Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 1017-1032 e-ISSN: 2656-8632
RESEARCH ARTICLE OPEN ACCESS

Wavelength Configuration and Signal Duration for Low-Complexity PPG-Based Anemia Detection: A Preliminary Validation Study

Mulia Rahmah, Fatma Indriani, Rudy Herteno, Radityo Adi Nugroho, and Irwan Budiman

Department of Computer Science, Lambung Mangkurat University, Banjarbaru, Indonesia
Corresponding author: Fatma Indriani (e-mail: f.indriani@ulm.ac.id), **Author(s) Email**: Mulia Rahmah (e-mail: muliarahmah130@gmail.com), Rudy Herteno (e-mail: rudy.herteno@ulm.ac.id), Radityo Adi Nugroho (e-mail: radityo.adi@ulm.ac.id), Irwan Budiman (e-mail: irwan.budiman@ulm.ac.id)

Abstract **Anemia remains a major global health problem, while standard diagnosis still depends on invasive hemoglobin testing, which may be less practical for repeated and resource-limited screening. Photoplethysmography (PPG) offers a potential non-invasive alternative, but the contribution of different wavelength configurations to anemia classification remains unclear. This preliminary subject-based validation study evaluated the effect of PPG wavelength configuration and recording duration on low-complexity anemia classification. A public dataset containing green, red, and infrared PPG recordings from 52 subjects was used, consisting of 42 normal and 10 anemia subjects. Eight morphological and temporal features were extracted from each wavelength. Seven signal configurations, namely Green, Red, IR, Green+Red, Green+IR, Red+IR, and all channels, were evaluated across 30, 45, 60, and 90 s recording durations. Support Vector Machine, Logistic Regression, Random Forest, and Extra Trees classifiers were trained using class-weighted learning and assessed with 5-fold subject-based cross-validation to reduce subject-level data leakage. The Red+IR configuration with a class-weighted SVM at 90 s achieved the best pooled performance, with a macro F1-score of 0.754, F1-Anemia of 0.588, anemia recall of 0.500, anemia precision of 0.714, accuracy of 0.769, and an error rate of 0.231. Fold-wise analysis showed substantial variability, with a macro F1-score of 0.617 ± 0.251, sensitivity of 0.467 ± 0.506, specificity of 0.846 ± 0.144, ROC-AUC of 0.864 ± 0.150, and PR-AUC of 0.694 ± 0.344. These findings suggest that adding more PPG wavelengths does not necessarily improve classification performance. However, the model still missed 5 of 10 anemia cases, and the limited anemia recall, small minority class, and demographic imbalance indicate that the results should be interpreted as preliminary and require validation on larger, more balanced datasets.**

Keywords **anemia; photoplethysmography; multispectral PPG; anemia screening; non-invasive detection; support vector machine; binary classification.**

I. Introduction

Anemia screening in low- and middle-income settings remains constrained by limited access to laboratory facilities and trained personnel, while the prevalence of anemia is still high among women of reproductive age and children [1][2]. Anemia can reduce physical capacity, cause fatigue, impair cognitive function, and increase the risk of health complications [3]. Early detection is therefore important to support timely intervention and prevent disease progression. At present, the diagnosis of anemia still relies mainly on hemoglobin measurements from venous or capillary blood samples [4]. Although this approach remains the clinical standard, it is invasive, requires trained personnel, and is less practical for repeated or large-scale screening [5]. These limitations motivate the development of simpler, faster, and more deployable non-invasive screening approaches. In this context, photoplethysmography (PPG) is a promising technique because it records peripheral blood-volume changes non-invasively using low-cost optical instrumentation that can be integrated into portable or wearable devices [6][7].

Previous studies have explored non-invasive hemoglobin estimation and anemia detection using smartphone imaging and wearable optical systems [8][9][10]. Other studies have examined hyperspectral imaging, diffuse reflectance spectroscopy, and PPG-

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

based analysis [11][12][13]. Among these approaches, PPG is attractive because its waveform reflects not only pulsatile blood-flow dynamics but also morphological information influenced by vascular, tissue, and optical sensing characteristics [14]. Machine learning has therefore been increasingly used to extract clinically relevant patterns from PPG signals in healthcare applications [15]. In addition to anemia-related analysis, PPG has been applied to arrhythmia detection, microcirculation monitoring, and cuffless blood-pressure estimation [16][17][18]. In the context of anemia, earlier studies have demonstrated the feasibility of using PPG features for hemoglobin estimation and classification [19][20]. Recent studies have further extended this direction using multispectral PPG and ensemble-based models for hemoglobin estimation or anemia-related classification [21][22][23]. Multi-wavelength PPG is physiologically relevant because different wavelengths have different penetration depths and optical sensitivities to blood and tissue properties [24][25]. Similar ideas have also been used for glycated hemoglobin and VO2 estimation [26][27].

Despite these advances, an important gap remains in determining which PPG wavelength configuration is most useful for anemia classification [28][29]. Many previous studies have used a single PPG wavelength, a limited subset of optical channels, or multi-wavelength input as a fixed design, without systematically comparing whether one-channel, two-channel, or three-channel configurations provide the best classification trade-off [30][31]. Smartphone-based non-invasive anemia detection has also been reported, but this approach does not directly address the role of PPG wavelength configuration [32]. This issue is important because adding more wavelengths may increase hardware complexity without necessarily improving classification performance. Therefore, the novelty of this work lies not in proposing a new sensor or classifier, but in evaluating whether increasing wavelength complexity actually improves anemia classification performance. Specifically, this study compares seven PPG signal configurations: Green, Red, IR, Green+Red, Green+IR, Red+IR, and all channels. This comparison is designed to determine whether the best performance is obtained from a single-channel, dual-channel, or three-channel configuration. In addition, careful model evaluation is required in PPG-based machine-learning applications because performance can be influenced by feature selection, validation design, and data partitioning strategies [33]. Rigorous validation is also needed because data leakage can occur when multiple segments from the same subject appear in both training and test sets [34][35]. Therefore, subject-based data partitioning is needed to reduce overly optimistic performance estimates in supervised biomedical machine-learning studies [36].

This study addresses these gaps by examining how wavelength configuration and signal duration jointly affect machine-learning performance in preliminary PPG-based anemia classification. The dataset consisted of 52 subjects with triple-wavelength PPG recordings and reference hemoglobin values [37]. In addition, PPG waveform analysis can provide complementary biomarkers related to cardiovascular and vascular characteristics [38]. The role of signal duration also requires further investigation, because shorter recordings are more practical for screening, while longer recordings may provide more stable morphological features and signal-derived indices [39][40]. Remote PPG has also been investigated for hemoglobin-related assessment, supporting the broader potential of optical physiological signals for non-invasive screening applications [41].

Therefore, the aim of this study is to evaluate how different PPG wavelength configurations and recording durations affect the performance of preliminary anemia classification using classical machine learning models. Seven channel configurations representing single-wavelength, dual-wavelength, and three-wavelength inputs were evaluated across four recording durations: 30, 45, 60, and 90 s. The evaluated configurations were Green, Red, IR, Green+Red, Green+IR, Red+IR, and all channels. Four classical machine learning classifiers were compared, namely Support Vector Machine, Logistic Regression, Random Forest, and Extra Trees, using class-weighted learning to reduce the effect of class imbalance. All experiments used subject-based cross-validation to obtain a more reliable internal validation estimate and to prevent subject-level data leakage.

The specific contributions of this study are as follows:

1. A systematic comparison of Green, Red, IR, Green+Red, Green+IR, Red+IR, and all-channel PPG configurations was performed to identify which level of wavelength complexity provides the best anemia-classification performance.
2. The effect of recording duration was evaluated at 30, 45, 60, and 90 s under the same subject-based validation framework, allowing the interaction between wavelength configuration and signal length to be examined.
3. Four classical machine learning classifiers, namely Support Vector Machine, Logistic Regression, Random Forest, and Extra Trees, were evaluated using class-weighted learning and subject-based

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

cross-validation to reduce the effect of class imbalance and subject-level data leakage.
4. The best-performing configuration was interpreted using clinically relevant evaluation metrics, including anemia recall, specificity, balanced accuracy, ROC-AUC, PR-AUC, confusion matrices, baseline comparison, and fold-wise variability.

