Selection of noninvasive features in a wrist-based wearable sensors to predict blood glucose concentrations using machine learning algorithms

Brian Bogue-Jimenez [1], Xiaolei Huang [2], Douglas Powell [3] and Ana Doblas[1,*]

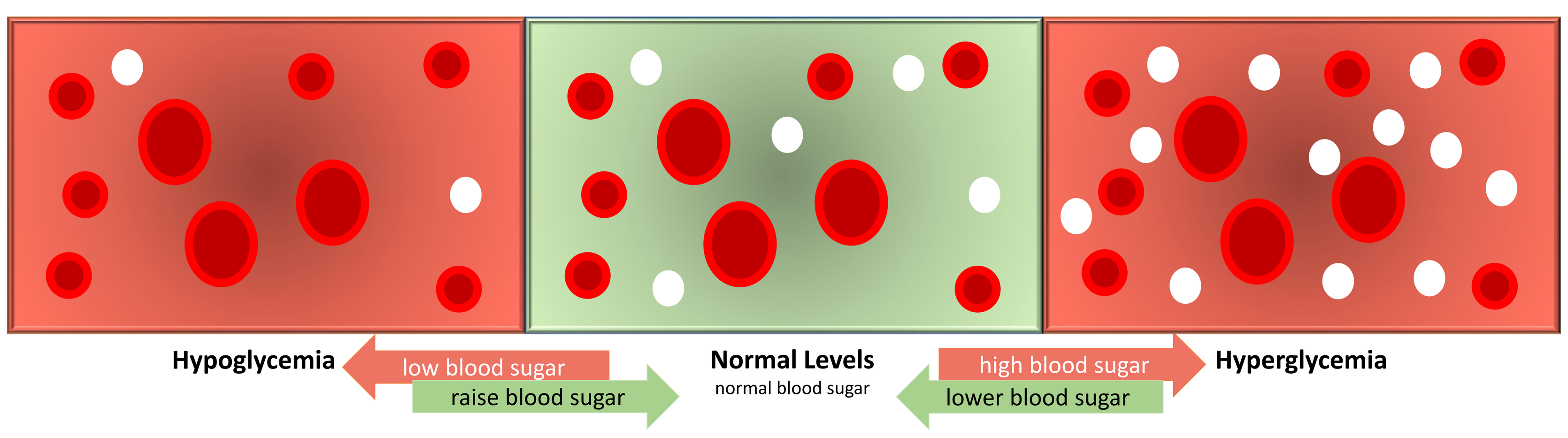
1 Department of Electrical and Computer Engineering, The University of Memphis, Memphis, USA; bbgjmnez@memphis.edu; adoblas@memphis.edu[2]
2 Department of Computer Science, The University of Memphis, Memphis, USA; xiaolei.huang@memphis.edu
3 College of Health Sciences, The University of Memphis, Memphis, USA; douglas.powell@memphis.edu

ABSTRACT

Glucose monitoring technologies allow users to monitor glycemic fluctuations (e.g., blood glucose levels). This is particularly important for individuals who suffer from diabetes mellitus (DM). Traditional self-monitoring blood glucose (SMBG) devices require the user to prick their finger and extract a blood drop to measure the blood glucose based on chemical reactions with the blood. Unlike traditional glucometer devices, noninvasive continuous glucose monitoring (NICGM) devices aim to solve these issues by consistently monitoring users' blood glucose levels (BGLs) without invasively acquiring a sample. In this work, we have investigated the feasibility of a novel approach to NICGM using multiple off-the-shelf wearable sensors and learning-based models (i.e., machine learning) to predict blood glucose. A multimodal dataset was created (e.g., the UofM dataset) by the research team. This UofM dataset consists of fourteen features provided by six sensors for studying possible relationships between glucose and noninvasive biometric measurements. Both datasets are passed through a machine learning pipeline that tests linear and non-linear models to predict BGLs from the set of noninvasive features. The results of this pilot study show that the combination of fourteen noninvasive biometric measurements with regression algorithms could lead to accurate BGL predictions within the clinical range; however, a larger dataset is required to make conclusions about the feasibility of this approach.

BACKGROUND

Figure 1. Normal levels of Blood Glucose vs Hypo/Hyper - glycemia



- ❖ Important for managing health conditions such as Diabetes [1].
- ❖ Diabetes has been estimated to effect 450 million worldwide [2].
- ❖ Management requires tracking glycemia

METHODOLOGY

Figure 2. Sensors and their features for the UofM dataset.

