scientific reports

OPEN

Clinical evaluation of a polarization-based optical noninvasive glucose sensing system

Ho Man Colman Leung[1]✉, Chengyue Gong[2], Luke Geiser[3], Emily E. Fivekiller[3], Nam Bui[4], Tam Vu[5], Temiloluwa Prioleau[5], Gregory P. Forlenza[3], Qiang Liu[2] & Xia Zhou[1]

Diabetes affects millions in the US, causing elevated blood glucose levels that could lead to complications like kidney failure and heart disease. Recent development of continuous glucose monitors has enabled a minimally invasive option, but the discomfort and social factors highlight the need for noninvasive alternatives in diabetes management. We propose a portable noninvasive glucose sensing system based on the glucose's optical activity property which rotates linearly polarized light depending on its concentration level. To enable a portable form factor, a light trap mechanism is used to capture unwanted specular reflection from the palm and the enclosure itself. We fabricate four sensing prototypes and conduct a 363-day multi-session clinical evaluation in real-world settings. 30 participants are provided with a prototype for a 5-day home monitoring study, collecting on average 8 data points per day. We identify the error caused by differences between the sensing boxes and the participants' improper usage. We utilize a machine learning pipeline together with Bayesian Ridge Regressor models and multiple-step data processing techniques to deal with the noisy data. Over 95% of the predictions fall within Zone A (clinically accurate) or B (clinically acceptable) of the Consensus Error Grid with a 0.24 mean absolute relative differences.

Diabetes is a chronic health condition that is currently affecting 38.4 million of people in the United States and the trend is increasing[1]. People with diabetes (PwD) often produce insufficient insulin, resulting in elevated blood glucose levels. Prolonged periods of high blood glucose level could damage the blood vessels which leads to complications such as kidney failure, heart disease, and stroke[2]. While intensive insulin treatment could reduce the risk of these complications as demonstrated in the landmark Diabetes Control and Complications Trial[3], it also increases the frequency of lower than normal blood glucose level, potentially causing cognitive impairment and seizures[4]. Given the lack of a cure for diabetes, it is essential for PwD to regularly monitor and manage their blood glucose level.

Blood glucose level is traditionally measured by pricking the fingertip to obtain a blood sample, which is then analyzed with a glucometer using an enzymatic test strip[5]. While accurate and effective for home monitoring, this method is invasive and inconvenient for PwD. Conversely, recent development has enabled a minimally invasive continuous glucose monitor (CGM) that reports blood glucose level regularly[6]. The device involves the insertion of a thin but flexible enzymatic sensor into the outer layer of the skin and is capable of measuring the glucose concentration typically for 7-14 days. However, discomfort from sensor insertion, potential bacterial infection[7], the physical and psychological burden of having it visibly attached, and the high financial burden of having the sensors replaced regularly have led to a relatively low adoption rate, underlining the necessity for a completely noninvasive approach[8].

Previous research has explored glucose sensing either directly or indirectly. Direct approaches involve studying the glucose present in biological fluids such as interstitial fluid[9–13], sweat[14–16], tears[17–19], and saliva[20,21]. These fluids are more accessible compared to blood, and their glucose concentration correlates with the blood glucose closely, making them of great interest to the noninvasive glucose sensing community[22]. Various techniques, including optical[9–12,23–26] and transdermal[27–29], are employed. Optical methods primarily exploit the

[1]Department of Computer Science, Columbia University, New York, NY 10027, USA. [2]Department of Computer Science, University of Texas at Austin, Austin, TX 78712, USA. [3]Barbara Davis Center for Diabetes, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA. [4]Department of Electrical Engineering, University of Colorado Denver, Denver, CO 80204, USA. [5]Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA. ✉email: colman.leung@columbia.edu

selective absorption properties of the spectrum to estimate blood glucose level. However, the limited penetration depth of light [30] and the overlapping absorption spectrum of other confounding substances [31] have impeded the development of a noninvasive system. Transdermal methods require the extraction of body fluids to the skin's surface using either a small current[27,28] or an electromagnetic field[29], and analyzing the fluids with conventional enzymatic glucose sensors. However, it takes time to extract enough fluid for measurement, leading to skin irritation from prolonged application of electric current to the skin. In contrast, indirect approaches infer blood glucose level from physiological signals or tissue properties. For instance, photoplethysmography (PPG) signals are used to predict blood glucose level by observing changes in blood viscosity caused by varying glucose concentrations[32–35]. However, confounding factors such as the concentration of red blood cells and plasma protein also affect blood viscosity[36]. Additionally, the signal is typically analyzed with deep learning models, and hence, the relationship between PPG signal and blood glucose level is not well-studied. Another example is bioimpedance spectroscopy, where alternating currents of various frequencies are applied to the skin to measure bioimpedances[37,38] and dielectric properties[39–41]. However, confounding factors such as body composition and hydration level, along with the insensitivity to changes in blood glucose level, result in unsatisfactory performance[42].

In this work, we propose a portable system for sensing glucose level noninvasively (Fig. 1). Our system comprises three components: the sensing box (Fig. 1c), a smartphone application (Fig. 1b), and a glucose level estimation pipeline (Fig. 1e). It harnesses the optical activity of glucose molecules to predict glucose concentration [43]. In essence, as glucose rotates the plane of polarization when linearly polarized light passes through it due to its optical activity, our system measures the optical rotation related to the glucose concentration in the skin using an optical polarimetry technique and infers the glucose level. Prior studies have explored

