GlucoSense: Non-Invasive Glucose Monitoring using Mobile Devices

Neha Sharma[1], Mariam Bebawy[1], Yik Yu Ng[2], Mohamed Hefeeda[1,3]
[1]Simon Fraser University, Burnaby, BC, Canada
[2]McGill University, Montreal, QC, Canada
[3]Qatar Computing Research Institute, Doha, Qatar

ABSTRACT

Regular glucose monitoring is crucial for diabetic patients to avoid the risk of health complications such as stroke, kidney failure, heart disease, and even death. Most current devices for measuring glucose are costly and painful. We propose GlucoSense, a non-invasive glucose sensing solution on mobile devices. GlucoSense builds on the fact that glucose is an optically active molecule, which interacts with various wavelengths. We first conduct spectral analysis to demonstrate the feasibility of measuring glucose in the visible and near-infrared range (400–1000 nm), which is the range available on mobile devices. We also identify the relative importance of various spectral bands in this range. We further propose multiple practical designs for obtaining the required spectral bands for measuring glucose. We then design GlucoSense exploiting the sensing capabilities of modern smartphones combined with machine learning models. We conduct an ethics-approved user study with a diverse set of participants in terms of age, sex, ethnicity, and body mass index (BMI). We compare GlucoSense against a widely-used, FDA-approved glucose measuring device. Our results show that 80.4% of GlucoSense predictions are within Zone A (clinically accurate), and the remaining 19.3% are in Zone B (clinically acceptable) of the Clarke Error Grid (CEG). In addition, 99.7% of the predictions are within the None and Slight risk zones of the Surveillance Error Grid (SEG), indicating their high accuracy. Both CEG and SEG are standard metrics for assessing glucose-measuring devices. These results were obtained by GlucoSense running on unmodified phones in realistic environments with diverse illuminations.

ACM MOBICOM '25, November 4–8, 2025, Hong Kong, China

ACM ISBN 979-8-4007-1129-9/2025/11…$15.00
https://doi.org/10.1145/3680207.3723472

CCS CONCEPTS

• **Human-centered computing** → *Ubiquitous and mobile computing*.

KEYWORDS

Blood Glucose, Mobile Health, Hyperspectral Imaging

ACM Reference Format:
Neha Sharma[1], Mariam Bebawy[1], Yik Yu Ng[2], Mohamed Hefeeda[1,3]. 2025. GlucoSense: Non-Invasive Glucose Monitoring using Mobile Devices. In *The 31st Annual International Conference on Mobile Computing and Networking (ACM MOBICOM '25), November 4–8, 2025, Hong Kong, China.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3680207.3723472

1 INTRODUCTION

Diabetes has become one of the leading causes of death globally [64]. According to the World Health Organization (WHO) [65], currently, there are around 422 million cases of diabetes in the world. Diabetes is caused by abnormal insulin levels in the body due to either the pancreas not producing enough insulin or the body cells not using it adequately. Insulin is a hormone that regulates glucose level by allowing cells to absorb it from the bloodstream to obtain energy or store it for future use. If the glucose level in the blood remains low or high for long periods, it could cause hypoglycemia or hyperglycemia, respectively, leading to severe medical conditions, including tissue damage, stroke, kidney failure, blindness, heart disease, and death if left untreated [22, 32].

Deficient production of insulin leads to diabetes Type 1, which is characterized by sudden drops in glucose levels. On the other hand, the ineffective use of insulin leads to diabetes Type 2, which is characterized by high levels of glucose. Both conditions do not have a cure and thus require regularly measuring glucose; at least four times a day [15] and up to ten times for patients with severe conditions [19].

Unfortunately, regularly checking blood glucose for most diabetic people is painful or at least inconvenient. Conventional devices for glucose monitoring use electrochemical methods, which require a small amount of blood to be drawn out of fingertips using automatic lancet devices [25]. Although accurate, this method is invasive and painful for patients to repeat multiple times daily. Alternate devices to

measure glucose are called continuous glucose monitoring (CGM) systems [21], which estimate glucose concentration in the interstitial fluid [34]. These systems contain biosensors with micro-needles that penetrate the skin. CGM systems provide glucose readings every 1–5 minutes but need to be replaced every 10–14 days. Both traditional and CGM systems cause discomfort and pose risks of potential infection and tissue damage [9]. Therefore, there is a need for non-invasive and cost-effective systems for measuring glucose to help millions of patients worldwide.

In this paper, we consider estimating glucose levels *using only smartphones*. This is a challenging research problem for multiple reasons. First, the sensing capabilities of regular smartphone (RGB) cameras are limited to the visible range in the electromagnetic spectrum. Whereas the most promising range to get information from deeper skin layers, where glucose can be measured, lies in the near-infrared (NIR) part of the spectrum. Second, the human skin is a highly absorbing and scattering medium containing many substances, e.g., collagen and elastin, which negatively interfere with the reflected signal. Third, people have diverse skin characteristics, e.g., skin tone and thickness, which affect the correlation between the reflected signal and glucose levels.

To address these challenges, we propose GlucoSense, a non-invasive glucose monitoring system on smartphones. As shown in Figure 1, GlucoSense has three main components. The first component comprises a mobile sensing module in the visible and near-infrared (VNIR) range. We propose using near-infrared signals captured by *depth-sensing* cameras on modern smartphones and regular RGB cameras. We demonstrate the potential of different depth-sensing technologies in modern smartphones, including Time-of-Flight, stereo, and structured light. This addresses the first challenge mentioned above. The second component is a deep learning model that converts the captured RGB and NIR signals to multiple (hyperspectral) bands in the VNIR (400–1000 nm) range. We analyze the importance of individual bands for measuring glucose, and we identify the crucial bands considering the diversity of people and the complexity of their skin tissues, which addresses the second and third challenges. The third component of GlucoSense is an estimation module that maps the recovered spectral information to glucose levels.

The contributions of this paper are as follows:

- We conduct spectral analysis of multiple subjects with and without diabetes using a hyperspectral camera in §4. Our analysis shows the potential of measuring glucose from the interstitial fluid in the skin using signals in the VNIR range, and it identifies the most important spectral bands to consider. This analysis is useful in its own right, especially for designers of devices that measure glucose.
- We propose *four* solutions for obtaining NIR signals on mobile devices in §5.2; two of them are readily available on recent phones, and the others are easily realizable.
- We design a machine learning model for estimating glucose levels, which is computationally efficient and does not require customization for different user groups in §5.3.
- We conduct a user study to demonstrate the accuracy of GlucoSense and compare it against a widely-used, FDA-approved glucose measuring device in §6. This study was approved by the Ethics Research Board of our institution. Our results show, for example, that 80.4% of GlucoSense predictions are within Zone A (clinically accurate), and the remaining 19.3% are in Zone B (clinically acceptable) of the Clarke Error Grid (CEG). In addition, 99.7% of our results are within the None and Slight risk zones of the Surveillance Error Grid (SEG). CEG and SEG are commonly used for evaluating glucose-measuring devices.

2 BACKGROUND AND RELATED WORK

Blood Glucose. Glucose is the main sugar found in our blood, and it is the primary source of energy [30, 43]. Our body breaks down the food we eat into glucose and releases it into our blood stream. Insulin, a hormone produced by the pancreas, is required to transport glucose from the blood stream to body cells. If the pancreas does not make enough insulin or our body cannot use the produced insulin, this leads to high glucose levels, resulting in diabetes. Prolonged high glucose levels can cause health complications such as stroke, kidney failure, heart disease, and even death. Thus, it is crucial for diabetic people to maintain their glucose levels within the normal range (70–180 mg/dL [15]).

Invasive and Minimally-Invasive Approaches. A common approach for measuring glucose is to draw a small blood sample from the fingertips. The blood sample is analyzed using a test strip, which contains a glucose oxidase enzyme that reacts to the glucose molecules and produces a proportional electric current [67]. The current is then mapped to readable glucose levels. Examples of commercial devices using this approach include Accu-Chek and True Metrix. This electrochemical approach provides accurate results, but it is invasive and uncomfortable.

CGM devices, on the other hand, periodically measure glucose every few minutes. These devices are based on the same electrochemical principle, but they measure the glucose level in the interstitial fluid (ISF) in the skin instead of the blood. ISF is primarily present in the lowermost skin layer of the dermis, which is 70% ISF by volume [50]. It has been shown that the ISF glucose level is highly correlated to blood glucose level but with an average lag time of 8–10 minutes [16]. CGM devices require inserting a sensor (tiny needle) into the skin, and thus, they are considered minimally invasive. These devices may cause skin irritation and discomfort

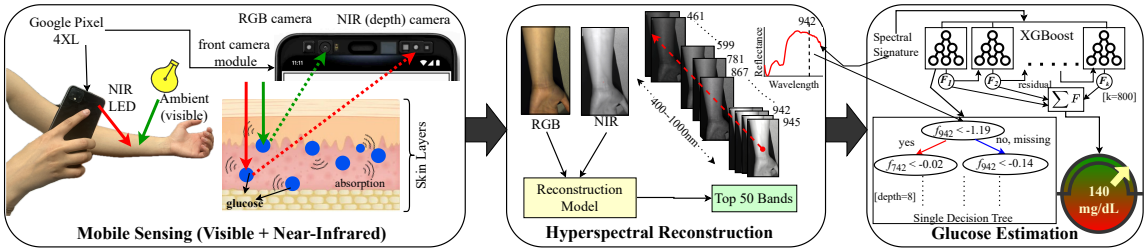


Figure 1: Overview of GlucoSense. Users take RGB and NIR images using their phones. A reconstruction model upscales these images across the spectral domain. The result is then fed to a glucose estimation model.

[59], and they may impose a financial burden on some patients since they need to be replaced every 10–14 days; the FreeStyle Libre device used in our study costs $130/unit.

Non-Invasive Approaches. Non-invasive glucose monitoring systems include transdermal, thermal, and optical techniques; comprehensive reviews can be found in [25, 37]. Transdermal techniques measure glucose molecules' electrical impedance, but they may cause skin irritation while passing the electric current through tissues. Thermal techniques measure heat generation, blood flow rate, and hemoglobin/oxyhemoglobin concentrations to estimate glucose levels, but the accuracy of these systems is significantly affected by environmental factors like temperature and humidity. Transdermal and thermal techniques are not practical to realize or approximate on smartphones.

Optical techniques include optical coherence tomography (OCT), polarimetry [39], photoplethysmography (PPG) [58], and spectroscopy [38]. All of them require special devices and sophisticated hardware setups. Among the optical techniques, spectroscopy is the most studied, and it has four main categories: Raman [38], photoacoustic [54], near-infrared (NIR) [28, 33], and mid-infrared (MIR) [40]. Raman spectroscopy measures the energy shift of lasers used to excite glucose molecules. Photoacoustic spectroscopy measures the absorption of glucose by using acoustic signals. NIR and MIR spectroscopy measure the reflectance from glucose across multiple wavelengths in different parts of the spectrum.

Multiple attempts have been made to build low-cost versions of the above spectroscopy systems, e.g., using small LEDs and photodiodes [45, 46, 57, 58, 66]. Most systems use a single wavelength to simplify the design. However, relying on a single wavelength may yield inaccurate measurements for people with different skin tones and thicknesses [11].

As analyzed in [20], wavelength selection for glucose measurements remains a challenge. Various works have used different wavelengths, some prioritized wavelengths that penetrate further in the skin, and others considered water absorption of wavelengths. Further, none of the prior works rigorously analyzed the impact of *individual* wavelengths in the 400–1000 nm range and *their combinations* on the accuracy of measuring glucose. The 400–1000 nm range is crucial because it is the sensitivity range of the RGB camera sensors on smartphones. Our analysis in §4.2 addresses this problem.

Measuring Blood Glucose on Smartphones. Multiple works considered employing smartphones in measuring glucose, including [27, 61]. GlucoScreen [61], for example, introduces a glucose test strip that works with smartphones, alleviating the need for specialized readers and reducing the cost. GlucoScreen, however, is still an invasive approach as it requires blood samples. SugarMate [27], on the other hand, is non-invasive. It builds a deep recurrent neural network model to predict blood glucose based on several factors entered by users and collected by the phone. Specifically, SugarMate requires users to *manually* input their daily food, drug, and insulin intake, which is tedious and error-prone.

