HEMOGLOBIN VALUE PREDICTION WITH BAYESIAN OPTIMIZATION ASSISTED MACHINE LEARNING MODELS

Koray AÇICI[1]

[1]Department of Artificial Intelligence and Data Engineering, Ankara University, Ankara, TÜRKİYE

ABSTRACT. This study presents a framework for predicting hemoglobin (Hb) levels utilizing Bayesian optimization-assisted machine learning models, incorporating both time-domain and frequency-domain features derived from photoplethysmography (PPG) signals. Hemoglobin, a crucial protein for oxygen and carbon dioxide transport in the blood, has levels that indicate various health conditions, including anemia and diseases affecting red blood cell production. Traditional methods for measuring Hb levels are invasive, posing potential risks and discomfort. To address this, a dataset comprising PPG signals, along with demographic data (gender and age), was analyzed to predict Hb levels accurately. Our models employ support vector regression (SVR), artificial neural networks (ANNs), classification and regression trees (CART), and ensembles of trees (EoT) optimized through Bayesian optimization algorithm. The results demonstrated that incorporating age and gender as features significantly improved model performance, highlighting their importance in Hb level prediction. Among the tested models, ANN provided the best results, involving normalized raw signals, feature selection, and reduction methods. The model achieved a mean squared error (MSE) of 1.508, root mean squared error (RMSE) of 1.228, and R-squared (R²) of 0.226. This study's findings contribute to the growing body of research on non-invasive hemoglobin prediction, offering a potential tool for healthcare professionals and patients for convenient and risk-free Hb level monitoring.

1. INTRODUCTION

The iron-rich protein hemoglobin (Hb), found inside red blood cells, is crucial for transporting oxygen and carbon dioxide throughout the body. Fundamentally, if Hb level is low, tissues cannot obtain necessary oxygen [1]. Hemoglobin levels play a

vital role in overall health, and abnormal levels can indicate underlying diseases. Especially, low Hb levels are a sign of anemia, a condition where the body doesn't have enough healthy red blood cells. Since iron is fundamental to produce hemoglobin, insufficient iron may lead to iron-deficiency anemia, which is the most prevalent form [2]. The shape of red blood cells is determined genetically. If an individual has a sickle cell disease, sickle-shaped red blood cells can block blood vessels, causing pain and tissue damage [3, 4]. Another genetic disorder, known as Thalassemia, affects the production of hemoglobin, resulting in lower levels than normal in the bodies of individuals [5]. It is known that some types of leukaemia can also affect the production of red blood cells, leading to anemia [6]. Healthy kidneys secrete a hormone known as erythropoietin (EPO), which aids in the stimulation of red blood cell creation. When the kidneys are damaged or in chronic kidney disease, they may not produce enough EPO and this situation leads to anemia [7].

According to the WHO guidelines anemia classification is based on hemoglobin levels [8]. Acceptable hemoglobin levels show difference for men and women. In mild anemia hemoglobin value is between 11 g/dL and 12.9 g/dL for men and between 11 g/dL and 11.9 g/dL for women. In moderate anemia, hemoglobin value is between 8 g/dL and 10.9 g/dL for both men and women. In severe anemia hemoglobin value is lower than 8 g/dL for both men and women. Therefore, the hemoglobin levels play an important role in people's lives. Especially for patients with hemoglobin related diseases, it is vital to measure hemoglobin values.

Traditionally, hemoglobin levels are measured through blood tests, which can be inconvenient. Invasive methods for measuring hemoglobin concentration (Hb) are generally safe, but there are some potential risks involved. Some low risks include the pain caused by the needle to draw blood, slight dizziness, light-headedness, and bruising at the puncture site. However, there are some serious risks associated with invasive Hb measurement, the probability of them occurring is very low. Infection, excessive bleeding, and fainting can be given as examples of potential serious risks. Machine learning offers inspiring possibilities for non-invasive hemoglobin prediction through various techniques. Leveraging the ubiquity of smartphones, researchers have developed hemoglobin prediction tools using smartphone cameras and built-in light sources. Techniques involve analyzing fingertip images or videos, focusing on color variations related to blood oxygenation [9]. Photoplethysmography (PPG) is a widely explored technique that uses light to measure blood volume changes in tissues. Machine learning algorithms can analyze features extracted from the PPG signal, such as pulse rate and amplitude, to predict hemoglobin levels [10, 11].

In this study, a machine learning framework was proposed including PPG signals. Our contribution to literature is two-fold. To the best of our knowledge, the dataset published by Abuzairi et al. was utilized in this study for the first time [12]. Second,

a Bayesian optimized method was applied in the training phases of machine learning algorithms.

The organization of the article is as follows: In Section 2, the related works are summarized. In Section 3, materials and methods including the dataset, machine learning, feature selection, feature reduction, and optimization algorithms are described. In Section 4, the experimental setup, the evaluation metrics, and the empirical results are presented. Section 5 concludes the article.

2. Related Works

There's been a growing focus on using machine learning for non-invasive hemoglobin prediction. The study by Dimauro et al. [13] proposed a non-invasive method for estimating hemoglobin (Hb) concentration based on digital images of the conjunctiva. This innovative approach aimed to assess anemia without requiring a blood sample, making it more convenient for patients and healthcare providers. Their prototype extracts essential information from colour values in acquired images of the conjunctiva. Participants were mainly recruited from Hematology Departments and a transfusion center in Italy. Each subject allowed one blood sample for laboratory Hb measurement, and simultaneously, images of their conjunctiva were acquired using the proposed device. Tests on a mix of 113 anemic and healthy individuals demonstrated a strong correlation between the device's estimated Hb value and the actual Hb value. A k-nearest neighbor (kNN) classification algorithm was employed to assess the anemic condition based on features extracted from the conjunctiva images. The study utilized the CIE L*a*b* color space for image analysis, focusing on extracting mean values of the a*, b* components, and the G value from the RGB components of the conjunctiva images. The methodology included filtering input data based on lightness (L) and RGB components to exclude pixels that were too dark or too bright, ensuring that only pixels allowing correct pallor evaluation were considered. Pearson Correlation Index between conjunctival a* mean values and measured Hb was found to be 0.726 for the full dataset indicating a strong correlation. The authors concluded that their proposed method and device could serve as an effective tool for non-invasive anemia screening and monitoring, with the potential for use both in medical settings and by patients at home.

Another study presented a method for the non-invasive diagnosis of anemia through Hb detection using a spectrophotometric system and a BP-ANN model [14]. In their study, the dataset consists of fingertip spectra from 109 volunteers, with 4 samples identified as outliers and removed, leaving 105 samples for the analysis. Samples were divided into calibration (53 samples), correction (26 samples), and prediction (26 samples) sets. A spectrophotometric system was developed, incorporating a broadband light source, grating spectrograph, and silicon photodiode

array for measuring the fingertip spectra. Principal Component Analysis (PCA) was employed to reduce the dimensionality of the collected spectra and eliminate redundant data. The principal components were then used as inputs to the BP-ANN model, with the optimal network structure having 9 input nodes (corresponding to the principal components), 11 hidden nodes, and 1 output node. The BP-ANN model was trained and validated using the calibration and correction sample sets, respectively, and tested with the prediction sample set. The correlation coefficient (CC) of the BP-ANN model established by this method was 0.94, indicating a strong correlation between the predicted and actual Hb levels. The study successfully demonstrated the feasibility of non-invasively predicting hemoglobin levels using a combination of PCA and BP-ANN, with satisfactory accuracy and robustness. However, the article did not explicitly provide metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), or the coefficient of determination (R^2) for the BP-ANN model's predictions, which are commonly used to evaluate the performance of regression models.