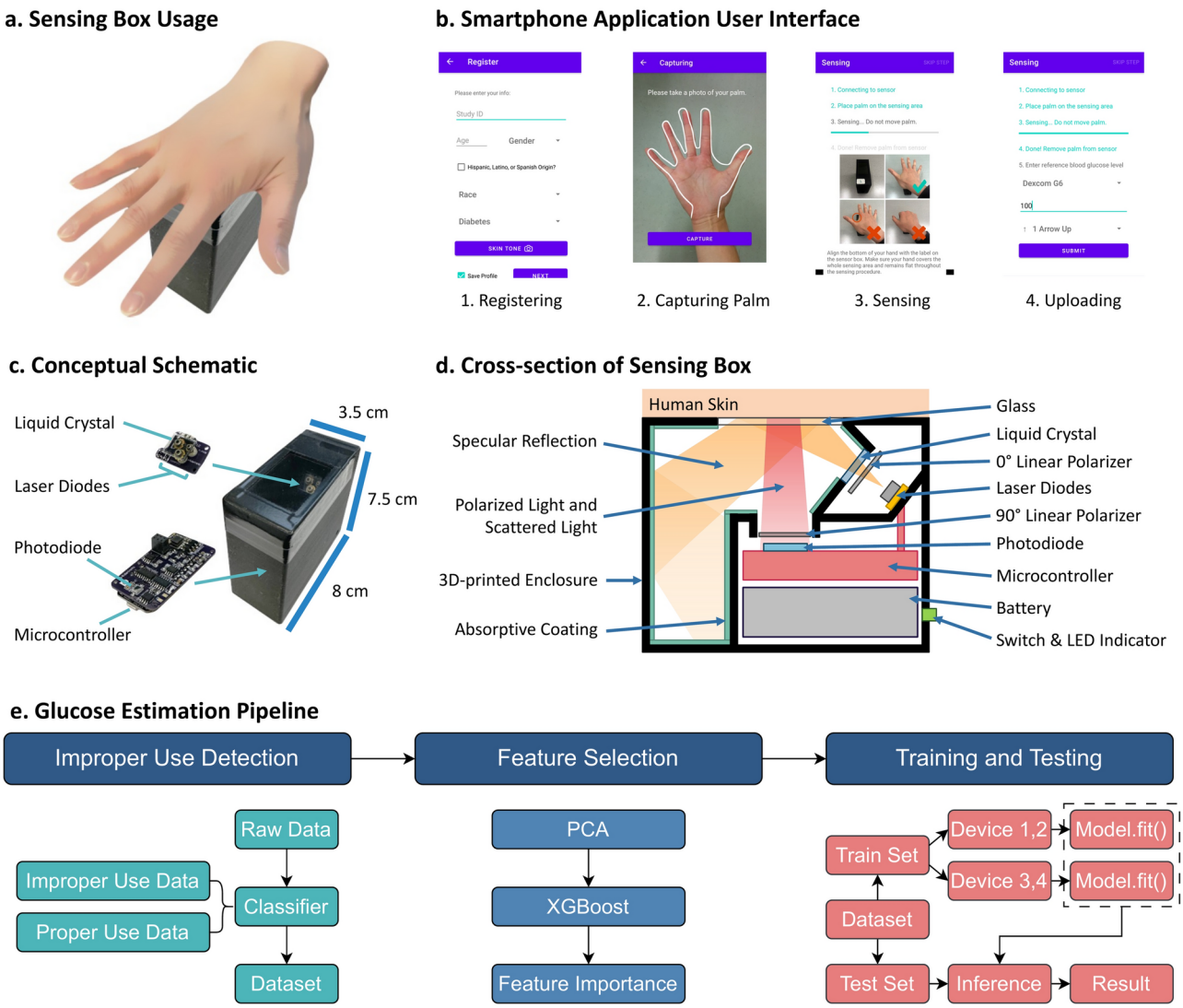


Fig. 1. System overview. (**a**) Palm's placement on the device. (**b**) Screenshot of Smartphone Application. (**c**) Sensing Box schematic. (**d**) Cross-section of the Sensing Box. (**e**) End-to-end glucose level estimation pipeline.

polarization-based approaches for glucose level inference by probing the aqueous humor in the eye[44,45], but mounting the system to the eye is impractical and inconvenient.

Our method is inspired by the polarimetry-based technique introduced by Li et al.[13]. While their system shows promising results, the sensing prototype is much larger, making it non-portable and inconvenient to carry around. Using demographics to infer skin tone may be sub-optimal as it only describes skin tones discretely. Finally, the lack of a clinical study across multiple days further limits its verifiability. Here, our goal is to develop a portable, miniaturized version of the prototype, extract features from palm images for skin tone analysis, and conduct a home monitoring clinical study where data is collected from PwD over time. The smaller form factor of the prototype poses an additional challenge of preventing light reflected from the skin from blinding the optical sensor, which we address by utilizing an optical light trap design.

To evaluate the system's performance in a real-world scenario, we conduct a 363-day clinical study in the wild where the participants take the device with them for home monitoring. This is particularly challenging as user may use the prototype improperly, which is reflected in the high noise level in the dataset, unlike most of the previous work[9,10,12,13,23,32] which are conducted in a lab settings. We handle the noisy data by identifying these improper cases and removing them from the dataset using both statistical and machine learning approaches. Over 95% of the predictions all in either Zone A (clinically accurate) or B (clinically acceptable) of the Consensus Error Grid with 0.24 mean absolute relative differences.

To summarize, our proposed system offers the following contributions:

1. A portable device capable of noninvasive gluocse level sensing.
2. A light trap mechanism to reduce specular reflection from saturating the optical sensor
3. An image feature extraction process for handling skin tone variation.
4. A glucose level estimation pipeline with improper use case detection.
5. A multi-day multi-session home monitoring evaluation demonstrating the effectiveness of our proposed solution.

Result

Sensing concept

Our system utilizes the optical activity displayed by glucose molecules and predicts glucose level by measuring the optical rotation induced by glucose molecules in the skin [43]. Essentially, the inherent chirality of glucose prompts its optical activity, resulting in the rotation of linearly polarized light directly proportional to the glucose concentration. This rotation is commonly measured through polarimetry techniques, where light passes through the first polarizer to become linearly polarized, interacts with glucose, passes through the second polarizer, and reaches the light sensor as depicted in Fig. 2a.

The intensity of the light reaching the sensor is proportional to the concentration of glucose. However, in practice, the highly scattering and absorptive properties of the skin pose a significant challenge[46]. In Fig. 2b, when linearly polarized light is directed toward the skin, most of the light is either absorbed (53%) or scattered (40%) in the skin. Subsequently, the remaining 5% of the light is specularly reflected, which could easily blind and saturate the sensor, leaving only 2% of the light still carrying useful polarization information. Therefore, to extract the polarization information, we must effectively mitigate the effects caused by scattered light and specular reflection.

For scattered light, we employ a setup similar to a simple polarimetry technique in Fig. 2a, but the incident light is positioned at a 45° angle and modulated by a liquid crystal, thereby probing the skin with light of two polarization directions. Then, using a similar but different algorithm described in[13], we can cancel out the effects caused by scattered light and retrieve the polarization information. Regarding specular reflection, we design a light trap mechanism to capture the specular light as shown in Fig. 1d. The specular light reflects away from the skin and enters the light trap, which is coated with light-absorbing material. As the light bounces around inside the light trap multiple times, it is sufficiently absorbed by the coating, effectively minimizing leakage from the light trap back to the sensing element. These two techniques allow us to extract the polarization information. Additionally, we collect the user's demographics and the image of the user's palm to handle diverse populations

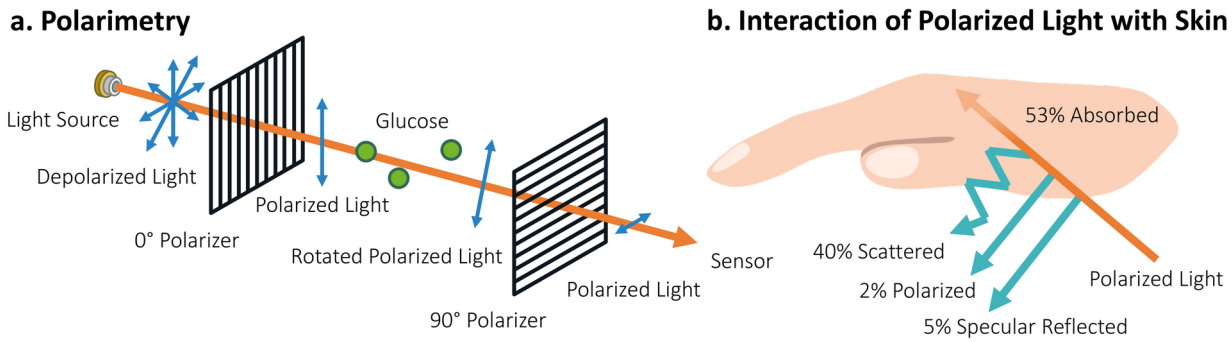


Fig. 2. System concept. (**a**) A basic polarimetry setup to measure optical rotation caused by glucose. (**b**) Interaction of polarized light with the human skin.

and skin variations. The features of the palm images are extracted by analyzing the image under various color and texture spaces. These three types of data are joint together to form the basis of the input to the machine learning model.

A machine learning pipeline is employed to predict the glucose level (Fig. 1e). Initially, improper usage is detected using a classifier trained on a use case dataset, which includes both proper and improper use cases. Then, the dimensionality of the input data is reduced via unsupervised Principal Component Analysis (PCA) [47] and supervised XGBoost [48] feature importance selection. Lastly, we conduct stratified training by grouping data points from sensing boxes 1 and 2 to one group, and those from sensing boxes 3 and 4 into another. A Bayesian Ridge Regressor is separately trained for each group, with the respective regressor utilized during inference. Details of the whole system are provided in the Methods Section.

Sensing box design and usage

The noninvasive glucose sensing system consists of two components: a sensing box and a smartphone application. The compact sensing box measures $8cm \times 7.5cm \times 3.5cm$, making it portable and convenient to carry (Fig. 1c). The sensing box comes with a smartphone application, which controls the sensing box and retrieves the sensor readings via Bluetooth (Fig. 1b).

To initiate the sensing procedure, the user first inputs their demographics into the smartphone application, including age, gender, race, ethnicity, and type of diabetes. Subsequently, the user captures a picture of their palm. They are given an option to save the information locally on the phone for future procedures. Upon confirming the correctness of the data, the application will instruct the sensing box to begin the sensing procedure and guide the user through the following steps:

1. Establishing a connection with the sensing box.
2. Placing the palm onto the sensing area.
3. Initiating the sensing sequence, where the sensing box begins to collect sensor readings.
4. Prompting the user to remove their palm.
5. Taking a measurement of their glucose level using the CGM and entering the value, CGM model, and trend arrows to the application.

