2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

# Prediction of Aqueous Glucose Concentration Using Hyperspectral Imaging

Chiao-Yi Wang*[1], Anjana Hevaganinge*[1], Dongyi Wang*[1,3], Mohamed Ali[1], Maurizio Cattaneo[2] and Yang Tao[1]

***Abstract*— Near infrared hyperspectral imaging (HSI) is an emerging optical imaging modality which boasts several advantages. Compared to conventional spectroscopy, HSI provides thousands of spectral samples with embedded spatial information in a single image. This allows the collection of high quality and high volume spectral signals in a short time. In this paper, transmissive HSI combined with Partial Least Squares Regression (PLSR) was used to non-invasively predict aqueous glucose concentration. Aqueous glucose samples are prepared with concentration ranging from 0 - 1000 mg/dL at intervals of 100 mg/dL and 100 - 300 mg/dL at intervals of 20 mg/dL. Our results are validated using leave-one-concentration-out cross validation, and demonstrate the feasibility of the proposed aqueous glucose concentration detection method using the combination of HSI and PLSR.**

## I. INTRODUCTION

Glucose concentration is a crucial parameter in biological, physiological and chemical domain. Monitoring and controlling glucose concentration is important in many applications, such as clinical blood glucose measurement [1] and bioreactor cultivation [2]. Continuously monitoring blood glucose is also meaningful for diabetes patients, which can help to determine the insulin uptake amount [3]. Moreover, glucose is a general supplement in cell culture, which further shows the importance to develop innovated tools to control and monitor the glucose concentration [2].

Several existing methods for measuring glucose concentration in bio-processing media include glucose biosensor [4], near-infrared spectroscopy (NIRS) [5]–[7] and Raman spectroscopy [8]. NIRS is a promising technique because it is a non-contact sensor with little sample preparation requirements. The principle of NIRS is based on the light absorption in the near-infrared region (800 to 2500 nm) by the observed samples [9]. These absorption signal intensities are related to the vibration of common functional groups, such as O-H, C-H and C=O. However, traditional NIRS only provides a single point of spectral information of the sample, which can be easily affected by environmental noise. Thus, corresponding chemometric glucose determination models usually suffer from poor repeatability.

*These authors contributed equally
[1]Fischell Department of Bioengineering University of Maryland College Park, College Park, MD 20742, USA {cyiwang, ahenagan, dywang, mali93, ytao}@umd.edu
[2]Founder & CEO of Artemis Biosystems, Cambridge, MA 02139 mcattaneo@artemisbiosystems.com
[3]Biological and Agricultural Engineering, University of Arkansas, Fayetteville, Arkansas, 72701

Recently, hyperspectral imaging (HSI) has gained favor in many fields and applications, such as disease diagnosis and surgical guidance instrumentation [10]. HSI has the capability to provide both spectral and spatial information of samples, combining the advantages of both spectroscopy and regular color imaging. This can potentially minimize the environmental interference in the NIRS method. In food engineering, studies demonstrate the feasibility of non-invasively measuring the nutrient contents of liquid samples using HSI [11]. However, few transmissive HSI studies were applied to quantitative chemical component estimation in aqueous samples due to strong water absorption [12]. This research is intended to explore the feasibility of glucose concentration estimation in aqueous solution using NIR HSI.

In this paper, the glucose concentration in aqueous solution with ranges from 0 to 1000 mg/dL was studied. Partial Least Square Regression (PLSR) was applied to extract glucose concentration from HSI. The results show the feasibility of using HSI to detect the aqueous glucose concentration. The HSI system structure and quantitative analysis are discussed based on several additional experiments, including a repeatability test and different resolution tests.

## II. METHOD

### A. Sample Preparation

Each sample was prepared by weighing (Mettler Toledo MS802S) an appropriate amount of D-glucose (Sigma-Aldrich, G8270, $\geq$ 99.5%) dissolved in 50 mL of deionized water. Each sample was shaken inside a centrifuge tube for 10 seconds to ensure that the glucose was fully dissolved inside the sample. Images were collected by two technicians over 5 days within 3 months to increase the randomness of the data, and to ensure the model generalization capabilities. Table I summarizes the range and minimum increment of glucose samples collected over different days.

