

The Journal of ExtraCorporeal Technology

Estimation of Hemoglobin Concentration at the Initiation of Cardiopulmonary Bypass Using Support Vector Regression --Manuscript Draft--

Manuscript Number:	ject250087R1
Full Title:	Estimation of Hemoglobin Concentration at the Initiation of Cardiopulmonary Bypass Using Support Vector Regression
Article Type:	Original Article
Keywords:	hemoglobin; support vector regression; Cardiopulmonary Bypass; machine learning; hemodilution
Corresponding Author:	Makoto Hibiya, Ph.D. Fujita Health University: Fujita Ika Daigaku Toyoake, Aichi JAPAN
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Fujita Health University: Fujita Ika Daigaku
Corresponding Author's Secondary Institution:	
First Author:	Harutoyo Hirano, PhD
First Author Secondary Information:	
Order of Authors:	Harutoyo Hirano, PhD Shun Takahashi, Bachelor Tetsuya Kamei, PhD Makoto Hibiya, Ph.D.
Order of Authors Secondary Information:	
Abstract:	<p>Background</p> <p>Hemodilution during cardiopulmonary bypass (CPB) is a standard perfusion strategy used to reduce blood viscosity and enhance microcirculatory flow. The hemodilution rate, expressed as hemoglobin (Hb) concentration, is a key control index in CPB and is currently estimated from total blood volume (TBV). The objective of this study was to propose a novel formula to accurately predict Hb concentration at the initiation of CPB (HbCPB), by incorporating circulating blood volume, laboratory data, physical measurements, and patient history.</p> <p>Methods</p> <p>We retrospectively analyzed 577 adult patients who underwent elective CPB at Fujita Health University Hospital from January 2016 to December 2020. Thirty-six preoperative variables—including demographics, laboratory data, circuit parameters, and indices such as TBV and ideal weight—were standardized. Categorical variables underwent one-hot encoding. We compared generalized linear models (GLM), support vector regression (SVR), and multilayer perceptron (MLP). Model performance was evaluated using the coefficient of determination (R²), mean squared error (MSE), and Bland–Altman analysis (bias and 95% limits of agreement, LoA). Predictions from two conventional TBV-based formulas were used as benchmarks.</p> <p>Results</p> <p>Of 993 screened cases, 577 met inclusion criteria (447 males, mean age 66.8 ± 11.7 years; 130 females, 69.5 ± 10.6 years). SVR on standardized predictors achieved the highest accuracy (R² = 0.498, MSE = 0.517), outperforming GLM (R² = 0.429,</p>

	<p>MSE=0.797) and MLP (R2 = 0.332, MSE=0.669). Conventional formulas showed lower performance (R2 = 0.325, MSE ≥ 1.48). Bland–Altman analysis for SVR demonstrated minimal bias (–0.0028g/dL) and narrower LoA (–1.42 to 1.41 g/dL) than conventional methods (bias –1.33 g/dL; LoA –3.49 to 0.83 g/dL).</p> <p>Conclusion</p> <p>These findings suggest that an SVR-based model improves prediction of HbCPB over conventional approaches, supporting optimized transfusion management and reduced hemodilution-related risks.</p>
<p>Response to Reviewers:</p>	<p>Response to comments of reviewers on the paper “Estimation of Hemoglobin Concentration at the Initiation of Cardiopulmonary Bypass Using Support Vector Regression” reference ject250087</p> <p>First of all, the authors thank the reviewers for their comments. The reviewer’s concerns are addressed the attached file.</p>



FUJITA HEALTH UNIVERSITY

Clinical and Educational Collaboration Unit, School of Medical
Sciences, Fujita Health University
1-98, Dengakugakubo, Kutsukake-cho, Toyoake, Aichi, Japan,
470-1192
e-mail: mhibiya@fujita-hu.ac.jp
Tel: +81-562-93-2654

December 1st, 2025

Dr. Raymond K. Wong

Editor-in-Chief

The Journal of ExtraCorporeal Technology

Dear Dr. Raymond K. Wong

We wish to resubmit the attached manuscript (ject250087) entitled “*Estimation of Hemoglobin Concentration at the Initiation of Cardiopulmonary Bypass Using Support Vector Regression*” for consideration as an original article in the *Journal of ExtraCorporeal Technology*.

The resubmission includes our responses to the reviewers’ comments (**Response_to_Reviewer.pdf**) and the revised manuscript with tracked changes (**Track_changes.pdf**), where modifications are highlighted in red. We sincerely thank you and the reviewers for your thoughtful suggestions and insights, which have enriched the manuscript and strengthened the presentation of our research.

Yours sincerely,

Makoto HIBIYA, Ph.D

Clinical and Educational Collaboration Unit, School of Medical Sciences, Fujita Health University

1-98, Dengakugakubo, Kutsukake-cho, Toyoake, Aichi, Japan, 470-1192

E-mail: mhibiya@shizuoka.ac.jp

Tel: +81-562-93-2654

[Reviewer 1]

Response to comments of **reviewer 1** on the paper “Estimation of Hemoglobin Concentration at the Initiation of Cardiopulmonary Bypass Using Support Vector Regression” reference [ject250087](#)

First of all, the authors thank the reviewer for their comments. The reviewer’s concerns are addressed below.

•**Comment (1)**

This research article discusses the development of a new method for estimating hemoglobin concentration at the initiation of cardiopulmonary bypass (HbCPB). HbCPB is generally determined by total blood volume (calculated using height, weight, and/or BSA depending on the formula), priming volume and pre-CPB hemoglobin. The authors included other patient-related data such as laboratory data and clinical history along with the traditional component and employed three different training models to determine which model potentially performs better than other models and conventional methods. The authors found that their models perform slightly better than the conventional methods in predicting HbCPB. This manuscript provides a new concept of predicting HbCPB, which can perform better than the conventional formula and implies that other patient related components likely contribute to HbCPB. This manuscript is potentially publishable in JECT but requires major rewriting for JECT readers to appreciate their work better.

•**Answer (1)**

We are grateful for the reviewer’s comprehensive assessment. Guided by these comments, we have revised the manuscript throughout and provided a point-by-point response below.

•**Comment (2)**

It seems to be beyond the capacity of JECT readers to understand the training models as described in the Materials and Methods section. The authors need to explain their "Analysis protocol" with plain language instead of scientific jargon. Most equations may not mean anything to JECT readers. The authors can remove several equations that are not necessary to explain their methodology in plain language.

Response (2)

In accordance with the reviewer’s suggestion, we have substantially simplified the “Analysis Protocol,” replacing technical jargon with plain language and removing equations that are not essential for understanding the methodology. We believe these revisions improve accessibility for JECT readers.

Before modification 2-1

P. 8

Linear model for hemoglobin concentration prediction at initiation of CPB

Conventional methods for estimating Hb_{CPB} at initiation (13, 14, 15) are typically classified as linear models. Linear modeling is appropriate when the relationship between explanatory variables and the target variable is approximately linear, or when a small number of predictors is used with a sufficiently large sample size. However, to the best of our knowledge,

no evidence confirms that a linear model is suitable for the variables used to estimate Hb_{CPB} . The first hypothesis of this study was that a linear model would provide an adequate framework for estimating Hb_{CPB} . A generalized linear model (GLM), a representative linear approach, is expressed by the following equation using explanatory variables \mathbf{x} and a coefficient vector $\boldsymbol{\beta}$:

$$y = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}, \quad \mathbf{x} = (x_1, x_2, \dots, x_p)^T, \quad \boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T \quad (5)$$

GLM assumes a linear relationship between the explanatory variables and the target variable. However, when real-world data exhibit nonlinearities or complex interactions, predictive performance may deteriorate. This occurs because linear models cannot adequately capture such relationships, leading to large model errors.

The GLM defined in Equation (5) was fitted to Hb_{CPB} using either the standardized predictor matrix $\mathbf{X} \in \mathbb{R}^{N \times P}$ or the PCA-transformed score matrix $\mathbf{P} \in \mathbb{R}^{N \times P_{PCA}}$. Parameter estimation was conducted by maximum likelihood under a Gaussian distribution with an identity link function, using MATLAB's `fitglm` function (MathWorks, Inc., Natick, MA, USA).

Nonlinear model for hemoglobin concentration prediction at initiation of CPB

The second hypothesis of this study was that nonlinear models would be more suitable for estimating Hb_{CPB} . Therefore, two basic nonlinear models—support vector regression (SVR) and multilayer perceptron (MLP)—were evaluated. SVR constructs a regression function using a radial basis function (RBF) kernel. Given an input data matrix $X \in \mathbb{R}^{n \times d}$ and a target vector $y \in \mathbb{R}^n$, the RBF kernel is defined as

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \quad (6)$$

A ε -SVR model is obtained by solving the convex optimization problem as follows:

$$\min_{w, b, \xi, \xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i + \xi_i^*, \quad (7)$$

$$\text{subject to } \begin{cases} y_i - (w^T \phi(x_i) + b) \leq \varepsilon + \xi_i, \\ (w^T \phi(x_i) + b) - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, n, \end{cases}$$

where ϕ is the mapping to the feature space, C is the penalty constant (BoxConstraint), and ε is the width of the epsilon-insensitive zone. The predicted Hb concentration $\hat{y}(x)$ is then expressed as a linear combination of support vectors, as follows:

$$\hat{y}(x) = \sum_{i \in S} \alpha_i K(x_i, x) + b, \quad S = \{i | \alpha_i \neq 0\}. \quad (8)$$

SVR can implicitly model complex nonlinear relationships in the input space through linear separation in a high-dimensional feature space. Its margin-maximization principle contributes to robust generalization performance. The multilayer perceptron (MLP) is a feedforward neural network with $L - 1$ hidden layers that employs nonlinear activation functions to approximate complex mappings (22). Given an input vector $\mathbf{X} \in \mathbb{R}^P$, the output of the l th layer is

$$\mathbf{h}^{(l)} = \phi(\mathbf{W}^{(l)} \mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}), \quad \mathbf{h}^{(0)} = \mathbf{X}. \quad (9)$$

where $\mathbf{W}^l \in \mathbb{R}^{n_l \times n_{l-1}}$ and $\mathbf{b}^l \in \mathbb{R}^{n_l}$ are the weight and bias vectors of layer l , and $\phi(\cdot)$ is the activation function. The output layer performs a linear transformation of the last hidden layer output $\mathbf{h}^{(L-1)}$. The output $\hat{y} \in \mathbb{R}^k$ corresponding

to the Hb concentration at the initiation of CPB is expressed as follows:

$$\hat{\mathbf{y}} = \mathbf{W}^{(L)}\mathbf{h}^{(L-1)} + \mathbf{b}^{(L)}, \quad (9)$$

where the parameters $\mathbf{W}^{(l)}, \mathbf{b}^l$ are learned via backpropagation or gradient descent methods. According to the universal approximation theorem, an MLP with a sufficiently large number of hidden units can approximate any continuous function, making it well-suited to capture complex nonlinearities and interactions. SVR was implemented to predict Hb_{CPB} using either the standardized predictor matrix $\mathbf{X} \in \mathbb{R}^{N \times P}$ or the PCA-transformed score matrix $\mathbf{P} \in \mathbb{R}^{N \times P_{\text{PCA}}}$. Model training was performed using MATLAB's `fitsvm` function with the RBF kernel defined in Equation (6). The parameters (BoxConstraint C, kernel scale parameter σ , and epsilon-insensitive zone width ε in Equation (7)) were optimized via 10-fold cross-validation using the expected-improvement-plus criterion. Model performance was evaluated using the cross-validated mean squared error (MSE) and the coefficient of determination (R^2).

The MLP was trained to predict Hb_{CPB} using either the standardized predictor matrix $\mathbf{X} \in \mathbb{R}^{N \times P}$ or the PCA-transformed score matrix $\mathbf{P} \in \mathbb{R}^{N \times P_{\text{PCA}}}$. A nested cross-validation scheme was employed: the data were divided into five outer folds ($K_{\text{outer}} = 5$), and within each outer fold, a five-fold inner cross-validation ($K_{\text{inner}} = 5$) was used to tune hyperparameters via grid search. The grid included hidden layer configurations: [4], [8], [16], [32], [32,16], [64,32], and [128,64]; training algorithms: Levenberg–Marquardt (23) and Resilient Backpropagation (24); learning rates: 10^{-3} and 10^{-4} ; and early stopping tolerances: 10 and 20. The average MSE across the inner folds was used to select the optimal configuration. The final network was then retrained using the Levenberg–Marquardt algorithm with μ -control parameters (initial $\mu = 10^{-3}$, increase factor $\mu_{\text{inc}} = 10$, decrease factor $\mu_{\text{dec}} = 0.1$, maximum $\mu_{\text{max}} = 10^{10}$), and L2 regularization with weight decay strength of 0.1 (25). To mitigate convergence to local minima, five random initializations were performed, and the model yielding the lowest validation MSE was retained. Model performance was reported as the outer-fold MSE and coefficient of determination (R^2).

After modification 2-1

P. 8

Linear model for predicting hemoglobin concentration at CPB initiation

Conventional methods for estimating Hb_{CPB} (17–19) are essentially linear models that incorporate a limited set of patient and circuit factors, such as priming volume. We first evaluated whether a linear approach would be adequate for our dataset. To do so, we fit a generalized linear model (GLM) using routinely available pre-CPB variables, including body size indices, laboratory values, and circuit settings. The GLM assumes that Hb_{CPB} varies in approximate proportion to changes in these inputs. Model performance was assessed using the coefficient of determination (R^2) and mean squared error (MSE). The explicit GLM equation and estimation procedures are provided in Supplementary Equation (S-i).

Nonlinear model for predicting hemoglobin concentration at CPB initiation

Because clinical data often contain nonlinear relationships and interactions that linear models may not capture, we also evaluated two standard nonlinear approaches: Support Vector Regression (SVR) and a Multilayer Perceptron (MLP).

SVR can model curved relationships between the inputs and Hb_{CPB} without the need to specify functional forms. We used a widely adopted SVR implementation and tuned its hyperparameters by cross-validation to reduce overfitting;

further details are provided in the Supplement (equations s-ii, iii). The MLP is a simple neural-network model that learns patterns by stacking a small number of layers. We compared several compact architectures and training configurations and selected the model that performed best in cross-validation. Training details (architecture and stopping criteria) are summarized in the Supplement (equation s-iv).

For some analyses, we also used a compressed set of input variables derived from principal component analysis to reduce redundancy among predictors. We retained the smallest number of components that preserved approximately 90% of the total variance. Technical details are provided in Supplement S2.

Model training, validation, and reporting

For all models, we standardized inputs when appropriate, tuned model settings by cross-validation, and summarized performance using R^2 and MSE. Because correlation-based indices do not measure exact agreement, we also evaluated agreement between predicted and measured Hb_{CPB} with Bland–Altman analysis (bias and 95% limits of agreement).

Software and reproducibility

Analyses were performed with commonly used, **off-the-shelf software** (e.g., MATLAB equivalents or open-source alternatives). Parameter settings and code snippets sufficient for reproduction are listed in the Supplement.

Addition 2-2

P. 26

Supplemental Material

S1. Generalized Linear Model (GLM)

For standardized predictors $X \in \mathbb{R}^{N \times P}$ and target $y \in \mathbb{R}^N$ (Hb_{CPB}), we fit a Gaussian-identity GLM:

$$y = \beta_0 + X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I) \quad (\text{s - i})$$

Parameters (β_0, β) were estimated by maximum likelihood (equivalently, least squares under (i)). Implementation used MATLAB fitglm (MathWorks) with default options.

S2. Support Vector Regression (SVR)

We used ε -SVR with radial basis function (RBF) kernel as shown in the following equation:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \quad (\text{s - ii})$$

A ε -SVR model is obtained by solving the convex optimization problem as follows:

$$\min_{w, b, \xi, \xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i + \xi_i^*, \quad (\text{s - iii})$$

$$\text{subject to } \begin{cases} y_i - (w^T \phi(x_i) + b) \leq \varepsilon + \xi_i, \\ (w^T \phi(x_i) + b) - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, n, \end{cases}$$

where ϕ is the mapping to the feature space, C is the penalty constant (BoxConstraint), and ε is the width of the epsilon-

insensitive zone. The predicted Hb concentration $\hat{y}(x)$ is then expressed as a linear combination of support vectors, as follows:

$$\hat{y}(x) = \sum_{i \in S} \alpha_i K(x_i, x) + b, \quad S = \{i | \alpha_i \neq 0\}. \quad (s - iv)$$

We tuned $(C, \sigma, \gamma, \varepsilon)$ by cross-validation. Implementation used MATLAB fitcsvm with RBF kernel; expected-improvement-plus criterion in 10-fold CV guided selection.

S3. Multilayer Perceptron (MLP)

We used a feed-forward network with $L - 1$ hidden layers. For an input $x \in \mathbb{R}^P$, the layer outputs were defined as:

$$\mathbf{h}^{(l)} = \phi(\mathbf{W}^{(l)} \mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}), \quad \mathbf{h}^{(0)} = \mathbf{X}. \quad (s - v)$$

with the final prediction given by:

$$\hat{\mathbf{y}} = \mathbf{W}^{(L)} \mathbf{h}^{(L-1)} + \mathbf{b}^{(L)}, \quad (s - vi)$$

The weights $\{W^{(l)}, b^{(l)}\}$ were learned by backpropagation using either the Levenberg–Marquardt or Resilient Backpropagation algorithm (31, 32), with learning rates 10^{-3} or 10^{-4} and early-stopping patience of 10 or 20 epochs. Model selection followed a nested cross-validation scheme with 5 outer folds and a 5-fold inner grid search over hidden-layer sizes [4], [8], [16], [32], [32,16], [64,32], [128,64], algorithm $\in \{\text{LM, RProp}\}$, learning rate $\in \{10^{-3}, 10^{-4}\}$, and early-stopping tolerance $\in \{10, 20\}$. On the full standardized predictor matrix X , the selected configuration was [4] (one hidden layer with 4 neurons) trained with LM at a learning rate of 10^{-3} , early stopping of 10 epochs, and L2 weight decay of 0.1 (33); five random initializations were run and the model with the lowest validation MSE was retained.

S4. Dimensionality Reduction via PCA.

Continuous variables were z-score standardized and categorical variables were one-hot encoded. Principal component analysis (PCA) was then applied to the full predictor matrix to obtain the score matrix $P \in \mathbb{R}^{N \times P_{\text{PCA}}}$. The retention rule was prespecified as “keep the smallest number of components whose cumulative explained variance reaches at least 90%.” Component loadings were computed to support interpretability.

S5. Cross-validation, Metrics, and Comparative Baselines.

For model validation, SVR/MPL used standard train–validation splits with reporting on the final fit; SVR employed 10-fold cross-validation for hyperparameter selection and cross-validated error estimates; and MLP used nested 5×5 cross-validation for model selection with outer-fold performance reported. Model performance was summarized uniformly using R^2 and MSE, and agreement between predicted and measured values was assessed with Bland–Altman analysis (bias and 95% limits of agreement). Two published TBV-based formulas were included as comparative baselines; all models (baselines and proposed) were evaluated under the same preprocessing and validation procedures.

•Comment (3)

Additionally, equation numbering is confusing with the reference numbering. Equation numbering should be different from the reference numbering (a, b, c and so on or I, ii, iii ...).

1
2 **Response (3)**

3 In accordance with the reviewer's suggestion, we have reformatted all equation numbering to i, ii, iii, ... throughout the
4 manuscript to avoid any confusion with the reference numbering.
5
6
7

8 **Before modification 3-1**

9
10 **P. 6**

11 BMI, ideal body weight (BW_i)(14), TBV (11) and BSA (17) were calculated as follows (Equations (1)–(4)):

$$12 \quad BMI = \frac{w}{h^2}, \quad (1)$$

$$13 \quad BW_i = 22h^2, \quad (2)$$

$$14 \quad TBV = 80w, \quad (3)$$

$$15 \quad BSA = 0.7184 \times 10^{-4} \times h^{0.725} \times w^{0.425}, \quad (4)$$

16
17
18
19
20
21 **After modification 3-1**

22
23 **P. 6**

24 BMI, ideal body weight (BW_i)(14), TBV (11) and BSA (17) were calculated as follows (Equations (i)–(iv)):

$$25 \quad BMI = \frac{w}{h^2}, \quad (i)$$

$$26 \quad BW_i = 22h^2, \quad (ii)$$

$$27 \quad TBV = 80w, \quad (iii)$$

$$28 \quad BSA = 0.7184 \times 10^{-4} \times h^{0.725} \times w^{0.425}, \quad (iv)$$

29
30
31
32
33
34 **Before modification 3-2**

35
36 **P. 12**

37 To benchmark our method against conventional prediction models, the same clinical dataset was fitted using the standard
38 hematocrit-based prediction formula (14), which is expressed as:

$$39 \quad Ht_{CPB} = Ht_{beforeCPB} \times \left(1 - \frac{PV}{PV+TBV}\right), \quad (11)$$

40 where Ht_{CPB} is the hematocrit value under CPB, $Ht_{beforeCPB}$ is the hematocrit value before CPB, PV is Priming Volume.
41 TBV was substituted from two approaches: Equation (3) (15) and the following equation (14):

$$42 \quad TBV = \begin{cases} 70 \times BW & (\text{Age} < 65) \\ 60 \times BW & (\text{Age} \geq 65) \end{cases} \quad (12)$$

43 Estimated Hb concentrations based on Equation (11) and their Bland–Altman statistics, mean difference, and 95 % LoA,
44 were compared with those obtained for our approach.
45
46
47
48
49
50
51
52

53 **After modification 3-2**

54
55 **P. 11**

56 To benchmark our method against conventional prediction models, the same clinical dataset was fitted using the standard
57 hematocrit-based prediction formula (14), which is expressed as:

$$58 \quad Ht_{CPB} = Ht_{beforeCPB} \times \left(1 - \frac{PV}{PV+TBV}\right), \quad (v)$$

where Ht_{CPB} is the hematocrit value under CPB, $Ht_{beforeCPB}$ is the hematocrit value before CPB, PV is Priming Volume.

