



Innovation in Blood Sugar and Cholesterol Measurement Tools Based on Integrated Sensors for Non-Invasive Health Monitoring

Arnita^{1*}, Hemawan Syahputra², Faridawaty Marpaung³, Marlina Setia Sinaga⁴, Agus Junaidi⁵,
Fanny Ramadhani⁶, M. Yani⁷, GiaColin Alfaro Samuel⁸, Zakiy Maulana Pulungan⁹,
Dinda Alexa Nabila Utomo¹⁰

^{1, 2, 3, 4, 6, 8, 9, 10}Department of Mathematic, Universitas Negeri Medan, Indonesia

⁵Department of Electrical Engineering, Universitas Negeri Medan, Indonesia

⁷Department of Mechanical Engineering, Universitas Muhammadiyah Sumatera Utara, Indonesia

Abstract.

Purpose: Non-invasive monitoring of blood glucose and cholesterol has become increasingly important due to the growing prevalence of diabetes and cardiovascular diseases and the limitations of conventional invasive measurement methods, which require blood sampling, cause discomfort, and increase the risk of infection. Although various non-invasive techniques have been proposed, most existing studies focus on a single biomarker or a single sensing modality, limiting their robustness and practical implementation for real-time health monitoring. This study aims to develop an Internet of Things (IoT)-based non-invasive monitoring system capable of simultaneously estimating blood glucose and cholesterol levels by integrating physiological multisensors with machine learning. The main contribution of this research is the development of a portable platform that combines Near-Infrared (NIR) sensors, photoplethysmography (PPG), and skin temperature sensing with a Random Forest regression model for simultaneous glucose and cholesterol prediction.

Methods: The research method used is the experimental research and development (R&D) method with the ADDIE model (Analysis, Design, Development, Implementation and Evaluation). The system developed uses a combination of multi-wavelength near-infrared (NIR), photoplethysmography (PPG) and skin-temperature sensors and a Random Forest Regressor to estimate blood glucose and cholesterol levels. Sensor data were pre-processed by missing value handling, outlier removal using Interquartile Range (IQR) method and Min-Max normalisation. Hyper parameters of Random Forest were optimised using GridSearchCV and model performance was evaluated using R^2 , MAE and MSE. The optimized models were subsequently integrated into the non-invasive monitoring prototype and evaluated in terms of prediction error, signal stability, response time, and user perception.

Result: In the experiment, the optimized Random Forest model achieved R^2 of 0.4515, MAE of 0.1386 and MSE of 0.0299 in glucose prediction, and R^2 of 0.4703, MAE of 0.0804 and MSE of 0.0150 in cholesterol prediction. The experimental results showed an average prediction error of ± 8 mg/dL for glucose and ± 10 mg/dL for cholesterol at system level. The standard deviation of the sensor signals was maintained at 0.03-0.05 and the average system response time was 2.8s. Moreover, the user evaluation has provided a usability score of 4.75 out of 5 which is a high level of usability of the system. The feature importance analysis of the optimized Random Forest model showed that age, nir680 and nir610 were the most important predictors for glucose estimation, while nir680, nir860 and nir610 were the most important predictors for cholesterol estimation. The results are consistent with the multisensor design of the proposed system, where PPG and NIR signals provide complementary physiological information for non-invasive estimation. The combination of physiological multisensor data with an optimized Random Forest regression model, supported by technical testing, fast system response, stable sensor signals, and high usability, looks very promising for the simultaneous estimation of glucose and cholesterol levels without touching the person and for the development of portable, real-time, IoT-enabled personalized healthcare systems.

Novelty: The novelty of this research lies in the development of a portable IoT-based device that integrates physiological multisensors and machine learning for simultaneous and non-invasive glucose and cholesterol estimation with better interpretability.

Keywords: Physiological sensors, Non-invasive, Blood sugar and cholesterol measurement, Random forest regressor

Received February 2026 / **Revised** September 2026 / **Accepted** September 2026

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



* Corresponding author.

Email addresses: arnita@unimed.ac.id (Arnita)

DOI: [10.15294/sji.v13i4.42177](https://doi.org/10.15294/sji.v13i4.42177)

INTRODUCTION

The Indonesian Health Survey 2023 data also showed that non-communicable diseases were the main causes of physical disabilities, including motion, sight and hearing disabilities. Chronic conditions like diabetes, high blood pressure and high cholesterol often trigger a chain of events that can quickly escalate into much more serious medical emergencies. Poor blood pressure, blood sugar control can lead to devastating results like stroke, kidney failure or heart disease in the long run. Abnormal cholesterol levels are also linked to the risk of serious vascular and cardiovascular complications [1]. If you have diabetes or are at high risk of developing diabetes, it is a good idea to check your diabetes and cholesterol levels. Monitoring your blood sugar helps you to manage and prevent your diabetes, and avoid serious complications. Knowing your cholesterol level can help you reduce your risk of heart disease, stroke and other blood vessel problems. Both tests are important for early detection, prevention and management of chronic diseases [2] [3]. Real time health monitoring uses several sensor based and embedded computing human condition monitoring systems. One of the research work proposed early warning system consists of monitoring system with sensors and computing devices to identify the critical conditions of the user and send the alerts online [4].

Based on the measurement method, there are two types of blood sugar and cholesterol measuring instruments, namely measuring instruments developed using invasive and non-invasive methods. Invasive methods are measuring instruments that require a blood sample from a skin prick to obtain measurement results [5]. Non-invasive methods are measurement methods that do not require taking a blood sample or a needle prick to obtain measurement results. This method uses infrared light, optical sensors, or electromagnetic waves to measure blood sugar or cholesterol levels directly through the skin or body tissue. Some people, especially those with needle phobia or some medical conditions (such as blood clotting disorders), find finger prick blood sampling painful and uncomfortable. However, invasive methods are often used because they are more accurate. If the instrument or skin area is not sterile, the small puncture can pose a risk of infection [6]. Furthermore, because invasive methods require a specific test strip for each measurement (e.g., blood glucose, cholesterol, or uric acid), they are less effective and efficient [9]. Consequently, a dependable and effective method of measuring blood sugar and cholesterol must be developed. Therefore, it is necessary to develop an effective and efficient blood sugar and cholesterol measurement tool that does not induce discomfort or risk of infection to the user. In this research, the proposed method is a non-invasive one.

Scientists are still working on and testing non-invasive ways to check blood sugar and cholesterol levels. Traditional methods are not suitable for prediction of medical conditions such as diabetes because of high non-linear and complex nature of medical data. In one study, using the Random Forest algorithm with data-balancing techniques to improve the interpretability of the models, better predictions and more confidence in medical decisions were obtained [7]. The end result is to allow checking of cholesterol and blood sugar levels without drawing blood. "This could be easier and more convenient for patients." There are several strategies and studies in this area, but non-invasive methods cannot totally replace invasive assessments yet. Reference [8] investigated the prediction of glucose and cholesterol levels using statistical modeling and machine learning based on historical data. This is one of many papers on the subject of how to make a blood sugar and cholesterol testing device using a cutting mechanism.