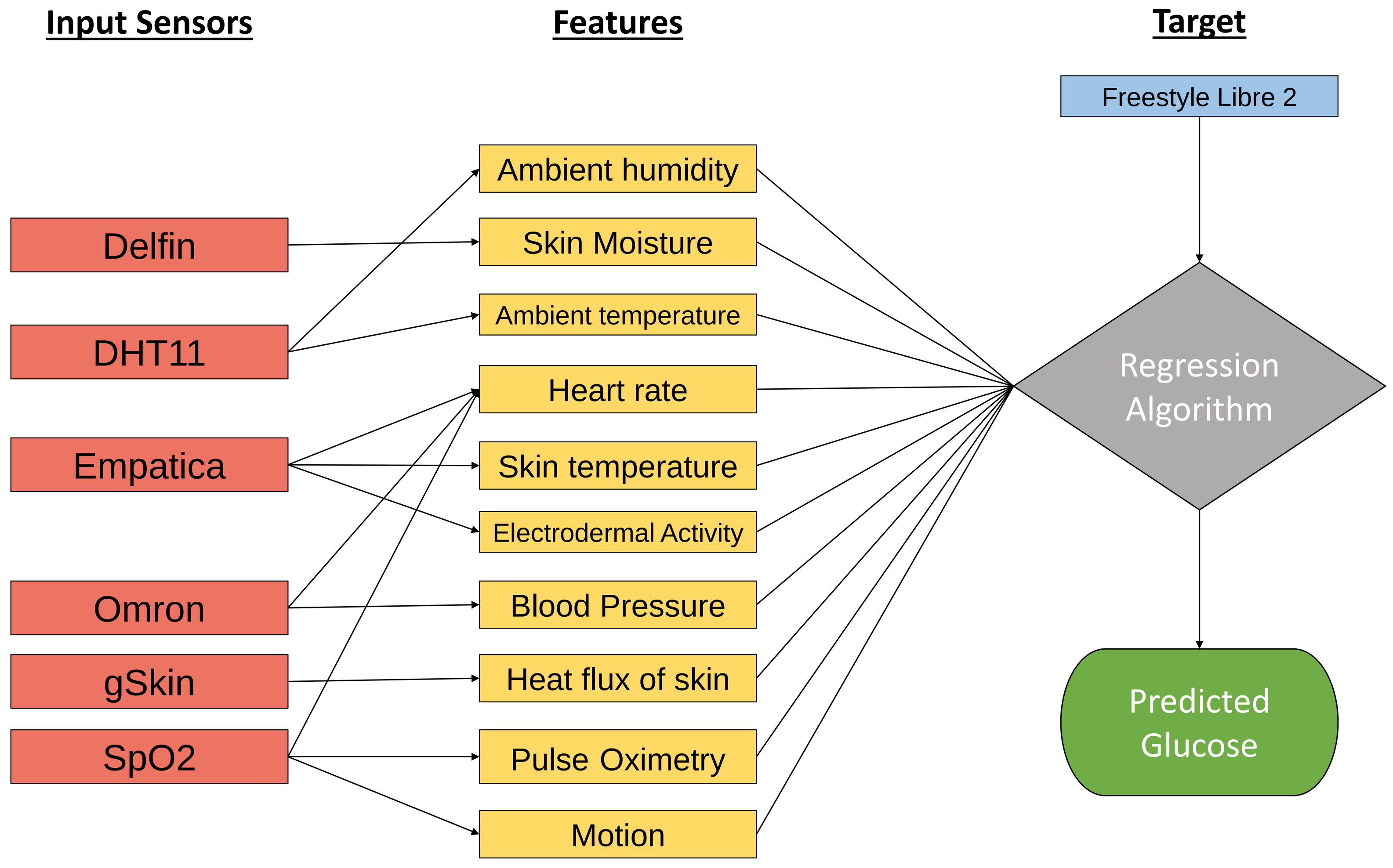