TABLE I
RANGE AND MINIMUM INCREMENT OF GLUCOSE SAMPLES

| Date | Concentration Range (mg/dL) | Min Increment (mg/dL) |
|---|---|---|
| 10/20/2020 | 100 - 1000 | 100 |
| 11/08/2020 | 500 | N/A |
| 12/12/2020 | 100 - 1000 | 100 |
| 12/22/2020 | 100 - 300 | 20 |

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

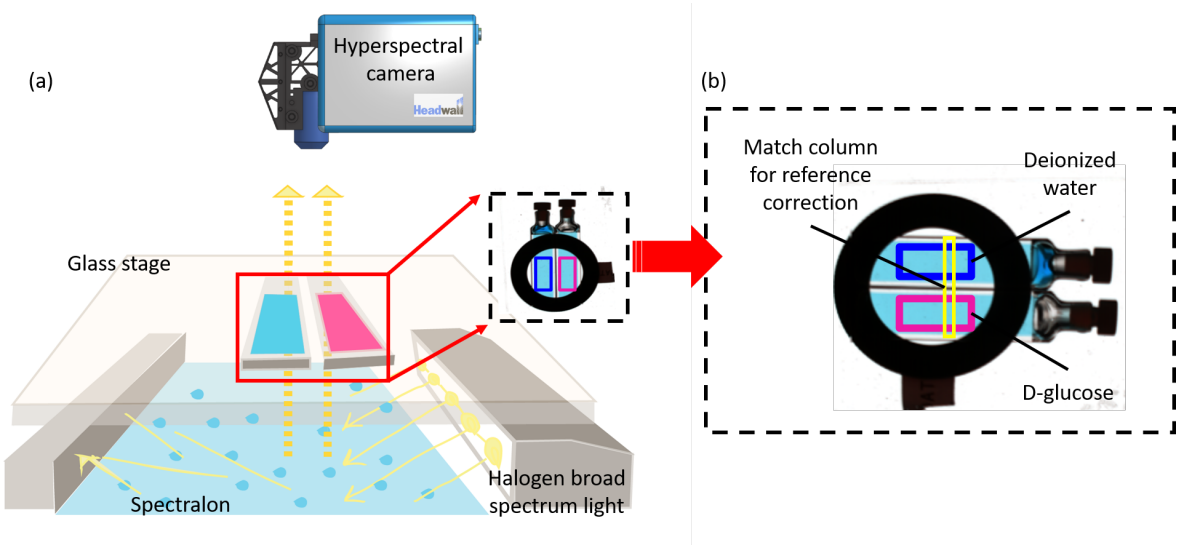


Fig. 1. (a) HSI system (b) The sample image. Blue area represents the reference deionized water. Pink area represents the D-Glucose sample. Yellow area represents the column matching of deionized water and D-Glucose sample. Reference correction was done by the subtraction of matching column.

### B. Experimental Setup and Data Acquisition

The NIR transmission images (900-2500 nm) of each samples were measured using the Headwall Micro - Hyperspec SWIR line scan camera. The light source is Halogen broad spectrum light. The hyperspectral image cubes were generated based on the push-broom principle. The glucose samples were loaded into a micro cuvette with 1mm light path (Hellma Macro cells No. 110-1-40), which is manufactured by UV/VIS (200-2500 nm) transparent high performance quartz glass in the red bounding box of Fig. 1(a). The measurements were taken at ambient room temperature. Before each days' experiment, the white and dark calibrations of the camera were conducted. The dark calibration was implemented with the camera shutter closed. The white calibration image was acquired by letting the light shotting through the station glass directly. For each glucose sample, an identical micro cuvette filled with DI water was put aside as the reference sample.

### C. Imaging Processing and Data Analysis

Each hyperspectral image cube contained at least 2000 sample pixels. The average of pixels in each column was computed, named as column average signals. This procedure yielded approximately 80 spectra signals per image. As described in Section B, each sample image contained the glucose sample as well as pure deionized water sample. The column average of glucose signals was subtracted by the average of water signals from the identical column as reference correction. In Fig. 1(b), the yellow bounding box represent an example column used for the described water correction. To establish the relationship between the spectral signal and the glucose concentration, a mathematical model Partial Least Square Regression (PLSR) model [13] was trained based on the ground truth of glucose concentration and validated by leave-one-concentration-out (LOCO) cross validation. The basic concept of LOCO is that one concentration group was chosen as testing data for each fold cross validation, and the left concentration groups were used for the model training.

