Supplementary material for
"Efficient nonparametric statistical inference on population feature importance using Shapley values"

Brian D. Williamson [*1] **Jean Feng** [*2]

1. Proof of Theorem 1

We consider the decomposition

$$\psi_{m,n} - \psi_{0,0} = (\psi_{0,n} - \psi_{0,0}) + (\psi_{m,0} - \psi_{0,0}) + r_{m,n}, \quad (1)$$

where

$$\psi_{m,0} = \underset{\psi \in \mathbb{R}^{p+1}}{\operatorname{argmin}} E_{Q_m}\left[(Z(S)\psi - v_{0,S})^2\right] \text{ subject to } G\psi = c_0 \quad (2)$$

and $r_{m,n} := (\psi_{m,n} - \psi_{m,0}) - (\psi_{0,n} - \psi_{0,0})$.

We first control the first term in (1). Since Shapley values are defined as a linear combination of the predictiveness vector, let the matrix $B(p)$ encode these weights. Note that this matrix only depends on p. The first row of $B(p)$, denoted $[B(p)]_1$, is given by $[B(p)]_1 = z(\emptyset)$. The matrix entry in row $j = 2, \ldots, p+1$ and column $i = 1, \ldots, 2^p$ is

$$[B(p)]_{ji} := \frac{1}{p}(-1)^{I\{(j-1)\notin s_i\}} \binom{p-1}{|s_i| - I\{(j-1) \in s_i\}}^{-1},$$

where column i corresponds to subset $s_{(i)}$. Then

$$\psi_{0,0} = B(p)v_0 \text{ and}$$
$$\psi_{0,n} := B(p)v_n.$$

Under the collection of conditions implied by (A1)–(A4) and (B1)–(B2) for each subset $s \in \mathcal{S}$, a straightforward application of the functional delta method and Theorem 2 of Williamson et al. (2020) yields that $\psi_{0,n}$ is an asymptotically linear estimator of $\psi_{0,0}$ with influence function given by

$$\phi_{0,1} : o \mapsto B(p)\dot{V}_0(o), \quad (3)$$

where $\dot{V}_0$ is the influence function of v_n and is defined in the main manuscript.

We now control the second term in (1). We use the equivalent weighted least squares formulation of the Shapley values,

$$\psi_{0,0} = \underset{\psi : G\psi = c_0}{\operatorname{argmin}} E_{Q_0}\left(Z(S)\psi - v_{0,S}\right)^2 \text{ and} \quad (4)$$

$$\psi_{m,0} = \underset{\psi : G\psi = c_0}{\operatorname{argmin}} E_{Q_m}\left(Z(S)\psi - v_{0,S}\right)^2. \quad (5)$$

We write the the QR decomposition of $G^\top$ as

$$G^\top = U\begin{bmatrix} R \\ 0 \end{bmatrix} = \begin{bmatrix} U_1 & U_2 \end{bmatrix}\begin{bmatrix} R \\ 0 \end{bmatrix},$$

where U is an orthonormal matrix and R is an upper triangular matrix. U_1 is a 2-column orthogonal matrix corresponding to the column space of $G^\top$ and U_2 is a $(p-1)$-column orthogonal matrix corresponding to its null space. As such, we can reparameterize the constrained least squares problems in (4) and (5) using the vector $\theta \in \mathbb{R}^{p+1}$, where $\psi = U\theta$. The constraint $G\psi = c_0$ implies that

$$\begin{bmatrix} R^\top & 0 \end{bmatrix}\theta = R^\top\theta_1 = c_0, \quad (6)$$

where θ_1 is the first 2 elements of θ. Thus θ_1 is fixed by the constraint, while θ_2 is not constrained. This implies that the solutions to (4) and (5) correspond to θ with θ_1 as the solution to (6) and θ_2 as the solution to the unconstrained least squares problems

$$\theta_{2,0} = \underset{\theta_2 \in \mathbb{R}^{p-1}}{\operatorname{argmin}} E_{Q_0}\left(Z(S)(U_1\theta_1 + U_2\theta_2) - v_{0,S}\right)^2 \text{ and}$$

$$\theta_{2,m} = \underset{\theta_2 \in \mathbb{R}^{p-1}}{\operatorname{argmin}} E_{Q_m}\left(Z(S)(U_1\theta_1 + U_2\theta_2) - v_{0,S}\right)^2.$$

A straightforward application of Theorem 5.23 in van der Vaart (2000) yields that $\theta_{2,m}$ is an asymptotically linear

[*]Equal contribution [1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA [2]Department of Biostatistics, University of Washington, Seattle, WA. Correspondence to: Brian D. Williamson <bwillia2@fredhutch.org>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.

