TOWARDS THE DEVELOPMENT OF SUBJECT-INDEPENDENT INVERSE METABOLIC MODELS

Seyedhooman Sajjadi[1], *Anurag Das*[1], *Ricardo Gutierrez-Osuna*[1], *Theodora Chaspari*[1], *Projna Paromita*[1], *Laura E. Ruebush*[2], *Nicolaas E. Deutz*[2], *Bobak J. Mortazavi*[1]

[1]Department of Computer Science & Engineering, Texas A&M University
[2]Center for Translational Research on Aging and Longevity, Texas A&M University
{hooman130, bobakm}@tamu.edu

ABSTRACT

Diet monitoring is an important component of interventions in type 2 diabetes, but is time intensive and often inaccurate. To address this issue, we describe an approach to monitor diet automatically, by analyzing fluctuations in glucose after a meal is consumed. In particular, we evaluate three standardization techniques (baseline correction, feature normalization, and model personalization) that can be used to compensate for the large individual differences that exist in food metabolism. Then, we build machine learning models to predict the amounts of macronutrients in a meal from the associated glucose responses. We evaluate the approach on a dataset containing glucose responses for 15 participants who consumed 9 meals. Three techniques improve the accuracy of the models: subtracting the baseline glucose, performing z-score normalization, and scaling the amount of macronutrients by each individuals' body mass index.

Index Terms— Continuous glucose monitors, diet monitoring, meal macronutrients, machine learning

1. INTRODUCTION

An essential component of clinical interventions for type-2 diabetes is monitoring dietary intake. However, this requires manual entry of each meal's nutritional content, which is time consuming and often inaccurate [1, 2]. Various technologies have been explored to capture dietary intake, such as wearable sensing [3] or computer vision [4, 5]. These methods reduce the burden to the user, but at present they are inaccurate and unreliable [6].

A unique and unexplored opportunity to solve the problem of automatic diet monitoring is the use of continuous glucose monitors (CGMs). A CGM measures glucose in the interstitial fluid every 5-15 minutes, thus providing a detailed pattern of fluctuations in glucose levels throughout the day. These glucose fluctuations in response to a meal, known as the post-prandial glucose response (PPGR), depend on the macronutrient composition of the meal (i.e., carbohydrates, protein, fat). Specifically, adding protein, fat, or fiber to a meal generally yields smaller spikes and lengthier responses [7, 8]. This suggests that the shape of the PPGR can be used to estimate the macronutrients in the meal. In what follows, we will refer to such models as inverse metabolic models (IMMs), in contrast with direct models that predict PPGRs given the macronutrients in a meal; see Related Work section.

A major hurdle to develop IMMs is that there exists significant inter-individual variability in the glucose response to a meal. In a landmark study, Zeevi et al. [9] tracked the glucose response of 800 participants for one week while participants kept detailed records of their diet, and found significantly different responses to identical meals. To address this issue, one might be inclined to build a personalized IMM for each individual. However, this requires a large amount of training data for each person, in the form of a variety of meals with the corresponding PPGRs, which may be impractical. Instead, we explore three complementary approaches (baseline correction, feature normalization, and personalization) to reduce individual differences in PPGRs, so that data from multiple participants can be pooled to train subject-independent IMMs. In baseline correction, we account for the pre-meal glucose levels, either by subtraction or division. In feature normalization, we scale the range of the feature space relative to the minima/maxima of the data or their mean/standard deviation. Finally, in model personalization we account for the body composition of each person, by using their body mass index as an auxiliary input to the model or as a scaling variable. We evaluated the approaches on a dataset consisting of PPGRs of 15 participants who consumed 9 meals of known macronutrient composition[1]. Our results show that the three approaches significantly improve the accuracy of IMMs using a leave-one-subject-out cross-validation procedure.

This work was supported, in part, by funding from the National Science Foundation (Award #2014475) and Engineering Research Center for Precise Advanced Technologies and Health Systems for Underserved Populations (PATHS-UP) (Award #1648451).

[1]A preliminary version of this study, involving 9 subjects, showed poor results due to individual variability [10]. Here, we increase the number of subjects, address individual variability, and propose an updated IMM.

