IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 16, 2022, accepted March 2, 2022, date of publication March 11, 2022, date of current version March 25, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3158945

Compact Instrumentation for Accurate Detection and Measurement of Glucose Concentration Using Photoacoustic Spectroscopy

FAHEEM SHAIKH[1], (Student Member, IEEE), NOAH HAWORTH[2], RILEY WELLS[2], JODI BISHOP[1], SHRE K. CHATTERJEE[3], SANKHA BANERJEE[2], AND SOUMYASANTA LAHA[1], (Member, IEEE)

[1]Department of Electrical and Computer Engineering, California State University, Fresno, Fresno, CA 93740, USA
[2]Department of Mechanical Engineering, California State University, Fresno, Fresno, CA 93740, USA
[3]Southampton Business School, University of Southampton, Highfield Campus, Southampton SO17 1BJ, U.K.

Corresponding author: Soumyasanta Laha (laha@csufresno.edu)

This work was supported by the California State University Program for Education & Research in Biotechnology (CSUPERB) - New Investigator Grant Award 2019 of the California State University.

ABSTRACT In this work, a novel *compact* and accurate glucose concentration measurement system is developed using the well-established photoacoustic Near Infra-Red spectroscopy. The proposed *in-vitro* instrumentation methods are in a small form factor, making it a viable candidate and precursor for an *in-vivo* non-invasive *wearable* blood glucose monitoring in the near future. The accuracy comes from the phase sensitive detection of the electrical signal. This detection technique uses an off-the shelf modulator/demodulator integrated circuit configured as a lock-in amplifier to increase the signal to noise ratio multifold. No prior work on photoacoustic spectroscopy, has taken advantage of this detection methodology in such a small form factor. The dimension of the lock-in-amplifier is 13mm x 10.65mm x 2.65mm. The maximum linear dimension of the exciting laser is 5.6 mm. The acoustic sensor (transducer) has a dimension of 42mm x 12mm. Furthermore, the measurement and analyses of the observed data uses multiple stochastic and machine learning techniques to bring out the best correlation fit between the glucose concentration and a specific feature of the electrical signal. With these methods and techniques, a strong correlation was confirmed between the glucose concentration and the amplitude of the electrical signal. The computed correlation coefficient between the signal amplitude and glucose concentration is 97% while the p-value is 5.6E-6. To the best of our knowledge, this is the first work to report photoacoustic spectroscopy for glucose concentration measurement in a *compact* form, with lock-in amplifier and aided with machine learning algorithms for future application as a *wearable* device.

INDEX TERMS Photoacoustic NIR spectroscopy, non-invasive glucose monitoring, lock-in-amplifier, machine learning.

I. INTRODUCTION

The current standard of successful management and effective treatment of diabetes relies on using an invasive finger pricking approach. Besides pain, discomfort and cost, the finger pricking protocol is prone to cause infection and produces potentially unsafe bio-waste in the form of testing strips and lancets. More importantly, drawing and testing blood requires pre-emptive action on the part of the patient or caretaker. For patients with complex medical conditions and aggressive forms of type-2 diabetes, continuous monitoring is desired so that unexpected changes in blood glucose levels, known to cause death due to sudden hypoglycemia during sleep, can be detected. Very recently, multiple commercial devices on continuous glucose monitoring have come up but none of them are non-invasive and wearable and need to be periodically implanted subcutaneously [1]–[4]. Besides the painful insertion and excessive cost (>∼$500), discomfort and the

The associate editor coordinating the review of this manuscript and approving it for publication was Sotirios Goudos.

risk due to the presence of a foreign body under the skin always remains active. Research findings thus far suggest that a workable non-invasive solution is feasible and multiple techniques using optical, electromagnetic, photoacoustic are being pursued to build a glucose monitor that is accurate, wearable as an accessory, low-cost (<$100) and continuous (every 1-minute max). However, they falter in one or many of these factors: size, portability, accuracy, reliability, cost, limited sensitivity below 70 mg/dl, operational complexity, and more [5]–[15].

In this paper, an *in-vitro* compact instrumentation for glucose measurement using Near Infra-Red (NIR) photoacoustic spectroscopy (PAS) has been reported to mitigate some of these above issues in *in-vivo* non-invasive blood glucose monitoring. The clear need of a non-invasive blood glucose monitoring device is what motivates the current work. The phase sensitive detection of the experimental data enhances accuracy, reliability and safety of the glucose measurement system by increasing the SNR multifold, an inherent quality attributed to phase sensitive detection. The phase sensitive detection uses a modulator/demodulator integrated circuit (IC) configured as a lock-in amplifier (LIA). It is important to assert that no earlier study on PAS uses LIA in this small form factor of an integrated circuit. The small form factor of the proposed experimental method with the use of an LIA IC is the basis for design of an accurate wearable *in-vivo* non-invasive blood glucose monitoring device using photoacoustic spectroscopy in the near future. Moreover, the data has been further measured and analyzed by a multitude of stochastic and machine learning techniques to determine the signal feature that has the highest correlation with the glucose concentration. This systematic, comprehensive measurement approach using a multitude of stochastic and machine learning techniques has not been not reported in the past for similar experimental studies to the best of our knowledge. The photoacoustic experimental technique is depicted in Figure 1.

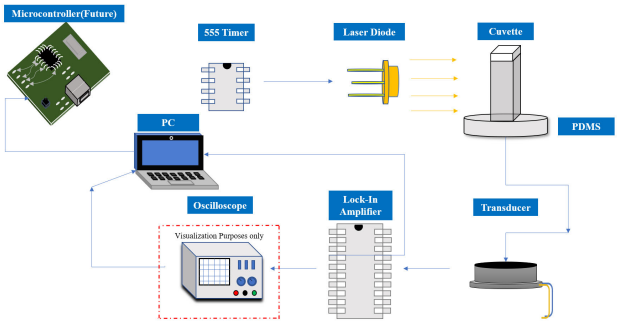


FIGURE 1. **The proposed compact instrumentation for the *in-vitro* glucose measurement with phase sensitive detection using PAS. All the instruments are portable and can be replicated for wearable *in-vivo* glucose monitoring.**

II. RELATED WORK

Almost all prior works in PAS have used benchtop measurement instruments and/or the expensive mid-IR spectroscopic approach to determine the glucose concentration such as the one reported by Zhang [16]. In [5], the signal detection uses a wideband low-noise amplifier (LNA), which can be susceptible to noise, unlike the phase sensitive lock-in amplifier IC used in this paper. Very recently, multiple stochastic and machine learning techniques have been reported for NIR optical spectroscopy [17]–[20] but none of them was extended to the photoacoustic approach. The machine learning techniques reported in [17] mentions using complex algorithms such as Random Forest Regression, Extra Trees Regression, Support Vector Machine (SVM) Regression, etc. Similarly, convolutional neural network (CNN) was used for NIR signal classification for different glucose concentration in the range of 50 – 430mg/dl as reported in [18]. NIR spectroscopy was used in combination with SVM to study the changes in water absorption for the early diagnosis of diabetes [20]. When compared to these reported articles, ordinary linear least squares (LLS) regression technique has been successfully used in this paper to model the relationship between the glucose concentrations and the voltage outputs of the designed circuit. The simple LS model, was chosen to be appropriate in this case due to the data collected from the experiments were suitable for such simple modelling. Also, the validation is modelling has been made robust using standard train / test splits made the associated model more robust. Additionally, the usage of simple machine learning technique such as LS enables the future embedding of such models within a wearable device due to its low complexity (hence low computations costs, thereby resulting in low power requirements). Further discussions on related work are continued in Section VII by including numerical data of the physical dimensions and experimental results from stochastic analysis in Table 8.

