EMITOR: Jurnal Teknik Elektro

A Portable IoT-Based System for Non-Invasive Early Stroke Risk Prediction Using Photoplethysmography and Logistic Regression

Agrippina W. R. G.[1*], M. Rochmad[1], Rizka S.[1], Rika R.[1], Kemalasari[1], Hanif J. I.[1], Khoironi[2]
[1]Department of Electrical Engineering – Politeknik Elektronika Negeri Surabaya
[2]Department of Informatics and Computer Engineering – Politeknik Elektronika Negeri Surabaya
Surabaya, Indonesia
*agrippina@pens.ac.id

Abstract – *Stroke is one of the leading causes of death and long-term disability in Indonesia. However, early stroke risk detection remains limited because conventional examination methods are often invasive, expensive, and impractical for routine health monitoring. This study develops a portable, non-invasive stroke risk prediction system using the MAX30105 sensor and photoplethysmography (PPG) signals to estimate blood pressure, blood glucose, and cholesterol levels. The measurement data are calibrated using a multi-layer perceptron (MLP), while logistic regression is employed to classify stroke risk into two categories: "Yes" and "No." The system is implemented using an ESP32 microcontroller and integrated with an Internet of Things (IoT) platform, enabling real-time measurement results to be displayed on both a liquid crystal display (LCD) and an Android application. Experimental results show that blood glucose measurement using the MAX30105 sensor achieved an accuracy of 86.60%, while cholesterol measurement achieved an accuracy of 95.07%. The accuracies of systolic and diastolic blood pressure measurements were 95.06% and 97.24%, respectively. Furthermore, the heart rate measurements deviated from the reference device by an average of 0.87%, corresponding to a precision level of 99.13%. The stroke risk classification model achieved an accuracy of 85.71%. These findings demonstrate that the proposed system provides reliable and consistent performance and has the potential to serve as a practical solution for non-invasive health monitoring and early stroke risk prediction.*

Keywords – *Stroke; non-invasive measurement; MAX30105; photoplethysmography; logistic regression.*

I. Introduction

STROKE is a non-communicable disease and one of the leading causes of mortality and disability worldwide. Nearly 15 million people worldwide experience a stroke each year [1]. According to the 2023 China Stroke Prevention and Control Report, China has the largest number of stroke patients worldwide, with an increasing incidence among younger populations [2]. In Indonesia, the number of stroke patients has reached approximately 500,000 and is predicted to continue increasing annually. According to data from the Indonesian Stroke Foundation, Indonesia ranked first in the number of stroke patients in 2009. Stroke in Indonesia predominantly affects elderly individuals, accounting for approximately 35.8% of cases, while 12.9% of affected patients die from the disease [3].

Stroke risk factors can be classified into two categories: non-modifiable and modifiable risk factors. Non-modifiable risk factors include age, sex, race, family history, and a history of previous stroke. Meanwhile, modifiable risk factors include hypertension, diabetes mellitus, smoking, obesity, and dyslipidemia [4]. These findings indicate that physiological and metabolic parameters, such as blood pressure, blood glucose levels, and cholesterol levels, can serve as important indicators for estimating an individual's risk of stroke.

Public awareness regarding regular health monitoring remains relatively low. Most commercially available blood glucose monitoring devices currently use invasive measurement techniques. Health parameter measurements can also be conducted through laboratory examinations; however, discomfort caused by needles, lengthy laboratory procedures, delayed results, and relatively high costs may discourage individuals from routinely monitoring stroke-related risk factors [5]. Furthermore, repeated finger-pricking procedures may cause tissue damage, infection, and reduced patient compliance [6]. Therefore, smart devices capable

The manuscript was received on July 14, 2026, revised on July 29, 2026, and published online on July 31, 2026. Emitor is a Journal of Electrical Engineering at Universitas Muhammadiyah Surakarta with ISSN (Print) 1411 – 8890 and ISSN (Online) 2541 – 4518, holding Sinta 3 accreditation. It is accessible at https://journals2.ums.ac.id/index.php/emitor/index.

of acquiring photoplethysmography (PPG) signals provide a promising opportunity for the development of non-invasive blood glucose monitoring systems [7].

Gusti et al. employed PPG signals combined with a Multi-Layer Perceptron (MLP) model to non-invasively estimate blood glucose, cholesterol, and uric acid levels. The proposed method achieved accuracies of 90.4%, 90.2%, and 87.7% for blood glucose, cholesterol, and uric acid estimation, respectively [8]. Rokhana et al. demonstrated that logistic regression could be employed to support early disease diagnosis [9]. Previous studies have also shown that an MLP model can process and learn complex nonlinear relationships from PPG signals acquired using the MAX30105 sensor [8]. Meanwhile, logistic regression provides a simple yet effective approach for calculating classification probabilities and is widely used because of its relatively straightforward implementation [10]. Another study investigated stroke prediction using logistic regression combined with several data preprocessing techniques and achieved an accuracy of 86% [11].

Based on these limitations and previous findings, this study proposes the development of a portable Internet of Things (IoT)-based system for non-invasive early stroke risk prediction by integrating PPG sensors, an MLP model, and logistic regression. The MLP model is used to process PPG signals and estimate multiple health parameters, including blood pressure, cholesterol levels, and blood glucose levels. Subsequently, logistic regression is employed to calculate the probability of stroke risk and classify users into at-risk and not-at-risk categories.