2. RELATED WORK

To our knowledge, ours is the first attempt at building inverse metabolic models. In contrast, much interest has recently been raised in the direct problem: predicting the PPGR to a meal given its macronutrient composition. The most well-known study is Zeevi et al. [9]. In this study, the authors developed a machine learning model (based on boosted decision trees) that could predict the PPGR of a meal given its contents. To account for individual differences, the model used phenotype variables (e.g., anthropometric variables, blood panels, gut microbiota) as additional inputs. Motivated by these findings, Tily et al. [11] tracked the glycemic response of 550 adults for up to two weeks, while they consumed a set of standardized meals carefully designed to cover a broad range of proportions of carbohydrates, proteins, fats, and fiber. Then, they built a multilevel mixed-effects regression model to predict PPGRs. This allowed the authors to quantify the relative influence of meal composition, anthropometric, gut microbiome and lifestyle variables in postprandial glucose.

Our work is related to artificial pancreas research [12]: to infer food intake from CGM data. In the artificial pancreas, the goal is to control an insulin pump, which administers doses of insulin according to a preestablished insulin-to-carbohydrate ratio. In contrast, we aim to estimate not just carbohydrates, but also protein and fat, which also affect the glycemic response of a meal. In addition, being a control problem, the artificial pancreas is sensitive to lags [13]: to prevent large glucose responses after a meal, an artificial pancreas must make a decision based on the early part of the glucose response. In contrast, we can afford to exploit information in the entire glucose response curve to predict the full macronutrient composition of a meal. While precise estimates are the goals, coarse estimates may still be suitable for meal macronutrients (i.e., low, medium, high) similar to suitable findings in caloric estimation work [14, 15].

3. METHODS

3.1. Dataset

Before presenting the computational methods, we describe the experimental dataset of PPGRs that was used in this study, as we believe this helps fully appreciate the problem. To collect the dataset, we recruited 15 healthy subjects, ages 60-85 years and BMI of 25-35. Each subject participated in 9 study days in which they consumed a predefined meal in a randomized design. Each study day lasted approximately 8 hours. The procedures on the study days were identical, with the only change being the macronutrient composition of the meal taken (e.g., varying amounts of carbohydrates, proteins, and fats). Subjects were asked to fast for at least 8 hours prior to the meal intake on each study day so that the first blood glucose reading would be their fasting glucose level. After taking a baseline blood sample the morning of a study visit, each subject consumed a predefined meal. Subjects remained in a sedentary state and were not allowed to consume any other food for the next 8 hours. This study was approved by the Texas A&M Institutional Review Board (#2017-0886). Table 1 shows the composition of each of the nine meals. Each meal had a known amount of carbohydrates (C), protein (P), and fat (F), which we denote as CxPxFx, where x represents the amount of each macronutrient in the meal (1: low; 2: medium; 3: high). A total of 4 meals (from the total set of 135 meals) were excluded due to missing and noisy data.

Table 1. Composition of meals for C, P, and F

Meal	C (g)	P (g)	F (ml)
C1P1F1	52.25	15	13
C1P2F2	52.25	30	26
C2P2F1	94.75	30	13
C2P1F2	94.75	15	26
C2P2F2	94.75	30	26
C2P2F3	94.75	30	52
C2P3F2	94.75	60	26
C3P2F2	179.75	30	26
C3P3F3	179.75	60	52

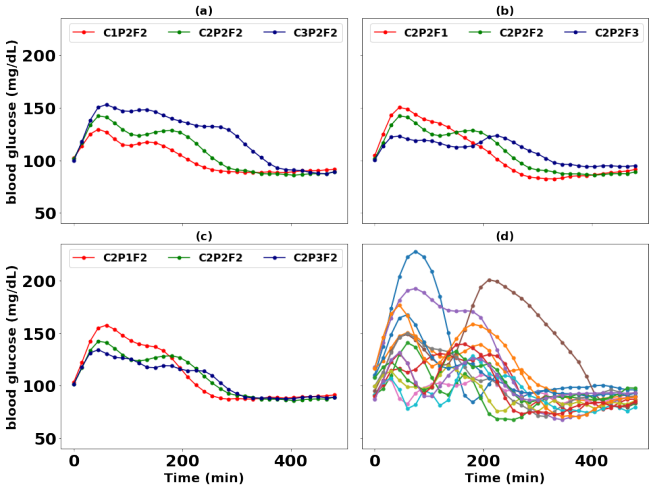

Fig. 1. (a) PPGR at increasing levels of C (P and F fixed). (b) PPGR at increasing levels of P. (c) PPGR at increasing levels of F. (d) PPGR of C2P2F2 meal for all participants