III. FUNDAMENTALS OF PAS AND LIA

The premise of the current study is primarily based upon the applications of PAS for detection of glucose concentration and LIA for high accuracy electrical signal detection. The following two subsections elaborate on the fundamental theory and properties of these two applications.

A. PAS AND DETECTION OF GLUCOSE CONCENTRATION

PAS is a hybrid technique which can be applied to determine the concentration of glucose in a solution. Unlike NIR optical spectroscopy, the NIR PAS is an ultra-sensitive method that can be used to study weak bulk and surface absorption in liquids, to evaluate the level of absorbed energy. This approach uses optical excitation and mechanical detection where the optical energy is converted into an acoustic energy by a multistage energy conversion process [21]. A laser pulsed at a high frequency is allowed to fall on the glucose solution for the optical excitation of the glucose molecules. The absorbed optical energy, determined by the optical absorption coefficient, from the laser leads to the localized heating of the solution, which produces a small temperature rise, resulting in the volumetric thermal expansion of the optical interaction region. The associated ultrasonic pressure pulse is

then measured by an ultrasonic piezoelectric transducer. The optical absorption coefficient has a strong correlation with the level of glucose concentration at a wavelength of 980 nm, where glucose molecules have strong overtones [21].

B. LOCK-IN AMPLIFIER

LIAs require a frequency reference and use the technique of phase-sensitive detection (PSD) to single out the component of the signal at that specific reference frequency and phase. The PSD can detect the signal with an extremely narrow bandwidth (< 1 Hz) which increases the overall Signal to Noise Ratio (SNR) of the incoming signal. This makes possible to detect a signal with very high accuracy as investigated in this paper. An input signal of $x_{sig} = V_{sig}sin(\omega_{sig}t + \theta_{sig})$ and a reference signal of $x_{ref} = V_{ref}\,sin(\omega_{ref}t + \theta_{ref})$ has been assumed, where the symbols have their usual meanings. The LIA amplifies the signal and then multiplies it by the lock-in reference using a modulator circuit. The resulted PSD signal is the product of the above two signals and is given by:

$$x_{psd} = V_{sig}V_{ref}\,sin(\omega_{sig}t + \theta_{sig})sin(\omega_{ref}t + \theta_{ref}) \quad (1)$$

Using trigonometric identities eqn. (1) can be written as,

$$x_{psd} = \frac{1}{2}V_{sig}V_{ref}\,cos([\omega_{sig}t - \omega_{ref}t] + \theta_{sig} - \theta_{ref}) - \frac{1}{2}V_{sig}V_{ref}\,cos([\omega_{sig}t + \omega_{ref}t] + \theta_{sig} + \theta_{ref}) \quad (2)$$

After choosing the reference frequency equal to the signal frequency and low pass filtering, the retrieved PSD signal from eqn. (2) becomes

$$x_{psd} = \frac{1}{2}V_{sig}V_{ref}\,cos(\theta_{sig} - \theta_{ref}) \quad (3)$$

The resultant signal from eqn. (3) is observed for the detection of glucose level.

IV. PROPOSED INSTRUMENTATION AND DETECTION METHOD

The setup of these components is explained in detail in the following sub-sections.

A. IN-VITRO EXPERIMENTAL THEORY AND METHOD

In this work (See Fig. 1), the laser diode at a wavelength of 980 nm is pulsed at 1 MHz using a timer IC. The laser beam is then directed into a glucose solution contained within a quartz cuvette from a distance of ∼4 mm. The cuvette sits on a Polydimethylsiloxane (PDMS) layer, which itself sits on an ultrasonic transducer. The PDMS layer acts as a transfer medium for the acoustic waves within the glucose solution to pass through to the transducer from the cuvette. The ultrasonic transducer then converts the acoustic waves into electrical signals, which are then sent to the LIA IC for signal detection with minimized interference from uncorrelated noise sources. The analog signals obtained from the output of the LIA IC are then observed on a digital oscilloscope after sampling. The digital data are then imported from the oscilloscope to a computer (microcontroller unit) for digital signal processing and stochastic data analyses. It is important to re-assert that, the analog signal detecting instrument in this experimental set-up uses a LIA in the form of an IC, replacing the traditional desk top based LIA that is primarily used for astronomical computations thus making it suitable for potential wearable applications of non-invasive $in-vivo$ blood glucose detection.

B. ELECTRONIC CIRCUIT DESIGN

The schematic of the circuit used to excite the glucose solution as well as detect the amplitude of the electrical signal for different glucose concentrations is shown in Fig. 2. A timer IC is configured to produce optical pulses via a laser diode at a frequency of 1 MHz as the transducer used for the experiment has a peak resonance at this frequency. This is an astable configuration where the output voltage alternates between VCC and 0 volts continuously. Since the timer equations do not accurately model the duty cycle and frequency observed experimentally, values for R_a, R_b, C_1, and C_2 were determined through trial and error with variable resistors and different capacitors. The final combination of resistance and capacitance that yields the desired results can be seen in Table 1. The matching of the pulsed frequency of the timer with the peak observed frequency of the transducer ensured a strong signal was observed.

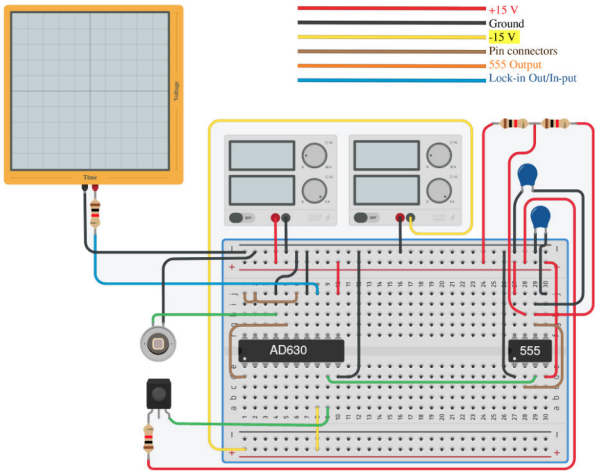


FIGURE 2. The electronics circuit for the optical stimulation and the phase sensitive detection of the proposed compact instrumentation system. The LIA IC is shown as AD630 in the diagram. (Tinker Cad Diagram).

TABLE 1. Resistor and Capacitor Values for CMOS Timer.

Symbol	Experimental Values
R_a	76.5 Ω
R_b	235 Ω
C_1	1 nF
C_2	1 nF

The laser diode has a NIR wavelength of 980 nm and a power output of 10 mW. The pulsed beam is directed at a

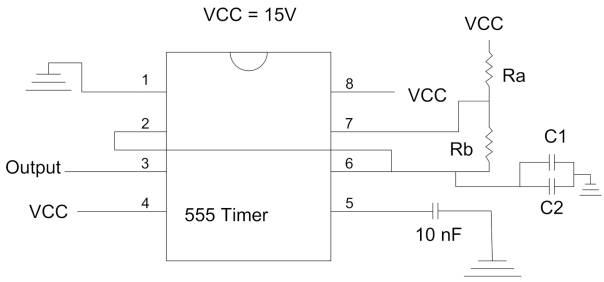
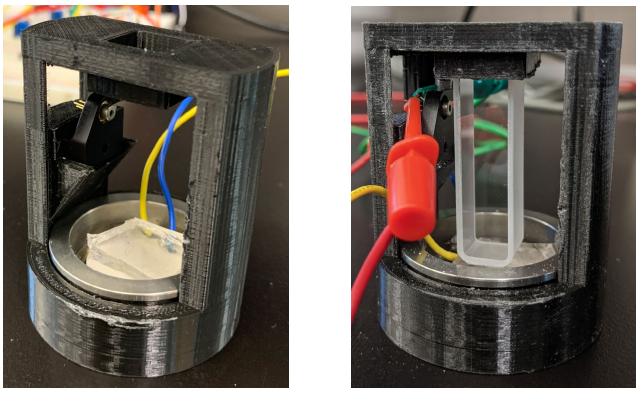


FIGURE 3. LMC555 CMOS Timer Setup.