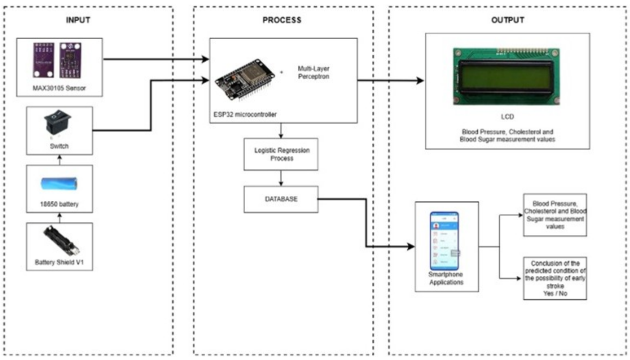


Figure 1: Architecture of the proposed IoT-based early stroke risk prediction system.

Unlike previous studies that primarily focused on monitoring a single health parameter or employed invasive diagnostic procedures, the proposed study integrates multi-parameter non-invasive sensing with lightweight machine-learning algorithms to provide rapid, real-time, and accurate stroke risk assessment. Furthermore, the proposed system is designed as a portable and low-cost preventive healthcare solution, making it suitable for continuous public health monitoring and practical implementation in everyday environments.

II. Research Methods

This section describes the proposed system architecture, the Multi-Layer Perceptron model used to estimate health parameters, the logistic regression model used to classify stroke risk, and the software design of the mobile monitoring application.

i. System Architecture

This study proposes an Internet of Things (IoT)-based early stroke detection system designed to estimate blood pressure, cholesterol, and blood glucose levels and to classify stroke risk in real time. The system consists of three main stages: input or data collection, processing or data preprocessing and classification, and output or IoT-based monitoring and visualization. The overall system architecture is presented in Figure 1.

As illustrated in Figure 1, the system operates through four main stages: data acquisition, health parameter estimation, data transmission, and stroke risk classification. First, the user places a finger on the MAX30105 sensor, which captures infrared (IR) signals and heart rate (HR). The system is powered by an 18650 battery, while the Battery Shield V1 regulates the voltage to $5, V$ to supply power to the MAX30105 sensor and a 16×2 LCD.

The MAX30105 optical sensor is employed as a non-invasive measurement device that detects infrared light reflected from the fingertip [12]. The acquired sensor data are transmitted to the ESP32 microcontroller through an Inter-Integrated Circuit (I^2C) interface. The ESP32 enables the acquisition and processing of PPG signals. Furthermore, its relatively low cost supports the development of an affordable and portable system [13].

The ESP32 subsequently processes the average IR and HR values together with the user's age using a pre-trained Multi-Layer Perceptron (MLP) model. The model estimates systolic blood pressure, diastolic blood pressure, blood glucose, and cholesterol levels. The estimated results are displayed on the LCD and simultaneously transmitted to a cloud database through a Wi-Fi connection.

Finally, the mobile application retrieves the stored data and applies a logistic regression model to classify the user's stroke risk. The application calculates the probability of stroke based on the estimated health parameters and user information. The predicted result is then displayed as an early stroke risk assessment.

ii. Multi-Layer Perceptron

Artificial Neural Networks (ANNs) are widely used for automatic learning and pattern classification. In medical applications, ANNs can support physiological variable prediction and disease diagnosis [14]. A Multi-Layer Perceptron is a commonly used ANN architecture consisting of an input layer, one or more hidden layers, and an output layer. Each layer contains computational elements known as neurons, and each neuron is associated with a specific weight and bias.

The input layer consists of n input variables expressed as $X = x_1, x_2, \ldots, x_n$, while the output layer consists of m output variables expressed as $Y = y_1, y_2, \ldots, y_m$. The best-performing model is determined by comparing several neural network architectures with different numbers of neurons in their hidden layers [15].

The model training process uses supervised learning with the Bayesian Regularization Backpropagation algorithm. Before the training process, the input data are normalized to the range of -1 to 1. The normalization procedure is defined in Equation (1) [16].

$$x_{\text{norm}} = 2\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right) - 1, \tag{1}$$

where x_{norm} is the normalized value, x is the original data value, $x_{\min}$ is the minimum data value, and $x_{\max}$ is the maximum data value.

Each hidden layer applies the hyperbolic tangent activation function, tanh, to model nonlinear relationships between the input and output variables. The general relationship between the network input and output is formulated in Equation (2) [17].

$$y = \omega * 2 \tanh\left(\omega * 1x + b * 1\right) + b * 2, \tag{2}$$

where $\omega * 1$ represents the weights connecting the input layer to the hidden layer, $\omega * 2$ represents the weights connecting the hidden layer to the output layer, $b * 1$ is the hidden-layer bias, $b * 2$ is the output-layer bias, x is the network input, and y is the network output.

The model is evaluated using the Mean Squared Error (MSE), which measures the average squared difference between the actual target values and the predicted values. The MSE is calculated using Equation (3) [17].

$$E_d = \frac{1}{N}\sum_{i=1}^{N}\left(t_i - y_i\right)^2, \tag{3}$$

where E_d is the Mean Squared Error, N is the number of observations, t_i is the actual target value for the i-th observation, and y_i is the corresponding model prediction.

The weight and bias update mechanism follows the Levenberg–Marquardt approach, which combines the gradient descent and Gauss–Newton methods. This algorithm is known for its relatively high convergence speed [18]. The parameter update mechanism is expressed in Equation (4).

$$\omega * \text{new} = \omega * \text{old} - \left(J^T J + \mu I\right)^{-1} J^T e, \tag{4}$$

where $\omega * \text{new}$ is the updated weight vector, $\omega * \text{old}$ is the previous weight vector, J is the Jacobian matrix, J^T is the transpose of the Jacobian matrix, μ is the damping parameter, I is the identity matrix, and e is the error vector.