The remainder of this paper is organized as follows. Section II describes the research method, including the dataset, data processing, classification procedure, and evaluation strategy. Section III presents the experimental results, including comparisons of channel configurations, performance across signal durations, a baseline comparison, and additional validation. Section IV discusses the findings in terms of model performance, signal configuration, and study limitations. Finally, Section V concludes the paper.

II. Method

A. Dataset

The overall experimental workflow is illustrated in Fig. 1. The diagram clarifies the complete pipeline from public PPG-Hb dataset selection, duration-based segmentation, preprocessing and signal-quality checking, PPG feature extraction, wavelength-configuration construction, class-weighted classification, subject-based validation, and final performance evaluation. This block diagram also separates the main experimental factors, namely recording duration, wavelength setting, model selection, and leakage-control strategy.

This study was conducted in two experimental stages. The first stage used features extracted from the full 90 s recording to compare classification models and signal combinations. The second stage evaluated the effect of signal duration by analyzing configurations of 30 s, 45 s, 60 s, and 90 s. This two-stage design allowed the best-performing model from the initial experiment to be examined more closely under different duration settings.

The task was formulated as a binary classification problem between anemia and normal classes. Subjects were labeled as anemic when Hb < 12 g/dL and as normal when Hb ≥ 12 g/dL. A total of 52 subjects were successfully processed, consisting of 42 normal subjects and 10 anemia subjects. This distribution indicates a clear class imbalance, with anemia representing only a small minority of the dataset. Preliminary exploratory analysis also showed a demographic imbalance: all anemia cases were female, and the average age of the anemia group was higher than that of the normal group. The mean age was 22.3 years in the normal group and 28.8 years in the anemia group, with a Mann-Whitney U test indicating a statistically significant age difference between the two groups ($p = 0.0006$). No missing values or infinite values were found in the generated feature dataset.

Although age and sex were available in the raw dataset, they were not used as predictors in the final model. The exclusion of age and sex was intended to reduce the direct demographic dependence of the classifier. However, this step does not fully eliminate demographic confounding, as sex and age may still indirectly influence PPG morphology. Therefore, the results should be interpreted cautiously, particularly because the classifier may still learn signal patterns associated with demographic differences rather than anemia-specific physiological characteristics. For this

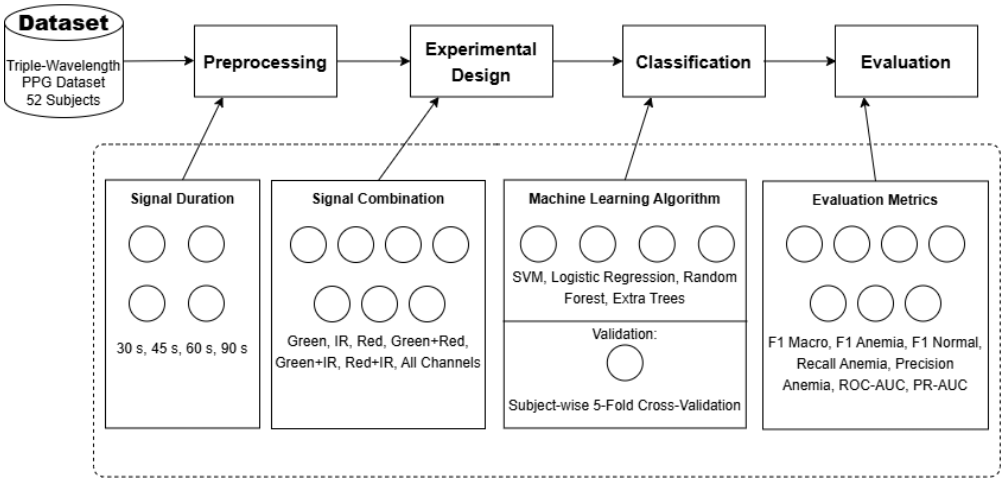

Fig. 1. Flowchart of research experimental design for anemia detection based on multispectral PPG signals

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

reason, this study focused only on features derived from the PPG signals, which aligns with the objective of evaluating a low-complexity, signal-only classification approach.

As the output of the data-preparation stage, the pipeline generated multiple feature datasets for each duration configuration. The 30 s configuration produced 156 segments, the 45 s configuration produced 104 segments, and the 60 s and 90 s configurations produced 52 segments each. Each dataset contained PPG features derived from the green, red, and infrared channels together with metadata required for subject-based validation, including subject identity and fold assignment. These datasets were then used to compare model performance across channel combinations and signal durations for preliminary non-invasive anemia classification.

B. Data Processing

Each raw recording contains three PPG channels: LEDC1 (green, 535 nm), LEDC2 (infrared, 880 nm), and LEDC3 (red, 660 nm) [37]. These wavelengths differ in tissue penetration depth and optical sensitivity, which is central to the channel-comparison design of this study [23]. The analysis used the filtered version of the public PPG dataset. No additional digital filtering, motion-artifact removal, or baseline correction was applied beyond the dataset preparation stage and the peak/trough selection procedure used during feature extraction. Signal quality was checked at the feature-table level by verifying the presence of missing values and infinite values. Segments with insufficient detected peaks or troughs were handled by assigning zero values to the corresponding temporal or amplitude feature. This limited signal-quality procedure is acknowledged as a methodological limitation.

For a duration L and sampling frequency fs = 100 Hz, the number of samples in each segment was computed using Eq. (1), and the non-overlapping segment signal for channel c and segment s was defined using Eq. (2). In Eq. (1), Ns denotes the number of samples in one segment, L denotes the selected recording duration in seconds, and fs denotes the sampling frequency in hertz. In Eq. (2), $x_{c,s}[n]$ denotes the nth sample of segment s from PPG channel c, x_c [ ] denotes the original filtered signal of that channel, n is the within-segment sample index from 0 to Ns - 1, and s identifies the non-overlapping segment order.

$$N_s = L \times f_s \quad (1)$$

$$x_{c,s}[n] = x_c[n + s \cdot N_s], \quad n = 0,1,\dots,N_s - 1 \quad (2)$$

The 90 s recordings were segmented non-overlappingly into three 30 s windows, two 45 s windows, or single 60 s and 90 s windows. These configurations corresponded to 3000, 4500, 6000, and 9000 samples, respectively. Peak and trough candidates were detected from the filtered PPG signal using the SciPy find_peaks function. Peak detection used a minimum distance of 50 samples and a prominence of 5. At 100 Hz, the distance parameter corresponds to 0.5 s, which reduces the number of physiologically implausible detections above approximately 120 beats per minute. The prominence value was selected empirically to suppress low-amplitude fluctuations while preserving dominant systolic peaks. These parameters were kept fixed across wavelengths and durations to avoid channel- or duration-specific overfitting.

$$p_i = argmax_n x_{c,s}[n] \quad (3)$$

$$t_i = argmin_n x_{c,s}[n] \quad (4)$$

In Eq. (3) and Eq. (4), pi and ti denote the sample indices of the ith detected systolic peak and diastolic trough, respectively. The operators $argmax_n$ and $argmin_n$ return the sample index with the maximum or minimum amplitude within the local search interval, while $x_{c,s}[n]$ represents the PPG amplitude at sample n for channel c and segment s.

$$A_i = x_{c,s}[p_i] - x_{c,s}[t_i] \quad (5)$$

$$AC_{\text{mean}} = \frac{1}{K}\sum_{i=1}^{K} A_i \quad (6)$$

$$DC_{\text{mean}} = \frac{1}{K}\sum_{i=1}^{K} x_{c,s}[t_i] \quad (7)$$

$$\frac{AC}{DC} = \frac{AC_{\text{mean}}}{DC_{\text{mean}}+\varepsilon} \quad (8)$$

$$SD_A = \sqrt{\frac{\sum_{i=1}^{K}(A_i - AC_{\text{mean}})^2}{K-1}} \quad (9)$$

The amplitude component Ai in Eq. (5) - Eq. (9). Eq. (5) calculates the beat-level amplitude component as the difference between the systolic peak and diastolic trough amplitudes. Eq. (6) computes AC_{mean} as the average pulsatile amplitude across valid beats. Eq. (7) defines DC_{mean} as the average trough or baseline-related component. Eq. (8) calculates the AC/DC ratio by normalizing the pulsatile component against the baseline component, with ε added to prevent division by zero. Eq. (9) computes SD_A , which represents the variability of beat-level amplitude components across valid beats. In these equations, K denotes the number of valid detected beats in a segment.