TABLE 2. Sample Concentration.

Trials per Concentration	Sugar Concentrations (mg/dL)
15	0, 60, 100, 140, 180, 220, 260, 300



(a) Holder w/o cuvette **(b)** Holder w/Cuvette

FIGURE 4. Holder in different configurations.

quartz cuvette containing glucose solutions of various concentrations. Mechanical vibrations induced by photoacoustic effects are transformed into electrical signals through a transducer, which is then detected by a modulator/demodulator integrated circuit that is configured to work as a lock-in amplifier for phase sensitive detection. The lock-in-amplifier takes in a reference signal directly from the timer at 1 MHz and correlates with the input signal received from the transducer at the same frequency for the phase sensitive detection. To reiterate, the lock-in amplifier which improves the signal to noise ratio multifold using this phase sensitive detection approach has not been used before for such photoacoustic signal detection in the form factor of an IC. The physical dimensions of the exciting laser, transducer and lock-in amplifier are specified in Table 8.

C. PREPARATION OF GLUCOSE SOLUTION

A fixture is designed, and 3D printed to hold the laser diode near the quartz cuvette as depicted in Fig. 4. The cuvette is not in place in Fig 4a to give an unobscured view of the PDMS. The cuvette can be seen in place in Fig. 4b. Variable samples of glucose (dextrose) solutions are created by mixing deionized (DI) water and powdered glucose to prepare concentration ranging from 0 mg/dL to 300 mg/dL. The DI water was measured to 2 dL and heated on a hotplate to 50°C. Glucose was added in increments to achieve the specific concentrations as seen in Table 2. A total of 15 trials were conducted for each concentration for greater reliability.

A magnetic stirring rod was used to mix the solution at 360 rpm. The magnetic stirring rod was used in conjunction with the hotplate to allow for continuous mixing without direct user intervention. Glucose solutions are stored inside quartz cuvettes which sit above a 5 mm PDMS transfer layer. After all voltages are recorded for each specific concentration, the solution in the 4mL cuvette was added back to the main solution and allowed to remix to keep future concentrations accurate.

D. INSTRUMENTS DESCRIPTION

This procedure utilizes many different instruments to acquire pertinent data. Each instrument is defined by its specific application and some characteristics below.

- **SIGLENT SPD3303X-E programmable direct current(DC) power supply**
  - Power supply for Timer, LNA, and Laser Diode
  - Key Specifications
    - 100/120/220/230 V compatible design to meet the needs of different power grids
    - Intelligent temperature-controlled fan effectively reduces fan noise
- **LMC555 CMOS Timer**
  - Sends periodic current to LP980P010 laser diode using a clock cycle
  - Key Specifications
    - Industry's Fastest Astable Frequency of 3 MHz
    - Output Fully Compatible With TTL and CMOS Logic at 5 V Supply
- **LMF56 diode mount**
  - Post-Mountable Laser Diode Mount for Ø5.6 mm Packages, 8-32 Tap
- **L980PO10 Laser Diode**
  - Sends a pulsating beam into a concentrated glucose solution made of DI water and glucose sugar
  - Key Specifications
    - Wavelength 980 nm
    - Output Power 0.01 W
    - Operating Voltage 1.5 to 2.2 V
    - Operating Current 0.025 to 0.04 A
- **4mL LAB4US standard quartz cuvette**
  - Solution container
  - Key Specifications
    - Spectral Range: 190-2500 nm
    - Volume: 3.5 mL

- **PDMS Transfer Layer**
  - Transfer medium for acoustic waves
  - Key Specifications
    - Cured from SYLCAP 284-F silicone elastomer kit
- **Steiner and Martins, Inc. ultrasonic transducer**
  - Converts acoustic signals into electrical signals
  - Key Specifications
    - 1 MHz resonant frequency
    - Resonant impedance Zm: $\leq 20\ \Omega$
- **AD630 Modulator/Demodulator configured as Lock-in Amplifier** [22]
  - Improves SNR of output signal
    - 2 MHz channel bandwidth
    - 2 differential input stages
    - Can recover signal from 100 dB noise
    - 100 μV channel offset
- **RIGOL DS1054 UltraVision oscilloscope**
  - Signal analysis tool to visualize output signal
  - Key Specifications
    - Up to 30,000 wfms/s Waveform capture rate
    - 4 Channels, 50 MHz Bandwidth

Other equipment used includes a hot plate, magnetic stirring rod, 200 mL beaker, 10 mL beaker, sonicator, desiccation chamber, fume hood, and an enclosed weigh-scale. Variable resistors, capacitors, and numerous wires are also used to connect and configure the electronic circuit used in this procedure. The electronic circuit includes a modulator/demodulator IC as a LIA for signal detection brings a major innovation to this work.

V. EXPERIMENTAL RESULTS

The electrical data captured from the output of the LIA IC is observed in an oscilloscope as depicted in Fig. 5. The sampled data obtained from the oscilloscope are recorded for stochastic analysis. Each observed wavelet corresponds to a single laser pulse. The time axis below spans six microseconds, during which time the cuvette experienced six laser pulses.

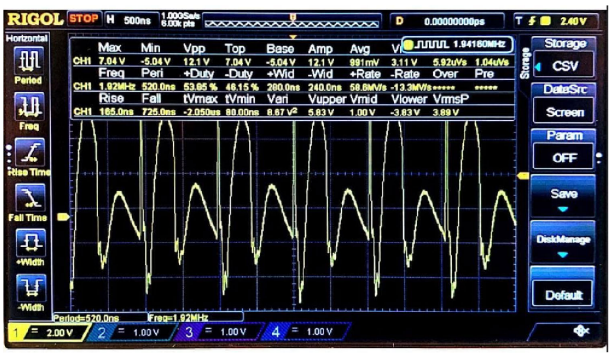


FIGURE 5. The electrical signal from the LIA output as observed in the Oscilloscope.

To bring out the best correlation efficiency between the signal and the glucose concentration, five different attributes of the signal were initially considered. These five attributes of the signal are the Amplitude (Max), Root Mean Square (RMS), Minimum (Min), Range and DC Average (Mean). The observed results of these five attributes are recorded in Table 3. It is important to note that each measurement of the signal attributes at a specific concentration is an aggregate of 15 trials. The consideration of multiple signal attributes and repetition of the experiment multiple times is to validate the accuracy and reliability of the proposed instrumentation system.

TABLE 3. Observed voltage at the oscilloscope of the five different signal attributes (Each measurement of the signal attributes at a specific concentration is an aggregate of 15 trials).

