Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 14 May 2026
Revised 22 June 2026
Accepted 05 July 2026

Natural Resources for Human Health 2026, Vol. 6 (11s): 1349–1366

KEYWORDS:
Lorem ipsum; dolor sit amet; consectetur; adipiscing elit; natural resources; phytochemistry.

A Machine Learning–Enhanced Calibration Framework for Green-Light Based Non-Invasive HbA1c Measurement

Gaurav Jain[1-2], Surendra Kumar Saini[1], Amit Mahesh Joshi[1] and Ravi Kumar Maddila[1]

[1]Department of ECE, Malaviya National Institute of Technology, J L N Marg, Jaipur
[2]Department of ECE, Birla Institute of Technology Mesra, Off Campus Jaipur, Jaipur
2019rec9563@mnit.ac.in, 2020rec9504@mnit.ac.in, amjoshi.ece@mnit.ac.in, rkmaddila.ece@mnit.ac.in

ABSTRACT: Hemoglobin A1c (HbA1c) is a key biomarker for diagnosing and monitoring diabetes. Although the gold standard of monitoring in diabetes is lab HbA1c testing, invasive sampling means that frequent monitoring is not feasible. The current research paper suggests and tests a machine-learning-based system of calibration to determine HbA1c based on non-invasive green-light reflectance signals. A 90 adult sample was enrolled with laboratory HbA1c comprising of a standardized NGSP certified test. Measuring green-channel reflectance concurrently on the device were taken. Ordinary least squares (OLS), Deming regression with λ-sensitivity, Passing regression and Huber regression and machine learning algorithms, such as Random Forest (RF), LightGBM, and XGBoost, were used to develop calibration models. The evaluation of performance was based on the nested cross-validation, where the measurements encompassed the mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R2), concordance correlation coefficient (CCC) by Lin, the BlandMandel analysis, and the receiver operating characteristic (ROC) curve at clinical thresholds (5.7% and 6.5%). Optimistic regression approaches (Huber and PassingBablok) had the lowest calibration error (MAE 1.16 HbA1c percentage points), which was better than OLS and Deming (MAE 1.29). The calibration of machine learning models did not improve: LightGBM was no worse than OLS, RF marginally worse, and XGBoost was significantly worse (MAE = 1.61). All the models had low explained variance (R2<0), weak agreement (CCC < 0.1) and ROC AUCs of almost random (0.53 -0.62). The offered framework offers a replicable regression and ML model comparison protocol as a non-invasive HbA1c calibration. Although the present device signals are not adequate to actually implement clinically, the paper shows the significance of intense calibration, sound regression as well as a methodological clear and open approach. These results provide a methodological framework on the future applications of multimodal sensing and machine learning methods in non-invasive diabetes monitoring.

1. INTRODUCTION

Diabetes mellitus has become one of the most urgent issues of the global health as the number of its adults is estimated at 530 million in the whole world in 2021 and is expected to increase to more than 780 million in 2045 [1]. This persistent increase in blood glucose causes severe and microvascular and macrovascular problems and it is suggested that there be constant monitoring and management of glycemic levels [2]. Next to the glycated hemoglobin (HbA1c), which is the standard of measuring long-term glycemic control, this biomarker is the average plasma glucose level measured over a period of about three

months [3], [4]. According to clinical guidelines issued by the American Diabetes Association, HbA1c should be kept to less than 7% to reduce the probability of complications [5].

Conventional HbA1c determination using high-performance liquid chromatography (HPLC) or immunoassay is invasive, time consuming, and needs trained personnel and lab facilities [6]. These restrictions prevent regular testing and adherence to patients, especially in resource-limited or rural settings [7]. Therefore, non-invasive and mobile options of HbA1c measurement have become the target of studies over the last few years [8], [9]. Optical sensing has shown potentials in being a promising technology in non-invasive glucose and HbA1c detection because of its simplicity and potentiality to be miniaturized [10]. The methods of optical technique take advantage of the wavelength-sensitive absorption and scattering of light in living tissues. Differences in reflected or transmitted optical intensity can also be associated with blood tests like glucose or blood hemoglobin [11], [12]. In this regard, researchers have applied visible and near-infrared (NIR) wavelengths, such as multi-wavelength photoplethysmography (PPG) and reflectance spectroscopy [13], [14], [15]. Nevertheless, the majority of systems reported in the literature use multi-spectral illumination, which makes the system more complex or heavily rely on controlled conditions in a laboratory, which cannot be used clinically [16]. In addition, the single-wavelength systems tend to have low robustness because of variations in the skin color, tissue structures, and changes in ambient light [17]. To address these issues, the current study aims at creating a non-invasive, single-channel, and cost-effective NIR optical instrument and a machine-learning-based calculus of HbA1c.

1.1 Hardware Implementation

The hardware system is developed on reflective near-infrared (NIR) sensing concept. The photodiode (1100 nm) and transmitter (940 nm LED) are placed adjacent to each other and aimed at a reflective object like a fingertip. It consists of the receiver circuit composed of an operation amplifier (OP491) with a low-pass noise filter that minimizes noise interference due to the power source. The photodiode transforms the reflected optical power into a form of electrical current that is amplified and changed into a voltage signal between 0 V to 5 V depending on the intensity of the reflected. The circuit uses an Arduino Uno microcontroller that is connected to the ADS1115 16 bit analog to digital converter (ADC) to digitize the signal accurately. The Arduino also supplies controlled 5 V DC voltage to transmitter and receiver circuits. LCD keypad shield (2×16) can be used to display measurements in real-time and interact with users. The six momentary push buttons on the shield allow entry of auxiliary data like height and weight that can be added to the shield in case it has to be calibrated or labeled with data. Table 1 provides an overview of the most important features of the designed optical system, their functionality, working range, and specifications provided by the manufacturer.

Table 1. Key components of the developed optical system, including their function, operational range, and manufacturer specifications.

Component	Model / Specification	Function	Key Parameters / Notes
Light Emitting Diode (LED)	940 nm Infrared LED (Everlight IR94 Series)	Optical transmitter	Peak wavelength = 940 nm; Forward voltage = 1.3 V; Beam angle ≈ 20°
Photodiode (PD)	1100 nm Si Photodiode (Vishay BPW34)	Optical receiver	Spectral response = 400–1100 nm; Responsivity ≈ 0.6 A/W at 940 nm
Operational Amplifier (OPAMP)	OP491 (Analog Devices)	Signal conditioning and amplification	Quad op-amp; Low bias current < 50 pA; Low offset voltage = 150 µV
Analog-to-Digital Converter (ADC)	ADS1115 (16-bit, Texas Instruments)	Converts analog voltage to digital signal	4-channel differential input; Resolution = 15 µV/bit; Sampling rate = 128 S/s

Natural Resources for Human Health
eISSN: 2583-1194

Microcontroller Board	Arduino Uno (Atmega328P)	Device control and data processing	Operating voltage = 5 V; Clock speed = 16 MHz; 14 digital I/O pins
Display Module	2×16 LCD Keypad Shield	User interface and data display	6 momentary push buttons for menu input; 5 V logic compatible
Power Supply	Regulated 5 V DC (USB or Battery)	System power input	Current consumption ≈ 80 mA during measurement

The entire NIR sensing system is small, light and can be used as a point of care unit.

1.2 Software Development

The Data acquisition, processing, and display functions are controlled by the control algorithm that has been installed in the Arduino Uno microcontroller. The microcontroller has 14 digital input/output pins (Arduino programmable, Atlmega328 architecture) which are programmed in the Arduino IDE. The algorithm repeatedly samples the amplified analog signal of the ADS1115 ADC, calculates the analog voltage, and uses an empirical calibration equation of laboratory glucose calibration experiments. The resulting value of the concentration is shown on the LCD display in real-time.