$$PI_i = \frac{p_{i+1} - p_i}{f_s} \quad (10)$$

$$PI_{\text{mean}} = \frac{1}{K-1}\sum_{i=1}^{K-1} PI_i \quad (11)$$

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

$$PI_{SD} = \sqrt{\frac{\sum_{i=1}^{K-1}(PI_i - PI_{mean})^2}{K-2}} \quad (12)$$

$$SRT_i = \frac{p_i - t_i}{f_s} \quad (13)$$

$$DFT_i = \frac{t_{i+1} - p_i}{f_s} \quad (14)$$

$$SRT_{mean} = \frac{1}{K}\sum_{i=1}^{K} SRT_i \quad (15)$$

$$DFT_{mean} = \frac{1}{K-1}\sum_{i=1}^{K-1} DFT_i \quad (16)$$

Temporal features were calculated from the positions of systolic peaks and diastolic troughs. In Eq. (10), PI_i denotes the peak interval of the ith beat, p_i and p_{i+1} 1 denote two adjacent systolic peak indices, and fs converts the interval from samples into seconds. In Eq. (11) and Eq. (12), PI_{mean} and PI_{SD} denote the mean and standard deviation of the peak intervals across valid beats. In Eq. (13) and Eq. (14), SRT_i denotes systolic rise time from trough t_i to peak p_i, whereas DFT_idenotes diastolic fall time from peak p_i to the next trough t_{i+1}. In Eq. (15) and Eq. (16), SRT_{mean} and DFT_{mean} represent the average systolic rise time and diastolic fall time across beats.

All extracted features were compiled into a feature table that also contained metadata, including SubjectID, SegmentID, Fold, Ground_Truth, and Anemic labels. In the 90 s dataset, each subject produced one data unit, whereas in the 30 s, 45 s, and 60 s configurations each subject could contribute multiple segments depending on the segmentation scheme. These metadata were retained to support subject-based validation and to ensure that all segments from the same subject were assigned to the same validation fold.

C. Classification and Evaluation

To evaluate the influence of multispectral channels, seven feature scenarios were defined: Green, Red, IR, Green+Red, Green+IR, Red+IR, and All. These scenarios were used to compare the contributions of single-channel, two-channel, and three-channel configurations to anemia classification performance. The evaluated classifiers were Support Vector Machine (SVM), Logistic Regression, Random Forest, and Extra Trees. SVM was implemented with a radial basis function kernel to model non-linear decision boundaries, while Logistic Regression served as a linear baseline model. Random Forest and Extra Trees were included as tree-based ensemble models to provide a comparison with non-linear ensemble classifiers. Feature normalization using StandardScaler was applied to SVM and Logistic Regression, whereas the tree-based models were trained without additional scaling.

To address class imbalance, all classifiers were trained using class-weighted learning with class_weight="balanced". The class weight for class j was computed using Eq. (17). In this equation, w_j denotes the weight assigned to class j during model training, N denotes the total number of training samples in the current fold, C denotes the number of target classes, and N_j denotes the number of training samples belonging to class j.

$$w_j = \frac{N}{C \times N_j} \quad (17)$$

$$z_k = \frac{x_k - \mu_k}{\sigma_k} \quad (18)$$

$$K(x_i, x_j) = exp\left(-\gamma \left|\left|x_i - x_j\right|\right|^2\right) \quad (19)$$

$$f(x) = sign(\sum_i \alpha_i y_i K(x_i, x) + b) \quad (20)$$

The mean μ_k and standard deviation σ_k were estimated only from the training fold and then applied to the corresponding test fold to avoid data leakage. Eq. (18) defines z-score standardization, where x_k is the original value of feature k, μ_k is the training-fold mean of the feature k, σ_k is the corresponding training-fold standard deviation, and z_k is the standardized feature value. Eq. (19) defines the radial basis function kernel used by SVM, where $K(x_i, x)$ measures the similarity between the training vector xi and the input vector x, γ controls the kernel width, and $\left|\left|x_i - x_j\right|\right|^2$ is the squared Euclidean distance. Eq. (20) defines the SVM decision function, where αi denotes the learned support-vector coefficient, yi is the class label of the ith support vector, and b is the bias term, $sign(.)$ converts the decision value into a class prediction, and $f(x)$ denotes the predicted decision output.

These classifiers were selected to maintain a low-complexity and reproducible evaluation framework, because the main objective of this study was to examine the effects of wavelength configuration and signal duration rather than to maximize performance through extensive model optimization. Gradient boosting models were not included as primary classifiers because the dataset contained only 52 subjects, including 10 anemia subjects, which may increase the risk of overfitting and fold-dependent instability in more aggressively optimized ensemble models. Probability-calibrated classifiers were also not emphasized because reliable calibration requires sufficient minority-class samples in each validation fold, whereas the anemia class in this study was very small. Therefore, model evaluation focused on class-weighted classical classifiers, fold-wise variability, ROC-AUC, PR-AUC, and confusion-matrix-based metrics. Future work with larger datasets should

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

investigate gradient boosting, calibrated classifiers, threshold adjustment, and cost-sensitive optimization more systematically.

In the initial experiment, all channel combinations were evaluated using the four classifiers on the full 90 s recording. Based on the initial-stage results, the best-performing classifier, i.e., SVM, was carried forward to the duration analysis. To avoid data leakage, all experiments used subject-based validation. Each subject was assigned to one of five folds using Stratified Group KFold so that all segments originating from the same subject were always placed in the same fold. This procedure ensured that samples from the same subject did not appear simultaneously in the training and test sets. Model evaluation was performed using a 5-fold cross-validation scheme based on PredefinedSplit with predetermined subject folds. The same validation framework was applied consistently to the 30 s, 45 s, 60 s, and 90 s datasets, ensuring comparisons across durations were made under identical subject-based conditions. Performance was evaluated primarily using macro F1-score because it gives equal importance to the normal and anemia classes under an imbalanced class distribution. The confusion-matrix definitions and evaluation metrics were calculated using Eq. (21) - (29), which are standard diagnostic-classification metrics [33][37][39].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (21)$$

$$\text{Error rate} = \frac{FP+FN}{TP+TN+FP+FN} \quad (22)$$

$$\text{Precision}_c = \frac{TP_c}{TP_c+FP_c} \quad (23)$$

$$\text{Recall}_c = \frac{TP_c}{TP_c+FN_c} \quad (24)$$

$$F1_c = \frac{2\times\text{Precision}_c\times\text{Recall}_c}{\text{Precision}_c+\text{Recall}_c} \quad (25)$$

$$\text{Macro F1} = \frac{F1_{\text{Normal}}+F1_{\text{Anemia}}}{2} \quad (26)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (27)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (28)$$

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity}+\text{Specificity}}{2} \quad (29)$$

The evaluation metrics were defined in Eq. (21) - Eq. (29). Eq. (21) defines accuracy as the proportion of correctly classified samples, while Eq. (22) defines the error rate as the proportion of incorrect predictions. Eq. (23) defines class-wise precision, which measures the proportion of predicted class-c samples that are correct. Eq. (24) defines class-wise recall, which measures the proportion of actual class-c samples that are correctly detected. Eq. (25) defines class-wise F1-score as the harmonic mean of precision and recall. Eq. (26) defines macro F1-score as the average of the F1-scores for the normal and anemia classes. Eq. (27) defines sensitivity, which corresponds to anemia recall in this study. Eq. (28) defines specificity, which measures the correct identification of normal subjects. Eq. (29) defines balanced accuracy as the average of sensitivity and specificity. In these equations, TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

The complete subject-based evaluation procedure is summarized in Algorithm 1. This algorithm describes the sequence from duration-based segmentation, peak and trough detection, feature extraction, channel-set construction, subject-based fold assignment, class-weighted classifier training, and final performance evaluation.

Algorithm 1. Subject-based wavelength and duration evaluation procedure.

Input:

PPG recordings, Hb labels, subject IDs, durations, channel sets, and classifiers

Output:

Fold-wise metrics, pooled metrics, and confusion matrices

1. Segment each 90 s recording into 30, 45, 60, or 90 s windows.
2. Detect peaks and troughs with fixed distance and prominence.
3. Extract amplitude and temporal features using Eq. (5) - (16).
4. Build seven channel sets: Green, Red, IR, G+R, G+IR, R+IR, and All.
5. Assign all windows from one subject to the same validation fold.
6. Train each classifier with class weighting using Eq. (17).
7. Test on the held-out subject fold.
8. Compute Eq. (21) - (29), ROC-AUC, PR-AUC, and confusion matrix.
9. Compare channels, durations, baselines, and fold-wise variability.