TBV was substituted from two approaches: Equation (iii) (15) and the following equation (14):

$$TBV = \begin{cases} 70 \times BW & (\text{Age} < 65) \\ 60 \times BW & (\text{Age} \geq 65) \end{cases} \quad (\text{vi})$$

Estimated Hb concentrations based on Equation (v) and their Bland–Altman statistics, mean difference, and 95 % LoA, were compared with those obtained for our approach.

Before modification 3-3

P. 12

Conversely, biases from conventional models were confirmed (reference (11): $p < 0.01, d = -1.21$; Equation (11): $p < 0.01, d = -0.617$).

After modification 3-3

P. 14

Conversely, biases from conventional models were confirmed (reference (11): $p < 0.01, d = -1.21$; Equation (xi): $p < 0.01, d = -0.617$).

●Comment (4)

Even though authors employed complicated training models, PC1 contributes 20% and PC1 to PC4 contribute to 50% (I assume that dots in Figure 2 represent PC1 through PC17 from left to right, which needs to be described in the figure legend). Components of PC1 and PC2 are mainly weight, height, TBV, BSA, BMI, ideal weight, and obesity. It indicates that total blood volume based on weight, height and/or BSA is still the most important component in estimating HbCPB in their models. I am not sure whether the authors are able to identify other components that improved the HbCPB prediction in their models compared to the conventional formula.

Response (4)

We appreciate the reviewer's insightful observation. We have revised the Figure 2 legend to explicitly state that the dots represent the cumulative inclusion of PCs (PC1 to PC17 from left to right). We agree that PC1 and PC2 are predominantly composed of weight, height, TBV, BSA, BMI, ideal weight, and obesity, supporting the interpretation that TBV-related factors remain central to Hb_{CPB} estimation. Although our models incorporating PCs 1–17 outperformed the conventional formula, we cannot presently attribute this improvement to specific PCs without further ablation or feature-contribution analyses. We have added this clarification to the Limitations section and plan to address it in future work.

Before modification 4-1

P. 6

Figure 2: Cumulative contribution rate. Top 17 principal components were required to achieve 90 % cumulative explained variance.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

After modification 4-1

P. 6

Figure 2. Cumulative explained variance. PCs are added sequentially from PC1 to PC17 (left to right); 90% cumulative explained variance is achieved with the first 17 PCs.

Before modification 4-2

P. 16

In this study, the first 17 principal components were adopted, collectively accounting for 90% of the cumulative contribution rate in the predictor matrix. While this approach preserved most of the predictor information, it may have failed to isolate features most relevant to Hb_{CPB} . Furthermore, the first principal component accounted for more than 20% of the total variance, indicating the presence of latent factors common to blood laboratory data and patient background characteristics that drive coherent variation across multiple variables. As a result, the PCR model may not have adequately captured the most predictive features of Hb_{CPB} , which likely contributed to its reduced performance, as reflected in the lower coefficient of determination and higher MSE compared with the GLM using the full predictor set. Traditional predictors used to estimate Hb_{CPB} —including body weight, priming volume (static circuit volume), and preoperative Hb concentration—were embedded within the principal components retained by PCR and were selected as input features for both the SVR and MLP models. Nevertheless, several of the principal components used in Hb_{CPB} prediction did not primarily reflect these conventional parameters. Therefore, Hb_{CPB} cannot be fully explained by body weight, priming volume, and preoperative Hb concentration alone, as suggested by conventional approaches. These findings highlight the influence of additional confounding variables, including age and comorbidities.

After modification 4-2

P. 15

In this study, the first 17 principal components were adopted, collectively accounting for 90% **cumulative explained variance** in the predictor matrix. While this approach preserved most of the predictor information, it may have failed to isolate features most relevant to Hb_{CPB} . Furthermore, the first principal component accounted for more than 20% of the total variance, **and PC1 and PC2 were primarily composed of weight, height, TBV, BSA, BMI, ideal weight, and obesity measures, supporting the interpretation that TBV-related factors remain central to Hb_{CPB} estimation.** As a result, the PCR model may not have adequately captured the most predictive features of Hb_{CPB} , **likely contributing** to its reduced performance, as reflected in the lower coefficient of determination and higher MSE compared with the GLM using the full predictor set. Traditional predictors used to estimate Hb_{CPB} —including body weight, priming volume (static circuit volume), and preoperative Hb concentration—were embedded within the principal components retained by PCR and were selected as input features for both the SVR and MLP models. Nevertheless, several of the principal components used in Hb_{CPB} prediction did not primarily reflect these conventional parameters. Therefore, Hb_{CPB} cannot be fully explained by body weight, priming volume, and preoperative Hb concentration alone, as suggested by conventional approaches. **At the same time, while models using PCs 1–17 showed better performance than the conventional formula, we cannot presently ascribe the improvement to specific components without dedicated ablation or feature-contribution analyses.** These findings highlight the influence of additional confounding variables, including

age and comorbidities.

Addition 4-3

P. 18

Limitations

While the proposed models using PCs 1–17 outperformed the conventional formula, we cannot, at present, attribute this improvement to specific principal components without dedicated ablation or feature-contribution analyses. We acknowledge this as a limitation and plan to investigate the contribution of individual components and underlying variables in future work.

•Comment (5)

All scattered plots in Figure 4 need linear regression analysis to provide the slope and Y-intercept. Better estimation of HbCPB will result in the slope close to 1 and Y-intercept close to 0. R or R² are indicating the correlation but not the exact match. Just by eyeballing, the slope is likely higher than 1 in Figure 2 a, b, and c while lower than 1 in Figure 2 d and e.

Authors used both R in Figure 4 and R² in other places. It should be better to use only one to be consistent.

Response (5)

In accordance with the reviewer's recommendation, we have added ordinary least-squares linear regression analyses to all scatter plots in Figure 4 and have reported the slope and y-intercept for each panel. We agree that a slope close to 1 and an intercept close to 0 better reflect accurate estimation. Our results showed that the slopes in Figures 2(a), 2(b), and 2(c) were close to 1, whereas those in Figures 2(d) and 2(e) were not. Because R and R^2 quantify correlation and model fit rather than agreement, we further assessed accuracy using Bland–Altman bias and limits of agreement (Figure 5). To avoid confusion and ensure consistency, all correlation metrics throughout the manuscript are now reported exclusively as R^2 .

Before modification 5-1

P. 14

Figure 4 shows the comparison between measured Hb_{CPB} and predicted values based on pre-CPB variables, while Table 3 summarizes model performance indices. A GLM fitted to the standardized predictor matrix $\mathbf{X} \in \mathbb{R}^{N \times P}$ achieved a coefficient of determination of $R_{GLM}^2 = 0.429$, adjusted $R_{GLM}^2 = 0.369$, and $MSE_{GLM} = 0.797$. However, multicollinearity in the input matrix was confirmed, resulting in a lack of full rank. Applying the same GLM to PCA-transformed predictors $\mathbf{X}_{PCA} \in \mathbb{R}^{N \times P_{PCA}}$ yielded $R_{PCR}^2 = 0.371$ (adjusted $R_{PCR}^2 = 0.352$) and $MSE_{PCR} = 0.808$. SVR on \mathbf{X} with optimized hyperparameters $C = 2.279$, $\gamma = 16.81$, and $\epsilon = 0.594$ yielded $R_{SVR}^2 = 0.498$ and 10-fold CV $MSE_{SVR} = 0.517$, outperforming the GLM. When SVR was applied to \mathbf{X}_{PCA} with $C = 1.662$, $\gamma = 14.28$, and $\epsilon = 0.0084$, performance decreased to $R_{SVR_{PCA}}^2 = 0.39$ and CV $MSE_{SVR_{PCA}} = 0.52$. The MLP model, tuned via nested 5-fold CV, selected one hidden-layer architecture of 4 neurons, with Resilient Backpropagation, a learning rate of 10^{-3} , and early stopping tolerance of 10. On \mathbf{X} , this configuration yielded $R_{MLP}^2 = 0.332 \pm 0.033$ and $MSE_{MLP} = 0.669 \pm 0.086$. When

1 applied to X_{PCA} , the MLP achieved $R^2 = 0.352 \pm 0.054$ and $MSE = 0.646 \pm 0.068$. The coefficients of determination from
2 both MLP configurations were lower than those from SVR. Previously published Hb_{CPB} prediction formulas yielded R^2
3 $= 0.3225$, $MSE = 2.98$ based on (15), and $R^2 = 0.3245$, $MSE = 1.48$ based on (14), both substantially lower than SVR
4 performance.
5

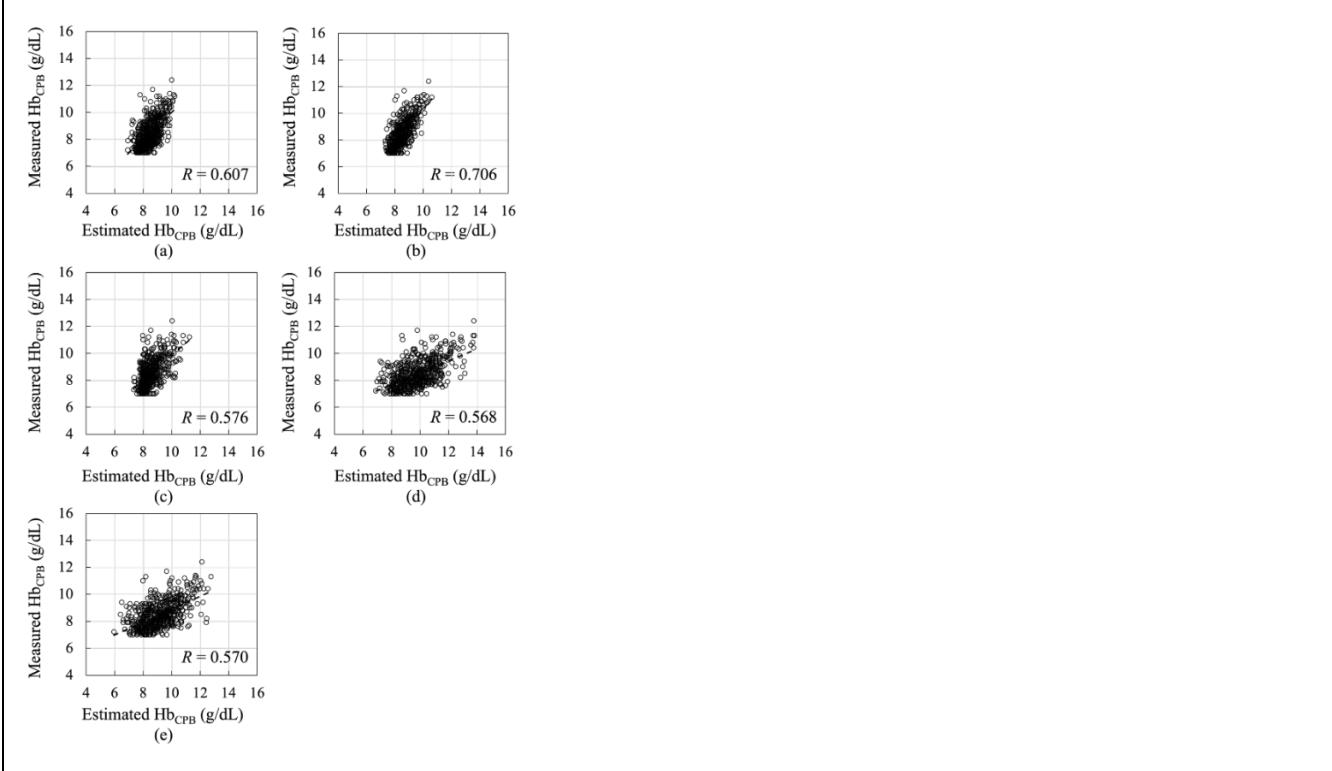
6 7 8 **After modification 5-1**

9 10 **P. 12**

11 Figure 4 shows the comparison between measured Hb_{CPB} and predicted values based on pre-CPB variables,
12 while Table 3 summarizes model performance indices. For each scatter plot in Figure 4, the ordinary least-squares
13 regression slopes and y-intercepts—where better estimation is reflected by a slope close to 1 and a y-intercept close to 0—
14 are displayed. The slopes and intercepts (panels a–e) are as follows: (a) slope = 1.00, intercept = 4.00×10^{-5} ; (b) slope =
15 1.22, intercept = -1.88 ; (c) slope = 0.92, intercept = 0.68; (d) slope = 0.44, intercept = 4.18; (e) slope = 0.48, intercept =
16 4.13. A GLM fitted to the standardized predictor matrix $X \in R^{N \times P}$ achieved a coefficient of determination of
17 $R_{GLM}^2 = 0.429$, adjusted $R_{GLM}^2 = 0.369$, and $MSE_{GLM} = 0.797$. However, multicollinearity in the input matrix was
18 confirmed, resulting in a lack of full rank. Applying the same GLM to PCA-transformed predictors $X_{PCA} \in R^{N \times P_{PCA}}$
19 yielded $R_{PCR}^2 = 0.371$ (adjusted $R_{PCR}^2 = 0.352$) and $MSE_{PCR} = 0.808$. SVR on X with optimized
20 hyperparameters $C = 2.279$, $\gamma = 16.81$, and $\varepsilon = 0.594$ yielded $R_{SVR}^2 = 0.498$ and 10-fold CV $MSE_{SVR} = 0.517$,
21 outperforming the GLM. When SVR was applied to X_{PCA} with $C = 1.662$, $\gamma = 14.28$, and $\varepsilon = 0.0084$, performance
22 decreased to $R_{SVR_{PCA}}^2 = 0.39$ and CV $MSE_{SVR_{PCA}} = 0.52$. The MLP model, tuned via nested 5-fold CV, selected one
23 hidden-layer architecture of 4 neurons, with Resilient Backpropagation, a learning rate of 10^{-3} , and early stopping
24 tolerance of 10. On X , this configuration yielded $R_{MLP}^2 = 0.332 \pm 0.033$ and $MSE_{MLP} = 0.669 \pm 0.086$. When applied
25 to X_{PCA} , the MLP achieved $R^2 = 0.352 \pm 0.054$ and $MSE = 0.646 \pm 0.068$. The coefficients of determination from
26 both MLP configurations were lower than those from SVR. Previously published Hb_{CPB} prediction formulas yielded
27 $R^2 = 0.3225$, $MSE = 2.98$ based on (19), and $R^2 = 0.3245$, $MSE = 1.48$ based on (18), both substantially lower than
28 SVR performance.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

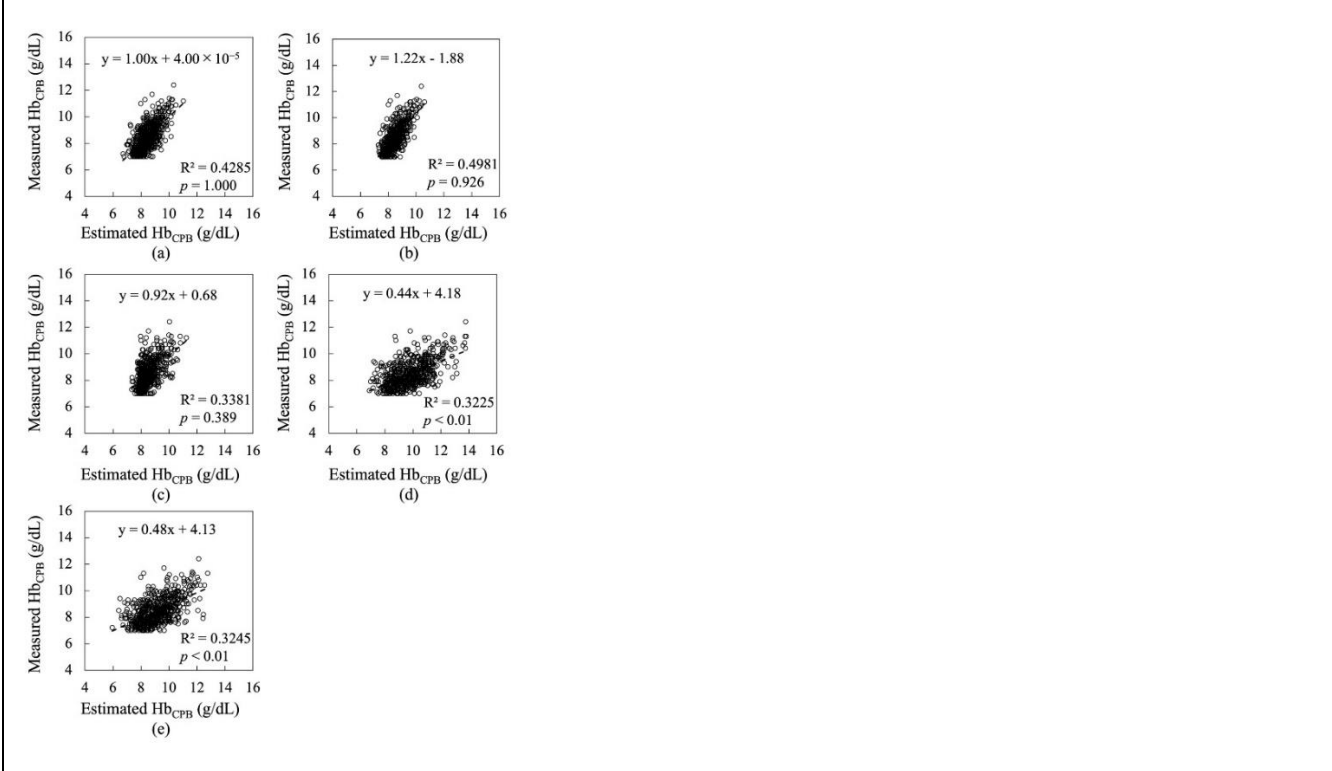
Before modification 5-2

P. 5



After modification 5-2

P. 5



1
2
3
4
5 **●Comment (6)**

P2 in abstract "This objective of this study". The objective?

6
7
8
9
10
11 **Response (6)**

We acknowledge that this issue resulted from a typographical error. We apologize for the oversight and have corrected it in the revised manuscript. Kindly verify the updated version.

12
13 **Before modification 6-1**

14 **P. 2**

This objective of this study was to propose a novel formula to accurately predict Hb concentration at the initiation of CPB (Hb_{CPB}), by incorporating circulating blood volume, laboratory data, physical measurements, and patient history.

18
19
20 **After modification 6-1**

21 **P. 2**

The objective of this study was to propose a novel formula to accurately predict Hb concentration at the initiation of CPB (Hb_{CPB}), by incorporating circulating blood volume, laboratory data, physical measurements, and patient history.

26
27
28
29 **●Comment (7)**

P2 in abstract, "PCA" is not well defined and not used again in the abstract.

32
33
34 **Response (7)**

We appreciate the reviewer's observation. To avoid introducing an undefined acronym and to improve the clarity of the abstract, we have removed the reference to PCA and revised the wording accordingly.

38
39
40 **Before modification 7-1**

41 **P. 2**

We compared generalized linear models (GLM), support vector regression (SVR), and multilayer perceptron (MLP), each trained on the full predictor set and on principal components (PCA; cumulative variance $\geq 90\%$). Model performance was evaluated using the coefficient of determination (R^2), mean squared error (MSE), and Bland–Altman analysis (bias and 95% limits of agreement, LoA). Predictions from two conventional TBV-based formulas were used as benchmarks.

45
46
47
48
49
50
51 **After modification 7-1**

52 **P. 2**

We compared generalized linear models (GLM), support vector regression (SVR), and multilayer perceptron (MLP), ~~each trained on the full predictor set and on principal components (PCA; cumulative variance $\geq 90\%$)~~. Model performance was evaluated using the coefficient of determination (R^2), mean squared error (MSE), and Bland–Altman analysis (bias and 95% limits of agreement, LoA). Predictions from two conventional TBV-based formulas were used as benchmarks.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

•Comment (8)

P4 in introduction "Hemodilution during CPB is a standard perfusion strategy used to reduce blood viscosity and enhance microcirculatory flow" needs references

Response (8)

In response to the reviewer's comment, we have added the relevant references and updated the manuscript accordingly. We would appreciate the reviewer's confirmation.

Before modification 8-1

P. 4

Hemodilution during CPB is a standard perfusion strategy used to reduce blood viscosity and enhance microcirculatory flow.

After modification 8-1

P. 4

Hemodilution during CPB is a standard perfusion strategy used to reduce blood viscosity and enhance microcirculatory flow (5).

Addition 8-2

P. 21

(5) Taniguchi FP, Martins AS. Hemodilution, kidney dysfunction and cardiac surgery. Einstein (Sao Paulo). 2009;7(1 Pt 1):103-7.

•Comment (9)

P4 in introduction "However, anemia resulting from hemodilution may cause ischemic organ injury, increase mortality, and elevate stroke risk" needs references

Response (9)

In response to the reviewer's comment, we have added the relevant references and updated the manuscript accordingly. We would appreciate the reviewer's confirmation.

Before modification 9-1

P. 4

However, anemia resulting from hemodilution may cause ischemic organ injury, increase mortality, and elevate stroke risk

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

After modification 9-1

P. 4

However, anemia resulting from hemodilution may cause ischemic organ injury, increase mortality, and elevate stroke risk (6, 7, 8).

Addition 9-2

P. 21

(6) Ranucci M, Conti D, Castelvechio S, et al. Hematocrit on cardiopulmonary bypass and outcome after coronary surgery in nontransfused patients. *Ann Thorac Surg.* 2010; 89 (1): 11-17.

(7) Karkouti K, Djaiani G, Borger MA, et al. Low hematocrit during cardiopulmonary bypass is associated with increased risk of perioperative stroke in cardiac surgery. *Ann Thorac Surg.* 2005; 80(4): 1381-1387.

(8) Tsui AKY, Dattani ND, Marsden PA, et al. Reassessing the risk of hemodilutional anemia: Some new pieces to an old puzzle. *Can J Anaesth.* 2010; 57(8): 779-791.