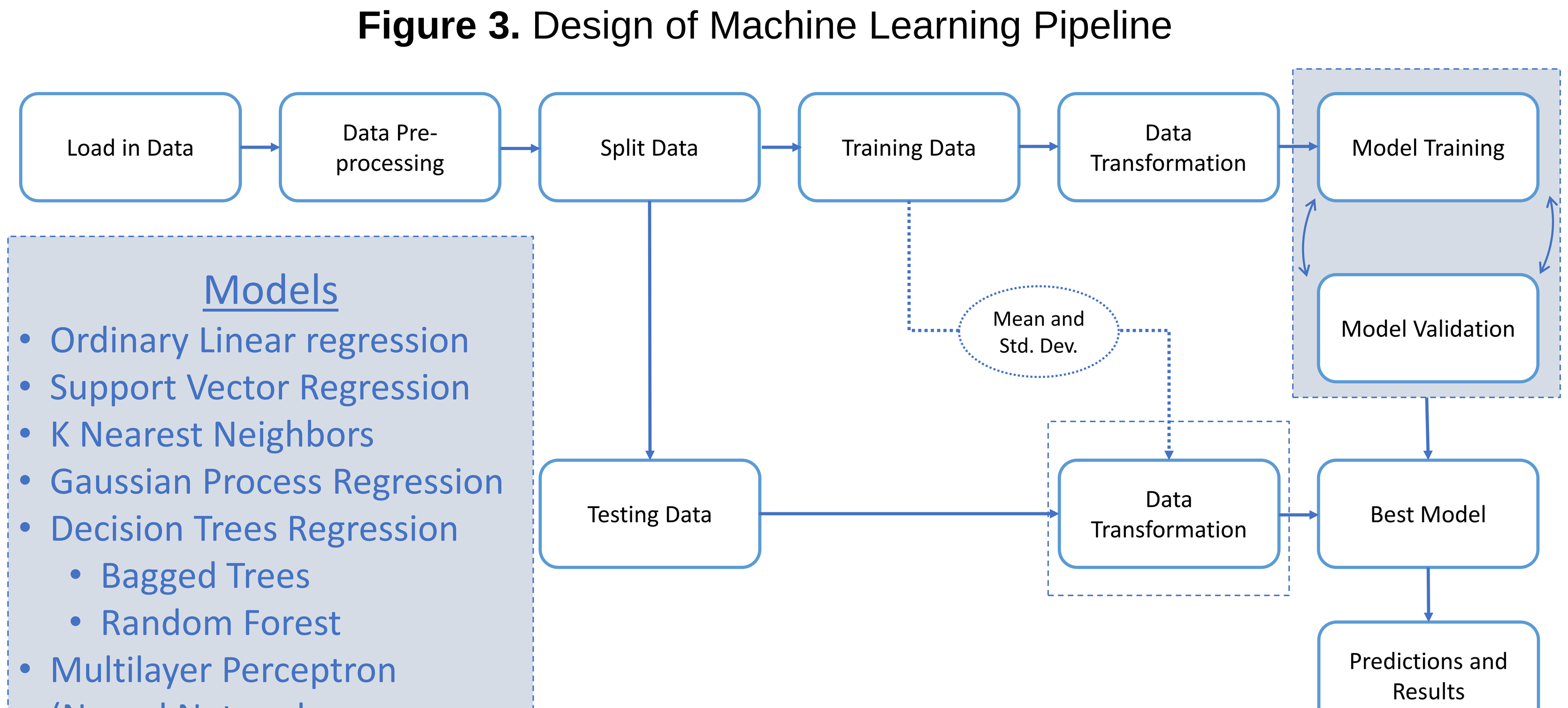