The whole sensing procedure takes around 1 minute. After that, the user's demographics and sensor readings are uploaded to a database for future analysis.

Glucose level estimation performance

We collaborated with the Barbara Davis Center for Diabetes at University of Colorado Anschutz Medical Campus to conduct a clinical study on PwD, following approval from the Institutional Review Board. Participants were provided with sensing boxes for home monitoring during the study period. They were advised to perform data collection procedures eight times a day: upon waking up, before and after each meal, and before bed. Our dataset contains 962 data points from 22 participants. The study protocol, the participant demographics, and dataset preprocessing are detailed in the Method Section.

Evaluation metric

We evaluate the system's performance using the reference glucose values retrieved from the CGM and plot the results on a Consensus Error Grid (CEG)[49] (Fig. 3). The choice of CGM based on two key factors: (1) it provides a more suitable ground truth than fingerstick tests as they are both measuring glucose at the interstitial fluid

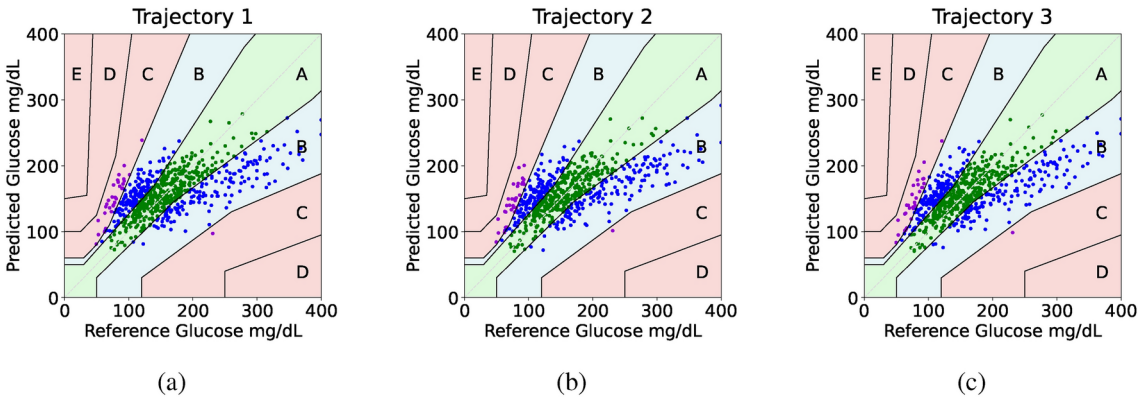


Fig. 3. The predicted glucose values generated using three random seeds are plotted against the reference glucose values onto three CEGs. The points in green are predictions that fall in Zone A, which are considered clinically accurate. The points in blue are those in Zone B, which are clinically acceptable. The predictions that fall into Zone C are marked as purple, which predicted a higher or lower glucose value with respect to the reference value. The percentage of predictions that fall into Zone A or B are 95.29%, 95.14%, and 95.29%, respectively. Conversely, 4.71%, 4.86%, and 4.71% of predictions fall into Zone C. There are no predictions in Zone D and E.

layer, and (2) it imposes a much lower overhead on participants compared to invasive fingerstick tests while also eliminating input errors by directly exporting the glucose data. Although measuring glucose in the interstitial fluid instead of blood can introduce an approximately 10% error in estimating blood glucose levels [50], fingerstick tests can have inaccuracies ranging from 5.6% to 20.8% depending on the brand and model[51]. Therefore, CGMs offer an acceptable level of accuracy comparing to fingerstick tests.

The CEG is a widely used metric in diabetes research where the predicted glucose values obtained from our sensing system are plotted against the reference glucose values retrieved from the CGM. It is divided into five zones: A-E. Zone A indicates clinically accurate predictions, with no effect on clinical action. Zone B corresponds to altered clinical action in which the errors in the predicted glucose value affect the treatment decision. Zone C represents predictions of higher or lower glucose values, leading to overcorrection. In Zone D, actual glucose values are either too high or too low, while predicted glucose values fall within the healthy range. This fails to detect abnormal blood glucose level, which poses risks to the user. Finally, Zone E indicates incorrect decisions due to reporting opposite events, leading to anti-correction. Additionally, we analyze the results by calculating the Pearson/Spearman correlation coefficients and the mean absolute relative difference (MARD)[22] between the predicted blood glucose level and the reference level. MARD is calculated as follows:

$$MARD = \frac{1}{n}\sum^{n}\frac{|\text{Predicted Value} - \text{Reference Value}|}{\text{Reference Value}},$$

which explains the mean percentage difference between predicted and reference glucose level.

Overall performance

We present the performance of our system using different models and feature processing procedures; see Table 1. We use leave-one-out cross-validation to maximize the use of available training samples while ensuring temporal independence, as the data points from each user are spaced apart by an average of 1.98 hours throughout the day. Overall, we observe that simple Bayesian ridge regression with improper use detection, image features, feature selection, and stratified training achieves the best performance. In particular, this model outperforms other machine learning models such as Gaussian process regression, XGBoost, and Neural Network. Ridge regression performs better than other methods due to the multicollinearity nature of the data, making the model less prone to overfitting. This is because the raw sensor readings consist of data collected from three laser diodes, two liquid crystal states, and 100 power intensities, resulting in a dimension of 600 that are highly correlated with each other. Despite using PCA to reduce the dimensionality, the remaining features are still correlated to each other, making ridge regression ideal for our scenario. On the other hand, Bayesian linear models excel in situations with high uncertainty. Specifically, the use of the sensing boxes varies among individuals despite prior training. This results in a better performance compared to other models. Regarding neural-network-based methods, the lower performance may be attributed to the high dimensionality of the data and the relatively small dataset, leading to overfitting.

Moreover, we plot the prediction results onto the CEG using three of the random seeds (Fig. 3). These random seeds affect the initialization weights and the optimization trajectory and therefore come to different results. Among the three CEG plots, 95.29%, 95.14%, and 95.29% of predictions lie within Zone A or B, which are clinically accurate and clinically acceptable. Only 4.71%, 4.86%, and 4.71% of predictions are in Zone C. No predictions are in Zone D and E. These results indicate that, our system can consistently get promising results and is robust to random seeds.

Ablation study

In this section, we remove each individual component to understand its effect on the system's overall performance.

Improper use detection Identifying and filtering out improper use cases greatly improves the performance of the system, reducing MARD from 0.27 to 0.24 and increasing Pearson correlation from 0.59 to 0.67. This improvement stems from the fact that improperly collected data points reside in a distinct feature space compared to the properly collected ones. Due to the dataset's limited size, there is insufficient data to capture the improper data feature space, resulting in high prediction error for these data points.

Method	MARD	Pearson	Spearman
Bayesian ridge regression	**0.24**	**0.67**	**0.63**
w/o Improper use detection	0.27	0.59	0.54
w/o Image features	0.25	0.66	0.63
w/o Feature selection	0.26	0.62	0.57
w/o Stratified training	0.32	0.47	0.43
Gaussian process regression	0.28	0.58	0.54
XGBoost	0.28	0.56	0.53
Neural network	0.29	0.51	0.48

Table 1. System's performance with different models and features processing procedures. Significant values are in bold. The values are obtained by averaging the result from three trajectories.

	Sensing Box				Gender		Race			
	1	2	3	4	M	F	A	B/AA	H/L	W
MARD	0.234	0.260	0.243	0.243	0.269	0.237	0.201	0.228	0.252	0.247
Pearson	0.708	0.679	0.649	0.613	0.665	0.672	0.339	0.683	0.698	0.571
Spearman	0.611	0.647	0.676	0.521	0.653	0.619	0.302	0.664	0.703	0.535

Table 2. The performance of the system with respect to each sensing box, gender, and race. The MARD, Pearson correlation, and Spearman correlation are listed. The genders include Male (M) and Female (F), and the races listed are Asian (A), Black or African American (B/AA), Hispanic or Latino (H/L), and White (W).