Concentration (mg/dL)	Max(V)	RMS(V)	Min(V)	Range(V)	Mean(V)
0	7.628	3.713	-3.880	11.508	1.227
60	8.018	3.986	-3.872	11.890	1.479
100	8.174	4.032	-4.000	12.174	1.451
140	8.583	4.238	-3.971	12.554	1.561
180	8.508	4.117	-4.056	12.564	1.440
220	8.964	4.438	-4.073	13.037	1.683
260	9.122	4.503	-3.988	13.110	1.703
300	9.256	4.547	-4.080	13.336	1.737

VI. DATA MEASUREMENT AND STOCHASTIC ANALYSIS

The features of the observed experimental data were extracted in the beginning. After that multiple models using ordinary linear-least squares(OLS) regression were fitted for analysis. Firstly, the OLS regression method was used to determine any preliminary correlations between glucose concentration and the output voltage. Secondly, the data was separated into training and testing sets, to measure unbiased correlations and test the trained model. These separate datasets, commonly used in the field of machine learning, were used to check the robustness of the trained regression model generated. Thirdly, a p-value test is conducted to determine probability of null hypotheses occurring within the scope of the experiment. With these methods and techniques, a strong correlation was confirmed between the glucose concentration and the signal amplitude.

A. FEATURE EXTRACTION

As stated above, 15 pre-processed data points were recorded for each concentration measured. The output voltage deviation over the set of 15 trials at 0 mg/dL is illustrated in Fig. 6. The red line in the shaded area represents the average voltage over 15 trials whereas the shaded area represents the voltage spread for all trials. The voltage of 'Min' remained relatively constant with changes in solution concentration and is believed to represent the undisturbed baseline signal output of the system and hence it is not considered for further analysis. Since 'Min' does not show a trend, 'Range' is not considered for further analysis as it also depends upon 'Min'.

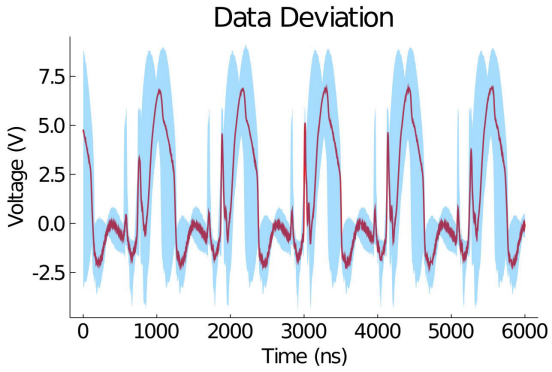


FIGURE 6. Deviation in the signal voltage.

TABLE 4. Standard deviation of the extracted signal attributes.

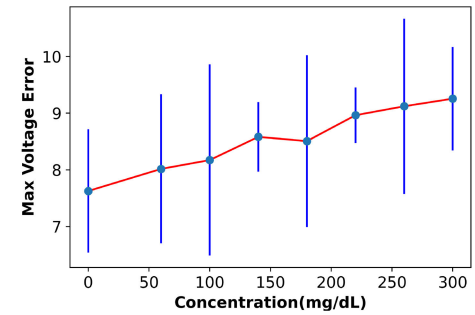
Concentration (mg/dL)	Max(V)	Mean(V)	RMS(V)
0	1.06	0.23	0.45
60	1.27	0.26	0.50
100	1.62	0.34	0.67
140	0.61	0.17	0.29
180	1.46	0.28	0.58
220	0.49	0.15	0.25
260	1.53	0.33	0.66
300	0.82	0.18	0.36

Therefore, only the three signal attributes: Max, Mean and RMS voltages of the wavelet were extracted from the signal for each data-set and considered for further stochastic studies. Table 4 shows the standard deviations of these attributes for each concentration obtained from the mean of 15 trials. The highest standard deviations are obtained for the 'Max' for all concentrations while 'Mean' and 'RMS' are comparatively less. Alongside standard deviations, error bar plots are developed as seen in Fig. 7. Although the standard deviation for 'Max' was little too high, considering the observations from the error bar plots, 'Max' was considered for the next level of stochastic analysis alongside 'Mean' and 'RMS'.

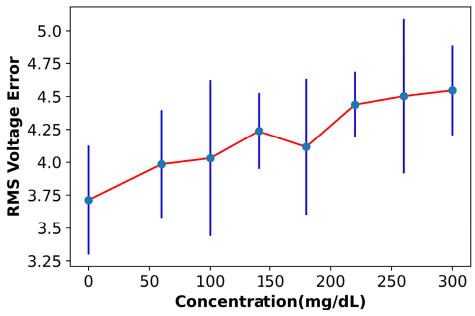
B. LEAST SQUARES REGRESSION

The processed dataset used for the linear regression model is illustrated in Fig. 8. The OLS regression was implemented using scikit-learn module in Python [23] to fit the training data to create the model. Then the concentrations from the test data was predicted using the fitted (trained) model. The regression model considered concentration level as the independent variable (x-axis), and the five signal attributes calculated from the raw voltage output as the dependent variables (y-axis).

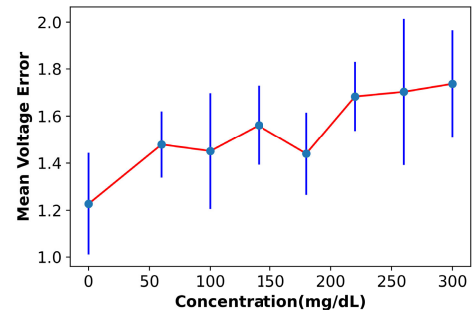
Table 5 provides the correlation coefficients of the ordinary least-squared models utilizing three different signal attributes: 'Max', 'RMS' & 'Mean'. These correlations were computed using the function *r2_score* built within scikit-learn module. This is given in equation 2, where y_i represents the processed y values and ŷ represents the predicted y-value.



(a) Max Error Bar Plot



(b) RMS Error Bar Plot



(c) Mean Error Bar Plot

FIGURE 7. Error Bar Plots for the three signal attributes: (a) Max, (b) RMS & (c) Mean.

TABLE 5. Correlation Coefficients.

Signal Attributes	Correlation Coefficient
Max	97%
RMS	93%
Mean	83%

Equation 3 gives the least square regression.

$$\bar{y} = \frac{1}{n}\Sigma_{i=1}^{n} y_i \tag{4}$$

$$\Sigma_{i=1}^{n}(y - \hat{y}_i)^2 = \Sigma_{i=1}^{n}\epsilon_i^2 \tag{5}$$

$$R^2(y, \hat{y}) = 1 - \frac{\Sigma_{i=1}^{n}(y - \hat{y})^2}{\Sigma_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

The red dots show the processed measurements recorded while the blue line represents the best-fit according to the least-square regression models previously established.