The complete MLP training and prediction process is presented in Figure 2. The process includes data normalization, network training, weight and bias optimization, model evaluation, and health parameter prediction.

As shown in Figure 2, the trained MLP models are used to estimate four health parameters: blood glucose, cholesterol, systolic blood pressure, and diastolic blood pressure. Each parameter is estimated using a different nonlinear model.

The training process uses the supervised learning method with the Bayesian Regularization Backpropagation algorithm, with a weight and bias update mechanism that follows the Levenberg–Marquardt approach, as shown in Fig. 2. The following are the four nonlinear equations used in the prediction process:

1. Nonlinear equation of blood sugar model:

$$y = \frac{1}{0.03390}\left(0.3960 + \mathbf{a}^{\mathsf{T}} \tanh(\mathbf{W}\mathbf{x} + \mathbf{b}) + 1\right) + 87 \text{ mg/dL} \tag{5}$$

$$\mathbf{W} = \begin{bmatrix} -3.427 \times 10^{-4} & -2.945 \times 10^{-4} & 1.079 \times 10^{-4} \\ 3.452 \times 10^{-1} & 1.151 \times 10^{-1} & -9.910 \times 10^{-1} \\ 2.763 \times 10^{-4} & 2.374 \times 10^{-4} & -8.701 \times 10^{-5} \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} 2.528 \times 10^{-4} \\ 1.007 \times 10^{-1} \\ -2.038 \times 10^{-4} \end{bmatrix} \qquad \mathbf{a} = \begin{bmatrix} -2.362 \times 10^{-3} \\ -8.149 \times 10^{-1} \\ -1.045 \times 10^{-3} \end{bmatrix}$$

2. Nonlinear equation of cholesterol model:

$$y = \frac{1}{0.02817}\left(0.36485 + \mathbf{a}^{\mathsf{T}} \tanh(\mathbf{W}\mathbf{x} + \mathbf{b}) + 1\right) + 151 \text{ mg/dL} \tag{6}$$

$$\mathbf{W} = \begin{bmatrix} 9.509 \times 10^{-1} & -4.408 \times 10^{-1} & 1.181 \times 10^{-3} \\ -1.870 \times 10^{-4} & 5.513 \times 10^{-5} & 2.248 \times 10^{-4} \\ 4.378 \times 10^{-1} & 5.208 \times 10^{-1} & 5.283 \times 10^{-1} \\ -2.185 \times 10^{-2} & 3.419 \times 10^{-3} & 5.732 \times 10^{-3} \\ -1.089 \times 10^{-1} & 2.714 \times 10^{-1} & 1.789 \times 10^{0} \\ 4.221 \times 10^{-1} & -1.615 \times 10^{0} & -8.395 \times 10^{-1} \\ -4.525 \times 10^{-3} & 2.436 \times 10^{-4} & 4.424 \times 10^{-3} \\ 2.075 \times 10^{-1} & 6.127 \times 10^{-1} & -2.031 \times 10^{0} \\ -1.216 \times 10^{-3} & -5.984 \times 10^{-4} & 5.739 \times 10^{-4} \\ -6.046 \times 10^{-4} & -4.994 \times 10^{-4} & 5.900 \times 10^{-4} \\ -5.373 \times 10^{-1} & 3.804 \times 10^{-1} & 1.036 \times 10^{-1} \\ -1.478 \times 10^{-4} & -5.485 \times 10^{-5} & 1.059 \times 10^{-4} \end{bmatrix}$$

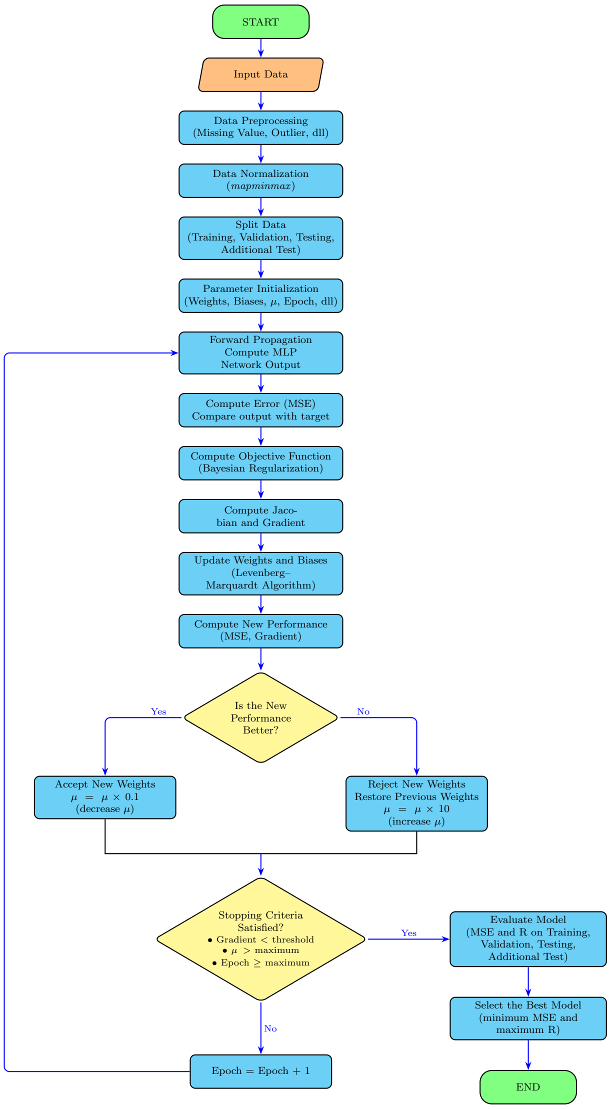


Figure 2: Process flow of the Multi-Layer Perceptron model.