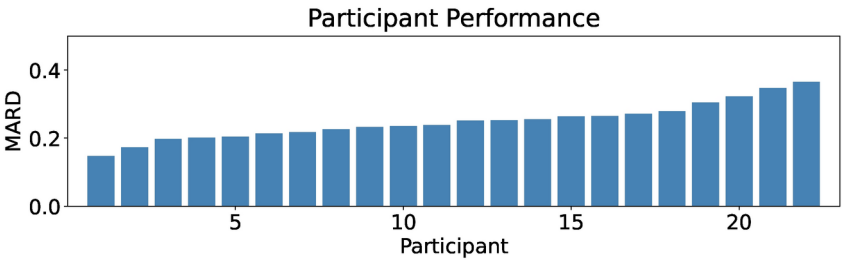


Fig. 4. The performance of each participant. Each bar corresponds to the averaged MARD from the trajectories for a single participant. The performances are sorted in an ascending order.

Image features Incorporating features derived from the palm image marginally enhances the system's performance. These features, obtained from various color and texture spaces, describe the participant's skin tone and texture in a continuous manner. However, demographic features, especially age and race, may also correlate with these image features. For example, aging can lead to changes such as thinner skin, uneven pigmentation, and decreased water content[52], while race influences skin tone. Thus, having both palm image features and demographic features only provides a minor improvement to the model performance.

Feature selection The feature selection procedure improves the performance of the system, reducing the MARD from 0.26 to 0.24 and increasing the Pearson correlation from 0.62 to 0.67. This indicates that using too many features with an insufficient amount of training data could easily lead to overfitting. PCA is employed to reduce dimensionality, which can mitigate multicollinearity by creating orthogonal variables to capture most of the data variance. Additionally, features are further filtered by selecting only those with the highest importance weights using XGBoost, effectively enhancing system's performance.

Stratified training Training a separate model for each group of sensing boxes significantly impacts the overall performance. If a joint model is trained for all sensing boxes instead, the MARD increases from 0.24 to 0.32 while the Pearson correlation and Spearman correlation drop from 0.67 to 0.47 and 0.63 to 0.43, respectively. One potential reason for this is that the features of various sensing boxes occupy different subspaces in the feature space. Due to the small amount of training data, the machine learning model cannot learn these complex patterns by itself. Hence, manually separating the data produces better results.

Sensitivity analysis

We conduct a sensitive study to assess how different sensing boxes, users and races affect the performance of the system. The MARD, the Pearson correlation, and the Spearman correlation concerning each variable are tabulated in Table 2.

Sensing box Figure 5a demonstrates the MARD scores for different sensing boxes, ranging from 0.234 to 0.260. While sensing box 2 yields the best results, all the other three sensing boxes perform similarly. The standard deviation (s.t.d.) of these four sensing boxes is 0.009, indicating that our system does not exhibit strong bias regarding sensing boxes.

User In Fig. 4, results for different users are visualized. Users are sorted by their MARD performance, which ranges from 0.148 to 0.365. The overall s.t.d. is 0.05, suggesting that our machine learning model does exhibit strong bias. Possible reasons for the performance gap include data noise and the lack of data collected from diverse demographics. A CEG plot for each user is also available in Figure S1 at the Supplementary Information 1.

Demographics Performance for different genders and races is examined in Table 2 and Fig. 5b and c. MARD for male and female participants is 0.269 and 0.237, respectively, while MARD for Asian, Black or African American, Hispanic or Latino, and White are 0.201, 0.228, 0.252, 0.247 respectively. The s.t.d. for gender is around

0.016 and for race 0.02, indicating that our system does not exhibit strong bias towards a particular demographics category.

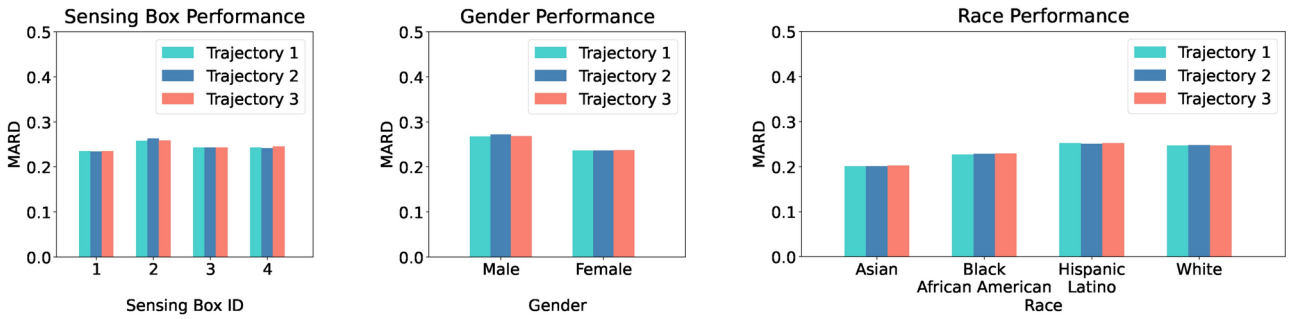
Although male participants have a higher MARD compared to female participants, the correlations are similar, suggesting that the poorer performance may be due to the imbalance of our dataset. There are 16 female participants but only 6 male participants. Regarding the performance between races, it is noteworthy that Asians have the lowest MARD, but also a lower correlation coefficient. This is likely due to having only a single Asian participant, and upon closer inspection, we found a narrow range in the reference glucose values. As noise in collected data introduces random fluctuations, it leads to a lower correlation coefficient. Overall, our model demonstrates robustness across different demographic groups.

Comparison to a prior work We develop a non-invasive glucose sensing system that is portable and convenient to use, which is inspired by a prior work[13]. In particular, we realize a much smaller form factor by utilizing a light trap mechanism to capture the unwanted specular reflection light and enable a better representation of skin tone and condition using features extracted from palm image. Furthermore, our clinical study allows the patients to bring the sensing prototype back home and use the system in an unsupervised manner. This is closer to a real use case scenario and more convenient for the participants to collect more data points. While most participants have been properly placing their palms for data collection, we observed a noticeable fraction of the collected data to be noisy and captured improperly. The resulting degradation of data quality subsequently affects the system's performance. In contrast, the data collection process in the previous study is conducted in a controlled lab setting, where each patient used our sensing prototype under our supervision during their doctor office visit to ensure proper and correct prototype usage. Additionally, our sensing box is significantly smaller than the prior work ($8cm \times 7.5cm \times 3.5cm$ vs $17cm \times 10cm \times 5cm$). Despite the use of the light trap mechanism, the small form factor still introduces additional specular reflection to the sensor which increases the amount of noise captured by the sensor. Therefore, the combined effects resulted in a lower performance.

Discussion

We study the efficacy of the polarization-based technique for glucose level inference by proposing a portable sensing system with a dedicated machine learning model to handle improper use cases and evaluating it with a clinical study. The multi-session clinical evaluation of 22 people with Type 1 Diabetes results in over 95% of predictions in Zone A and B, and a 0.24 MARD. The improper usage of the sensing prototype in a home monitoring setting has introduced significant noise affecting sensing accuracy. Overall we observe that use of light polarization properties for sensing glucose level holds promises. Next we discuss future research directions to improve sensing performance.

First, future studies can further improve the dataset scale, participant diversity, and the quality of ground-truth data to enhance the sensing performance. Our current dataset has limited data on 22 user population. The dataset size leads to more features extracted from the sensor data than the number of data points. As such, complex models such as neural networks tend to overfit quickly and we can only resolve to simpler machine learning models such as regression which are less effective against noise. Moreover, each participant collected on average 40 data points only, which is insufficient for training a personalized model given the high level of noise. We currently use a leave-one-out cross-validation strategy to train the model due to the limited dataset. This is a valid approach as the data points are spaced several hours apart, indicating that a similar number of calibrations may be needed to achieve the same performance for new users. With a larger dataset, we could explore alternative split strategies, such as leaving out an entire day or an entire participant to better assess the model's generalizability. As for the reference glucose data, although CGMs report interstitial fluid glucose levels that are highly correlated to fingerstick tests[53], it still inherently contains a delay in glucose level responses. For future studies, we can include fingerstick tests as our ground truth for training the model. Additionally, we conduct a separate study involving a participant without diabetes using fingerstick tests. We collect 40 pairs of data points using our prototype and the Contour Next One glucometer. Our results shows a MARD of 0.182 with all predictions falling within Zone A and Zone B of the CEG (Refer to Fig. S2 in Supplementary Information



(a) Different Sensing Box (b) Different Gender (c) Different Race

Fig. 5. The performances of the system regarding different (**a**) sensing boxes, (**b**) genders, and (**c**) races. Each bar corresponds to a trajectory, and they are grouped by (**a**) sensing box ID, (**b**) genders, and (**c**) races.

