosio, and Vassilis Kostakos. 2017. Towards Commoditised Near Infrared Spectroscopy. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 515–527. https://doi.org/10.1145/3064663.3064738

[23] Simon Klakegg, Chu Luo, Jorge Goncalves, Simo Hosio, and Vassilis Kostakos. 2016. Instrumenting Smartphones with Portable NIRS. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA, 618–623. https://doi.org/10.1145/2968219.2971590
[24] J. Lester, D. Tan, S. Patel, and A. J. B. Brush. 2010. Automatic classification of daily fluid intake. In *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*. 1–8. https://doi.org/10.4108/ICST.PERVASIVEHEALTH2010.8906
[25] Pei-Shih Liang, Tu San Park, and Jeong-Yeol Yoon. 2014. Rapid and reagentless detection of microbial contamination within meat utilizing a smartphone-based biosensor. *Scientific Reports* 4 (05 08 2014), 5953 EP –. https://doi.org/10.1038/srep05953
[26] Ai Liu, Gang Li, Zhigang Fu, Yang Guan, and Ling Lin. 2018. Non-linearity correction in NIR absorption spectra by grouping modeling according to the content of analyte. *Scientific reports* 8, 1 (2018), 8564.
[27] Prinya Masawat, Antony Harfield, and Anan Namwong. 2015. An iPhone-based digital image colorimeter for detecting tetracycline in milk. *Food Chemistry* 184 (2015), 23–29. https://doi.org/10.1016/j.foodchem.2015.03.089
[28] Hidenori Matsui, Takahiro Hashizume, and Koji Yatani. 2018. Al-light: An Alcohol-Sensing Smart Ice Cube. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 126 (Sept. 2018), 20 pages. https://doi.org/10.1145/3264936
[29] Gerard Meurant. 1995. *Handbook of milk composition*. Elsevier.
[30] Simon Mezgec and Barbara Koroušić Seljak. 2017. NutriNet: A Deep Learning Food and Drink Image Recognition System for Dietary Assessment. *Nutrients* 9, 7 (2017). https://doi.org/10.3390/nu9070657
[31] Alan S. Morris, Reza Langari, and Alan S. Morris. 2016. *Chapter 21 - Summary of Other Measurements*. Academic Press, Boston, 633–672. https://doi.org/10.1016/B978-0-12-800884-3.00021-6
[32] Metrohm NIRSystems. 2002. A guide to near-infrared spectroscopic analysis of industrial manufacturing processes.
[33] Noboru Ohta and Alan Robertson. 2006. *Colorimetry: fundamentals and applications*. John Wiley & Sons.
[34] World Health Organization. 2015. *Guideline: sugars intake for adults and children*. World Health Organization.
[35] Brian G Osborne. 2006. Near-infrared spectroscopy in food analysis. *Encyclopedia of analytical chemistry: applications, theory and instrumentation* (2006).
[36] John Parrish. 2012. *UV-A: Biological effects of ultraviolet radiation with emphasis on human responses to longwave ultraviolet*. Springer Science & Business Media.
[37] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research* 12, Oct (2011), 2825–2830.
[38] Tauhidur Rahman, Alexander T. Adams, Perry Schein, Aadhar Jain, David Erickson, and Tanzeem Choudhury. 2016. Nutrilyzer: A Mobile System for Characterizing Liquid Food with Photoacoustic Effect. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM (SenSys '16)*. ACM, New York, NY, USA, 123–136. https://doi.org/10.1145/2994551.2994572
[39] Giovanni Rateni, Paolo Dario, and Filippo Cavallo. 2017. Smartphone-Based Food Diagnostic Technologies: A Review. *Sensors* 17, 6 (2017). https://doi.org/10.3390/s17061453
[40] Yves Roggo, Pascal Chalus, Lene Maurer, Carmen Lema-Martinez, Aurélie Edmond, and Nadine Jent. 2007. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of pharmaceutical and biomedical analysis* 44, 3 (2007), 683–700.
[41] Allan Rosencwaig and Allen Gersho. 1976. Theory of the photoacoustic effect with solids. *Journal of Applied Physics* 47, 1 (2019/08/04 1976), 64–69. https://doi.org/10.1063/1.322296
[42] Naresh C. Saha and Harland G. Tompkins. 1992. Titanium nitride oxidation chemistry: An x-ray photoelectron spectroscopy study. *Journal of Applied Physics* 72, 7 (2019/08/13 1992), 3072–3079. https://doi.org/10.1063/1.351465
[43] Abraham Savitzky and Marcel JE Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* 36, 8 (1964), 1627–1639.
[44] Maria Lúcia F Simeone, Rafael AC Parrella, Robert E Schaffert, Cynthia MB Damasceno, Michelle CB Leal, and Celio Pasquini. 2017. Near infrared spectroscopy determination of sucrose, glucose and fructose in sweet sorghum juice. *Microchemical Journal* 134 (2017), 125–130.
[45] Douglas A Skoog, F James Holler, and Stanley R Crouch. 2017. *Principles of instrumental analysis*. Cengage learning.
[46] DF Swinehart. 1962. The beer-lambert law. *Journal of chemical education* 39, 7 (1962), 333.
[47] Eizo Taira, Masami Ueno, Kwantri Saengprachatanarug, and Yoshinobu Kawamitsu. 2013. Direct sugar content analysis for whole stalk sugarcane using a portable near infrared instrument. *Journal of Near Infrared Spectroscopy* 21, 4 (2013), 281–287.
[48] Arno Villringer, J Planck, C Hock, L Schleinkofer, and U Dirnagl. 1993. Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. *Neuroscience letters* 154, 1-2 (1993), 101–104.
[49] World Health Organization and others. 2003. Diet, Nutrition and the Prevention of Chronic Diseases: Report of a Joint WHO/FAO Expert Consultation. (2003).