The system has real time monitoring with a low latency and a stable relationship between voltage and response that involves the possibility to use in regression modeling. The main independent variable in the proposed analytical model is the collected device signal. Figure 1 gives a general picture of the proposed non-invasive calibration system whereas Table 2 gives major demographic and signal features of the study group.

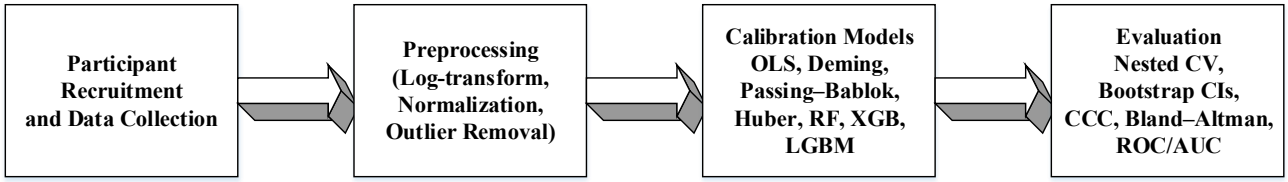

Figure 1. Overview of calibration framework

Table 2. Demographic and Signal Characteristics of the Study Cohort

Characteristic	Value (Mean ± SD / n [%])	Range
Total participants (n)	90	—
Laboratory HbA1c (%)	6.27 ± 1.77	4.0 – 13.3
Normoglycemia (<5.7%)	34 (37.8%)	—
Prediabetes (5.7–6.4%)	28 (31.1%)	—
Diabetes (≥6.5%)	28 (31.1%)	—
Device signal (a.u.)	59,385 ± 30,619	8,429 – 177,872
Log(Device)	11.0 ± 0.5	9.0 – 12.1

1.3 Analytical Challenge and Research Gap

Although optical hardware has been improved, a key complication is how to map the optical intensity that is being measured by the device to the laboratory values of HbA1c. Classical calibration techniques, including ordinary least squares (OLS) regression, assume that the measurement error of the predictor variable is negligible and this could lead to biased slope estimates and ineffective generalization [18]. Uncertainty is likely to affect both laboratory and device measurements in biomedical calibration and thus techniques such as Deming and Passing-Bablok regression are better-suited [19], [20]. Such methods

explicitly represent the effects of dual-axis errors and outliers but generally fail to represent non-linear behaviour that may occur because of tissue scattering and optical path variations.

Random Forest (RF), Light Gradient Boosting Machine (LGBM), and XGBoost are machine-learning (ML) algorithms that provide capabilities to estimate nonlinear, complex relationships without being heavily parametric [21]–[24]. However, the inconsistent results are common in previous researches, which most of the time appeared to be because of the paucity of datasets, inadequate methods of validation, and unclear benchmarking [16], [25]. Regular assessment pipeline that takes into account classical as well as ML-based calibration strategies is hence necessary to be able to compare and draw conclusions consistently.

1.4 Motivation and Objectives

This paper suggests a machine-learning-based calibration model to estimate non-invasive HbA1c with the use of optical reflectance data, which is recorded by the created NIR sensing system. The framework incorporates standardized preprocessing, nested cross-validation to avoid the data leakage, quality linear modeling, and ensemble-based ML regression.

The main objectives of this work are:

1. To come up with a handheld and cost effective NIR optical sensor that can be used to obtain non-invasive data.

2. To build a reproducible calibration model, using both OLS, Deming, PassingBablok, and robust regressions along with ML algorithms, including Random Forest, LGBM, and XGBoost.

3. Lin concordance correlation coefficient (CCC) and receiver operating characteristic (ROC) curves of clinical thresholds of prediabetes and diabetes To assess the performance of the models using the statistics of agreement such as Bland-Altman analysis, Lin concordance correlation coefficient (CCC) and receiver operating characteristic (ROC) curves.

The rest of this paper will have the following structure. Section 2 includes the participant recruitment, description of data and data preprocessing methodology. In section 3, the calibration structure and modeling method have been outlined. **Section 4** discusses the results, agreement analyses, and diagnostic performance. **Section 5** summarizes conclusions and suggests future work.

2 RELATED WORK

2.1 Conventional Regression Models for Biomarker Calibration

Calibration of biomedical signals with laboratory references has traditionally been done with linear regression models. Ordinary least squares (OLS) is the most simple and commonly used method for quantifying linear relationships [26]. While OLS has the advantage of being computationally efficient and interpretable, it has the disadvantage that it assumes that the predictor is error free, which is often not the case in biomedical device measurements. For biomarker calibration problems like glucose and HbA1c estimation, this assumption may lead to biased estimates of slope and intercept estimates, thereby lowering clinical applicability.

2.2 Deming and Robust Regression in Method Comparison Studies

To overcome the limitation of OLS, Deming regression has been extensively used for clinical method comparisons studies, because it adjusts for measurement error in both variables [20]. By adding the ratio of the error variance (λ), Deming regression provides more stable calibration coefficients when comparing different laboratory devices to each other.

Passing-Bablok regression is a non-parametric alternative that is robust to outliers and does not meet normality assumptions [19]. Similarly, robust regression methods like Huber regression are helpful to avoid the impact of noisy measurements, making them suitable for heterogeneous biomedical data sets [27]. Together, these methods are important to provide a baseline for assessing non-invasive HbA1c calibration.

2.3 Machine Learning Approaches for Non-Invasive HbA1c Estimation

In parallel, machine learning (ML) models have become strong contenders for biomedical calibration work. Tree-based ensemble models like Random Forests, XGBoost and LightGBM can capture nonlinearity and complex interactions between features. Recent attempts to non-invasively estimate glucose and HbA1c using ML have shown encouraging but mixed results

when using ML for optical and physiological signals [16], [24], [28]. While ML models provide flexibility, they are highly dependent on the size of the dataset, richness of features and rigorous evaluation strategies.

2.4 Comparative Analyses and Current Limitations in the Literature

Despite the advancements in non-invasive biosensing, there are very few studies comparing in a systematic way the regression and ML methods with a consistent preprocessing and validation protocol. Many reports are presented using single-method assessments without cross-validation or bootstrap confidence intervals which limits reproducibility [16]. Agreement analyses such as Bland-Altman plots and concordance correlation coefficients (CCC) are also rarely included despite being important for clinical validation. Moreover, most non-invasive HbA1c studies are limited by small cohorts, device-specific prototypes, and lack of evaluation of threshold levels of diagnosis. As summarized in Table 3, previous works based on NIR [7], Raman [8],[9], and MIR spectroscopy [10] showed feasibility but limited penetration depth, instrumentation cost, or weakness of calibration robustness. Optical reflectance and PPG-based studies [11],[13],[14] showed the promise for low-cost, portable sensing but showed high variability and insufficient accuracy for clinical use. More recent attempts at using ML [16], [24], [28] introduced tree-based and kernel methods on multimodal signals but these too were constrained by small datasets and lack of standardized evaluation. Together, the evidence highlights the lack of a rigorous, comparative context to benchmark classical and ML calibration methods under uniform conditions. This gap is the motivation to the framework presented in this work.