In addition to macro F1-score, clinically relevant metrics were reported, including anemia sensitivity/recall, specificity, balanced accuracy, ROC-AUC, PR-AUC, and confusion matrices. Sensitivity was defined as the recall for the anemia class, while specificity was defined as the ability of the model to correctly identify normal subjects. Balanced accuracy was used to summarize the average performance across both classes. ROC-AUC and PR-AUC were

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

included to evaluate the ranking performance of the classifier, with PR-AUC considered particularly relevant because the anemia class was small and clinically important.

All segments from the same subject were assigned to the same fold to prevent subject leakage. Although StratifiedGroupKFold was used, the small number of anemia subjects led to unequal minority-class distributions across folds, with one to three anemia subjects per test fold. The subject-based fold composition used in the cross-validation procedure is presented in Table 1. This table is provided to clarify the distribution of anemia subjects across folds, because the limited number of anemia cases may affect fold-wise performance stability. The resulting performance values were compared across classifiers, channel combinations, and signal durations to examine how wavelength selection and recording duration jointly affect classification performance, and whether a lower-complexity configuration can match or exceed a higher-complexity one. The interpretation of model performance emphasized anemia-class recall and fold-wise variability, because high overall performance may still be clinically limited if the model fails to identify a substantial proportion of anemia cases.

Table 1. Subject-based fold composition.

Fold	Total subjects	Anemia subjects
Fold 1	11	1
Fold 2	10	3
Fold 3	10	1
Fold 4	10	2
Fold 5	11	3

III. Results

A. Channel Configuration Comparison

A total of 52 subjects were successfully processed during feature extraction, yielding 24 PPG features across three wavelengths together with the metadata required for analysis. The class distribution consisted of 42 normal subjects (80.8%) and 10 anemic subjects (19.2%), with a mean hemoglobin level of 14.17 ± 2.62 g/dL and a mean age of 23.58 ± 5.09 years. Table 2 summarizes the participant demographics, class distribution, and data-quality indicators used in this study.

The initial experiment compared seven PPG signal configurations, consisting of three single-channel configurations (Green, Red, and IR), three dual-channel configurations (Green+Red, Green+IR, and Red+IR), and one three-channel configuration (All: Green+Red+IR). This comparison was designed to evaluate whether anemia classification benefited more from one-channel, two-channel, or three-channel PPG input. Fig. 2 reports the five best-performing signal configurations and classifier pairs from the initial experiment.

As shown in Fig. 2, the pooled cross-validated result

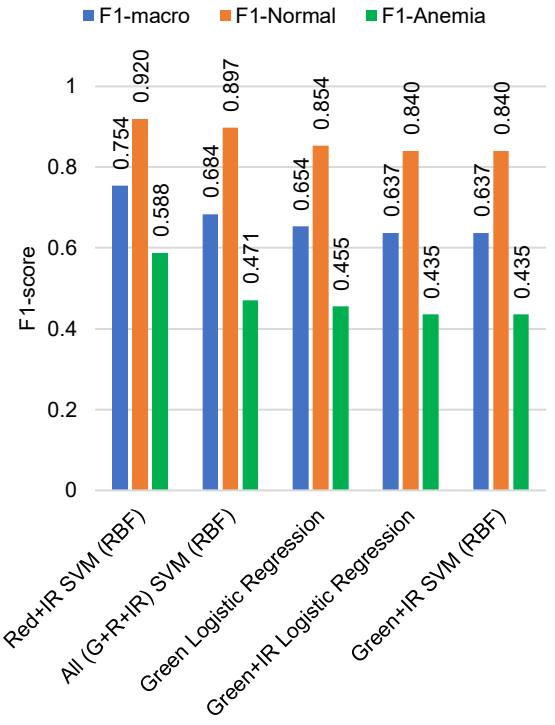

Fig. 2. Comparison of macro F1-score among the five best-performing signal combination and classifier configurations in the initial anemia classification experiment.

indicated that the Red+IR configuration with SVM achieved the highest performance at 90 s, with a macro F1-score of 0.754. This result was mainly supported by strong normal-class performance, with F1-Normal of 0.920, while anemia-class performance remained lower, with F1-Anemia of 0.588, Recall-Anemia of 0.500, and Precision-Anemia of 0.714. The pooled confusion matrix showed 35 true negatives, 7 false positives, 5 false negatives, and 5 true positives, corresponding to an accuracy of 0.769 and an error rate of 0.231. The second-best configuration was All (Green+Red+IR) with SVM, which achieved a macro F1-score of 0.684.

Table 2 shows that the dataset was highly imbalanced and demographically uneven, while Fig. 2 indicates that adding the green channel to Red+IR did

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

not improve performance. The All configuration achieved a lower macro F1-score than Red+IR, suggesting that the green wavelength provided limited additional discriminative value at this feature-extraction level. These results indicate that wavelength selection may be more important than wavelength count, and that a dual-channel configuration can offer a better accuracy-complexity trade-off than using all available PPG channels.

Table 2. Summary of participant demographics and data quality metrics used in this study.

Parameter	Value
Total subjects	52
Normal subjects	42 (80.8%)
Anemic subjects	10 (19.2%)
Female subjects	38
Male subjects	14
Mean hemoglobin (all subjects)	14.17 ± 2.62 g/dL
Mean hemoglobin (female)	13.27 ± 2.15 g/dL
Mean hemoglobin (male)	16.61 ± 2.22 g/dL
Mean age	23.58 ± 5.09 years
Mean age (normal)	22.3 years
Mean age (anemic)	28.8 years
Mann-Whitney U test (age)	p = 0.0006
Missing values	None
Infinite values	None

Although Red+IR achieved the highest pooled macro F1-score, the fold-wise analysis showed substantial variability. In particular, anemia sensitivity varied strongly across folds, reflecting instability due to the limited number of anemia subjects. For the Red+IR configuration at 90 s, the fold-wise macro F1-score was 0.617 ± 0.251, sensitivity was 0.467 ± 0.506, specificity was 0.846 ± 0.144, balanced accuracy was 0.656 ± 0.277, ROC-AUC was 0.864 ± 0.150, and PR-AUC was 0.694 ± 0.344. The total confusion matrix across folds consisted of 35 true negatives, 7 false positives, 5 false negatives, and 5 true positives. These results show that the model still missed 5 of 10 subjects with anemia; therefore, the performance should be interpreted as preliminary rather than clinically sufficient for screening.

B. Performance Across Signal Durations

Signal duration affected performance unevenly across configurations, as shown in Table 3. Red+IR improved monotonically across all four durations, increasing from a macro F1-score of 0.543 at 30 s to 0.754 at 90 s. This represented a gain of 0.211 points, the largest improvement among all evaluated configurations. A similar but smaller trend was observed for the All configuration, where macro F1-score increased from 0.546 at 30 s to 0.684 at 90 s. In contrast, the single-channel configurations showed smaller and less consistent changes across durations. Table 4 reports the corresponding recall values for the anemia class.

Table 3. Macro F1-score of the SVM model across signal combinations and recording durations.

Combination	30 s	45 s	60 s	90 s
Green	0.458	0.456	0.546	0.538
Red	0.534	0.525	0.511	0.512
IR	0.553	0.529	0.590	0.563
Green+Red	0.497	0.510	0.530	0.612
Green+IR	0.560	0.567	0.596	0.637
Red+IR	0.543	0.590	0.612	0.754
All	0.546	0.483	0.597	0.684

The macro F1-score results across signal combinations and recording durations are summarized in Table 3, while the corresponding anemia-class recall values are presented in Table 4. These two tables are interpreted together because macro F1-score reflects overall class-balanced performance, whereas anemia recall reflects the clinically important minority class. Among the single-channel configurations, IR produced the highest anemia recall at 90 s, reaching 0.600. This suggests that the infrared channel carries more anemia-

Table 4. Recall of the anemia class across signal combinations and recording durations.

Combination	30 s	45 s	60 s	90 s
Green	0.300	0.300	0.500	0.400
Red	0.400	0.400	0.400	0.300
IR	0.467	0.450	0.500	0.600
Green+Red	0.200	0.250	0.200	0.400
Green+IR	0.267	0.300	0.400	0.500
Red+IR	0.300	0.350	0.400	0.500
All	0.167	0.100	0.300	0.400

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

relevant information than either the green or the red channel alone. However, the Red+IR configuration provided the most favorable trade-off between overall macro F1-score and anemia-class performance. Although its anemia recall was 0.500, it achieved substantially higher overall classification performance than IR alone.