The goal of this work is to build a glucose-measuring solution on unmodified smartphones. It also conducts a detailed analysis to rank wavelengths based on their importance in measuring glucose, which provides a systematic method for designing future glucose measurement systems.

3 FOUNDATIONS AND CHALLENGES

We discuss the principles and challenges of measuring blood glucose using spectral analysis on regular phones.

Light Transport Theory and Measuring Glucose. Glucose is an optically active molecule with a chemical composition of $C_6H_{12}O_6$. When light interacts with a skin tissue containing glucose, the covalent bonds (C-H, O-H) inside the glucose molecules vibrate. This causes attenuation of the incident light due to scattering and absorption. A widely accepted model to quantify the interaction of glucose with different light wavelengths can be summarized by the light transport theory using the following equation [20]:

$$I = I_0 e^{\mu L}, \tag{1}$$

where I is the reflected light intensity, I_0 is the incident light intensity, and L is the optical path length inside the tissue. Attenuation of light inside the tissue depends on the attenuation coefficient μ, which is the sum of the absorption coefficient μ_a and the scattering coefficient μ_s' [20]. μ_a depends on the wavelength of the used light and the glucose concentration. Thus, the reflected light, which is inversely proportional

to the absorption, can determine glucose concentration in skin tissue. This, however, faces multiple challenges, which we describe in the following.

Challenge 1: Complexity of Human Skin. We provide a simplified illustration of human skin in Figure 2. The outermost layer is the epidermis (about 0.1 mm thick), which provides a waterproof barrier and creates skin tones based on different melanin concentrations. Beneath the epidermis is the dermis (about 2 mm thick), which contains connective tissues and sweat glands. It contains the interstitial fluid surrounding the cells, which is used by CGM devices such as [7]. The deepest layer is the hypodermis (2–7 mm thick), which mainly consists of fat tissues that store energy and nutrients, providing insulation from cold temperatures and protection from injuries [10]. As Figure 2 shows, the human skin is a complex and highly absorbing medium. It scatters most of the incident light and thus decreases the signal-to-noise (SNR) of the reflected light. This makes it harder to capture the small variations in the reflected signal caused by the presence of glucose deep in the skin tissues.

Challenge 2: Diversity of Humans. Humans are quite diverse along many dimensions, including skin tone, sex, and weight. This diversity presents a major challenge for developing a scalable, non-invasive glucose monitoring solution. For example, among different users, the optical path L in Equation (1) varies due to different skin thicknesses, which changes the amount of reflected light even for the same glucose concentration, making it harder to estimate glucose concentrations accurately. The optical path also depends on the wavelength used since the penetration depth of light in skin tissues varies across the electromagnetic spectrum. Multiple studies have shown that the penetration depth generally increases with increasing the wavelength [13, 41]. Skin thickness and skin tone (melanin concentration) both affect the ratio of the absorbed/reflected light. Existing systems (e.g., CGM devices) generally require calibration using blood tests to improve the estimation across different users.

We propose using multiple wavelengths in the whole 400–1000 nm range to address the challenges of human diversity and skin complexity. Multiple wavelengths can reveal information from different layers of the skin, establishing a stronger correlation between glucose concentration and the reflected light intensity. In §4.1, we first analyze the correlation between different wavelengths and changes in glucose levels in controlled settings. This is done to establish the feasibility of our approach. We then recruit a diverse set of participants (31 in total) with and without diabetes, considering different age groups, skin tones, biological sexes, and Body Mass Indexes (BMIs), to understand the impact of different wavelengths on measuring various glucose concentrations. We present a systematic way to select the most important wavelengths for estimating glucose. This analysis also helps in designing future glucose-measuring devices.

Challenge 3: Lack of Information in the Infrared Range. Glucose is an optically active molecule, and according to [53], it has shown strong absorption characteristics in the NIR (700–1000 nm) range. Although the CMOS sensors of RGB cameras on smartphones have spectral response in the VNIR (400–1000 nm) range, they utilize cut-off filters to truncate all signals beyond 700 nm. This is done to improve the visual quality of RGB images. To address this problem, we propose different camera solutions that use NIR signals and RGB images to extend the sensing capabilities of smartphones. Specifically, we propose using off-the-shelf depth sensing cameras (e.g., Time-of-Flight) available on modern smartphones and used in applications like face recognition and augmented reality. We also design camera systems using infrared-enabled cameras that are widely used for surveillance. We present the details of our designs in §5.2.

Challenge 4: Limited Number of Captured Wavelengths. Regular RGB cameras capture only three channels (or bands). While these RGB bands enable humans to perceive the captured scene, they are insufficient to conduct spectral analysis of different skin layers to estimate glucose levels. Spectral analysis is typically performed by expensive hyperspectral cameras or spectrometers that capture many (200+) equally-spaced, narrow bands in the entire VNIR spectral range. To address this problem, in §5.3, we leverage a deep-learning model to convert RGB and NIR signals into multiple narrow bands, which allows the creation of accurate *spectral signatures* essential for assessing glucose levels.

4 SPECTRAL ANALYSIS AND BAND SELECTION

In this section, we first conduct controlled experiments to demonstrate the feasibility of measuring different glucose levels using a hyperspectral camera operating in the 400–1000 nm range. This is unlike many prior works, e.g., [62], which used cameras operating in ranges not available on smartphones. Then, we analyze the relative importance of bands for glucose measurements.

4.1 Spectral Analysis to Measure Glucose

We analyze different glucose concentrations in *water* using a high-end hyperspectral camera. We note that this study is difficult to do with humans, as the glucose level may not change quickly, and it is hard to cover the entire glucose range. Nonetheless, studying glucose concentrations in water provides good approximations of glucose concentrations in blood and interstitial fluid in the skin, as water is the dominant component in both.

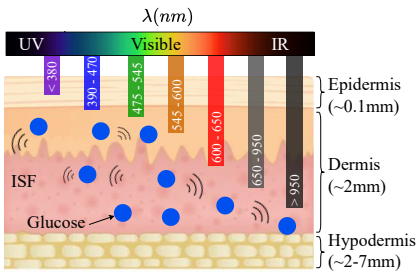


Figure 2: Penetration of wavelengths in skin tissues.

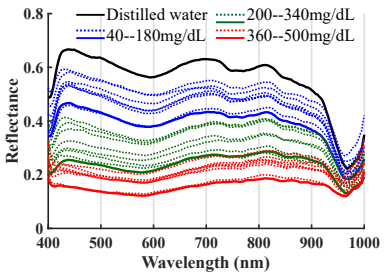


Figure 3: Measuring glucose using spectral analysis.

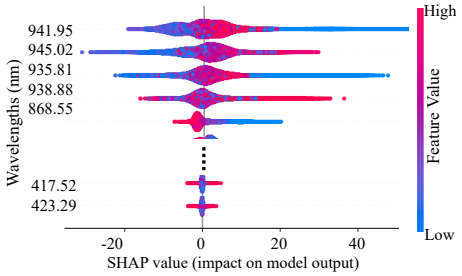


Figure 4: Relative importance of spectral bands.

We prepared glucose solutions using distilled water as a dissolving medium for the dextrose powder (pure form of sugar) to achieve concentrations ranging from 40 to 500 mg/dL with 20 mg/dL increments, which cover the entire practical glucose range. In total, we prepared 25 samples in 30 ml glass bottles, one of which was plain water.

Our experimental setup consists of the hyperspectral camera (model: Specim IQ and costs about $35K), a halogen light source, and 25 bottles with different glucose concentrations. The camera captures 204 wavelengths in the 400–1000 nm range, each with a spatial resolution of 512 × 512 pixels. The halogen light source is recommended by the camera manufacturer because it emits power across the 400–1000 nm range, unlike common sources such as LED and fluorescent. We captured a hyperspectral image for each glucose concentration and computed a *spectral signature* for it. The spectral signature is the normalized reflectance across all 204 wavelengths. It is computed per pixel and typically averaged across multiple pixels in a small square area of the captured scene. In our experiments, the signatures are averaged across all pixels in 8 × 8 pixel areas.

We plot the spectral signatures of all 25 samples in Figure 3. The black curve represents the spectral signature of pure distilled water. To facilitate visualizing the results, we divide and color code the glucose concentrations into three ranges: 40–180 (blue), 200–340 (green), and 360–500 mg/dL (red). We draw the curve corresponding to the highest glucose concentration in each range as a solid line, while others are dotted. The results in Figure 3 indicate that as the glucose concentration increases, the absorption across different bands increases, and the corresponding reflectance decreases with respect to pure water. Thus, spectral analysis is a *potential* solution for detecting glucose concentrations.

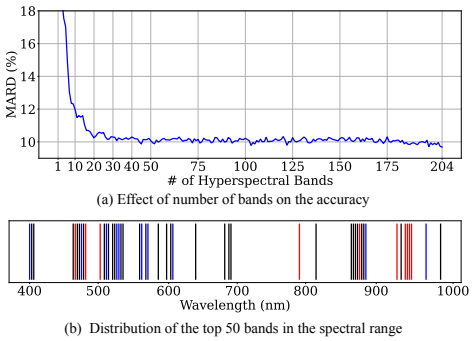
We note that the human skin is much more complex than simple glucose solutions in transparent bottles. Therefore, a detailed analysis of how various wavelengths contribute to measuring glucose in human skin is needed, which we conduct in the next section.

4.2 Band (Wavelength) Selection

Relative Importance of Bands. We first analyze the relative importance of spectral bands on the accuracy of measuring glucose from the ISF in human skin. We use a subset of the data collected during our user study (detailed in §6.1). This dataset has many reference glucose readings measured by a CGM device from 31 diverse participants over multiple weeks. Each reading is paired with a hyperspectral image taken at the same time for the participant's inner wrist.

We implemented a supervised learning model to map the spectral signatures computed from the hyperspectral images to their corresponding reference glucose readings. We designed gradient-boosting decision trees to capture the relationship between each spectral band and the glucose reading. To analyze each band's contribution in predicting glucose, we use the SHAP (SHapley Additive exPlanations) [42]. SHAP uses a game theoretic approach that measures each player's contribution to the outcome. Our model treats each spectral band as a player/feature contributing to the output (glucose estimation). Specifically, we compute a band's contribution (SHAP value) by first calculating the average prediction of the model across the entire dataset, referred to as the baseline. Then, we compute the marginal contribution of the band by considering all possible subsets of bands. This involves adding/removing the band to a given subset and calculating the change in prediction from the baseline. Then, we average these changes across all possible subsets. We note that there are 204! subsets in each case, which is infeasible to try. We utilize the approximation method in [42] to explore the relevant subset of bands. The run time for each case is several hours on a decent workstation. We repeat this computationally expensive process for each of the 204 bands captured by the hyperspectral camera in the 400–1000 nm range.

Using these experiments, we compute a ranking for each of the 204 bands based on how much the model performance gets affected by eliminating that particular band. For illustration, we show the top five and bottom two bands in Figure 4. The x-axis denotes the SHAP values, which measure how much a band contributes to pushing the model's prediction



(a) Effect of number of bands on the accuracy

(b) Distribution of the top 50 bands in the spectral range

Figure 5: Impact of the number of bands on the accuracy of glucose estimation and the locations of the most important bands in the 400-1000 nm range.

away from its expected value (the baseline). Dots clustered around zero reflect samples where the band has little impact, meaning the prediction stayed close to the baseline. Red (blue) dots mean that the band pushed the prediction higher (lower) than the baseline. To illustrate, consider the top 941.95 nm band. It has the largest mean SHAP value in the positive direction (>40) and is color-coded blue. This means when the reflectance value of this band is low, its contribution to glucose prediction is high.

Effect of Number of Bands. The above SHAP analysis provides a *relative* importance for each of the 204 spectral bands on the accuracy of predicting glucose. We further analyze the effect of the number of bands on the accuracy. In other words, we analyze the *marginal* contribution to the prediction accuracy as we successively consider more bands.