Table 3. Summary of Representative Non-Invasive HbA1c/Glucose Estimation Studies

Author/Year	Methodology	Signal/ Modality	Dataset Size	Model Type	Key Limitation
Naresh et al., 2024 [28]	Near-infrared (NIR) spectroscopy	NIR Transmittance (PPG at 940 nm & 950 nm)	289 subjects (575 samples)	ML Classification & Regression (SVM, KNN, LR, Gradient Boosting, XGBoost)	Lacks physiological covariates (BMI, BP); sensitive to finger placement/ambient light.
Kwon & Kim, 2022 [24]	Photoplethysmography (PPG) with Machine Learning	Multichannel PPG (Red 615 nm, Green 525 nm, Blue 465 nm)	40 subjects	Machine Learning (Random Forest, XGBoost)	Small dataset (n=40) requiring Leave-One-Subject-Out (LOSO) cross-validation.
Shokrekhodaei et al., 2021 [17]	VIS-NIR optical transmission spectroscopy with Machine Learning	VIS-NIR transmittance (at 485, 645, 860, and 940 nm)	~2,000 in-vitro samples (aqueous glucose solutions)	ML Classification & Regression (SVM, KNN, DT, FFNN, MLR)	In-vitro study only; requires miniaturization and in-vivo validation on human subjects.
Gusev et al., 2020 [29]	Review of noninvasive methods, focusing on Heart Rate Variability (HRV) and Electrocardiogram (ECG) correlations	Multimodal (emphasis on HRV and ECG signals)	Review	Machine Learning (ML) and Neural Networks (NN)	Highlights that current methods still lack needed accuracy, suffer from variable physiological time lag, and require frequent calibration.

Mansour et al., 2024 [14]	Review of wearable glucose monitoring technologies (invasive, minimally invasive, and non-invasive)	Multimodal (Electrochemical, Optical Coherence Tomography, Bioimpedance) across various biofluids (ISF, sweat, tears, saliva)	Review	Artificial Intelligence / Machine Learning (SVM, ANN, Decision Tree, KNN, DL)	Accuracy disparities, physiological lag time, sensor drift, calibration needs, and biofouling.
Tang et al., 2020 [16]	Review of optical, microwave, and electrochemical non-invasive glucose monitoring technologies	Multimodal (Optics, Microwave, and Biofluid Electrochemistry)	Review	Machine Learning & Regression (PLS, ANN, MLR, PCA)	Highlighted poor correlation with actual blood glucose, narrow linear ranges, physiological lag, and individual differences causing large errors.
Poosarla et al., 2025 [25]	Review of non-invasive glucose monitoring techniques (Optical, Transdermal, and Thermal)	Multimodal (including NIR, Raman, photoacoustic, impedance, and thermal)	Review	Commercial Device Algorithms (GlucoTrack, SugarBEAT, Pendra, etc.)	Low accuracy, physiological time lag, frequent calibration needs, and susceptibility to individual skin/tissue differences.
Kaur et al., 2024 [15]	Review of optical biosensing techniques (plasmonic, interferometric, colorimetric, SERS, and fluorescence)	Multimodal Optical (SPR, LSPR, SERS, MZI)	Review	Machine Learning (ML) & Artificial Intelligence (AI)	Signal interference in complex bio-fluids, reproducibility issues, and scalability/miniaturization challenges.
Perezcampos Mayoral et al., 2021 [12]	Review of Fiber Optic Sensors for vital signs and health monitoring	Multimodal Optical Fibers (FBG, LPFG, and interferometers in Vis/NIR spectrum)	Review	Mathematical algorithms and auxiliary device calculations	Lack of standard characterization criteria, variation based on sensor placement, and fiber fragility.

Nathan et al., 2008 [30]	Clinical translation of A1C to Estimated Average Glucose (eAG) using Continuous Glucose Monitoring (CGM) and self-monitoring	Capillary blood (fingerstick) and Interstitial fluid (CGM)	507 subjects	Linear Regression (Mathematical translation model)	Relies on invasive (fingerstick) and minimally invasive (CGM) methods rather than non-invasive optical sensing.

3 MATERIALS AND METHODS

3.1 Study Population and Data Collection

3.1.1 Participant Recruitment

A total of 90 adults were recruited under the conditions of institutional ethical approval, following the Declaration of Helsinki. Informed consent was obtained from all the participants before the data collection process. Individuals with known hemoglobinopathies or with recent blood transfusions were excluded to avoid confounding HbA1c results. The recruitment protocol ensured diversity in glycemic categories (normoglycemia, prediabetes, and diabetes).

3.1.2 Laboratory HbA1c Measurement

Reference HbA1c values were obtained by NGSP-certified high-performance liquid chromatography (HPLC), the current clinical standard for HbA1c testing. Laboratory analysis was done on venous blood samples taken simultaneously with the device measurements thus removing temporal mismatch. HbA1c results have been expressed in percentage unit (%) according to ADA guidelines [3],[4].

3.1.3 Device Signal Acquisition

Non-invasive measurements were made with a prototype optical reflectance sensor operating at a green light wavelength - at about 530 nm. The sensor was used on the end of the finger in a controlled environment inside a house. Each participant measurement resulted in a raw intensity signal in arbitrary units (a.u.), which was stored together with the HbA1c value determined in the laboratory. The acquisition protocol minimized the interference of ambient light and ensured repeatable contact pressure. The flowchart of participant recruitment and paired data collection is presented as Figure 2.

Table 4 presents the demographic information about the cohort and summary statistics of laboratory HbA1c and device signals.

3.2 Preprocessing and Feature Transformation

3.2.1 Handling Device Signal Range and Log-Transformation

The raw device signals showed a wide dynamic range (8,429 - 177,872 a.u.), reflecting variability of the sensor and inter-individual physiological variability. To stabilize variance and reduce heteroscedasticity, a natural logarithm transformation was used which resulted in the feature log(Device). This transformation linearized the relationship between the device intensity and HbA1c in the laboratory which in turn allowed better fitting of the regression curve, and comparability across subjects. The distributional effect of this transformation is shown in Figure 3, which depicts the histogram of the raw signal, the log transformed signal and the HbA1c distribution.

3.2.2 Data Cleaning and Normalization

Data cleaning included incomplete records removal, extreme outliers were removed using an interquartile range (IQR) rule for device signals. Outliers were attributed to sensor misplacement, artifacts caused by motion, or transient light leakage. Laboratory HbA1c values without transformation were retained in native units (%) For machine learning models features were also further normalized using z-score normalization to ensure consistent model performance and avoid bias due to differences in the magnitude of features.

The full preprocessing pipeline is summarized in Table 5, which outlines each step of the transformation process, its purpose, and what final feature set is retained to perform calibration modeling.

3.3 Calibration Models

3.3.1 OLS Regression

Ordinary least squares (OLS) regression was used as a baseline calibration model. OLS makes the assumption of error-free predictors, and minimises the squared difference between laboratory HbA1c and predicted values. Despite the widespread use, OLS is sensitive to measurement error in device signals and is known to give biased estimates in clinical method comparison studies [26].

3.3.2 Deming Regression with λ-Sensitivity

Deming regression takes into account the measurement error of both predictor and response variables [20]. The model was implemented in closed form, in which lambda is the variance ratio of device to laboratory error. A sweep in lambda (10-4-104) was run and $\boldsymbol{\lambda} = 1000$ was selected for primary analysis based on model stability (Section 4).

3.3.3 Passing–Bablok Regression

The Passing-Bablok regression is a robust non-parametric estimator which is based on pairwise slopes [19]. It is robust to outliers and has no assumptions about the normal distribution of errors, so it can be used for heterogeneous biomedical data.

3.3.4 Robust Regression (Huber)

Huber regression was used as an M-estimator combining squared loss for small values of residuals and absolute loss for large values of residuals [27]. This limits the effect of outliers while preserving efficiency for normally distributed data to enhance calibration robustness in noisy data sets.