BSA-coated gold nanoclusters were used in study [9] for the investigation of non-invasive optical measurement technique for the determination of glucose and cholesterol excluding interactions. Previous studies have found that currently there is no effective and universal non-invasive method for measuring glucose and cholesterol. Moreover, there is no research on the simultaneous measurement of glucose and cholesterol on a single non-invasive platform. The existing solutions do not meet the requirements of the invasive blood glucose and cholesterol measurement devices such as portability, low cost and accuracy. A non-invasive method, such as spectroscopy or skin sensors that can be integrated with wearable devices, is needed to develop a device that can measure blood glucose and cholesterol at the same time. Moreover, technologies that allow for simultaneous measurement of glucose and cholesterol with a single non-invasive device should be explored. Although the glucose and cholesterol measurements have been studied using invasive and non-invasive methods in many research works, however, the majority still focus on a single biomarker or they use only one type of sensor. This results in some limitations in the system flexibility, the user comfort and the robustness to the physiological signal change. As recent reviews and studies indicate [10], [11], optical sensor systems for real-time monitoring and non-invasive wearables still present challenges in terms of accuracy, signal stability and sensitivity to noise and physiological artifacts.

Moreover, although machine learning algorithms have been widely used in health monitoring based on physiological data, most of the studies are still focusing on classification tasks or not combining different physiological and optical parameters simultaneously on one platform [12], [13]. Thus, the research works combining the integration of physiological multi-sensors and multi-wavelength optical sensors with robust machine learning algorithms for simultaneous glucose and cholesterol estimation are still relatively few [10], [14]

In view of these research gaps, this research work aims to design a non-invasive measurement system for glucose and cholesterol levels. The system comprises physiological multisensors, including multi-wavelength near-infrared (NIR) sensors, photoplethysmography (PPG), and skin temperature sensors, and employs a random forest regression algorithm. The main contribution of this study is to develop an IoT based wearable device capable of measuring glucose and cholesterol levels simultaneously without blood sampling. Moreover, the system performance is evaluated in terms of the prediction accuracy, the signal stability and the response time. The novelty of this work is that we combine optical and physiological multisensors in a single non-invasive system and propose a machine learning approach that can capture nonlinear relations between parameters, thus resulting in more accurate and stable estimates that can be used in real-time IoT-based non-invasive health monitoring [12], [14].

METHODS

The study is conducted using the experimental and technological engineering approach of research and development (R&D). To ensure a systematic and focused innovation process we use the ADDIE (Analysis, Design, Development, Implementation and Evaluation) approach in the development process. To support intelligent health monitoring systems, this method is used in the development and manufacture of non-invasive health assessment instruments that combine multiple physiological sensors with machine learning and Internet of Things (IoT) technologies. The most important data used to create the predictive model came from direct measurements of respondents' blood sugar and cholesterol levels, as well as measurements using an infrared thermometer, a near-infrared sensor, and the MAX30100 sensor. The metrics listed in Table 1 below will result from the sensor measurement data:

Table 1. Sensor measurement result parameters.

Parameter Name	Sensor Source	Data Type / Unit	Meaning and Physiological Function
Temperature (°C)	Infrared Thermometer (IR)	Degrees Celsius (°C)	Measures the surface temperature of the skin, reflecting peripheral blood perfusion and metabolic activity. Temperature may decrease in high glucose or cholesterol levels due to decreased micro blood flow.
BPM (Beats Per Minute)	MAX30100 (PPG sensor)	Beats Per Minute	Shows heart rate frequency. Obtained from the PPG (AC component) signal. Used to assess cardiovascular health and control blood circulation.
SpO ₂ (Oxygen Saturation)	MAX30100 (PPG sensor)	Percentage (%)	The percentage of hemoglobin oxygenation in the blood is based on the red/IR PPG ratio. May affect the interpretation of tissue perfusion and energy metabolism.
ppg_red	MAX30100 (red LED, 660 nm)	Intensity value (ADC count)	Intensity value (ADC The intensity of red light reflection from tissue. Related to the level of reduced hemoglobin (deoxyhemoglobin). Used to calculate SpO ₂ and assess changes in blood volume. count)
ppg_ir	MAX30100 (infrared LED, 880 nm)	Intensity value (ADC count)	The intensity of IR light reflection from tissue. Indicates variations in blood flow in capillaries. The primary data for calculating BPM and perfusion variations (AC/DC).
nir610	Sensor NIR (Near Infrared, $\lambda = 610$ nm)	Reflectance intensity	Light reflectance at a wavelength of 610 nm, sensitive to tissue pigments and water. Can be used for basic tissue spectrum analysis.
nir680	Sensor NIR ($\lambda = 680$ nm)	Reflectance intensity	Wavelength 680 nm — close to the peak absorption of hemoglobin, helps correct for the influence of blood on the spectral signal.
nir730	Sensor NIR ($\lambda = 730$ nm)	Reflectance intensity	Captures variations in absorption by water and fat tissues; begins to be sensitive to lipid components.
nir760	Sensor NIR ($\lambda = 760$ nm)	Reflectance intensity	Wavelength is important for the absorption of oxygenated hemoglobin (oxy-Hb). It can help assess local tissue oxygenation.
nir810	Sensor NIR ($\lambda = 810$ nm)	Reflectance intensity	The 810 nm spectrum is often used to detect glucose because it is sensitive to the vibrations of the –OH group in glucose.
nir860	Sensor NIR ($\lambda = 860$ nm)	Reflectance intensity	This region shows a mild uptake peak by cholesterol and triglycerides; suitable for cholesterol prediction features.
Manual Glucose	Laboratory reference measuring instrument / glucometer	mg/dL	Blood glucose levels obtained invasively (manually) are used as ground truth for training and validating the prediction model.
Manual Cholesterol	Laboratory reference measuring tool / cholesterol meter	mg/dL	The total cholesterol level value measured manually/invasively becomes the reference label (target) in the machine learning model..

Development stage

The stages of research and development of measuring instruments for measuring blood sugar and cholesterol levels consist of:

Analysis

This stage aims to identify system requirements and specifications. The analysis is conducted by reviewing the limitations of conventional measuring instruments, which are still invasive and require blood draws, causing discomfort for users. Non-invasive technology based on Near-Infrared (NIR) has been widely developed due to its ability to detect glucose levels without blood draws [15]. This stage resulted in system requirements specifications and a preliminary design of the device architecture.