We train 204 glucose prediction models; the model's design is described in §5.4. Each prediction model takes a different number of spectral bands as input. The first model takes only the top 1 band (942 nm) according to the SHAP analysis. The second model takes the top 2 bands (942 and 945 nm), the third model takes the top 3 bands (942, 945, and 936 nm), and so on, until the last model takes all 204 bands as inputs. We assess the accuracy of the prediction model using the absolute relative difference (ARD) between the predicted and reference glucose values, normalized by the reference value. We plot the mean of the ARD (referred to as MARD) achieved by all models in Figure 5.a. The results indicate virtually no performance gain from bands beyond the top 50, as the MARD stabilizes around 10% for 50+ bands. In contrast, each of the top 10 bands provides a significant performance gain; in total, the MARD decreased from around 19% (for the top 1 band) to less than 12% (for the top 10 bands). Adding the next 10 bands reduced MARD by only 1–1.5%, while the following 30 bands provided tiny gains.

We present the distribution of the top 50 bands for glucose prediction across the 400–1000 nm range in Figure 5.b. For easier identification of the important bands and where they are in the spectral range, we color-code the bands as: (i) top 10: red, (ii) 11–20: blue, and (ii) 21–50: black. Two observations can be made on this figure. First, five spectral ranges can be identified, which are: [920–950], [850–880], [780–810], [450–480], and [550–600]. These ranges capture various information for glucose measurements. For example, the [920–950] range has the least water absorption and thus has the highest correlation to the glucose concentration. The [450–480] range captures the melanin information, which is crucial for handling user diversity. The [850-880], [780–810], and [550–600] ranges are important for light penetration in different levels of the dermis, which help in handling various skin thicknesses.

The second observation on Figure 5.b is that multiple bands are close to each other in the spectral range. For example, four of the top 10 bands (942, 945, 936, 939 nm) are all within 9 nm range. Further, the spectral distance between the neighboring bands is only 3 nm. This fine granularity is only possible with high-end spectral cameras, because their design includes complex optical elements and narrow band-pass filters. In contrast, mobile devices typically have simpler hardware and can only capture wider bands. Given these physical/practical limitations, our results in Figure 5.b can help in selecting the most important bands to capture in future devices designed for glucose measurements. For example, if a device would capture only five bands, our analysis indicates that a single band should be chosen from each of the above-mentioned spectral ranges to increase the system's robustness to various issues. If the device would capture more than five bands, the extra bands should first be allocated to the more important ranges (e.g., ones with more red lines).

Summary. Since the SHAP analysis is model agnostic [42] and rigorously examines all band combinations, the results in this section offer a systematic approach to rank spectral bands in order of importance for glucose prediction. This is useful for designing future glucose measurement devices with minimal hardware cost by selecting only the most essential bands. Our analysis also shows the distribution of bands across the 400–1000 nm range, which can provide further guidelines on selecting bands given the physical limitations of the glucose measurement devices compared to high-end hyperspectral cameras.

5 DESIGN OF THE PROPOSED SYSTEM

5.1 Overview

Main Components. An overview of GlucoSense is presented in Figure 1. It contains three main components. The first is a mobile sensing solution to capture visible and NIR signals on phones. For obtaining NIR signals on smartphones, we present *four* possible solutions in §5.2. The second component is a spectral reconstruction machine learning model

to convert captured sparse signals into rich spectral bands. This is presented in §5.3. The third component is a machine learning model to estimate glucose concentration from the reconstructed spectral signals, which is detailed in §5.4.

Operation. At a high level, GlucoSense operates as follows. The two machine learning models are first trained on a workstation. Then, they are uploaded to a phone. To estimate glucose, a user takes RGB and NIR images of the inner part of their wrist by the phone. These images are then upscaled to multiple spectral bands by the reconstruction model. Then, the reconstructed bands are fed to the glucose estimation model, which outputs the predicted value in mg/dL.

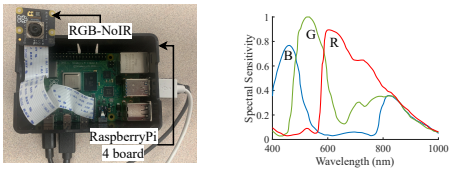
5.2 Infrared Sensing on Phones

We present four camera designs to obtain NIR signals on smartphones: (i) Using full spectrum RGB cameras with no IR filter (RGB-NoIR), (ii) Designing camera sensors with custom RGB-NIR filter (Custom), (iii) Using stereo NIR depth cameras (RGB+NIR), and (iv) Using the time-of-flight (ToF) sensors on modern smartphones (RGB+ToF). We analyze the pros and cons of each solution in the following, and we experimentally evaluate their performance in §6. We note that the RGB+NIR design is the only solution that works on current, unmodified phones. Thus, when we refer to the performance of GlucoSense, we mean the results achieved by the RGB+NIR design.

Using Full Spectrum RGB Cameras. Regular RGB cameras consists of an IR filter which is typically a thin film attached to the camera sensor. Removing the IR cutoff is a simple hardware change as shown in [5]. RGB sensors with no IR filters are called *full spectrum RGB*. These sensors are already used in surveillance cameras [2] for day/night video capturing and infrared photography. To realise our ideas for GlucoSense, we assemble an image capturing system using a *Raspberry Pi Camera Module 3 NoIR* [4], as shown in Figure 6, in combination with off-the-shelf low-powered 850*nm* LED (emission curve shown in Figure 7.a). We choose this illumination wavelength because the highest spectral response of the RGB-NoIR sensor is at $\approx$ 850*nm* in the NIR range.

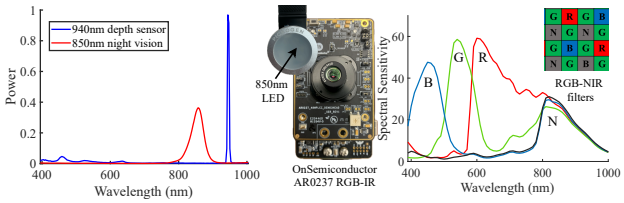
The advantages of using RGB-NoIR sensors are the simplicity and existence of commercial cameras without IR filters [4]. Also, this is a *single sensor* solution, where we capture visible and NIR signals on the same sensor. Removing the IR filter, however, damages the visual quality of regular images. Thus, this solution is more suitable for imaging systems designed especially for glucose monitoring or to bring this solution to wearable devices, e.g., smartwatches.

Designing Camera Sensors with Custom RGB-NIR Filter. Full spectrum RGB (NoIR) sensors capture NIR signals mixed with RGB bands (Figure 6.b). This provides limited information in the NIR range. Multiple alternative *single sensor* camera designs have been proposed in the literature to allow for capturing an *explicit* NIR band with minimal or no damage to the RGB image quality. This is realised by replacing the traditional Bayer color filter array with a custom-designed wavelength filter array that has an explicit filter for the NIR band as illustrated in Figure 7.b. A comprehensive performance analysis of various filter arrangements is presented in [44]. The authors demonstrate that the 4×4 filter pattern shown in Figure 7.b results in the best overall performance in terms of obtaining an explicit NIR band with no damage to the RGB bands.



(a) RGB-NoIR Camera (b) Sensitivity

Figure 6: Full spectrum RGB camera with no IR filter.



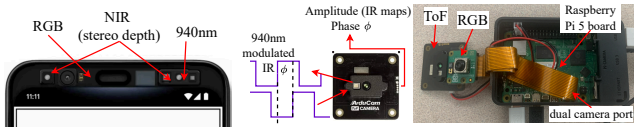
(a) NIR Illuminations (b) RGB-NIR Camera and its Sensitivity

Figure 7: Camera with custom RGB-NIR filter and separate NIR illuminations.

To evaluate GlucoSense, we assembled a camera system with a commercial custom sensor board, *Model AR0237 RGB-IR* from ON Semiconductor [3]. This sensor and its sensitivity are shown in Figure 7.b. The sensitivity was obtained from the manufacturer's data sheets. The manufacturer also provided us with its custom image acquisition and processing software. We capture raw RGB and NIR signals and then process them to get standard RGB and NIR bands. This imaging setup is also used with 850*nm* LED similar to the previous setup in combination with ambient light for visible range.

The advantages of using the custom RGB-NIR sensor include better spectral reconstruction compared to RGB-NoIR sensor and maintaining the quality of RGB images. This is also a *single* sensor approach but the disadvantages are the complexity and cost of manufacturing imaging sensors with custom filters. This solution can be useful for designing future smartphone cameras and specialized sensing systems for mobile health applications such as GlucoSense.

Using Stereo NIR Depth Cameras. Depth sensing technologies on smartphones have evolved significantly over the past few years. A depth sensor in a mobile camera is a component that measures the distance between the camera



(a) Google Pixel 4XL (b) RGB + Time-of-Flight (ToF)

Figure 8: Dual camera systems: (a) RGB+NIR (depth) and (b) RGB+ToF sensor.

and various objects in the scene. The depth information is valuable for several applications such as face recognition and augmented reality. Manufacturers use different depth sensing technology but most techniques use high power infrared LED (≈930– 980 nm) and an infrared camera to capture the scene. The depth sensing camera modules on latest smartphones provide an opportunity to get NIR signals without any hardware changes which is a major advantage. However, this is a *multi-sensor* approach where we get NIR signals from depth camera and RGB from regular RGB camera.

For evaluating GlucoSense, we use the front camera module of Google Pixel 4XL which consists of stereo vision depth sensing module [26]. As illustrated in Figure 8.a, we propose to use one NIR camera with in-built 940 nm LED (Figure 7.a) to get NIR signals and RGB signals from front facing RGB camera. Using a multi-sensor approach requires additional post processing to resize and align the images, as the two cameras produce different resolutions and are physically apart on the phone. In addition, the NIR camera is front-facing, which poses a slight inconvenience when capturing images. To address this issue, we design our mobile application, to capture images after a short timer (3 sec), where the user clicks on the start button and then turns the camera around to capture the RGB and NIR images.

Using Time-of-Flight (ToF) Sensors. ToF sensors can be categorized into two types: direct (dToF) and indirect (iToF). dToF directly calculates the time it takes for the emitted infrared signal to reach back to the sensor. Then, using the time and speed of light, it calculates depth maps. This approach is used in LiDAR sensors on recent Apple's iPhones 12 Pro and beyond. On the other hand, an iToF camera measures the phase offset between emitted and reflected light [47, 60] to calculate depth maps. A number of Android smartphones (e.g., Huawei P/Mate series, Samsung S/Note series [36, 51, 56]) are equipped with iToF cameras for various applications such as FaceID and AR/VR.

We can directly use IR signals from iToF cameras, which are generally present in the smartphone back camera module. The raw phase and amplitude components from iToF cameras are inaccessible on Android smartphones. So, we designed a dual-camera imaging system on Raspberry Pi using an off-the-shelf iToF sensor (Arducam [6]) and a regular RGB camera [4] as illustrated in Figure 8.b. This imaging system is used under ambient light and in-built 940 nm LED (Figure 7.a) already present in the iToF sensor. This is also a *multi-sensor* approach that can be used directly in smartphones without any hardware changes but will require post-processing based on the sensor characteristics of both RGB and iToF.

5.3 Spectral Reconstruction

The spectral reconstruction model in GlucoSense is built on top of the state-of-the-art MobiSpectral model [52], which extends MST++ [14] to reconstruct bands in the entire visible and NIR range. Specifically, MobiSpectral takes four bands (three RGB bands and one NIR band) and produces 204 bands equally distributed in the 400–1000 nm range. These bands were shown to be accurate relative to the actual bands captured by a hyperspectral camera. The reconstruction model is designed based on transformers, where each spectral feature is treated as a token, and self-attention is calculated along the spectral dimension.

In GlucoSense, we make minor changes to the MobiSpectral model. First, we do not reconstruct all bands. Instead, we reconstruct only the most important bands based on the analysis in §4.2. Further, these top bands are not equally spaced in the 400–1000 nm range, as shown in Figure 5. Reconstructing only the relevant bands reduces the computational resources for training the reconstruction model and improves its accuracy. In addition, we proposed four designs for obtaining NIR signals on phones in §5.2. Three of these designs (Custom, RGB+NIR, and ToF) produce four bands, which the reconstruction model can readily use. However, the fourth (RGB-NoIR) produces only three channels, which are the RGB channels mixed with NIR signals because the camera does not have an IR filter. In this case, we change the structure of the reconstruction model to take three channels. We provide more details on training the reconstruction model and demonstrate its accuracy in Appendix A.1.