•Comment (10)

P7 in Analysis protocol "One-hot encoding was applied to 15 categorical variables (e.g., sex, type of surgery, and presence of comorbidities), and z-score standardization was applied to 20 continuous variables (e.g., laboratory data)." needs references

Response (10)

In response to the reviewer's comment, we have added the relevant references and updated the manuscript accordingly. We would appreciate the reviewer's confirmation.

Before modification 10-1

P. 7

One-hot encoding was applied to 15 categorical variables (e.g., sex, type of surgery, and presence of comorbidities), and z-score standardization was applied to 20 continuous variables (e.g., laboratory data).

After modification 10-1

P. 7

One-hot encoding was applied to 15 categorical variables (e.g., sex, type of surgery, and presence of comorbidities), and z-score standardization was applied to 20 continuous variables (e.g., laboratory data) (24, 25).

Addition 10-2

P. 24

(24) Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings.* 2022; 3(1): 91-99.

●Comment (11)

P15 in Bland-Altman analysis, bias and LoA of Figure 5 a, b, c, and d are described but not e.

Response (11)

We acknowledge the omission and apologize for the oversight. We have added the Bland-Altman bias and limits of agreement for Figure 5e and updated the corresponding description. Kindly verify the changes.

Before modification 11-1

P. 15

Figure 5 displays the Bland-Altman analysis results. The bias between predicted and measured GLM values was -5×10^{-7} g/dL (95% CI: -0.0654 to 0.0654 g/dL), with 95% LoA of [-1.57 g/dL, 1.57 g/dL]. For SVR, the bias was -0.0028 g/dL (95% CI: -0.0619 to 0.0562g/dL) with LoA of [-1.41 g/dL, 1.41 g/dL]. For MLP, the bias was -0.029 g/dL (95% CI: -0.0963 to 0.0375g/dL), and LoA was [-1.63 g/dL, 1.57 g/dL]. In contrast, the conventional method had a bias of -1.33 g/dL (95% CI: -1.41 to -1.24 g/dL) and LoA of [-3.49 g/dL, 0.831 g/dL]. The 95% CI for bias did not include zero, indicating systematic error. In comparison, CIs from GLM, SVR, and MLP included zero, suggesting no statistically significant bias and the absence of systematic error. Biases between measured and predicted Hb_{CPB} from GLM, SVR, and MLP were not statistically significant (GLM: $p = 1.000$, $d = -6.5 \times 10^{-7}$; SVR: $p = 0.926$, $d = -0.004$; MLP: $p = 0.389$, $d = -0.036$). Conversely, biases from conventional models were confirmed (reference (11): $p < 0.01$, $d = -1.21$; Equation (11): $p < 0.01$, $d = -0.617$).

After modification 11-1

P. 14

Figure 5 displays the Bland-Altman analysis results. The bias between predicted and measured GLM values was -5×10^{-7} g/dL (95% CI: -0.0654 to 0.0654 g/dL), with 95% LoA of [-1.57 g/dL, 1.57 g/dL]. For SVR, the bias was -0.0028 g/dL (95% CI: -0.0619 to 0.0562g/dL) with LoA of [-1.41 g/dL, 1.41 g/dL]. For MLP, the bias was -0.029 g/dL (95% CI: -0.0963 to 0.0375g/dL), and LoA was [-1.63 g/dL, 1.57 g/dL]. In contrast, the conventional **methods** had a bias of -1.33 g/dL (95% CI: -1.41 to -1.24 g/dL) and LoA of [-3.49 g/dL, 0.831 g/dL] **for the reference (11) and a bias of -0.64 g/dL (95% CI: -0.72 to -0.555 g/dL) and LoA of $[-2.67$ g/dL, 1.39 g/dL] for the Equation (v), respectively.** The 95% CI for bias did not include zero, indicating systematic error. In comparison, CIs from GLM, SVR, and MLP included zero, suggesting no statistically significant bias and the absence of systematic error. Biases between measured and predicted Hb_{CPB} from GLM, SVR, and MLP were not statistically significant (GLM: $p = 1.000$, $d = -6.5 \times 10^{-7}$; SVR: $p = 0.926$, $d = -0.004$; MLP: $p = 0.389$, $d = -0.036$). Conversely, biases from conventional models were confirmed (reference (11): $p < 0.01$, $d = -1.21$; Equation (v): $p < 0.01$, $d = -0.617$).

1
2
3
4
5 **●Comment (12)**

The Reference section is not organized in the JECT format.

6
7
8 **Response (12)**

In accordance with the reviewer's comment, we have reformatted the Reference section to comply with the JECT guidelines. In addition, former item (10), which cited a handbook, has been replaced with an original research article, and the in-text citations and numbering have been updated accordingly. We would appreciate the reviewer's confirmation.

11
12 **Before modification 12-1**

13
14 **P. 20**

- 15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- (1) Hirata Y, Hirahara N, Murakami A, Motomura N, Miyata H, Takamoto S. Current status of cardiovascular surgery in Japan: analysis of data from Japan Cardiovascular Surgery Database in 2015, 2016. Japanese Journal of Cardiovascular Surgery. 2019; 48(1): 1-5.
 - (2) Tomonobu A, Kiyoharu N, Norimichi H, Noboru M, Hiroaki M, Shin'ichi T. Current Status of Cardiovascular Surgery in Japan: Analysis of Data from Japan Cardiovascular Surgery Database in 2015, 2016. Japanese Journal of Cardiovascular Surgery. 2019; 48(1): 11-7.
 - (3) Shimizu H, Hirahara N, Motomura N, Miyata H, Takamoto S. Current Status of Cardiovascular Surgery in Japan: Analysis of Data from Japan Cardiovascular Surgery Database in 2015, 2016. Japanese Journal of Cardiovascular Surgery. 2019; 48(1): 18-24.
 - (4) Saito A, Hirahara N, Motomura N, Miyata H, Takamoto S. Current Status of Cardiovascular Surgery in Japan: Analysis of Data from Japan Cardiovascular Surgery Database in 2015, 2016. Japanese Journal of Cardiovascular Surgery. 2019; 48(1): 6-10.
 - (5) Boer C, Meesters MI, Milojevic M, Benedetto U, Bolliger D, von Heymann C, et al. 2017 EACTS/EACTA Guidelines on patient blood management for adult cardiac surgery. Journal of Cardiothoracic and Vascular Anesthesia. 2018; 32(1): 88-120.
 - (6) Ferraris VA, Brown JR, Despotis GJ, Hammon JW, Reece TB, Saha SP, et al. 2011 Update to The Society of Thoracic Surgeons and the Society of Cardiovascular Anesthesiologists Blood Conservation Clinical Practice Guidelines**The International Consortium for Evidence Based Perfusion formally endorses these guidelines. The Annals of Thoracic Surgery. 2011; 91(3): 944-82.
 - (7) Murphy GS, Hessel EA, Groom RC, et al. Optimal perfusion during cardiopulmonary bypass: an evidence-based approach. Anesthesia & Analgesia. 2009; 108(5): 1394-417.
 - (8) Gravlee GP. Cardiopulmonary bypass: principles and practice. Lippincott Williams & Wilkins; 2008.
 - (9) Hall JE. Guyton and Hall Textbook of Medical Physiology. 14th ed. Philadelphia, PA: Elsevier; 2020.
 - (10) Yamaguchi A, Momose N. Jinko Shinpai Handbook 3rd Ed. Tokyo: Chugai Igakusha; 2020 (in Japanese).
 - (11) Hilberath J, Thomas ME, Smith T, Jara C, Fitzgerald D, Wilusz K, et al. Blood volume measurement by hemodilution: association with valve disease and re-evaluation of the Allen Formula. Perfusion. 2015; 30(4): 305-11.
 - (12) Hasegawa T, Iba Y, Naraoka S, Nakajima T, Hashimoto S, Murohashi T, et al. Improvement of predicted hematocrit values after the initiation of cardiopulmonary bypass in cardiovascular surgery. Journal of Artificial Organs. 2021; 25:

1-8.

- (13) Trowbridge, Cody, Stammers, Alfred, Klayman, Myra, Brindisi, Nicholas. A Novel Calculation to Estimate Blood Volume and Hematocrit During Bypass. *J Extra Corpor Technol.* 2008; 40(1): 61-4.
- (14) Muraki R, Hiraoka A, Nagata K, Nakajima K, Oshita T, Arimichi M, et al. Novel method for estimating the total blood volume: the importance of adjustment using the ideal body weight and age for the accurate prediction of haemodilution during cardiopulmonary bypass. *Interactive CardioVascular and Thoracic Surgery.* 2018; 27(6): 802-7.
- (15) Erpicum M, Dardenne N, Hans G, Larbuisson R, Defraigne JO. Prediction of the post-dilution hematocrit during cardiopulmonary bypass. Are new formulas needed? *Perfusion.* 2016; 31(6): 458-64.
- (16) Hibiya M, Kamei T, Kubota S, et al. Study profile of the perfusion registry in Japan. *Japanese Journal of Extra-Corporeal Technology.* 2018; 45(1): 1-7.
- (17) Du Bois D, Du Bois EF. A formula to estimate the approximate surface area if height and weight be known. *Archives of Internal Medicine.* 1916; 17(6): 863-71.
- (18) Japan Society for the Study of Obesity. GUIDELINES FOR THE MANAGEMENT OF OBESITY DISEASE 2022. Tokyo: Life Science Publishing; 2022 (in Japanese).
- (19) Shin DA, Lee JC, Shin H, Cho YJ, Kim HC. Point-of-care testing of plasma free hemoglobin and hematocrit for mechanical circulatory support. *Scientific Reports.* 2021; 11(1): 3788.
- (20) Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res.* 2003; 3(null): 1157–1182.
- (21) Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.* 2016; 374(2065): 20150202.
- (22) Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Networks.* 1989; 2(5): 359-66.
- (23) Marquardt DW. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics.* 1963; 11(2): 431-41.
- (24) Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: *IEEE International Conference on Neural Networks*; 1993. p. 586-91 vol.1.
- (25) Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics.* 1970; 12(1): 55-67.

After modification 12-1

P. 21

- (1) Hirata Y, Hirahara N, Murakami A, Motomura N, Miyata H, Takamoto S. **Current status of cardiovascular surgery in Japan: analysis of data from Japan cardiovascular surgery database in 2015, 2016. 1. congenital heart surgery.** *Jpn J Cardiovasc Surg.* 2019; 48(1): 1-5 (in Japanese).
- (2) Abe T, Nakano K, Hirahara N, Motomura N, Miyata H, Takamoto S, **Current Status of cardiovascular surgery in Japan: analysis of data from Japan cardiovascular surgery database in 2015, 2016. 3-Valvular heart surgery.** *Jpn J Cardiovasc Surg.* 2019; 48(1): 11-7 (in Japanese).
- (3) Shimizu H, Hirahara N, Motomura N, Miyata H, Takamoto S. **Current status of cardiovascular surgery in Japan: analysis of data from Japan cardiovascular surgery database in 2015, 2016. 4-Thoracic aortic surgery.** *Jpn J Cardiovasc*

Surg. 2019; 48(1): 18-24 (in Japanese).

- (4) Saito A, Hirahara N, Motomura N, Miyata H, Takamoto S. Current status of cardiovascular surgery in Japan: analysis of data from Japan cardiovascular surgery database in 2015, 2016. 2. Isolated Coronary Artery Bypass Surgery. 2- Isolated coronary artery bypass surgery. *Jpn J Cardiovasc Surg.* 2019; 48(1): 6-10 (in Japanese).
- (5) Taniguchi FP, Martins AS. Hemodilution, kidney dysfunction and cardiac surgery. *Einstein (Sao Paulo).* 2009;7(1 Pt 1):103-7.
- (6) Ranucci M, Conti D, Castelvechio S, et al. Hematocrit on cardiopulmonary bypass and outcome after coronary surgery in nontransfused patients. *Ann Thorac Surg.* 2010; 89(1): 11-17.
- (7) Karkouti K, Djaiani G, Borger MA, et al. Low hematocrit during cardiopulmonary bypass is associated with increased risk of perioperative stroke in cardiac surgery. *Ann Thorac Surg.* 2005; 80(4): 1381-1387.
- (8) Tsui AKY, Dattani ND, Marsden PA, et al. Reassessing the risk of hemodilutional anemia: Some new pieces to an old puzzle. *Can J Anaesth.* 2010; 57(8): 779-791.
- (9) Boer C, Meesters MI, Milojevic M et al. 2017 EACTS/EACTA Guidelines on patient blood management for adult cardiac surgery. *J Cardiothorac Vasc Anesth.* 2018; 32(1): 88-120.
- (10) Ferraris VA, Brown JR, Despotis GJ et al. 2011 Update to The Society of Thoracic Surgeons and the Society of Cardiovascular Anesthesiologists Blood Conservation Clinical Practice Guidelines**The International Consortium for Evidence Based Perfusion formally endorses these guidelines. *Ann Thorac Surg.* 2011; 91(3): 944-982.
- (11) Murphy GS, Hessel EA, Groom RC et al. Optimal perfusion during cardiopulmonary bypass: an evidence-based approach. *Anesth Analg.* 2009; 108(5): 1394-1417.
- (12) Gravlee GP. Cardiopulmonary bypass: principles and practice. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008:403-409.
- (13) Hall JE. Guyton and Hall Textbook of Medical Physiology. 14th ed. Philadelphia, PA: Elsevier; 2020: 305-310.
- (14) Sonoki T, Mushino T, Ueda Y et al., Use guidelines on red blood cell preparation based on scientific grounds (the revision third edition). *Jpn J Transfusion Cell Therapy.* 2024; 70(6): 579—596 (in Japanese).
- (15) Hilberath J, Thomas ME, Smith T et al. Blood volume measurement by hemodilution: association with valve disease and re-evaluation of the Allen Formula. *Perfusion.* 2015; 30(4): 305-311.
- (16) Hasegawa T, Iba Y, Naraoka S et al. Improvement of predicted hematocrit values after the initiation of cardiopulmonary bypass in cardiovascular surgery. *J Artif Organs.* 2021; 25: 1-8.
- (17) Trowbridge C, Stammers A, Klayman M, Brindisi N. A Novel Calculation to Estimate Blood Volume and Hematocrit During Bypass. *J Extra Corpor Technol.* 2008; 40(1): 61-4.
- (18) Muraki R, Hiraoka A, Nagata K et al. Novel method for estimating the total blood volume: the importance of adjustment using the ideal body weight and age for the accurate prediction of haemodilution during cardiopulmonary bypass. *Interact Cardiovasc Thorac Surg.* 2018; 27(6): 802-7.
- (19) Erpicum M, Dardenne N, Hans G, Larbuisson R, Defraigne JO. Prediction of the post-dilution hematocrit during cardiopulmonary bypass. Are new formulas needed? *Perfusion.* 2016; 31(6): 458-464.
- (20) Hibiya M, Kamei T, Kubota S et al. Study profile of the perfusion registry in Japan. *Jpn J Extra Corpor Technol.* 2018; 45(1): 1-7.
- (21) Du Bois D, Du Bois EF. A formula to estimate the approximate surface area if height and weight be known. *Arch Int*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

[Med.](#) 1916; 17(6): 863-871.

- (22) Japan Society for the Study of Obesity. Guidelines for the management of obesity disease 2022. Tokyo: Life Science Publishing; 2022: 1-7 (in Japanese).
- (23) Shin DA, Lee JC, Shin H, Cho YJ, Kim HC. Point-of-care testing of plasma free hemoglobin and hematocrit for mechanical circulatory support. [Sci Rep.](#) 2021; 11(1): 3788.
- (24) Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings.* 2022; 3(1): 91-99.
- (25) Hancock JT, Khoshgoftaar TM. Survey on categorical data for neural networks. *J Big Data.* 2020; 7: 28.
- (26) Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res.* 2003; 3: 1157–1182.
- (27) Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. [Philos Trans Royal Soc A: Math, Phys Eng Sci.](#) 2016; 374(2065): 20150202.
- (28) J.M. Bland, D.G. Altman, Comparing methods of measurement: why plotting difference against standard method is misleading, *Lancet.* 1995; 346(8982): 1085-1087.
- (29) Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999; 8: 135–60.
- (30) Marquardt DW. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. [J Soc Indust Appl Math.](#) 1963; 11(2): 431-41.
- (31) Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. [In: IEEE Int Conf Neural Netw; 1993; 1: 586-91.](#)
- (32) Hoerl AE, Kennard RW. [Ridge regression: biased estimation for nonorthogonal problems.](#) *Technometrics.* 1970; 12(1): 55-67.

[Reviewer 2]

Response to comments of **reviewer 2** on the paper “Estimation of Hemoglobin Concentration at the Initiation of Cardiopulmonary Bypass Using Support Vector Regression” reference ject250087

First of all, the authors thank the reviewer for their comments. The reviewer’s concerns are addressed below.

•Comment (1)

In order to maximize the utility of the manuscript within the target publication, please consider adding to the discussion: How might a computationally intensive method such as this find its way into the operating room?

Response (1)

Thank you for the suggestion. We have added a paragraph to the Discussion describing a practical pathway for clinical deployment. Model development and training are performed on a workstation or cloud environment due to computational requirements. Once trained, however, inference is lightweight and can be executed on standard devices (e.g., a laptop or smartphone), making in-the-OR use feasible. The model can run locally with low latency and without the need for specialized hardware. We also note that integration into perfusion workflows (e.g., as a simple calculator or as an EMR/perfusion-system plugin) and consideration of governance aspects (validation, monitoring, and data security) constitute part of our future work, which has been added to the Limitations section.

Before modification 1-1

P. 16

Therefore, Hb_{CPB} cannot be fully explained by body weight, priming volume, and preoperative Hb concentration alone, as suggested by conventional approaches. These findings highlight the influence of additional confounding variables, including age and comorbidities.

After modification 1-1

P. 15

Therefore, Hb_{CPB} cannot be fully explained by body weight, priming volume, and preoperative Hb concentration alone, as suggested by conventional approaches. **At the same time, while models using PCs 1–17 showed better performance than the conventional formula, we cannot presently ascribe the improvement to specific components without dedicated ablation or feature-contribution analyses.** These findings highlight the influence of additional confounding variables, including age and comorbidities.

Although model development and training require computational resources (workstation- or cloud-based), inference is lightweight and can be executed on common devices (e.g., a standard laptop or smartphone) with low latency and no specialized hardware. This enables point-of-care use in the operating room as a simple calculator or as a module integrated into existing perfusion workflows.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Addition 1-2

P. 18

Limitations

Although inference is feasible on standard devices, clinical integration remains ongoing. Both EHR-based and device-based (perfusion console) deployments are feasible, but practical adoption will require: (A) interoperability with hospital systems and/or perfusion consoles, (B) a user interface suited to intraoperative workflow, (C) prospective validation with monitoring for model drift, and (D) security and access control consistent with institutional policies. We plan to evaluate these aspects and conduct prospective usability and safety assessments prior to routine operating-room adoption.

●Comment (2)

Could this be implemented within an EHR, or would it need to be added to one of the existing perfusion hardware devices (VIPER, etc.)?

Response (2)

Both deployment pathways are feasible. The model can be integrated within an EHR (e.g., as a calculator embedded in an order set, flowsheet, or CDS module) or on existing perfusion hardware (e.g., as a plug-in for devices such as VIPER). In both scenarios, model training is performed offline (workstation or cloud), while inference is lightweight and can run on standard hardware. We have added text clarifying these options and noted implementation and validation as part of future work in the Limitations section.

Before modification 2-1

P. 16

These findings highlight the influence of additional confounding variables, including age and comorbidities.

After modification 2-1

P. 16

These findings highlight the influence of additional confounding variables, including age and comorbidities. Although model development and training require computational resources (workstation- or cloud-based), inference is lightweight and can be executed on common devices (e.g., a standard laptop or smartphone) with low latency and no specialized hardware. This enables point-of-care use in the operating room as a simple calculator or as a module integrated into existing perfusion workflows. Accordingly, the trained model can be delivered either (i) within the EHR as a point-of-care calculator or clinical decision support widget that auto-populates inputs from routinely collected pre-CPB data, or (ii) on perfusion hardware (e.g., as a vendor plug-in) to present predictions alongside pump parameters. After deployment, the model can operate offline, and input variables can be populated automatically from routine pre-CPB data, minimizing manual entry and cognitive load.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Addition 2-2

P. 18

Limitations

While inference is feasible on standard devices, clinical integration remains ongoing work. Both EHR-based and device-based (perfusion console) deployments are feasible, but practical adoption will require (A) interoperability with hospital systems and/or perfusion consoles, (B) a user interface suited to intraoperative workflow, (C) prospective validation with monitoring for model drift, and (D) security and access control consistent with institutional policies. We plan to evaluate these aspects and conduct prospective usability and safety assessments prior to routine operating-room adoption.

●Comment (3)

The R2 value of the best performing method (while notably better than existing common methods) seemed somewhat modest - please add this to the discussion, along with references to comparable tools and their performance.

Response (3)

Thank you for this observation. We note that agreement is more appropriately assessed using Bland–Altman (BA) analysis rather than correlation alone. In our data, BA analysis demonstrates smaller bias and narrower limits of agreement for the novel methods (Fig. 3(a)–(c)) compared with the conventional formulas (Fig. 3(d)–(e)), indicating improved agreement despite moderate R² values. We have cited BA as a complementary evaluation tool and included references to the seminal methodology papers by Bland and Altman (1995, 1999). The Discussion has been revised accordingly.

Addition 3-1

P. 17

Overall, SVR and MLP, which perform nonlinear function estimation and embed variable selection within the model, outperformed linear models, including conventional approaches, in predicting outcomes from high-dimensional medical data.

Although the best-performing method yielded a modest R², agreement relevant to bedside use is better captured by Bland–Altman analysis (28, 29): the new methods demonstrated minimal bias and narrower limits of agreement than the conventional formulas (Figure 5), indicating improved agreement despite a moderate R² compared with the previous methods. Finally, while models using multiple principal components outperformed the conventional formula, we cannot presently ascribe the improvement to specific components without dedicated ablation or feature-contribution analyses; this is noted as a limitation and will be addressed in future work.