Fig. 1 illustrates the average PPGR across all participants while we change the C, P, and F concentrations from low to medium to high (leaving the others fixed). This result shows that increasing the amount of macronutrients alters the shape of the PPGR, which supports our hypothesis that the PPGR may be used to estimate macronutrients. On the flip side, Fig. 1(d) shows the PPGR to the intermediate meal (C2P2F2) of all 15 participants, which illustrates the significant inter-subject variability and the challenge of developing an accurate, subject-independent model.

3.2. Overview of signal processing methods

Our computational approach to analyzing PPGRs consists of four steps: data preprocessing, feature extraction, standardization, and model training. In a first step, we preprocess the

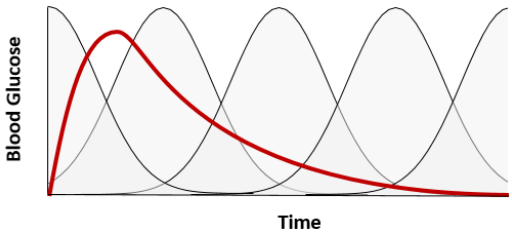

Fig. 2. Extraction of gAUC features using 5 kernels

raw PPGRs with a Kalman filter [16] to de-noise the signal and handle missing values. Next, we extract features to capture the shape of the PPGRs. Namely, we place a family of Gaussian kernels uniformly over the time axis in the PPGR, and compute the Gaussian area-under-the-curve (gAUC) [17], as illustrated in Fig. 2 and defined as:

$$x(i) = \int_0^T g(t) e^{\frac{1}{2\sigma_k^2}(t-T_k)} dt \quad (1)$$

where each gAUC feature $x(i)$ is computed from PPGR $g(t)$, T is the duration of the PPGR, T_k is the center position of the kernel in the time domain, and σ_k^2 is its spread. These gAUC features capture the initial rise time in glucose, duration of the elevated glucose levels, and the recovery back to the baseline glucose level. We evaluated families of 3, 5, and 9 kernels as well as their combinations, and found the combination of 3 + 5 kernels yield the best performance. Thus, the rest of this paper uses 3 + 5 kernels. [2]

In a third step, we apply the standardization methods to reduce individual differences in these PPGR features, which we describe in the next section. In a final step, we train an IMM to predict the macronutrient composition of the meals from the resulting features. In this work, we use gradient descent boosting (XGBoost) [18] to build the IMMs, and we compare it against Linear Regression (LR) as a baseline technique. We evaluate the various IMMs through Pearson's correlation coefficient and Root Mean Squared Relative Error. Since the three macronutrients have different ranges of quantities, using the RMSRE (instead of the more conventional RMSE) aids in comparing performance across the three macronutrients. We trained IMMs in a leave-one-subject-out fashion and performed a training-set cross-validation for hyperparameter tuning {estimators: 1-300, learning rate: 0.1-0.15, and maximum depth: 1-4}.[3]

3.3. Reducing individual differences in post-prandial glucose responses

3.3.1. Baseline Correction

The first technique consists of correcting for the pre-meal glucose level, which is unique to each individual. The rationale behind this method is that differences in PPGRs when two individuals eat the same meal may be due to differences in fasting glucose. Thus, these baseline correction techniques allow the gAUC features to represent relative changes to blood glucose, rather than absolute quantities. Following [17], we consider two baseline correction techniques: subtracting the fasting glucose from all readings, and dividing all readings by the fasting glucose level. In our study, the fasting (or baseline) glucose level is computed as the average of the last three measurements before the meal is consumed.

[2]kernels out of this range did not have significant impact on the results.

[3]We also experimented using Neural networks [10] and Adaboost models, but there was no significant difference in the results.