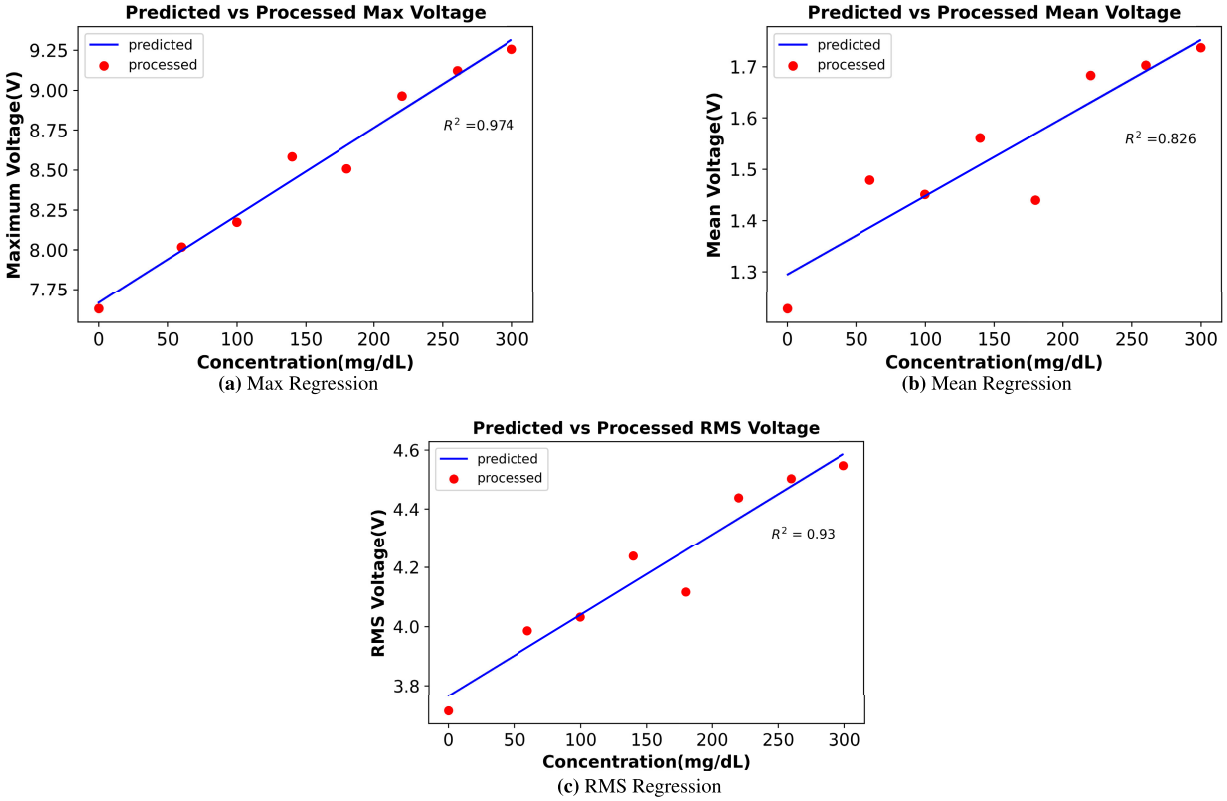


FIGURE 8. Linear regression calculations using three signal attributes: (a) Max, (b) Mean & (c) RMS.

TABLE 6. Correlation between the concentration and the independent variables extracted from the signals.

Signal Attributes	Unbiased R^2
Max	97.0%
RMS	85.9%

TABLE 7. p-value.

Signal Attributes	p-value
Max	5.59E-6
RMS	1.09E-04

Interestingly, the correlation coefficient for the 'Max' was found to be highest at around 97% followed by 'RMS' at 93% and 'Mean' at 87%. Hence, the signal amplitude is identified to be the best candidate among the three extracted signal features from this study. After conducting linear regression tests, the processed data-set is separated into train and test sets to develop a second model to support the results obtained from this study of least-squares regression model. In this approach, the 'Mean' performed poorly and not taken into consideration for further analysis.

C. TRAIN/TEST SPLIT OF CORRELATIONS

The scikit-learn module had an inbuilt function to split the data into training and testing, which was used to separate the data into a 80:20 split (i.e. 80% training and 20% testing sets). The majority of the available data was used for training and a smaller portion of the data is used for testing [24]. Table 6 shows the un-biased coefficients where the model is tested against the test data.

The coefficients for the Max and RMS demonstrate a strong linear correlation with concentration. As can be observed from Table 6, the 'Max' gives the strongest correlation, conforming to the least squares regression study.

D. CALCULATED PROBABILITY

In addition to linear regression and train and split modeling, p-value statistics were used to estimate whether the correlations held true to an asserted hypothesis as it is an important indicator of whether the correlations between two variables are due to chance or whether there is enough evidence to disprove the opposing side. In our case, p-value statistics are based upon hypothesis testing of the experiments [25], within which, the null hypothesis *'no correlation exists between any of the four signal attributes and varying glucose concentration'* is tested against the alternate hypothesis. The alternate hypothesis states that there is some form of correlation between varying glucose concentration and the four signal attributes. Noting the threshold of significance of 0.05 [26], the p-values seen in Table 7 provide support for the alternative hypothesis as they both fall well below this threshold value, with 'Max' falling by four orders of magnitude. Hence, 'Max' or the signal amplitude was confirmed to be the chosen feature of the signal for obtaining the glucose concentration.

E. EXTENDED STOCHASTIC ANALYSIS

In order to remove any form of algorithmic bias, four additional regression algorithms were implemented to check the correlation between the glucose concentration (treated as the independent variable) and the three extracted features (Max, RMS and Mean - treated as the target variable). These additional methods are:

Support Vector Regression(SVR) - uses the same principles as the SVM for classification, and introduces the constraint for the maximum error allowed in the predicted value of each of the training data point not to exceed ϵ [30]–[32].

Support Vector Regression problem formulation is often best derived from a geometrical perspective, the continuous-valued function being approximated can be written as seen in equation 7.

$$\begin{aligned} y = f(x) &= < w, x > + b \\ &= \sum_{j=1}^{M} w_j x_j; b \in \mathbb{R}; x, w \in \mathbb{R}^M \end{aligned} \tag{7}$$

For multidimensional input data, the following multivariate equation 8 is used.

$$f(x) = \begin{pmatrix} w \\ b \end{pmatrix}^T \begin{pmatrix} x \\ 1 \end{pmatrix} = w^T x + b; x, w \in \mathbb{R}^{M+1} \tag{8}$$

The objective is to minimize the error between the predicted value and the actual output for a given input. SVR adopts an ϵ-insensitive loss function, which penalizes predictions for being farther than ϵ from the ideal output.

K-nearest neighbor regression – similar to K-NN for classification, the predicted values are based on the average of the neighbors. In this case, the number of neighbors chosen were two as default [33], [34].

K-NN regression uses the same distance functions as K-NN classification, that is Euclidean, Manhattan or Minkowski for continuous variables. In our case, Euclidean distance was chosen as default.

Least Absolute Shrinkage and Selection Operator (LASSO) Regression – this is a type of linear regression where the data values shrink to the center or mean to avoid overfitting the data [35].

In Lasso regression, a L1 regularization which introduces additional information to prevent overfitting is employed. Consequently, a model containing all possible predictors can be fitted and lasso can be used to perform variable selection by regularizing the coefficient estimates (shrinks the towards zero). The objective is to minimize both the residual sum of squares (RSS) and the sum of the absolute value of coefficients. This minimization is shown below in equation 9.

$$\begin{aligned} &\sum_{i=1}^{n} (y_i - (\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}))^2 + \alpha \sum_{j=1}^{p} |\beta_j| \\ &= RSS + \alpha \sum_{j=1}^{p} |\beta_j \end{aligned} \tag{9}$$

Here, n represents the number of observations, p denotes the number of variables that are available in the dataset. x_{ij} represents the value of the j^{th} variable for the i^{th} observation, where i = 1,2,3…n, and j = 1,2,3…p. α takes a range of values between 0 to ∞, with an increase in its value reducing the flexibility of the lasso regression fit. This leads to decreased variance but increased bias.