Addition 3-2

P. 24

(28) J.M. Bland, D.G. Altman, Comparing methods of measurement: why plotting difference against standard method is misleading, *Lancet*. 1995; 346(8982). 1085-1087.

(29) Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999; 8: 135–60.

•Comment (4)

Also please consider adding a discussion of the costs and benefits of implementing a computationally intensive algorithm such as this. We can see that the performance is somewhat better with your approach, however implementing it would involve considerable software programming expense - how large of a benefit might one expect for patients and cardiac programs by engaging in such an expense?

Response (4)

Thank you for this important point. We have added a paragraph to the Discussion outlining both costs and benefits. Model development and training are computationally intensive but can be performed cost-effectively on a workstation or cloud platform using widely available open-source tools. Once trained, inference is lightweight and can run on standard hospital devices (e.g., laptop or tablet) without specialized hardware, keeping deployment costs low. The anticipated clinical benefits include more accurate estimation of Hb_{CPB} at pump initiation, supporting tighter transfusion management and potentially reducing unnecessary transfusions and anemia-related risks, thereby improving care quality and workflow consistency. We also note that prospective evaluation of clinical impact and a formal cost-effectiveness analysis are planned as part of future work.

Addition 4-1

P. 17

Model development and training require computational resources; these activities can be conducted cost-effectively on a workstation or in the cloud using open-source machine-learning frameworks. After training, inference is lightweight, runs on commodity devices without accelerators, and can be embedded in existing EHR or perfusion-system workflows with minimal latency. From a clinical perspective, more accurate prediction of Hb_{CPB} at CPB initiation supports optimized transfusion strategies—reducing the likelihood of both under- and over-transfusion—and may help lower hemodilution-related risks while standardizing practice across teams. We therefore view the cost–benefit balance as favorable, while acknowledging that prospective evaluation of clinical outcomes and a formal cost-effectiveness analysis are warranted in future work.

Estimation of Hemoglobin Concentration at the Initiation of Cardiopulmonary Bypass Using Support Vector Regression

Harutoyo Hirano^{1,*}, PhD, Shun Takahashi^{2,*}, Tetsuya Kamei³, PhD, Makoto Hibiya^{4,†}, PhD

1. Department of Medical Equipment Engineering, Clinical and Educational Collaboration Unit,
Faculty of Medical Sciences, Fujita Health University, Toyoake, Aichi, Japan

2. Department of Biomedical Engineering, Graduate School of Medical Science, Fujita Health
University, Toyoake, Aichi, Japan

3. Fundamental Education Department, Faculty of Medical Sciences, Fujita Health University,
Toyoake, Aichi, Japan

4. Department of Clinical Engineering, Clinical and Educational Collaboration Unit, Faculty of
Medical Sciences, Fujita Health University, Toyoake, Aichi, Japan

** These two authors contributed equally to this work.*

† Corresponding author: mhibiya@fujita-hu.ac.jp


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65


**Estimation of Hemoglobin Concentration at the Initiation of
Cardiopulmonary Bypass Using Support Vector Regression**

1
2
3 **Keywords**
4
5

6 hemoglobin, support vector regression, cardiopulmonary bypass, machine learning, hemodilution
7
8
9

10
11
12
13 **Abstract**
14
15

16
17 **Background:** Hemodilution during cardiopulmonary bypass (CPB) is a standard perfusion strategy
18 used to reduce blood viscosity and enhance microcirculatory flow. The hemodilution rate, expressed
19 as hemoglobin (Hb) concentration, is a key control index in CPB and is currently estimated from total
20 blood volume (TBV).  **The objective** of this study was to propose a novel formula to accurately predict
21 Hb concentration at the initiation of CPB (Hb_{CPB}) by incorporating circulating blood volume,
22 laboratory data, physical measurements, and patient history.
23
24
25
26
27
28
29
30
31
32
33
34
35
36

37 **Methods:** We retrospectively analyzed 577 adult patients who underwent elective CPB at Fujita
38 Health University Hospital from January 2016 to December 2020. Thirty-six preoperative variables—
39 including demographics, laboratory data, circuit parameters, and indices such as TBV and ideal
40 weight—were standardized. Categorical variables underwent one-hot encoding. We compared
41 generalized linear models (GLM), support vector regression (SVR), and multilayer perceptron (MLP).
42 
43 Model performance was evaluated using the coefficient of determination (R^2), mean squared error
44 (MSE), and Bland–Altman analysis (bias and 95% limits of agreement [LoA]). Predictions from two
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3 conventional TBV-based formulas were used as benchmarks.
4
5

6 **Results:** Of 993 screened cases, 577 met inclusion criteria (447 males, mean age 66.8 ± 11.7 years;
7
8
9 130 females, 69.5 ± 10.6 years). SVR on standardized predictors achieved the highest accuracy ($R^2 =$
10
11 0.498, MSE = 0.517), outperforming GLM ($R^2 = 0.429$, MSE = 0.797) and MLP ($R^2 = 0.332$, MSE
12
13 = 0.669). Conventional formulas showed lower performance ($R^2 = 0.325$, MSE ≥ 1.48). Bland-
14
15
16 Altman analysis for SVR demonstrated minimal bias (-0.0028 g/dL) and narrower LoA (-1.42 to 1.41
17
18
19 g/dL) than conventional methods (bias -1.33 g/dL; LoA -3.49 to 0.83 g/dL).
20
21
22
23
24

25 **Conclusion:** These findings suggest that an SVR-based model improves prediction of Hb_{CPB} over
26
27
28 conventional approaches, supporting optimized transfusion management and reduced hemodilution-
29
30
31 related risks.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3 **Introduction**
4
5
6

7 Japan's population is aging rapidly, and the proportion of elderly patients undergoing cardiac surgery
8
9
10 with cardiopulmonary bypass (CPB) has steadily increased over the past decade (1–4). As this trend
11
12
13 is expected to continue, developing appropriate CPB technologies for older patients has become
14
15
16 imperative. Hemodilution during CPB is a standard perfusion strategy used to reduce blood viscosity
17
18 and enhance microcirculatory flow (5). However, anemia resulting from hemodilution may cause
19
20
21 ischemic organ injury, increase mortality, and elevate stroke risk (6, 7, 8). Current guidelines
22
23
24 recommend maintaining hematocrit (Ht) levels $\geq 21\%$ and/or hemoglobin (Hb) concentrations ≥ 7.0
25
26
27 g/dL during CPB to minimize hemodilution-related complications (9, 10, 11). To assess the
28
29
30 hemodilution rate, total blood volume (TBV) is estimated using body weight in kilograms (kg). TBV
31
32
33 estimates are typically set at 70 mL/kg for American adult males, 65 mL/kg for females (12, 13), and
34
35
36
37
38 80 mL/kg for Japanese adults (14). This method is commonly applied to all adults, regardless of age.
39
40
41 However, several studies have shown discrepancies between predicted post-dilution Ht or Hb values
42
43
44 and actual Hb concentrations at the initiation of CPB (Hb_{CPB}) (15–17). Revised formulas
45
46
47
48 incorporating sex and age have been proposed to improve TBV estimates (18). Nonetheless, these
49
50
51 methods often lack applicability to older patients, as they do not adequately account for age-related
52
53
54 TBV decline (19).
55
56

57 This study aims to develop and propose a novel formula to predict Hb_{CPB} more accurately by
58
59
60
61
62
63
64
65

1
2
3 incorporating TBV, laboratory data, physical measurements, and clinical history.
4
5
6
7
8
9

10 **Materials and Methods**

11 **Clinical data**

12
13
14
15
16
17
18 This study analyzed data that were prepared to register in the Perfusion Database maintained by the
19
20
21 Japanese Society of Extra-Corporeal Technology in Medicine (JaSECT) for patients who underwent
22
23
24 cardiac surgery with CPB at Fujita Health University Hospital between January 1, 2016, and December
25
26
27 31, 2020. Only elective cases were included. Patients with congenital heart disease or those who
28
29
30 underwent repeat surgeries were excluded to avoid duplicate entries, particularly in preoperative
31
32
33 background data. Additionally, cases involving blood priming or with hemoglobin (Hb) concentrations
34
35
36 below 7 g/dL at CPB initiation were excluded. Details of the data cleaning process are shown in Figure
37
38
39
40 1. The JaSECT Perfusion Database collects data across six domains: patient demographics; circuit and
41
42
43 priming fluid details; CPB parameters; fluid volume input and output management; laboratory data
44
45
46 and outcomes. These domains consist of 84 multiple-choice items and 159 numeric variables, totaling
47
48
49 243 parameters (20). From in-hospital data for this database, we extracted 35 pre-CPB parameters
50
51
52 along with Hb_{CPB} for model development, as described in the following section.
53
54
55

- 56 ● **Demographic and clinical variables:** sex; age at surgery; surgical procedure; height; weight;
57
58
59
60
61
62
63
64
65

1
2
3 left ventricular ejection fraction; and risk factors, including congestive heart failure, chronic
4
5
6 respiratory disease, smoking history, diabetes mellitus, arrhythmia, hypertension, dyslipidemia,
7
8
9 cardiovascular or extracardiac vascular disease, cerebrovascular disease, renal dysfunction, and
10
11
12 chronic dialysis.

- 13
14
15
16 ● **Blood laboratory data obtained by the time of admission to the operating room:** glucose
17
18 (mg/dL); potassium (mEq/L); lactate (mg/dL); creatinine (mg/dL); Hb (g/dL); results of arterial
19
20
21 blood gas analysis (pH, PaO₂(mmHg), PaCO₂(mmHg), HCO₃⁻(mEq/L)) and activated clotting
22
23
24 time (sec)
- 25
26
27
28 ● **Variables described the CPB circuit configuration:** biocompatible coating part; planned
29
30
31 priming volume (Static circuit volume); and priming fluid composition.
- 32
33
34
35 ● **Variables obtained from formulas using collected data:** ideal body weight; TBV; body mass
36
37
38 index (BMI); obesity classification; and body surface area (BSA).

39
40
41 BMI, ideal body weight (BW_i) (18), TBV (11) and BSA (21) were calculated as follows

42
43
44 (Equations (i)–(iv)):

$$45 \text{ BMI} = \frac{w}{h^2},$$

46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Reviewer1 3-1
(i)

$$66 \text{ BW}_i = 22h^2, \quad \text{(ii)}$$

$$67 \text{ TBV} = 80w, \quad \text{(iii)}$$

$$68 \text{ BSA} = 0.7184 \times 10^{-4} \times h^{0.725} \times w^{0.425}, \quad \text{(iv)}$$

1
2
3 where w is the body weight (kg) and h is the height (m).
4
5

6 Obesity was classified according to the Japanese Society for the Study of Obesity Guidelines
7
8
9 (22) as follows: underweight ($\text{BMI} < 18.5 \text{ kg/m}^2$), normal weight ($18.5 \leq \text{BMI} < 25 \text{ kg/m}^2$), obesity
10
11
12 class I ($25 \leq \text{BMI} < 30 \text{ kg/m}^2$), class II ($30 \leq \text{BMI} < 35 \text{ kg/m}^2$), class III ($35 \leq \text{BMI} < 40 \text{ kg/m}^2$), and
13
14
15 class IV ($\text{BMI} \geq 40 \text{ kg/m}^2$). Hematocrit (Ht) values measured before and at CPB initiation were
16
17
18 converted to Hb concentrations using the equation: $\text{Ht} \approx 3 \times \text{Hb}$ (23).
19
20
21
22
23
24
25

26 **Analysis protocol**

27 28 29 **Preprocessing of eligible Clinical Data**

30
31
32
33 The 35 parameters described in the previous section were considered candidate explanatory variables
34
35
36 for estimating Hb_{CPB} . One-hot encoding was applied to 15 categorical variables (e.g., sex, type of
37
38
39 surgery, and presence of comorbidities), and z-score standardization was applied to 20 continuous
40
41
42 variables (e.g., laboratory data) (24, 25). The interactions among variables, and between each variable
43
44
45 and hemoglobin concentration, remain unknown. In general, including variables with low or no
46
47
48 relevance to the target variable may impair the generalizability of a predictive model and lead to
49
50
51 overfitting. It is therefore advisable to limit explanatory variables to those that carry significant
52
53
54 information about the outcome (26). However, this restriction may exclude potentially informative
55
56
57 features. To address this, we prepared two datasets for analysis: the full standardized dataset and a
58
59
60
61
62
63
64
65

Reviewers' comments: Reviewer1 10-1

1
2
3 dimensionally reduced dataset, described below.
4
5

6 Principal Component Analysis (PCA) (27) was applied to the standardized explanatory matrix
7
8
9 $\mathbf{X} \in \mathbb{R}^{N \times P}$ (where N is the number of subjects and P is the number of variables) to mitigate
10
11
12 multicollinearity. Eigenvalue decomposition was performed on the covariance matrix $\Sigma = \frac{1}{N-1} \mathbf{X}^T \mathbf{X}$,
13
14
15 and the smallest number of principal components K was selected such that the cumulative
16
17
18 contribution of the eigenvalues $\lambda_{k=1}^P$ exceeded 90 %. Next, the principal component score matrix
19
20
21 $\mathbf{P} = [\mathbf{t}_1, \dots, \mathbf{t}_K] = \mathbf{X}[\mathbf{v}_1, \dots, \mathbf{v}_K] \in \mathbb{R}^{N \times K}$ was calculated using the principal component vector $\mathbf{v}_{k=1}^K$.
22
23
24

25 After PCA, \mathbf{P} was used as the explanatory variable for the subsequent model training.
26
27

28
29 
30
31

32 **Linear model for predicting hemoglobin concentration at CPB initiation** 33 34

35
36 Conventional methods for estimating Hb_{CPB} (17–19) are essentially linear models that incorporate a
37
38
39 limited set of patient and circuit factors, such as priming volume. We first evaluated whether a linear
40
41
42 approach would be adequate for our dataset. To do so, we fit a generalized linear model (GLM) using
43
44
45 routinely available pre-CPB variables, including body size indices, laboratory values, and circuit
46
47
48 settings. The GLM assumes that Hb_{CPB} varies in approximate proportion to changes in these inputs.
49
50
51 Model performance was assessed using the coefficient of determination (R^2) and mean squared error
52
53
54 (MSE). The explicit GLM equation and estimation procedures are provided in Supplementary
55
56
57
58 Equation (S-i).
59
60
61
62
63
64
65

Nonlinear model for predicting hemoglobin concentration at CPB initiation

Because clinical data often contain nonlinear relationships and interactions that linear models may not capture, we also evaluated two standard nonlinear approaches: Support Vector Regression (SVR) and a Multilayer Perceptron (MLP).

SVR can model curved relationships between the inputs and Hb_{CPB} without the need to specify functional forms. We used a widely adopted SVR implementation and tuned its hyperparameters by cross-validation to reduce overfitting; further details are provided in the Supplement (equations s-ii, iii). The MLP is a simple neural-network model that learns patterns by stacking a small number of layers. We compared several compact architectures and training configurations and selected the model that performed best in cross-validation. Training details (architecture and stopping criteria) are summarized in the Supplement (equation s-iv).

For some analyses, we also used a compressed set of input variables derived from principal component analysis to reduce redundancy among predictors. We retained the smallest number of components that preserved approximately 90% of the total variance. Technical details are provided in Supplement S2.

Model training, validation, and reporting

For all models, we standardized inputs when appropriate, tuned model settings by cross-validation, and summarized performance using R^2 and MSE. Because correlation-based indices do not measure exact agreement, we also evaluated agreement between predicted and measured Hb_{CPB} with Bland–Altman analysis (bias and 95% limits of agreement).

Software and reproducibility

Analyses were performed with commonly used, off-the-shelf software (e.g., MATLAB equivalents or open-source alternatives). Parameter settings and code snippets sufficient for reproduction are listed in the Supplement.

Evaluation of the proposed method

To evaluate the performance of the trained models, predictions on the validation and hold-out test sets were computed and compared against the reference values of Hb_{CPB} . As this is a regression task, three metrics were used to quantify prediction accuracy: MSE, mean absolute error, and the coefficient of determination (R^2). In addition, Bland–Altman analysis was performed to further examine the agreement between predicted and measured hemoglobin concentrations. From this analysis, the mean difference (bias) and 95% limits of agreement (LoA), defined as ± 1.96 times the standard deviation of

1
2
3 the differences, were calculated.
4
5

6 To benchmark our method against conventional prediction models, the same clinical dataset was
7
8 fitted using the standard hematocrit-based prediction formula (18), which is expressed as:
9
10

$$Ht_{CPB} = Ht_{beforeCPB} \times \left(1 - \frac{PV}{PV + TBV}\right),$$

(v) Reviewer1 3-2
↓

11 where Ht_{CPB} is the hematocrit value under CPB, $Ht_{beforeCPB}$ is the hematocrit value before CPB,
12
13 PV is Priming Volume. TBV was substituted from two approaches: Equation (iii) (19) and the
14
15 following equation (18):
16
17
18
19
20
21
22

$$TBV = \begin{cases} 70 \times BW & (\text{Age} < 65) \\ 60 \times BW & (\text{Age} \geq 65) \end{cases} \quad (vi)$$

23
24
25
26
27
28 Estimated Hb concentrations based on Equation (v) and their Bland–Altman statistics, mean difference,
29
30 and 95 % LoA, were compared with those obtained for our approach.
31
32
33
34
35
36
37
38

39 Results

40
41
42 Data from 577 patients, including 447 males (mean age 66.1 ± 11.8 years) and 130 females (mean age
43
44
45 69.5 ± 10.6 years), were included in the analysis. Subject characteristics are summarized in Table 1.
46
47
48

49 In terms of obesity classification, 31 patients (5.4%) were underweight, 361 (62.6%) were standard
50
51 weight, 146 (25.3%) were classified as obese class I, 33 (5.7%) as class II, 5 (0.8%) as class III, and 1
52
53 (0.2%) as class IV. The surgical procedures included coronary artery bypass grafting (CABG) in 149
54
55 patients (25.8%), valve surgery in 215 (37.3%), combined CABG and valve procedures in 67 (11.6%),
56
57
58
59
60
61
62
63
64
65


1
2
3 combined CABG and other procedures in 27 (4.7%), aortic surgery in 91 (15.8%), and other
4
5
6 procedures in 28 (4.8%). Regarding biopassive coating, it was absent in 2 patients (0.3%), applied to
7
8
9 limited components in 34 (5.9%), applied to all components except the cannulae in 484 (84.0%), and
10
11
12 applied tip-to-tip in 57 (9.8%). The main priming solution was bicarbonate Ringer's in 540 patients
13
14
15 (93.6%). Other solutions included saline in 17 (3.0%), lactated Ringer's in 2 (0.3%), acetate Ringer's
16
17
18 in 3 (0.5%), starch or dextran in 3 (0.5%), and other crystalloid-only solutions in 12 (2.1%). A
19
20
21 centrifugal pump was employed in all cases.


22
23
24
25 Figure 2 shows the scree plot of the principal components derived from PCA of the predictor matrix
26
27
28 $\mathbf{X} \in \mathbb{R}^{N \times P}$, which comprises z-score standardized continuous variables and one-hot encoded
29
30
31 categorical variables. From this plot, it was confirmed that 17 components were required to reach a
32
33
34 cumulative explained variance of 90%. Figure 3 illustrates the five variables with the highest absolute
35
36
37 loadings for each principal component. These results indicate that body weight was captured primarily
38
39
40 by PC1, priming volume (static circuit volume) by PCs 5, 7, 9, and 11, and preoperative hemoglobin
41
42
43 concentration by PCs 5, 8, 9, 12, 14, and 16. Table 2 presents the results of the generalized linear
44
45
46 model fitted to the PCA-transformed predictors. The analysis revealed that intraoperative Hb_{CPB} was
47
48
49 primarily determined by six principal components: PC1, PC2, PC5, PC8, PC12, and PC14.
50
51
52

53
54 Figure 4 shows the comparison between measured Hb_{CPB} and predicted values based on pre-
55
56
57 CPB variables, while Table 3 summarizes model performance indices. For each scatter plot in Figure
58
59
60
61
62
63
64
65

1
2
3 4, the ordinary least-squares regression slopes and y-intercepts—where better estimation is
4
5
6 reflected by a slope close to 1 and a y-intercept close to 0—are displayed. The slopes and intercepts
7
8
9 (panels a–e) are as follows: (a) slope = 1.00, intercept = 4.00×10^{-5} ; (b) slope = 1.22, intercept
10
11 = -1.88 ; (c) slope = 0.92, intercept = 0.68; (d) slope = 0.44, intercept = 4.18; (e) slope = 0.48,
12
13 intercept = 4.13. A GLM fitted to the standardized predictor matrix $\mathbf{X} \in \mathbb{R}^{N \times P}$ achieved a coefficient
14
15 of determination of $R_{\text{GLM}}^2 = 0.429$, adjusted $R_{\text{GLM}}^2 = 0.369$, and $\text{MSE}_{\text{GLM}} = 0.797$. However,
16
17 multicollinearity in the input matrix was confirmed, resulting in a lack of full rank. Applying the same
18
19 GLM to PCA-transformed predictors $\mathbf{X}_{\text{PCA}} \in \mathbb{R}^{N \times P_{\text{PCA}}}$ yielded $R_{\text{PCR}}^2 = 0.371$ (adjusted $R_{\text{PCR}}^2 =$
20
21 0.352) and $\text{MSE}_{\text{PCR}} = 0.808$. SVR on \mathbf{X} with optimized hyperparameters $C = 2.279$, $\gamma = 16.81$, and
22
23 $\varepsilon = 0.594$ yielded $R_{\text{SVR}}^2 = 0.498$ and 10-fold CV $\text{MSE}_{\text{SVR}} = 0.517$, outperforming the GLM. When
24
25 SVR was applied to \mathbf{X}_{PCA} with $C = 1.662$, $\gamma = 14.28$, and $\varepsilon = 0.0084$, performance decreased to
26
27 $R_{\text{SVR}_{\text{PCA}}}^2 = 0.39$ and CV $\text{MSE}_{\text{SVR}_{\text{PCA}}} = 0.52$. The MLP model, tuned via nested 5-fold CV, selected
28
29 one hidden-layer architecture of 4 neurons, with Resilient Backpropagation, a learning rate of 10^{-2} ,
30
31 and early stopping tolerance of 10. On \mathbf{X} , this configuration yielded $R_{\text{MLP}}^2 = 0.332 \pm 0.033$ and
32
33 $\text{MSE}_{\text{MLP}} = 0.669 \pm 0.086$. When applied to \mathbf{X}_{PCA} , the MLP achieved $R^2 = 0.352 \pm 0.054$ and MSE
34
35 = 0.646 ± 0.068 . The coefficients of determination from both MLP configurations were lower than
36
37 those from SVR. Previously published Hb_{CPB} prediction formulas yielded $R^2 = 0.3225$, MSE =
38
39 2.98 based on (19), and $R^2 = 0.3245$, MSE = 1.48 based on (18), both substantially lower than SVR
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3 performance.