Table 2. Results of the base models using LR and XGBoost

	Correlation			Mean RMSRE (std)		
Model	**C**	**P**	**F**	**C**	**P**	**F**
LR	0.40	0.15	0.31	0.44(0.18)	0.57(0.21)	0.55(0.21)
XGBoost	**0.55**	**0.42**	**0.39**	**0.41(0.17)**	**0.51(0.15)**	**0.51(0.16)**

Correlation significance: $p < 0.001$, except LR-P $p < 0.05$, and LR-F $p < 0.1$

Table 3. Results of baseline correction using XGBoost

	Correlation			Mean RMSRE (std)		
Model	**C**	**P**	**F**	**C**	**P**	**F**
None	0.55	0.42	0.39	0.41(0.17)	0.51(0.15)	0.51(0.16)
Subtraction	**0.61**	**0.48**	**0.48**	**0.35(0.20)**	**0.50(0.13)**	**0.49(0.15)**
Division	0.59	0.49	0.47	0.34(0.21)	0.49(0.12)	0.51(0.13)

Correlation significance: $p < 0.001$

3.3.2. Feature normalization

A complementary approach is to perform feature normalization [19]. In this study, we consider two techniques: min-max normalization and z-score normalization. Briefly, min-max normalization scales each feature (i.e., a gAUC) between 0 and 1 by subtracting its minimum value and dividing by the range, whereas z-score normalization subtracts the mean of each feature and divides it by the standard deviation. The normalizing variables (i.e., minimum, maximum, mean and standard deviation) can also be computed across all features (i.e., a common mean for all 8 gAUC features). We shall refer to the former form of normalization as feature-wise normalization, and the latter as curve-wise normalization.

3.3.3. Personalization

Our third standardization approach attempts to account for differences in body composition, which is known to impact the absorption rate of macronutrients [9, 11] and therefore can lead to differences in PPGRs across individuals. As an example, a given meal is likely to result in a larger glucose excursion on a 5'2" 120 lb person than on a 6'5" 220 lb person. We consider two measures of body composition: Lean Body Mass (LBM) and Body Mass Index (BMI), and examine two approaches to incorporate this information into the models: (1) adding these variables as auxiliary inputs, and (2) dividing the amount of macronutrients by them. Note that, in the second approach, we are then predicting the amount of macronutrients relative to an individual's body composition, rather than the absolute macronutrient amount.

Table 4. Results of feature normalization using XGBoost

	Correlation			Mean RMSRE (std)		
Model	**C**	**P**	**F**	**C**	**P**	**F**
None	0.61	0.48	0.48	0.35(0.20)	0.50(0.13)	0.49(0.15)
min-max*	0.77	**0.48**	0.64	0.28(0.16)	**0.47(0.17)**	0.41(0.14)
min-max	0.76	0.38	0.49	0.27(0.12)	0.51(0.16)	0.51(0.19)
z-score*	**0.83**	0.43	**0.65**	**0.22(0.10)**	0.50(0.12)	**0.40(0.14)**
z-score	0.79	0.46	0.63	0.26(0.08)	0.51(0.14)	0.46(0.14)

*: Feature-wise, Correlation significance: $p < 0.001$

Table 5. Results of two meal-normalization using XGBoost

	Correlation			Mean RMSRE (std)		
Model	**C**	**P**	**F**	**C**	**P**	**F**
None	0.61	0.48	0.48	0.35(0.20)	0.50(0.13)	0.49(0.15)
all meals	0.83	0.43	0.65	0.22(0.10)	0.50(0.12)	0.40(0.14)
two meals	0.74	0.57	0.67	0.28(0.12)	0.47(0.19)	0.40(0.15)

Correlation significance: $p < 0.001$

4. RESULTS

In a first step, we compared XGBoost against LR. For this purpose, no standardization was applied to the data. We used a Fisher r-to-z transformation to compare the statistical significance between correlations, and a difference of means test to compare RMSRE. Results are shown in Table 2. XGBoost outperformed LR, implying a higher-dimensional, nonlinear relationship between PPGRs and macronutrients. The increase in correlation had p-values of p=0.057, p=0.009, and p=0.230 for C, P, and F, respectively. This shows a statistically significant improvement in P, with the F correlation moving from the $p < 0.05$ to the $p < 0.001$ significance level. Similarly, the RMSRE improvements had p=0.16, p=0.007, and p=0.08 for C, P, and F, respectively. Therefore, all subsequent evaluations will use XGBoost. The number of estimators was around 200 for C and 60 for P and F. Also, depth of the trees were 1 for C and 3 for P and F. The learning rate varied between 0.1 to 0.12.