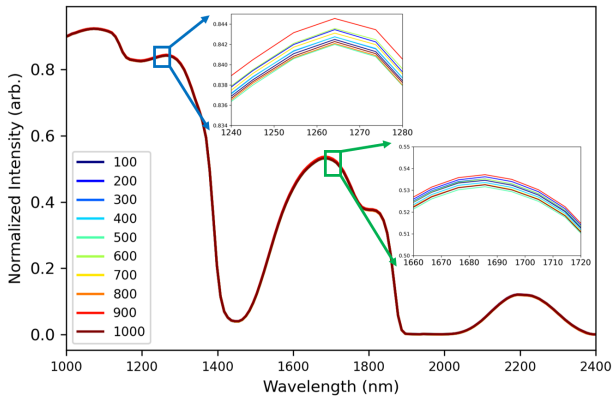


Fig. 2. Original spectra of 10 samples of glucose aqueous concentrations. The glucose concentration are in the range from 100 to 1000 mg/dL, with 100 mg/dL interval. The NIR spectra range are between 1000 to 2400 nm.

## III. RESULTS

### A. Leave-One-Concentration-Out Cross-Validation (LOCO) Results

In the NIR range (900 – 2500 nm), water dominates the absorption of the spectrum comparing to glucose absorption. Fig. 2 shows the original spectra of glucose samples. The original spectra of glucose with different concentration are very similar. To extract the glucose concentration from the spectra, water reference correction was done in imaging processing, as mentioned in Section 2.2. Fig. 3 shows the spectra (glucose concentration = 100 to 1000 mg/dL) after water reference correction. As seen in Fig. 3, most of the NIR spectra look different and there is no obvious trend corresponding to glucose concentration. It is difficult to predict the glucose concentration directly through the intensity of

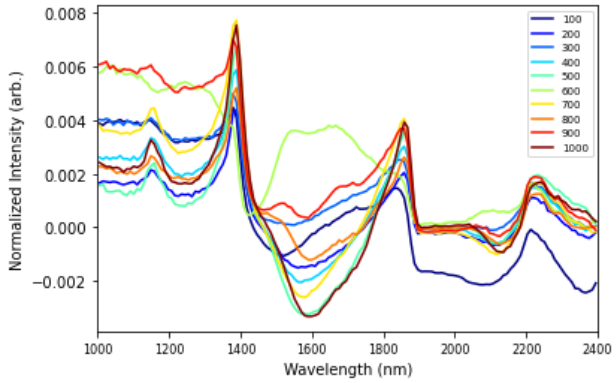


Fig. 3. NIR spectra after reference correction

TABLE II
LOCO CROSS-VALIDATION RESULTS OF THE PLSR MODEL. GLUCOSE CONCENTRATIONS ARE BETWEEN 100 AND 1000 MG/DL, WITH 100 MG/DL INTERVAL.

| Ground truth (mg/dL) | Prediction (mg/dL) | STD (mg/dL) |
|---|---|---|
| 100 | 106.73 | 53.81 |
| 200 | 257.21 | 77.91 |
| 300 | 304.58 | 64.52 |
| 400 | 403.31 | 49.56 |
| 500 | 516.10 | 58.55 |
| 600 | 545.29 | 51.96 |
| 700 | 657.90 | 60.93 |
| 800 | 804.07 | 56.63 |
| 900 | 902.03 | 50.08 |
| 1000 | 980.91 | 54.32 |

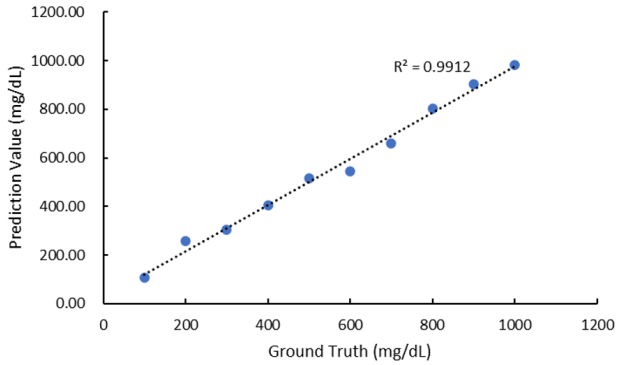


Fig. 4. The prediction results of LOCO cross-validation. Glucose concentrations are between 100 and 1000 mg/dL, with 100 mg/dL interval.