4
5
6 Figure 5 displays the Bland–Altman analysis results. The bias between predicted and measured
7
8
9 GLM values was -5×10^{-7} g/dL (95% CI: -0.0654 to 0.0654 g/dL), with 95% LoA of [-1.57 g/dL,
10
11
12 1.57 g/dL]. For SVR, the bias was -0.0028g/dL (95% CI: -0.0619 to 0.0562g/dL) with LoA of [-1.41
13
14 g/dL, 1.41 g/dL]. For MLP, the bias was -0.029 g/dL (95% CI: -0.0963 to 0.0375g/dL), and LoA was
15
16 [-1.63 g/dL, 1.57 g/dL]. In contrast, the conventional methods had a bias of -1.33 g/dL (95% CI: -
17
18
19 1.41 to -1.24 g/dL) and LoA of [-3.49 g/dL, 0.831 g/dL] for the reference (11) and a bias of -0.64
20
21 
22 g/dL (95% CI: -0.72 to -0.555 g/dL) and LoA of [-2.67g/dL, 1.39g/dL] for the Equation (v),
23
24
25
26
27
28
29 respectively. The 95% CI for bias did not include zero, indicating systematic error. In comparison, CIs
30
31 from GLM, SVR, and MLP included zero, suggesting no statistically significant bias and the absence
32
33
34 of systematic error. Biases between measured and predicted Hb_{CPB} from GLM, SVR, and MLP were
35
36
37 not statistically significant (GLM: $p = 1.000, d = -6.5 \times 10^{-7}$; SVR: $p = 0.926, d =$
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153

1
2
3 across all evaluation metrics compared with the GLM. In PCR, dimensionality reduction is driven
4
5
6 solely by the variance structure of the predictor matrix; therefore, principal components that explain
7
8
9 substantial variance but have limited correlation with the response variable may still be retained as
10
11
12 leading components. In this study, the first 17 principal components were adopted, collectively
13
14
15 accounting for 90% cumulative explained variance in the predictor matrix. While this approach
16
17
18  Reviewer1 4-2
19 preserved most of the predictor information, it may have failed to isolate features most relevant to
20
21
22 Hb_{CPB}. Furthermore, the first principal component accounted for more than 20% of the total variance,
23
24
25 and PC1 and PC2 were primarily composed of weight, height, TBV, BSA, BMI, ideal weight, and
26
27
28 obesity measures, supporting the interpretation that TBV-related factors remain central to Hb_{CPB}
29
30
31 estimation. As a result, the PCR model may not have adequately captured the most predictive features
32
33
34 of Hb_{CPB}, likely contributing to its reduced performance, as reflected in the lower coefficient of
35
36
37 determination and higher MSE compared with the GLM using the full predictor set. Traditional
38
39
40 predictors used to estimate Hb_{CPB}—including body weight, priming volume (static circuit volume),
41
42
43 and preoperative Hb concentration—were embedded within the principal components retained by PCR
44
45
46 and were selected as input features for both the SVR and MLP models. Nevertheless, several of the
47
48
49 principal components used in Hb_{CPB} prediction did not primarily reflect these conventional
50
51
52 parameters. Therefore, Hb_{CPB} cannot be fully explained by body weight, priming volume, and
53
54
55 preoperative Hb concentration alone, as suggested by conventional approaches. At the same time,
56
57
58
59
60
61
62
63
64
65


1 while models using PCs 1–17 showed better performance than the conventional formula, we
 2
 3
 4
 5
 6 cannot presently ascribe the improvement to specific components without dedicated ablation or
 7
 8
 9
 10 feature-contribution analyses. These findings highlight the influence of additional confounding
 11
 12
 13 variables, including age and comorbidities.


14
 15
 16 Although model development and training require computational resources (workstation- or
 17
 18
 19 cloud-based), inference is lightweight and can be executed on common devices (e.g., a standard
 20
 21
 22 laptop or smartphone) with low latency and no specialized hardware. This enables point-of-care
 23
 24
 25 use in the operating room as a simple calculator or as a module integrated into existing perfusion
 26
 27
 28 workflows. Accordingly, the trained model can be delivered either (i) within the EHR as a point-

29
 30
 31 of-care calculator or clinical decision support widget that auto-populates inputs from routinely
 32
 33
 34 collected pre-CPB data, or (ii) on perfusion hardware (e.g., as a vendor plug-in) to present
 35
 36
 37 predictions alongside pump parameters. After deployment, the model can operate offline, with
 38
 39
 40 input variables populated automatically from routine pre-CPB data, minimizing manual entry and
 41
 42
 43 cognitive load.
 44
 45

46
 47
 48 Among the nonlinear models, SVR demonstrated a higher coefficient of determination and lower
 49
 50
 51 MSE compared with both the GLM and PCR, indicating that SVR had superior predictive power. In
 52
 53
 54 contrast, applying PCA prior to SVR led to a decreased coefficient of determination, likely due to the
 55
 56
 57 exclusion of Hb_{CPB}-related features during the PCA process, as these were among the bottom 10% of
 58
 59
 60
 61
 62
 63
 64
 65


1
2
3 principal components that were subsequently discarded. The MLP model, incorporating Levenberg–
4
5
6 Marquardt–based parameter optimization and early stopping, achieved stable training. When applied
7
8
9 to the standardized predictor matrix, MLP showed slightly lower performance than SVR but still
10
11
12 achieved a moderate goodness-of-fit. Overall, SVR and MLP, which perform nonlinear function
13
14
15 estimation and embed variable selection within the model, outperformed linear models, including
16
17
18 conventional approaches, in predicting outcomes from high-dimensional medical data.
19
20
21


22 Although the best-performing method yielded a modest R^2 , agreement relevant to bedside use
23
24  is better captured by Bland–Altman analysis (28, 29): the new methods demonstrated minimal bias
25
26
27
28 and narrower limits of agreement than the conventional formulas (Figure 5), indicating improved
29
30
31 agreement despite a moderate R^2 compared with previous methods. Finally, while models using
32
33
34 multiple principal components outperformed the conventional formula, we cannot presently ascribe
35
36
37 the improvement to specific components without dedicated ablation or feature-contribution analyses;
38
39
40
41 this is noted as a limitation and will be addressed in future work.
42
43

44 Model development and training require computational resources; these activities can be
45
46  conducted cost-effectively on a workstation or in the cloud using open-source machine-learning
47
48
49 frameworks. After training, inference is lightweight, runs on commodity devices without accelerators,
50
51
52 and can be embedded in existing EHR or perfusion-system workflows with minimal latency. From a
53
54
55 clinical perspective, more accurate prediction of Hb_{CPB} at CPB initiation supports optimized
56
57
58
59
60
61
62
63
64
65

1
2
3 transfusion strategies—reducing the likelihood of both under- and over-transfusion—and may help
4
5
6 lower hemodilution-related risks while standardizing practice across teams. We therefore view the
7
8
9 cost–benefit balance as favorable, while acknowledging that prospective evaluation of clinical
10
11
12 outcomes and a formal cost-effectiveness analysis are warranted in future work.
13
14
15
16
17
18

19 Limitations

20
21 While the proposed model ^{DC-1-17} outperformed the conventional formula, we cannot, at
22  Reviewer1 4-3
23
24 present, attribute this improvement to specific principal components without dedicated ablation or
25
26
27 feature-contribution analyses. We acknowledge this as a limitation and plan to investigate the
28
29
30 contribution of individual components and underlying variables in future work.
31
32
33

34 Although inference is feasible on standard devices, clinical integration remains ongoing. Both
35  Reviewer2 1-2, 2-2
36
37 EHR-based and device-based (perfusion console) deployments are feasible, but practical adoption will
38
39
40 require: (A) interoperability with hospital systems and/or perfusion consoles, (B) a user interface
41
42
43 suited to intraoperative workflow, (C) prospective validation with monitoring for model drift, and (D)
44
45
46 security and access control consistent with institutional policies. We plan to evaluate these aspects and
47
48
49 conduct prospective usability and safety assessments prior to routine operating-room adoption.
50
51
52
53
54
55
56
57
58
59
60
61
62

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Funding

The authors received no funding to complete this research.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

Data available upon request from the corresponding author.

Author Contributions

H. H. contributed to the conceptualization and design of the study, performed data analysis, drafted the initial manuscript, and participated in the review and revision of the final version. S. T. conducted statistical analysis, contributed to data interpretation, and assisted in drafting the manuscript. T.K. participated in data analysis and provided a critical review of the manuscript. M. H. contributed to the study's conceptualization and design, provided overall supervision, and critically reviewed and revised the manuscript. All authors read and approved the final manuscript.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65


Ethics Approval

This study was approved by the Ethics Review Committee of Fujita Health University (Approval No. HM21-381).

Acknowledgement

The authors express their sincere gratitude to Tomoaki Yamashiro, Yusuke Nakamura, and the Clinical Engineering Department staff at Fujita Health University Hospital for their meticulous efforts in data collection.


1
2
3 **References**
4
5
6


7 (1) Hirata Y, Hirahara N, Murakami A, Motomura N, Miyata H, Takamoto S. Current status of
8 
9 cardiovascular surgery in Japan: analysis of data from Japan cardiovascular surgery database in
10 2015, 2016. 1. congenital heart surgery. Jpn J Cardiovasc Surg. 2019; 48(1): 1-5 (in Japanese).
11
12
13
14

15
16 (2) Abe T, Nakano K, Hirahara N, Motomura N, Miyata H, Takamoto S, Current Status of
17 cardiovascular surgery in Japan: analysis of data from Japan cardiovascular surgery database in
18 2015, 2016. 3-Valvular heart surgery. Jpn J Cardiovasc Surg. 2019; 48(1): 11-7 (in Japanese).
19
20
21
22
23
24

25
26 (3) Shimizu H, Hirahara N, Motomura N, Miyata H, Takamoto S. Current status of cardiovascular
27 surgery in Japan: analysis of data from Japan cardiovascular surgery database in 2015, 2016. 4-
28 Thoracic aortic surgery. Jpn J Cardiovasc Surg. 2019; 48(1): 18-24 (in Japanese).
29
30
31
32
33
34

35 (4) Saito A, Hirahara N, Motomura N, Miyata H, Takamoto S. Current status of cardiovascular
36 surgery in Japan: analysis of data from Japan cardiovascular surgery database in 2015, 2016. 2.
37 Isolated Coronary Artery Bypass Surgery. 2-Isolated coronary artery bypass surgery. Jpn J
38 Cardiovasc Surg. 2019; 48(1): 6-10 (in Japanese).
39
40
41
42
43
44

45 
46 (5) Taniguchi FP, Martins AS. Hemodilution, kidney dysfunction and cardiac surgery. Einstein (Sao
47 Paulo). 2009;7(1 Pt 1):103-7.
48
49
50
51
52

53 
54 (6) Ranucci M, Conti D, Castelvechio S, et al. Hematocrit on cardiopulmonary bypass and outcome
55 after coronary surgery in nontransfused patients. Ann Thorac Surg. 2010; 89(1): 11-17.
56
57
58
59
60
61
62

- 1
2
3 (7) Karkouti K, Djaiani G, Borger MA, et al. Low hematocrit during cardiopulmonary bypass is
4
5
6 associated with increased risk of perioperative stroke in cardiac surgery. *Ann Thorac Surg.* 2005;
7
8
9 80(4): 1381-1387.
10
11
12 (8) Tsui AKY, Dattani ND, Marsden PA, et al. Reassessing the risk of hemodilutional anemia: Some
13
14
15 new pieces to an old puzzle. *Can J Anaesth.* 2010; 57(8): 779-791.
16
17
18 (9) Boer C, Meesters MI, Milojevic M et al. 2017 EACTS/EACTA Guidelines on patient blood
19
20
21 management for adult cardiac surgery. *J Cardiothorac Vasc Anesth.* 2018; 32(1): 88-120.
22
23
24 (10) Ferraris VA, Brown JR, Despotis GJ et al. 2011 Update to The Society of Thoracic Surgeons and
25
26
27 the Society of Cardiovascular Anesthesiologists Blood Conservation Clinical Practice
28
29
30 Guidelines**The International Consortium for Evidence Based Perfusion formally endorses these
31
32
33 guidelines. *Ann Thorac Surg.* 2011; 91(3): 944-982.
34
35
36 (11) Murphy GS, Hessel EA, Groom RC et al. Optimal perfusion during cardiopulmonary bypass: an
37
38
39 evidence-based approach. *Anesth Analg.* 2009; 108(5): 1394-1417.
40
41
42 (12) Gravlee GP. Cardiopulmonary bypass: principles and practice. 3rd ed. Philadelphia, PA:
43
44
45 Lippincott Williams & Wilkins; 2008: 416-417.
46
47
48 (13) Hall JE. Guyton and Hall Textbook of Medical Physiology. 14th ed. Philadelphia, PA: Elsevier;
49
50
51 2020: 305-310.
52
53
54 (14) Sonoki T, Mushino T, Ueda Y et al., Use guidelines on red blood cell preparation based
55
56
57
58
59
60
61
62
63
64
65

1
2
3 on scientific grounds (the revision third edition). *Jpn J Transfusion Cell Therapy*. 2024; 70(6):
4
5
6 579—596 (in Japanese).
7

8
9
10 (15)Hilberath J, Thomas ME, Smith T et al. Blood volume measurement by hemodilution: association
11
12 with valve disease and re-evaluation of the Allen Formula. *Perfusion*. 2015; 30(4): 305-311.
13
14

15
16 (16)Hasegawa T, Iba Y, Naraoka S et al. Improvement of predicted hematocrit values after the
17
18 initiation of cardiopulmonary bypass in cardiovascular surgery. *J Artif Organs*. 2021; 25: 1-8.
19
20

21
22 (17)Trowbridge C, Stammers A, Klayman M, Brindisi N. A Novel Calculation to Estimate Blood
23
24 Volume and Hematocrit During Bypass. *J Extra Corpor Technol*. 2008; 40(1): 61-4.
25
26

27
28 (18)Muraki R, Hiraoka A, Nagata K et al. Novel method for estimating the total blood volume: the
29
30 importance of adjustment using the ideal body weight and age for the accurate prediction of
31
32 haemodilution during cardiopulmonary bypass. *Interact Cardiovasc Thorac Surg*. 2018; 27(6):
33
34
35
36
37
38 802-7.
39

40
41 (19)Ercicum M, Dardenne N, Hans G, Larbuisson R, Defraigne JO. Prediction of the post-dilution
42
43 hematocrit during cardiopulmonary bypass. Are new formulas needed? *Perfusion*. 2016; 31(6):
44
45
46
47
48 458-464.
49

50
51 (20)Hibiya M, Kamei T, Kubota S et al. Study profile of the perfusion registry in Japan. *Jpn J Extra*
52
53
54
55
56
57
58
59
60
61
62
63
64
65
Corpor Technol. 2018; 45(1): 1-7.

(21)Du Bois D, Du Bois EF. A formula to estimate the approximate surface area if height and weight

1
2
3 be known. *Arch Int Med*. 1916; 17(6): 863-871.

4
5
6 (22)Japan Society for the Study of Obesity. Guidelines for the management of obesity disease 2022.

7
8
9 Tokyo: Life Science Publishing; 2022: 1-7 (in Japanese).

10
11
12 (23)Shin DA, Lee JC, Shin H, Cho YJ, Kim HC. Point-of-care testing of plasma free hemoglobin and

13
14
15 hematocrit for mechanical circulatory support. *Sci Rep*. 2021; 11(1): 3788.

16
17 ✓ Reviewer1 10-2

18
19 (24)Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation

20
21
22 techniques. *Global Transitions Proceedings*. 2022; 3(1): 91-99.

23
24
25 (25)Hancock JT, Khoshgoftaar TM. Survey on categorical data for neural networks. *J Big Data*. 2020;

26
27
28 7: 28.

29
30
31 (26)Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res*. 2003;

32
33
34 3: 1157–1182.

35
36
37 (27)Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos*

38
39
40
41
42 *Trans Royal Soc A: Math, Phys Eng Sci*. 2016; 374(2065): 20150202.

43
44 ✓ Reviewer2 3-2

45
46 (28)J.M. Bland, D.G. Altman, Comparing methods of measurement: why plotting difference against

47
48
49 standard method is misleading, *Lancet*, 1995; 346(8982): 1085-1087.

50
51 (29)Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med*

52
53
54
55 *Res*. 1999; 8: 135–60.

56
57 (30)Marquardt DW. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J Soc Indust*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Appl Math. 1963; 11(2): 431-41.

(31)Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: IEEE *Int Conf Neural Netw*; 1993; 1: 586-91.

(32)Hoerl AE, Kennard RW. *Ridge regression: biased estimation for nonorthogonal problems.* *Technometrics.* 1970; 12(1): 55-67.

Supplemental material

S1. Generalized Linear Model (GLM)

For standardized predictors $X \in \mathbb{R}^{N \times P}$ and target $y \in \mathbb{R}^N$ (Hb_{CPB}), we fit a Gaussian-identity GLM:

$$y = \beta_0 + X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I). \quad (\text{s-i})$$

Parameters (β_0, β) were estimated by maximum likelihood (equivalently, least squares under (i)).

Implementation used MATLAB fitglm (MathWorks) with default options.

S2. Support Vector Regression (SVR)

We used ε -SVR with radial basis function (RBF) kernel as shown in the following equation:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \quad (\text{s-ii})$$

A ε -SVR model is obtained by solving the convex optimization problem as follows:

$$\min_{w, b, \xi, \xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i + \xi_i^*, \quad (\text{s-iii})$$

$$\text{subject to } \begin{cases} y_i - (w^T \phi(x_i) + b) \leq \varepsilon + \xi_i, \\ (w^T \phi(x_i) + b) - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, n, \end{cases}$$

where ϕ is the mapping to the feature space, C is the penalty constant (BoxConstraint), and ε is

the width of the epsilon-insensitive zone. The predicted Hb concentration $\hat{y}(x)$ is then expressed as

a linear combination of support vectors, as follows:

$$\hat{y}(x) = \sum_{i \in S} \alpha_i K(x_i, x) + b, \quad S = \{i | \alpha_i \neq 0\}. \quad (\text{s-iv})$$

We tuned $(C, \sigma, \gamma, \varepsilon)$ by cross-validation. Implementation used MATLAB fitsvm with RBF kernel;

1
2
3 expected-improvement-plus criterion in 10-fold CV guided selection.
4
5
6
7
8

9 **S3. Multilayer Perceptron (MLP)**

10
11
12 We used a feed-forward network with $L - 1$ hidden layers. For an input $x \in \mathbb{R}^P$, the layer outputs
13
14
15
16 were defined as:

$$17 \mathbf{h}^{(l)} = \phi(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}), \quad \mathbf{h}^{(0)} = \mathbf{X}. \quad (\text{s-v})$$

18
19
20
21
22 with the final prediction given by:

$$23 \hat{\mathbf{y}} = \mathbf{W}^{(L)}\mathbf{h}^{(L-1)} + \mathbf{b}^{(L)}, \quad (\text{s-vi})$$

24
25
26
27
28 The weights $\{W^{(\ell)}, b^{(\ell)}\}$ were learned by backpropagation using either the Levenberg–Marquardt or
29
30 Resilient Backpropagation algorithm (31, 32), with learning rates 10^{-3} or 10^{-4} and early-stopping
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

1
2
3 **S4. Dimensionality Reduction via PCA.**
4
5

6 Continuous variables were z-score standardized and categorical variables were one-hot encoded.
7

8
9 Principal component analysis (PCA) was then applied to the full predictor matrix to obtain the score
10 matrix $P \in \mathbb{R}^{N \times P_{\text{PCA}}}$. The retention rule was prespecified as “keep the smallest number of components
11
12 whose cumulative explained variance reaches at least 90%.” Component loadings were computed to
13
14 support interpretability.
15
16
17
18
19
20
21
22
23
24

25 **S5. Cross-validation, Metrics, and Comparative Baselines.**
26

27 For model validation, SVR/MLP used standard train–validation splits with reporting on the final fit;
28
29 SVR employed 10-fold cross-validation for hyperparameter selection and cross-validated error
30
31 estimates; and MLP used nested 5×5 cross-validation for model selection with outer-fold performance
32
33 reported. Model performance was summarized uniformly using R^2 and MSE, and agreement
34
35 between predicted and measured values was assessed with Bland–Altman analysis (bias and 95%
36
37 limits of agreement). Two published TBV-based formulas were included as comparative baselines; all
38
39 models (baselines and proposed) were evaluated under the same preprocessing and validation
40
41
42
43
44
45
46
47
48
49
50
51 procedures.
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Table 1