Bayesian Ridge Regression – Bayesian Ridge Regression (BRR), also known as Tikhonov Regularization, is a classical regularization technique and is essentially an OLS linear regression method but with a tuneable additive L2 norm penalty term embedded within the function. In BRR, the output y is assumed to be drawn from a probability distribution rather than estimated as a single value such as in OLS. This involves estimating a probabilistic model of the regression, where the coefficient weights are near to zero for stability of the model [36]. Mathematically, y is assumed to be normally distributed around as seen in equation 10.

$$p(y|X, w, \alpha = N(y|Xw, \alpha) \tag{10}$$

The prior for the coefficient is given by the Gaussian relation seen in equation 11.

$$p(w|\alpha) = Nw|0, \alpha^{-1} I_p \tag{11}$$

Here, $\alpha \geq 0$ is the complexity parameter which controls the amount of shrinkage affecting the coefficient's robustness to collinearity. The parameters w,α, λ are estimated jointly while fitting the model, and α, λ (the regularization parameters) are estimated by maximizing the log marginal likelihood [37].

Along with these four additional regression algorithms, the test samples were varied between one and four. That is, the train-test split was increased in steps to 40:60, thereby reducing the samples used for training. This was done to check the suitability of the models in terms of less training samples, their robustness to bias and an understanding of which method / algorithm can be taken up as a future work for building the prototype.

These results in terms of accuracy, as shown in Figure 9, was calculated by comparing the actual vs. predicted value of the target variable (i.e. Max, RMS or Mean). The total accuracy was summed over all the number of test samples. That is, the results report how many test samples were correctly predicted out of the total test samples provided. Thus, as an example – Lasso could only predict 2/3 test samples correctly using RMS as target variable, Bayesian Ridge could predict 3/3 test samples correctly when using Max as target variable and so on. From all of these subfigures, it can be concluded that K-NN, Lasso and Bayesian Ridge regression shows a robust performance with 100% accuracy for 'Max' as a predicted variable for all test samples. This extended analysis thus validates further 'Max' to be the chosen feature of the signal for obtaining the glucose concentration.

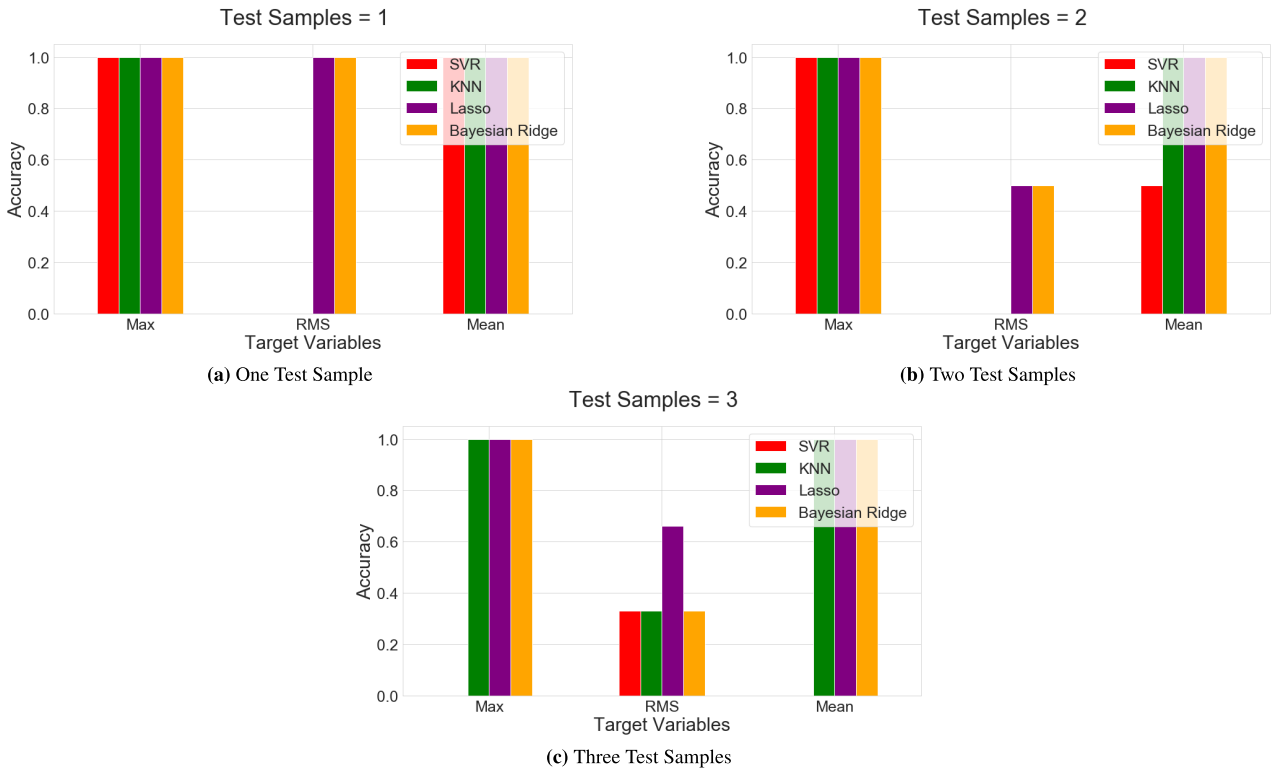


FIGURE 9. Additional regression algorithms using (a) One Test Sample, (b) Two test samples, (c) Three test samples. Additional regression algorithms using (a) One Test Sample, (b) Two test samples, (c) Three test samples. The accuracy on the y-axis shows how many test samples were correctly predicted out of the total given test samples.

TABLE 8. Comparison of Glucose Concentration Detection Methods using NIR Photoacoustic and Optical Spectroscopy Techniques.

Publication	Glucose Detection Method	Laser Size(mm)	Laser Wavelength(nm)	Amplifier Type	Amplifier Size(mm)	ML algorithm
[16]	PAS	Tabletop	1600	USPA[14] (Olympus 5662)	127 x 76 x 51 (LWH)	LS Regression(96.68%)
[5]	PAS	Matchbook	850/1500	LNA (Model351A)	79.3 x 38.1 x 27.7 (LWH)	LS Regression
[17]	NIR-Op	182 x 110 x 47	900-2200	N/A	N/A	PLSR[1],SVMR[5](98.5%),RF[3] ETRXG[3],PCA-NN[4](98.5%)
[27]	NIR-Op	50 x 40 x 20	485,645,940	N/A	N/A	MLR (91%)[5] PLSR[1],PCR[6] FFNR[7] DNN[8] SVM[5](99%),RF[2]
[28]	PAS	Unspecified	1382/1610	LIA (SR844)	495.3 x 432 x 133.4 (LWH)	N/A
[29]	PAS	H:110 x 65 x 52 C:190 x 130 x 40	MID-IR	LIA (SR810)	495.3 x 432 x 133.4 (LWH)	PCA[9], PLSR[1],LOOCV[10]
This work	PAS	Bd[11]:5.6, Cd[12]: 4.70 Ch[13]:4.83	980	LIA (AD630)	13 x 10.65 x 2.65 (LWH)	LS[2] Regression(97.4%) Train Test(97%) SVR, KNN LASSO, BRR

Partial Least Squares Regression[1], Random Forest[2], Extreme Gradient Boosting[3], Principle Component Analysis Neural Networks[4], Multiple Linear Regression[5], Principle Component Regression[6], Feed-Forward Neural Network[7], Deep Neural Network[8], Principle Component Analysis[9], Leave-One-Out Cross-Validation[10], Base Diameter[11], Cone Diameter[12], Cap height[13], Ultrasonic Preamplifier[14]

VII. COMPARATIVE ANALYSIS

To bring out the advantage and better understand the compactness of the proposed experimental method, Table 8 is included for comparative analysis. As can be seen from the table, the dimensions of the experimental components used in this work are significantly smaller than similar works earlier reported. The form factor of this work makes the experimental approach applicable to design a *wearable* blood glucose measurement device in the future. Two recent works on PAS [28], [29] used LIA, albeit with bench-top experimental instruments thus making it not viable for wearable applications. Two other recent works on PAS [5], [16] neither

use an LIA nor the dimensions are of smaller scale to that of the wearable form factor. Out of these, two have used LS regression only, while a third one did not use any for data measurement analysis. The stochastic coefficient of 97% reported in this work is comparable to earlier works. The physical dimensions of the laser is also significantly smaller in this work and of the wearable form factor. The comparison of the laser dimension include two very recent works on NIR optical spectroscopy (NIR-Op) [17], [27] as well. The physical dimension of 42mm (diameter) x 12mm (height) of the acoustic sensor (transducer) also conforms to the wearable form factor.