4.1. Baseline Correction

Results for baseline correction are shown in Table 3, and compared against performing no correction. Both methods moderately improve model performance, and the subtraction method modestly outperforms the division method. Subtraction increases the correlation for C from 0.55 to 0.61 (p=0.23), and F from 0.42 to 0.48 (p=0.27), and also reduces the relative error (difference in mean RMSRE with p=0.008, p=0.56, and p=0.29 for C, P, and F respectively). In summary, accounting for fasting glucose, whether by subtraction or division, improves correlation and reduces error. While both are sufficient, we use subtraction moving forward.

4.2. Normalization

Results with feature normalization are shown in Table 4. Normalization sees large improvements in the prediction of C and F, and marginal improvements for P, with feature-wise normalization providing larger improvements than curve-wise. This suggests that each gAUC has variation that differs significantly for feature-wise versus curve-wise normalization. The improvements in correlation and RMSRE for C and F are statistically significant ($p < 0.001$). The results in Table 4 were normalized by using all the meals. Table 5 normalizes using only two meals, selected for calibration with widest variation in gAUC features: C1P1F1 and C3P3F3. The RMSRE improvements for C and F are still statistically significant ($p < 0.001$) over no normalization.

Table 6. Results of BMI/LBM utilization using XGBoost

BMI/LBM	Correlation			Mean RMSRE (std)		
Inclusion	**C**	**P**	**F**	**C**	**P**	**F**
None	0.61	0.48	0.48	0.35(0.20)	0.50(0.13)	0.49(0.15)
As inputs	0.60	0.51	0.52	0.35(0.19)	0.48(0.13)	0.47(0.15)
Scale by LBM	0.61	0.47	0.56	0.32(0.09)	0.49(0.18)	0.48(0.16)
Scale by BMI	**0.68**	**0.55**	**0.56**	**0.31(0.04)**	**0.39(0.10)**	**0.39(0.10)**

Correlation significance: $p < 0.001$

4.3. Personalization

Table 6 shows the results of using BMI and LBM to personalize the IMMs. Adding these physiological features as auxiliary inputs to XGBoost does not improve performance. However, scaling the target values by BMI enhances both the correlation and RMSRE for all macronutrients (p=0.16, p=0.22, p=0.19 for correlation of C, P, and F, respectively; p=0.02, $p < 0.001$, $p < 0.001$ for RMSRE of C, P, and F, respectively). However, when scaling by BMI after z-score normalization of all meals, we see a drop in performance for C and F. Although label scaling does not perform as well as normalization, the improvement of results suggests it as a substantial cause of inter-subject variability.

5. LIMITATIONS AND FUTURE WORK

Our study was conducted in a controlled setting, where participants remained stationary following a meal's consumption. Therefore, the model does not account for physical activity that may follow a meal, which is known to reduce postprandial glucose. Thus, future work is needed to evaluate our approach in more naturalistic settings and with a larger variety of meals. Work is underway to conduct experiments in which participants consume multiple solid and liquid meals over an extended period while carrying out their daily lives.

6. CONCLUSION

This work examined several techniques to reduce individual differences in PPGRs towards estimating meal macronutrients. We evaluated techniques to correct baseline glucose, normalize features and personalize models to account for physiological variables. Our results show promise for three techniques: subtracting baseline glucose, computing z-scores and scaling the amount of macronutrients by BMI. These models may eventually help health practitioners to monitor diet without adding burden to their patient's lives.

7. REFERENCES

[1] Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty, "Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 3207–3216.

[2] Chia-Fang Chung, Elena Agapie, Jessica Schroeder, Sonali Mishra, James Fogarty, and Sean A Munson, "When personal tracking becomes social: Examining the use of instagram for healthy eating," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 1674–1687.

[3] Haik Kalantarian, Nabil Alshurafa, Tuan Le, and Majid Sarrafzadeh, "Monitoring eating habits using a piezoelectric sensor-based necklace," *Computers in biology and medicine*, vol. 58, pp. 46–55, 2015.