Among the dual-channel configurations, Green+IR showed relatively consistent performance across durations, while Green+Red improved gradually but remained inferior to Red+IR at 90 s. Overall, these findings suggest that wavelength configuration and signal duration are coupled design variables. The benefit of longer recordings was most pronounced for Red+IR, while single-channel configurations showed comparatively limited improvement as recording duration increased. However, statistical testing showed that the observed duration-related improvement should be interpreted with caution. Wilcoxon signed-rank testing across folds did not show statistically significant differences between shorter durations and 90 s for Red+IR macro F1-score. The p-values were 0.3125 for 30 s versus 90 s, 0.6250 for 45 s versus 90 s, and 0.7500 for 60 s versus 90 s. Therefore, the observed improvement with longer duration should be interpreted as a descriptive trend rather than a statistically confirmed effect.

C. Baseline Comparison and Additional Validation

To determine whether the proposed model provided meaningful performance beyond simple decision rules, the Red+IR SVM model at 90 s was compared with two baseline approaches: majority-class prediction and stratified random prediction. Table 5 reports this baseline comparison using macro F1-score, F1-Anemia, sensitivity, specificity, balanced accuracy, and PR-AUC.

As shown in Table 5, compared with the majority-class and stratified random baselines, the proposed Red+IR SVM model improved macro F1-score and anemia-class detection. The Red+IR SVM achieved a macro F1-score of 0.617 ± 0.251, compared with 0.445 ± 0.029 for the majority-class baseline and 0.431 ± 0.107 for the stratified random baseline. The improvement was particularly clear in PR-AUC, where Red+IR SVM achieved 0.694 ± 0.344, compared with 0.193 ± 0.096 and 0.208 ± 0.116 for the two baselines. However, the remaining false negatives indicate that the model still misses a substantial proportion of anemia cases. Taken together, the results indicate that Red+IR at 90 s produced the best pooled performance among the evaluated configurations. Nevertheless, the high fold-wise variability, limited anemia recall, and non-significant duration differences indicate that the findings remain preliminary and should be validated using larger and more demographically balanced datasets.

Table 5. Additional validation and baseline comparison for the 90 s Red+IR SVM model.

Model	Macro F1-score	F1-Anemia	Sensitivity	Specificity	Balanced Accuracy	PR-AUC
Majority-class baseline	0.445 ± 0.029	0.000	0.000	1.000	0.500	0.193 ± 0.096
Stratified random baseline	0.431 ± 0.107	0.080 ± 0.179	0.067 ± 0.149	0.783 ± 0.060	0.425 ± 0.101	0.208 ± 0.116
Red+IR SVM	0.617 ± 0.251	0.380 ± 0.415	0.467 ± 0.506	0.846 ± 0.144	0.656 ± 0.277	0.694 ± 0.344

IV. Discussion

A. Model Performance

The results show that the Red+IR configuration with class-weighted SVM produced the strongest pooled performance among the evaluated configurations. This configuration achieved a macro F1-score of 0.754, higher than the All configuration with SVM (0.684). The difference of 0.070 indicates that adding the green channel to the Red+IR pair did not improve the classification result in this dataset. This finding supports the interpretation that wavelength selection was more important than wavelength count for the evaluated handcrafted PPG features. The performance of the best model was mainly supported by strong normal-class classification. Red+IR SVM achieved F1-Normal of 0.920, while F1-Anemia was lower at 0.588. This gap indicates that the model was more reliable in identifying normal subjects than anemic subjects. The pooled confusion matrix further supports this interpretation, with 35 true negatives, 7 false positives, 5 false negatives, and 5 true positives. Therefore, although the pooled accuracy reached 0.769, the anemia recall was only 0.500, meaning that the model correctly detected only 5 of 10 anemia cases. From a screening perspective, the anemia recall rate of 0.500 remains limited. In anemia screening, false negatives are clinically important because missed cases of anemia may delay further examination and

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

intervention. Therefore, the best-performing Red+IR SVM model should not be interpreted as clinically ready for screening. Instead, it should be interpreted as a preliminary finding that Red+IR may be a promising wavelength configuration for further investigation. The fold-wise results also show that the model performance was unstable. For Red+IR SVM at 90 s, the fold-wise macro F1-score was 0.617 ± 0.251, sensitivity was 0.467 ± 0.506, specificity was 0.846 ± 0.144, balanced accuracy was 0.656 ± 0.277, ROC-AUC was 0.864 ± 0.150, and PR-AUC was 0.694 ± 0.344. The large standard deviation of sensitivity indicates that anemia detection varied substantially across validation folds. This instability is consistent with the small number of anemia subjects, where each fold contained only one to three anemia subjects. Compared with the baseline models, Red+IR SVM showed meaningful improvement. The macro F1-score of Red+IR SVM was 0.617 ± 0.251, compared with 0.445 ± 0.029 for the majority-class baseline and 0.431 ± 0.107 for the stratified random baseline. The improvement was more pronounced in PR-AUC, where Red+IR SVM achieved 0.694 ± 0.344, while the majority-class and stratified random baselines achieved only 0.193 ± 0.096 and 0.208 ± 0.116, respectively. This indicates that the proposed model learned discriminative information beyond simple class-prior prediction. However, the remaining false negatives indicate that this improvement was insufficient for reliable clinical screening.

B. Signal Configuration and Duration

The comparison across wavelength configurations indicates that more channels did not automatically produce better performance. The best pooled result was obtained using Red+IR, not the full Green+Red+IR configuration. Red+IR achieved a macro F1-score of 0.754 at 90 s, while the All configuration achieved 0.684. This means that the three-channel configuration was 0.070 lower despite using more wavelength information. Therefore, the additional green channel may have introduced redundant or less discriminative information under the current feature-extraction approach. Among the single-channel configurations, IR showed the highest anemia recall at 90 s, reaching 0.600. However, its macro F1-score was only 0.563, lower than that of Red+IR. This suggests that IR alone may capture anemia-related information, but its overall class balance was weaker than the Red+IR combination. In contrast, Red+IR achieved a better overall trade-off, with macro F1-score of 0.754 and anemia recall of 0.500. This result suggests that the complementary information from red and infrared wavelengths was more useful than using either channel alone.

The duration analysis also showed that recording length influenced the classification result, but the effect was configuration-dependent. For Red+IR, macro F1-score increased from 0.543 at 30 s to 0.590 at 45 s, 0.612 at 60 s, and 0.754 at 90 s. This increase of 0.211 from 30 s to 90 s was the largest improvement among the evaluated configurations. The All configuration also improved from 0.546 at 30 s to 0.684 at 90 s, but the improvement was smaller than that of Red+IR. These findings suggest that longer recordings may provide more stable morphological and temporal PPG features, especially when the selected wavelengths are complementary. However, the duration-related improvement should be interpreted cautiously. The Wilcoxon signed-rank test did not show a statistically significant difference between shorter durations and 90 s for the Red+IR macro F1-score. The p-values were 0.3125 for 30 s versus 90 s, 0.6250 for 45 s versus 90 s, and 0.7500 for 60 s versus 90 s. Therefore, the observed improvement from 30 s to 90 s should be interpreted as a descriptive trend rather than a statistically confirmed effect. This is important because the small sample size and the limited number of subjects with anemia reduce the statistical power of the comparison. Overall, the signal-configuration results indicate that Red+IR may provide a more favorable accuracy-complexity trade-off than the full three-channel configuration. From a low-complexity device perspective, this finding is relevant because using two wavelengths may reduce hardware complexity compared with using three wavelengths. Nevertheless, because the anemia recall remained limited and the fold-wise variability was high, this result should be treated as preliminary evidence for wavelength selection rather than as final device-level validation.

The comparison with previous studies serves as contextual positioning rather than direct ranking, since the studies differ in modality, task, validation strategy, dataset size, and metrics. Some estimate hemoglobin through regression, whereas this study evaluates binary anemia classification using signal-only handcrafted PPG features. Compared with image-based anemia screening, the present Red+IR SVM model is more conservative in performance. Dimauro et al. [19] reported higher sensitivity and specificity using conjunctiva-related image features, and Saleh et al. [30] reported substantially higher classification accuracy when PPG signals were combined with clinical data. In contrast, the present study intentionally excluded demographic and clinical predictors and focused on a smaller signal-only setting; therefore, its lower anemia recall highlights both the difficulty of the task and the preliminary nature of the findings.