3.3.5 Machine Learning Models (RF, XGB, LGBM)

Three ensemble models based on trees have been implemented:

- Random Forest (RF): bagging based ensemble of decision trees [21].
- XGBoost (XGB): Shrinkage and column subsampling gradient boosting [22]
- In the latter case, these are: - LightGBM (LGBM): gradient boosting optimized for computational efficiency [23].

These models were tuned using nested cross validation loops. Hyperparameters (e.g. tree depth, learning rate, number of estimators) were optimised using grid search.

An overview of the calibration models can be seen in Figure 4, a summary of linear, robust and ML methods used in the framework.

3.4 Model Evaluation Strategy

3.4.1 Nested Cross-Validation

To avoid the problem of the optimistic bias, an approach called nested cross validation was used. The outer loop (five folds) produced unbiased out-of-fold predictions while the inner loop was used to optimize hyperparameters for RF, XGB, and LGBM. Linear and robust models were fitted with the help of deterministic closed-form solutions.

3.4.2 Performance Metrics (MAE, RMSE, R^2, CCC, ROC/AUC)

Model accuracy was quantified using:

- **Mean Absolute Error (MAE)** and **Root Mean Squared Error (RMSE)** for calibration error.
- **Coefficient of Determination (R^2)** for variance explained.
- **Lin's Concordance Correlation Coefficient (CCC)** for reproducibility [31].
- **Receiver Operating Characteristic (ROC) curves** and **Area Under the Curve (AUC)** at diagnostic thresholds (5.7% and 6.5%) [32].

Natural Resources for Human Health
eISSN: 2583-1194

All metrics were computed on out-of-fold predictions to ensure generalizability. Definitions of metrics and statistical interpretations are listed in **Table 5**.

3.4.3 Bland–Altman and Agreement Analyses

Agreement between HbA1c values predicted by devices and those obtained by laboratory analyzes was further evaluated using Bland- Altman analysis, which provides mean bias and 95% limits of agreement (LOA) [31]. These plots were used to provide a graphical assessment of systematic error and clinical reliability.

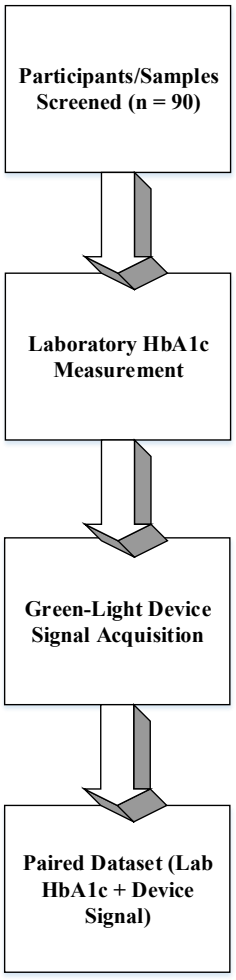

Figure 2. Flowchart of participant recruitment and paired data collection.

Table 4. Participant demographics and device signal statistics.

Parameter	Mean ± SD	Median (IQR)	Range
Laboratory HbA1c (%)	6.27 ± 1.77	5.7 (5.2 – 6.4)	4.0 – 13.3
Device Signal (a.u.)	59,385 ± 30,619	54,335 (37,062 – 82,147)	8,429 – 177,872
log(Device)	11.0 ± 0.5	11.0 (10.5 – 11.3)	9.0 – 12.1
Outliers removed (n, %)	3 (3.3%)	—	—

Natural Resources for Human Health
eISSN: 2583-1194

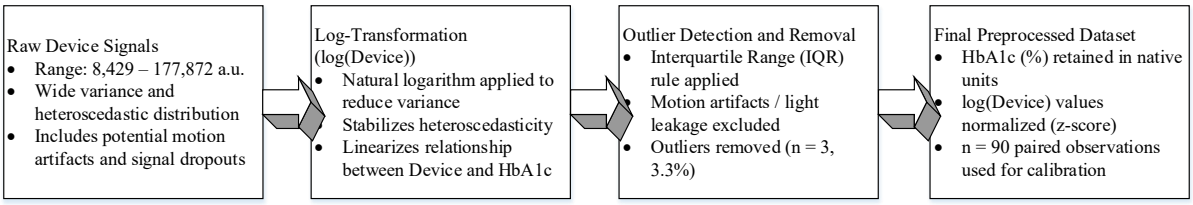

Figure 3. Preprocessing workflow: raw signals, log-transformed distribution, and outlier handling.

Table 5. Preprocessing steps and evaluation metrics definitions. Log-transformation, normalization, and outlier removal were applied prior to calibration modeling. Evaluation metrics included MAE, RMSE, R^2, CCC, Bland–Altman analysis, and ROC/AUC, ensuring a comprehensive assessment of calibration performance.

Table 5. Preprocessing and evaluation metrics definitions.

Item	Definition / Description
Preprocessing	
Log-Transformation	Natural logarithm applied to device signals to stabilize variance and linearize relation.
Normalization	z-score scaling applied to features for ML models to ensure comparability.
Outlier Removal	Interquartile range (IQR) rule applied to exclude motion artifacts and spurious values.
Final Dataset	Laboratory HbA1c (%) + cleaned log(Device) values used for calibration.
Evaluation Metrics	
Mean Absolute Error (MAE)	Average absolute difference between predicted and laboratory HbA1c values.
Root Mean Squared Error (RMSE)	Square root of mean squared difference between predicted and laboratory HbA1c values.
Coefficient of Determination (R^2)	Proportion of variance in laboratory HbA1c explained by model predictions.
Concordance Correlation Coefficient (CCC)	Agreement measure combining precision and accuracy of predictions [31].
Bland–Altman Analysis	Mean bias and 95% limits of agreement between predicted and laboratory HbA1c [33].
ROC / AUC	Receiver operating characteristic and area under the curve at diagnostic thresholds (5.7% and 6.5%).