Figure 3. Design of Machine Learning Pipeline



Models
- Ordinary Linear regression
- Support Vector Regression
- K Nearest Neighbors
- Gaussian Process Regression
- Decision Trees Regression
  - Bagged Trees
  - Random Forest
- Multilayer Perceptron (Neural Network)

Metrics
- Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

- Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

- ❖ Best model for each dataset chosen according to reproducible protocol
- ❖ Tested several common Machine Learning algorithms that can accommodate for bot linear and non-linear relationships
- ❖ Used K-Fold Cross Validation to ensure robustness against overfitting

RESULTS

- ❖ Glucose values show moderate correlation with heart rate, motion moisture ventral, and EDA for both subjects 1 and 2.
- ❖ Heart rate is highly correlated to EDA, and blood pressure.

Figure 4. The correlation coefficient between target glucose concentrations and the selected fourteen features for both subjects from the UofM dataset to identify linear relationship between features and potential redundant features

(a)

Features	Glucose	SpO2	Heart rate	Motion	Amb. Humidity	Amb. Temp.	Moisture: Ventral	Moisture: Dorsal	Heat Flux	BVP	IBI	Skin Temp.	EDA	Systolic	Diastolic
Glucose	1.00														
SpO2	-0.02	1.00													
Heart rate	-0.32	-0.07	1.00												
Motion	-0.24	0.06	0.48	1.00											
Amb. Humidity	-0.08	0.18	0.02	-0.17	1.00										
Amb. Temp.	-0.10	0.13	-0.05	-0.18	0.86	1.00									
Moisture:Ventral	-0.24	0.05	0.22	-0.04	0.79	0.62	1.00								
Moisture:Dorsal	-0.19	-0.19	0.42	0.14	-0.10	-0.21	0.16	1.00							
Heat Flux	-0.45	0.08	0.35	0.44	-0.10	-0.30	0.03	0.09	1.00						
BVP	-0.04	-0.12	0.15	-0.02	-0.07	-0.05	-0.03	0.09	-0.02	1.00					
IBI	0.05	-0.04	0.06	-0.07	-0.08	-0.23	0.06	0.27	0.06	0.09	1.00				
Skin Temp.	-0.05	0.23	-0.10	-0.30	0.16	0.25	0.10	0.04	-0.43	-0.01	-0.04	1.00			
EDA	-0.45	-0.10	0.71	0.47	-0.04	-0.17	0.30	0.57	0.49	0.10	0.21	-0.05	1.00		
Systolic	-0.09	-0.02	0.71	0.46	-0.17	-0.38	0.12	0.49	0.32	0.10	0.24	-0.15	0.61	1.00	
Diastolic	-0.08	-0.09	0.39	0.14	-0.07	-0.27	0.10	0.42	0.30	-0.01	0.29	-0.12	0.48	0.47	1.00

(b)

Features	Glucose	SpO2	Heart rate	Motion	Amb. Humidity	Amb. Temp.	Moisture: Ventral	Moisture: Dorsal	Heat Flux	BVP	IBI	Skin Temp.	EDA	Systolic	Diastolic
Glucose	1.00														
SpO2	0.40	1.00													
Heart rate	-0.37	-0.48	1.00												
Motion	-0.32	-0.31	0.62	1.00											
Amb. Humidity	0.08	-0.12	0.05	0.08	1.00										
Amb. Temp.	-0.06	0.09	-0.27	-0.09	0.47	1.00									
Moisture:Ventral	-0.21	-0.33	0.55	0.34	-0.04	-0.29	1.00								
Moisture:Dorsal	-0.28	-0.10	0.32	0.19	0.06	0.20	-0.05	1.00							
Heat Flux	0.12	-0.01	-0.02	-0.02	0.45	-0.09	-0.01	0.11	1.00						
BVP	0.25	0.20	-0.10	-0.03	-0.05	0	-0.13	-0.04	-0.03	1.00					
IBI	0.06	0.18	-0.27	-0.05	-0.14	-0.10	-0.16	-0.14	-0.18	-0.11	1.00				
Skin Temp.	-0.61	-0.40	0.58	0.42	-0.06	0.02	0.38	0.36	0.13	-0.10	-0.23	1.00			
EDA	-0.38	-0.37	0.87	0.54	-0.07	-0.43	0.62	0.33	-0.01	-0.01	-0.13	0.57	1.00		
Systolic	-0.18	-0.51	0.84	0.52	0.01	-0.43	0.53	0.22	-0.10	-0.05	-0.11	0.32	0.84	1.00	
Diastolic	-0.36	-0.43	0.83	0.53	-0.04	-0.39	0.59	0.29	-0.07	-0.05	-0.08	0.48	0.87	0.85	1.00