A study, demonstrating the feasibility of measuring hemoglobin levels noninvasively using a standard smartphone's built-in RGB camera and white LED flash, was presented by Wang et al. [15]. The study involved 32 participants, providing a dataset for evaluating the proposed hemoglobin measurement system. Hemoglobin levels were compared against measurements, taken by a device known for optical hemoglobin measurement, to validate the smartphone-based approach. The proposed system extracted features from the PPG signals, focusing on the ratio of peak to trough intensities across different wavelengths (color channels), to assess blood absorption characteristics indicative of hemoglobin levels. A linear regression model that correlates the features extracted from the PPG signals to hemoglobin levels was employed in the study. Although the document did not specifically mention feature reduction techniques, it highlighted the importance of adjusting color channel gain to balance signal contributions from each channel, effectively optimizing the feature set for regression analysis. A Pearson correlation of 0.62 with the reference device was reported, indicating a moderate positive correlation between the smartphone-based measurements and the reference hemoglobin levels. Additionally, an RMSE value of 1.27 g/dL demonstrated the typical deviation of the smartphone-based hemoglobin estimates from the reference measurements.

A comprehensive study on non-invasively predicting hemoglobin levels using PPG signals and various machine learning algorithms was contributed by Kavsaoglu et al. [16]. The dataset included data from 33 individuals. PPG signals were collected for each participant over 10 periods. Additionally, gender, height, weight, and age were added as features, which increased the total number of features to 44. Hemocue Hb-201TM device was utilized simultaneously with PPG signal collection as a reference for Hb levels. In the study Classification and Regression Trees (CART),

Least Squares Regression (LSR), Generalized Linear Regression (GLR), Multivariate Linear Regression (MVLR), Partial Least Squares Regression (PLSR), Generalized Regression Neural Network (GRNN), Multilayer Perceptrons (MLP), and Support Vector Regression (SVR) machine learning algorithms were utilized to predict Hb levels. 40 characteristic features were derived from the PPG signal, including time-domain features from the signal and its first and second derivatives. RELIEF based feature selection (RFS) and Correlation-based feature selection (CFS) were utilized to reduce feature dimensions to 10 and 11 features, respectively. As performance metrics, MAE, MSE, R^2, RMSE, Mean Absolute Percentage Error (MAPE), and Index of Agreement (IA) were taken into consideration to calculate the effectiveness of the algorithms. RFS-assisted SVR provided promising results with the lowest MSE of 0.0027. The study demonstrated that machine learning techniques could effectively predict hemoglobin levels non-invasively using PPG signals and selected characteristic features, offering a viable method for continuous, pain-free monitoring of hemoglobin levels.

Another study that exploited Artificial Neural Network (ANN) architecture to focus on developing a non-invasive method for estimating blood hemoglobin levels was presented by Hasan et al. [17]. Their study involved 75 adults, with hemoglobin levels ranging from 7.6 to 13.5 g/dL. The data collection was performed by using 10-second fingertip videos recorded with a smartphone, resulting in 300 frames per video. The participants' ages ranged from 20 to 56 years. For feature extraction, RGB pixel intensities were obtained from 100 area blocks in each frame. Then, ANN was utilized to build a prediction model for hemoglobin values. A correlation rank order of 0.93 between the predicted hemoglobin values by the model and the gold standard was noted, signifying a high level of predictive accuracy. Additionally, the dataset was divided into 2 categories for classification purposes. Finally, the proposed method demonstrated 94% sensitivity and 96% specificity performance.

El-Kenawy et al., presented a study on using machine learning techniques for estimating Hemoglobin levels and classifying Anemia based on hematological parameters [18]. Their dataset consisted of 9004 records, which were split into training (75%) and testing (25%) data. The training dataset included 6753 records, while the testing dataset had 2251 records. Z-score Normalization was applied for standardizing the data. Some parameters like gender and age were omitted due to incomplete data. ANN, LR, and Random Forest (RF) regressors were employed to estimate Hb levels. The RF model outperformed other regression models in estimating Hemoglobin levels with the lowest RMSE (0.0123) and MAE (0.0435). For anemia classification, several classifiers were tested. A hybrid classifier combining Decision Tree (DT), Naive Bayes (NB), and ANN, optimized through weighted average probabilities, obtained the best performance with an RMSE value of 0.0838 and a MAE value of 0.0159. The study demonstrated that machine learning

techniques, particularly ensemble methods like RF for regression and a hybrid model for classification, can effectively estimate Hemoglobin levels and classify anemia types.

A novel approach for estimating Hb levels non-invasively by using PPG signals captured at four different wavelengths was presented by Chen et al. [19]. Their dataset consisted of 58 volunteers, aged between 21-27, with an approximately equal male-to-female ratio. The signals were collected at a 200 Hz sampling rate for 1 minute. For the preprocessing stage, a second-order Butterworth bandpass filter was implemented to process the raw PPG signal, removing high-frequency noise and motion artifacts. 160 morphological and time-domain feature parameters from the PPG signal across four channels were extracted. To identify the most relevant features, reliefF, Chi-square Score, and Information Gain methods were employed. Three machine learning algorithms, Logistic Regression (LR), SVR, and eXtreme Gradient Boosting (XGBoost), were utilized to obtain models. The XGBoost model, utilizing the top 30 features selected via the Chi-square method, achieved the best performance with a R^2 value of 0.997, a RMSE value of 0.762, and a MAE value of 0.325. The utilization of XGBoost, in combination with carefully selected PPG signal features, represented a novel contribution to the field of non-invasive hemoglobin measurement, showcasing the potential for clinical application.

Another approach to non-invasively predict Hb concentrations by using PPG signals was contributed by Peng et al. [10]. The research included 249 volunteers, with 199 samples allocated to a training set and 50 samples to a test set. An eight-wavelength PPG signal acquisition system, alongside a reference value of Hb concentration from an automatic blood cell analyzer were utilized for data collection. 56 feature values were extracted from the PPG signals, considering both physiological and demographical (age and gender) data. A Recursive Feature Elimination (RFE) algorithm was employed to choose the most contributive features for Hb prediction. An ensemble model combining several independent Extreme Learning Machine (ELM) models was established to enhance prediction stability and accuracy. A RMSE value of 1.72 and a PCC value of 0.76, indicating a strong correlation between predicted and actual Hb values, were achieved at the end of the experiments. Additionally, the proposed model outperformed other regression models (LR, SVR, RF, and traditional ELM) in terms of RMSE and PCC. The study introduced an ensemble approach to the ELM algorithm for improved prediction accuracy and stability, showcasing potential for broader clinical application and research into non-invasive biomarker detection.