Natural Resources for Human Health
eISSN: 2583-1194

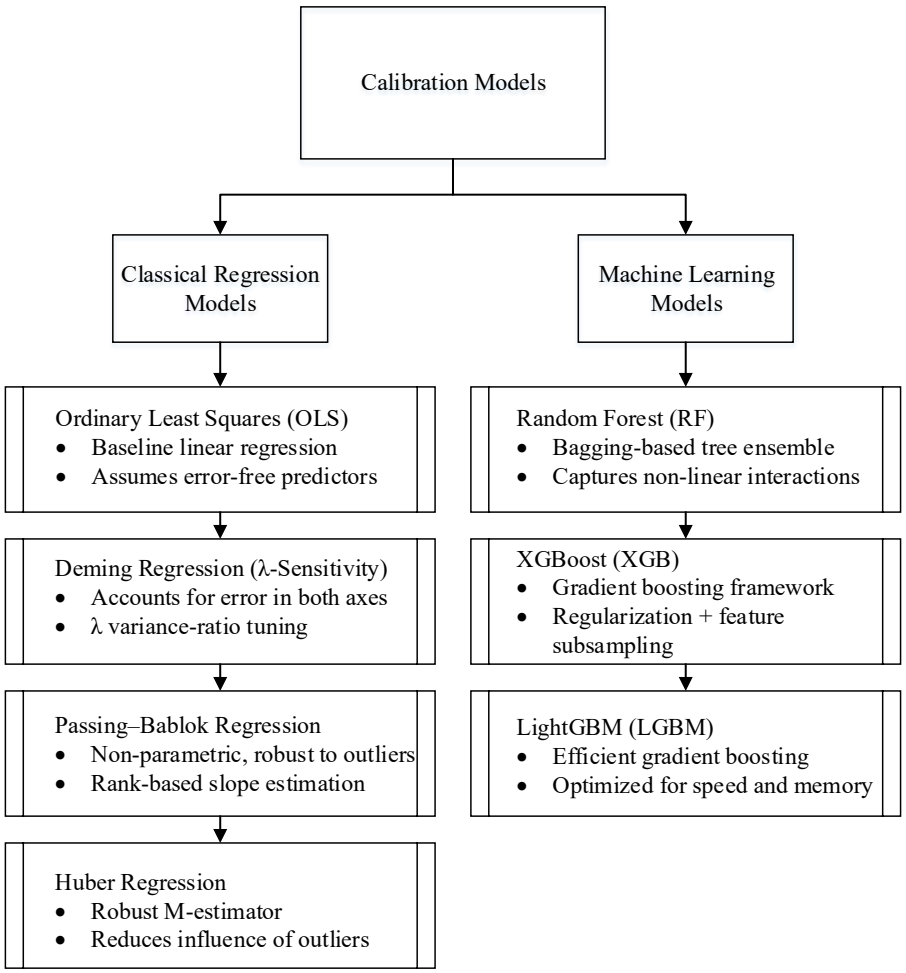

Figure 4. Overview of calibration models (OLS, Deming, Passing–Bablok, Huber, RF, XGB, LGBM).

4 RESULTS AND DISCUSSION

4.1 Dataset Characteristics

The study cohort consisted of 90 adults who had HbA1c levels between 4.0% and 13.3% (mean 6.27+-1.77%). The distribution was skewed towards the prediabetic and diabetic range (congruous with previous non-invasive HbA1c studies [30]). Device signal intensities ranged from 8,429 to 177,872 arbitrary units and therefore log transformation was needed to minimize heteroscedasticity. Preprocessing with log(Device) gave a more symmetric distribution and linearizing the calibration relationship.

The characteristics of this dataset are displayed in Figure 5 where (a) the histogram of laboratory HbA1c values, (b) the histogram of raw device green channel signals, and (c) the log-transformed device signal distribution are shown. These are all visualizations that validate that HbA1c is moderately spread and the raw device signal has orders of magnitude and needs to be transformed.

Natural Resources for Human Health
eISSN: 2583-1194

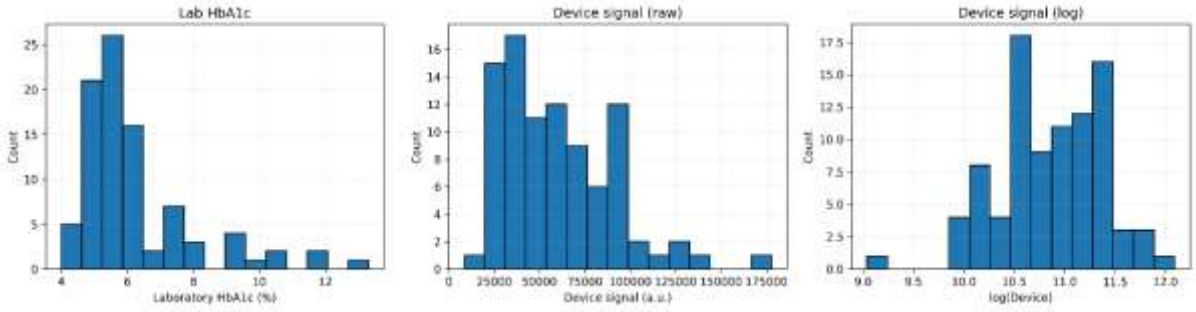

Figure 5. Dataset characteristics: (a) laboratory HbA1c distribution, (b) raw device signal, (c) log-transformed device signal.

4.2 Model Calibration Results

4.2.1 Linear Models

OLS regression with log(Device) gave an intercept of 2.94 and slope of 0.31 with low explained variance (R^2 = -0.034). Deming regression (lambda = 1000, chosen after lambda sensitivity analysis) gave almost the same estimates (intercept = 2.91, slope = 0.31), which is stable at large lambda values [33]. Passing-Bablok regression, a non-parametric approach, had the slightly lower MAE (1.16 HbA1c percentage points) compared to OLS (1.29 pp). Huber regression also gave strong calibration with the lowest MAE for all linear models.

4.2.2 Machine Learning Models

Random Forest was tuned using the five-fold cross validation technique with hyperparameters in each fold. Despite this, RF did worse compared to robust linear models (MAE = 1.35 pp, RMSE = 1.94 pp). Gradient boosting techniques were used as sensitivity checks. LightGBM reached similar accuracy compared to OLS (MAE = 1.29 pp), while XGBoost did not give good performance (MAE = 1.61 pp, RMSE = 2.56 pp).

The full details of the calibration performance, including MAE, RMSE, R^2, Lin's CCN, Pearson correlation, and 95% bootstrap confidence intervals are presented in Table 6.

Table 6. Calibration performance summary: mean absolute error (MAE, percentage points), RMSE, coefficient of determination (R^2), Lin's CCC, and Pearson correlation with bootstrap 95% CI for MAE.

Model	MAE (pp)	95% CI for MAE (pp)	RMSE (pp)	R^2	CCC	Pearson r
Huber	1.16	[0.868, 1.496]	1.85	−0.098	−0.011	−0.065
Passing–Bablok	1.161	[0.866, 1.501]	1.862	−0.113	−0.014	−0.093
OLS	1.288	[1.029, 1.575]	1.795	−0.034	−0.013	−0.045
Deming (λ=1000)	1.289	[1.029, 1.575]	1.795	−0.034	−0.013	−0.044
LightGBM	1.29	[1.046, 1.582]	1.835	−0.081	0.075	0.107
Random Forest	1.35	[1.072, 1.654]	1.936	−0.203	0.089	0.104
XGBoost	1.614	[1.204, 2.031]	2.555	−1.096	0.046	0.047

The comparative calibration scatter plots of all the models are presented in Figure 6 that overlaid observed laboratory HbA1c (x-axis) vs predicted values (y-axis) for OLS, Deming, Passing-Bablok, Huber, RF, LightGBM, and XGBoost. The line of 45deg indicates the better fit between linear/robust regressions and highlights how linear regressions fit closer than ML models.