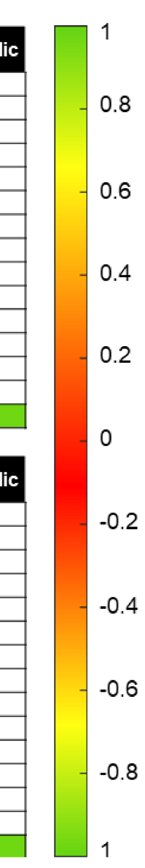
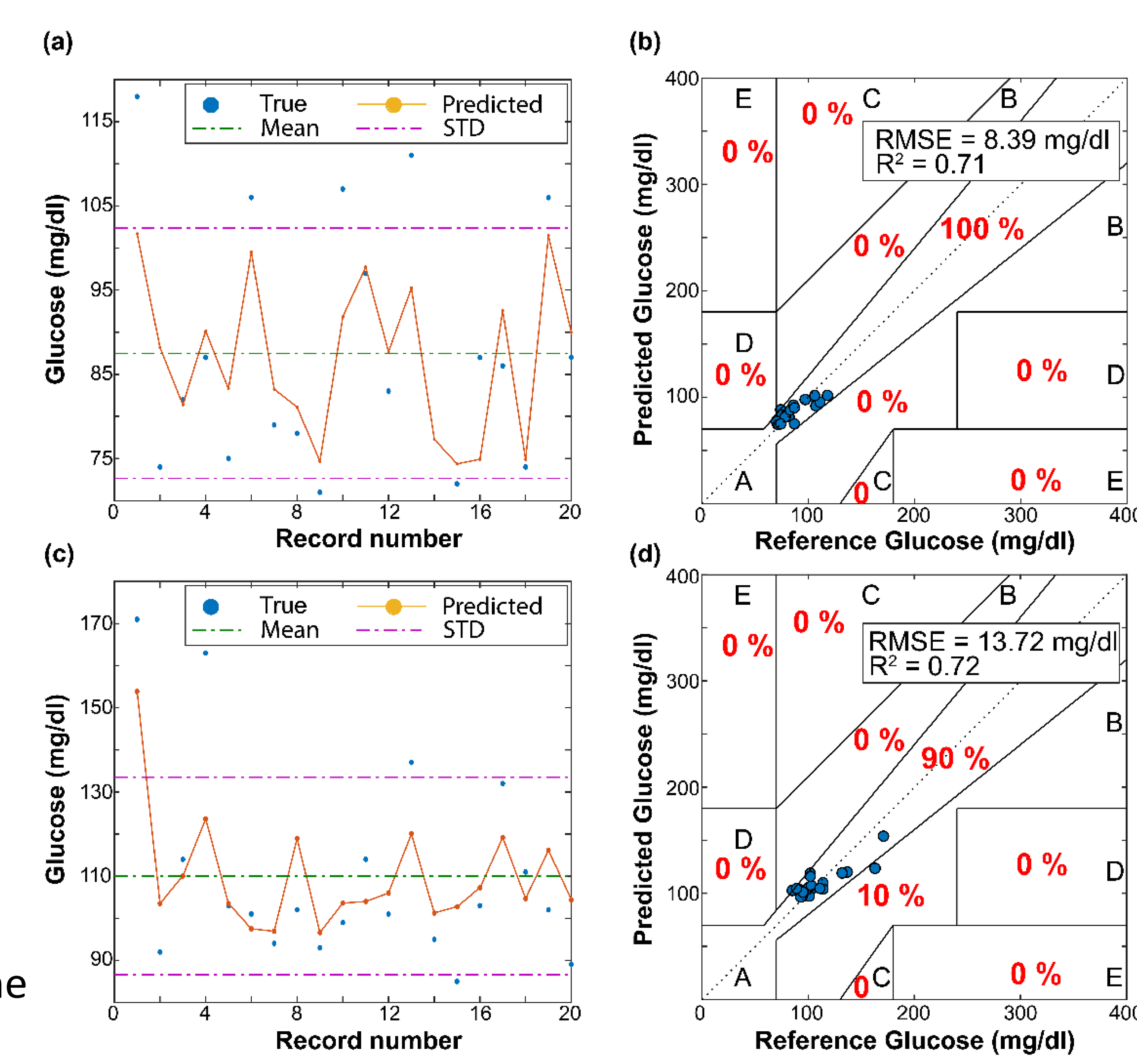


Figure 5. Performance of the best-trained model in unseen instances for both subjects from the UofM dataset. Panels (a) and (c) show the predicted and target glucose concentrations from unseen data. The Clarke Error Grids are shown in panels (b) and (d). RMSE and R^2 values of the best model are reported in panels (b) and (d). We also have reported the percentage of instances (red font) falling in each region of the Clarke Error Grid.



- ❖ Best model for Subject 1: Bagged Trees Ensemble
- ❖ Best model for Subject 2: Gaussian Process Regression (rational quadratic)
- ❖ All predictions fall within the clinical range (regions A & B)

CONCLUSIONS

- ❖Pilot study to validate a multi-sensor array with wearable sensors can predict blood glucose using a regression-based models.
- ❖Our dataset contains up to 15 features (Fig. 4).
- ❖Tested our hypothesis on UofM datasets
- ❖Future work involves the increasing of the number of subjects and samples.

REFERENCES

[1] W. V. Gonzales, A. T. Mobashsher, and A. Abbosh, "The progress of glucose monitoring—A review of invasive to minimally and non-invasive techniques, devices and sensors," *Sensors*, vol. 19, no. 4, pp. 1–45, 2019.

[2] I. Harman-Boehm, A. Gal, A. M. Raykhman, J. D. Zahn, E. Naidis, and Y. Mayzel, "Noninvasive glucose monitoring: A novel approach (Glucotrack)," *J. Diabetes Sci. Technol.*, vol. 3, no. 2, pp. 253–260, 2009.

ACKNOWLEDGEMENT

This work is supported by the University of Memphis. B. Bogue-Jimenez would like to acknowledge the support assistance of several faculty mentors at within the institution, and the mentorship of Drs. Ana Doblas, Xiaolei Huang, and Douglas Powell. Without all these components, none of this would have been possible

Optical Imaging Research Lab

Driven by doing.

THE UNIVERSITY OF MEMPHIS.