Design

This stage focuses on creating the conceptual and technical design of the device, including the development of a system block diagram consisting of a sensor module, ESP32 microcontroller, IoT module, and user interface. The sensor circuit design is designed to obtain stable physiological signals, while machine learning algorithms such as Random Forest are prepared to predict glucose and cholesterol levels from sensor data. This approach aligns with research findings showing that the combination of machine learning methods and optical sensors can improve the accuracy of non-invasive measurements [16]. This stage produces a system blueprint and hardware and software designs.

Development

In the development phase, hardware was assembled using an NIR sensor (610–860 nm), a MAX30100 sensor (PPG and SpO₂), and an IR thermometer for skin temperature measurement. Software (firmware) was developed on a microcontroller for sensor data acquisition and transmission to an IoT server. Machine learning models were applied to predict glucose and cholesterol levels based on physiological features, with the device calibration process carried out using manual data from standard medical devices. The result of this phase was a non-invasive device prototype ready for testing.

Implementation

The implementation phase involves testing the prototype in real-world conditions by collecting data from several respondents with varying health conditions. Sensor data is sent in real time via an IoT platform to ensure system stability. Testing is conducted by comparing the device's predictions with manual measurements [17]. This phase produces an integrated dataset and initial system performance test results. As for the justification of model selection, respondent characteristics, and potential data bias, it is explained as follows:

a. Justification for model selection

The Random Forest Regressor algorithm was chosen in this study because of its ability to model nonlinear relationships and complex interactions between physiological parameters generated by non-invasive sensors, which are generally difficult to represent using a simple linear approach [17], [16]. Data from photoplethysmography (PPG), multi-wavelength near-infrared (NIR), and skin temperature sensors are heterogeneous, correlated, and susceptible to noise and physiological artifacts, requiring the use of a robust model that accounts for data variability [18], [19]. Random Forest uses a bootstrap aggregation (bagging) mechanism and random feature selection, both of which have been proven effective in increasing prediction stability and reducing the risk of overfitting in sensor-based health data [16], [19]. In addition, this algorithm provides feature importance analysis, allowing identification of the contribution of each sensor parameter to the prediction results, thus supporting model interpretability and the physiological relevance of the estimation results [17], [7].

b. Respondent characteristics and subject

Condition Variation. This study involved adult respondents aged 20–55 years and physiological conditions, including subjects with glucose and cholesterol levels within and above normal limits. This variation in subject conditions was intended to ensure that the prediction model was not trained solely on a single physiological group but could capture general patterns that represent the population of potential users of non-invasive health monitoring systems. Measurements were conducted under resting conditions and in a relatively controlled environment to minimise the influence of extreme physical activity, stress, and changes in environmental temperature on PPG and NIR signals.

c. Potential data bias and mitigation strategies

Potential data bias in this study may arise from the limited number of participants, uneven distribution of glucose and cholesterol levels, and differences in individual characteristics such as gender, skin temperature, and hydration status. These factors are known to affect the quality of non-invasive

physiological signals and the performance of predictive models [17], [16]. To minimise this potential bias, several steps were implemented, including normalising sensor data and removing outliers. These procedures have been shown to improve the stability and consistency of models using sensor-based health data [18]. After preprocessing, 129 valid observations were retained. The dataset was randomly divided into training (80%, $n = 103$) and testing (20%, $n = 26$) subsets to provide an objective evaluation of model performance [19]. Furthermore, the Random Forest algorithm employs bootstrap sampling, which reduces the influence of individual observations and enhances model generalisation, particularly when dealing with unevenly distributed datasets [7].

Evaluation

The evaluation phase aims to assess the system's accuracy, stability, and feasibility. The performance of the machine learning model was evaluated using R^2 , MAE and MSE metrics [16]. In addition, the tool's validity and reliability were verified by comparing sensor measurements with standard medical devices and by user evaluations of usability, comfort, and the user interface. This evaluation method aligns with the ADDIE-based R&D framework, which is often used in research on the development of health technologies [20].

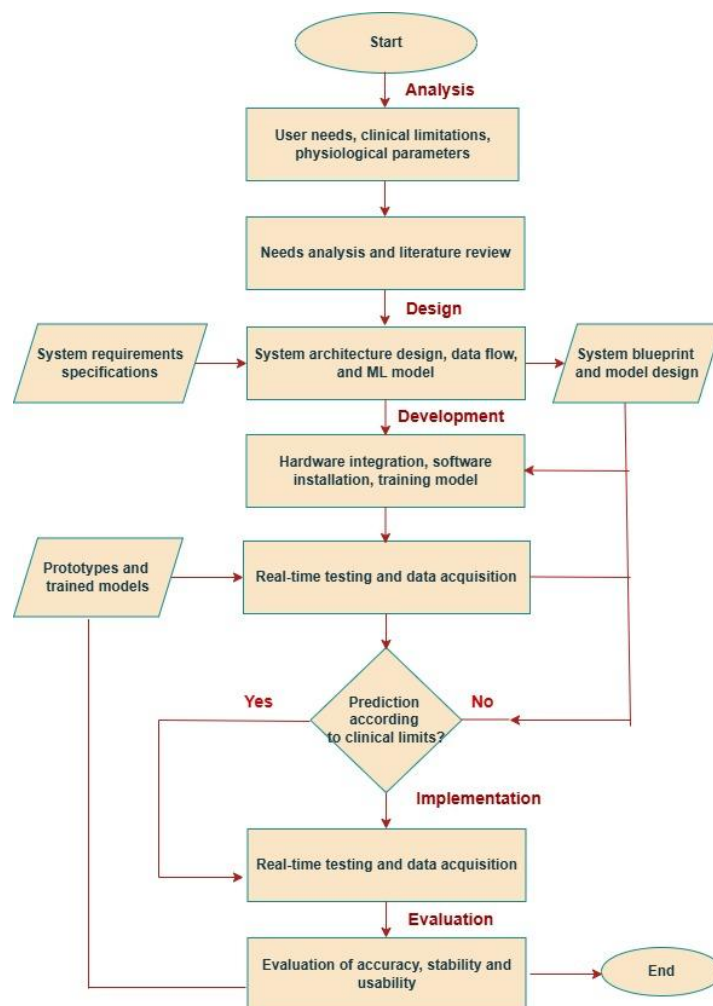


Figure 1. Flowchart of ADDIE development stages

Stages of predicting blood sugar and cholesterol levels using the random forest algorithm.

Initialization and formation of training data

A sensor dataset containing physiological parameters such as skin temperature, BPM, SpO₂, ppg_red, ppg_ir, and NIR intensity (610–860 nm) is used as input variables (X), while Glucose_manual and Cholesterol_manual are used as target variables (y). The data is then divided into two parts, namely training data and testing data, with a ratio of 80%:20%. Next, the number of trees to be built in the model

(*n_estimators*) is determined, for example, 100 or 500 trees. This process is the initial stage in the formation of a Random Forest model for predictions based on physiological data [21].