5.4 Glucose Estimation

We design a supervised learning regression model to map the reconstructed spectral bands to glucose values. The model is based on XGBoost (extreme gradient boosting) [17], which is an optimized implementation of gradient boosting [31]. Gradient boosting sequentially builds an ensemble of weak learners, typically decision trees. Each tree is trained to correct the residual errors of the previous ensemble using gradient descent. At each step, a new decision tree is fit to the negative gradient of the loss function, and its predictions are scaled by a learning rate before being added to the overall model. XGBoost introduces several enhancements to gradient boosting. For example, it employs second-order optimization using both gradients and Hessians (second derivatives) to make more precise updates, leading to faster convergence. It also

includes various regularization methods, e.g., tree pruning and random subsampling, to prevent overfitting.

For our glucose estimation model, each tree in the ensemble strives to capture different aspects of the relationship between the spectral bands and glucose levels. The first tree may capture the strongest patterns, while subsequent trees focus on more subtle relationships and error correction. The learning rate of the model controls how each new tree contributes to the ensemble, providing a parameter to balance learning speed and stability. The model has 800 trees, providing substantial capacity to learn complex patterns.

In addition, we use the SHAP values, described in §4.2, to create an initial subset of spectral bands that have the strongest overall impact on the model's predictions across all samples. This allows us to reduce the dimensionality of the input data from 204 bands to the 50 most influential bands, which significantly accelerates training and improves accuracy. Once the subset of bands is selected, the model builds its trees independently using XGBoost's various optimization methods. Specifically, at each node in each tree, XGBoost decides which feature to split on by calculating the potential gain (improvement in accuracy) that would result from each possible split. This gain is based on the reduction in training loss that would occur after making that split. Therefore, while SHAP values determine which bands are available to the model, the actual order and frequency of band usage in the tree structure are determined by XGBoost's internal optimization process, which may differ from the original SHAP-based ranking. This two-step approach combines the benefits of SHAP-based feature selection with XGBoost's sophisticated tree-building algorithm. We provide more details and illustrations on the glucose estimation model in Appendix A.2. In addition, our preliminary experimentation evaluated various regression models, including SVM, KNN, PLSR, and neural networks. Our results, also presented in Appendix A.2, showed that the gradient boosting model provides higher accuracy and requires less computing resources for training and inference.

6 EVALUATION

In this section, we evaluate the performance of GlucoSense using a user study with diverse participants and compare it against an FDA-approved glucose-measuring device. Some results are presented in the appendix due to space limitations.

6.1 Description of the User Study

This user study has been approved by the Research Ethics Board of our university. Since it involves humans using medical devices, everybody involved in this research project had to go through rigorous training and national certification to ensure the safety, comfort, and rights of the participants. This is in addition to the usual measures for anonymizing, storing, and sharing the collected datasets. The approval process and experiments took multiple months to complete.

Participants. Table 1 summarizes the participants' demographic information. Specifically, we enrolled 31 participants between 18 and 59 years old. Six participants have diabetes (3 Type I and 3 Type II), and one is prediabetic. The remaining 24 participants are healthy with normal glucose levels. We also strived to ensure other aspects of participants' diversity, including body mass index (BMI) and ethnicity. BMI correlates with skin thickness, affecting the absorption and reflection of different wavelengths. We calculate the BMI for each participant based on their height and weight. Among the 31 participants, 19 are categorized as Healthy (BMI between 18.5 and 24.9), 9 as Overweight (BMI between 25 and 29.9), and 3 as Obese (BMI above 30). Ethnicity, on the other hand, is a complex construct that reflects multiple attributes, including genetic, cultural, and social aspects [55]. We considered these ethnic groups—White, Black, Middle-Eastern, East Asian, and South-East Asian—based on previous studies ([23, 49]) that highlight ethnic variability in glucose metabolism and prediabetes prevalence. We note that although ethnicity does not necessarily correlate with skin tone [63], the participants in our study were quite diverse and covered a wide range of skin tones.

The study excluded pregnant and breastfeeding women and people with known allergies to medical-grade adhesives. We discuss the limitations and extensions of this study and GlucoSense in general in §6.6.

Reference Glucose Levels. Participants agreed to wear a CGM device throughout the study. The device chosen for the study is Freestyle Libre 2. This widely-used device is approved by the US FDA, meets the requirements of the ISO 15197:2013 standard [1], and has been shown to yield high accuracy [29]. The device is typically worn on the upper part of the arm. It has a tiny needle that reaches the interstitial fluid in the skin to measure the glucose level once every minute. It is paired with a mobile application, which displays the glucose reading and periodically archives the data on the cloud. Like other CGM devices, the lifetime of the Freestyle Libre 2 is up to 14 days, after which it has to be replaced.

Frequency of Participant Visits and Study Length. Each participant agreed to visit our lab 4 to 6 times on different days while wearing the glucose monitoring device. These visits did not have to occur on consecutive days. However, they must occur while the monitoring device is still valid, i.e., the last visit must be within 14 days from the first one. Each visit lasted 30–60 minutes. Participant visits were scheduled to cover a wide range of glucose levels. Specifically, the visits were arranged at different times throughout most of the day, including evenings. They were also arranged shortly before/after meals and occasionally 2–3 hours after meals.

Participant Demographics (Total = 31)			
Biological Sex	Male: 14	Female: 17	
Age	19–25: 18	25–40: 8	40–59: 5
BMI	Healthy: 19	Overweight: 9	Obese: 3
Diabetic Profile	Normal: 24 Prediabetic: 1	Type I: 3 Type II: 3	
Ethnicity	White: 8 East Asian: 9 Black: 1	Middle-Eastern: 9 South-East Asian: 4	

Table 1: Demographic information of the participants.

6.2 Hardware and Software of the Testbed

Image Capturing Hardware. Our hardware setup includes five image capturing systems. The first is a high-end hyperspectral camera (Model: Specim IQ), which is a line scanning camera that captures 204 spectral bands with a spatial resolution of 512×512 pixels. This camera takes 20–30 seconds to produce a single hyperspectral image, as it has to scan the captured scene mechanically. The remaining four imaging systems are: (i) RGB+NIR: unmodified Google Pixel 4XL smartphone with RGB and NIR cameras, (ii) RGB-NoIR: Raspberry Pi RGB camera sensor without IR filter, (iii) RGB+ToF: Raspberry Pi RGB camera sensor and time of flight (ToF) sensor, and (iv) Custom: camera sensor with a custom color filter array (model: OnSemiconductor AR0237 RGB-IR).

Software and Automation Scripts. The hyperspectral camera came with its software tools to capture images. Similarly, we used existing applications on the Google Pixel phone to capture RGB and NIR images. However, we wrote software to program the other three capturing systems: RGB-NoIR, RGB+ToF, and Custom. In addition, we wrote automation scripts to control all five imaging systems so that they capture right after each other. This is important for the convenience of the participants, as they will not have to hold their arms for a long time. It is also critical to ensure that the glucose level has not changed while capturing successive images.

Illumination Sources. We have four illumination settings: halogen, LED, fluorescent (CFL), and arbitrary. Halogen sources emit power across most of the 400–1000 nm spectral range, which is suitable for hyperspectral cameras. We use a 250 watt halogen bulb. We also use multiple off-the-shelf LED and CFL bulbs widely used in regular environments such as homes. In addition, we mix several sources, including LED, CFL, light bulbs in the ceiling of our lab, and natural sunlight coming from the lab windows. We refer to this illumination setting as arbitrary, which stresses our system and demonstrates its robustness.

6.3 Data Collection and Model Training

Data Collection. We invite one participant at a time for 30–60 min data collection sessions. The participant sits in a chair with a relaxing posture. We capture their inner wrist using our five imaging systems successively and under all considered illumination settings, typically within a few seconds. The ideal working distance for each imaging system is 20-40cm. Right after capturing these images, we log the glucose reading from the mobile app of the CGM device and associate it with the captured images; we tag all of them with the same timestamp. The timestamp granularity is 1 minute, which is the finest granularity at which the CGM device can record glucose.

We repeat the capturing process every 5–10 minutes within each session, where the participant may rest between capturing and/or eat/drink some snacks/juices we make available. Since we invited each participant for 4–6 sessions on different days, we collected between 48 to 80 glucose readings and images from each. All participants received a $100 amazon gift card each as a token of appreciation upon completing the study.

In total, we collected **1,752** reference glucose readings from the CGM device and their corresponding images captured by our five imaging systems. The distribution of these 1,752 glucose readings is: (i) 21.2%: Low (61–94 mg/dL), (ii) 41.6%: Normal (95–130 mg/dL), (iii) 26.0%: High (131–180 mg/dL), and (iv)11.2%: Abnormally High (181–258 mg/dL).

Model Training and Inference. GlucoSense has two machine learning models: spectral reconstruction and glucose estimation. The reconstruction model is trained on the hyperspectral images dataset. It does not use or see images captured by the other four (mobile) imaging systems. Once trained, this model is used to reconstruct 50 spectral bands from the images captured by the mobile imaging systems.

The glucose estimation model is trained on the glucose readings and their corresponding mobile images. Specifically, for each of the four imaging systems (RGB+NIR, RGB-NoIR, RGB+ToF, and Custom), we train a separate model on 80% of the corresponding glucose readings and images, and we use the remaining 20% for testing. We ensure here to include all samples captured under considered illuminations for each glucose measurement in training as well in testing. So, that model can learn spectral information for different illuminations for each glucose value.

The experiments proceed as follows. We first train the reconstruction and glucose estimation models as described above. Then, for each mobile imaging system, e.g., RGB+NIR, we take a sample from the test dataset, which is a pair of RGB and NIR images in this case. We use the trained reconstruction model to create 50 spectral bands from the sample. Then we feed the 50 spectral bands to the glucose estimation model, which produces a predicted value for the glucose level. This predicted value is then compared against the corresponding reference value measured by the CGM device at the time of taking the sample.

6.4 Accuracy of GlucoSense

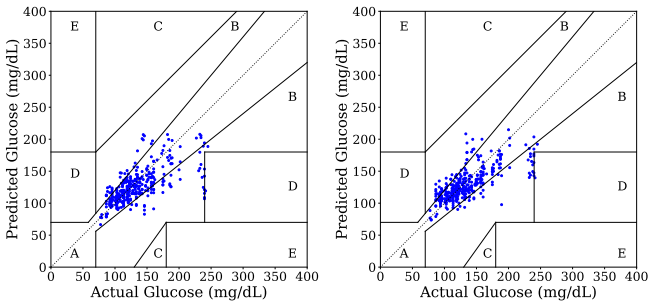
We examine the accuracy of GlucoSense in estimating the glucose levels from mobile images captured by the four proposed systems: RGB+NIR, RGB-NoIR, RGB+ToF, and Custom. Simply comparing predicted values against reference values does not provide a sufficient understanding of the results in the case of glucose analysis. This is because errors have varying clinical significance at different glucose levels. For example, a specific error value, e.g., ±20 mg/dL, may be clinically accepted for glucose levels > 200 mg/dL (diabetic), but the same error would pose a significant risk of misdiagnosis at levels around 130 mg/dL (border of normal and diabetic).

To address this serious issue, error grids have been proposed for glucose analysis. There are three known error grids [35]: Clark Error Grid (CEG), Consensus (aka Parkes) Error Grid (PEG), and Surveillance Error Grid (SEG). CEG is the oldest and has been in use since the late 1980s. It has nine risk zones: A, B, C, D, and E, where some of the zones are repeated at different locations, as shown in Figure 9. PEG addresses some of the shortcomings of CEG, has eight risk zones (also named A–E), and has been adopted by the ISO standard for evaluating glucose devices [1]. SEG is the newest, and as analyzed in [35], it provides the finest granularity in analyzing the accuracy. Specifically, it offers a color-coded *continuous* scale for assessing the risk associated with errors in measuring glucose levels. In Figure 10, we demonstrate PEG (A–E zones separated by lines) overlaid on top of SEG (areas with different shades of colors).

We analyze our results using CEG as it is still the most familiar to researchers in this area. We also analyze our results in the combined PEG/SEG for completeness and compatibility with various evaluation methods [1, 24].