2 for more details). These findings are comparable to ours and other prior studies that sense the glucose in the interstitial fluid layer in real-world conditions[11,26].

Second, the techniques of palm feature extraction can be optimized with a larger dataset. In the current study, we extract the features from the palm by converting the whole image into various color and texture spaces. These features are useful to describe the skin tone and the texture of the palm. However, since the images are taken manually, they are captured under different lighting conditions. The features are extracted from the whole image, but only a portion of the image is the palm. Finally, spatial features such as pigmentation and damaged skin are not captured due to the dataset size that prevents us from building an end-to-end neural network to extract palm features efficiently. With a larger dataset, we can explore more on the palm feature extraction, and hopefully further improve the performance of the system. In particular, we can take a high resolution image of the palm and employ a more sophisticated neural network to perform segmentation. This allows us to extract useful information such as the skin surface and the sweat glands as the uneven nature of the skin and the secretions of sweat could affect how light is reflected from the skin[54]. To achieve better self-supervised inference for low-level vision features, e.g., brightness, lightness, recently-developed large-scale vision foundation models[55] may be an alternative choice and candidate for feature extraction.

Third, the system can be further optimized to support continuous monitoring and reduce the likelihood of improper usage. Specifically, to further reduce the prototype size and make it wearable, we can consider a skin-attachable sensor form factor[56]. The idea is to place the light source and sensor directly in contact with the skin so that all the light that reaches to the sensor must come from the skin. This allows us to remove the 45° incident light source design and the light trap which is one of the biggest components inside the sensing box, significantly reducing the sensing box into a much smaller form factor. The small form factor makes it more convenient for the user to have the sensor placed on the skin 24/7, realizing the continuous monitoring feature. It can also reduce the chance of improper usage, as the optical components are directly touching the skin, where the propagation of light rays will be relatively constant throughout the sensing session. On the software side, currently the smartphone application is only used to collect and upload the sensing box measurements to a server for offline analysis. For a real-time glucose sensing system, a web service API can be implemented to receive the sensing box readings and perform glucose estimation. The simple machine learning model only requires 0.01 seconds on a computer to perform inference, enabling a continuous monitoring application. Finally, the user experience of various form factors should be explored to maximize the system's usability. This allows us to better understand the user's preferences such as the shape, size, and location of the sensors, resulting in a more comfortable and convenient system.

Finally, future research can dive deeper into the impact of confounding factors. For example, other substances in the interstitial fluid such as collagen and albumin are also optically active [57]. While the concentration of these substances could be relatively constant for a person, they may be quite different compared to other people. These individual differences will introduce errors to our prediction model. Moreover, the visible light can be absorbed by the haemoglobin and melanin inside the skin[58], which adds another layer of noise to the signal. This is a challenging problem that have been affecting all optical techniques. One potential solution is to probe the skin with more wavelengths and intensity levels as each substance can have a unique optical rotation profile. Another direction is to combine with other techniques to better separate confounding factors from glucose molecules. For example, infrared spectroscopy can be utilized as each substance has a unique absorption spectrum which differs with the specific rotation. Environmental factors, such as temperature and humidity, can influence skin condition. Although the polarimetry-based technique has demonstrated robustness against temperature changes and pressure variations in controlled lab settings [13], the higher variability of these factors in real-world environments could impact the system's accuracy. In particular, the clinical study spans an entire year, capturing data across different seasons and weather conditions. To mitigate this, one solution is to incorporate additional temperature and humidity sensors to monitor and account for these environmental changes. Furthermore, the skin-attachable sensor form factor mentioned earlier could help shield the skin from external environmental influences, providing more consistent measurement conditions. Research efforts can focus on efficient hardware design and algorithmic solutions to fuse various techniques for more reliable noninvasive glucose sensing.

Methods

Clinical study protocol

We initially recruited 30 participants for the study, but three dropped out after the start of the study. Subsequently, we excluded data from five participants due to a majority of their collected data exhibiting signs of improper usage. Consequently, our dataset contains 962 data points from 22 participants. The demographics of these participants are outlined in Table 3. Participants were familiarized with the data collection procedure and given the opportunity to practice scanning their palms on the first day, where then they take the device with them and perform the remaining scans at home. Each participant was instructed to take measurements approximately eight times daily over a 5-day period, including measurements taken upon waking up, before and after meals, and before bedtime. The participants also used their CGM to collect the glucose level data, which was manually entered into the system as reference glucose values. Additionally, participants' CGM data was retrieved from Clarity or Libreview CSV files. As a token of appreciation, each participant received a $75 Amazon Gift Card upon completion of the study. This study was approved by the Institutional Review Board of Barbara Davis Center for Diabetes at University of Colorado Anschutz Medical Campus and was conducted in compliance with applicable guidelines and regulations. Informed consent were obtained from all participants.

Features extraction

We explore the features extracted from the sensor readings and the palm images that are utilized for glucose level prediction.

	Age		Gender		Race group				Skin tone		CGM		
Demographics	Median	IQR	Male	Female	A	B/AA	H/L	W	Mean	SD	FL1	FL2	DG6
No. of Users	28.5	12.5	6	16	1	4	3	15	1.13	0.34	1	1	19

Table 3. Demographics of the participants recruited for the clinical study. The race groups are Asian (A), Black or African American (B/AA), Hispanic or Latino (H/L), and White (W). The CGMs are Freestyle Libre 1 (FL1), Freestyle Libre 2 (FL2), and Dexcom G6 (DG6). The skin tone is represented by the Fitzpatrick Scale, where 1 is the lightest and 6 is the darkest. They are obtained by extracting the palm region and analyzing the image in the LAB color space. Note that the palm is generally lighter compared to other part of the skin due to a different melanosome distribution[59], resulting in a generally lower Fitzpatrick Scale value.

Polarization-based optical features

As glucose is optically active, the plane of polarization will be rotated when linearly polarized light passes through them. For example, it rotates a light beam of 589 nm clockwise by $52.7^{c}irc\,\mathrm{dm}^{-1}(\mathrm{g/mL})^{-1}$ at $20^{c}irc$C[43]. This rotation α is linearly proportional to the glucose concentration C which can be explained by the following equation.

$$\alpha = R(\lambda, T) \cdot C \cdot L, \tag{1}$$

where $R(\lambda, T)$ is the rotary power of glucose with respect to the wavelength λ and temperature T, and L is the optical path length[60]. In our setup, we utilize light of two polarization states. The sensor readings are denoted as I_{on} and I_{off} when the liquid crystal is on and off respectively. We can calculate a feature F such that

$$F = \frac{I_{\mathrm{on}} - I_{\mathrm{off}}}{I_{\mathrm{on}} + I_{\mathrm{off}}} = \frac{I_{\mathrm{p}}(-\cos 2\alpha) - I_{\mathrm{s}}}{I_{\mathrm{r}}} = \frac{I_{\mathrm{p}}}{I_{\mathrm{r}}}(-\cos 2\alpha) - \frac{I_{\mathrm{s}}}{I_{\mathrm{r}}}.$$

Approximating the cosine function with Taylor series,

$$F \approx \frac{I_{\mathrm{p}}}{I_{\mathrm{r}}}\left(\frac{\alpha^2}{2} - 1\right) - \frac{I_{\mathrm{s}}}{I_{\mathrm{r}}},$$

where $\frac{I_{\mathrm{p}}}{I_{\mathrm{r}}}$ and $\frac{I_{\mathrm{s}}}{I_{\mathrm{r}}}$ correspond to the ratio of polarized light and specular light to the total reflected light respectively. These two terms are related to the design of the system such as the placement of the optical components and, therefore, are constant. Then, the feature value F that we calculated has an interesting physical property as it has a direct correlation with the square of the rotation angle α. A full derivation can be found at Supplementary Information 3.