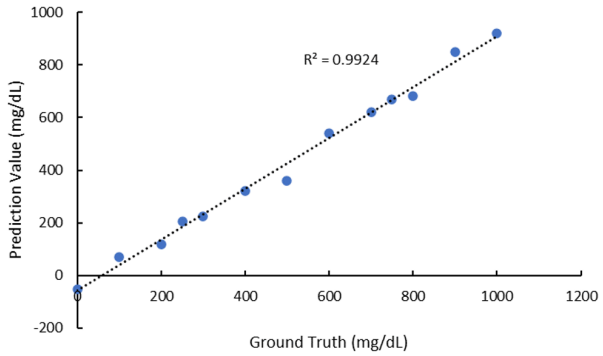


Fig. 5. The prediction results of different data sets. The samples in training data set and testing data set were prepared in different dates to show the repeatability of sample preparation.

the spectra. However, statistical models such as PLSR can overcome this problem [7].

Table II shows the LOCO cross-validation results of the PLSR model, including the prediction value and standard deviation (STD). The prediction value was an average of the prediction of all sample data in same image. The STD was the standard deviation of all sample data in same image. The root-mean-squared error (RMSE) is 29.60 mg/dL, which is calculated based on the average of the prediction values. The $R^2$ is 0.99, shown in Fig. 4. It shows that the prediction and the ground true are highly correlated.

### *B. Prediction base on the data from different dates*

To validate the generalization of glucose concentration prediction, the PLSR model was used to predict different glucose samples captured by different people from different date. The results are shown in Fig. 5. The $R^2$ and RMSE are 0.99 and 79.94 mg/dL respectively, which show that the PLSR model is still feasible for the prediction of glucose concentration in different data set with the same concentration range.

### *C. Repeatability Test*

To validate the repeatability of sample preparation and the stability of the hyperspectral imaging system, the PLSR model was also used to test a data set of 24 different images with the same glucose concentration (500 mg/dL). The results are shown in Fig. 6. Each predicted value is the average prediction from all column average samples from the same image. As seen in the figure, all prediction values are between 480 to 515 mg/dL, and the STD is 10.08 mg/dL. The data shows both the repeatability and the stability of the system.

### *D. Higher Resolution Test*

To evaluate the feasibility of this model for high resolution glucose concentration predictions, we trained another PLSR model using sample data (12/22/2020 experiment, shown in Table I) with minimal concentration increment 20 mg/dL. Similar to section III-A, LOCO cross-validation was used to validate the PLSR model. The cross-validation results are shown in Table III and Fig.7. Although the RMSE is 37.5 mg/dL which is larger than the concentration increment, the predicted $R^2$ is 0.93. It shows that there is high correlation between the ground truth and the prediction value. However, the accuracy is not high comparing to the results in Section 3.1, which can be explained as the lack of training data. Besides, the precision of scale for sample preparation may also be one of the factors. Further improvements of the data collection procedures are in progress.

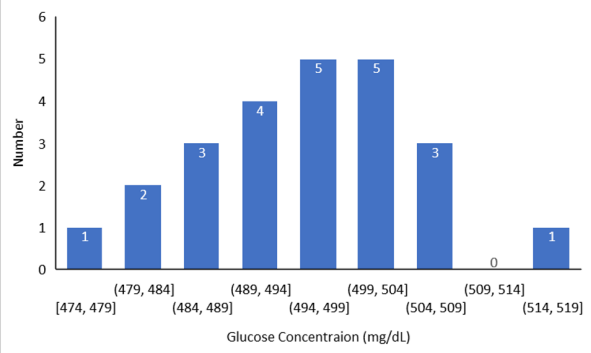


Fig. 6. The prediction results of 24 different images with same glucose concentration (500 mg/dL). Most of the predictions are in the range from 490 to 510 mg/dL.

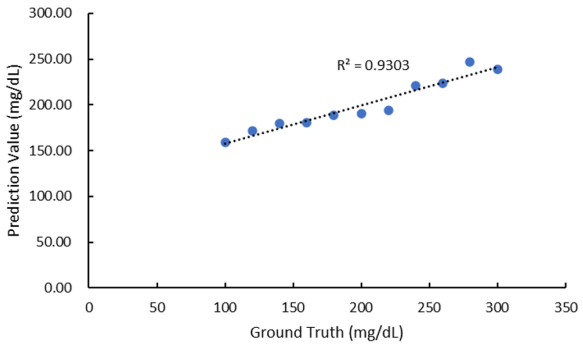


Fig. 7. The prediction results of LOCO cross-validation in higher resolution. Glucose concentrations are between 100 and 300 mg/dL, with 20 mg/dL interval.