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

Table 6. Comparison with previous non-invasive anemia and hemoglobin-related studies.

Study	Modality / Method	Main Task	Reported Performance	Comparison with This Study
Dimauro et al. [19]	Conjunctiva image-based machine learning	Anemia classification	Sensitivity = 79%, specificity = 74%	Dimauro et al. reported higher anemia sensitivity than the present work, but used image-based pallor features rather than PPG waveform features.
Chen et al. [13]	Four-wavelength PPG with feature selection and XGBoost	Hemoglobin estimation	R^2 = 0.997, RMSE = 0.762 g/L, MAE = 0.325 g/L	Chen et al. focused on Hb regression, while the present study focused on binary anemia classification and wavelength-configuration comparison.
Zhu et al. [21]	Multispectral PPG device with AdaBoost	Hemoglobin estimation	MAE = 2.67 g/L, R^2 = 0.91	Zhu et al. demonstrated the feasibility of multispectral PPG for Hb estimation, whereas the present study evaluated whether fewer wavelength channels could still support anemia classification.
Peng et al. [22]	Eight-wavelength PPG with ensemble extreme learning machine	Hemoglobin estimation	RMSE = 1.72 g/dL, PCC = 0.76	Peng et al. used a higher-complexity eight-wavelength setup, while the present study evaluated a lower-complexity three-wavelength PPG dataset.
Ni et al. [29]	Red and infrared PPG features with machine learning	Hemoglobin estimation	RMSE ≈ 0.321 g/dL, R^2 = 0.967, MRE = 2.46%	Ni et al. also highlighted the relevance of Red+IR information, supporting the finding that red and infrared wavelengths may contain useful hemoglobin-related signal characteristics.
Saleh et al. [30]	PPG signal and clinical data with machine learning	Anemia classification	Accuracy = 100%	Saleh et al. reported very high anemia-classification performance using PPG and clinical data, while the present study used signal-only handcrafted PPG features and obtained more limited anemia recall.
Mulia Rahmah et al. / This study	Triple-wavelength PPG features with class-weighted SVM	Anemia classification and wavelength-configuration comparison	Macro F1-score = 0.754, F1-Anemia = 0.588, anemia recall = 0.500, accuracy = 0.769, error rate = 0.231, ROC-AUC = 0.864 ± 0.150, PR-AUC = 0.694 ± 0.344	The present study showed that Red+IR produced the best performance among the evaluated configurations, but the model still missed 5 of 10 anemia cases; therefore, the result should be interpreted as preliminary rather than clinically ready.

A different pattern appears when the present results are viewed beside hemoglobin-estimation studies. Chen et al. [13], Zhu et al. [21], and Peng et al. [22] reported strong regression performance using four-wavelength, multispectral, or eight-wavelength optical configurations with optimized machine-learning

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

models. Those studies support the general value of optical PPG information for hemoglobin-related assessment, but they do not answer the same question addressed here: whether adding wavelength complexity improves binary anemia classification under a low-complexity, subject-based validation design. The closest methodological direction is the use of red and infrared information for hemoglobin-related analysis. Ni et al. [29] emphasized the use of red and infrared PPG features for non-invasive hemoglobin estimation, which is consistent with the present observation that Red+IR produced the best pooled macro F1-score among the evaluated channel settings. Even so, the present model still missed half of the anemia cases, so the Red+IR finding should be interpreted as a promising wavelength-selection result rather than as evidence of clinical screening readiness. Table 6 summarizes related non-invasive anemia detection and hemoglobin estimation studies. The comparison positions the present work as a preliminary, signal-only, subject-based validation study whose main contribution is the evaluation of wavelength-configuration complexity rather than the claim of superior diagnostic performance.

This study has several limitations, including a small sample size. First, the anemia class contained only 10 subjects, while the normal class contained 42 subjects. This imbalance limited minority-class learning and contributed to high fold-wise variability. The sensitivity standard deviation of 0.506 indicates that anemia detection was highly dependent on fold composition. Second, the dataset was demographically imbalanced. All anemia subjects were female and, on average, older than the normal group, with a mean age of 28.8 years in the anemia group and 22.3 years in the normal group. Although age and sex were excluded from the predictor set, demographic confounding cannot be ruled out, as they may still influence PPG morphology indirectly.

Third, the signal-quality procedure was limited. The analysis used the filtered version of the public dataset, but no additional motion-artifact rejection, independent signal-quality index, or advanced baseline correction was applied. Signal quality was primarily assessed by inspecting missing and infinite values in the extracted feature table. Fourth, the validation was internal and subject-based, but no external dataset was used. Therefore, generalizability across different populations, devices, acquisition protocols, and measurement environments remains unknown. Finally, the proposed model should not be interpreted as a direct hemoglobin-estimation method or a clinically validated screening tool. The best model still missed 5 of 10 anemia cases, and the duration-related improvement was not statistically significant. Future work should validate the Red+IR configuration using larger, externally collected, and demographically balanced datasets. Stronger signal-quality control, artifact handling, threshold optimization, imbalance-aware learning, and external clinical validation are required before this approach can be considered for practical screening.

V. Conclusion

This preliminary subject-based validation study evaluated the effects of PPG wavelength configuration and recording duration on the performance of anemia classification. Seven signal configurations were compared, including Green, Red, IR, Green+Red, Green+IR, Red+IR, and All, across 30, 45, 60, and 90 s recordings. The Red+IR configuration with class-weighted SVM at 90 s achieved the highest pooled performance in this dataset, with a macro F1-score of 0.754, F1-Anemia of 0.588, anemia recall of 0.500, anemia precision of 0.714, pooled accuracy of 0.769, error rate of 0.231, ROC-AUC of 0.864 ± 0.150, and PR-AUC of 0.694 ± 0.344. These results suggest that a carefully selected two-channel configuration may provide a better accuracy-complexity trade-off than a full three-channel configuration; however, the model still missed 5 of 10 anemia cases, so the findings should be interpreted as preliminary evidence for wavelength-configuration selection rather than clinical screening readiness. The duration analysis showed a descriptive improvement from 30 s to 90 s for Red+IR, but statistical testing did not confirm a significant difference across durations. Future work should validate the Red+IR configuration using larger, externally collected, and demographically balanced datasets, with stronger signal-quality control, artifact handling, imbalance-aware optimization, and external clinical validation before practical screening use is considered.

Acknowledgment

The authors thank the Department of Computer Science, Lambung Mangkurat University, for the academic support and facilities that supported this research.

Funding

This research received no specific grant from any public, commercial, or not-for-profit funding agency.

Data Availability

The source dataset used in this study is described in [37] and is publicly available through Mendeley Data:

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

https://data.mendeley.com/datasets/8xbbck45tt/1. The feature dataset generated during this study is available at: https://drive.google.com/drive/folders/1hKe5SBSQnRfssnkLpjrqWejdvJIkfS3-?usp=sharing.

Code Availability

The Python scripts and notebooks used for feature extraction, data processing, classification experiments, and performance evaluation in this study are available in the same repository as the generated feature dataset, as provided in the Data Availability section.

Author Contribution

MR contributed to the conceptualization of the study, data processing, experimentation, analysis and interpretation of the results, and drafting of the manuscript. FI contributed through supervision, methodological validation, critical review of the study, and editing of the manuscript. RH contributed to the interpretation of the results, scientific review, and refinement of the manuscript. RAN and IB contributed to the discussion of the results, critical review, and revision of the manuscript. All authors reviewed and approved the final version of the manuscript and agreed to be responsible for all aspects of the work, ensuring its integrity and accuracy.

Declarations

Ethical Approval

This study did not involve new recruitment, new biological sample collection, or direct interaction with human participants. The analysis was conducted as a secondary analysis of a publicly available and anonymized PPG-hemoglobin dataset. Therefore, no additional ethical approval was required for the present study.

The original data article reported that informed consent was obtained from all participants, that the data were anonymized, that no personal identifiers, such as phone numbers or email addresses, were collected, and that participants were allowed to withdraw at any point during the study. The original data collection was approved by the Faculty of Nursing Ethics Committee, Universitas Indonesia, with Institutional Review Board number KET-090/UN2.F12.D1.2.1/PPM.00.02/2023.