A study, incorporates deep neural semantic segmentation and convolutional neural networks (CNNs), was presented by Chen et al. [20]. The study involved images of 1065 patients undergoing surgery. Hemoglobin levels among these patients ranged from 6 to 18 moL/L. The dataset was balanced by using the SMOTE algorithm due

to the original imbalance, where most patients had normal Hb levels. Deep neural segmentation was utilized to identify the palpebral conjunctiva region from images, ensuring the focus on relevant features for Hb prediction. CNNs and an ensemble of ELM were employed to predict Hb values. The proposed model obtained an R^2 value of 0.512, indicating a strong predictive capability. The explained variance score (EVS) reached 0.535, and MAE was 1.521, demonstrating the accuracy of the prediction model. Compared to other methods like decision trees (DT), LR, and SVR, the suggested approach demonstrated enhanced performance in terms of R^2, EVS, and MAE. The research demonstrated the potential of using deep learning and image analysis for non-invasive hemoglobin level prediction. Additionally, the study highlighted the importance of causal knowledge in improving prediction accuracy and reducing the impact of pseudo-correlation noise in the images.

Kwon and Kim proposed a non-invasive method for estimating glycated hemoglobin (HbA1c) levels using PPG signals [21]. Their dataset was derived from 40 volunteers, including their PPG signals and corresponding HbA1c levels, measured invasively. Additional data such as body mass index (BMI), finger width (FW), and SpO2 levels were collected. For the experiments, a custom-developed device that measures PPG signals through both reflective and transmissive methods was utilized. 18 features were initially considered, based on physiological characteristics, signal-directed characteristics, and physical parameters. 7 key features were ultimately selected for their importance in estimating HbA1c levels, including zero-crossing rate (ZCR), power spectral density (PSD) variance, and FW. For feature selection, RFE and importance analysis were employed to identify the most contributive features for HbA1c prediction. The study utilized RF and XGBoost ensemble models for the prediction of HbA1c levels based on the extracted PPG signal features. XGBoost model showed superior performance with a PCC value of 0.957 for the reflection method including FW as a feature. For diabetes classification, XGBoost also outperformed RF, Beer–Lambert Model, and Photon-Diffusion Model. The study demonstrated that XGBoost model can provide a promising tool for diabetes management without the need for invasive blood samples.

3. MATERIAL AND METHODS

3.1. **Dataset.** The dataset utilized in this study comprises PPG signals, gender, age, and Hb value, designed for research into non-invasive hemoglobin measurement using machine learning [12]. The dataset includes 68 participants (56% female, 44% male) between the ages of 18 and 65. A total of 816 data points were collected, corresponding to 12 data points per participant. Red and infra -red light intensity values, measured by the PPG sensor in arbitrary units (a.u.), are represented as

numerical data (float). Gender corresponds to the categorical variable indicating each participant's sex. Age corresponds to the numerical variable (integer), indicating the participant's age in years. Hemoglobin (Hb) corresponds to the target variable indicating the concentration of hemoglobin in blood, measured in grams per deciliter (g/dL) as a numeric data type (float). Hb values were measured invasively. For each participant, raw PPG signals were gathered every 40 milliseconds across a span of 10 seconds. These signals were then averaged into 12 sets of red and infra-red data to create the dataset. The study acknowledges potential biases in PPG signal measurements due to vibrations, movements, and subjects' skin tones, and the dataset's generalizability to different populations. The dataset is publicly accessible for further research and development in the field of non-invasive hemoglobin measurement and is hosted on Mendeley Data.

3.2. **General Framework.** The first stage in the study is to prepare the features to feed the machine learning algorithms. The raw signals, time-domain features, and frequency-domain features are utilized to build different models. Time-domain and frequency-domain features are extracted from the raw signals. Additionally, normalization is applied to improve the model performance and have interpretable results, while feature selection and feature reduction algorithms are applied to decrease the dimension of the samples. The second stage is to train the machine learning algorithms. In this stage, the aim is to optimize the hyper-parameters for a machine learning algorithm. Cross-validation technique is applied on the training data to obtain the performance values for the related hyper-parameters then this information is utilized by the Bayesian optimization algorithm to fine-tune the hyper-parameters. The final stage includes the test phase. After building the models, they are tested with the independent test set to obtain the performance results. The general framework is given in Figure 1.
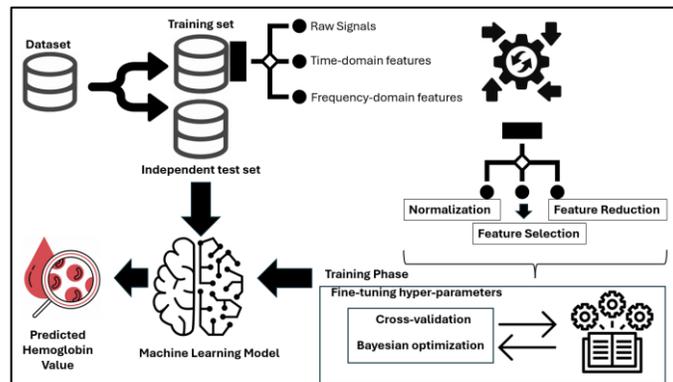

FIGURE 1. The general framework.

3.**3**. **Normalization.** Many machine learning algorithms perform better or require that the input data be normally distributed. Standardizing variables helps meet these assumptions or improve the algorithm's performance. Z-score normalization process, utilized in this study, involves transforming the original dataset so that the mean of the transformed data is 0 and the standard deviation is 1 for each feature. Z-score normalization is a valuable preprocessing step that can improve model performance. Its formula is given as follows:

$$z = \frac{(x-\mu)}{\sigma} \tag{1}$$

where, x represents the original value to be normalized, μ represents the mean of the selected feature, and σ represents standard deviation of the same feature.

3.4. **Feature Extraction.** Time-domain and frequency-domain features were extracted from the raw signals to compare them and to feed machine learning regression algorithms. Time-domain features utilized in the study are mean, root mean square (RMS), standard deviation, shape factor, signal-to-noise ratio (SNR), signal to noise and distortion ratio (SINAD), peak value, crest factor, clearance factor and impulse factor. The set of frequency-domain features include mean frequency, median frequency, bandpower, occupied bandwidth power bandwidth, peak amplitude, peak location, and power spectral density (PSD) estimate.

3.5. **Feature Selection.** In our study, rReliefF feature selection algorithm was utilized. It is used primarily to identify relevant features that contribute significantly to the prediction of the output variable. Instead of looking for nearest neighbours within the same class or different classes, rReliefF for regression searches for k nearest neighbours based on the closeness of their response values [22]. The algorithm assesses how well a feature can discriminate between instances that are near each other in the feature space but have different response values. The process has 4 main steps. First, weights of each feature are set to 0. Second, an instance is selected randomly from the dataset. Third, for the selected instance, a set of nearest neighbours is found based on the feature space. Finally, for each feature, its weight is updated based on how much the feature values for selected instance and its nearest neighbours differ, considering the differences in their response values. The intuition is that if small differences in a feature correspond to large differences in the response variable for otherwise similar instances, then the feature is important for predicting the response.