$$\mathbf{b} = \begin{bmatrix} -5.249\times10^{-1} \\ 5.570\times10^{-4} \\ -4.385\times10^{-1} \\ 4.182\times10^{-2} \\ 5.980\times10^{-1} \\ 2.642\times10^{-1} \\ 1.210\times10^{-2} \\ -1.483\times10^{0} \\ 1.974\times10^{-3} \\ 3.104\times10^{-3} \\ -4.862\times10^{-1} \\ 5.316\times10^{-4} \end{bmatrix} \qquad \mathbf{a} = \begin{bmatrix} -9.593\times10^{-1} \\ 6.124\times10^{-4} \\ 6.564\times10^{-1} \\ 4.510\times10^{-2} \\ 1.489\times10^{0} \\ 1.104\times10^{0} \\ 1.341\times10^{-2} \\ 1.801\times10^{0} \\ 2.384\times10^{-3} \\ 3.228\times10^{-3} \\ -6.210\times10^{-1} \\ 5.622\times10^{-4} \end{bmatrix}$$

3. Nonlinear equation of systolic blood pressure model:

$$y = \frac{1}{0.03390}\left(-0.09215 + \mathbf{a}^{\mathsf{T}}\tanh(\mathbf{Wx}+\mathbf{b}) + 1\right) + 106 \text{ mmHg} \quad (7)$$

$$\mathbf{W} = \begin{bmatrix} 2.519\times10^{-1} & -3.089\times10^{-1} & -7.481\times10^{-2} \\ -2.356\times10^{-1} & 2.941\times10^{-1} & 8.410\times10^{-2} \\ 2.095\times10^{-1} & -2.642\times10^{-1} & -8.741\times10^{-2} \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} 8.063\times10^{-2} \\ -6.744\times10^{-2} \\ 5.520\times10^{-2} \end{bmatrix} \qquad \mathbf{a} = \begin{bmatrix} -4.364\times10^{-1} \\ 4.025\times10^{-1} \\ -3.648\times10^{-1} \end{bmatrix}$$

4. Nonlinear equation of diastolic blood pressure model:

$$y = \frac{1}{0.125}\left(-0.39449 + \mathbf{a}^{\mathsf{T}}\tanh(\mathbf{Wx}+\mathbf{b}) + 1\right) + 78 \text{ mmHg} \quad (8)$$

$$\mathbf{W} = \begin{bmatrix} -1.767\times10^{-4} & 1.015\times10^{-3} & 1.726\times10^{-3} \\ 1.653\times10^{-1} & -5.337\times10^{-1} & -6.653\times10^{-1} \\ -1.638\times10^{-4} & 9.413\times10^{-4} & 1.600\times10^{-3} \\ 1.829\times10^{-4} & -1.051\times10^{-3} & -1.787\times10^{-3} \\ -1.833\times10^{-4} & 1.053\times10^{-3} & 1.791\times10^{-3} \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} -8.658\times10^{-4} \\ -1.929\times10^{-1} \\ -8.026\times10^{-4} \\ 8.963\times10^{-4} \\ -8.982\times10^{-4} \end{bmatrix} \qquad \mathbf{a} = \begin{bmatrix} 2.888\times10^{-3} \\ -8.128\times10^{-1} \\ 2.103\times10^{-3} \\ -2.553\times10^{-3} \\ 2.604\times10^{-3} \end{bmatrix}$$

iii. Stroke Classification Using Logistic Regression

Regression is a statistical method used to determine the relationship between one or more predictor variables and a response variable. Predictor variables may also be referred to as explanatory, independent, or input variables, whereas the response variable may be referred to as the dependent or output variable [19].

Based on the measurement scale and type of response variable, logistic regression can be classified into binary, multinomial, and ordinal logistic regression. Binary logistic regression is widely used when the response variable contains two categories represented by the values 0 and 1 [20]. In this study, the two categories represent users who are not at risk of stroke and users who are at risk of stroke.

The logistic regression model calculates the probability of a stroke event and transforms the probability into a logit value. The logit function is expressed in Equation (9) [21].

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n, \quad (9)$$

where p is the probability of stroke, X_n represents the predictor variables, and β_n represents the regression coefficient associated with each predictor variable. The predictor variables used in this study include age, gender, stroke history, number of cigarettes smoked, Rapid Arterial Occlusion Evaluation (RACE) scale score, heart rate, systolic blood pressure, diastolic blood pressure, blood glucose, and cholesterol.

The equivalent probability function derived from Equation (9) is presented in Equation (10) [21].

$$p(Y=1) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}, \quad (10)$$

where $p(Y = 1)$ represents the probability that the user belongs to the stroke-risk category.

The resulting probability is compared with a classification threshold of 0.5. A probability of $p \geq 0.5$ is classified as 'Yes" or at risk of stroke, whereas a probability of $p < 0.5$ is classified as 'No" or not at risk of stroke.

The intercept and regression coefficients implemented in the logistic regression model are presented in Table 1.

Table 1: Logistic regression model coefficients.

Coefficient	Value
β_0	−95.5978
β_1	−0.0811
β_2	0.1454
β_3	0.4719
β_4	0.1092
β_5	0.7465
β_6	−0.2435
β_7	0.1641
β_8	0.8893
β_9	−0.1059
β_{10}	0.1635

As summarized in Table 1, the intercept β_0 and the ten predictor coefficients are substituted into Equation (10) to calculate the probability of stroke risk.