Palm image features

The features from the collected palm images of the participants are extracted. These features aim to provide a comprehensive representation of the palm, capturing both color and texture information. First, in the HSV color space, we generate the features: mean hue (H), saturation (S), and value (V), which describe the dominant color, vibrancy, and overall brightness, respectively. Second, we convert the images to the LAB color space which consists of the mean lightness (L) and the mean values of the green-red (A) and blue-yellow (B) components, offering a color representation aligned with human perception. Third, Local Binary Pattern (LBP) features capture texture information by encoding local pixel variations into a histogram, reflecting the frequency of specific texture patterns. By combining these features, it generates a rich, multidimensional image descriptor of the palm, effectively capturing the skin tones and textures and enhancing the overall accuracy of the system.

System design

We fabricate four sensing boxes that are wireless and portable using off-the-shelf components. Figure 1 shows the overview of our noninvasive glucose sensing system. Figure 1b–e illustrate our system, which contains three components: a sensing box, a smartphone application, and our glucose estimation pipeline. The usage of the sensing box is shown on Fig. 1a.

Sensing box

The sensing box, contained within a 3D-printed enclosure measuring at $8cm \times 7.5cm \times 3.5cm$, integrates electronic components and a rechargeable battery. At its core lies the optical unit, capable of emitting laser light at three wavelengths (658 nm, 520 nm, 450 nm) and two polarization states (0°, 90°) using three laser diodes and a liquid crystal. The laser light intensity is modulated by the control unit using a switch and a potentiometer whereas the liquid crystal is driven by an integrated chip. On the other hand, the reflected light is detected via an optical sensor and amplifiers. Both the optical and the sensing components are managed by a microcontroller powered by a 3.7V lithium polymer. The total cost of the sensing box is around $210 with the majority goes to the laser diodes ($130) and the microcontroller ($30). This can be further reduced through mass production.

Due to its small profile, the light source, the optical sensor, and the skin are brought very close together. This results in the optical sensor capturing a lot of light that is readily reflected off the surface of both the palm and the sensing box. This reflected light has not interacted with the glucose inside the skin and significantly increases the

portion of noise to the sensor reading. We employ a structural design that implements a light trap mechanism to guide the specular light away from the sensor. Figure 1d shows the cross-section view of the sensing box. The light trap is positioned such that the specular light will enter the light trap and then bounce around the trap multiple times. Additionally, the light trap is coated with light-absorbing material to maximize light absorption, effectively minimizing the leakage from the light trap back to the sensing element. The implementation and the hardware details are described in the Supplementary Information 4.

Smartphone application
Our dedicated Android smartphone application, developed in Java, facilitates seamless communication with the sensing box via Bluetooth using the UART protocol. It sends commands to the sensing box to initiate the sensing procedure which controls the liquid crystal and laser diodes, and retrieve the readings back from the sensing box. The application also provides a user-friendly interface for capturing palm images and collecting the user's demographics, including age, gender, race, ethnicity, and the type of diabetes (Figure 1(b)). When the application receives all the required data, it uploads them to a secure database for future analysis. Additionally, the Internet connection status, the Bluetooth connection status, and the sensing box battery percentage are displayed to the user. This ensures that users are aware of the system's status, ensuring the sensing box has sufficient power and connectivity for proper functionality, and enabling successful upload of data to the database.

Glucose estimation pipeline
The glucose level estimation pipeline depicted in Fig. 1e comprises three key stages. Initially, improper use cases are identified through a combination of statistical analysis and machine learning techniques to filter out unreliable data points. Following this, the optimal set of features are selected using PCA and XGBoost Importance Selection. Finally, the model is trained and evaluated using leave-one-out cross-validation. This approach is particularly well-suited for the small dataset, as it maximizes the use of available training samples, thereby improving the reliability of the model's performance evaluation. Furthermore, it is valid in this context because the data points collected by each user are spaced several hours apart. This temporal separation ensures independence between training and testing samples, effectively eliminating the risk of data leakage. The entire pipeline is highly efficient, taking approximately 0.01 seconds to complete a single glucose estimation.

Improper usage detection
Improper usage detection involves identifying instances where users fail to adhere to the sensing procedure. Our sensing box requires a user to fully cover the sensing area with their palm while remaining fully flat and stationary for around 1 minute. While ensuring adherence to the correct procedure is feasible under supervision, it becomes challenging in unsupervised home monitoring scenarios. Despite the convenience and flexibility of allowing users to perform sensing at their discretion which allows more data points to be collected, the quality of the data, however, may suffer when the users do not strictly adhere to the protocol. Here, we identify four types of improper usage scenarios and address them with various techniques:

- *Curved hand* Users often relax their hand instead of keeping it flat, creating an air pocket between the palm and the glass. This alters the geometry of the light's path, resulting in readings that diverge into a different feature space. To tackle this issue, we collect a separate dataset with users intentionally relaxing their hands on the sensing box and train a classifier to detect those scenarios. Combining this improper use-case dataset with the clinical study dataset, we employ PCA to visualize the data into a 2D feature space (Fig. 6). Notably, the light blue points representing the improper use cases are separable from the rest. We filter out all the data points that are classified as improper use data (red points) with a confidence score greater than 0.705 using a binary support vector classifier.
- *Hand moved during sensing* During a standard sensing procedure, the user's hand remains stationary while the power of the laser diodes gradually increases. This results in a relatively smooth increase in reflected light intensity over time. However, if the user's hand moves during the procedure, it causes a sudden spike in the sensor reading as ambient light is leaked into the sensing box. We detect these spikes by calculating changes in reflected light intensity and comparing them to a predefined threshold.
- *Sensing area not fully covered* When the user's hand does not cover the sensing box entirely, some ambient light enters the sensing box and fully saturates the optical sensor. We can simply remove these data points as the sensing box could not capture anything useful under such conditions.
- *Incorrect input of CGM data* The ground truth is entered by the user after performing a sensing procedure. However, errors may occur due to human error. To rectify this, we extract corresponding CGM values from Clarity or Libreview CSV files. If there are two readings within 5 minutes before and after the sensing procedure, we linearly interpolate these two values to obtain the corrected ground truth. Otherwise, we use the user-entered value as the ground truth.

Our original dataset contains 1147 data points from 27 users, where 89 improper use cases are detected by the algorithm above. Upon analyzing the improper data, we notice 5 users who have more than 10 improper use cases, and subsequently removed from the dataset, resulting in a total of 962 data points in the dataset.

Feature selection After the improper data is removed, we perform feature selection to identify an optimized set of features for maximum performance. Three types of data are used: the user's demographics, palm image features, and sensor values features. The user demographics consist of 14 features, including age, gender, and races, which are collected through the smartphone application interface as previously described. The palm image is also obtained through the smartphone application, where 1,034 features are extracted using various color and

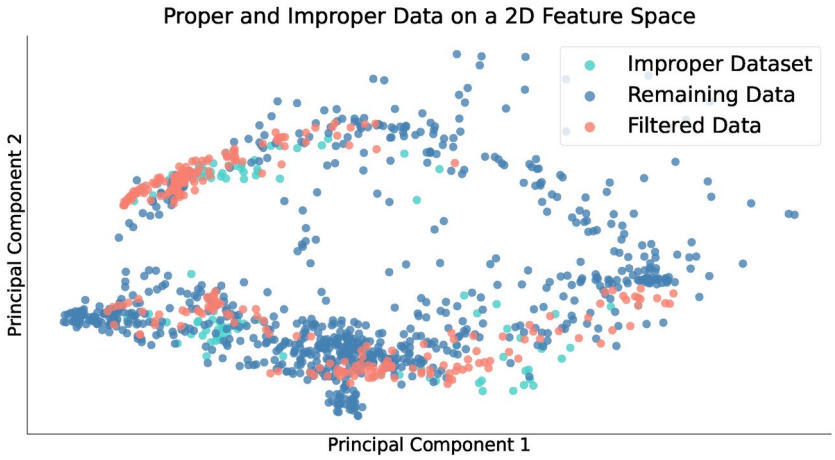


Fig. 6. The improper use dataset and the main dataset are plotted after applying PCA in a 2D feature space. The points in light blue are the noisy data that we collected in the improper use dataset and the red points are filtered out, leaving the dark blue data points in the main dataset. PC1 and PC2 denote the top-2 principal components.

texture spaces. The sensor values are obtained by emitting light from three individual laser diodes using 100 intensity configurations and two polarization states, resulting in a total of 600 raw values. Additionally, nine features are calculated from the raw sensor readings. Combining these three types of data results in a total of 1,657 features.