TABLE III

LOCO CROSS-VALIDATION RESULTS OF THE PLSR MODEL. GLUCOSE CONCENTRATIONS ARE BETWEEN 100 AND 300 MG/DL, WITH 20 MG/DL INTERVAL.

| Ground truth (mg/dL) | Prediction (mg/dL) | STD (mg/dL) |
|---|---|---|
| 100 | 159.03 | 36.00 |
| 120 | 172.03 | 34.86 |
| 140 | 179.31 | 38.21 |
| 160 | 180.48 | 43.36 |
| 180 | 188.27 | 37.12 |
| 200 | 190.31 | 35.05 |
| 220 | 194.45 | 46.37 |
| 240 | 221.06 | 43.44 |
| 260 | 223.75 | 33.30 |
| 280 | 247.31 | 43.41 |
| 300 | 238.71 | 36.40 |

## IV. CONCLUSION AND DISCUSSION

To the best of our knowledge, this is the first study which applies HSI to quantize aqueous glucose concentration. Unlike NIRS, HSI allows the acquisition of thousands of spectra in a single shot. This saves significant time for collecting spectral data sets. The method presented of combining HSI and PLSR can serve as an online method for the non-invasive prediction of aqueous glucose concentration.

Although the raw spectra of glucose at varying concentrations look nearly identical, PLSR can identify the distinguishing features of these spectra to produce a reliable and accurate predictions. Glucose concentration within 100 mg/dL resolution are predicted with $R^2$ value larger than 0.99 according to the cross-validation results. The repeatability of glucose sample preparation and the stability of our HSI system are proven through this study. The next steps will include improving the experimental process and enhancing the accuracy of high resolution glucose concentration predictions using more advanced machine learning models. Overall, the study shows the feasibility of this method and develops a new useful tool for sensing aqueous glucose concentration non-invasively.

## ACKNOWLEDGMENT

The authors would like to acknowledge the funding support from the National Institute for Innovation in Manufacturing Biopharmaceuticals (NIIMBL, Grant number: PC2.1-058) and the assistance of our partners at Genentech and Artemis Biosystems for the study.

## REFERENCES

[1] Villena Gonzales, Wilbert, Ahmed Toaha Mobashsher, and Amin Abbosh. "The progress of glucose monitoring—A review of invasive to minimally and non-invasive techniques, devices and sensors." Sensors 19.4 (2019): 800.

[2] Kozma, B., Hirsch, E., Gergely, S., Párta, L., Pataki, H., & Salgó, A. (2017). On-line prediction of the glucose concentration of CHO cell cultivations by NIR and Raman spectroscopy: comparative scalability test with a shake flask model system. Journal of Pharmaceutical and Biomedical Analysis, 145, 346-355.

[3] Klonoff, D. C. (2005). Continuous glucose monitoring: roadmap for 21st century diabetes therapy. Diabetes care, 28(5), 1231-1239.

[4] Yoo, Eun-Hyung, and Soo-Youn Lee. "Glucose biosensors: an overview of use in clinical practice." Sensors 10.5 (2010): 4558-4576.

[5] Qiu, Jiang, Mark A. Arnold, and David W. Murhammer. "On-line near infrared bioreactor monitoring of cell density and concentrations of glucose and lactate during insect cell cultivation." Journal of Biotechnology 173 (2014): 106-111.

[6] Rondonuwu, F. S., A. Setiawan, and F. F. Karwur. "Determination of glucose concentration in aqueous solution using FT NIR spectroscopy." Journal of Physics: Conference Series. Vol. 1307. No. 1. IOP Publishing, 2019.

[7] Mekonnen, Bitewulign Kassa, et al. "Accurate prediction of glucose concentration and identification of major contributing features from hardly distinguishable near-infrared spectroscopy." Biomedical Signal Processing and Control 59 (2020): 101923.

[8] Abu-Absi, Nicholas R., et al. "Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe." Biotechnology and bioengineering 108.5 (2011): 1215-1221.

[9] Manley, Marena. "Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials." Chemical Society Reviews 43.24 (2014): 8200-8214.

[10] Lu, Guolan, and Baowei Fei. "Medical hyperspectral imaging: a review." Journal of biomedical optics 19.1 (2014): 010901.

[11] Baiano, Antonietta. "Applications of hyperspectral imaging for quality assessment of liquid based and semi-liquid food products: A review." Journal of Food Engineering 214 (2017): 10-15.

[12] Thyholt, Kari, and Tomas Isaksson. "Near infrared spectroscopy of dry extracts from high moisture food products on solid support—a review." Journal of Near Infrared Spectroscopy 5.4 (1997): 179-193.

[13] Abdi, Hervé. "Partial least square regression (PLS regression)." Encyclopedia for research methods for the social sciences 6.4 (2003): 792-795.