Bootstrapping (random sampling)

From the training data, the algorithm randomly generates several subsets of the data with replacement. Each subset of the data is used to train a unique decision tree. The unsampled samples are called Out-of-Bag (OOB) and are used for internal validation of the model [22]. According to [23], When making splits, each tree in the forest only takes into account a random subset of features after being trained on a random sample of the data (bootstrap sampling). This strategy helps the model become more stable and resistant to overfitting.

Random feature selection

At this stage, only a subset of the total features is randomly selected to be used in forming nodes in each tree. For example, if the total features are 10, the algorithm only selects approximately $\sqrt{10} \approx 3$ random features for each node. This approach reduces correlation between trees and improves the model's generalization ability [24]. Similarly, [25] explains that by generating an uncorrelated forest of decision trees, feature randomization enhances the robustness and generalization of the model.

Decision tree formation

Each decision tree is constructed using a randomly selected subset of data and features. The process of forming nodes involves selecting the best features and split points according to specific rules. Mean Squared Error (MSE) is used for regression, and Gini Impurity or Entropy [26] is used for classification. The process continues until certain conditions are met, such as *max_depth* or *min_samples_split*. This stage makes a group of decision trees that don't depend on each other.

Forest formation

The model builds a "forest," a group of independent decision trees, after all the trees are built. The final prediction is based on each tree's calculations. According to Analytics [27], for regression tasks, Random Forest averages the predictions from individual trees to obtain the final output.

Forecasting (averaging or voting)

The following equation is used to average all of the predictions for each tree in order to determine the final result for glucose and cholesterol level forecasts:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (1)$$

where *N* is the number of trees in the forest, and *y_i* is the predicted result of the *i*-th tree [21].

Model evaluation

The model was evaluated by comparing the predicted results to actual (manual) data using metrics such as Mean Absolute Error (MAE), Mean Square Error (MSE), and *R*². A low error value indicates a good level of accuracy in non-invasively estimating glucose and cholesterol levels [26].

Interpretation and feature importance

After the model is created, a feature significance analysis is performed to determine the sensor features that most influence the prediction results. According to [27], "Random Forest provides a built-in method to evaluate variable importance by measuring the mean decrease in impurity or accuracy." You can see the results as a bar chart or SHAP summary plot to make it easier to understand what each sensor does [28].

Validation & generalization

The last step is to test the model on a test set to ensure it isn't overfitting and that it generalizes well to new data. This validation is also important for assessing how reliable the system is when used in real-world IoT-based measuring devices that don't harm people.

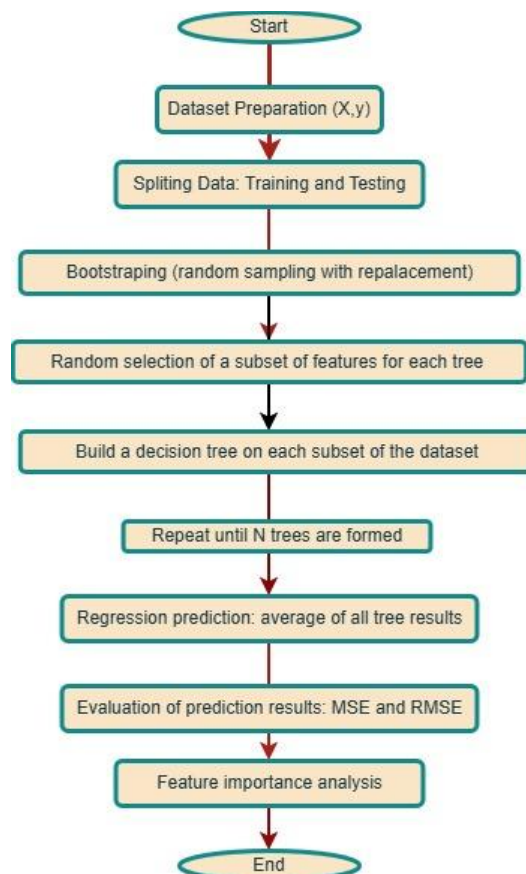


Figure 2. Flowchart of the random forest regressor algorithm

RESULT AND DISCUSSION

Analysis stage

The analysis phase was conducted to determine whether a way to measure things without harming them was needed. The literature review and observations showed that the primary issue was the community's dependence on invasive measurement tools necessitating blood sampling, rendering them inefficient and uncomfortable. A solution was needed in the form of a needle-free (non-invasive) glucose and cholesterol measurement tool capable of providing rapid, highly accurate estimates.

Whatever tools and materials are needed to develop a smart device for measuring blood sugar and cholesterol levels using this non-invasive method, they consist of:

Table 2. Tools and materials needed to develop a non-invasive blood sugar and cholesterol measuring tool

No	Tools and materials	Usability
1	Hardware	
a	Microcontroller (ESP32)	Used to control the system, process sensor data, and execute the prediction model.
b	NIR (Near-Infrared) Optical Sensor	Used to acquire near-infrared optical signals associated with physiological parameters for non-invasive glucose and cholesterol estimation.
c	Pulse Oximeter Sensor (MAX30100)	Used to measure heart rate and oxygen saturation and to provide additional physiological parameters.
d	PIR Motion Sensor (HC-SR501)	Used to detect motion and help identify movement that may affect sensor measurements.
e	IR Thermometer (MLX90614)	Used to measure body or skin temperature as an additional physiological parameter.
f	OLED Display	Used to display sensor readings and predicted glucose and cholesterol levels.
g	Button/Pushbutton	Used to control user input and operate or initiate measurement functions.
h	Power Supply/Battery Pack	Used to provide electrical power to the device.
i	Case/Enclosure (3D Printed)	Used to protect the electronic components and provide a practical and ergonomic device enclosure.
j	Bluetooth/Wi-Fi Communication	Used to transmit data between the device, smartphone, or cloud platform.
k	Laptop/PC	Used to program the microcontroller, process data, and develop the machine-learning model.

2	Software used	Usability
a	Arduino IDE / Python	To program the microcontroller
b	Jupyter Notebook	Used as an interactive development environment for data preprocessing, exploratory data analysis, model development, training, and evaluation
c	TensorFlow / Scikit-Learn	Machine learning libraries used to create predictive models
d	Excel / Google Sheets	To perform initial data analysis
3	Materials used:	Usability
a	Blood Sugar and Cholesterol Sample Data (Dataset)	Data collected from patients or taken from open-source sources for training machine learning models.
b	Jumper Cables and Breadboard	To assemble hardware components.
c	Resistors, Capacitors, and Other Electronic Components	To support the hardware circuit.
d	3D Printing Filament (PLA/ABS)	Materials for making device casing.
e	Adhesive Tape and Soldering Kit	For assembly and connection of components.
f	Wearable Strap	To make it easier for users to use the device.