Given the relatively large feature dimension compared to the dataset size, we employ unsupervised feature selection using PCA for continuous features, effectively reducing dimensionality from 1,657 to 300 dimensions. To further refine feature selection, we select the most important features by filtering out those with the importance score lower than 0.01 using XGBoost.

Machine learning model Four sensing boxes are manually produced for the clinical study, but the slight differences in the sensing boxes have caused the collected data points to reside in distinct feature spaces. Observing varying degrees of similarity among different sensing boxes, we perform stratified training by grouping data points from sensing box 1 and 2 into one group, and sensing box 3 and 4 into another group. During training, we separately train a model for each group and select the appropriate model during inference.

We explore the performance of a diverse range of models, spanning from simple ridge regressors to sophisticated neural networks. Ultimately, we employ the Bayesian Rigid Regression model with hyperparameters $\alpha_1 = 2.5 \times 10^{-6}, \alpha_2 = 2.5 \times 10^{-6}, \lambda_1 = 5 \times 10^{-7}, \lambda = 5 \times 10^{-7}$, which controls the shape for the prior Gamma distribution. For the data selection model, we utilize a gradient boosting tree regressor with 0.1 learning rate, 80 estimators, and the MSE (mean squared errors) loss. Features are whitened for all non-categorical features, while categorical features are transformed into binary discrete values. This strategic approach contributes to the robustness and effectiveness of our system, particularly in the face of data limitations.

User experience study After the clinical study, we asked the participants if they have any comments and suggestions for improving the system. In general, the participants think it was easy and comfortable to use. It becomes straightforward once the scanning procedure is explained and demonstrated clearly at the beginning. A minor inconvenience of the system is the occasional disconnection of the sensing box with the smartphone, which requires the users to turning both devices off and on. The participants like the portability of the device, which can be carried around. The proposed system is preferred over the finger prick tests because of its noninvasive nature, but the current form factor is still larger than the CGMs, which the participants favor more. Overall, the participants think the device would be useful for glucose management, and they look forward to the future version with a wearable form factor.

Data availability

The datasets generated during and/or analyzed during the current study are available in the Zenodo repository: https://zenodo.org/doi/10.5281/zenodo.13138650.

Received: 31 July 2024; Accepted: 28 February 2025
Published online: 14 March 2025

References

1. Centers for Disease Control and Prevention. *National Diabetes Statistics Report*. https://www.cdc.gov/diabetes/php/data-research/index.html (2021). Accessed 30 July 2024.
2. ElSayed, N. A. et al. 10. cardiovascular disease and risk management: Standards of care in diabetes-2023. *Diabetes Care* **46**, S158–S190 (2023).

3. Group D. R. The diabetes control and complications trial (dcct): Design and methodologic considerations for the feasibility phase. *Diabetes* **35**, 530–545 (1986).
4. ElSayed, N. A. et al. Facilitating positive health behaviors and well-being to improve health outcomes: Standards of care in diabetes-2023. *Diabetes Care* **46**, S68–S96 (2023).
5. Forlenza, G. P., Kushner, T., Messer, L. H., Wadwa, R. P. & Sankaranarayanan, S. Factory-calibrated continuous glucose monitoring: How and why it works, and the dangers of reuse beyond approved duration of wear. *Diabetes Technol. Ther.* **21**, 222–229 (2019).
6. Kim, J., Campbell, A. S. & Wang, J. Wearable non-invasive epidermal glucose sensors: A review. *Talanta* **177**, 163–170 (2018).
7. Waghule, T. et al. Microneedles: A smart approach and increasing potential for transdermal drug delivery system. *Biomed. Pharm.* **109**, 1249–1258 (2019).
8. Christiansen, M. et al. A new-generation continuous glucose monitoring system: Improved accuracy and reliability compared with a previous-generation system. *Diabetes Technol. Ther.* **15**, 881–888 (2013).
9. Han, G. et al. Noninvasive blood glucose sensing by near-infrared spectroscopy based on plsr combines sae deep neural network approach. *Infrared Phys. Technol.* **113**, 103620 (2021).
10. Srichan, C. et al. Non-invasively accuracy enhanced blood glucose sensor using shallow dense neural networks with nir monitoring and medical features. *Sci. Rep.* **12**, 1769 (2022).
11. Pleus, S. et al. Proof of concept for a new raman-based prototype for noninvasive glucose monitoring. *J. Diabetes Sci. Technol.* **15**, 11–18 (2021).
12. Pai, P. P. et al. Cloud computing-based non-invasive glucose monitoring for diabetic care. *IEEE Trans. Circuits Syst. I Regul. Pap.* **65**, 663–676 (2017).
13. Li, T. et al. Noninvasive glucose monitoring using polarized light. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 544–557 (2020).
14. Lin, P.-H., Sheu, S.-C., Chen, C.-W., Huang, S.-C. & Li, B.-R. Wearable hydrogel patch with noninvasive, electrochemical glucose sensor for natural sweat detection. *Talanta* **241**, 123187 (2022).
15. Li, Q.-F., Chen, X., Wang, H., Liu, M. & Peng, H.-L. Pt/mxene-based flexible wearable non-enzymatic electrochemical sensor for continuous glucose detection in sweat. *ACS Appl. Mater. Interfaces* **15**, 13290–13298 (2023).
16. Asaduzzaman, M. et al. A hybridized nano-porous carbon reinforced 3d graphene-based epidermal patch for precise sweat glucose and lactate analysis. *Biosens. Bioelectron.* **219**, 114846 (2023).
17. Elsherif, M., Hassan, M. U., Yetisen, A. K. & Butt, H. Wearable contact lens biosensors for continuous glucose monitoring using smartphones. *ACS Nano* **12**, 5452–5462 (2018).
18. Kim, S., Jeon, H.-J., Park, S., Lee, D. Y. & Chung, E. Tear glucose measurement by reflectance spectrum of a nanoparticle embedded contact lens. *Sci. Rep.* **10**, 8254 (2020).
19. Zhou, F. et al. Flexible electrochemical sensor with fe/co bimetallic oxides for sensitive analysis of glucose in human tears. *Anal. Chim. Acta* **1243**, 340781 (2023).
20. Diouf, A., Bouchikhi, B. & El Bari, N. A nonenzymatic electrochemical glucose sensor based on molecularly imprinted polymer and its application in measuring saliva glucose. *Mater. Sci. Eng. C* **98**, 1196–1209 (2019).
21. Hu, S. et al. Enzyme-free tandem reaction strategy for surface-enhanced raman scattering detection of glucose by using the composite of au nanoparticles and porphyrin-based metal-organic framework. *ACS Appl. Mater. Interfaces* **12**, 55324–55330 (2020).
22. Leung, H. M. C., Forlenza, G. P., Prioleau, T. O. & Zhou, X. Noninvasive glucose sensing in vivo. *Sensors* **23**, 7057 (2023).
23. Li, N. et al. A noninvasive accurate measurement of blood glucose levels with raman spectroscopy of blood in microvessels. *Molecules* **24**, 1500 (2019).
24. Aloraynan, A., Rassel, S., Kaysir, M. R. & Ban, D. Dual quantum cascade lasers for noninvasive glucose detection using photoacoustic spectroscopy. *Sci. Rep.* **13**, 7927 (2023).
25. Park, E.-Y., Baik, J., Kim, H., Park, S.-M. & Kim, C. Ultrasound-modulated optical glucose sensing using a 1645 nm laser. *Sci. Rep.* **10**, 13361 (2020).
26. Pors, A. et al. Accurate post-calibration predictions for noninvasive glucose measurements in people using confocal raman spectroscopy. *ACS Sens.* **8**, 1272–1279 (2023).
27. Bandodkar, A. J. et al. Tattoo-based noninvasive glucose monitoring: A proof-of-concept study. *Anal. Chem.* **87**, 394–398 (2015).
28. De la Paz, E. et al. Extended noninvasive glucose monitoring in the interstitial fluid using an epidermal biosensing patch. *Anal. Chem.* **93**, 12767–12775 (2021).
29. Hakala, T. A. et al. Sampling of fluid through skin with magnetohydrodynamics for noninvasive glucose monitoring. *Sci. Rep.* **11**, 7609 (2021).
30. Ash, C., Dubec, M., Donne, K. & Bashford, T. Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods. *Lasers Med. Sci.* **32**, 1909–1918 (2017).
31. Chen, J., Arnold, M. A. & Small, G. W. Comparison of combination and first overtone spectral regions for near-infrared calibration models for glucose and other biomolecules in aqueous solutions. *Anal. Chem.* **76**, 5405–5413 (2004).
32. Hina, A. & Saadeh, W. A noninvasive glucose monitoring soc based on single wavelength photoplethysmography. *IEEE Trans. Biomed. Circuits Syst.* **14**, 504–515 (2020).
33. Gupta, S. S., Kwon, T.-H., Hossain, S. & Kim, K.-D. Towards non-invasive blood glucose measurement using machine learning: An all-purpose ppg system design. *Biomed. Signal Process. Control* **68**, 102706 (2021).
34. Lee, P.-L., Wang, K.-W., Hsiao, C.-Y. A non-invasive blood glucose estimation system using dual-channel ppgs and pulse-arrival velocity. *IEEE Sens. J.* (2023).
35. Hammour, G. & Mandic, D. P. An in-ear ppg-based blood glucose monitor: A proof-of-concept study. *Sensors* **23**, 3319 (2023).
36. Nader, E. et al. Blood rheology: Key parameters, impact on blood flow, role in sickle cell disease and effects of exercise. *Front. Physiol.* **10**, 1329 (2019).
37. Li, J. et al. An approach for noninvasive blood glucose monitoring based on bioimpedance difference considering blood volume pulsation. *IEEE Access* **6**, 51119–51129 (2018).
38. Pedro, B. G., Marcôndes, D. W. C. & Bertemes-Filho, P. Analytical model for blood glucose detection using electrical impedance spectroscopy. *Sensors* **20**, 6928 (2020).
39. Obaid, S. M., Elwi, T. & Ilyas, M. Fractal minkowski-shaped resonator for noninvasive biomedical measurements: Blood glucose test. PIER (2021)
40. Hanna, J. et al. Wearable flexible body matched electromagnetic sensors for personalized non-invasive glucose monitoring. *Sci. Rep.* **12**, 14885 (2022).
41. Zhang, M. et al. Microfluidic microwave biosensor based on biomimetic materials for the quantitative detection of glucose. *Sci. Rep.* **12**, 15961 (2022).
42. Deurenberg, P., Weststrate, J., Paymans, I. & Van der Kooy, K. Factors affecting bioelectrical impedance measurements in humans. *Eur. J. Clin. Nutr.* **42**, 1017–1022 (1988).
43. Windholz, M. The merck index an encyclopedia of chemical and drugs. In *The Merck Index an Encyclopedia of Chemical and Drugs*, 15–1313 (publisherMerck & Co., Inc., 1976).
44. Cameron, B.D., Gorde, H., Cote, G.L. Development of an optical polarimeter for in-vivo glucose monitoring. In *Optical Diagnostics of Biological Fluids IV*, vol. 3599, 43–49 (SPIE, 1999).