3.6. **Feature Reduction.** In our study, Principal Component Analysis (PCA) algorithm was utilized for dimensionality reduction. PCA works by identifying the

axes (principal components) that maximize the variance in a dataset [23]. These principal components are orthogonal to each other, which guarantees that they capture distinct aspects or patterns within the data. The first principal component captures the most variance, the second captures the second most, and so on, allowing for dimensionality reduction by selecting a subset of components to retain while minimizing information loss. The principal components serve as novel features that can be used in a regression model. These features represent linear combinations of the initial variables and are chosen since they explain the maximum amount of variance in the data.

3.7. **Machine Learning Regression Algorithms.**

3.7.1. *Support Vector Machines (SVM).* Support Vector Regression (SVR) extends the concept of SVM from classification to regression problems. It incorporates the core principles of SVMs to handle regression, providing a unique approach to predict continuous outcomes. Unlike traditional regression methods that aim to minimize the error between the predicted and actual outcomes, SVR focuses on ensuring that errors do not exceed a certain threshold [24]. This is achieved by fitting the best line or hyperplane within a predefined margin of tolerance, effectively capturing as many data points as possible while ignoring errors that are within the acceptable range. This guarantees that the model does not excessively react to minor fluctuations in the training data, leading to more stable and generalizable predictions.

In regression tasks, SVR is employed by choosing a type of kernel (linear, polynomial, or radial basis function) to transform the original data into a higher-dimensional space where a linear regression surface seems likely to fit better. The SVR model then focuses on minimizing the error for only those data points that fall outside the epsilon margin, ignoring errors within the margin. This approach allows the SVR to balance the intricacy of the model and the extent to which deviations exceeding epsilon are acceptable.

3.7.2. Artificial Neural Networks (ANNs). ANNs are a foundational element of machine learning and artificial intelligence, drawing inspiration from the human brain's architecture and operations. When applied to regression tasks, they're often referred to as neural network regressors. An ANN consists of interconnected processing units or nodes, called neurons. There are simply 3 different layers in an ANN. The input layer takes the features, the neurons in the hidden layers operate on the features, and the output layer produces the predicted value [25]. Every link between neurons has a corresponding weight, which is adjusted during the learning process. In a regression context, ANNs are designed to predict continuous outcomes based on input features, as opposed to classifying inputs into categories. The aim of

an ANN regressor is to learn a mapping from inputs to a continuous output, minimizing the disparity between forecasted and real values across a training dataset. This involves adjusting the weights of the connections in the network to minimize a loss function, a measure of prediction error, through a process known as backpropagation.

The model's complexity and capacity can be adjusted by varying the number of hidden layers and neurons within them, allowing ANNs to model complex, nonlinear relationships that might be difficult for other regression techniques to capture. However, they require careful tuning of hyperparameters and feature scaling, especially as model complexity increases.

3.7.3. Classification and Regression Trees (CART). Classification and Regression Trees (CART) is a decision tree learning technique that can be used for both classification and regression predictive modelling problems. The method involves splitting data into subsets based on the value of input features, leading to a tree-like model of decisions and their possible consequences [26]. The main goal of CART is to develop a model capable of predicting the value of a target variable by deriving straightforward decision rules from the features present in the data.

In regression tasks, CART involves building a decision tree to model the relationship between the features of data and a continuous target variable. The data is split at nodes based on feature values, aiming to minimize the variance of the target variable within each node. The process continues until a stopping criterion is met, like a maximum depth of the tree or a minimum number of samples in a node. The outcome is a model where each leaf node represents a prediction value based on the input features.

3.7.4. *Ensembles of Trees (EoT)*. Ensembles of trees are advanced machine learning techniques that combine multiple decision trees to create a more powerful model. These models are used for both classification and regression tasks. The core idea behind ensemble methods is to leverage the collective power of multiple models to achieve better accuracy and performance than any single model could on its own. Ensemble methods involve the integration of multiple decision trees to form a stronger predictor. Bootstrap Aggregating (Bagging) and Least Squares Boosting (LSBoost) can be given as examples for the most common ensemble methods [27,28]. LSBoost is a gradient boosting method that uses least squares loss to improve models' predictions iteratively. Bagging involves training multiple models in parallel, each on a random subset of the data (with replacement), and then aggregating their predictions. This approach is effective in reducing variance and overfitting. In EoT, a higher number of decision trees (learners) can increase the accuracy but may also lead to increased computational complexity and the risk of overfitting. The minimum leaf size in trees refers to the smallest number of observations that must be present in the leaf (terminal node) of a tree. Setting a higher minimum leaf size can help prevent overfitting by ensuring that the trees are not too deep or overly complex, which might make them sensitive to noise in the training data. By aggregating the predictions of multiple trees, ensembles can capture more complex patterns in the data, reduce the risk of overfitting, and handle variance better.

3.8. **Bayesian Optimization.** Bayesian optimization is a strategy used for optimizing objective functions that are expensive to evaluate [29]. It's particularly useful when dealing with black-box functions where the underlying mathematical form is unknown and derivatives are not available, making traditional optimization methods unsuitable. Bayesian optimization is widely used in machine learning and hyperparameter tuning where simulations or experiments are costly and time-consuming. Surrogate Model and Acquisition Function are crucial components in Bayesian optimization. They work together to efficiently find the minimum or maximum of an expensive function. Bayesian optimization builds a probabilistic model of the objective function, called the surrogate model, to approximate the true function. This model is used to make predictions about the function's behaviour and estimate the uncertainty of those predictions. Gaussian Processes (GP) are the most used surrogate models in Bayesian optimization owing to their capability to model the uncertainty of predictions. The acquisition function is used to decide where to sample next. It determines the trade-off between exploration (sampling where the model is uncertain) and exploitation (sampling where the model predicts high values). The acquisition function is chosen to be easily maximized unlike the original objective function. While Bayesian optimization can be used for regression, it does not directly target minimizing MSE or RMSE during the optimization process.

Instead, it focuses on finding the model parameters that have the highest posterior probability given the data and any prior beliefs. In other words, Bayesian optimization is a specific technique used to optimize expensive functions where the goal is to minimize or maximize the function's output. Here, the acquisition function within Bayesian optimization considers the uncertainty of the surrogate model to choose the next data point that will be most informative for finding the minimum or maximum.

One popular acquisition function is the Expected Improvement (EI), which measures the expected amount of improvement over the current best observation at a given point. The EI for a point x can be computed as follows while assuming minimization:

$$EI(x) = (\mu(x) - f(x^+) - \xi)\Phi(Z) + \sigma(x)\phi(Z) \tag{2}$$
$$Z = \begin{cases} \frac{\mu(x) - f(x^+) - \xi}{\sigma(x)}, & \sigma(x) > 0 \\ 0, & \sigma(x) \leq 0 \end{cases}$$

where, $\mu(x)$ is the mean prediction of the surrogate model at x, $f(x^+)$ is the value of the best sample observed so far, ξ is a small positive number to encourage exploration, $\sigma(x)$ is the standard deviation of the prediction at x, Φ and ϕ represent the cumulative distribution function and probability density function of the standard normal distribution, respectively. The term Z is used to calculate the expected improvement. If the predictive uncertainty at x $(\sigma(x))$ is zero, implying no uncertainty in the model's prediction at x, Z is set to 0 since the formula aims to prevent division by zero. Z plays a crucial role in quantifying how much improvement a new sample is expected to provide over the current best observation, adjusted for the level of uncertainty in the prediction at that point. This standardization allows the EI formula to balance exploration and exploitation by taking into account both the average prediction and the uncertainty of the prediction.