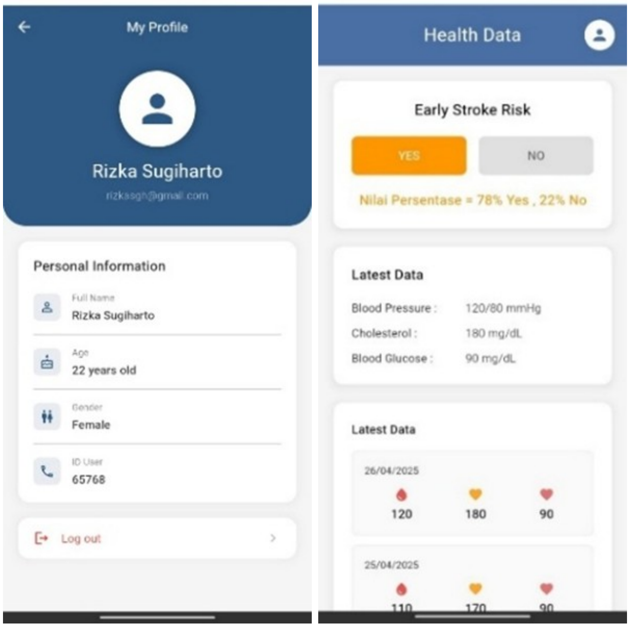
iv. Software Design

The database in the proposed system is implemented using Firebase Cloud Firestore to support real-time data storage and synchronization. Firebase provides real-time database services, authentication, and notification capabilities that support efficient IoT-based health monitoring and timely responses [22].

Firebase functions as middleware connecting the ESP32 microcontroller to the Flutter-based mobile application. The transmitted sensor data include systolic blood pressure, diastolic blood pressure, cholesterol, blood glucose, and heart rate. Data acquired by the MAX30105 sensor are processed using the MLP model on the ESP32 to generate estimates of the health parameters. The mobile application subsequently combines the estimated sensor data with user profile data and applies logistic regression to detect and classify stroke risk.

The main interfaces of the developed mobile application are presented in Figure 3. Figure 3(a) displays real-time monitoring data and stroke risk classification results, while Figure 3(b) displays the user profile page.

As shown in Figure 3, the mobile application was



(a) (b)

Figure 3: Mobile application interfaces: (a) real-time monitoring and stroke risk classification results and (b) user profile page.

developed using Flutter and operates on Android devices. The application serves as the primary user interface for displaying sensor measurements and stroke risk classification results. Its main features include a user profile page, real-time monitoring, measurement history, and stroke risk prediction based on sensor readings and information entered through the application. The profile page also provides a unique identification number for each user.

III. Results and Discussion

This section presents the data collection procedure, MAX30105 sensor performance, Multi-Layer Perceptron performance, logistic regression classification results, and interpretation of the factors associated with stroke risk.

i. Data Collection

The training data were collected through direct measurements involving 50 participants aged between 20 and 70 years. All participants provided informed consent before the data acquisition process. The collected data consisted of input data obtained from the MAX30105 sensor and participant age, while the target data were acquired using the Elvasense 3-in-1 measurement device.

All measurements were conducted in a calm and relaxed environment to minimize noise and improve data quality. Additional participant information, including age, gender, smoking status, and the number of cigarettes smoked per day, was also collected to support the stroke risk analysis.

The model performance was evaluated using a confusion matrix, accuracy, precision, sensitivity, specificity, F1-score, and Receiver Operating Characteristic (ROC) analysis. The evaluation results indicated that the proposed system achieved an accuracy of 82%, a sensitivity of 50%, and a specificity of 80%. These findings indicate that the model was capable of distinguishing between stroke-risk and non-stroke-risk conditions with relatively good performance.

ii. MAX30105 Sensor Performance

The performance of the MAX30105 sensor was quantitatively evaluated across participants of different ages. The heart rate measurements produced by the sensor were compared with measurements obtained using a reference sphygmomanometer. The comparison results are presented in Figure 4.

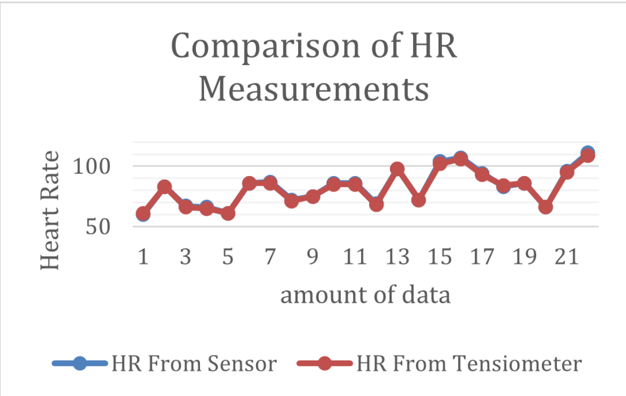


Figure 4: Comparison of heart rate measurements obtained using the MAX30105 sensor and the reference device.

As shown in Figure 4, the MAX30105 sensor was capable of detecting heart rate values that were nearly identical to those obtained using the reference device. The average measurement error was 0.87%, corresponding to a precision level of 99.13%. These results indicate that the MAX30105 sensor provides sufficiently precise measurements for implementation in a heart rate monitoring system.

iii. Multi-Layer Perceptron Performance

The search for the optimal Multi-Layer Perceptron model was conducted using a supervised learning approach with the Bayesian Regularization Backpropagation algorithm. Blood glucose, cholesterol, systolic blood pressure, and diastolic blood pressure were trained separately using a predetermined dataset.