Index	Total (n=577)	Male (n=447)	Female (n=130)
Demographics			
Age of surgery (years)	66.9 ± 11.6	66.1 ± 11.8	69.5 ± 10.6
Height (cm)	163.0 ± 9.2	167.0 ± 7.31	153.0 ± 6.42
Weight (kg)	63.6 ± 12.3	66.2 ± 11.5	54.8 ± 10.5
Body mass index (kg/m ²)	23.8 ± 3.8	23.8 ± 3.66	23.5 ± 4.16
Body surface area (m ²)	1.68 ± 0.18	1.74 ± 0.16	1.50 ± 0.14
Activated clotting time (s)	117.0 ± 17.3	118.0 ± 17.0	114.0 ± 18.0
LV fraction			
Good	202 (35.0%)	140 (31.3%)	62 (47.7%)
Med	336 (58.2%)	275 (61.5%)	61 (46.9%)
Bad	39 (6.8%)	32 (7.2%)	7 (5.4%)
Laboratory Data			
Creatinine (mg/dL)	0.97 [0.47]	1.01 [0.44]	0.76 [0.39]
Glucose (mg/dL)	109 [26]	109 [26]	110 [23]
Potassium (mEq/L)	3.9 [0.5]	3.9 [0.6]	3.8 [0.6]
Lactic acid (mg/dL)	5.5 [3.4]	5.6 [3.3]	5.0 [3.6]
Hb (g/dL, before CPB)	12.6 [1.8]	12.7 [1.85]	11.9 [1.6]
Hb (g/dL, initiation of CPB)	8.3 [1.4]	8.5 [1.4]	7.8[1.1]
pH	7.39 [0.07]	7.39 [0.07]	7.40 [0.07]
Po ₂ (mmHg)	450 [219]	456 [214]	419 [263]
Pco ₂ (mmHg)	44.2 [9.30]	44.9 [9.40]	42.3 [8.15]
HCO ₃ ⁻ (mEq/L)	26.8 [3.30]	26.8 [3.30]	26.8 [3.75]
Blood Volume Estimations			
Ideal weight (kg)	58.9 ± 6.6	61.1 ± 5.4	51.3 ± 4.3
TBV (mL)	5088 ± 981	5293 ± 923	4384 ± 839
Priming Volumes			
Static circuit vol. (mL)	1406 ± 125	1412 ± 128	1386 ± 112
Medical History			
	Yes	No	Unknown
CHF	32 (5.6%)	535 (92.7%)	10 (1.7%)
CLD	6 (1.0%)	560 (97.1%)	11 (1.9%)
Smoking	230 (39.9%)	332 (57.5%)	15 (2.6%)
Diabetes	170 (29.5%)	396 (68.6%)	11 (1.9%)
Arrhythmia	37 (6.4%)	530 (91.9%)	10 (1.7%)
Hypertension	389 (67.4%)	181 (31.4%)	7 (1.2%)
Hyperlipidemia	260 (45.1%)	307 (53.2%)	10 (1.7%)
Non-cardiac VD	11 (1.9%)	553 (95.8%)	13 (2.3%)
Cerebral VD	16 (2.8%)	549 (95.1%)	12 (2.1%)
Renal failure	107 (18.6%)	463 (80.2%)	7 (1.2%)
Dialysis	52 (9.0%)	518 (89.8%)	7 (1.2%)

The data, except for laboratory data, are presented as mean ± S.D. or number (%). Laboratory data are presented as medians [IQR]. LV: left ventricular; CHF: chronic heart failure; CLD: chronic lung disease; VD: vascular disease; Hb: hemoglobin; TBV: total blood volume; CPB: cardiopulmonary bypass.

Table 2

Component	Estimate	p-value
Intercept	8.48	$< 1 \times 10^{-308}$
PC1	0.173	3.56×10^{-29}
PC2	-0.0665	2.15×10^{-3}
PC3	0.0207	3.70×10^{-1}
PC4	-0.0329	1.76×10^{-1}
PC5	0.245	4.79×10^{-17}
PC6	0.0588	6.19×10^{-2}
PC7	-0.0612	6.16×10^{-2}
PC8	-0.1579	7.66×10^{-6}
PC9	-0.0149	6.76×10^{-1}
PC10	-0.0466	2.11×10^{-1}
PC11	-0.0510	1.86×10^{-1}
PC12	0.251	4.86×10^{-10}
PC13	-0.0396	3.29×10^{-1}
PC14	-0.2184	8.88×10^{-7}
PC15	0.0533	2.55×10^{-1}
PC16	0.0738	1.57×10^{-1}
PC17	0.0822	1.63×10^{-1}

Table 3

Model	MSE	R^2 (adjusted R^2)
GLM	0.797	0.429 (0.369)
PCR	0.808	0.371 (0.352)
SVR	0.517	0.489
SVR with PCA	0.520	0.390
MLP	0.669 ± 0.086	0.332 ± 0.033
MLP with PCA	0.646 ± 0.068	0.352 ± 0.054

MSE: mean squared error; R^2 : coefficient of determination; GLM: general linear model; PCR: principal component regression; SVR: support vector regression; PCR: principal component analysis; MLP: multilayer perceptron

Legends

List of Tables

Table 1: Clinical characteristics/Baseline clinical characteristics of the study population

Table 2: Regression results of PCA components for predicting Hb_{CPB}

Table 3: Comparison of prediction performance of hemoglobin prediction models

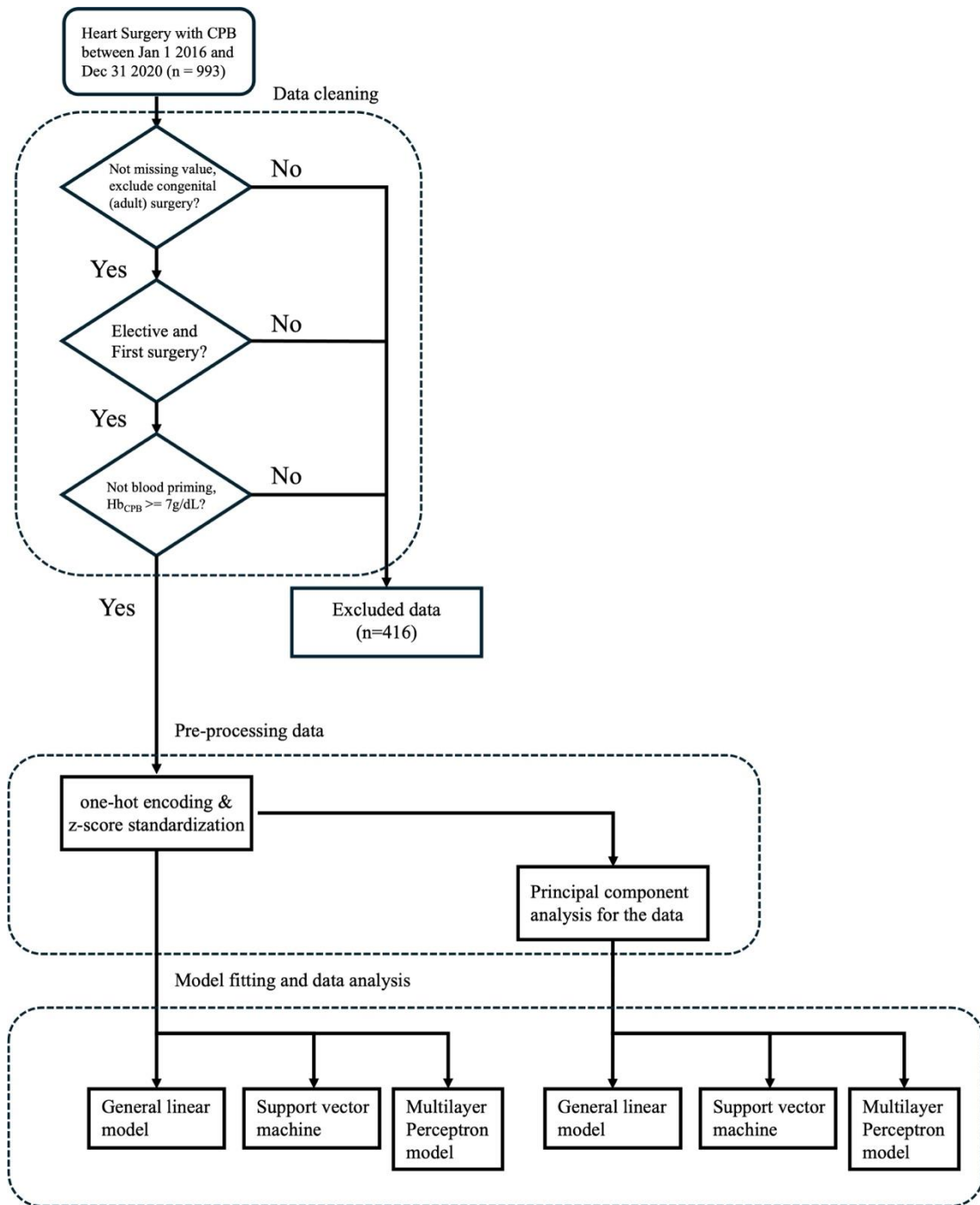


Figure 1

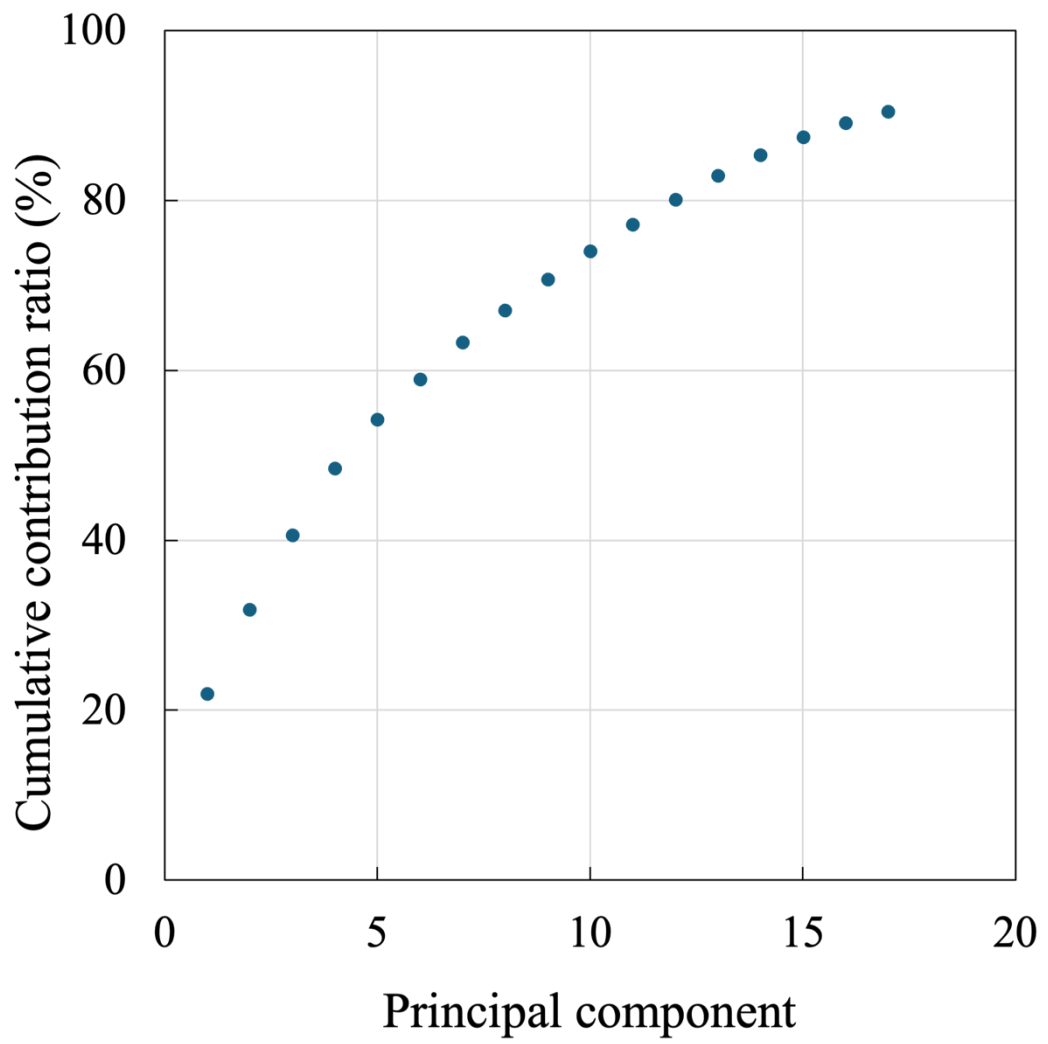


Figure 2

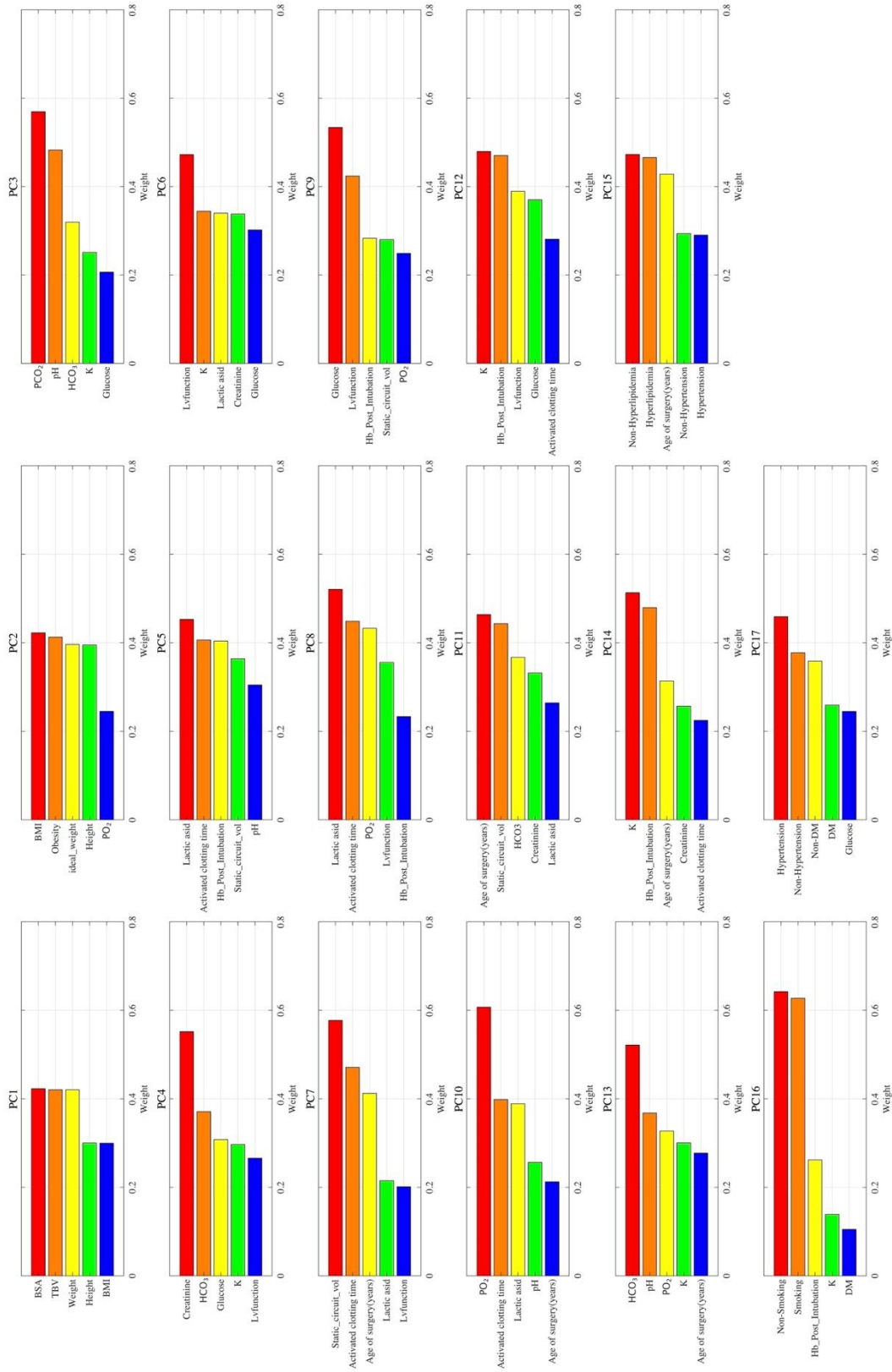


Figure 3

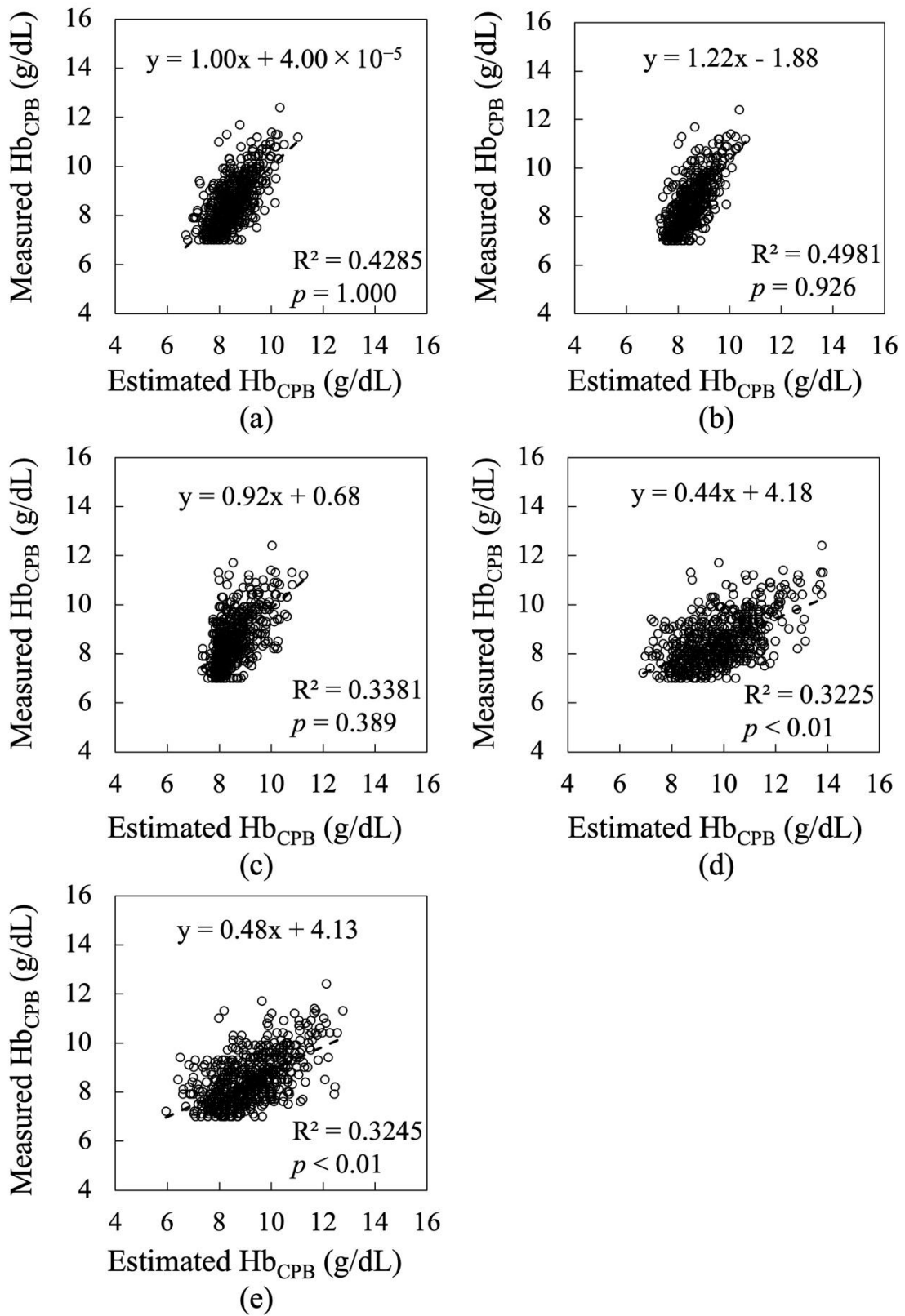


Figure4

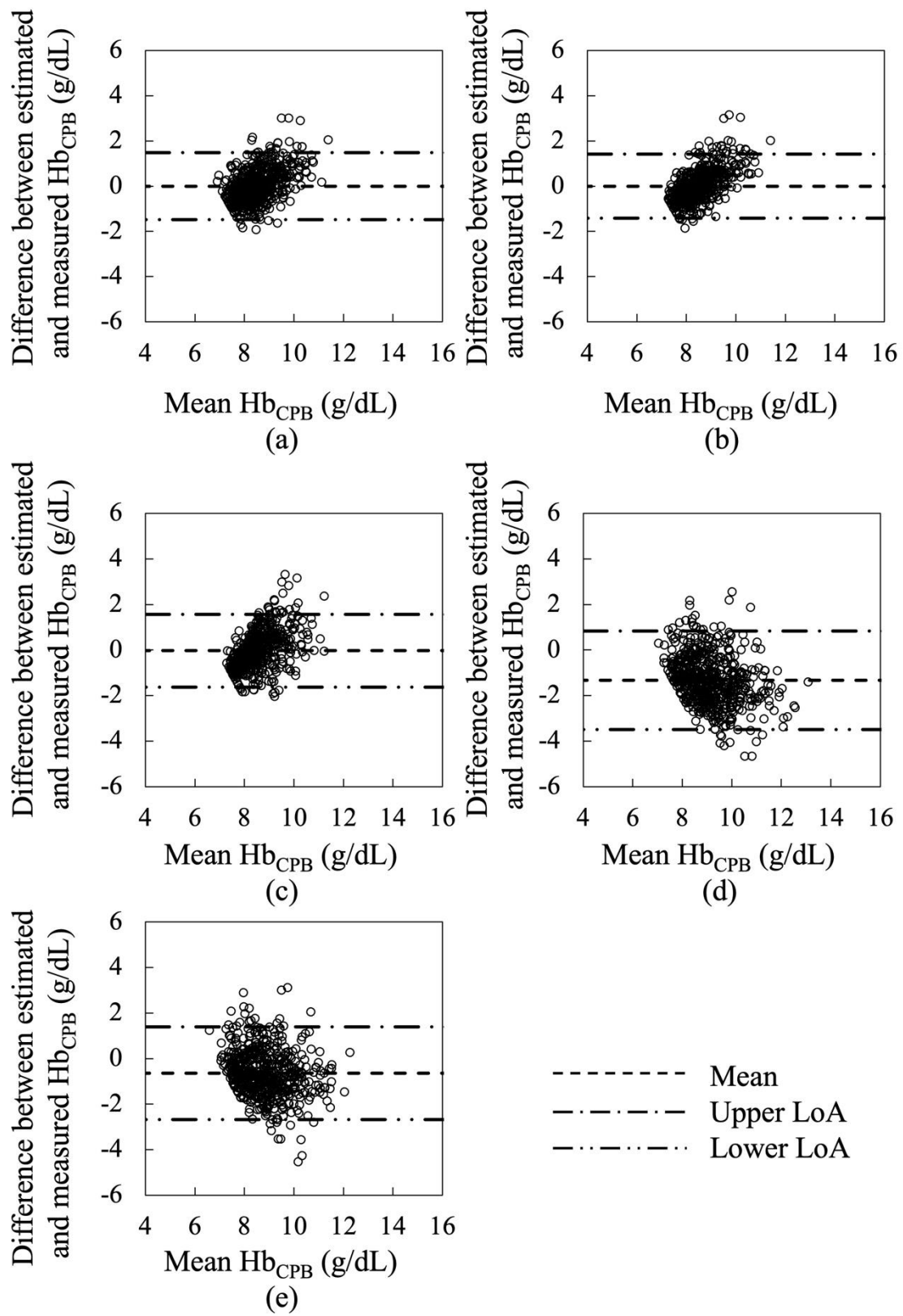


Figure 5

Legends

List of Figures

Figure 1: Analysis flowchart

Figure 2: Cumulative explained variance. Principal components (PCs) are added sequentially from PC1 to PC17 (left to right); 90% cumulative explained variance is achieved with the first 17 PCs

Figure 3: Weight of principal component analysis

Figure 4: Correlation between measured Hb_{CPB} and estimated Hb_{CPB} : (a) Hb_{CPB} was estimated based on GLM, (b) that based on SVR, (c) that based on MLP, (d) that calculated based on reference (11), and (e) that based on Equation (11).