Design stage

In this stage, the smart device system architecture is designed, consisting of a sensor module integrated with a microcontroller (ESP32) that sends data in real time to an IoT-based server. The user interface design is developed to display estimated glucose and cholesterol levels via a web-based or mobile application. The design of the system flow includes:

- Getting data from sensors like PPG, NIR, and temperature.
- Storing data in a cloud database.
- Processing data with a machine learning pipeline that includes preprocessing, training, and testing.
- Showing the results of the predictions on the screen.

The drawing for the tool system flow design is:

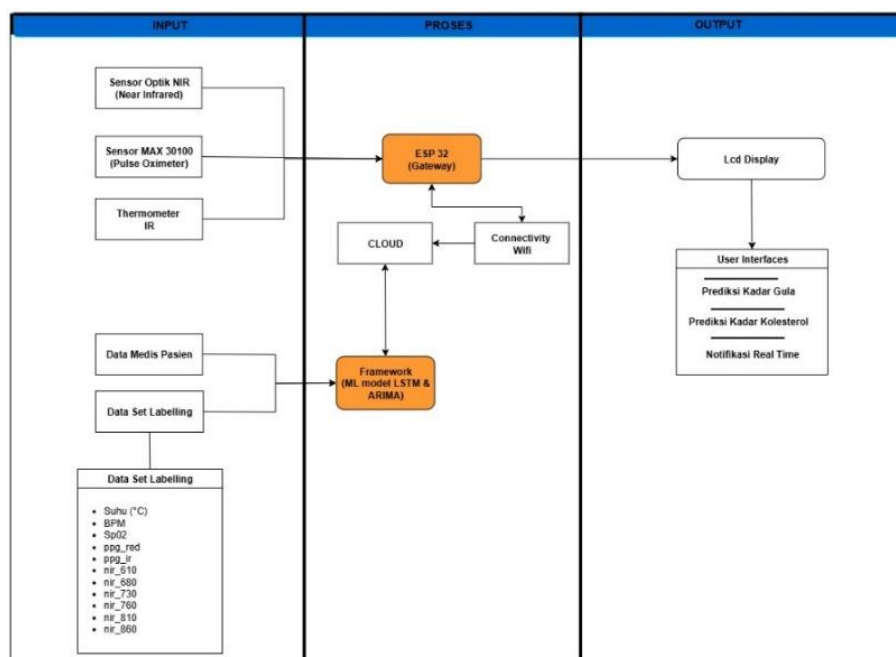


Figure 3. Block diagram of the architecture of smart devices

Development phase

In the development phase a prototype of the proposed non-invasive measurement device was built and the required physiological sensors were incorporated and tested. Data were collected from 133 adult participants, age range 20–55 years, with varying physiological profiles including blood glucose and cholesterol levels from normal to high. All participants underwent physiological measurements that consisted of the acquisition of multi-sensor data for the estimation of blood glucose and cholesterol levels. The tests were conducted in a controlled indoor environment with the participants at rest, to reduce the physiological variability unrelated to the metabolic conditions. Reference values for blood glucose and cholesterol were obtained using certified invasive medical devices and served as ground-truth target variables for the development of a supervised machine learning model.

The predictor variables were collected with a multi-sensor platform composed of a multi-wavelength Near-Infrared (NIR) sensor, a MAX30100 photoplethysmography (PPG) sensor, and an infrared (IR) thermometer. The physiological parameters recorded were temperature (°C), heart rate (BPM), oxygen saturation (SpO₂), PPG_Red, PPG_IR and NIR signals at 610, 680, 730, 760, 810 and 860 nm. The full list of physiological parameters is shown in Table 1. For building the supervised regression dataset, a corresponding reference measurement was obtained for each sensor measurement.

Before developing the model, the dataset was subjected to a series of preprocessing steps to ensure data quality and consistency. The steps were (1) synchronizing the multisensor measurements with the reference measurements, (2) handling missing values, (3) detecting and removing outliers using the Interquartile Range (IQR) method, (4) Min–Max normalizing the predictor variables, and (5) chronologically splitting the dataset into training (80%) and testing (20%) subsets. Summary of preprocessing results is listed in Table 2.

Table 2. Summary of data preprocessing

Preprocessing Step	Result
Initial observations	133
Missing values	Glucose (4), Age (4)
Missing-value treatment	[isi sesuai perlakuan sebenarnya]
Outlier detection	IQR method
Final observations	129
Feature normalization	Min–Max Scaling [0, 1]
Data splitting	80% training / 20% testing

The data set contained 133 observations and four missing values were identified in the Glucose and Age variables. Exploratory data analysis also indicated the presence of extreme observations in several physiological parameters, especially in the PPG signals, NIR reflectance measurements, glucose and age. The IQR method was used to identify and remove extreme observations. After cleaning the data and removing outliers, a total of 129 valid observations were retained for the next model development. Then, the predictor variables were normalized to the [0, 1] range using Min–Max normalization. The preprocessing approach minimized the effect of extreme observations while maintaining the overall distributional properties of physiological characteristics, as illustrated in Figure 4. The resulting dataset was then used to train the Random Forest model and to perform hyperparameter optimization using GridSearchCV. The testing subset was reserved for the final evaluation of the model.

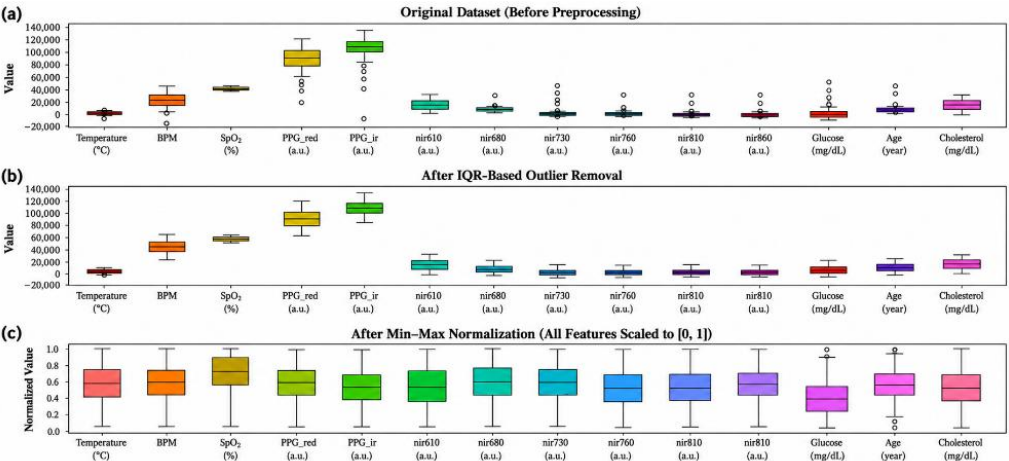


Figure 4. Distribution of physiological sensor features throughout the preprocessing stages: (a) original dataset before preprocessing, showing the initial distribution and extreme observations; (b) dataset after outlier removal using the Interquartile Range (IQR) method, showing a reduction in extreme values; and (c) physiological features after Min–Max normalization, with all variables scaled to the [0, 1] range prior to Random Forest model training.