This approach enables the model to learn complex relationships between the input and target variables, thereby improving the accuracy of the resulting predictions. The comparison between the MLP model predictions and invasive reference measurements is presented in Table 2.

Table 2: Comparison of the accuracy of MLP predictions and invasive reference measurements.

Evaluation Metric	Blood Glucose	Cholesterol	Systolic	Diastolic
Average error (%)	13.3988	4.9300	4.9396	2.7603
Accuracy (%)	86.6012	95.0700	95.0604	97.2397

As summarized in Table 2, the proposed MLP model achieved accuracies of 86.60% for blood glucose, 95.07% for cholesterol, 95.06% for systolic blood pressure, and 97.24% for diastolic blood pressure.

Although the proposed approach is similar to those used in previous studies, this study employed a larger training dataset with greater data variability. Consequently, the prediction accuracies for cholesterol and blood pressure were improved compared with those reported in previous studies.

iv. Regression Performance for Stroke Diagnosis

The logistic regression model was implemented using JupyterLab Portable and the Python programming language. The implementation began by initializing the required libraries, including `pandas`, `numpy`, `scikit-learn`, `matplotlib`, and `seaborn`.

The measurement dataset was subsequently loaded, and its structure was examined to ensure that all variables were correctly organized. The input variables, denoted by X, consisted of age, gender, stroke history, number of cigarettes smoked, RACE scale score, heart rate, systolic blood pressure, diastolic blood pressure, blood glucose, and cholesterol. The output variable, denoted by y, represented the stroke-risk classification target. An example of the logistic regression probability calculation is presented in below.

$$p(Y) = \frac{\exp(z)}{1+\exp(z)}$$

$$\begin{aligned} z = {} & -95.59 + (-0.08 \times 58) + (0.14 \times 1) + (0.47 \times 1) \\ & + (0.10 \times 0) + (0.74 \times 2) + (-0.24 \times 75) + (0.16 \times 136) \\ & + (0.88 \times 85) + (-0.10 \times 120) + (0.1 \times 232) \end{aligned}$$

It shows the probability calculation obtained by substituting the participant's health parameters and profile information into the logistic regression model.

The resulting probability is compared with a classification threshold of 0.5. When the probability value satisfies $p(Y = 1) \geq 0.5$, the participant is classified as being at risk of stroke. Conversely, when $p(Y = 1) < 0.5$, the participant is classified as not being at risk of stroke.

The "stroke risk" label used in this study was obtained through participant interviews and medical history records. Participants who had previously been diagnosed with stroke or demonstrated clinical indications associated with stroke risk based on previous medical examinations were categorized as being at risk. Participants without such a history or clinical indication were categorized as not being at risk.

The logistic regression testing results for female and male participants are presented in Table 3.

Table 3: Logistic regression testing results for female and male participants. SH: stroke history (1 = yes, 0 = no); CS: cigarettes smoked; RACE: Rapid Arterial oCclusion Evaluation scale; HR: heart rate (bpm); SBP: systolic blood pressure (mmHg); DBP: diastolic blood pressure (mmHg); BG: blood glucose (mg/dL); CHOL: total cholesterol (mg/dL).

Age	SH	CS	RACE	HR	SBP	DBP	BG	CHOL	Diagnosis
Female Participants									
22	1	0	0	79	118	80	124	166	No Risk
28	0	0	0	77	111	79	119	164	No Risk
25	0	0	0	83	129	82	135	206	No Risk
66	0	0	0	88	165	87	134	184	No Risk
50	0	0	0	83	158	94	124	188	At Risk
58	0	0	0	73	138	85	119	169	No Risk
58	0	0	0	75	134	84	95	220	At Risk
Male Participants									
21	0	10	0	102	134	84	146	227	No Risk
52	0	36	0	83	121	83	120	192	No Risk
28	0	0	0	85	121	80	124	188	No Risk
58	1	0	2	75	136	85	120	232	At Risk
25	0	10	0	71	111	80	87	162	No Risk
44	1	36	0	89	169	101	141	208	At Risk
64	0	24	2	74	150	94	116	203	At Risk

As shown in Table 3, the logistic regression model classified each participant into either the 'At Risk" or 'No Risk" category based on the probability generated from the participant's physiological and demographic variables.

The model performance was assessed using accuracy, precision, recall, and F1-score metrics supported by a confusion matrix. The evaluation results showed that the model achieved an accuracy of 85.7% and demonstrated good performance in identifying non-stroke cases. However, the recall value for the stroke-risk class remained relatively low, indicating that the model was not fully effective in identifying all positive cases.

The complete classification report generated using JupyterLab Portable is presented in Table 4.

As reported in Table 4, the logistic regression model achieved an overall accuracy of 85.71%, with a weighted precision of 0.88, weighted recall of 0.86, and weighted F1-score of 0.84.

For the non-stroke-risk class, denoted by class 0, the model achieved a precision of 0.83, recall of 1.00,

Table 4: Logistic regression classification results obtained using JupyterLab Portable.

Class	Precision	Recall	F1-Score	Support
0	0.83	1.00	0.91	5
1	1.00	0.50	0.67	2
Accuracy	–	–	0.86	7
Macro average	0.92	0.75	0.79	7
Weighted average	0.88	0.86	0.84	7

and F1-score of 0.91. The recall value of 1.00 indicates that all actual non-stroke-risk samples in the testing data were correctly identified.