VIII. CONCLUSION AND FUTURE WORK

An *in-vitro* glucose measurement system is developed using PAS NIR spectroscopy. The detection mechanism of the electrical signal uses the phase sensitive LIA in the form-factor of an integrated circuit to increase the SNR for better accuracy and reliability. No prior work on PAS, has taken advantage of this detection methodology in such a small form factor, thus making it a potential candidate for an *in-vivo* accurate non-invasive wearable glucose monitoring device. The dimension of the lock-in-amplifier and the maximum linear dimension of the laser are 13mm x 10.65mm x 2.65mm and 5.6 mm respectively. The dimension of 42mm x 12mm of the transducer also conforms to the wearable form factor.

The measurement and data analyses use a multitude of stochastic and machine-learning techniques to determine the signal amplitude ('Max') as the signal feature that has the highest correlation with the glucose concentration. The measured correlation coefficient between the signal amplitude and glucose concentration is 97% while the p-value is 5.6E-6. To the best of our knowledge, this is the first work to report photoacoustic spectroscopy for glucose concentration measurement in a *compact* form, with lock-in amplifier and aided with machine learning algorithms for future application as a *wearable* device.

To develop the proposed instrumentation into an *in-vivo* wearable biomedical device for continuous glucose monitoring addressing all the safety protocols and medical standardization rules, the use of light-emitting-diodes (LED) replacing lasers are currently being investigated. Besides safety, to make the device energy efficient, affordable, and wearable, low power microcontrollers are considered for data processing and analyses. The design will be similar to a wristwatch and will provide enough space for the components used in this procedure to integrate as a wearable biomedical device. The laser diode, the crystal cuvette containing glucose solution, the computer are to be substituted with LED, user's wrist/fingertip and a microcontroller respectively.

CODE, DATA AND MATERIALS

Datasets and code used and analyzed for this work can be made available from the corresponding author upon request.

REFERENCES

[1] (2022). *The Guardian Connect System*. [Online]. Available: https://www.medtronicdiabetes.com/products/guardian-connect-continuous-glucose-monitoring-system
[2] (2019). *Glucotrack|Your Track to Health!*. [Online]. Available: http://www.glucotrack.com/about-glucotrack
[3] (2021). *Dexcom Continuous Glucose Monitoring*. [Online]. Available: https://www.dexcom.com/g6-cgm-system
[4] (2021). *Freestyle Libre 14 Day System*. [Online]. Available: https://www.freestylelibre.us/system-overview/freestyle-14-day.html
[5] P. P. Pai, P. K. Sanki, A. De, and S. Banerjee, "NIR photoacoustic spectroscopy for non-invasive glucose measurement," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 7978–7981.
[6] S. Saha, H. Cano-Garcia, I. Sotiriou, O. Lipscombe, I. Gouzouasis, M. Koutsoupidou, G. Palikaras, R. Mackenzie, T. Reeve, P. Kosmas, and E. Kallos, "A glucose sensing system based on transmission measurements at millimetre waves using micro strip patch antennas," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, Dec. 2017.
[7] P. H. Siegel, Y. Lee, and V. Pikov, "Millimeter-wave non-invasive monitoring of glucose in anesthetized rats," in *Proc. 39th Int. Conf. Infr., Millim., THz Waves (IRMMW-THz)*, Sep. 2014, pp. 1–2.
[8] P. H. Siegel, A. Tang, G. Virbila, Y. Kim, M. C. F. Chang, and V. Pikov, "Compact non-invasive millimeter-wave glucose sensor," in *Proc. 40th Int. Conf. Infr., Millim., THz Waves (IRMMW-THz)*, Aug. 2015, pp. 1–3.
[9] M. J. Tierney, J. A. Tamada, R. O. Potts, L. Jovanovic, and S. Garg, "Clinical evaluation of the GlucoWatch® biographer: A continual, non-invasive glucose monitor for patients with diabetes," *Biosensors Bioelectron.*, vol. 16, nos. 9–12, pp. 621–629, Dec. 2001.
[10] S. Laha, S. Kaya, N. Dhinagar, Y. Kelestemur, and V. Puri, "A compact continuous non-invasive glucose monitoring system with phase-sensitive front end," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2018, pp. 1–4.
[11] (2014). *Introducing Our Smart Contact Lens Project*. [Online]. Available: https://blog.google/alphabet/introducing-our-smart-contact-lens/
[12] S. Haxha and J. Jhoja, "Optical based noninvasive glucose monitoring sensor prototype," *IEEE Photon. J.*, vol. 8, no. 6, pp. 1–11, Dec. 2016.
[13] L. W. Ishle, "Non-invasive blood glucose monitoring system," U.S. Patent 6 949 070, 2005.
[14] J. Y. Sim, C.-G. Ahn, E.-J. Jeong, and B. K. Kim, "*In vivo* microscopic photoacoustic spectroscopy for non-invasive glucose monitoring invulnerable to skin secretion products," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, Dec. 2018.
[15] H. M. E. Ciurczak and K. Bynum, "Near infrared blood glucose monitoring system," U.S. Patent 6 675 030, 2004.
[16] R. Zhang, F. Gao, X. Feng, S. Liu, R. Kishor, Y. Luo, and Y. Zheng, "Noninvasive photoacoustic measurement of glucose by data fusion," *Analyst*, vol. 142, no. 16, pp. 2892–2896, 2017.
[17] B. K. Mekonnen, W. Yang, T.-H. Hsieh, S.-K. Liaw, and F.-L. Yang, "Accurate prediction of glucose concentration and identification of major contributing features from hardly distinguishable near-infrared spectroscopy," *Biomed. Signal Process. Control*, vol. 59, May 2020, Art. no. 101923.
[18] B. K. Mekonnen, W. Yang, T.-H. Hsieh, S.-K. Liaw, and F.-L. Yang, "Classification of near-infrared spectroscopic glucose concentrations using convolutional neural network," in *Proc. Opto-Electron. Commun. Conf. (OECC)*, Oct. 2020, pp. 1–3.
[19] Q. Hao, J. Zhou, L. Zhou, L. Kang, T. Nan, Y. Yu, and L. Guo, "Prediction the contents of fructose, glucose, sucrose, fructo-oligosaccharides and iridoid glycosides in *Morindaofficinalis* radix using near-infrared spectroscopy," *Spectrochimica Acta A, Mol. Biomolecular Spectrosc.*, vol. 234, Jun. 2020, Art. no. 118275.
[20] Y. Li, L. Guo, L. Li, C. Yang, P. Guang, F. Huang, Z. Chen, L. Wang, and J. Hu, "Early diagnosis of type 2 diabetes based on near-infrared spectroscopy combined with machine learning and aquaphotomics," *Frontiers Chem.*, vol. 8, p. 1133, Dec. 2020.
[21] H. A. MacKenzie, H. S. Ashton, S. Spiers, Y. Shen, S. S. Freeborn, J. Hannigan, J. Lindberg, and P. Rae, "Advances in photoacoustic noninvasive glucose testing," *Clin. Chem.*, vol. 45, no. 9, pp. 1587–1595, 1999.
[22] *Balanced Modulator/Demodulator*, Analog Devices, Norwood, MA, USA, 2015.
[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 10, pp. 2825–2830, Jul. 2017.