Consent for Publication

All authors have read and approved the final version of the manuscript and agreed to its publication.

Competing Interests

The authors declare that they have no competing interests.

Use of AI Tools

AI-assisted tools were used only to support language editing, grammar checking, and manuscript clarity. The authors reviewed, verified, and approved all AI-assisted edits, and the scientific content, analysis, results, and conclusions remain the authors' responsibility.

References

[1] G. A. Stevens *et al.*, "National, regional, and global estimates of anaemia by severity in women and children for 2000-19: a pooled analysis of population-representative data," *Lancet Glob. Health*, vol. 10, no. 5, pp. e627-e639, 2022, doi: 10.1016/S2214-109X(22)00084-5.

[2] S. Safiri *et al.*, "Burden of anemia and its underlying causes in 204 countries and territories, 1990-2019: results from the Global Burden of Disease Study 2019," *J. Hematol. Oncol.*, vol. 14, no. 1, p. 185, Nov. 2021, doi: 10.1186/s13045-021-01202-2.

[3] S. Y. Hess, A. Owais, M. E. D. Jefferds, M. F. Young, A. Cahill, and L. M. Rogers, "Accelerating action to reduce anemia: Review of causes and risk factors and related data needs," *Ann. N. Y. Acad. Sci.*, vol. 1523, no. 1, pp. 11-23, May 2023, doi: 10.1111/nyas.14985.

[4] M. N. Garcia-Casal, O. Dary, M. E. Jefferds, and S. Pasricha, "Diagnosing anemia: Challenges selecting methods, addressing underlying causes, and implementing actions at the public health level," *Ann. N. Y. Acad. Sci.*, vol. 1524, no. 1, pp. 37-50, Jun. 2023, doi: 10.1111/nyas.14996.

[5] V. S. Reddy *et al.*, "Comparison of hemoglobin measurements from venous and capillary blood from the same individual using HemoCue 301 and automated hematology analyzer in a cross-sectional community-based study in India," *Am. J. Clin. Nutr.*, vol. 123, no. 1, p. 101119, Jan. 2026, doi: 10.1016/j.ajcnut.2025.11.009.

[6] J. Park, H. S. Seok, S.-S. Kim, and H. Shin, "Photoplethysmogram Analysis and Applications: An Integrative Review," *Front. Physiol.*, vol. 12, p. 808451, Mar. 2022, doi: 10.3389/fphys.2021.808451.

[7] K. B. Kim and H. J. Baek, "Photoplethysmography in Wearable Devices: A Comprehensive Review of Technological Advances, Current Challenges, and Future

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

Directions," *Electronics (Basel).*, vol. 12, no. 13, p. 2923, Jul. 2023, doi: 10.3390/electronics12132923.

[8] S. Suner *et al.*, "Prediction of anemia and estimation of hemoglobin concentration using a smartphone camera," *PLoS One*, vol. 16, no. 7, p. e0253495, Jul. 2021, doi: 10.1371/journal.pone.0253495.

[9] M. K. Hasan *et al.*, "Noninvasive Hemoglobin Level Prediction in a Mobile Phone Environment: State of the Art Review and Recommendations," *JMIR Mhealth Uhealth*, vol. 9, no. 4, p. e16806, Apr. 2021, doi: 10.2196/16806.

[10] G. Zuccotti *et al.*, "Feasibility of a Noncontact Photoplethysmography-Based Mobile App for Noninvasive Hemoglobin Monitoring: Exploratory Observational Study," *JMIR Form. Res.*, vol. 10, p. e78820, Feb. 2026, doi: 10.2196/78820.

[11] C. Pinto, J. Parab, and G. Naik, "Non-invasive hemoglobin measurement using embedded platform," *Sens. Biosensing Res.*, vol. 29, p. 100370, Aug. 2020, doi: 10.1016/j.sbsr.2020.100370.

[12] B. Yakimov, K. Buiankin, G. Denisenko, Y. Shitova, A. Shkoda, and E. Shirshin, "Diffuse reflectance spectroscopy and RGB-imaging: a comparative study of non-invasive haemoglobin assessment," *Sci. Rep.*, vol. 14, no. 1, p. 22874, Oct. 2024, doi: 10.1038/s41598-024-73084-6.

[13] Z. Chen, H. Qin, W. Ge, S. Li, and Y. Liang, "Research on a Non-Invasive Hemoglobin Measurement System Based on Four-Wavelength Photoplethysmography," *Electronics (Basel).*, vol. 12, no. 6, p. 1346, Mar. 2023, doi: 10.3390/electronics12061346.

[14] L. Chen *et al.*, "A Four-Wavelength Photoplethysmography dataset for non-invasive hemoglobin assessment," *Sci. Data*, vol. 13, no. 1, p. 564, Mar. 2026, doi: 10.1038/s41597-026-06945-6.

[15] R. Ranjith, S. Priya, A. S. Kaviya Dharshini, and J. B. Jeeva, "Non-invasive hemoglobin measurement using optical method," *Heliyon*, vol. 10, no. 15, p. e35777, Aug. 2024, doi: 10.1016/j.heliyon.2024.e35777.

[16] H. Gruwez *et al.*, "Real-world validation of smartphone-based photoplethysmography for rate and rhythm monitoring in atrial fibrillation," *Europace*, vol. 26, no. 4, p. euae065, Apr. 2024, doi: 10.1093/europace/euae065.

[17] Y. Hu, A. Hu, and S. Song, "Photoplethysmography for Assessing Microcirculation in Hypertensive Patients After Taking Antihypertensive Drugs: A Review," *J. Multidiscip. Healthc.*, vol. 17, pp. 263-274, Jan. 2024, doi: 10.2147/JMDH.S441440.

[18] M. Elgendi *et al.*, "Recommendations for evaluating photoplethysmography-based algorithms for blood pressure assessment," *Communications Medicine*, vol. 4, no. 1, p. 140, Jul. 2024, doi: 10.1038/s43856-024-00555-2.

[19] G. Dimauro, M. E. Griseta, M. G. Camporeale, F. Clemente, A. Guarini, and R. Maglietta, "An intelligent non-invasive system for automated diagnosis of anemia exploiting a novel dataset," *Artif. Intell. Med.*, vol. 136, p. 102477, Feb. 2023, doi: 10.1016/j.artmed.2022.102477.

[20] M. H. Chowdhury *et al.*, "Estimating Blood Pressure from the Photoplethysmogram Signal and Demographic Features Using Machine Learning Techniques," *Sensors*, vol. 20, no. 11, p. 3127, Jun. 2020, doi: 10.3390/s20113127.

[21] J. Zhu *et al.*, "A Non-Invasive Hemoglobin Detection Device Based on Multispectral Photoplethysmography," *Biosensors (Basel).*, vol. 14, no. 1, p. 22, Dec. 2023, doi: 10.3390/bios14010022.

[22] F. Peng, N. Zhang, C. Chen, F. Wu, and W. Wang, "Ensemble Extreme Learning Machine Method for Hemoglobin Estimation Based on PhotoPlethysmoGraphic Signals," *Sensors*, vol. 24, no. 6, p. 1736, Mar. 2024, doi: 10.3390/s24061736.

[23] V. V. Lychagov, V. M. Semenov, E. K. Volkova, D. I. Chernakov, J. Ahn, and J. Y. Kim, "Noninvasive Hemoglobin Measurements With Photoplethysmography in Wrist," *IEEE Access*, vol. 11, pp. 79636-79647, 2023, doi: 10.1109/ACCESS.2023.3300293.

[24] M. A. Almarshad, M. S. Islam, S. Al-Ahmadi, and A. S. BaHammam, "Diagnostic Features and Potential Applications of PPG Signal in Healthcare: A Systematic Review," *Healthcare*, vol. 10, no. 3, p. 547, Mar. 2022, doi: 10.3390/healthcare10030547.

[25] P. H. Charlton *et al.*, "The 2023 wearable photoplethysmography roadmap," *Physiol. Meas.*, vol. 44, no. 11, p. 111001, Nov. 2023, doi: 10.1088/1361-6579/acead2.

[26] S. Hossain, C. A. Haque, and K.-D. Kim, "Quantitative Analysis of Different Multi-Wavelength PPG Devices and Methods for Noninvasive In-Vivo Estimation of Glycated Hemoglobin," *Applied Sciences*, vol. 11, no. 15, p. 6867, Jul. 2021, doi: 10.3390/app11156867.