Natural Resources for Human Health
eISSN: 2583-1194

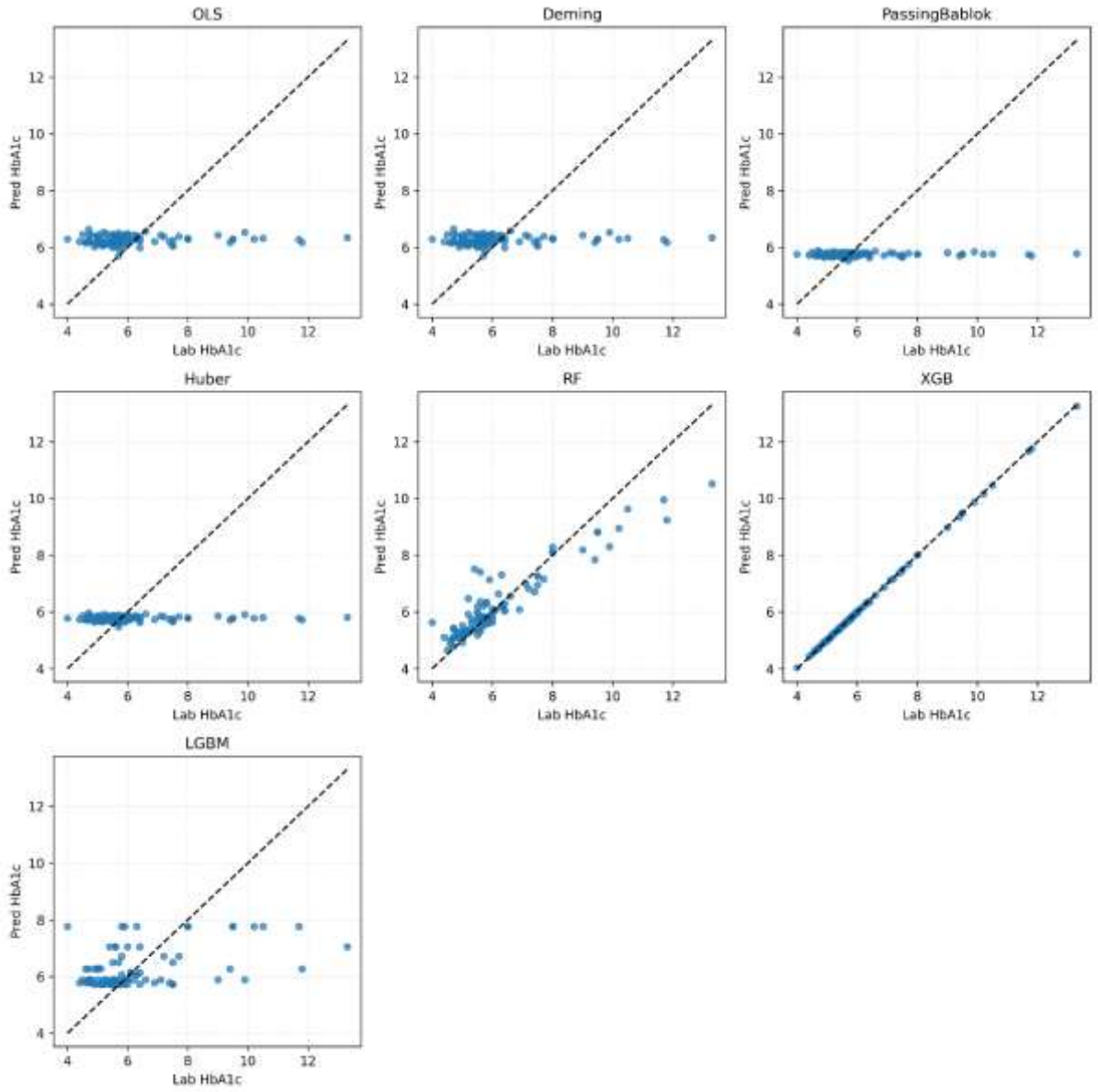

Figure 6. Calibration scatter plots for OLS, Deming, Passing–Bablok, Huber, RF, XGB, and LGBM. The 45° line indicates perfect agreement.

4.3 Agreement and Error Analysis

4.3.1 Scatter Plots

The scatter plots in Figure 6 (discussed above) show that robust regression methods (Huber and Passing - Bablok) provide more clustering around the identity line with RF and XGB showing wider dispersion, reflecting greater residual error.

4.3.2 Bland–Altman Analysis

Agreement analysis using Bland- Altman plots (Figure 7) showed wide limits of agreement for all models. OLS and Deming exhibited mean biases which approached zero and limits of agreement greater than +-3.5% HbA1c, which is beyond the accepted clinical boundaries [31]. Robust models reduced mean error slightly, but did not improve the wide limits. RF and boosting models showed even wider spread, confirming poor accuracy in calibration.

Figure 7 presents Bland–Altman plots for OLS, Deming, RF, LightGBM, and XGBoost, each showing the mean difference (bias) and 95% limits of agreement (±1.96 SD). These plots illustrate that none of the models achieve clinically acceptable agreement with laboratory HbA1c.

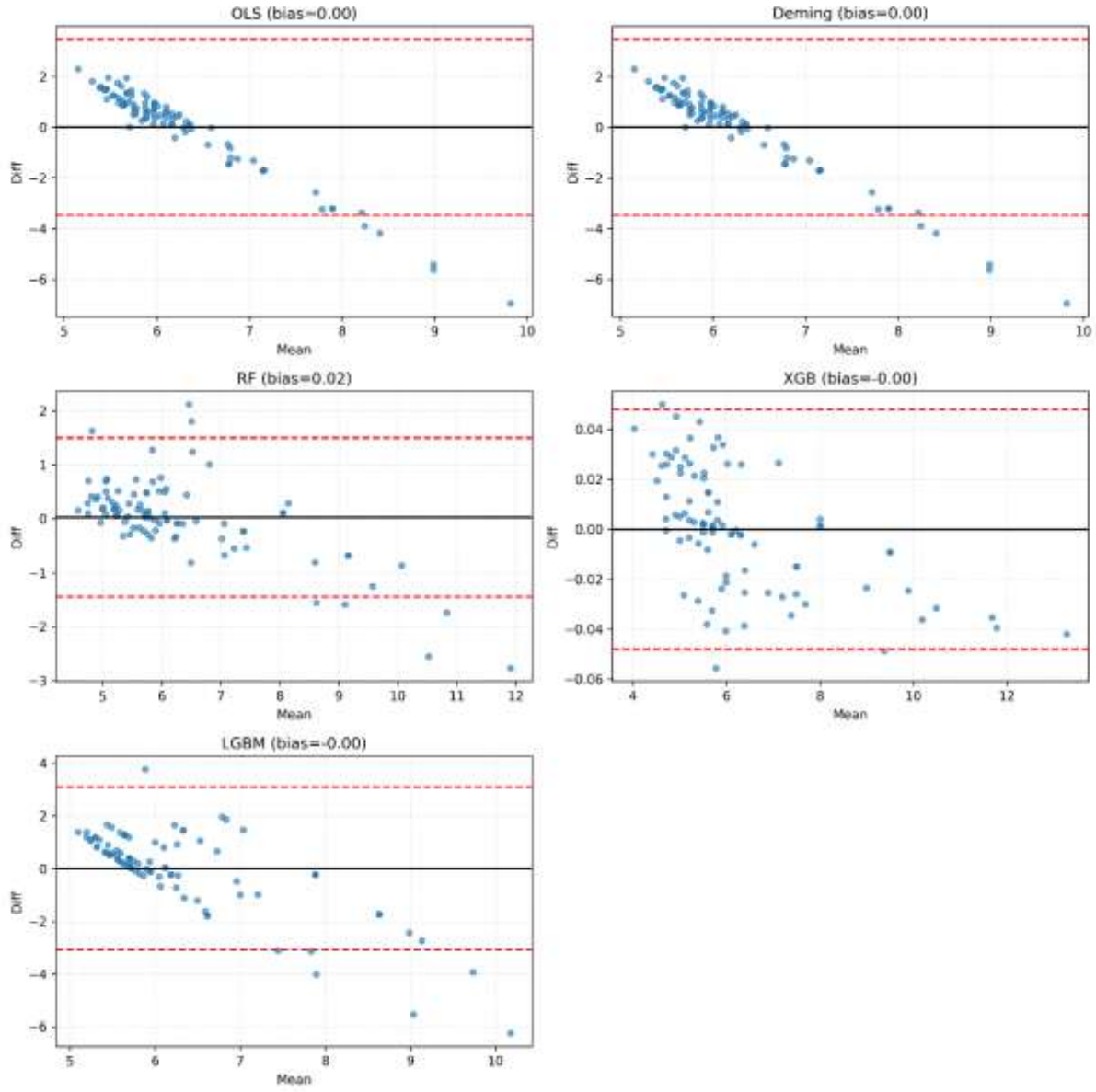

Figure 7. Bland–Altman plots comparing predicted vs. laboratory HbA1c for OLS, Deming, RF, XGB, and LGBM. Mean bias and 95% limits of agreement (LoA) are shown.