[4] Shaobo Fang, Zeman Shao, Deborah A Kerr, Carol J Boushey, and Fengqing Zhu, "An end-to-end image-based automatic food energy estimation technique based on learned energy distribution images: Protocol and methodology," *Nutrients*, vol. 11, no. 4, pp. 877, 2019.

[5] Haik Kalantarian, Nabil Alshurafa, and Majid Sarrafzadeh, "A survey of diet monitoring technology," *IEEE Pervasive Computing*, vol. 16, no. 1, pp. 57–65, 2017.

[6] Hamid Hassannejad, Guido Matrella, Paolo Ciampolini, Ilaria De Munari, Monica Mordonini, and Stefano Cagnoni, "Automatic diet monitoring: a review of computer vision and wearable sensor-based methods," *International journal of food sciences and nutrition*, vol. 68, no. 6, pp. 656–670, 2017.

[7] Thomas MS Wolever and Claudia Bolognesi, "Prediction of glucose and insulin responses of normal subjects after consuming mixed meals varying in energy, protein, fat, carbohydrate and glycemic index," *The Journal of nutrition*, vol. 126, no. 11, pp. 2807–2812, 1996.

[8] Andreas Rytz, Dorothée Adeline, Kim-Anne Lê, Denise Tan, Lisa Lamothe, Olivier Roger, and Katherine Macé, "Predicting glycemic index and glycemic load from macronutrients to accelerate development of foods and beverages with lower glucose responses," *Nutrients*, vol. 11, no. 5, pp. 1172, 2019.

[9] David Zeevi, Tal Korem, Niv Zmora, David Israeli, Daphna Rothschild, Adina Weinberger, Orly Ben-Yacov, Dar Lador, Tali Avnit-Sagi, Maya Lotan-Pompan, et al., "Personalized nutrition by prediction of glycemic responses," *Cell*, vol. 163, no. 5, pp. 1079–1094, 2015.

[10] Zepeng Huo, Bobak J Mortazavi, Theodora Chaspari, Nicolaas Deutz, Laura Ruebush, and Ricardo Gutierrez-Osuna, "Predicting the meal macronutrient composition from continuous glucose monitors," in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE, 2019, pp. 1–4.

[11] Hal Tily, Ally Perlina, Eric Patridge, Stephanie Gline, Matvey Genkin, Vishakh Gopu, Haely Lindau, Alisson Sjue, Niels Klitgord, Momchilo Vuyisich, et al., "Gut microbiome activity contributes to individual variation in glycemic response in adults," *bioRxiv*, p. 641019, 2019.

[12] Odd Martin Staal, Steinar Sælid, Anders Lyngvi Fougner, and Øyvind Stavdahl, "Meal estimation from continuous glucose monitor data using kalman filtering and hypothesis testing," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 5654–5661.

[13] Lauren M Huyett, Eyal Dassau, Howard C Zisser, and Francis J Doyle III, "Glucose sensor dynamics and the artificial pancreas: the impact of lag on sensor measurement and controller performance," *IEEE Control Systems Magazine*, vol. 38, no. 1, pp. 30–46, 2018.

[14] Jon Noronha, Eric Hysen, Haoqi Zhang, and Krzysztof Z Gajos, "Platemate: crowdsourcing nutritional analysis from food photographs," in *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 2011, pp. 1–12.

[15] Weiqing Min, Shuqiang Jiang, Linhu Liu, Yong Rui, and Ramesh Jain, "A survey on food computing," *ACM Computing Surveys (CSUR)*, vol. 52, no. 5, pp. 1–36, 2019.

[16] Odd Martin Staal, Steinar Sælid, Anders Fougner, and Øyvind Stavdahl, "Kalman smoothing for objective and automatic preprocessing of glucose data," *IEEE journal of biomedical and health informatics*, vol. 23, no. 1, pp. 218–226, 2018.

[17] Ricardo Gutierrez-Osuna and H Troy Nagle, "A method for evaluating data-preprocessing techniques for odour classification with an array of gas sensors," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 29, no. 5, pp. 626–632, 1999.

[18] Tianqi Chen and Carlos Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[19] Andreas Hierlemann and Ricardo Gutierrez-Osuna, "Higher-order chemical sensing," *Chemical reviews*, vol. 108, no. 2, pp. 563–613, 2008.