Figure 5: Bland–Altman analysis between measured Hb_{CPB} and estimated Hb_{CPB} : (a) estimated Hb_{CPB} was estimated based on GLM, (b) that based on SVR, (c) that based on MLP, (d) that calculated based on reference (11), and (e) that based on equation (11).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

**Estimation of Hemoglobin Concentration at the Initiation of
Cardiopulmonary Bypass Using Support Vector Regression**

1
2
3 **Keywords**
4
5

6 hemoglobin, support vector regression, cardiopulmonary bypass, machine learning, hemodilution
7
8
9

10
11
12
13 **Abstract**
14
15

16
17 **Background:** Hemodilution during cardiopulmonary bypass (CPB) is a standard perfusion strategy
18 used to reduce blood viscosity and enhance microcirculatory flow. The hemodilution rate, expressed
19 as hemoglobin (Hb) concentration, is a key control index in CPB and is currently estimated from total
20 blood volume (TBV). **The objective** of this study was to propose a novel formula to accurately predict
21 Hb concentration at the initiation of CPB (Hb_{CPB}) by incorporating circulating blood volume,
22 laboratory data, physical measurements, and patient history.
23
24
25
26
27
28
29
30
31
32
33
34
35
36

37 **Methods:** We retrospectively analyzed 577 adult patients who underwent elective CPB at Fujita
38 Health University Hospital from January 2016 to December 2020. Thirty-six preoperative variables—
39 including demographics, laboratory data, circuit parameters, and indices such as TBV and ideal
40 weight—were standardized. Categorical variables underwent one-hot encoding. We compared
41 generalized linear models (GLM), support vector regression (SVR), and multilayer perceptron (**MLP**).
42 Model performance was evaluated using the coefficient of determination (R^2), mean squared error
43 (MSE), and Bland–Altman analysis (bias and 95% limits of agreement [LoA]). Predictions from two
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3 conventional TBV-based formulas were used as benchmarks.
4
5

6 **Results:** Of 993 screened cases, 577 met inclusion criteria (447 males, mean age 66.8 ± 11.7 years;
7
8
9
10 130 females, 69.5 ± 10.6 years). SVR on standardized predictors achieved the highest accuracy ($R^2 =$
11
12 0.498, MSE = 0.517), outperforming GLM ($R^2 = 0.429$, MSE = 0.797) and MLP ($R^2 = 0.332$, MSE
13
14 = 0.669). Conventional formulas showed lower performance ($R^2 = 0.325$, MSE ≥ 1.48). Bland-
15
16 Altman analysis for SVR demonstrated minimal bias (-0.0028 g/dL) and narrower LoA (-1.42 to 1.41
17
18 g/dL) than conventional methods (bias -1.33 g/dL; LoA -3.49 to 0.83 g/dL).
19
20
21
22
23
24

25 **Conclusion:** These findings suggest that an SVR-based model improves prediction of Hb_{CPB} over
26
27 conventional approaches, supporting optimized transfusion management and reduced hemodilution-
28
29 related risks.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Introduction

Japan's population is aging rapidly, and the proportion of elderly patients undergoing cardiac surgery with cardiopulmonary bypass (CPB) has steadily increased over the past decade (1–4). As this trend is expected to continue, developing appropriate CPB technologies for older patients has become imperative. Hemodilution during CPB is a standard perfusion strategy used to reduce blood viscosity and enhance microcirculatory flow (5). However, anemia resulting from hemodilution may cause ischemic organ injury, increase mortality, and elevate stroke risk (6, 7, 8). Current guidelines recommend maintaining hematocrit (Ht) levels $\geq 21\%$ and/or hemoglobin (Hb) concentrations ≥ 7.0 g/dL during CPB to minimize hemodilution-related complications (9, 10, 11). To assess the hemodilution rate, total blood volume (TBV) is estimated using body weight in kilograms (kg). TBV estimates are typically set at 70 mL/kg for American adult males, 65 mL/kg for females (12, 13), and 80 mL/kg for Japanese adults (14). This method is commonly applied to all adults, regardless of age. However, several studies have shown discrepancies between predicted post-dilution Ht or Hb values and actual Hb concentrations at the initiation of CPB (Hb_{CPB}) (15–17). Revised formulas incorporating sex and age have been proposed to improve TBV estimates (18). Nonetheless, these methods often lack applicability to older patients, as they do not adequately account for age-related TBV decline (19).

This study aims to develop and propose a novel formula to predict Hb_{CPB} more accurately by

1
2
3 incorporating TBV, laboratory data, physical measurements, and clinical history.
4
5
6
7
8
9

10 **Materials and Methods**

11 **Clinical data**

12
13
14
15
16
17
18 This study analyzed data that were prepared to register in the Perfusion Database maintained by the
19
20
21 Japanese Society of Extra-Corporeal Technology in Medicine (JaSECT) for patients who underwent
22
23
24 cardiac surgery with CPB at Fujita Health University Hospital between January 1, 2016, and December
25
26
27 31, 2020. Only elective cases were included. Patients with congenital heart disease or those who
28
29
30 underwent repeat surgeries were excluded to avoid duplicate entries, particularly in preoperative
31
32
33 background data. Additionally, cases involving blood priming or with hemoglobin (Hb) concentrations
34
35
36 below 7 g/dL at CPB initiation were excluded. Details of the data cleaning process are shown in Figure
37
38
39
40
41 1. The JaSECT Perfusion Database collects data across six domains: patient demographics; circuit and
42
43
44 priming fluid details; CPB parameters; fluid volume input and output management; laboratory data
45
46
47 and outcomes. These domains consist of 84 multiple-choice items and 159 numeric variables, totaling
48
49
50 243 parameters (20). From in-hospital data for this database, we extracted 35 pre-CPB parameters
51
52
53 along with Hb_{CPB} for model development, as described in the following section.

- 54 ● **Demographic and clinical variables:** sex; age at surgery; surgical procedure; height; weight;

1
2
3 left ventricular ejection fraction; and risk factors, including congestive heart failure, chronic
4
5
6 respiratory disease, smoking history, diabetes mellitus, arrhythmia, hypertension, dyslipidemia,
7
8
9 cardiovascular or extracardiac vascular disease, cerebrovascular disease, renal dysfunction, and
10
11
12 chronic dialysis.

- 13
14
15 ● **Blood laboratory data obtained by the time of admission to the operating room:** glucose
16
17 (mg/dL); potassium (mEq/L); lactate (mg/dL); creatinine (mg/dL); Hb (g/dL); results of arterial
18
19 blood gas analysis (pH, PaO₂(mmHg), PaCO₂(mmHg), HCO₃⁻(mEq/L)) and activated clotting
20
21 time (sec)
22
23
24
25
26
27
28 ● **Variables described the CPB circuit configuration:** biocompatible coating part; planned
29
30 priming volume (Static circuit volume); and priming fluid composition.
31
32
33
34
35 ● **Variables obtained from formulas using collected data:** ideal body weight; TBV; body mass
36
37 index (BMI); obesity classification; and body surface area (BSA).

38
39
40
41 BMI, ideal body weight (BW_i) (18), TBV (11) and BSA (21) were calculated as follows

42
43
44 (Equations (i)–(iv)):

$$45 \text{ BMI} = \frac{w}{h^2}, \quad (i)$$

$$46 \text{ BW}_i = 22h^2, \quad (ii)$$

$$47 \text{ TBV} = 80w, \quad (iii)$$

$$48 \text{ BSA} = 0.7184 \times 10^{-4} \times h^{0.725} \times w^{0.425}, \quad (iv)$$

1
2
3 where w is the body weight (kg) and h is the height (m).
4
5

6 Obesity was classified according to the Japanese Society for the Study of Obesity Guidelines
7
8
9 (22) as follows: underweight ($\text{BMI} < 18.5 \text{ kg/m}^2$), normal weight ($18.5 \leq \text{BMI} < 25 \text{ kg/m}^2$), obesity
10
11
12 class I ($25 \leq \text{BMI} < 30 \text{ kg/m}^2$), class II ($30 \leq \text{BMI} < 35 \text{ kg/m}^2$), class III ($35 \leq \text{BMI} < 40 \text{ kg/m}^2$), and
13
14
15 class IV ($\text{BMI} \geq 40 \text{ kg/m}^2$). Hematocrit (Ht) values measured before and at CPB initiation were
16
17
18 converted to Hb concentrations using the equation: $\text{Ht} \approx 3 \times \text{Hb}$ (23).
19
20
21
22
23
24
25

26 **Analysis protocol**

27 28 29 **Preprocessing of eligible Clinical Data**

30
31
32
33 The 35 parameters described in the previous section were considered candidate explanatory variables
34
35
36 for estimating Hb_{CPB} . One-hot encoding was applied to 15 categorical variables (e.g., sex, type of
37
38
39 surgery, and presence of comorbidities), and z-score standardization was applied to 20 continuous
40
41
42 variables (e.g., laboratory data) (24, 25). The interactions among variables, and between each variable
43
44
45 and hemoglobin concentration, remain unknown. In general, including variables with low or no
46
47
48 relevance to the target variable may impair the generalizability of a predictive model and lead to
49
50
51 overfitting. It is therefore advisable to limit explanatory variables to those that carry significant
52
53
54 information about the outcome (26). However, this restriction may exclude potentially informative
55
56
57 features. To address this, we prepared two datasets for analysis: the full standardized dataset and a
58
59
60
61
62
63
64
65

1
2
3 dimensionally reduced dataset, described below.
4
5

6 Principal Component Analysis (PCA) (27) was applied to the standardized explanatory matrix
7
8
9 $\mathbf{X} \in \mathbb{R}^{N \times P}$ (where N is the number of subjects and P is the number of variables) to mitigate
10
11
12 multicollinearity. Eigenvalue decomposition was performed on the covariance matrix $\Sigma = \frac{1}{N-1} \mathbf{X}^T \mathbf{X}$,
13
14
15 and the smallest number of principal components K was selected such that the cumulative
16
17
18 contribution of the eigenvalues $\lambda_{k=1}^P$ exceeded 90 %. Next, the principal component score matrix
19
20
21 $\mathbf{P} = [\mathbf{t}_1, \dots, \mathbf{t}_K] = \mathbf{X}[\mathbf{v}_1, \dots, \mathbf{v}_K] \in \mathbb{R}^{N \times K}$ was calculated using the principal component vector $\mathbf{v}_{k=1}^K$.
22
23
24
25 After PCA, \mathbf{P} was used as the explanatory variable for the subsequent model training.
26
27
28
29
30
31
32

33 **Linear model for predicting hemoglobin concentration at CPB initiation**

34
35

36 Conventional methods for estimating Hb_{CPB} (17–19) are essentially linear models that incorporate a
37
38
39 limited set of patient and circuit factors, such as priming volume. We first evaluated whether a linear
40
41
42 approach would be adequate for our dataset. To do so, we fit a generalized linear model (GLM) using
43
44
45 routinely available pre-CPB variables, including body size indices, laboratory values, and circuit
46
47
48 settings. The GLM assumes that Hb_{CPB} varies in approximate proportion to changes in these inputs.
49
50
51 Model performance was assessed using the coefficient of determination (R^2) and mean squared error
52
53
54 (MSE). The explicit GLM equation and estimation procedures are provided in Supplementary
55
56
57
58 Equation (S-i).
59
60
61
62
63
64
65

Nonlinear model for predicting hemoglobin concentration at CPB initiation

Because clinical data often contain nonlinear relationships and interactions that linear models may not capture, we also evaluated two standard nonlinear approaches: Support Vector Regression (SVR) and a Multilayer Perceptron (MLP).

SVR can model curved relationships between the inputs and Hb_{CPB} without the need to specify functional forms. We used a widely adopted SVR implementation and tuned its hyperparameters by cross-validation to reduce overfitting; further details are provided in the Supplement (equations s-ii, iii). The MLP is a simple neural-network model that learns patterns by stacking a small number of layers. We compared several compact architectures and training configurations and selected the model that performed best in cross-validation. Training details (architecture and stopping criteria) are summarized in the Supplement (equation s-iv).

For some analyses, we also used a compressed set of input variables derived from principal component analysis to reduce redundancy among predictors. We retained the smallest number of components that preserved approximately 90% of the total variance. Technical details are provided in Supplement S2.

Model training, validation, and reporting

For all models, we standardized inputs when appropriate, tuned model settings by cross-validation, and summarized performance using R^2 and MSE. Because correlation-based indices do not measure exact agreement, we also evaluated agreement between predicted and measured Hb_{CPB} with Bland–Altman analysis (bias and 95% limits of agreement).

Software and reproducibility

Analyses were performed with commonly used, off-the-shelf software (e.g., MATLAB equivalents or open-source alternatives). Parameter settings and code snippets sufficient for reproduction are listed in the Supplement.

Evaluation of the proposed method

To evaluate the performance of the trained models, predictions on the validation and hold-out test sets were computed and compared against the reference values of Hb_{CPB} . As this is a regression task, three metrics were used to quantify prediction accuracy: MSE, mean absolute error, and the coefficient of determination (R^2). In addition, Bland–Altman analysis was performed to further examine the agreement between predicted and measured hemoglobin concentrations. From this analysis, the mean difference (bias) and 95% limits of agreement (LoA), defined as ± 1.96 times the standard deviation of

1
2
3 the differences, were calculated.
4
5

6 To benchmark our method against conventional prediction models, the same clinical dataset was
7
8
9 fitted using the standard hematocrit-based prediction formula (18), which is expressed as:
10
11

$$12 \quad Ht_{CPB} = Ht_{beforeCPB} \times \left(1 - \frac{PV}{PV + TBV}\right), \quad (v)$$

13 where Ht_{CPB} is the hematocrit value under CPB, $Ht_{beforeCPB}$ is the hematocrit value before CPB,
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

$$25 \quad TBV = \begin{cases} 70 \times BW & (\text{Age} < 65) \\ 60 \times BW & (\text{Age} \geq 65) \end{cases}, \quad (vi)$$

29 Estimated Hb concentrations based on Equation (v) and their Bland–Altman statistics, mean difference,
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

40 **Results**

41
42
43 Data from 577 patients, including 447 males (mean age 66.1 ± 11.8 years) and 130 females (mean age
44
45
46 69.5 ± 10.6 years), were included in the analysis. Subject characteristics are summarized in Table 1.
47
48

49 In terms of obesity classification, 31 patients (5.4%) were underweight, 361 (62.6%) were standard
50
51
52 weight, 146 (25.3%) were classified as obese class I, 33 (5.7%) as class II, 5 (0.8%) as class III, and 1
53
54
55 (0.2%) as class IV. The surgical procedures included coronary artery bypass grafting (CABG) in 149
56
57
58 patients (25.8%), valve surgery in 215 (37.3%), combined CABG and valve procedures in 67 (11.6%),
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
4

1
2
3 combined CABG and other procedures in 27 (4.7%), aortic surgery in 91 (15.8%), and other
4
5
6 procedures in 28 (4.8%). Regarding biopassive coating, it was absent in 2 patients (0.3%), applied to
7
8
9 limited components in 34 (5.9%), applied to all components except the cannulae in 484 (84.0%), and
10
11
12 applied tip-to-tip in 57 (9.8%). The main priming solution was bicarbonate Ringer's in 540 patients
13
14
15 (93.6%). Other solutions included saline in 17 (3.0%), lactated Ringer's in 2 (0.3%), acetate Ringer's
16
17
18 in 3 (0.5%), starch or dextran in 3 (0.5%), and other crystalloid-only solutions in 12 (2.1%). A
19
20
21 centrifugal pump was employed in all cases.
22
23
24

25 Figure 2 shows the scree plot of the principal components derived from PCA of the predictor matrix
26
27 $\mathbf{X} \in \mathbb{R}^{N \times P}$, which comprises z-score standardized continuous variables and one-hot encoded
28
29 categorical variables. From this plot, it was confirmed that 17 components were required to reach a
30
31
32 cumulative explained variance of 90%. Figure 3 illustrates the five variables with the highest absolute
33
34
35 loadings for each principal component. These results indicate that body weight was captured primarily
36
37
38 by PC1, priming volume (static circuit volume) by PCs 5, 7, 9, and 11, and preoperative hemoglobin
39
40
41 concentration by PCs 5, 8, 9, 12, 14, and 16. Table 2 presents the results of the generalized linear
42
43
44 model fitted to the PCA-transformed predictors. The analysis revealed that intraoperative Hb_{CPB} was
45
46
47 primarily determined by six principal components: PC1, PC2, PC5, PC8, PC12, and PC14.
48
49
50
51
52
53

54 Figure 4 shows the comparison between measured Hb_{CPB} and predicted values based on pre-
55
56
57 CPB variables, while Table 3 summarizes model performance indices. For each scatter plot in Figure
58
59
60
61
62
63
64
65

1
2
3 4, the ordinary least-squares regression slopes and y-intercepts—where better estimation is
4
5
6 reflected by a slope close to 1 and a y-intercept close to 0—are displayed. The slopes and intercepts
7
8
9 (panels a–e) are as follows: (a) slope = 1.00, intercept = 4.00×10^{-5} ; (b) slope = 1.22, intercept
10
11
12 = -1.88 ; (c) slope = 0.92, intercept = 0.68; (d) slope = 0.44, intercept = 4.18; (e) slope = 0.48,
13
14
15 intercept = 4.13. A GLM fitted to the standardized predictor matrix $\mathbf{X} \in \mathbb{R}^{N \times P}$ achieved a coefficient
16
17
18 of determination of $R_{\text{GLM}}^2 = 0.429$, adjusted $R_{\text{GLM}}^2 = 0.369$, and $\text{MSE}_{\text{GLM}} = 0.797$. However,
19
20
21 multicollinearity in the input matrix was confirmed, resulting in a lack of full rank. Applying the same
22
23
24 GLM to PCA-transformed predictors $\mathbf{X}_{\text{PCA}} \in \mathbb{R}^{N \times P_{\text{PCA}}}$ yielded $R_{\text{PCR}}^2 = 0.371$ (adjusted $R_{\text{PCR}}^2 =$
25
26
27 0.352) and $\text{MSE}_{\text{PCR}} = 0.808$. SVR on \mathbf{X} with optimized hyperparameters $C = 2.279$, $\gamma = 16.81$,
28
29
30 and $\varepsilon = 0.594$ yielded $R_{\text{SVR}}^2 = 0.498$ and 10-fold CV $\text{MSE}_{\text{SVR}} = 0.517$, outperforming the GLM.
31
32
33
34 When SVR was applied to \mathbf{X}_{PCA} with $C = 1.662$, $\gamma = 14.28$, and $\varepsilon = 0.0084$, performance
35
36
37 decreased to $R_{\text{SVR}_{\text{PCA}}}^2 = 0.39$ and CV $\text{MSE}_{\text{SVR}_{\text{PCA}}} = 0.52$. The MLP model, tuned via nested 5-fold
38
39
40 CV, selected one hidden-layer architecture of 4 neurons, with Resilient Backpropagation, a learning
41
42
43 rate of 10^{-2} , and early stopping tolerance of 10. On \mathbf{X} , this configuration yielded $R_{\text{MLP}}^2 = 0.332 \pm$
44
45
46 0.033 and $\text{MSE}_{\text{MLP}} = 0.669 \pm 0.086$. When applied to \mathbf{X}_{PCA} , the MLP achieved $R^2 = 0.352 \pm 0.054$
47
48
49 and $\text{MSE} = 0.646 \pm 0.068$. The coefficients of determination from both MLP configurations were
50
51
52
53 lower than those from SVR. Previously published Hb_{CPB} prediction formulas yielded $R^2 = 0.3225$,
54
55
56
57 $\text{MSE} = 2.98$ based on (19), and $R^2 = 0.3245$, $\text{MSE} = 1.48$ based on (18), both substantially lower
58
59
60
61
62
63
64
65