[27] C.-T. Hsiao, C. Tong, and G. L. Coté, "Machine Learning-Based VO2 Estimation Using a Wearable Multiwavelength

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

Photoplethysmography Device," *Biosensors (Basel).*, vol. 15, no. 4, p. 208, Mar. 2025, doi: 10.3390/bios15040208.

[28] T. Abuzairi, E. Vinia, M. A. Yudhistira, M. Rizkinia, and W. Eriska, "A dataset of hemoglobin blood value and photoplethysmography signal for machine learning-based non-invasive hemoglobin measurement," *Data Brief*, vol. 52, p. 109823, Feb. 2024, https://doi.org/10.1016/j.dib.2023.109823.

[29] B. Ni *et al.*, "An approach to machine learning-based non-invasive hemoglobin estimation using multi-wavelength PPG signal features," *Front. Physiol.*, vol. 17, p. 1637455, Apr. 2026, doi: 10.3389/fphys.2026.1637455.

[30] N. Saleh, A. M. Salaheldin, Y. Ismail, and H. M. Afify, "Classification of anemic condition based on photoplethysmography signals and clinical dataset," *Biomedical Engineering / Biomedizinische Technik*, vol. 70, no. 4, pp. 359-370, Aug. 2025, doi: 10.1515/bmt-2024-0433.

[31] L. Liu, Z. Wang, X. Zhang, Y. Zhuang, and Y. Liang, "A Novel Model for Noninvasive Haemoglobin Detection Based on Visibility Network and Clustering Network for Multi-Wavelength PPG Signals," *Algorithms*, vol. 18, no. 2, p. 75, Feb. 2025, doi: 10.3390/a18020075.

[32] R. G. Mannino *et al.*, "Smartphone app for non-invasive detection of anemia using only patient-sourced photos," *Nat. Commun.*, vol. 9, no. 1, p. 4924, Dec. 2018, doi: 10.1038/s41467-018-07262-2.

[33] C. El-Hajj and P. A. Kyriacou, "A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure," *Biomed. Signal Process. Control*, vol. 58, p. 101870, Apr. 2020, doi: 10.1016/j.bspc.2020.101870.

[34] G. Brookshire *et al.*, "Data leakage in deep learning studies of translational EEG," *Front. Neurosci.*, vol. 18, p. 1373515, May 2024, doi: 10.3389/fnins.2024.1373515.

[35] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *Patterns*, vol. 4, no. 9, p. 100804, Sep. 2023, doi: 10.1016/j.patter.2023.100804.

[36] F. Del Pup, A. Zanola, L. F. Tshimanga, A. Bertoldo, L. Finos, and M. Atzori, "The role of data partitioning on the performance of EEG-based deep learning models in supervised cross-subject analysis: A preliminary study," *Comput. Biol. Med.*, vol. 196, p. 110608, Sep. 2025, doi: 10.1016/j.compbiomed.2025.110608.

[37] T. Abuzairi and D. B. Maharani, "A hemoglobin concentration dataset derived from triple-wavelength photoplethysmography for machine learning applications," *Data Brief*, vol. 63, p. 112241, Dec. 2025, doi: 10.1016/j.dib.2025.112241.

[38] A. J. W. Mathieu *et al.*, "Advanced waveform analysis of the photoplethysmogram signal using complementary signal processing techniques for the extraction of biomarkers of cardiovascular function," *JRSM Cardiovasc. Dis.*, vol. 13, Feb. 2024, doi: 10.1177/20480040231225384.

[39] E. Mejía-Mejía and P. A. Kyriacou, "Duration of photoplethysmographic signals for the extraction of Pulse Rate Variability Indices," *Biomed. Signal Process. Control*, vol. 80, p. 104214, Feb. 2023, doi: 10.1016/j.bspc.2022.104214.

[40] N. Sviridova, T. Zhao, A. Nakano, and T. Ikeguchi, "Photoplethysmogram Recording Length: Defining Minimal Length Requirement from Dynamical Characteristics," *Sensors*, vol. 22, no. 14, p. 5154, Jul. 2022, doi: 10.3390/s22145154.

[41] S. Y. L. Tan *et al.*, "Remote Photoplethysmography Technology for Blood Pressure and Hemoglobin Level Assessment in the Preoperative Assessment Setting: Algorithm Development Study," *JMIR Form. Res.*, vol. 9, p. e60455, Jun. 2025, doi: 10.2196/60455.

Author Biography

Mulia Rahmah is an undergraduate student in the Department of Computer Science, Lambung Mangkurat University, Banjarbaru, Indonesia, enrolled since 2022. Her academic interests include data science, biomedical signal processing, physiological signal analysis, and machine learning for healthcare applications. In this study, she contributed to organizing the public dataset, preparing duration-based feature tables, performing data preprocessing, conducting feature engineering, implementing the classification experiments, and evaluating model performance using diagnostic metrics. She also supported the experimental design, subject-based validation, reproducible Python-based experiments, and manuscript revision in response to reviewer comments.

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718

Her current research focuses on developing data-driven methods to support preliminary healthcare screening and biomedical signal analysis. She may be contacted at muliarahmah130@gmail.com. Her ORCID ID is 0009-0002-2791-9589.

Fatma Indriani is a lecturer in the Department of Computer Science, Lambung Mangkurat University, Banjarbaru, Indonesia. She completed her undergraduate studies in Informatics at the Bandung Institute of Technology and later obtained a master's degree from Monash University, Australia, in 2012. She earned her doctoral degree in Bioinformatics from Kanazawa University, Japan, in 2022. Her research interests include applied data science, bioinformatics, computational analysis, and interdisciplinary applications of machine learning. In this study, she contributed to research supervision, formulation of the research direction, validation of the experimental design, interpretation of biomedical signal-processing results, and critical revision of the manuscript. She also ensured that the analysis and conclusions were presented cautiously, considering the small sample size and class imbalance. She can be contacted at f.indriani@ulm.ac.id. Her ORCID ID is 0009-0006-7180-6708.

Rudy Herteno is a lecturer at the Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University, Banjarbaru, Indonesia. He completed his undergraduate studies in Computer Science at Lambung Mangkurat University in 2011 and obtained his master's degree in Informatics from STMIK Amikom University in 2017. Before becoming a lecturer, he worked as a software developer and gained professional experience in developing software systems, particularly for local government institutions. His research interests include software engineering, software defect prediction, data-driven systems, and deep learning. In this manuscript, he contributed to the methodological review, the interpretation of machine learning results, and the refinement of the discussion on classification performance and model limitations. His expertise supported the reproducibility and clarity of the experimental workflow. He can be contacted at rudy.herteno@ulm.ac.id. His ORCID ID is 0000-0003-0637-8090.

Radityo Adi Nugroho is an assistant professor in the Department of Computer Science, Lambung Mangkurat University, Banjarbaru, Indonesia. He received his bachelor's degree in Informatics from the Islamic University of Indonesia and his master's degree in Computer Science from Gadjah Mada University. His research interests include software defect prediction, computer vision, machine learning, and applied intelligent systems. In this study, he contributed to reviewing the experimental design, discussing the implications of wavelength selection, and refining the interpretation of the results in the context of low-complexity biomedical computing. He also supported revising the manuscript to improve technical clarity, consistency, and scientific presentation. His academic background in machine learning and computer vision strengthened the discussion of computational performance and practical implementation. He can be contacted at radityo.adi@ulm.ac.id. His ORCID ID is 0000-0002-7326-7668.

Irwan Budiman is a lecturer at Lambung Mangkurat University, Banjarbaru, Indonesia. He earned his bachelor's degree in Informatics Engineering from Universitas Islam Indonesia, Yogyakarta, and completed his master's degree in Information Systems at Diponegoro University, Semarang. His research interests include data mining, human-computer interaction, applied business intelligence, information systems, and e-government system development. In this manuscript, he contributed to the critical review of the research design and evaluation framework, the discussion of practical implications, and the final revision of the article. His academic experience in data mining and information systems supported positioning this study as a preliminary, data-driven biomedical signal analysis work rather than as a clinically validated screening device. He also helped strengthen the clarity of the study's limitations and future research directions. He can be contacted at irwan.budiman@ulm.ac.id. His ORCID ID is 0000-0002-0514-7429.

Manuscript received 8 April 2026; Revised 5 June 2026; Accepted 25 June 2026; Available online 1 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1718