Results on the Clark Error Grid. In Figure 9, we present sample results for the accuracy achieved by GlucoSense running on RGB+NIR and RGB+ToF imaging systems in CEG. The results of the other two camera systems (RGB-NoIR and Custom) are presented in Appendix A.3. The figure demonstrates the high accuracy achieved by GlucoSense. For example, Figure 9.a shows that 99.7% of the predicted values of GlucoSense for the RGB+NIR case (which is the unmodified Google Pixel phone) fall within zones A and B of the Clark error grid. Note that in each sub-figure, we plot 350 predicted glucose values, many of them are overlapping. These are the whole test dataset of the glucose estimation machine learning model (20% of 1,752 = 350), as described in §6.3.

Interpretations of the various zones CEG in Figure 9 are as follows [18, 35]. Glucose predictions in zone A are considered clinically accurate, and in zone B, they are clinically acceptable. In contrast, predictions in zone C may lead to unnecessary treatment, whereas in zone D, they fail to detect



(a) RGB+NIR **(b) RGB+ToF**

Figure 9: Clarke error grid analysis of GlucoSense on (a) Google Pixel, (b) Raspberry Pi camera with time of flight sensor. Almost all predicted glucose values fall in regions A and B (clinically accurate/acceptable).

hypoglycemia or hyperglycemia. Predictions in zone E lead to incorrect treatment decisions in the opposite direction.

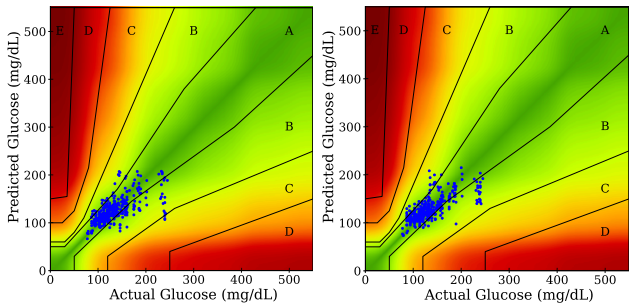
Table 6 (Appendix A.3) summarizes the results for all four mobile imaging systems using the Clark error grid. As the table shows, at most 0.6% of the GlucoSense predictions occurred outside zones A or B for any of the proposed imaging systems.

Results on the Consensus and Surveillance Error Grids. We present sample analysis for the RGB+NIR and RGB+ToF cases in Figure 10 and summarize all results in Table 2. The remaining results are in Appendix A.3. Interpretations of zones in PEG are similar to those in CEG. For SEG, green areas represent no risk of using the predicted glucose values, whereas yellow and red areas represent moderate and high risk, respectively. Similar to the case for CEG, we plot in Figure 10 all 350 predicted glucose values in the test dataset; many of them are also overlapping in this case.

The results in Figure 10 and Table 2 demonstrate the accuracy of GlucoSense and confirm its potential suitability for practical deployment. For example, Table 2 shows that 99.7% of GlucoSense predictions in the RGB+ToF case are within zones A and B. Further, 77.4% of the predictions introduce no risk, whereas 22.6% of them pose only a slight risk.

Quantifying Absolute Relative Difference (ARD). In addition to error grids, the absolute relative difference (ARD) is occasionally used in the literature to assess the accuracy of glucose monitoring systems. ARD is computed as the absolute difference between the predicted and reference glucose values, normalized by the reference value.

We report the mean of the ARD values (MARD) of all considered four mobile imaging systems in Figure 11.a. We also show the MARD achieved by the hyperspectral camera as the lower bound on MARD that can be achieved in the 400-1000 nm range. Further, we show the 95% confidence interval as error bars for each case. The figure shows that



(a) RGB+NIR **(b) RGB+ToF**

Figure 10: Consensus error grid overlaid on surveillance error grid for the analysis of GlucoSense on (a) Google Pixel, (b) Raspberry Pi camera with time of flight sensor. All predicted glucose values fall in regions A and B and within the no (green) or slight (yellow) risk areas.

	Cons. Error Grid (%)			Surv. Error Grid (%)		
	A	B	C	None	Slight	Mod.
RGB+NIR	80.0	17.8	2.2	76.0	23.7	0.3
RGB+ToF	**81.3**	**18.4**	**0.3**	**77.4**	**22.6**	**0.0**
RGB-NoIR	60.4	34.9	4.7	59.3	40.4	0.3
Custom	73.6	25.8	0.6	73.5	26.5	0.0

Table 2: Summary of the results using Consensus and Surveillance error grids. At least 95% of GlucoSense predictions in all cases fall within zones A and B, and 99% of them have no or slight risk.

the RGB+NIR and RGB+ToF achieve MARD of about 13%, which is close to the lower bound achieved by the expensive hyperspectral camera. On the other hand, the Custom and RGB-NoIR imaging systems produce MARD of 15% and 20%, respectively. We attribute the relatively lower performance in these two cases to the weak NIR signals captured by the Custom and RGB-NoIR systems.

Reference Comparison to FDA and ISO Requirements. To shed more light on our results, we compare them to globally recognized requirements and standards: US FDA requirements [24] and ISO 15197:2013 standard [1].

We first note that there are two broad categories of systems for measuring glucose: blood glucose monitoring (BGM) and continuous glucose monitoring (CGM). BGM systems, e.g., Accu-Chek, measure glucose directly from the blood using, for example, medical test strips to obtain small blood samples after finger pricking. CGM systems, e.g., Free Libre 2, measure glucose within the interstitial fluid in the skin. The ISO 15197:2013 standard [1] defines requirements for BGM systems but does not target CGM. The US FDA defines requirements for CGM systems [24].

To be clinically accurate, one of the main FDA requirements for CGM devices is that at least 87% of all predicted glucose values should be within ±20% of the reference values. We plot the CDF of all ARD values resulted from running GlucoSense on different imaging systems in Figure 11.b. The figure shows that the hyperspectral camera meets this requirement, as it results in 88% of the predictions having 20% or less absolute relative error. On the other hand, two of the proposed mobile imaging systems (RGB+NIR and RGB+ToF) result in 80% of all predictions within the ±20% error margin, which is a small gap of 7% from the FDA main requirement.

The ISO 15197:2013 standard [1] for BGM systems has two main accuracy requirements: (i) at least 99% of the glucose readings must fall within zones A and B on the Consensus Error Grid, and (ii) at least 95% of the readings should be within ±15 mg/dL of the reference measurement for glucose concentrations < 100 mg/dL and within ±15% for concentrations ≥ 100 mg/dL. As shown in Figure 10, GlucoSense running on a simple, unmodified smartphone (the RGB+NIR case) with no external attachments and no direct access to blood can meet the first requirement of the ISO standard. But it does not yet meet the second requirement.

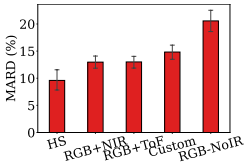
In our future work, we will explore various optimizations to hopefully achieve the second requirement of the ISO standard and shrink the gap with the FDA main requirement mentioned above, as well as achieve the few other range-specific FDA requirements stated in [24].

6.5 Robustness Analysis

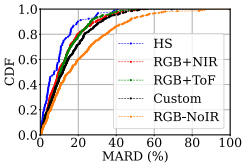
Robustness to User Diversity. Figure 12 shows the MARD and 95% confidence intervals for different genders, ethnic groups, ages, BMI categories, and diabetes groups on the RGB+NIR case. Overall, the MARD is similar across diverse participants. We detected less error in the black ethnic group. This can be attributed to the low sample size (1 person) in this category. Due to local demographics, we could not recruit more in this community. We will conduct more experiments in the future to validate such cases. We also identified slightly more error in subjects having diabetes as compared to normal subjects. This can be attributed to the fact that the total participants having diabetes (7 people) is considerably less than the normal (24 people). We analyze the generalizability of GlucoSense by using the leave-one-out training approach and showed the positive effect of calibration in Appendix A.3

Robustness to Illuminations. We evaluate the performance of GlucoSense under different illuminations, and the results (Figure 13) demonstrate its robustness as the accuracy did not significantly change across different illuminations.

Identification of Optimal Body Site. We experimented with different locations on the body to identify the most convenient location(s) for participants to take pictures while



(a) Average



(b) CDF

Figure 11: Distributions of the glucose measurement errors for different camera systems. HS: Hyperspectral.

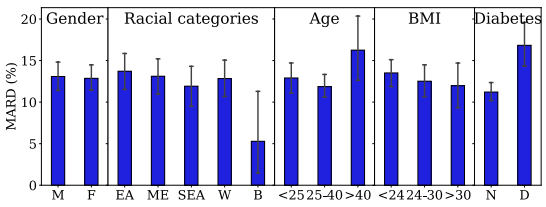


Figure 12: Impact of user diversity. EA: East Asian, ME: Middle Eastern, SEA: South East Asian, W: White, and B: Black.

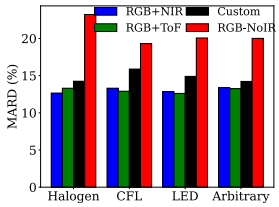


Figure 13: Robustness to different illuminations.

optimizing the accuracy of GlucoSense. Through discussions with participants and multiple trials, we concluded that the *inner part of the wrist* is the best location. This is because the skin in this location is the thinnest, maximizing the reflected signals from the inner layers of the skin. In addition, this location typically has the least amount of hair, which reduces obstruction during image capturing. Further, the melanin content in this location is lower than in other parts of the arm, reducing the impact of skin tone on the accuracy.

6.6 Limitations and Practical Considerations

Limitations of our User Study. We compared our glucose predictions against reference values obtained by the Free Libre 2 device. While this device is approved by the FDA and is reported to be among the most accurate in the market, it still has some errors compared to directly measuring glucose from the blood. In the future, we plan to collaborate with diabetes medical specialists and to compare GlucoSense directly against the ground-truth blood glucose levels.

In addition, while our user study lasted for months, it contained data from only 31 participants. A larger number of diverse participants would better assess the potential of GlucoSense. For example, more participants are needed to cover the broad spectrum of human skin tones. This is in addition to the diversity in ethnicity because ethnicity is an inadequate proxy for skin tone [63].

Practical Considerations. GlucoSense requires NIR signals on phones. Most recent phones have NIR cameras and/or time-of-flight sensors. However, some phone manufacturers, e.g., Apple, do not currently expose the APIs for accessing such sensors to external developers. We hope that the results in this paper motivate more phone manufacturers to allow accessing their NIR camera modules.

For deployment, GlucoSense does not need a hyperspectral camera or hyperspectral images; it requires only RGB and NIR images captured by regular cameras. The spectral reconstruction model in GlucoSense does need hyperspectral images for training but not for inference. Our hyperspectral images dataset, which will be open-source, presents a good start for training. When larger hyperspectral image datasets become available, the reconstruction model can be fine-tuned on them to improve the accuracy.

Finally, we note that GlucoSense is a non-invasive system for measuring glucose. This alleviates the pain for many diabetic patients, especially those who use blood tests through finger pricking. GlucoSense also reduces the cost compared to using CGM devices, which are fairly expensive and need to be periodically replaced. In addition, CGM devices may increase the chances of skin infection/irritation because of their prolonged use. GlucoSense, on the other hand, requires patients to actively take pictures and measure glucose, which may be inconvenient for some patients compared to CGM devices that continuously monitor glucose.

7 CONCLUSIONS AND FUTURE WORK

Diabetes is a chronic disease that affects hundreds of millions of people worldwide. To avoid severe health complications, diabetic patients must closely monitor their glucose levels through multiple measurements each day. Most current commercial glucose monitoring systems are painful and costly. We presented GlucoSense, a non-invasive glucose measurement system that runs on smartphones. We first demonstrated the feasibility of measuring glucose in the visible and NIR range (400–1000 nm) using a hyperspectral camera. We also proposed a systematic approach to rank wavelengths in order of importance for designing low-cost glucose-measurement devices. We then presented multiple camera designs to obtain NIR signals on phones. We conducted a user study to analyze the performance and robustness of GlucoSense. The results demonstrated that GlucoSense achieves high accuracy in measuring glucose on a diverse set of participants, close to the clinically accepted measurements that require expensive/invasive devices.

In the future, we plan to conduct a clinical study in collaboration with medical experts to include larger and more diverse user populations, compare against glucose measurements obtained in medical labs directly from blood, and customize our system to support various patient needs, e.g., focus on specific glucose ranges.