For the stroke-risk class, denoted by class 1, the model achieved a precision of 1.00, indicating that every sample predicted as being at risk of stroke was correctly classified. However, the recall value was only 0.50, indicating that the model detected only half of the actual stroke-risk cases. Consequently, the F1-score for the stroke-risk class was 0.67.

These results indicate that although the proposed model achieved relatively good overall classification performance, its sensitivity to stroke-risk cases requires further improvement. Increasing the quantity and class balance of the training data may improve the model's ability to detect positive stroke-risk cases.

v. Analysis of Stroke Risk Factors

Based on the logistic regression results, age was identified as the most dominant factor influencing increased stroke risk in both male and female participants. Increasing age, particularly above 50 years, tends to be accompanied by increased blood pressure, blood glucose, and cholesterol levels because of declining blood vessel function and metabolic regulation.

Bosu et al. explained that increasing age causes reduced blood vessel elasticity and impaired metabolic regulation, thereby increasing hypertension and stroke risk among older adults [23].

Among male participants, smoking contributed to increased stroke risk through blood vessel damage and increased blood pressure. Nuraisyah et al. reported a significant relationship between smoking, lifestyle, and hypertension in older adults, which may consequently increase stroke risk [24]. Among female participants, age and family history appeared to have a greater influence on the predicted stroke risk.

Compared with the Decision Tree method, logistic regression demonstrated comparable and relatively stable predictive performance while providing a clearer interpretation of the contribution of each independent variable through its regression coefficients.

A previous study concerning the application of logistic regression and Decision Tree models to predict activities of daily living in patients with stroke reported that both methods produced similar ROC performance. However, logistic regression was considered more systematic and suitable for clinical interpretation because it explained the relationships between risk factors and stroke outcomes more consistently [25].

Therefore, stroke risk is influenced by interactions among physiological, demographic, and lifestyle factors. Age appears to be a dominant factor that strengthens the influence of other variables, while logistic regression provides interpretability and reliability that may support clinical decision-making.

IV. Conclusion

The results of this study indicate that the proposed non-invasive measurement system using the MAX30105 sensor achieved relatively high accuracy for multiple health parameters. The system achieved accuracies of 86.60% for blood glucose, 95.07% for cholesterol, 95.06% for systolic blood pressure, and 97.24% for diastolic blood pressure.

Furthermore, the relatively low measurement variation indicates that the proposed system has good precision and is capable of producing consistent measurements. These findings demonstrate the potential reliability of the system for repeated health monitoring.

The implementation of the Multi-Layer Perceptron model showed that each health parameter required a different optimal hidden-layer configuration. This finding indicates that model complexity should be adjusted to the characteristics of each parameter to achieve optimal predictive performance. The variation in model architecture emphasizes the importance of adaptive model design in multi-parameter health monitoring systems.

In the proposed system, the MLP and logistic regression models perform complementary functions. The MLP is employed as a nonlinear regression model to estimate physiological parameters, including blood pressure, blood glucose, and cholesterol, from the IR signal and heart rate measured by the MAX30105 sensor.

The estimated physiological parameters, together with other stroke risk factors, are subsequently used as input features for the logistic regression model. The logistic regression model calculates the probability of stroke and classifies the user into a stroke-risk category. This sequential integration enables the system to transform raw sensor signals into clinically meaningful information before performing stroke risk assessment.

The logistic regression model demonstrated an accuracy of 85.71%, with weighted precision, recall, and F1-score values of 0.88, 0.86, and 0.84, respectively. These results indicate that the proposed system has promising potential as a non-invasive health monitoring and early stroke risk assessment tool.

Nevertheless, further development is required to improve prediction accuracy and minimize measurement errors. One limitation of this study is that the reference measurements for blood glucose and cholesterol were obtained using commercial portable devices rather than laboratory-grade biochemical analyses. This limitation may affect clinical-level accuracy and influence the predictive performance of the model.

In addition, the limited number of participants and physiological variations may restrict the generalization capability of the proposed system. Future work will therefore focus on collecting a larger and more diverse dataset to improve the generalization performance of both the MLP and logistic regression models.

Future studies will also employ laboratory-based reference measurements for biochemical parameters, investigate more advanced machine learning algorithms for stroke risk prediction, and validate the proposed system in broader clinical settings to enhance its accuracy and practical applicability.

Acknowledgment

This research was supported by the 2026 Internal Research Grant of Politeknik Elektronika Negeri Surabaya (PENS) under Contract No. 012/SP2H/PENS/PPK/2026.

References

[1] World Health Organization Regional Office for the Eastern Mediterranean (WHO EMRO), "Stroke, cerebrovascular accident," Health Topics, 2026, accessed: July 28, 2026. https://www.emro.who.int/health-topics/stroke-cerebrovascular-accident/

[2] W.-J. Tu and L.-D. Wang, "China stroke surveillance report 2021," *Military Medical Research*, vol. 10, 2023, article No. 33.

[3] R. S. Syareza, R. Oktiasari, P. Madona, E. Susianti, and M. Sahar, "Alat bantu terapi pasca stroke untuk tangan," *Jurnal Politeknik Caltex Riau*, vol. 4, no. 5, 2018.

[4] R. T. Pinzon, *Awas Stroke*. Yogyakarta, Indonesia: Betha Grafika, 2016.

[5] A. Kusuma, "Jutaan orang fobia jarum suntik, akankah jadi kendala vaksinasi COVID-19?" Kompas, Jan. 2021.