When evaluating the objective function is time-consuming, it's beneficial to incorporate the evaluation time into the acquisition function. The Expected Improvement Per Second Plus (EIPS) is a variant of the EI that accounts for the evaluation time, aiming to maximize the efficiency of the optimization process in terms of the improvement gained per unit of time. The EIPS acquisition function can be formulated as:

$$EIPS(x) = \frac{EI(x)}{t(x)} \tag{3}$$

where, $E(x)$ is the expected improvement at point x and $t(x)$ is the expected

evaluation time for point x. This formulation encourages selecting points that are not only expected to yield high improvement but also are quicker to evaluate, thus optimizing the efficiency of the Bayesian optimization process [30].

4. RESULTS

4.1. **Experimental Setup.** The dataset utilized in this study was split into training and independent test sets. The training set comprised 70% of the data (48 samples), while the independent test set included the remaining 30% (20 samples).

Given the relatively small size of our dataset, Leave-One-Out (LOO) cross-validation is a suitable choice for optimizing the hyperparameters of our machine learning algorithms during the training phase. In LOO strategy, each sample is used once as a validation case, while the remaining part of the training set is used to obtain a model. This process is repeated for every sample in the training set. Finally, by averaging the performances of all validation samples, hyper-parameters of a machine learning algorithm are determined. With the integration of Bayesian optimization algorithm, hyper-parameters are optimized, leading to improved model performance.

5 setups were prepared utilizing the same machine learning algorithms but with different features, in order to compare the performances in terms of evaluation metrics.

For each machine learning algorithm, the hyper-parameters were fine-tuned by utilizing Bayesian optimization. These hyper-parameters were box constraint (cost), epsilon, and kernel function (linear, Gaussian, quadratic, and cubic) regarding SVR model; number of hidden layers, size of each layer, activation function (sigmoid, rectified layer unit), and regularization strength for ANN model; minimum leaf size for CART model; ensemble method (LSBoost or Bag), number of learners, learning rate, minimum leaf size, and number of features to sample for EoT model.

4.2. **Evaluation Metrics**. Mean Absolute Error (MAE) quantifies the average size of the mistakes in a series of forecasts, disregarding their sign. It calculates the average of the absolute differences between the forecasted and the actual values. Below is the formula for MAE:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad (4)$$

where n, y_i, and $\hat{y}_i$ represent the number of observations, the actual value of the observation, and the predicted value, respectively. The lower the MAE, the better, with 0 being the ideal score.

Mean Squared Error (MSE) evaluates the average of the squared discrepancies, differences between the predicted values and the true values. The formula for MSE is provided below:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5}$$

where, n, y_i, and $\hat{y}_i$ represent the same meanings in Equation (4). It penalizes larger errors more severely than smaller ones, due to the squaring of each term. A smaller MSE signifies a closer match to the actual data, where a score of 0 represents an ideal fit.

Root Mean Squared Error (RMSE) calculates the square root of the mean of the squared deviations between the predicted values and the actual observations. It provides an indication of the dispersion of these residuals, essentially showing the degree to which the data clusters around the best fit line. Below is the formula for RMSE:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{6}$$

where, n, y_i, and $\hat{y}_i$ represent the same meanings in Equation (5), but the whole formula is under a square root. RMSE is the square root of MSE, bringing the error metric back to the same units as the target variable. It similarly penalizes larger errors more than smaller ones. smaller RMSE value suggests a more accurate model, with 0 being the ideal score. RMSE is sensitive to outliers.

R-squared (R^2) quantifies the fraction of variance in the dependent variable that can be explained by the independent variables. The formula for R^2 is as follows:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2} \tag{7}$$

where, n, y_i, and $\hat{y}_i$ represent the same meanings in Equation (4), $\bar{y}$ represents the mean of the actual values. SS_{res} is the sum of squares of residuals, which measures the variability of the prediction errors. SS_{tot} is the total sum of squares, which measures the total variability of the observed data around the mean. The nearer R^2 approaches 1, the greater the proportion of variance in the dependent variable explained by the model, signifying a stronger model fit. An elevated R^2 value does not automatically mean the model is the most effective or accurate in its predictions. In models where the predictions are worse than merely estimating the average of the observed values, R^2 can be negative.

4.3. **Empirical Results.** The results of the first, second, third, fourth, and fifth setups are given in Table 1, Table 2, Table 3, Table 4, and Table 5, respectively.

TABLE 1. Regression results for raw signals, age and gender.

| | Raw signals + Age + Gender (26 features) | | | | | | | |
| | Validation | | | | Test | | | |
Model	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
SVR	1.419	3.152	1.775	0.157	**1.016**	**1.783**	**1.335**	**0.085**
ANN	1.232	2.111	1.453	0.435	1.579	4.1	2.024	-1.01
CART	1.099	2.052	1.432	0.451	1.254	2.96	1.721	-0.518
EoT	**0.913**	**1.732**	**1.316**	**0.537**	1.17	2.529	1.591	-0.297
	RReliefF (19 features)							
SVR	1.434	2.939	1.714	0.214	**1.044**	**1.859**	**1.363**	**0.046**
ANN	**1.075**	**1.729**	**1.315**	**0.537**	1.677	4.523	2.126	-1.32
CART	1.12	2.099	1.448	0.439	1.331	2.972	1.724	-0.525
EoT	1.147	2.712	1.646	0.275	1.422	3.769	1.941	-0.933
	PCA (95% variance, 2 components)							
SVR	**1.506**	**3.320**	**1.822**	**0.112**	1.247	2.443	1.563	-0.253
ANN	1.532	3.562	1.887	0.047	**1.101**	**2.149**	**1.466**	**-0.102**
CART	1.52	3.693	1.921	0.012	1.293	3.035	1.742	-0.557
EoT	1.63	3.804	1.95	-0.017	1.225	2.799	1.673	-0.436
	RReliefF + PCA							
SVR	**1.504**	**3.314**	**1.82**	**0.113**	1.24	2.403	1.55	-0.233
ANN	1.638	4.052	2.013	-0.083	**1.131**	**2.026**	**1.423**	**-0.039**
CART	1.576	3.58	1.892	0.042	1.307	2.526	1.589	-0.296
EoT	1.576	3.739	1.933	4.4e-16	1.219	2.289	1.513	-0.174

According to the Table 1, the best performance values were obtained as 1.016, 1.783, 1.335, and 0.085 in terms of MAE, MSE, RMSE, and R^2, respectively for the independent test set by utilizing raw signals, age, and gender information as features and SVR as regressor.