REFERENCES

[1] 2013. *ISO 15197:2013 In vitro diagnostic test systems — Requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus*. Technical Report. International Standardization Organization (ISO). https://www.iso.org/standard/54976.html

[2] 2023. *IR in surveillance Day-and-night cameras and OptimizedIR*. Technical Report. https://www.axis.com/dam/public/da/54/a6/ir-in-surveillance-en-US-191142.pdf

[3] 2024. ON Semiconductor's CMOS Image Sensor Model AR0237 RGB-IR. https://www.onsemi.com/products/sensors/image-sensors/AR0237

[4] 2024. Raspberry Pi Cameras. https://www.raspberrypi.com/documentation/accessories/camera.html

[5] 2024. Removal of IR Filter. https://www.raspberrypi.com/documentation/accessories/camera.html#filter-removal

[6] 2024. ToF Camera for Raspberry Pi. https://www.arducam.com/time-of-flight-camera-raspberry-pi/

[7] Abbott. 2024. Freestyle Libre 2. https://www.freestyle.abbott/ca-en/learn/sensor-based-glucose-monitoring.html

[8] Mahmoud Afifi and Michael S Brown. 2020. Deep White-Balance Editing. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'20)*. IEEE Explore, Seattle, WA, USA, 1394–1403. https://doi.org/10.1109/CVPR42600.2020.00147

[9] Israr Ahmed, Nan Jiang, Xinge Shao, Mohamed Elsherif, Fahad Alam, Ahmed Salih, Haider Butta, and Ali K. Yetisen. 2022. Recent advances in optical sensors for continuous glucose monitoring. , 1098–1125 pages. https://doi.org/10.1039/d1sd00030f

[10] Kamal Alhallak. 2021. Skin, Light and their Interactions, an In-Depth Review for Modern Light-Based Skin Therapies. *Clinical Dermatology & Therapy* 7, 2 (9 2021), 1–15. https://doi.org/10.24966/CDT-8771/100081

[11] Murad Althobaiti. 2022. Citation: Althobaiti, M. In Silico Investigation of SNR and Dermis Sensitivity for Optimum Dual-Channel Near-Infrared Glucose Sensor Designs for Different Skin In Silico Investigation of SNR and Dermis Sensitivity for Optimum Dual-Channel Near-Infrared Glucose Sensor Designs for Different Skin Colors. (2022). https://doi.org/10.3390/bios

[12] Boaz Arad and et al. 2022. NTIRE 2022 Spectral Recovery Challenge and Data Set. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'22) Workshops*. 863–881.

[13] Caerwyn Ash, Michael Dubec, Kelvin Donne, and Tim Bashford. 2017. Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods. *Lasers in Medical Science* 32, 8 (11 2017), 1909–1918. https://doi.org/10.1007/s10103-017-2317-4

[14] Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. 2022. MST++: Multi-Stage Spectral-Wise Transformer for Efficient Spectral Reconstruction. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'22) Workshops*. 745–755.

[15] CDC. 2024. Manage Blood Sugar. https://www.cdc.gov/diabetes/managing/manage-blood-sugar.html

[16] Eda Cengiz and William V Tamborlane. 2009. *A Tale of Two Compartments: Interstitial Versus Blood Glucose Monitoring*. Technical Report. https://pmc.ncbi.nlm.nih.gov/articles/PMC2903977/

[17] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A scalable tree boosting system. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Vol. 13-17-August-2016. 785–794. https://doi.org/10.1145/2939672.2939785

[18] Koushik Chowdhury, Anuj Srivastava, Neeraj Sharma, and Shiru Sharma. 2015. Error Grid Analysis of Reference and Predicted Blood Glucose Level Values as Obtained from the Normal and Prediabetic Human Volunteers. *American Journal of Biomedical Engineering* 5, 1 (2015), 6–14. https://doi.org/10.5923/j.ajbe.20150501.02

[19] Leszek Czupryniak, László Barkai, Svetlana Bolgarska, Agata Bronisz, Jan Broz, Katarzyna Cypryk, Marek Honka, Andrej Janez, Mladen Krnic, Nebojsa Lalic, Emil Martinka, Dario Rahelic, Gabriela Roman, Tsvetalina Tankova, Tamás Várkonyi, Bogumił Wolnik, and Nadia Zherdova. 2014. Self-monitoring of blood glucose in diabetes: From evidence to clinical reality in central and eastern europe - Recommendations from the international central-eastern european expert group. , 460–475 pages. https://doi.org/10.1089/dia.2013.0302

[20] Nicholas B. Davison, Christopher J. Gaffney, Jemma G. Kerns, and Qiandong D. Zhuang. 2022. Recent Progress and Perspectives on Non-Invasive Glucose Sensors. *Diabetology* 3, 1 (1 2022), 56–71. https://doi.org/10.3390/diabetology3010005

[21] Andrew DeHennis and Mark Mortellaro. 2020. CGM sensor technology. In *Glucose Monitoring Devices: Measuring Blood Glucose to Manage and Control Diabetes*. Elsevier, 111–134. https://doi.org/10.1016/B978-0-12-816714-4.00006-5

[22] Nuha A. Elsayed, Grazia Aleppo, Vanita R. Aroda, Raveendhara R. Bannuru, Florence M. Brown, Dennis Bruemmer, Billy S. Collins, Sandeep R. Das, Marisa E. Hilliard, Diana Isaacs, Eric L. Johnson, Scott Kahan, Kamlesh Khunti, Mikhail Kosiborod, Jose Leon, Sarah K. Lyons, Mary Lou Perry, Priya Prahalad, Richard E. Pratley, Jane Jeffrie Seley, Robert C. Stanton, and Robert A. Gabbay. 2023. 10. Cardiovascular Disease and Risk Management: Standards of Care in Diabetes—2023. *Diabetes Care* 46 (1 2023), S158–S190. https://doi.org/10.2337/dc23-S010

[23] Taynara Formagini, Joanna Veazey Brooks, Andrew Roberts, Kai Mc Keever Bullard, Yan Zhang, Ryan Saelee, and Matthew James O'Brien. 2023. Prediabetes prevalence and awareness by race, ethnicity, and educational attainment among U.S. adults. *Frontiers in Public Health* 11 (2023). https://doi.org/10.3389/fpubh.2023.1277657

[24] Guido Freckmann, Stefan Pleus, Mike Grady, Steven Setford, and Brian Levy. 2019. Measures of Accuracy for Continuous Glucose Monitoring and Blood Glucose Monitoring Devices. , 575–583 pages. https://doi.org/10.1177/1932296818812062

[25] Wilbert Villena Gonzales, Ahmed Toaha Mobashsher, and Amin Abbosh. 2019. The progress of glucose monitoring—A review of invasive to minimally and non-invasive techniques, devices and sensors. https://doi.org/10.3390/s19040800

[26] Google Research. 2020. uDepth: Real-time 3D Depth Sensing on the Pixel 4. (2020). https://research.google/blog/udepth-real-time-3d-depth-sensing-on-the-pixel-4/

[27] Weixi Gu, Yuxun Zhou, Zimu Zhou, Xi Liu, Han Zou, Pei Zhang, Costas J. Spanos, and Lin Zhang. 2017. SugarMate. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (9 2017), 1–27. https://doi.org/10.1145/3130919

[28] Guang Han, Siqi Chen, Xiaoyan Wang, Jinhai Wang, Huiquan Wang, and Zhe Zhao. 2021. Noninvasive blood glucose sensing by near-infrared spectroscopy based on PLSR combines SAE deep neural network approach. *Infrared Physics and Technology* 113 (3 2021). https://doi.org/10.1016/j.infrared.2020.103620

[29] Kevin Hanson, Mark Kipnes, and Hien Tran. 2024. Comparison of Point Accuracy Between Two Widely Used Continuous Glucose Monitoring Systems. *Journal of Diabetes Science and Technology* 18, 3 (5 2024), 598–607. https://doi.org/10.1177/19322968231225676

[30] Paris J Hantzidiamantis and Sarah L Lappin. 2023. *Physiology, Glucose*. StatPearls Publishing, Treasure Island (FL). http://europepmc.org/abstract/MED/31424785http://europepmc.org/books/NBK545201https://www.ncbi.nlm.nih.gov/books/NBK545201

[31] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (second ed.). Springer.

[32] Healthline. 2024. Effect of low blood sugar. https://www.healthline.com/health/low-blood-sugar-effects-on-body

[33] Prateek Jain, Ravi Maddila, and Amit M. Joshi. 2019. A precise non-invasive blood glucose measurement system using NIR spectroscopy and Huber's regression model. *Optical and Quantum Electronics* 51, 2 (2 2019). https://doi.org/10.1007/s11082-019-1766-3

[34] Jayoung Kim, Alan S. Campbell, and Joseph Wang. 2018. Wearable non-invasive epidermal glucose sensors: A review. *Talanta* 177 (1 2018), 163–170. https://doi.org/10.1016/j.talanta.2017.08.077

[35] David C. Klonoff, Courtney Lias, Robert Vigersky, William Clarke, Joan Lee Parkes, David B. Sacks, M. Sue Kirkman, and Boris Kovatchev. 2014. The surveillance error grid. *Journal of Diabetes Science and Technology* 8, 4 (2014), 658–672. https://doi.org/10.1177/1932296814539589

[36] Larry Li. 2014. *Time-of-Flight Camera – An Introduction.* Technical Report. Texas Instruments. https://www.ti.com/lit/wp/sloa190b/sloa190b.pdf

[37] Ho Man Colman Leung, Gregory P. Forlenza, Temiloluwa O. Prioleau, and Xia Zhou. 2023. Noninvasive Glucose Sensing In Vivo. https://doi.org/10.3390/s23167057

[38] Nan Li, Hang Zang, Huimin Sun, Xianzhi Jiao, Kangkang Wang, Timon Cheng Yi Liu, and Yaoyong Meng. 2019. A noninvasive accurate measurement of blood glucose levels with Raman spectroscopy of blood in microvessels. *Molecules* 24, 8 (4 2019). https://doi.org/10.3390/molecules24081500

[39] Tianxing Li, Derek Bai, Temiloluwa Prioleau, Nam Bui, Tam Vu, and Xia Zhou. 2020. Noninvasive glucose monitoring using polarized light. In *SenSys 2020 - Proceedings of the 2020 18th ACM Conference on Embedded Networked Sensor Systems.* Association for Computing Machinery, Inc, 544–557. https://doi.org/10.1145/3384419.3430720

[40] Sabbir Liakat, Kevin A. Bors, Laura Xu, Callie M. Woods, Jessica Doyle, and Claire F. Gmachl. 2014. Noninvasive in vivo glucose sensing on human subjects using mid-infrared light. *Biomedical Optics Express* 5, 7 (7 2014), 2397. https://doi.org/10.1364/boe.5.002397

[41] Li Lin, Haoqi He, Ruiyang Xue, Yumin Zhang, Ziwen Wang, Shuming Nie, and Jian Ye. 2023. Direct and quantitative assessments of near-infrared light attenuation and spectroscopic detection depth in biological tissues using surface-enhanced Raman scattering. *Med-X* 1, 1 (8 2023). https://doi.org/10.1007/s44258-023-00010-2

[42] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems.* Curran Associates Inc., 4768–4777. https://doi.org/doi/10.5555/3295222.3295230

[43] Bilal Ahmad Malik and Mohammed Benaissa. 2012. Glucose Chemistry. In *Dietary Sugars: Chemistry, Analysis, Function and Effects*, J R Alonso-Fernandez, D J Timson, A Szilagyi, Ralph Louis Obendorf, M F Abdelmalek, Mitch McGrath, Laid Boukraâ, M D Gastrich, Anna Haukioja, Jorgen Jensen, M Erkkola, K A Amin, M F Varela, Fabio Vieira dos Santos, L M Moreira, Lee Hansen, Marybeth B Hall, Tomomi Fujisawa, D S Millington, E Denova-Gutierrez, Mohammed Benaissa, D Centonze, Dirk A L Aarts, J J Gagliardino, M L Sanz, Leon A Terry, A Abdul-Aziz, M M F Choi, E M S M Gaspar, J L Todoli, K Sugimoto, W Surareungchai, L M Moreira, Jeanette N Keith, N Silanikove, David S Newburg, Odilia I Bermudez, Andrea L Deierlein, K McCowen, M J Rodriguez-Yoldi, John Griffith Jones, C A Pinkert, N Tamaki Tamaki, S Navarro, C Cheeseman, M Soupioni, Jose Manuel Pingarron, Luca Valgimigli, and Victor R Preedy (Eds.). The Royal Society of Chemistry, 0. https://doi.org/10.1039/9781849734929-00077

[44] Yusuke Monno, Hayato Teranaka, Kazunori Yoshizaki, Masayuki Tanaka, and Masatoshi Okutomi. 2019. Single-Sensor RGB-NIR Imaging: High-Quality System Design and Prototype Implementation. *IEEE Sensors Journal* 19, 2 (2019), 497–507.