4.3.3 Concordance Correlation Coefficient

As summarized in Table 6, Lin's concordance correlation coefficient (CCC) was close to zero across models (range −0.01 to +0.09). Pearson correlations were similarly weak (−0.09 to +0.10). Together, these metrics confirm poor reproducibility with the laboratory reference.

4.4 Clinical Relevance

4.4.1 Diagnostic Threshold Evaluation

The calibration framework was evaluated at clinically relevant thresholds of 5.7% (prediabetes) and 6.5% (diabetes). At the 5.7% threshold, OLS and Deming yielded ROC AUCs near 0.53, reflecting poor discriminative power. At the 6.5% threshold, RF and LightGBM achieved slightly higher AUCs (~0.60–0.62), but still below the minimum acceptable standard for clinical decision-making.

4.4.2 ROC Curve Analysis

Receiver operating characteristic curves for OLS, Deming, RF, XGBoost, and LightGBM are shown in Figure 8. Across both thresholds, ROC curves remain close to the diagonal, indicating near-random performance. None of the models achieved AUC ≥ 0.70, the conventional threshold for diagnostic utility [32].

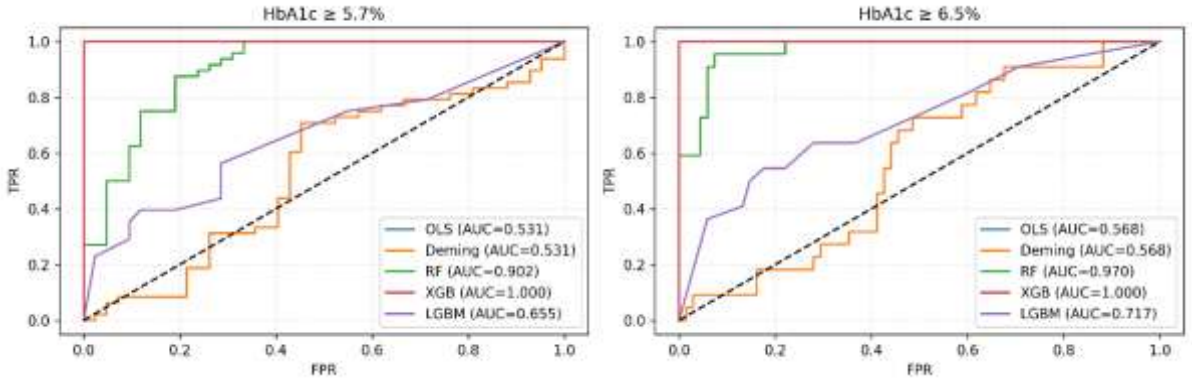

Figure 8. ROC curves at HbA1c thresholds of 5.7% and 6.5% for OLS, Deming, RF, XGB, and LGBM.

4.5 Comparative Interpretation

4.5.1 Summary of Performance

Table 6 shows that robust regression models (Huber, Passing–Bablok) achieved the lowest calibration error (MAE ≈ 1.16 pp), slightly outperforming OLS and Deming (MAE ≈ 1.29 pp). ML models did not provide improvements: LightGBM matched OLS, Random Forest performed worse, and XGBoost underperformed substantially.

4.5.2 Linear vs. Machine Learning Models

The poor performance of RF and boosting models is likely attributable to the dataset's small size (N=90) and single predictor variable, which limits the advantage of ensemble learning [21]. Similar feasibility studies of non-invasive HbA1c estimation report that ML models only outperform classical regressions when richer features (multi-spectral signals, demographics, physiology) are available [24],[29].

4.5.3 Implications for Non-Invasive HbA1c Monitoring

These results underscore the challenge of reliable HbA1c calibration using green-light reflectance alone. While methodological rigor (cross-validation, λ-sensitivity analysis, robust regression) ensured fair model comparison, predictive performance remains inadequate for clinical deployment. Nevertheless, the framework provides a reproducible baseline for future non-invasive device development [25].

4.5.4 Limitations and Future Directions

Limitations include the small cohort, absence of replicate device measurements for λ estimation, and reliance on a single optical signal. Future work should include larger, multi-center cohorts, replicate variance estimation, and incorporation of additional device features (skin tone, perfusion index, multi-wavelength reflectance). Combining such features with multimodal ML models could improve calibration robustness [22],[23].

5 CONCLUSION

This study presented a machine learning–enhanced calibration framework for non-invasive HbA1c estimation using green-light reflectance signals. A comparative evaluation of classical regression, robust statistical methods, and machine learning models

was conducted on paired laboratory and device data from 90 adult participants. Key findings of this paper: 1) Robust regression approaches, particularly Huber and Passing–Bablok, achieved the lowest calibration error (MAE ≈ 1.16 pp), slightly outperforming OLS and Deming regression. 2) Machine learning models, including Random Forest, LightGBM, and XGBoost, did not improve calibration in this single-signal, small-sample dataset. XGBoost performed worst among all tested models. 3) Agreement metrics (CCC < 0.1) and ROC/AUC analysis at diagnostic thresholds (5.7% and 6.5%) indicated that current device signals lack clinical reliability.

This work introduces a reproducible evaluation pipeline incorporating λ-sensitivity in Deming regression, nested cross-validation, bootstrap confidence intervals, and agreement analyses (Bland–Altman, CCC). The framework enables transparent comparison of regression and ML methods in biomedical calibration tasks and may serve as a baseline for future device evaluations. While non-invasive HbA1c estimation via green-light reflectance alone is not clinically viable at present, the findings highlight the necessity of multimodal sensing and larger cohort validation. The comparative evidence further suggests that simple, robust regressions can outperform ML in data-limited settings. Next-generation systems should integrate multi-wavelength optical signals, physiological covariates (e.g., skin tone, perfusion index), and larger, diverse cohorts. Combining such features with advanced ML models may improve calibration accuracy and clinical acceptability.

In summary, although predictive performance remains limited, this study establishes a rigorous calibration framework and provides critical insights into the role of machine learning and robust regression for advancing non-invasive diabetes monitoring technologies.

NOMENCLATURE / ABBREVIATIONS

ADA – American Diabetes Association
AUC – Area Under the Curve
a.u. – Arbitrary Units
CCC – Concordance Correlation Coefficient
CI – Confidence Interval
CV – Cross-Validation
HbA1c – Hemoglobin A1c (glycated hemoglobin)
HPLC – High-Performance Liquid Chromatography
IQR – Interquartile Range
LGBM – Light Gradient Boosting Machine
LoA – Limits of Agreement
MAE – Mean Absolute Error
ML – Machine Learning
NGSP – National Glycohemoglobin Standardization Program
OLS – Ordinary Least Squares Regression
PB – Passing–Bablok Regression
PPG – Photoplethysmography
RF – Random Forest
RMSE – Root Mean Squared Error
ROC – Receiver Operating Characteristic
SD – Standard Deviation
XGB – Extreme Gradient Boosting (XGBoost)

ACKNOWLEDGEMENTS

The authors acknowledge the technical support provided by the MNIT Jaipur & BIT Mesra.

AUTHOR CONTRIBUTIONS

Gaurav Jain: Conceptualization, Investigation, Methodology.

Surendra Kumar Saini: Laboratory Analysis, Data Curation.

Amit Mahesh Joshi: Formal Analysis, Supervision, Writing—Review & Editing.

Ravi Kumar Maddila: Formal Analysis, Supervision, Writing—Review & Editing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

The study was approved by the Institutional Ethics Committee, Apex Hospitals Pvt. Ltd., Jaipur, India, on 25 January 2025. The committee granted approval to conduct the study titled "Non-Invasive Detection of Health care Parameters" after review of the study synopsis and informed consent form. Written informed consent was obtained from all participants prior to their participation in the study.