TABLE 2. Regression results for normalized raw signals, age and gender.

Model	Normalized features (26 features)							
	Validation				Test			
	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
SVR	1.143	2.213	1.487	0.408	1.432	3.325	1.823	-0.705
ANN	**1.015**	**1.508**	**1.228**	**0.597**	1.718	4.338	2.082	-1.225
CART	1.099	2.052	1.432	0.451	1.254	2.96	1.72	-0.518
EoT	1.22	2.585	1.608	0.308	**1.083**	**2.314**	**1.521**	**-0.187**
	RRreliefF (19 features)							
SVR	1.132	1.991	1.411	0.467	**0.943**	**1.764**	**1.328**	**0.095**
ANN	1.092	2.016	1.42	0.461	1.136	1.806	1.344	0.073
CART	**1.054**	**1.937**	**1.392**	**0.482**	1.254	2.96	1.72	-0.518
EoT	**1.054**	2.095	1.447	0.439	1.277	2.76	1.661	-0.416
	PCA (95% variance, 3 components)							
SVR	**0.903**	**1.218**	**1.103**	**0.674**	1.706	4.605	2.146	-1.362
ANN	1.091	2.234	1.494	0.403	**1.000**	**1.556**	**1.247**	**0.202**
CART	1.122	2.063	1.436	0.448	1.116	2.178	1.475	-0.117
EoT	1.227	2.292	1.514	0.387	1.022	2.038	1.427	-0.045
	RRreliefF + PCA							
SVR	**0.96**	**1.4**	**1.184**	**0.625**	2.845	11.999	3.464	-5.155
ANN	1.042	2.152	1.467	0.424	**0.981**	**1.508**	**1.228**	**0.226**
CART	1.226	2.257	1.502	0.396	1.097	2.036	1.426	-0.044
EoT	1.255	2.508	1.583	0.329	1.072	2.163	1.471	-0.11

According to the Table 2, the best MAE value was obtained as 0.943 by utilizing SVR and feature selection. In terms of MSE, RMSE, and R^2, the best performance values were achieved as 1.508, 1.228, and 0.226, respectively for the independent test set by utilizing normalized features, ANN as regressor, and with the inclusion of the feature selection and reduction methods.

TABLE 3. Regression results for normalized raw signals and w/o age and gender.

| | Normalized features (24 features) | | | | | | | |
| | Validation | | | | Test | | | |
Model	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
SVR	1.463	3.34	1.827	0.107	**1.004**	**1.951**	**1.396**	**-0.001**
ANN	**1.295**	**2.572**	**1.603**	**0.312**	1.267	2.416	1.554	-0.239
CART	1.576	3.548	1.883	0.051	1.181	2.416	1.555	-0.239
EoT	1.677	3.817	1.953	-0.021	1.184	2.598	1.612	-0.333
	RReliefF (16 features)							
SVR	1.498	3.279	1.811	0.123	1.193	2.392	1.546	-0.227
ANN	**1.397**	**3.277**	**1.81**	**0.124**	**1.171**	**2.179**	**1.476**	**-0.117**
CART	1.575	3.547	1.883	0.051	1.181	2.416	1.554	-0.239
EoT	1.645	3.957	1.989	-0.058	1.225	2.482	1.575	-0.273
	PCA (95% variance, 2 components)							
SVR	**1.501**	**3.304**	**1.817**	**0.116**	1.236	2.397	1.548	-0.23
ANN	1.602	3.821	1.954	-0.021	1.256	2.45	1.565	-0.256
CART	1.576	3.739	1.933	0	1.258	2.39	1.546	-0.226
EoT	1.59	3.687	1.92	0.013	**1.182**	**2.233**	**1.494**	**-0.146**
	RReliefF + PCA							
SVR	**1.523**	**3.31**	**1.819**	**0.115**	1.23	2.401	1.549	-0.232
ANN	1.563	3.768	1.941	-0.007	**1.16**	**2.076**	**1.442**	**-0.066**
CART	1.602	3.547	1.883	0.051	1.361	3.143	1.772	-0.612
EoT	1.638	3.907	1.976	-0.044	1.259	2.451	1.565	-0.257

According to the Table 3, the best performance values were obtained as 1.004, 1.951, 1.396, and -0.001 in terms of MAE, MSE, RMSE, and R^2, respectively, by utilizing SVR and the normalized features without the gender and the age information for the independent test set.

TABLE 4. Regression results for normalized time-domain features.

	Normalized time-domain features (10 features)							
	Validation				Test			
Model	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
SVR	**1.434**	**2.819**	**1.679**	**0.246**	**1.183**	2.671	1.634	-0.371
ANN	1.51	3.151	1.775	0.157	1.238	**2.595**	**1.611**	**-0.331**
CART	1.519	3.233	1.798	0.135	1.199	2.608	1.615	-0.338
EoT	1.612	3.662	1.913	0.02	1.232	3.354	1.831	-0.721
	RReliefF (4 features)							
SVR	1.539	3.415	1.848	0.086	1.111	**2.089**	**1.445**	**-0.071**
ANN	**1.532**	**3.279**	**1.811**	**0.123**	**1.09**	2.124	1.457	-0.089
CART	1.587	3.492	1.868	0.066	1.199	2.608	1.615	-0.338
EoT	1.59	3.526	1.877	0.056	1.267	3.288	1.813	-0.686
	PCA (95% variance, 1 component)							
SVR	1.549	3.68	1.918	0.015	1.159	2.077	1.441	-0.065
ANN	1.521	3.671	1.916	0.018	**1.056**	1.968	1.402	-0.009
CART	**1.51**	**3.616**	**1.901**	**0.032**	1.269	2.258	1.502	-0.158
EoT	1.561	3.924	1.981	-0.049	1.145	**1.726**	**1.313**	**0.114**
	RReliefF + PCA (2 components)							
SVR	**1.358**	**3.116**	**1.765**	**0.166**	1.319	2.889	1.699	-0.482
ANN	1.575	3.572	1.89	0.044	**1.147**	**2.352**	**1.533**	**-0.206**
CART	1.562	3.61	1.9	0.034	1.291	2.723	1.65	-0.397
EoT	1.624	3.863	1.965	-0.033	1.236	2.89	1.7	-0.482

According to the Table 4, the best MAE value was obtained as 1.056 on normalized time-domain features, extracted from the raw signals, by utilizing ANN as regressor and the feature reduction method. The best performance values in terms of MSE, RMSE, and R^2 were achieved as 1.726, 1.313, and 0.114, respectively by utilizing EoT as regressor and the feature reduction method.