[45] Kazi Atique Moula Nabil, Md Ariful Islam, Abdullah Al Noman, and Mohammad Monirujjaman Khan. 2023. Development of A Smart Non-Invasive Glucose Monitoring System With SpO2 and BPM for Diabetic Patient. In *2023 IEEE 13th Annual Computing and Communication Workshop and Conference, CCWC 2023.* Institute of Electrical and Electronics Engineers Inc., 193–197. https://doi.org/10.1109/CCWC57344.2023.10099236

[46] Semi Oh, Min-Seok Kim, Beom-Rae Noh, Hyung-Hwan Baik, and Kyoung-Kook Kim. 2021. Non-Invasive Measurement of Glucose Concentration Using Red and Near-Infrared Light-Emitting Diodes. *Biomedical Sciences* 7, 2 (2021), 47. https://doi.org/10.11648/j.bs.20210702.12

[47] Paul O'Sullivan and Nicolas Le Dortz. 2021. Time of Flight System Design—Part 1: System Overview. *Analog Devices* 55 (7 2021). https://www.analog.com/en/resources/analog-dialogue/articles/time-of-flight-system-design-part-1-system-overview.html

[48] Ruiliang Pu. 2017. *Hyperspectral Remote Sensing: Fundamentals and Practices.* Routledge.

[49] Amena Sadiya, Vidya Jakapure, and Vijay Kumar. 2023. Ethnic Variability in Glucose and Insulin Response to Rice Among Healthy Overweight Adults: A Randomized Cross-Over Study. *Diabetes, Metabolic Syndrome and Obesity* 16 (2023), 993–1002. https://doi.org/10.2147/DMSO.S404212

[50] Pradnya P. Samant, Megan M. Niedzwiecki, Nicholas Raviele, Vilinh Tran, Juan Mena-Lapaix, Douglas I. Walker, Eric I. Felner, Dean P. Jones, Gary W. Miller, and Mark R. Prausnitz. 2020. Sampling interstitial fluid from human skin using a microneedle patch. *Science Translational Medicine* 12, 571 (11 2020). https://doi.org/10.1126/SCITRANSLMED.AAW0285

[51] Samsung ToF. 2024. Samsung ToF sensor. https://semiconductor.samsung.com/image-sensor/tof-sensor/

[52] Neha Sharma, Muhammad Shahzaib Waseem, Shahrzad Mirzaei, and Mohamed Hefeeda. 2023. MobiSpectral: Hyperspectral Imaging on Mobile Devices. In *Proceedings of ACM Conference on Mobile Computing and Networking (MobiCom'23) (ACM MobiCom '23).*

[53] Maryamsadat Shokrekhodaei, David P. Cistola, Robert C. Roberts, and Stella Quinones. 2021. Non-Invasive Glucose Monitoring Using Optical Sensor and Machine Learning Techniques for Diabetes Applications. *IEEE Access* 9 (2021), 73029–73045. https://doi.org/10.1109/ACCESS.2021.3079182

[54] Joo Yong Sim, Chang Geun Ahn, Eun Ju Jeong, and Bong Kyu Kim. 2018. In vivo Microscopic Photoacoustic Spectroscopy for Non-Invasive Glucose Monitoring Invulnerable to Skin Secretion Products. *Scientific Reports* 8, 1 (12 2018). https://doi.org/10.1038/s41598-018-19340-y

[55] Audrey Smedley and Brian D. Smedley. 2005. Race as biology is fiction, racism as a social problem is real : Anthropological and historical perspectives on the social construction of race. , 16–26 pages. https://doi.org/10.1037/0003-066X.60.1.16

[56] Sony. 2024. Time of Flight Image Sensor. https://www.sony-semicon.com/en/products/is/industry/tof.html

[57] Parama Sridevi, Shamsul Arefin, and Abu Shahadat. 2021. A Feasibility Study of Non-invasive Blood Glucose Level Detection Using Near-Infrared Optical Spectroscopy. *Bangladesh Journal of Medical Physics* 14, 1 (2021), 1–13.

[58] Ernia Susana, Kalamullah Ramli, Hendri Murfi, and Nursama Heru Apriantoro. 2022. Non-Invasive Classification of Blood Glucose Level for Early Detection Diabetes Based on Photoplethysmography Signal. *Information (Switzerland)* 13, 2 (2 2022). https://doi.org/10.3390/info13020059

[59] Yong Yi Tan, Enhui Suan, Gerald Choon Huat Koh, Suhana Binte Suhairi, and Shilpa Tyagi. 2024. Effectiveness of continuous glucose monitoring in patient management of Type 2 Diabetes Mellitus: an umbrella review of systematic reviews from 2011 to 2024. *Archives of Public Health* 82, 1 (12 2024). https://doi.org/10.1186/s13690-024-01459-2

[60] Tzu-Yu Wu. 2021. ToF System Design—Part 2: Optical Design for Time of Flight Depth Sensing Cameras. *Analog Devices* (2021). https://www.analog.com/en/resources/analog-dialogue/articles/tof-system-design-part-2-optical-design-for-time-of-flight-depth-sensing-cameras.html

[61] Anandghan Waghmare, Farshid Salemi Parizi, Jason Hoffman, Yuntao Wang, Matthew Thompson, and Shwetak Patel. 2023. GlucoScreen: A Smartphone-based Readerless Glucose Test Strip for Prediabetes Screening. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 1 (3 2023). https://doi.org/10.1145/3580855

[62] Adam Wawerski, Barbara Siemiątkowska, Michał Józwik, Bartłomiej Fajdek, and Małgorzata Partyka. 2024. Machine Learning Method and Hyperspectral Imaging for Precise Determination of Glucose and Silicon Levels. *Sensors* 24, 4 (2 2024). https://doi.org/10.3390/s24041306

[63] Vanessa R. Weir, Katelyn Dempsey, Judy Wawira Gichoya, Veronica Rotemberg, and An Kwok Ian Wong. 2024. A survey of skin tone assessment in prospective research. https://doi.org/10.1038/s41746-024-01176-8

[64] World Health Organization (WHO). 2024. The top 10 causes of death (WHO). https://www.who.int/news-room/fact-sheets/detail/diabetes

[65] World Heath Organizations (WHO). 2024. Diabetes.

[66] Jyoti Yadav, Asha Rani, Vijander Singh, and Bhaskar Mohan Murari. 2014. Near-infrared LED based Non-invasive Blood Glucose Sensor. In *International Conference on Signal Processing and Integrated Networks (SPIN)*.

[67] Eun Hyung Yoo and Soo Youn Lee. 2010. Glucose biosensors: An overview of use in clinical practice. *Sensors* 10, 5 (5 2010), 4558–4576. https://doi.org/10.3390/s100504558

A APPENDIX

This appendix contains more details and results that supplement the materials presented in the paper.

A.1 Details of the Spectral Reconstruction Model

Training the Reconstruction Model. A significant challenge in building a reconstruction model for mobile images is that we need paired data to train the model. That is, we need ground truth hyperspectral images captured by a hyperspectral camera that correspond to images captured by mobile devices of the same scene. This, however, is not practically possible because hyperspectral cameras have very different sensor characteristics, e.g., resolution and bit representation, than mobile cameras. To address this challenge, we first capture a large hyperspectral image dataset with our high-end hyperspectral camera. Then, we create the corresponding mobile image dataset by applying the camera's sensitivity function of the considered mobile device.

To illustrate, consider a mobile device that has the full spectrum RGB camera shown in Figure 6.a. We process the 200+ spectral bands of each image in the hyperspectral image dataset by applying the sensitivity function in Figure 6.b on it, which results in a *synthesized* image similar to the one that a real full spectrum RGB camera would have captured. Since the hyperspectral camera produces far more detailed information than other cameras, the synthesized images are fairly accurate. This image synthesization is commonly used to evaluate the performance of hyperspectral imaging systems [12, 14, 48, 52].

In addition, to simulate the camera processing pipeline of smartphone cameras and normalize the effect of different ambient light, we apply Automatic White Balancing (AWB) [8] to the synthesized RGB images. Since the input to our reconstruction model is RGB and NIR images, we choose a paired NIR band (940 nm) from the hyperspectral image. This NIR wavelength was selected based on the multiple depth sensing solutions on recent smartphones that use the same wavelength to mitigate the effect of sunlight.

Accuracy of the Reconstructed Bands. The reconstruction model produces 50 bands from which we extract spectral signatures representing different glucose concentrations. Thus, the accuracy of these bands is critical for the performance of GlucoSense.

We assess the accuracy of the reconstructed bands by comparing them against the ground truth ones captured by the hyperspectral camera. We train our reconstruction model on two settings: (i) Input RGB and NIR (4 bands) and (ii) Input full spectrum RGB (3 bands). In both settings, we reconstruct the top 50 bands selected based on the order of importance as discussed in §4.2. Unlike prior works [14, 52], these bands are not equally spaced in the 400–1000 nm spectral range.

	RMSE ↓	SAM ↓	SID ↓	PSNR ↑
RGB+NIR	0.0528	0.1035	0.0217	26.5
RGB-NoIR	0.1919	0.1790	0.1035	15.2

Table 3: Performance of the reconstruction model using two different inputs: (i) 4 bands (RGB+NIR) and (ii) 3 bands (RGB-NoIR).

We divide the hyperspectral image dataset into 75% training and 25% testing partitions. The training partition has data points from 23 (out of 31) participants, which belong to three (out of the five) ethnic groups: White (7), East Asian (8), and Middle Eastern (8). The test partition has data points from the remaining eight participants covering all five ethnic groups: White (1), East Asian (1), Middle Eastern (1), South East Asian (4), and Black (1). Notice that each participant has multiple data points collected over several days. Further, data from the Black and South East Asian were not seen during the training of the reconstruction model. This is done to stress the model and assess its robustness.

We report the overall performance of the two versions of the reconstruction model in Table 3 using four performance metrics commonly used in the literature [14, 52]. The RMSE (Root Mean Square Error) measures the pixel-wise *error* between reconstructed and ground truth bands, SAM (Spectral Angle Mapper) and SID (Spectral information divergence) measure the *error* in the spectral domain, and PSNR (Peak Signal-to-Noise Ratio) assesses the *quality* of the reconstructed bands.

The results in Table 3 show that the recovery is much better from 4 input bands (RGB and an independent NIR) compared to the 3 input bands (RGB-NoIR) having mixed RGB and NIR signals. For example, PSNR is decreased (26.5 to 15.2) for the RGB-NoIR version, indicating higher noise in the recovered signal. The low quality of the reconstructed bands affects the overall performance of GlucoSense, as was shown in §6.4.

We further analyze the performance on different ethnicity groups in the test partition using the RGB+NIR version of the model. The results, shown in Table 4, show that the performance of all metrics is near the average, even for the ethnicity groups not present in the training partition (South East Asian and Black). This demonstrates the robustness of the reconstruction model.

Finally, we present a sample result to visually illustrate the accuracy of the reconstructed bands in Figure 14. The figure shows the spectral profile of a participant in the test partition (Black) and compares it to the ground truth across all 50 bands. Specifically, we select a 64x64 patch of the image. Then, we perform spectral reconstruction for this patch to recover the top 50 bands. Then, we compute a spectral

	RMSE ↓	SAM ↓	SID ↓	PSNR ↑
White	0.0501	0.0859	0.0077	26.2
East Asian	0.0494	0.1228	0.0320	27.3
Middle Eastern	0.0516	0.1173	0.0452	27.6
South East Asian	0.0540	0.0938	0.0147	26.4
Black	0.0471	0.0957	0.0114	26.5

Table 4: Performance of the reconstruction model on various ethnic groups.