[6] E.-Y. Park, J. Baik, H. Kim, S.-M. Park, and C. Kim, "Ultrasound-modulated optical glucose sensing using a 1645 nm laser," *Scientific Reports*, vol. 10, 2020, article No. 13361. https://doi.org/10.1038/s41598-020-70299-4

[7] E. Vargov'a, A. Němcov'a, and Z. Nov'akov'a, "Non-invasive PPG-based estimation of blood glucose level,"

L'ekař a technika–Clinician and Technology, vol. 53, no. 1, pp. 19–24, 2023. https://doi.org/10.14311/CTJ.2023.1.04

[8] A. W. R. Gusti, Kemalasari, and J. S. P. C. Parasian, "Deteksi dini kesehatan berdasarkan nilai kadar gula darah, kolesterol, dan asam urat non-invasif dengan Multi-Layer Perceptron," *Jurnal Riset Rekayasa Elektro*, vol. 7, no. 2, pp. 127–136, Dec. 2025. https://doi.org/10.30595/jrre.v7i2.27002

[9] R. Rokhana, A. W. R. Gusti, and R. U. Salim, "Smart system for early diagnosis of gastroesophageal reflux disease (GERD)," *Journal of Applied Electrical Engineering*, vol. 9, no. 2, pp. 169–175, Dec. 2025.

[10] GeeksforGeeks, "Logistic regression in machine learning," GeeksforGeeks, Dec. 2025, published December 23, 2025; accessed July 28, 2026. https://www.geeksforgeeks.org/machine-learning/understanding-logistic-regression/

[11] M. Guhdar, A. I. Melhum, and A. L. Ibrahim, "Optimizing accuracy of stroke prediction using logistic regression," *Journal of Technology and Informatics (JoTI)*, vol. 4, no. 2, pp. 97–103, 2023. https://doi.org/10.37802/joti.v4i2.278

[12] D. Rahmadaniar, I. Salamah, and M. M. Rose, "Application of polynomial regression method in non-invasive measurement of blood sugar, cholesterol, and uric acid based on IoT," *INVOTEK: Jurnal Inovasi Vokasional dan Teknologi*, vol. 24, no. 3, pp. 187–196, Jul. 2025.

[13] L. Yan, Z. Long, J. Qian, J. Lin, S. Q. Xie, and B. Sheng, "Rehabilitation assessment system for stroke patients based on fusion-type optoelectronic plethysmography device and multi-modality fusion model: Design and validation," *Sensors*, vol. 24, no. 9, 2024, article No. 2925. https://doi.org/10.3390/s24092925

[14] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: A review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, pp. 159–190, 2006. https://doi.org/10.1007/s10462-007-9052-3

[15] K. Y. Chan, B. Abu-Salih, R. Qaddoura, A. M. Al-Zoubi, V. Palade, D.-S. Pham, J. Del Ser, and K. Muhammad, "Deep neural networks in the cloud: Review, applications, challenges and research directions," *Neurocomputing*, vol. 545, May 2023, article No. 126327.

[16] MathWorks, "mapminmax: Process matrices by mapping row minimum and maximum values to [-1, 1]," MathWorks Documentation, 2026, accessed: July 28, 2026. https://www.mathworks.com/help/deeplearning/ref/mapminmax.html

[17] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, Nov. 1994. https://doi.org/10.1109/72.329697

[18] C. Yildiz, K. Guney, M. Turkmen, and S. Kaya, "Neural models for coplanar strip line synthesis," *Progress In Electromagnetics Research*, vol. 69, pp. 127–144, 2007. https://doi.org/10.2528/PIER06120802

[19] Olah Data Skripsi, "Analisis regresi logistik SPSS," Mar. 2017, published March 17, 2017; accessed July 28, 2026. https://www.olahdataskripsi.com/2017/03/analisis-regresi-logistik-spss.html

[20] A. I. Sofiyat, A. Tjalla, and Mahdiyah, "Pemodelan regresi logistik biner terhadap penerimaan pegawai di PT XYZ Jakarta," *Matematika Sains*, vol. 1, no. 1, 2023.

[21] E. Roflin, F. Riana, E. Munarsih, Pariyana, and I. A. Liberty, *Regresi Logistik Biner dan Multinomial*. Indramayu, Indonesia: Penerbit NEM, 2021.

[22] A. Efendi, M. I. Ammarullah, I. G. T. Isa, M. P. Sari, J. N. Izza, Y. S. Nugroho, H. Nasrullah, and D. Alfian, "IoT-based elderly health monitoring system using firebase cloud computing," *Health Science Reports*, vol. 8, no. 3, 2025, article No. e70498. https://doi.org/10.1002/hsr2.70498

[23] W. K. Bosu, J. M. K. Aheto, E. Zucchelli, and S. T. Reilly, "Determinants of systemic hypertension in older adults in africa: A systematic review," *BMC Cardiovascular Disorders*, vol. 19, pp. 1–14, 2019, article No. 173. https://doi.org/10.1186/s12872-019-1147-7

[24] F. Nuraisyah, A. Irjayanti, M. Irmanto, and Noorkhalisah, "Smoking, physical inactivity and dietary with hypertension in the elderly: A systematic review and meta-analysis," *Jurnal Berkala Epidemiologi*, vol. 13, no. 3, pp. 289–298, Sep. 2025. https://doi.org/10.20473/jbe.v13i32025.289-298

[25] Q. Zhang, Z. Zhang, X. Huang, C. Zhou, and J. Xu, "Application of logistic regression and decision tree models in the prediction of activities of daily living in patients with stroke," *Computational and Mathematical Methods in Medicine*, vol. 2022, pp. 1–9, 2022, article ID 9662630. https://doi.org/10.1155/2022/9662630