TABLE 5. Regression results for normalized frequency-domain features.

| Model | Normalized frequency-domain features (136 features) | | | | | | | |
| | Validation | | | | Test | | | |
	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
SVR	1.574	3.762	1.939	-0.006	1.155	**2.061**	**1.435**	**-0.057**
ANN	1.571	3.404	1.845	0.089	1.246	2.983	1.727	-0.53
CART	**1.493**	**3.368**	**1.835**	**0.099**	1.282	2.853	1.689	-0.463
EoT	1.536	3.389	1.841	0.093	**1.125**	2.195	1.481	-0.126
RReliefF (2 features)								
SVR	**1.426**	**3.107**	**1.762**	**0.168**	1.241	2.734	1.653	-0.402
ANN	1.629	3.917	1.979	-0.047	**1.205**	**2.237**	**1.495**	**-0.147**
CART	1.515	3.231	1.797	0.135	1.397	3.759	1.938	-0.928
EoT	1.519	3.278	1.81	0.123	1.227	2.689	1.639	-0.379
PCA (95% variance, 2 components)								
SVR	**1.451**	**3.212**	**1.792**	**0.14**	1.446	3.081	1.755	-0.58
ANN	1.62	3.933	1.983	-0.052	1.48	3.542	1.882	-0.817
CART	1.535	3.252	1.803	0.13	**1.037**	2.146	1.465	-0.101
EoT	1.67	3.854	1.963	-0.031	1.082	**2.113**	**1.453**	**-0.084**
RReliefF + PCA (2 components)								
SVR	1.488	3.159	1.777	0.155	1.373	3.348	1.83	-0.718
ANN	**1.441**	**2.959**	**1.72**	**0.208**	1.456	3.723	1.929	-0.91
CART	1.468	3.384	1.839	0.094	**1.251**	**2.598**	**1.611**	**-0.332**
EoT	1.598	3.636	1.907	0.027	1.265	2.993	1.73	-0.535

According to the Table 5, the best MAE value was obtained as 1.037 on normalized frequency-domain features, extracted from the raw signals, by utilizing CART as regressor and the feature reduction method. The other best performance values were achieved as 2.061, 1.435, and -0.057 in terms of MSE, RMSE, and R^2, respectively, by utilizing SVR without the inclusion of the feature selection and reduction methods.

The best performance values, according to the independent test set results, are shown in Table 6, for each experimental setup.

TABLE 6. The best regression results for 5 experimental setups.

	Setup 1: Raw signals + Age + Gender (26 features)							
	Validation				Test			
Model	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE	R^2
SVR	1.419	3.152	1.775	0.157	**1.016**	**1.783**	**1.335**	**0.085**
Setup 2: Normalized features + RReliefF (19 features)								
SVR	1.132	1.991	1.411	0.467	**0.943**	1.764	1.328	0.095
Setup 2: Normalized features + RReliefF + PCA (3 components)								
ANN	1.042	2.152	1.467	0.424	0.981	**1.508**	**1.228**	**0.226**
Setup 3: Normalized features w/o Age & Gender (24 features)								
SVR	1.463	3.34	1.827	0.107	**1.004**	**1.951**	**1.396**	**-0.001**
Setup 4: Normalized time-domain features + PCA (1 component)								
ANN	1.521	3.671	1.916	0.018	**1.056**	1.968	1.402	-0.009
EoT	1.561	3.924	1.981	-0.049	1.145	**1.726**	**1.313**	**0.114**
Setup 5: Normalized frequency-domain features + PCA (2 components)								
CART	1.535	3.252	1.803	0.13	**1.037**	2.146	1.465	-0.101
Setup 5: Normalized frequency-domain features (136 features)								
SVR	1.574	3.762	1.939	-0.006	1.155	**2.061**	**1.435**	**-0.057**

According to the Table 6, for the first experimental setup, the best performance values were obtained by SVR regressor on raw signals with the inclusion of age and gender information. For the second experimental setup, the best performance value was obtained by SVR regressor on normalized features with the inclusion of the feature selection algorithm in terms of MAE, while the best performance values were obtained by ANN regressor with the inclusion of the feature selection and reduction algorithms in terms of MSE, RMSE, and R^2. For the third experimental setup, the best performance values were obtained by SVR regressor on normalized features without the inclusion of age and gender information. For the fourth experimental setup, the best performance value was obtained by ANN regressor on normalized time-domain features with the inclusion of the feature reduction algorithm in terms of MAE, while the best performance values were obtained by EoT regressor in terms of the other evaluation metrics. For the last experimental setup, the best performance value was obtained by CART regressor on frequency-domain features with the inclusion of the feature reduction algorithm in terms of MAE, while the best performance values were obtained by SVR regressor on frequency-domain features without the inclusion of the feature selection and reduction algorithms in terms of the other evaluation metrics.

In the first setup, for the SVR model, the hyperparameters were optimized by Bayesian optimization with a cost of 0.279, epsilon value of 1.116, and a kernel function of linear. In the second setup, for the SVR model, the hyperparameters were optimized with a cost of 0.207, epsilon value of 0.364, and a kernel function of

quadratic, while for the ANN model the number of hidden layers, the number of neurons, activation function, and regularization strength were optimized as 1, 1, sigmoid, and 1.385, respectively. In the third setup, for the SVR model, the hyperparameters were optimized with a cost of 0.001, epsilon value of 0.002, and a kernel function of quadratic. In the fourth setup, for the ANN model, the number of hidden layers, the number of neurons, activation function, and regularization strength were optimized as 1, 1, rectified layer unit, and 0.397, respectively. For the EoT model, the ensemble method, the number of learners, the learning rate, the minimum leaf size, and the number of the features to sample were optimized as LSBoost, 24, 0.998, 15, and 1, respectively. In the last setup, the minimum leaf size was optimized as 19 for the CART model, while the hyperparameters were optimized with a cost of 0.001, epsilon value of 0.013, and a kernel function of Gaussian, for the SVR model.

5. CONCLUSION

According to the data obtained from the experiments, it was observed that standardization on the features is an important preprocessing step. Experiments further showed that age and gender were informative features, as the performance of the regressors dropped when these features were removed from the feature vectors. When time-domain and frequency-domain features were used to feed the regressors without including age and gender information, the results showed that the time-domain features led to better performance than the frequency-domain-features. Based on R^2 evaluation metric, normalized raw signal and age-gender information can explain the dependent variable better than other features. It can be interpreted that the utilization of time and frequency-domain features is indicative of potential information loss relative to the raw signal. If we compare the models, it can be observed that while the SVR model stood out in 4 out of 5 different setups, the highest performance values were obtained with the ANN model. ANN provided the best performance values among the setups in terms of MSE, RMSE, and R^2 after normalization, feature selection, and reduction were applied. We hope the results obtained from the first utilization of this relevant dataset will be established as a benchmark, encouraging further research, and paving the way for achieving even better results in the future.

Declaration of Competing Interests The author declares no known competing interests.

REFERENCES

[1]   Cho, H., Lee, S.-R., Baek, Y., Anemia diagnostic system based on impedance measurement of red blood cells, *Sensors*, 21 (23) (2021), 1-12, https://doi.org/10.3390/s21238043.

[2]   Yap, B. K., Soair, S. N. M., Talik, N. A., Lim, W. F., Mei, I. L., Potential point-of-care microfluidic devices to diagnose iron deficiency anemia, *Sensors*, 18 (8) (2018), 1-17, https://doi.org/10.3390/s18082625.

[3]   Mandal, A. K., Mitra, A., Das, R., Sickle cell hemoglobin, In: Hoeger, U., Harris, J. (eds), Vertebrate and invertebrate respiratory proteins, lipoproteins and other body fluid proteins, Subcellular Biochemistry, vol 94, Springer, Cham, (2020), https://doi.org/10.1007/978-3-030-41769-7_12.