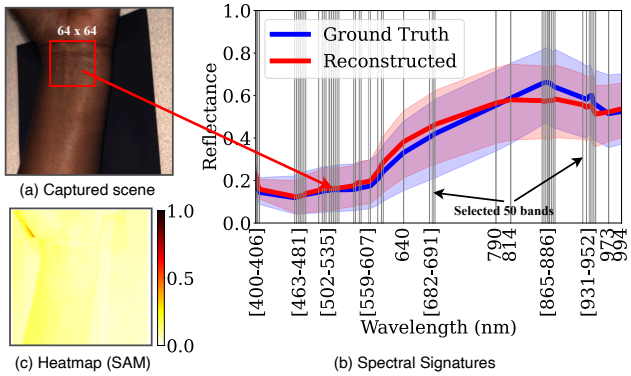


Figure 14: An example to visually illustrate the accuracy of the spectral reconstruction model in producing signatures close to the ground truth ones.

signature for each pixel in this patch. The spectral signature shows how the reflectance changes across the 50 bands. We plot the signatures created from reconstructed bands in red Figure 14.b. We also compute the spectral signatures for all pixels from the ground truth bands and plot them in blue. The thick red and blue lines represent the average of all reconstructed and ground truth signatures, respectively. The figure shows that the reconstructed and ground truth signatures overlap across most of the spectral range, demonstrating the high quality of the reconstructed bands even for a participant from an ethnic group not seen during training.

To further illustrate this accuracy, we plot the error (measured using the SAM metric) between the reconstructed and ground truth bands as a heat map in Figure 14.c. As the heat map shows, the error is close to 0 in most cases.

A.2 Details of the Glucose Estimation Model

Implementation Details and Hyperparameters. We use the XGBoost library [17] to implement our model. XGBoost is a powerful gradient-boosting framework optimized for efficiency and performance. The model uses 800 decision trees and incorporates several key regularization parameters to control model complexity and prevent overfitting. Specifically, the model is configured with the following parameters: maximum tree depth of 8, tree pruning parameter γ of 0.006, learning rate of 0.26, and subsample ratio of 0.89. The tree depth parameter effectively limits the complexity of each decision tree by restricting how deep it can grow, preventing the model from creating overly specific decision paths that might overfit the training data. The γ parameter implements a form of tree pruning by requiring a minimum loss reduction for further tree partitioning, helping to eliminate unnecessary complexity in the model structure. The learning rate scales the contribution of each tree, allowing for more conservative model updates and preventing any single tree from having too much influence on the final predictions. The subsample ratio parameter introduces an element of randomness, which helps create a more robust ensemble that's less likely to overfit specific patterns in the training data.

Our model further enhances XGBoost's regularization strategy by employing the SHAP values to select the top 50 most important spectral bands from the original 204 bands. This feature selection approach significantly reduces the dimensionality of the input data while retaining the most informative features for glucose prediction. Focusing on fewer, more relevant bands makes the model less susceptible to noise and more computationally efficient. The selected bands are likely to correspond to meaningful glucose absorption wavelengths, which improves the model's interpretability and ensures that predictions are based on genuine spectral relationships rather than spurious correlations.

The synergy between XGBoost's internal regularization mechanisms and SHAP-based feature selection creates a robust framework for glucose prediction. The initial feature selection process provides a first layer of complexity reduction by eliminating less important spectral bands, while XGBoost's regularization parameters fine-tune the model's behavior on the selected features. This multi-level approach to regularization results in a more stable and generalizable model. The reduced feature set also enables faster training times and more extensive hyperparameter tuning, making more efficient use of computational resources. Most importantly, focusing on the most predictive spectral bands and employing proper regularization techniques makes the model less susceptible to noise and artifacts, leading to better performance on unseen data.

Finally, the model is configured to run on GPU using the histogram-based algorithm for split finding, which is optimized for large datasets. This architecture balances model complexity with computational efficiency while maintaining strong predictive performance. Further, our model leverages XGBoost's various system optimizations, e.g., parallel processing and cache-aware access patterns, to improve training efficiency.

Illustrative Example. We present a simplified visualization of a single decision tree from our XGBoost-based glucose

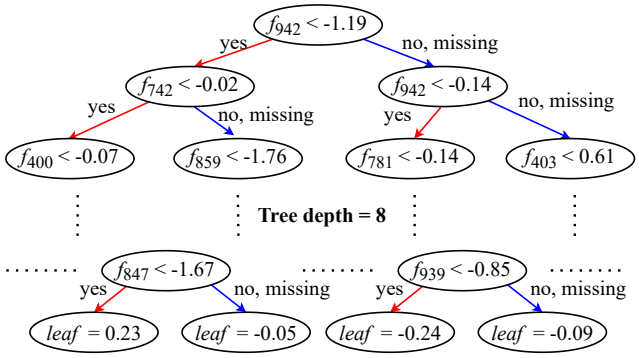


Figure 15: Illustration of one of the decision trees in our glucose prediction model.

prediction model in Figure 15. The complete model has an ensemble of 800 trees, each with a depth of up to 8. Each tree is a weak learner that attempts to correct the residual errors from previous trees' predictions. The final glucose prediction is computed as the sum of predictions from all trees multiplied by the learning rate (0.26 in our model) plus an initial base prediction. In this visualization, each node displays crucial information: the feature index (corresponding to a specific wavelength), the splitting condition, the prediction value, and the coverage (number of training samples). For instance, if a node shows '$f_{942} < 0.5$', it indicates that the model evaluates whether the reflectance at wavelength 942nm is less than 0.5. The 'leaf' value in terminal nodes represents this path's contribution to the final prediction. While this single tree provides insights into the model's decision-making process, it's important to note that the final prediction accuracy comes from the combined effect of all 800 trees working together.

Comparison against other Regression Models. We considered other widely adopted regression methods for our glucose prediction model, including K-Nearest-Neighbours (KNN), Random Forest (RF), Support Vector Machine Regressor (SVR), Partial Least Square Regressor (PLSR), and Neural Network (multilayer perceptron). We implemented all these models and evaluated their accuracy in predicting glucose. For the Neural Network case, we tried many designs with different numbers of layers and neurons in each layer. We present the results for a network with two layers, each with 100 neurons. Deeper networks quickly overfit our relatively small datasets and did not produce better results.

The results, summarized in Table 5, demonstrate that our XGBoost-based model outperforms all other machine learning regression models considered in the comparison. This is attributed to the fact that XGBoost is more likely to distinguish between numerical features than other models, as it minimizes the loss function by adding new decision trees that predict the residuals or errors made by the existing trees. XGBoost can also handle unbalanced datasets since it creates

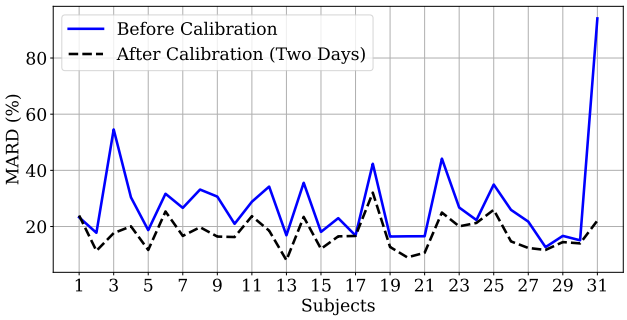


Figure 16: The impact of calibration on the performance of GlucoSense. The experiments are conducted using the leave-one-participant-out method.

an ensemble of decision trees for prediction. Furthermore, in our experiments, the training and testing of the XGBoost model were much faster than other models.

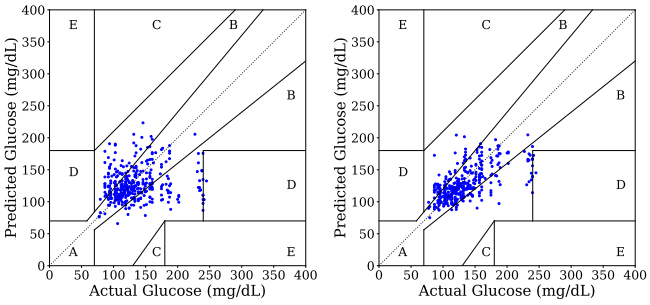
A.3 More Evaluation Results and Figures

Generalization and Calibration of GlucoSense. We assess the generalizability of GlucoSense and the impact of calibration on the performance. We train the regression model using the leave-one-out approach. That is, we train the model on 30 participants and test on the left-out participant. We repeat this process 31 times, each time testing on a different participant. We summarize the results in Figure 16, which show the MARD (blue curve) value in each case. The x-axis denotes the participant left out. The results show that the overall performance decreases due to the smaller number of participants. The performance is notably worse for some participants, e.g., 31 and 3. Participant 31 is the only black person in the user study. The model did not see any darker skin tone during training, which is the main reason behind the poor performance. Participant 3, on the other hand, has low glucose values (around 70 mg/dL) compared to the rest of the participants. Other participants either have glucose values in the normal range (70–180 mg/dL) for healthy people or high glucose values (> 180 mg/dL) for diabetic people. The training data has very few data points in the low glucose range. Datasets from larger and more diverse user populations will be needed to potentially deploy the GlucoSense in real life.

To partially mitigate the effect of our small datasets, we recommend a *calibration* step for GlucoSense. In this case, GlucoSense is initially fine-tuned on a few data points from the user. To demonstrate the impact of calibration, we fine-tuned GlucoSense on two days of the glucose readings from the left-out participant. Then, we assess the accuracy on the remaining days after fine-tuning and plot results in Figure 16 (black dashed curve). The figure shows performance improvement for all participants, including 31 and 3. We

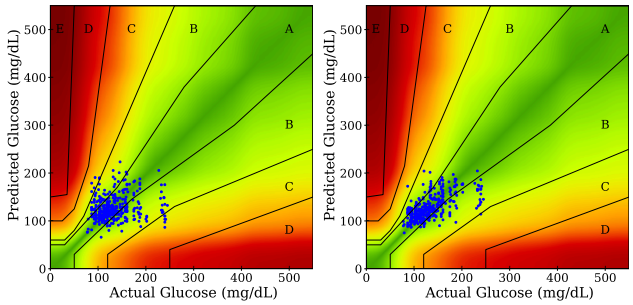
	Clark Error Grid (%)			Surv. Error Grid (%)			MARD
	A	B	D	None	Slight	Mod.	
XGBoost (ours)	80.4	19.3	0.3	76.0	23.7	0.3	0.1297
KNN	77.4	22.3	0.3	74.9	24.8	0.3	0.1339
RF	74.9	24.5	0.5	74.9	25.1	0.0	0.1376
SVR	70.5	29.0	0.5	64.7	35.3	0.0	0.1592
PLSR	57.3	42.1	0.6	51.2	48.8	0.0	0.1890
Neural Network	78.0	21.5	0.5	74.7	25.1	0.3	0.1380

Table 5: Performance of various machine learning regression models for glucose prediction.



(a) RGB-NoIR **(b) Custom**

Figure 17: Clarke error grid analysis of GlucoSense on (a) RGB-NoIR, (b) Custom sensor. Almost all predicted glucose values fall in regions A and B (clinically accurate/acceptable).



(a) RGB-NoIR **(b) Custom**

Figure 18: Consensus error grid overlaid on surveillance error grid for the analysis of GlucoSense on (a) RGB-NoIR, (b) Custom sensor. All predicted glucose values fall in regions A and B and within the no (green) or slight (yellow) risk areas.

	Clark Error Grid (%)		
	A	B	D
RGB+NIR	**80.4**	**19.3**	**0.3**
RGB+ToF	78.2	21.5	0.3
RGB-NoIR	59.9	39.6	0.5
Custom	71.8	27.6	0.6

Table 6: Summary of the results using Clarke error grid. In all cases, 99% of GlucoSense predictions fall within zones A and B, which are clinically accurate and acceptable, respectively.

note that participant 18 observed a relatively smaller performance improvement compared to other participants after calibration. After investigation, we found that most of their captured image had significant obstructions from the hair on their hand. Thus, we plan to implement image processing methods to reduce the impact of hair obstruction to address such cases.

In summary, the experiments in this section demonstrate the potential of the calibration step in generalizing GlucoSense to broader and more diverse user populations.