Figure 4 shows the preprocessing results of the physiological dataset. The original data set (Figure. 4(a)) had some outliers in the variables PPG, NIR, glucose and age. The IQR based outlier removal method was able to effectively reduce these outliers (Figure. 4(b)) maintaining the overall distribution of the data. Then

all the predictor variables are normalized by Min-Max normalization to bring the features to a common range of [0, 1] as shown in Fig. 6(c) The resulting normalized data set allows for a more homogenous feature space for training the Random Forest model and improves numerical stability during prediction.

GridSearchCV is used to tune the hyperparameters of the Random Forest regression model for the optimal combination of parameters for blood glucose and cholesterol level prediction based on multi-sensor physiological data. In the optimization process, different combinations of model parameters were tested, such as the number of choice trees, maximum feature selection strategy, minimum samples to split a node and minimum samples at leaf nodes. The best configuration of the model was 150 decision tree models ($n_estimators = 150$), $max_features = sqrt$, $min_samples_leaf = 2$ and $min_samples_split = 10$, which yielded the best predictive performance among all the tested combinations of parameters.

Table 3. Performance of the optimized random forest model

Target Variable	R ²	MAE	MSE
Glucose	0.4515	0.1386	0.0299
Cholesterol	0.4703	0.0804	0.015

The predictive performance of the optimized Random Forest model is summarized in Table 3. The proposed blood glucose prediction model achieved a coefficient of determination (R²) of 0.4515, a Mean Absolute Error (MAE) of 0.1386, and a Mean Squared Error (MSE) of 0.0299. The cholesterol prediction model achieved R², MAE and MSE of 0.4703, 0.0804 and 0.0150 respectively. The results imply that the optimized Random Forest model was able to explain around 45–47 % of the variability of the glucose and cholesterol values with relatively low prediction errors. In conclusion, the presented performance shows the possibility to combine physiological multisensor data with Random Forest regression to estimate simultaneously blood glucose and cholesterol levels in non-invasive way.

Following the integration and assembly of the sensing components, the developed prototype was fully assembled and prepared for experimental testing. Figure 5 shows the resulting prototype and the integration of the multisensor components into the proposed non-invasive measurement system.

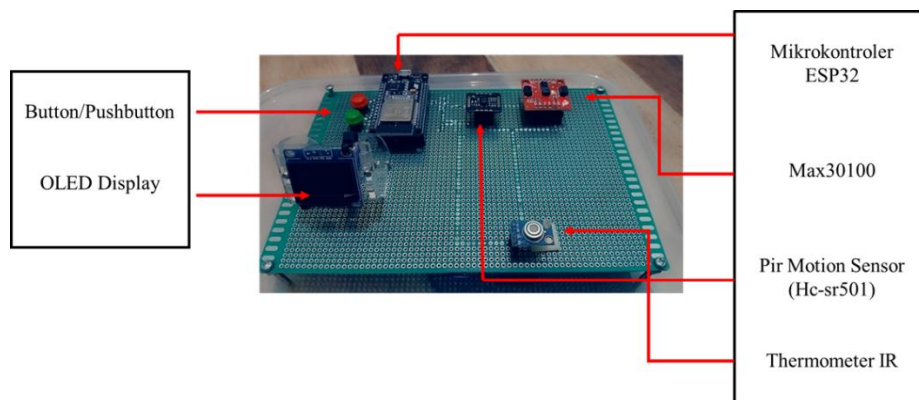


Figure 5. Result of putting together smart devices

A prototype of a health monitoring system using ESP32 microcontroller and several physiological sensors is shown in Figure 5. The circuit uses a MAX30100 sensor to detect heart rate and blood oxygen levels, an HC-SR501 PIR sensor for motion detection and an infrared thermometer to measure skin temperature without contact. The ESP32 processes the data of the measurements and displays them on an OLED screen. Push buttons turn on or control system functions. All the parts are installed on a test circuit board which acts as a portable device for real-time health monitoring.



Figure 5. Smart device prototype

The preceding image shows the first assembly of the final prototype of a non-invasive health measurement device and figure 5 shows the final prototype. In this picture, all components of the ESP32 microcontroller, MAX30100 sensor, HC-SR501 PIR sensor, infrared thermometer, OLED display and pushbutton are permanently installed in a sturdy and neat protective case. This image shows a complete integrated system for the real world. The previous image though, is still an open circuit on a perfboard for testing and wiring. The transition from open circuit to a closed device is a movement from development to implementation in the project. The new design is safer, more comfortable, and easier to carry.

Implementation stage

After the development of the prototype and the optimization of the prediction models using the Random Forest, the developed system was implemented as an integrated non-invasive measurement device. The trained models were integrated into the system workflow to enable processing of physiological sensor data and converting it to estimated blood glucose and cholesterol levels. The user presses the pushbutton to start the measurement, and the multisensor module measures PPG, NIR, oxygen saturation, heart rate and skin-temperature signals. The ESP32 acts as the main processor and controller, collecting sensor data and passing it onto the prediction pipeline. The acquired sensor data are preprocessed by the same preprocessing procedures defined during model development, including feature scaling using predefined Min–Max parameters. Then, the processed features are fed into the respective Random Forest regression models for glucose and cholesterol estimation. The prediction results are then shown on the OLED interface and transmitted through the wireless communication module for IoT-based monitoring, if applicable. Thus, the workflow of implementation combines the acquisition of physiological signals, preprocessing of data, machine-learning prediction and visualization of results into a single measurement process. The developed system was designed to work without blood sampling and to provide the estimated glucose and cholesterol values in a short time of processing. The complete operational workflow of the implemented device was then tested for technical performance and evaluated for user perception as described in the Evaluation Stage.

Evaluation stage

The developed system was evaluated from two complementary points of view: technical performance and user perception. Technical evaluation was done on prediction accuracy, sensor signal stability and system response time. User evaluation was on ease of use, comfort and practicality. The prototype was developed and assembled and then tested for technical performance to determine the system's ability to obtain physiological signals and provide glucose and cholesterol estimations within an acceptable processing time reliably. The results of the technical performance evaluation are presented in Table 4.