[4]   Telfer, P., Carvalho, S. J., Ruzangi, J., Arici, M., Binns, M., Beaubrun, A., Montealegre-Golcher, F., Rice, C. T., Were, J. J., Association between hemoglobin levels and end-organ damage in sickle cell disease: A retrospective linked primary and secondary care database analysis in England, *Hematol. Transfus. Cell Ther.*, 44 (Supplement 2) (2022), S10-S11.

[5]   Helmi, N., Bashir, M., Shireen, A., Ahmed, I. M., Thalassemia review: features, dental considerations and management, *Electron. Physician*, 9 (3) (2017), 4003-4008, https://doi.org/10.19082/4003.

[6]   Gaspar, B. L., Sharma, P., Das, R., Anemia in malignancies: Pathogenetic and diagnostic considerations, *Hematology*, 20 (1) (2015), 18-25, https://doi.org/10.1179/1607845414Y.0000000161.

[7]   Panjeta, M., Tahirović, I., Sofić, E., Ćorić, J., Dervišević, A., Interpretation of erythropoietin and haemoglobin levels in patients with various stages of chronic kidney disease, *J. Med. Biochem.*, 36 (2) (2017), 145-152, https://doi.org/10.1515/jomb-2017-0014.

[8]   World Health Organization, Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity, (2011). Available at: https://www.who.int/publications/i/item/WHO-NMH-NHD-MNM-11.1.   [Accessed March 2024].

[9]   Hasan, M. K., Aziz, M. H., Zarif, M. I. I., Hasan, M., Hashem, M., Guha, S., Love, R. R., Ahamed, S., Noninvasive hemoglobin level prediction in a mobile phone environment: State of the art review and recommendations, *JMIR mHealth and uHealth*, 9 (4) (2021), 1-24, https://doi.org/10.2196/16806.

[10] Peng, F., Zhang, N., Chen, C., Wu, F., Wang, W., Ensemble extreme learning machine method for hemoglobin estimation based on photoplethysmographic signals, *Sensors*, 24 (6) (2024), 1-14, https://doi.org/10.3390/s24061736.

[11] Zhu, J., Sun, R., Liu, H., Wang, T., Cai, L., Chen, Z., Heng, B., A non-invasive hemoglobin detection device based on multispectral photoplethysmography, *Biosensors*, 14 (1) (2024), 1-19, https://doi.org/10.3390/bios14010022.

[12] Abuzairi, T., Vinia, E., Yudhistira, M. A., Rizkinia, M., Eriska, W., A dataset of hemoglobin blood value and photoplethysmography signal for machine learning-based non-invasive hemoglobin measurement, *Data in Brief*, 52 (2024), 1-7, https://doi.org/10.1016/j.dib.2023.109823.

[13] Dimauro, G., Caivano, D., Girardi, F., A new method and a non-invasive device to estimate anemia based on digital images of the conjunctiva, *IEEE Access*, 6 (2018), 1-8, https://doi.org/10.1109/ACCESS.2018.2867110.

[14] Ding, H., Lu, Q., Gao, H., Peng, Z., Non-invasive prediction of hemoglobin levels by principal component and back propagation artificial neural network, *Biomed. Opt. Express*, 5 (2014), 1145-1152, https://doi.org/10.1364/BOE.5.001145.

[15] Wang, E. J., Li, W., Zhu, J., Rana, R., Patel, S. N., Noninvasive hemoglobin measurement using unmodified smartphone camera and white flash, *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* Jeju, Korea (South), (2017), 2333-2336, https://doi.org/10.1109/EMBC.2017.8037323.

[16] Kavsaoğlu, A., Polat, K., Hariharan, M., Non-invasive prediction of hemoglobin level using machine learning techniques with the PPG signal's characteristic features, *Appl. Soft Comput.*, 28 (2015), 433-441, https://doi.org/10.1016/j.asoc.2015.04.008.

[17] Hasan, M. K., Haque, M. M., Adib, R., Tumpa, J. F., Begum, A., Love, R. R., Kim, Y. L., Sheikh, I. A., SmartHeLP: Smartphone-based hemoglobin level prediction using an artificial neural network, *AMIA Annu. Symp. Proc.*, 2018, 535-544.

[18] El-kenawy, E. S. M. T., A machine learning model for hemoglobin estimation and anemia classification, *IJCSIS*, 17 (2) (2019), 100-108.

[19] Chen, Z., Qin, H., Ge, W., Li, S., Liang, Y., Research on a non-invasive hemoglobin measurement system based on four-wavelength photoplethysmography, *Electronics*, 12 (6) (2023), 1-12, https://doi.org/10.3390/electronics12061346.

[20] Chen, Y., Zhong, K., Zhu, Y., Sun, Q., Two-stage hemoglobin prediction based on prior causality, *Front. Public Health*, 10, (2022), 1-12, https://doi.org/10.3389/fpubh.2022.1079389.

[21] Kwon, T.-H., Kim, K.-D., Machine-learning-based noninvasive in vivo estimation of HbA1c using photoplethysmography signals, *Sensors*, 22 (8) (2022), 1-19, https://doi.org/10.3390/s22082963.

[22] Robnik-Sikonja, M., Kononenko, I., Theoretical and empirical analysis of ReliefF and RReliefF, *Mach. Learn.*, 53 (2003), 23-69.

[23] Greenacre, M., Groenen, P. J. F., Hastie, T. et al., Principal component analysis, *Nat. Rev. Methods Primers,* 2 (2022), 100, https://doi.org/10.1038/s43586-022-00184-w.

[24] Wang, Y. G., Wu, J., Hu, Z. H., McLachlan, G. J., A new algorithm for support vector regression with automatic selection of hyperparameters, *Pattern Recognit.*, 133 (2023), 1-9, https://doi.org/10.1016/j.patcog.2022.108989.

[25] Kufel, J., Bargieł-Łączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., Czogalik, Ł., Dudek, P., Magiera, M., Lis, A., et al., What is machine learning, artificial

neural networks and deep learning? Examples of practical applications in medicine, *Diagnostics*, 13 (15) (2023), 1-22, https://doi.org/10.3390/diagnostics13152582.

[26] Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J., *Classification and Regression Trees*, Boca Raton, FL: Chapman and Hall, 1984.

[27] Breiman, L., Bagging predictors, *Mach. Learn.*, 26 (1996), 123-140.

[28] Friedman, J., Hastie, T., Tibshirani, R., Additive logistic regression: A statistical view of boosting, *Ann. Stat.*, 28 (2) (2000), 337-407.

[29] Wang, X., Jin, Y., Schmitt, S., Olhofer, M., Recent advances in Bayesian optimization, *ACM Comput. Surv.*, 55 (13s) (2023), 1-36, https://doi.org/10.1145/3582078.

[30] Jones, D. R., Schonlau, M., Welch, W. J., Efficient global optimization of expensive black-box functions, *J. Glob. Optim.*, 13 (1998), 455-492.