45. Malik, B.H., Coté, G.L. Real-time dual wavelength polarimetry for glucose sensing. In *Optical Diagnostics and Sensing IX*, vol. 7186, 7–11 (SPIE, 2009).
46. Jacques, S. L., Lee, K. & Ramella-Roman, J. C. Scattering of polarized light by biological tissues. In *Saratov Fall Meeting'99: Optical Technologies in Biophysics and Medicine*, vol. 4001, 14–28 (SPIE, 2000).
47. Pearson, K. Liii. on lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **2**, 559–572 (1901).
48. Chen, T., Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794 (2016).
49. Parkes, J. L., Slatin, S. L., Pardo, S. & Ginsberg, B. H. A new consensus error grid to evaluate the clinical significance of inaccuracies in the measurement of blood glucose. *Diabetes Care* **23**, 1143–1148 (2000).
50. Kovatchev, B. P., Patek, S. D., Ortiz, E. A. & Breton, M. D. Assessing sensor accuracy for non-adjunct use of continuous glucose monitoring. *Diabetes Technol. Ther.* **17**, 177–186 (2015).
51. Ekhlaspour, L. et al. Comparative accuracy of 17 point-of-care glucose meters. *J. Diabetes Sci. Technol.* **11**, 558–566 (2017).
52. Farage, M. A., Miller, K. W., Elsner, P. & Maibach, H. I. Structural characteristics of the aging skin: a review. *Cutan. Ocul. Toxicol.* **26**, 343–357 (2007).
53. Matabuena, M. et al. Reproducibility of continuous glucose monitoring results under real-life conditions in an adult population: a functional data analysis. *Sci. Rep.* **13**, 13987 (2023).
54. Sim, J. Y., Ahn, C.-G., Jeong, E.-J. & Kim, B. K. In vivo microscopic photoacoustic spectroscopy for non-invasive glucose monitoring invulnerable to skin secretion products. *Sci. Rep.* **8**, 1059 (2018).
55. Kirillov, A. et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4015–4026 (2023).
56. Yang, J. C. et al. Electronic skin: recent progress and future prospects for skin-attachable devices for health monitoring, robotics, and prosthetics. *Adv. Mater.* **31**, 1904765 (2019).
57. Yasui, T., Tohno, Y. & Araki, T. Characterization of collagen orientation in human dermis by two-dimensional second-harmonic-generation polarimetry. *J. Biomed. Opt.* **9**, 259–264 (2004).
58. Anderson, R., Parrish, J. Optical properties of human skin. *The Science of Photomedicine*, 147–194 (1982).
59. Milburn, P. B., Sian, C. S. & Silvers, D. N. The color of the skin of the palms and soles as a possible clue to the pathogenesis of acral-lentiginous melanoma. *Am. J. Dermatopathol.* **4**, 429–434 (1982).
60. Wood, M.F., Ghosh, N., Guo, X., Vitkin, I. A. Towards noninvasive glucose sensing using polarization analysis of multiply scattered light. In *Handbook of Optical Sensing of Glucose in Biological Fluids and Tissues* **12** (2008).

Acknowledgements
This work is supported by the National Science Foundation under SenSE-2037267.

Author contributions
H.L. designed and fabricated the hardware. H.L. implemented the Android application. L.G. and E.F. performed the user study. H.L. and C.G. designed and implemented the feature extraction process. H.L. and C.G. performed the data preprocessing and data analysis. All authors contributed to the preparation of the manuscript.

Declarations

Competing interests
Forlenza conducts research supported by Medtronic, Dexcom, Abbott, Insulet, Tandem, Beta Bionics, and Lilly and has been a speaker/consultant/ad board member for Medtronic, Dexcom, Abbott, Insulet, Tandem, Beta Bionics, and Lilly. All other authors declare no competing interests.

Ethical approval
All methods and protocols included in this study are approved by the Institutional Review Board of Barbara Davis Center for Diabetes at University of Colorado Anschutz Medical Campus (COMIRB #22-0074).

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-92515-6.

Correspondence and requests for materials should be addressed to H.M.C.L.