estimator of $\theta_{2,0}$, with

$$\begin{aligned}&\sqrt{m}(\theta_{2,m} - \theta_{2,0})\\ &= -\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\Big[V^{-1}\left\{z(S_j)^\top(U_1\theta_1 + U_2\theta_{2,0}) - v_{0,S_j}\right\}\\ &\qquad \times U_2^\top z(S_j)\Big] + o_P(1),\end{aligned}$$

where $V = U_2^\top Z^\top W Z U_2$. Thus, $\psi_{m,0}$ is an asymptotically linear estimator of $\psi_{0,0}$, i.e.,

$$\sqrt{m}(\psi_{m,0} - \psi_{0,0}) = \frac{1}{\sqrt{m}}\sum_{j=1}^{m}\phi_{0,2}(S_j; v_0) + o_P(1) \quad (7)$$

where $\phi_{0,2}(S; v_0)$ is defined as

$$\phi_{0,2} : s \mapsto -U_2 V^{-1}\left[z(s)^\top\psi_{0,0} - v_{0,s}\right]U_2^\top z(s).$$

Finally, we control the remainder term $r_{m,n}$. By the KKT conditions in the main manuscript, we have that

$$\psi_{m,n} = C_2(Q_m)v_n$$

where $C_2(Q_m)$ is defined as

$$\begin{bmatrix}I_{p+1} & 0\end{bmatrix}\begin{bmatrix}2Z_m^\top W_m Z_m & G^\top\\ G & 0\end{bmatrix}^{-1}\begin{bmatrix}2\sqrt{W_m}\\ e_\emptyset\\ e_N - e_\emptyset\end{bmatrix}$$

and $e_s \in \{0,1\}^{2^p}$ is a one-hot vector for the set s. Likewise, define $C_2(Q_0)$ as

$$\begin{bmatrix}I_{p+1} & 0\end{bmatrix}\begin{bmatrix}2Z^\top W Z & G^\top\\ G & 0\end{bmatrix}^{-1}\begin{bmatrix}2\sqrt{W}\\ e_\emptyset\\ e_N - e_\emptyset\end{bmatrix}.$$

Then

$$\begin{aligned}r_{m,n} &= (\psi_{m,n} - \psi_{m,0}) - (\psi_{0,n} - \psi_{0,0})\\ &= \{C_2(Q_m) - C_2(Q_0)\}(v_n - v_0).\end{aligned}$$

Since the empirical distribution Q_m converges weakly to Q_0, then $C_2(Q_m) \to_p C_2(Q_0)$. Moreover, if (A1)–(A4) and (B1)–(B2) hold for each subset $s \in \mathcal{S}$, then $v_n - v_0 = O_p(n^{-1/2})$. Thus

$$r_{m,n} = o_P(n^{-1/2}). \quad (8)$$

In view of (3), (7), and (8), we can write

$$\begin{aligned}&\sqrt{n}(\psi_{m,n} - \psi_{0,0})\\ &= \sqrt{n}(\psi_{0,n} - \psi_{0,0}) + \sqrt{n}(\psi_{m,0} - \psi_{0,0}) + \sqrt{n}r_{m,n}\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{0,1}(O_i) + \frac{1}{\sqrt{n\gamma_n}}\sum_{i=1}^{n\gamma_n}\phi_{0,2}(S_i; v_0) + o_P(1).\end{aligned}$$

Because O and S are sampled independently and $\gamma_n \to_p \gamma$, then, by Slutsky's theorem, we have that

$$\begin{aligned}&\sqrt{n}(\psi_{m,n} - \psi_{0,0}) \to_d\\ &N\left[0, \mathrm{Cov}\{\phi_{0,1}(O)\} + \gamma^{-1}\,\mathrm{Cov}\{\phi_{0,2}(S; v_0)\}\right]\end{aligned}$$

Finally, note that if $\gamma_n \to \infty$, then the second term in the asymptotic variance is zero.

2. Additional technical details

2.1. Shapley values minimize a weighted least squares problem

Recall the classical Shapley formula: for $j = 1, \ldots, p$,

$$\psi_{0,j} = \frac{1}{p}\sum_{s \in N\setminus\{j\}}\binom{p-1}{|s|}^{-1}(v_{0,s\cup\{j\}} - v_{0,s}).$$

Our goal is to show that the solution x^* to the minimization problem

$$\begin{aligned}&\underset{x\in\mathbb{R}^{p+1}}{\text{minimize}}\ \frac{1}{2}\|\sqrt{W}(Zx - v_0)\|_2^2\\ &\text{subject to } \sum_{j=1}^{p}x_j = v_{0,N} - v_{0,\emptyset} \text{ and } x_0 = v_{0,\emptyset}\end{aligned}$$

satisfies $x_j^* = \psi_{0,j}$ for $j = 1, \ldots, p$.

Since the classical Shapley values in the first display and the solution to the constrained, weighted least squares problem are both linear in v_0, if we can prove that the two values are equivalent for all one-hot vectors $v_{(k)}$ for $k = 1, \ldots, 2^p$, then we will have proved that the two values are equal. Our first result provides the form of the classical Shapley values for a one-hot vector. As in the main manuscript, $s_{(k)}$ refers to the kth ordered subset of $N = \{1, \ldots, p\}$.