1
2
3 than SVR performance.
4
5

6 Figure 5 displays the Bland–Altman analysis results. The bias between predicted and measured
7
8
9 GLM values was -5×10^{-7} g/dL (95% CI: -0.0654 to 0.0654 g/dL), with 95% LoA of [-1.57 g/dL,
10
11
12 1.57 g/dL]. For SVR, the bias was -0.0028g/dL (95% CI: -0.0619 to 0.0562g/dL) with LoA of [-1.41
13
14
15 g/dL, 1.41 g/dL]. For MLP, the bias was -0.029 g/dL (95% CI: -0.0963 to 0.0375g/dL), and LoA was
16
17
18 [-1.63 g/dL, 1.57 g/dL]. In contrast, the conventional **methods** had a bias of -1.33 g/dL (95% CI: -
19
20
21 1.41 to -1.24 g/dL) and LoA of [-3.49 g/dL, 0.831 g/dL] **for the reference (11) and a bias of -0.64**
22
23
24 **g/dL (95% CI: -0.72 to -0.555 g/dL) and LoA of [-2.67g/dL, 1.39g/dL] for the Equation (v),**
25
26
27 **respectively.** The 95% CI for bias did not include zero, indicating systematic error. In comparison, CIs
28
29
30 from GLM, SVR, and MLP included zero, suggesting no statistically significant bias and the absence
31
32
33 of systematic error. Biases between measured and predicted Hb_{CPB} from GLM, SVR, and MLP were
34
35
36 not statistically significant (GLM: $p = 1.000, d = -6.5 \times 10^{-7}$; SVR: $p = 0.926, d =$
37
38
39 -0.004 ; MLP: $p = 0.389, d = -0.036$). Conversely, biases from conventional models were
40
41
42 confirmed (reference (11): $p < 0.01, d = -1.21$; Equation (v): $p < 0.01, d = -0.617$).
43
44
45
46
47
48
49
50
51

52 **Discussion**

53
54

55 The GLM demonstrated moderate goodness-of-fit despite the multicollinearity present in the input
56
57
58 parameter matrix $\mathbf{X} \in \mathbb{R}^{N \times P}$. In contrast, implementing PCR resulted in reduced model performance
59
60
61
62
63
64
65

1
2
3 across all evaluation metrics compared with the GLM. In PCR, dimensionality reduction is driven
4
5
6 solely by the variance structure of the predictor matrix; therefore, principal components that explain
7
8
9 substantial variance but have limited correlation with the response variable may still be retained as
10
11
12 leading components. In this study, the first 17 principal components were adopted, collectively
13
14
15 accounting for 90% cumulative explained variance in the predictor matrix. While this approach
16
17
18 preserved most of the predictor information, it may have failed to isolate features most relevant to
19
20
21 Hb_{CPB}. Furthermore, the first principal component accounted for more than 20% of the total variance,
22
23
24 and PC1 and PC2 were primarily composed of weight, height, TBV, BSA, BMI, ideal weight, and
25
26
27 obesity measures, supporting the interpretation that TBV-related factors remain central to Hb_{CPB}
28
29
30 estimation. As a result, the PCR model may not have adequately captured the most predictive features
31
32
33 of Hb_{CPB}, likely contributing to its reduced performance, as reflected in the lower coefficient of
34
35
36 determination and higher MSE compared with the GLM using the full predictor set. Traditional
37
38
39 predictors used to estimate Hb_{CPB}—including body weight, priming volume (static circuit volume),
40
41
42 and preoperative Hb concentration—were embedded within the principal components retained by PCR
43
44
45 and were selected as input features for both the SVR and MLP models. Nevertheless, several of the
46
47
48 principal components used in Hb_{CPB} prediction did not primarily reflect these conventional
49
50
51 parameters. Therefore, Hb_{CPB} cannot be fully explained by body weight, priming volume, and
52
53
54 preoperative Hb concentration alone, as suggested by conventional approaches. At the same time,
55
56
57
58
59
60
61
62
63
64
65

1
2
3 while models using PCs 1–17 showed better performance than the conventional formula, we
4
5
6 cannot presently ascribe the improvement to specific components without dedicated ablation or
7
8
9 feature-contribution analyses. These findings highlight the influence of additional confounding
10
11
12 variables, including age and comorbidities.
13
14

15
16 Although model development and training require computational resources (workstation- or
17
18 cloud-based), inference is lightweight and can be executed on common devices (e.g., a standard
19
20 laptop or smartphone) with low latency and no specialized hardware. This enables point-of-care
21
22 use in the operating room as a simple calculator or as a module integrated into existing perfusion
23
24 workflows. Accordingly, the trained model can be delivered either (i) within the EHR as a point-
25
26 of-care calculator or clinical decision support widget that auto-populates inputs from routinely
27
28 collected pre-CPB data, or (ii) on perfusion hardware (e.g., as a vendor plug-in) to present
29
30 predictions alongside pump parameters. After deployment, the model can operate offline, with
31
32 input variables populated automatically from routine pre-CPB data, minimizing manual entry and
33
34 cognitive load.
35
36
37
38
39
40
41
42
43
44
45
46

47
48 Among the nonlinear models, SVR demonstrated a higher coefficient of determination and lower
49
50 MSE compared with both the GLM and PCR, indicating that SVR had superior predictive power. In
51
52 contrast, applying PCA prior to SVR led to a decreased coefficient of determination, likely due to the
53
54 exclusion of Hb_{CPB} -related features during the PCA process, as these were among the bottom 10% of
55
56
57
58
59
60
61
62
63
64
65

1
2
3 principal components that were subsequently discarded. The MLP model, incorporating Levenberg–
4
5
6 Marquardt–based parameter optimization and early stopping, achieved stable training. When applied
7
8
9 to the standardized predictor matrix, MLP showed slightly lower performance than SVR but still
10
11
12 achieved a moderate goodness-of-fit. Overall, SVR and MLP, which perform nonlinear function
13
14
15 estimation and embed variable selection within the model, outperformed linear models, including
16
17
18 conventional approaches, in predicting outcomes from high-dimensional medical data.
19
20
21

22 Although the best-performing method yielded a modest R^2 , agreement relevant to bedside use
23
24
25 is better captured by Bland–Altman analysis (28, 29): the new methods demonstrated minimal bias
26
27
28 and narrower limits of agreement than the conventional formulas (Figure 5), indicating improved
29
30
31 agreement despite a moderate R^2 compared with previous methods. Finally, while models using
32
33
34 multiple principal components outperformed the conventional formula, we cannot presently ascribe
35
36
37 the improvement to specific components without dedicated ablation or feature-contribution analyses;
38
39
40 this is noted as a limitation and will be addressed in future work.
41
42
43

44 Model development and training require computational resources; these activities can be
45
46
47 conducted cost-effectively on a workstation or in the cloud using open-source machine-learning
48
49
50 frameworks. After training, inference is lightweight, runs on commodity devices without accelerators,
51
52
53 and can be embedded in existing EHR or perfusion-system workflows with minimal latency. From a
54
55
56 clinical perspective, more accurate prediction of Hb_{CPB} at CPB initiation supports optimized
57
58
59
60
61
62
63
64
65

1
2
3 transfusion strategies—reducing the likelihood of both under- and over-transfusion—and may help
4
5
6 lower hemodilution-related risks while standardizing practice across teams. We therefore view the
7
8
9 cost–benefit balance as favorable, while acknowledging that prospective evaluation of clinical
10
11
12 outcomes and a formal cost-effectiveness analysis are warranted in future work.
13
14
15
16
17
18

19 Limitations

20
21 While the proposed models using PCs 1–17 outperformed the conventional formula, we cannot, at
22
23
24 present, attribute this improvement to specific principal components without dedicated ablation or
25
26
27 feature-contribution analyses. We acknowledge this as a limitation and plan to investigate the
28
29
30 contribution of individual components and underlying variables in future work.
31
32
33

34 Although inference is feasible on standard devices, clinical integration remains ongoing. Both
35
36
37 EHR-based and device-based (perfusion console) deployments are feasible, but practical adoption will
38
39
40 require: (A) interoperability with hospital systems and/or perfusion consoles, (B) a user interface
41
42
43 suited to intraoperative workflow, (C) prospective validation with monitoring for model drift, and (D)
44
45
46 security and access control consistent with institutional policies. We plan to evaluate these aspects and
47
48
49 conduct prospective usability and safety assessments prior to routine operating-room adoption.
50
51
52
53
54
55
56
57
58
59
60
61
62

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Funding

The authors received no funding to complete this research.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

Data available upon request from the corresponding author.

Author Contributions

H. H. contributed to the conceptualization and design of the study, performed data analysis, drafted the initial manuscript, and participated in the review and revision of the final version. S. T. conducted statistical analysis, contributed to data interpretation, and assisted in drafting the manuscript. T.K. participated in data analysis and provided a critical review of the manuscript. M. H. contributed to the study's conceptualization and design, provided overall supervision, and critically reviewed and revised the manuscript. All authors read and approved the final manuscript.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Ethics Approval

This study was approved by the Ethics Review Committee of Fujita Health University (Approval No. HM21-381).

Acknowledgement

The authors express their sincere gratitude to Tomoaki Yamashiro, Yusuke Nakamura, and the Clinical Engineering Department staff at Fujita Health University Hospital for their meticulous efforts in data collection.

1
2
3 **References**
4
5
6

- 7 (1) Hirata Y, Hirahara N, Murakami A, Motomura N, Miyata H, Takamoto S. Current status of
8
9 cardiovascular surgery in Japan: analysis of data from Japan cardiovascular surgery database in
10
11 2015, 2016. 1. congenital heart surgery. *Jpn J Cardiovasc Surg.* 2019; 48(1): 1-5 (in Japanese).
12
13
14
15
16 (2) Abe T, Nakano K, Hirahara N, Motomura N, Miyata H, Takamoto S, Current Status of
17
18 cardiovascular surgery in Japan: analysis of data from Japan cardiovascular surgery database in
19
20 2015, 2016. 3-Valvular heart surgery. *Jpn J Cardiovasc Surg.* 2019; 48(1): 11-7 (in Japanese).
21
22
23
24
25
26 (3) Shimizu H, Hirahara N, Motomura N, Miyata H, Takamoto S. Current status of cardiovascular
27
28 surgery in Japan: analysis of data from Japan cardiovascular surgery database in 2015, 2016. 4-
29
30 Thoracic aortic surgery. *Jpn J Cardiovasc Surg.* 2019; 48(1): 18-24 (in Japanese).
31
32
33
34
35
36 (4) Saito A, Hirahara N, Motomura N, Miyata H, Takamoto S. Current status of cardiovascular
37
38 surgery in Japan: analysis of data from Japan cardiovascular surgery database in 2015, 2016. 2.
39
40 Isolated Coronary Artery Bypass Surgery. 2-Isolated coronary artery bypass surgery. *Jpn J*
41
42 *Cardiovasc Surg.* 2019; 48(1): 6-10 (in Japanese).
43
44
45
46
47
48 (5) Taniguchi FP, Martins AS. Hemodilution, kidney dysfunction and cardiac surgery. *Einstein (Sao*
49
50 *Paulo).* 2009;7(1 Pt 1):103-7.
51
52
53
54
55 (6) Ranucci M, Conti D, Castelvechio S, et al. Hematocrit on cardiopulmonary bypass and outcome
56
57 after coronary surgery in nontransfused patients. *Ann Thorac Surg.* 2010; 89(1): 11-17.
58
59
60
61
62
63
64
65

- 1
2
3 (7) Karkouti K, Djaiani G, Borger MA, et al. Low hematocrit during cardiopulmonary bypass is
4
5
6 associated with increased risk of perioperative stroke in cardiac surgery. *Ann Thorac Surg*. 2005;
7
8
9 80(4): 1381-1387.
10
11
12 (8) Tsui AKY, Dattani ND, Marsden PA, et al. Reassessing the risk of hemodilutional anemia: Some
13
14
15 new pieces to an old puzzle. *Can J Anaesth*. 2010; 57(8): 779-791.
16
17
18
19 (9) Boer C, Meesters MI, Milojevic M et al. 2017 EACTS/EACTA Guidelines on patient blood
20
21
22 management for adult cardiac surgery. *J Cardiothorac Vasc Anesth*. 2018; 32(1): 88-120.
23
24
25 (10) Ferraris VA, Brown JR, Despotis GJ et al. 2011 Update to The Society of Thoracic Surgeons and
26
27
28 the Society of Cardiovascular Anesthesiologists Blood Conservation Clinical Practice
29
30
31 Guidelines**The International Consortium for Evidence Based Perfusion formally endorses these
32
33
34
35 guidelines. *Ann Thorac Surg*. 2011; 91(3): 944-982.
36
37
38 (11) Murphy GS, Hessel EA, Groom RC et al. Optimal perfusion during cardiopulmonary bypass: an
39
40
41 evidence-based approach. *Anesth Analg*. 2009; 108(5): 1394-1417.
42
43
44 (12) Gravlee GP. Cardiopulmonary bypass: principles and practice. 3rd ed. Philadelphia, PA:
45
46
47 Lippincott Williams & Wilkins; 2008: 416-417.
48
49
50 (13) Hall JE. Guyton and Hall Textbook of Medical Physiology. 14th ed. Philadelphia, PA: Elsevier;
51
52
53
54 2020: 305-310.
55
56
57 (14) Sonoki T, Mushino T, Ueda Y et al., Use guidelines on red blood cell preparation based
58
59
60
61
62
63
64
65

1
2
3 on scientific grounds (the revision third edition). *Jpn J Transfusion Cell Therapy*. 2024; 70(6):
4
5
6 579—596 (in Japanese).
7

8
9
10 (15)Hilberath J, Thomas ME, Smith T et al. Blood volume measurement by hemodilution: association
11
12 with valve disease and re-evaluation of the Allen Formula. *Perfusion*. 2015; 30(4): 305-311.
13
14

15
16 (16)Hasegawa T, Iba Y, Naraoka S et al. Improvement of predicted hematocrit values after the
17
18 initiation of cardiopulmonary bypass in cardiovascular surgery. *J Artif Organs*. 2021; 25: 1-8.
19
20

21
22 (17)Trowbridge C, Stammers A, Klayman M, Brindisi N. A Novel Calculation to Estimate Blood
23
24 Volume and Hematocrit During Bypass. *J Extra Corpor Technol*. 2008; 40(1): 61-4.
25
26

27
28 (18)Muraki R, Hiraoka A, Nagata K et al. Novel method for estimating the total blood volume: the
29
30 importance of adjustment using the ideal body weight and age for the accurate prediction of
31
32 haemodilution during cardiopulmonary bypass. *Interact Cardiovasc Thorac Surg*. 2018; 27(6):
33
34 802-7.
35
36
37

38
39 (19)Epicum M, Dardenne N, Hans G, Larbuisson R, Defraigne JO. Prediction of the post-dilution
40
41 hematocrit during cardiopulmonary bypass. Are new formulas needed? *Perfusion*. 2016; 31(6):
42
43 458-464.
44
45
46
47

48
49 (20)Hibiya M, Kamei T, Kubota S et al. Study profile of the perfusion registry in Japan. *Jpn J Extra*
50
51 *Corpor Technol*. 2018; 45(1): 1-7.
52
53
54

55
56 (21)Du Bois D, Du Bois EF. A formula to estimate the approximate surface area if height and weight
57
58
59
60
61
62

1
2
3 be known. *Arch Int Med*. 1916; 17(6): 863-871.

4
5
6 (22)Japan Society for the Study of Obesity. Guidelines for the management of obesity disease 2022.

7
8
9 Tokyo: Life Science Publishing; 2022: 1-7 (in Japanese).

10
11
12 (23)Shin DA, Lee JC, Shin H, Cho YJ, Kim HC. Point-of-care testing of plasma free hemoglobin and

13
14
15 hematocrit for mechanical circulatory support. *Sci Rep*. 2021; 11(1): 3788.

16
17
18 (24)Maharana K, Mondal S, Nemade B. A review: Data pre-processing and data augmentation

19
20
21 techniques. *Global Transitions Proceedings*. 2022; 3(1): 91-99.

22
23
24 (25)Hancock JT, Khoshgoftaar TM. Survey on categorical data for neural networks. *J Big Data*. 2020;

25
26
27 7: 28.

28
29
30 (26)Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res*. 2003;

31
32
33 3: 1157–1182.

34
35
36 (27)Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos*

37
38
39 *Trans Royal Soc A: Math, Phys Eng Sci*. 2016; 374(2065): 20150202.

40
41
42 (28)J.M. Bland, D.G. Altman, Comparing methods of measurement: why plotting difference against

43
44
45 standard method is misleading, *Lancet*, 1995; 346(8982): 1085-1087.

46
47
48 (29)Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med*

49
50
51 *Res*. 1999; 8: 135–60.

52
53
54 (30)Marquardt DW. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J Soc Indust*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Appl Math. 1963; 11(2): 431-41.

(31) Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: *IEEE Int Conf Neural Netw; 1993; 1: 586-91.*

(32) Hoerl AE, Kennard RW. *Ridge regression: biased estimation for nonorthogonal problems.* *Technometrics.* 1970; 12(1): 55-67.

1
2
3 Supplemental material
4

5 **S1. Generalized Linear Model (GLM)**
6

7
8 For standardized predictors $X \in \mathbb{R}^{N \times P}$ and target $y \in \mathbb{R}^N$ (Hb_{CPB}), we fit a Gaussian-identity GLM:
9

10
11
$$y = \beta_0 + X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I). \quad (\text{s-i})$$

12
13

14 Parameters (β_0, β) were estimated by maximum likelihood (equivalently, least squares under (i)).
15
16

17 Implementation used MATLAB fitglm (MathWorks) with default options.
18
19
20
21
22
23

24 **S2. Support Vector Regression (SVR)**
25
26

27 We used ε -SVR with radial basis function (RBF) kernel as shown in the following equation:
28
29

30
$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \quad (\text{s-ii})$$

31
32
33

34 A ε -SVR model is obtained by solving the convex optimization problem as follows:
35

36
$$\min_{w, b, \xi, \xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i + \xi_i^*, \quad (\text{s-iii})$$

37
38
39
40 subject to
$$\begin{cases} y_i - (w^T \phi(x_i) + b) \leq \varepsilon + \xi_i, \\ (w^T \phi(x_i) + b) - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, n, \end{cases}$$

41
42
43
44

45 where ϕ is the mapping to the feature space, C is the penalty constant (BoxConstraint), and ε is
46
47

48 the width of the epsilon-insensitive zone. The predicted Hb concentration $\hat{y}(x)$ is then expressed as
49
50

51 a linear combination of support vectors, as follows:
52

53
$$\hat{y}(x) = \sum_{i \in \mathcal{S}} \alpha_i K(x_i, x) + b, \quad \mathcal{S} = \{i | \alpha_i \neq 0\}. \quad (\text{s-iv})$$

54
55
56

57 We tuned $(C, \sigma, \gamma, \varepsilon)$ by cross-validation. Implementation used MATLAB fitsvm with RBF kernel;
58
59
60
61
62
63
64
65

1
2
3 expected-improvement-plus criterion in 10-fold CV guided selection.
4
5
6
7
8

9 **S3. Multilayer Perceptron (MLP)**

10
11
12 We used a feed-forward network with $L - 1$ hidden layers. For an input $x \in \mathbb{R}^P$, the layer outputs
13
14
15
16 were defined as:

$$17 \mathbf{h}^{(l)} = \phi(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}), \quad \mathbf{h}^{(0)} = \mathbf{X}. \quad (\text{s-v})$$

18
19
20
21
22 with the final prediction given by:

$$23 \hat{\mathbf{y}} = \mathbf{W}^{(L)}\mathbf{h}^{(L-1)} + \mathbf{b}^{(L)}, \quad (\text{s-vi})$$

24
25
26
27
28 The weights $\{W^{(\ell)}, b^{(\ell)}\}$ were learned by backpropagation using either the Levenberg–Marquardt or
29
30 Resilient Backpropagation algorithm (31, 32), with learning rates 10^{-3} or 10^{-4} and early-stopping
31
32 patience of 10 or 20 epochs. Model selection followed a nested cross-validation scheme with 5 outer
33
34
35 folds and a 5-fold inner grid search over hidden-layer sizes [4], [8], [16], [32], [32, 16], [64, 32], [128,
36
37
38 64], algorithm $\in \{\text{LM, RProp}\}$, learning rate $\in \{10^{-3}, 10^{-4}\}$, and early-stopping tolerance \in
39
40
41 {10,20}. On the full standardized predictor matrix X , the selected configuration was [4] (one hidden
42
43
44 layer with 4 neurons) trained with LM at a learning rate of 10^{-3} , early stopping of 10 epochs, and L2
45
46
47 weight decay of 0.1 (33); five random initializations were run and the model with the lowest validation
48
49
50
51
52
53
54 MSE was retained.
55
56
57
58
59
60
61
62
63
64
65

1
2
3 **S4. Dimensionality Reduction via PCA.**
4
5

6 Continuous variables were z-score standardized and categorical variables were one-hot encoded.
7

8
9 Principal component analysis (PCA) was then applied to the full predictor matrix to obtain the score
10
11 matrix $P \in \mathbb{R}^{N \times P_{\text{PCA}}}$. The retention rule was prespecified as “keep the smallest number of components
12
13 whose cumulative explained variance reaches at least 90%.” Component loadings were computed to
14
15
16
17
18
19 support interpretability.
20
21
22
23
24

25 **S5. Cross-validation, Metrics, and Comparative Baselines.**
26

27
28 For model validation, SVR/MLP used standard train–validation splits with reporting on the final fit;
29
30
31 SVR employed 10-fold cross-validation for hyperparameter selection and cross-validated error
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000