[24] S. Theodoridis, A. Pikrakis, K. Koutroumbas, and D. Cavouras, *Introduction to Pattern Recognition: A MATLAB Approach*. New York, NY, USA: Academic, 2010.
[25] B. Dyson. (2014). *Everything you Know About the P-Value is Wrong*. [Online]. Available: https://www.tldrpharmacy.com/content/everything-you-know-about-the-p-value-is-wrong
[26] D. B. Panagiotakos, "The value of *p*-value in biomedical research," *Open Cardiovascular Med. J.*, vol. 2, p. 97, Nov. 2008.
[27] M. Shokrekhodaei, D. P. Cistola, R. C. Roberts, and S. Quinones, "Non-invasive glucose monitoring using optical sensor and machine learning techniques for diabetes applications," *IEEE Access*, vol. 9, pp. 73029–73045, 2021.
[28] Y. Tanaka, T. Tajima, M. Seyama, and K. Waki, "Differential continuous wave photoacoustic spectroscopy for non-invasive glucose monitoring," *IEEE Sensors J.*, vol. 20, no. 8, pp. 4453–4458, Apr. 2020.
[29] M. A. Pleitez, T. Lieblein, A. Bauer, O. Hertzberg, H. von Lilienfeld-Toal, and W. Mäntele, "*In vivo* noninvasive monitoring of glucose concentration in human epidermis by mid-infrared pulsed photoacoustic spectroscopy," *Anal. Chem.*, vol. 85, no. 2, pp. 1013–1020, Jan. 2013, doi: 10.1021/ac302841f.
[30] M. Awad and R. Khanna, *Support Vector Regression*. Berkeley, CA, USA: Apress, 2015, pp. 67–80, doi: 10.1007/978-1-4302-5990-9_4.
[31] R. McShinsky and B. Marshall, "Comparison of forecasting algorithms for type 1 diabetic glucose prediction on 30 and 60-minute prediction horizons," in *Proc. ECAI*, 2020, pp. 12–18.
[32] T. Hamdi, J. B. Ali, V. Di Costanzo, F. Fnaiech, E. Moreau, and J.-M. Ginoux, "Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm," *Biocybern. Biomed. Eng.*, vol. 38, no. 2, pp. 362–372, 2018, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0208521617302966
[33] S. Kohli, G. T. Godwin, and S. Urolagin, "Sales prediction using linear and KNN regression," in *Advances in Machine Learning and Computational Intelligence*, S. Patnaik, X.-S. Yang, and I. K. Sethi, Eds. Singapore: Springer, 2021, pp. 321–329.
[34] S. H. A. Faruqui, Y. Du, R. Meka, A. Alaeddini, C. Li, S. Shirinkam, and J. Wang, "Development of a deep learning model for dynamic forecasting of blood glucose level for type 2 diabetes mellitus: Secondary analysis of a randomized controlled trial," *JMIR mHealth uHealth*, vol. 7, no. 11, Nov. 2019, Art. no. e14452.
[35] M. Zanon, M. Riz, G. Sparacino, A. Facchinetti, R. E. Suri, M. S. Talary, and C. Cobelli, "Assessment of linear regression techniques for modeling multisensor data for non-invasive continuous glucose monitoring," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 2538–2541.
[36] G. A. Alonso-Silverio, V. Francisco-García, I. P. Guzmán-Guzmán, E. Ventura-Molina, and A. Alarcón-Paredes, "Toward non-invasive estimation of blood glucose concentration: A comparative performance," *Mathematics*, vol. 9, no. 20, p. 2529, Oct. 2021. [Online]. Available: https://www.mdpi.com/2227-7390/9/20/2529
[37] R. M. Neal, *Bayesian Learning for Neural Networks*, vol. 118. New York, NY, USA: Springer, 1996.

FAHEEM SHAIKH (Student Member, IEEE) received the B.S. degree in mechanical engineering from California State University, Fresno, in December 2020, where he is currently pursuing the M.S. degree in electrical engineering. His current research interests include biomedical devices, machine learning, and data science.

NOAH HAWORTH received the Bachelor of Science degree in mechanical engineering from California State University, where he is currently pursuing the Masters of Science degree in engineering with a focus on mechanical engineering. His research interests include biomedical devices and mechatronics. He is interested in pursuing an academic career in the future.

RILEY WELLS received the B.S. degree in mechanical engineering from California State University, Fresno, in 2020. His research interests include computer programming, artificial intelligence, and machine learning.

JODI BISHOP pursued the B.S. degree in electrical engineering from California State University, Fresno. She was involved in the early stages of the project.

SHRE K. CHATTERJEE received the B.Eng. degree in electronics and communications from Visvesvaraya Technological University, Belgaum, India, in 2004, and the M.Sc. degree in nanotechnology from the University of Southampton, U.K., in 2006, where he is currently pursuing the Ph.D. degree in the area of machine learning, signal processing and data science. He was involved in several high profile microfabrication related projects, specializing in active matrix, flexible, and rollable displays in industry. Following his Ph.D., he has worked as a Research Fellow with the University of Southampton in a few different projects. He is currently working in the industry as a Senior Engineer.

SANKHA BANERJEE received the M.S. and Ph.D. degrees in mechanical and aerospace engineering from Rutgers University. He is currently an Associate Professor with the Department of Mechanical Engineering, Fresno. He has extensive experience in the interdisciplinary areas of materials fabrication, plasma processing and characterization. He also founded the Energy Devices and Plasma Applications Laboratory, which is partially funded by the Department of Defense and Southern California Edison. He has worked with the Princeton Plasma Physics Laboratory on synthesis and surface modification of nanomaterials using thermal and micro-discharge plasmas. He has over 25 refereed journal publications and over 40 presentations in national/international conferences and technical meetings. He has also received several grants over $1 million from several federal, state, local agencies and the private industry. He actively collaborates with UC Merced, Rutgers University, and the State University of New York, Buffalo, in developing lead free ferroelectric perovskite oxides and halides for biomedical applications.

SOUMYASANTA LAHA (Member, IEEE) received the M.S. degree in embedded digital systems with distinction from the University of Sussex, U.K., in 2007, and the Ph.D. degree in electrical engineering from Ohio University, Athens, OH, USA, in 2014. He was a Visiting Researcher at Intel Inc., Austria, in 2016. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, California State University, Fresno, CA, USA. He has over 30 refereed publications in journals and conference proceedings. His current research interests include biomedical instrumentation for wearable health monitoring and energy-efficient mm-wave/sub-THz transceiver design in advanced CMOS technology for on-chip wireless communications. He was a recipient of the Stocker Research Fellowship from Ohio University and the New Investigator Research Grant Award from CSUPERB of the California State University.