Table 4. Results of system technical aspect testing

No.	Technical Indicator	Test Parameter	Average Result	Description
1	Prediction Accuracy	Mean glucose error	± 8 mg/dL	Within predefined tolerance
		Mean cholesterol error	± 10 mg/dL	Within predefined tolerance
2	Sensor Signal Stability	SD of PPG/NIR signals	0.03–0.05	Stable signal acquisition
3	System Response Time	Mean data processing time	2.8 s	Fast system response

The technical evaluation showed that the developed system achieved an average glucose measurement error of ± 8 mg/dL and a cholesterol measurement error of ± 10 mg/dL. According to [29], a glucose monitoring system can be considered clinically acceptable when most measurements fall within ± 15 mg/dL of the reference value. The obtained glucose error was therefore within this reported tolerance range. Previous

studies using PPG and NIR spectroscopy combined with machine learning reported glucose prediction errors of approximately $\pm 10\text{--}12$ mg/dL [18], [30]. The present results are within a comparable range. In addition, the standard deviation of the PPG/NIR signals ranged from 0.03 to 0.05, indicating relatively stable signal acquisition during measurement. This value was lower than the range of approximately 0.07–0.09 reported by [19] before signal-processing optimization. The mean system processing time was 2.8 s, which is a relatively short response for near real-time health monitoring applications. Previous studies of wearable and AI-based health monitoring systems have reported response or inference times ranging from milliseconds to several seconds, depending on the sensing and processing architecture [31].

Apart from the technical performance, user perception was evaluated to assess the usability, comfort and practicality of the developed device. The evaluation was conducted with 30 adult participants between 20–55 years old, purposively selected as potential users of non-invasive health monitoring technology. The sample included 17 males (56.7%) and 13 females (43.3%) who were from medical and non-medical fields. A five-point Likert scale was used to assess three aspects: ease of use, comfort, and practicality.

Table 5. User perception and usability evaluation of the developed device

No	Assessment Indicators	Maximum Score	Average Score	Description
1	Ease of use	5	4,8	Very easy to use
2	Convenience	5	4,7	Comfortable and painless
3	Practicality	5	4,8	Practical and efficient
Average total		5	4,75	Excellent

The results of the user evaluation presented in Table 5 showed mean scores of 4.8 for ease of use, 4.7 for comfort, and 4.8 for practicality, with an overall mean score of 4.75 out of 5. The results show that the developed device was perceived by the participants as easy to operate, comfortable to use during measurement and practical for non-invasive health monitoring. The high overall score also suggests positive user acceptance of the developed prototype.

To further investigate the contribution of individual sensor variables to the prediction models, feature importance analysis using Random Forest Regressor was performed. The analysis showed that NIR730 and PPG_IR were the most important features to predict glucose and NIR810 and skin temperature were the most important features to predict cholesterol. These results show that the contribution of both optical signals and physiological parameters to the prediction models can reveal the relative importance of the multisensor variables in the estimation of the target biomarkers.

CONCLUSION

The objective of this study was to develop and validate a portable, IoT-based non-invasive health monitoring device for simultaneous measurement of blood glucose and cholesterol by incorporating physiological and multi-wavelength NIR sensors along with Random Forest regression. The system is developed based on the ADDIE framework which includes requirements analysis, system design, prototype development, implementation and evaluation. The developed platform integrates ESP32 microcontroller, NIR sensors, MAX30100 PPG sensor, infrared temperature sensor, OLED display and wireless communication for supporting the non-invasive physiological data acquisition and real-time estimation. The optimized Random Forest models showed the capability to estimate both biomarkers from multisensor data. For the glucose model, $R^2 = 0.4515$, $MAE = 0.1386$, and $MSE = 0.0299$, while for the cholesterol model, $R^2 = 0.4703$, $MAE = 0.0804$, and $MSE = 0.0150$. The average measurement errors at system level were ± 8 mg/dL for glucose and ± 10 mg/dL for cholesterol. These results indicate that the multisensor approach in combination with the Random Forest is able to capture meaningful nonlinear relations between physiological and optical measurements and the reference biomarker values. The technical evaluation also demonstrated a relatively stable acquisition of PPG and NIR signals, with standard deviations ranging from 0.03 to 0.05 and an average system response time of 2.8 s. The user perception assessment with 30 participants resulted in an overall mean score of 4.75 out of 5 in ease of use, comfort and practicality, showing a high level of perceived usability of the developed prototype. Feature-importance analysis also showed differences in the contribution of different sensor variables to the two prediction tasks, with NIR- and PPG-related variables contributing significantly to glucose estimation, while NIR and skin-temperature variables were important for cholesterol estimation. The results demonstrate the technical feasibility of combining physiological multisensors, multi-wavelength NIR sensing, Random Forest regression and IoT technology into a single portable platform for non-invasive simultaneous estimation of blood glucose and

cholesterol. However, the moderate R^2 values and the limited data set suggest that the system should be considered as a research prototype and not as a replacement for clinically validated invasive measurements. Future studies should include larger and more heterogeneous population, wider ranges of glucose and cholesterol concentrations, independent external validation, better calibration and signal processing methods and comparison with clinically accepted reference methods. These developments are important for establishing the reliability, generalizability, and clinical applicability of the proposed non-invasive monitoring system.