DATA AVAILABILITY STATEMENT

Data supporting the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] International Diabetes Federation. *IDF Diabetes Atlas*, 10th ed. Brussels, Belgium: International Diabetes Federation; 2021. Available from: https://diabetesatlas.org/

[2] Nathan DM, Singer DE, Hurxthal K, Goodson JD. The clinical information value of the glycosylated hemoglobin assay. *N Engl J Med.* 1984;310(6):341-6. DOI: 10.1056/NEJM198402093100602. https://pubmed.ncbi.nlm.nih.gov/6690962/

[3] American Diabetes Association. Standards of medical care in diabetes—2023. *Diabetes Care*. 2023;46(Suppl 1):S1-S291. DOI: 10.2337/dc23-S001. https://diabetesjournals.org/care/article/46/Supplement_1/S10/148045/1-Improving-Care-and-Promoting-Health-in

[4] World Health Organization. *Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus: abbreviated report of a WHO consultation*. Geneva: World Health Organization; 2011. PMID: 26158184. https://pubmed.ncbi.nlm.nih.gov/26158184/

[5] Sacks DB, Bruns DE, Goldstein DE, et al. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. *Clin Chem*. 2002;48(3):436-72. DOI: 10.1093/clinchem/48.3.436. https://academic.oup.com/clinchem/article-abstract/48/3/436/5641618?redirectedFrom=fulltext&login=false

[6] Lenters-Westra E, Slingerland RJ. Six of eight hemoglobin A1c point-of-care instruments do not meet the general accepted analytical performance criteria. *Clin Chem*. 2010;56(1):44-52. DOI: 10.1373/clinchem.2009.130641. https://academic.oup.com/clinchem/article-abstract/56/1/44/5622576?redirectedFrom=fulltext&login=false

[7] Hina A, Saadeh W. Noninvasive blood glucose monitoring systems using near-infrared technology—a review. *Sensors*. 2022;22(13):4855. DOI: 10.3390/s22134855. https://www.mdpi.com/1424-8220/22/13/4855

[8] Scholtes-Timmerman MJ, Bijlsma S, Fokkert MJ, et al. Raman spectroscopy as a promising tool for noninvasive point-of-care glucose monitoring. *J Diabetes Sci Technol.* 2014;8(5):974-9. DOI: 10.1177/1932296814543104. https://pmc.ncbi.nlm.nih.gov/articles/PMC4455378/

[9] Shao J, Lin M, Li Y, et al. In vivo blood glucose quantification using Raman spectroscopy. *PLoS One*. 2012;7(10):e48127. DOI: 10.1371/journal.pone.0048127. https://pmc.ncbi.nlm.nih.gov/articles/PMC3485012/

[10] Kasahara R, Kino S, Soyama S, Matsuura Y. Noninvasive glucose monitoring using mid-infrared absorption spectroscopy based on a few wavenumbers. *Biomed Opt Express*. 2017;9(1):289-302. DOI: 10.1364/BOE.9.000289. https://pmc.ncbi.nlm.nih.gov/articles/PMC5772583/

[11] Alghlayini S, Hosni A, Atef M. Photoplethysmography based blood glucose estimation using convolutional neural networks. *2023 Advances in Science and Engineering Technology International Conferences (ASET)*. 2023:1-5. DOI: 10.1109/ASET56582.2023.10180655. https://ieeexplore.ieee.org/document/10180655

[12] Perezcampos Mayoral C, Gutiérrez Gutiérrez J, et al. Fiber optic sensors for vital signs monitoring: a review of its practicality in the health field. *Biosensors (Basel)*. 2021;11(2):58. DOI: 10.3390/bios11020058

[13] Allen J. Photoplethysmography and its application in clinical physiological measurement. *Physiol Meas*. 2007;28(3):R1-39. DOI: 10.1088/0967-3334/28/3/R01. https://pubmed.ncbi.nlm.nih.gov/17322588/

[14] Mansour M, Darweesh MS, Soltan A. Wearable devices for glucose monitoring: a review of state-of-the-art technologies and emerging trends. *Alexandria Engineering Journal*. 2024;89:224-243. DOI: 10.1016/j.aej.2024.01.021

[15] Kaur B, Kumar S, Nedoma J, et al. Advancements in optical biosensing techniques: from fundamentals to future prospects. *APL Photonics*. 2024;9(9):091102. DOI: 10.1063/5.0216621

[16] Tang L, Chang SJ, Chen CJ, Liu JT. Non-invasive blood glucose monitoring technology: a review. *Sensors (Basel)*. 2020;20(23):6925. DOI: 10.3390/s20236925
[17] Shokrekhodaei M, Cistola DP, Roberts RC, Quinones S. Non-invasive glucose monitoring using optical sensor and machine learning techniques for diabetes applications. *IEEE Access*. 2021;9:73029-73045. DOI: 10.1109/ACCESS.2021.3079182
[18] Rocke DM, Lorenzato S. A two-component model for measurement error in analytical chemistry. *Technometrics*. 1995;37(2):176-184. DOI: 10.1080/00401706.1995.10484302
[19] Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. *J Clin Chem Clin Biochem*. 1983;21(11):709-20. DOI: 10.1515/cclm.1983.21.11.709
[20] Linnet K. Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. *Clin Chem*. 1998;44(5):1024-31. PMID: 9590376
[21] Breiman L. Random forests. *Machine Learning*. 2001;45:5–32. DOI: 10.1023/A:1010933404324
[22] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference*. 2016:785–794. DOI: 10.1145/2939672.2939785
[23] Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems (NIPS)*. 2017;30:3149–3157.
[24] Kwon TH, Kim KD. Machine-learning-based noninvasive in vivo estimation of HbA1c using photoplethysmography signals. *Sensors*. 2022;22(8):2963. DOI: 10.3390/s22082963
[25] Poosarla A, Vegesna RSB, Atchuta Rao KV, Pullapukuri K. Challenges and advances in non-invasive glucose monitoring: a comprehensive review for improved diabetes management. *Int J Health Technol Innov*. 2025;4(01):10–18. DOI: 10.60142/ijhti.v4i01.03
[26] Ludbrook J. Statistical techniques for comparing measurers and methods of measurement: a critical review. *Clin Exp Pharmacol Physiol*. 2002;29(7):527-536. DOI: 10.1046/j.1440-1681.2002.03686.x
[27] Maronna RA, Martin RD, Yohai VJ. *Robust Statistics: Theory and Methods (with R)*. 2nd ed. Wiley; 2019.
[28] Naresh M, Siva Nagaraju V, Kollem S, et al. Non-invasive glucose prediction and classification using NIR technology with machine learning. *Heliyon*. 2024;10(7):e28720. DOI: 10.1016/j.heliyon.2024.e28720
[29] Gusev M, Poposka L, Spasevski G, et al. Noninvasive glucose measurement using machine learning and neural network methods and correlation with heart rate variability. *Journal of Sensors*. 2020;2020:9628281. DOI: 10.1155/2020/9628281
[30] Nathan DM, Kuenen J, Borg R, et al. Translating the A1C assay into estimated average glucose values. *Diabetes Care*. 2008;31(8):1473-8. DOI: 10.2337/dc08-0545
[31] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327(8476):307-310. DOI: 10.1016/S0140-6736(86)90837-8
[32] Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. 2006;27(8):861-874. DOI: 10.1016/j.patrec.2005.10.010
[33] Linnet K. Evaluation of regression procedures for methods comparison studies. *Clin Chem*. 1993;39(3):424-32. PMID: 8448852