REFERENCES

- [1] Lusia Kus Anna, "PENTINGNYA RUTIN CEK KOLESTEROL, HIPERTENSI, DAN GULA DARAH," Jakarta, Jul. 23, 2024. Accessed: Jan. 16, 2025. [Online]. Available: : <https://health.kompas.com/read/24G23103100568/pentingnya-rutin-cek-kolesterol-hipertensi-dan-gula-darah>
- [2] Rokom, "POLA HIDUP SEHAT DAN DETEKSI DINI BANTU KONTROL GULA DARAH PADA PENDERITA DIABETES," <https://sehatnegeriku.kemkes.go.id/baca/umum/20211115/3438859/pola-hidup-sehat-dan-deteksi-dini-bantu-kontrol-gula-darah-pada-penderita-diabetes/>. Accessed: Jan. 16, 2025. [Online]. Available: <https://sehatnegeriku.kemkes.go.id/baca/umum/20211115/3438859/pola-hidup-sehat-dan-deteksi-dini-bantu-kontrol-gula-darah-pada-penderita-diabetes/>
- [3] E. E. Rais, I. R. Aziz, and S. Surdianah, "Pemeriksaan total kolesterol pada sampel serum darah dengan menggunakan metode fotometrik di Balai Besar Laboratorium Kesehatan Makassar," *Filogeni: Jurnal Mahasiswa Biologi*, vol. 4, no. 1, pp. 19–27, Apr. 2024, doi: 10.24252/filogeni.v4i1.42013.
- [4] S. Syamsul Hidayat, I. Mujahidin, M. Faizin Zaini, and R. Suhartina, "Optimizing Fall Detection System as an Early Warning System for the Elderly to Enhance Quality of Life," *Scientific Journal of Informatics*, vol. 11, no. 2, 2024, doi: 10.15294/sji.v11i2.3719.
- [5] A. Saputra and B. Santoso, "EVALUATING THE EFFECTIVENESS OF MINIMALLY INVASIVE TECHNIQUES IN THE MANAGEMENT OF CHOLELITHIASIS AND ITS COMPLICATIONS," *International Journal of Intelligent Healthcare Analytics*, no. 4, Jul. 2024.
- [6] L. B. Palavani *et al.*, "Invasive and Noninvasive Techniques for Intracranial Pressure Monitoring After Decompressive Craniectomy: A Systematic Review and Meta-Analysis," Oct. 01, 2024, *Elsevier Inc.* doi: 10.1016/j.wneu.2024.06.118.
- [7] H. Aulia, A. Wibowo, and S. Sutrisno, "Integration of Random Forest, ADASYN, and SHAP for Diabetes Prediction and Interpretation," *Scientific Journal of Informatics*, vol. 12, no. 2, pp. 211–222, Jun. 2025, doi: 10.15294/sji.v12i2.24314.
- [8] U. Krishnamoorthy, V. Karthika, M. K. Mathumitha, H. Panchal, V. K. S. Jatti, and A. Kumar, "Learned prediction of cholesterol and glucose using ARIMA and LSTM models – A comparison," *Results in Control and Optimization*, vol. 14, Mar. 2024, doi: 10.1016/j.rico.2023.100362.
- [9] M. R. Krishnendu, S. Bhagat, V. Jain, D. Mehta, and S. Singh, "Paper immobilized BSA-decorated gold nanoclusters for single-step optical sensing of glucose and cholesterol without cross-reactivity," *Colloids Surf. B Biointerfaces*, vol. 245, Jan. 2025, doi: 10.1016/j.colsurfb.2024.114303.
- [10] M. Mansour, M. Saeed Darweesh, and A. Soltan, "Wearable devices for glucose monitoring: A review of state-of-the-art technologies and emerging trends," Feb. 01, 2024, *Elsevier B.V.* doi: 10.1016/j.aej.2024.01.021.
- [11] W. Oñate, E. Ramos-Zurita, J. P. Pallo, S. Manzano, P. Ayala, and M. V. Garcia, "NIR-Based Electronic Platform for Glucose Monitoring for the Prevention and Control of Diabetes Mellitus," *Sensors*, vol. 24, no. 13, Jul. 2024, doi: 10.3390/s24134190.
- [12] B. Shi *et al.*, "Assessing Elevated Blood Glucose Levels Through Blood Glucose Evaluation and Monitoring Using Machine Learning and Wearable Photoplethysmography Sensors: Algorithm Development and Validation," *JMIR AI*, vol. 2, no. 1, 2023, doi: 10.2196/48340.
- [13] Z. Nie, M. Rong, and K. Li, "Blood glucose prediction based on imaging photoplethysmography in combination with Machine learning," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, doi: 10.1016/j.bspc.2022.104179.
- [14] R. A. Fraser, R. J. Walker, J. A. Campbell, O. Ekwunife, and L. E. Egede, "Integration of artificial intelligence and wearable technology in the management of diabetes and prediabetes," *NPJ Digit. Med.*, vol. 8, no. 1, Dec. 2025, doi: 10.1038/s41746-025-02036-9.

- [15] J. Yadav, A. Rani, V. Singh, and B. M. Murari, "Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy," 2015, *Elsevier Ltd.* doi: 10.1016/j.bspc.2015.01.005.
- [16] S. V. K. R. Rajeswari and P. Vijayakumar, "Development of sensor system and data analytic framework for non-invasive blood glucose prediction," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-59744-7.
- [17] H. Agrawal, P. Jain, and A. M. Joshi, "Machine learning models for non-invasive glucose measurement: towards diabetes management in smart healthcare," *Health Technol. (Berl.)*, vol. 12, no. 5, pp. 955–970, Sep. 2022, doi: 10.1007/s12553-022-00690-7.
- [18] C. T. Yen, U. H. Chen, G. C. Wang, and Z. X. Chen, "Non-Invasive Blood Glucose Estimation System Based on a Neural Network with Dual-Wavelength Photoplethysmography and Bioelectrical Impedance Measuring," *Sensors*, vol. 22, no. 12, Jun. 2022, doi: 10.3390/s22124452.
- [19] W. K. Beh, Y. C. Yang, and A. Y. Wu, "Quality-Aware Signal Processing Mechanism of PPG Signal for Long-Term Heart Rate Monitoring †," *Sensors*, vol. 24, no. 12, Jun. 2024, doi: 10.3390/s24123901.
- [20] F. Hidayat, "MODEL ADDIE (ANALYSIS, DESIGN, DEVELOPMENT, IMPLEMENTATION AND EVALUATION) DALAM PEMBELAJARAN PENDIDIKAN AGAMA ISLAM ADDIE (ANALYSIS, DESIGN, DEVELOPMENT, IMPLEMENTATION AND EVALUATION) MODEL IN ISLAMIC EDUCATION LEARNING," *JIPAI; Jurnal Inovasi Pendidikan Agama Islam*, no. 1, Dec. 2021.
- [21] L. Breiman, "RANDOM FORESTS," *Random Forest*, vol. 7, no. 6, pp. 1–33, 2001, doi: 10.14569/ijacsa.2016.070603.
- [22] L. Breiman, *Random Forests*, vol. 45. 2001. doi: <https://doi.org/10.1023/A:1010933404324>.
- [23] Y. Chen, K. C. Cheung, R. Z. Sun, and S. C. P. Yam, "A user guide of CART and random forests with applications in FinTech and InsurTech," *Japanese Journal of Statistics and Data Science*, vol. 7, no. 2, pp. 999–1038, Nov. 2024, doi: 10.1007/s42081-024-00258-x.
- [24] E. Kavlakoglu, "What is random forest?" Accessed: Nov. 03, 2025. [Online]. Available: https://www.ibm.com/think/topics/random-forest?utm_source=chatgpt.com
- [25] Qi. Xiaokun; and He. Tian, "Analysis and Prediction of Energy Consumption in Neural Networks Based on Machine Learning," *Academic Journal of Computing & Information Science*, vol. 7, no. 4, 2024, doi: 10.25236/ajcis.2024.070412.
- [26] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012. doi: 10.1016/C2009-0-61819-5.
- [27] M. Yang, M. Zhao, D. Liu, M. Ma, and X. Su, "Improved Random Forest Method for Ultra-short-term Prediction of the Output Power of a Photovoltaic Cluster," *Front. Energy Res.*, vol. 9, Oct. 2021, doi: 10.3389/fenrg.2021.749367.
- [28] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Nov. 2017, [Online]. Available: <http://arxiv.org/abs/1705.07874>
- [29] ISO 15197, "In vitro diagnostic test systems — Requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus," 2013.
- [30] M. A. Al-Dhaheri, N. E. Mekkakia-Maaza, H. Mouhadjer, and A. Lakhdari, "Noninvasive blood glucose monitoring system based on near-infrared method," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 2, pp. 1736–1746, 2020, doi: 10.11591/ijece.v10i2.pp1736-1746.
- [31] M. Oloko-Oba, E. Esenogho, and K. Aruleba, "From Biosignals to Bedside: A Review of Real-Time Edge Machine Learning for Wearable Health Monitoring," May 01, 2026, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/bioengineering13050559.