Lemma 1. *For $j = 1, \ldots, p$, the classical Shapley value corresponding to one-hot vector $v_{(k)}$ is given by*

$$\begin{aligned}\psi_j(v_{(k)}) =& \frac{1}{p}1\{k \neq 1, k \neq 2^p\}\\ &\times\left[\binom{p-1}{|s_{(k)}|-1}^{-1}1\{j \in s_{(k)}\}\right.\\ &\left.\quad - \binom{p-1}{|s_{(k)}|}^{-1}1\{j \notin s_{(k)}\}\right]\\ &+ \frac{1}{p}(v_{(k),N} - v_{(k),\emptyset}).\end{aligned}$$

Proof. For $j = 1, \ldots, p$, the classical Shapley formula states that

$$\psi_j(v_{(k)}) = \frac{1}{p}\sum_{S\subseteq N\setminus\{j\}}\binom{p-1}{|S|}^{-1}(v_{(k),s\cup\{j\}} - v_{(k),s}).$$

Since $s_{(k)}$ corresponds to $v_{(k)}$, we have that

$$\psi_j(v_{(k)}) = \begin{cases} \frac{1}{p}\binom{p-1}{|s_{(k)}|-1}^{-1}(1-0) & \text{if } j \in s_{(k)} \\ \frac{1}{p}\binom{p-1}{|s_{(k)}|}^{-1}(0-1) & \text{if } j \notin s_{(k)} \end{cases}.$$

For $k = 2, \ldots, 2^p - 1$, we have that $v_{(k),N} - v_{(k),\emptyset} = 0$. Thus, the claim is proved for these values of k. Note that $\psi_j(v_{(1)}) = -1/p$, while for $k = 2^p$, $\psi_j(v_{(2^p)}) = 1/p$ by the definition above. Thus, the claim is proved for all k. □

Our next result provides the solution to the constrained, weighted least squares problem

$$\underset{x \in \mathbb{R}^{p+1}}{\text{minimize}} \|\sqrt{W}(Zx - v_{(k)})\|_2^2 \tag{9}$$
$$\text{subject to } \sum_{j=1}^p x_j = v_{(k),N} - v_{(k),\emptyset} \text{ and } x_0 = v_{(k),\emptyset}.$$

Lemma 2. *For $j = 1, \ldots, p$, the solution to (9) is given by*

$$\begin{aligned} x_j^*(v_{(k)}) =& \frac{1}{p} 1\{k \neq 1, k \neq 2^p\} \\ & \times \left[\binom{p-1}{|s_{(k)}|-1}^{-1} 1\{j \in s_{(k)}\} \right. \\ & \left. - \binom{p-1}{|s_{(k)}|}^{-1} 1\{j \notin s_{(k)}\}\right] \\ & + \frac{1}{p}(v_{(k),N} - v_{(k),\emptyset}). \end{aligned}$$

Proof. For ease of notation, we use x_j^* and $x_j^*(v_{(k)})$ interchangeably. Consider the Lagrangian of (9), given by

$$L(v_{(k)}, x, \lambda) = \|\sqrt{W}(Zx - v_{(k)})\|_2^2 + \lambda^\top(Gx - v_c),$$

where $G = \begin{bmatrix} z(\emptyset) \\ z(N) - z(\emptyset) \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 1 \end{bmatrix}$ and $v_c = \begin{bmatrix} v_{(k),\emptyset} \\ v_{(k),N} - v_{(k),\emptyset} \end{bmatrix}$. Setting the gradient of the Lagrangian equal to zero, we find that x^* must satisfy

$$\begin{aligned} \nabla_x L(v_{(k)}, x, \lambda) &= Z^\top W(Zx - v_{(k)}) + G^\top \lambda \overset{\text{set}}{=} 0 \\ &\Rightarrow 0 = Z^\top W(Zx^* - v_{(k)}) + G^\top \lambda^* \\ \nabla_\lambda L(v_{(k)}, x, \lambda) &= Gx - v_c \overset{\text{set}}{=} 0 \\ &\Rightarrow 0 = Gx^* - v_c. \end{aligned}$$

This yields that

$$Z^\top W Z x^* = Z^\top W v_{(k)} - G^\top \lambda^*. \tag{10}$$

Note that $Z^\top W v_{(k)} = w_{s_{(k)}} z(s_{(k)})$, where $w_{s_{(k)}}$ is the weight for subset $s_{(k)}$, and for ease of notation we set $w_S = \binom{p-2}{|S|-1}^{-1}$, with $w_\emptyset = 1$. We now find the value of λ^*. We denote the index of the first row of $Z^\top W Z x^*$ by zero, to match with x_0^*. Expanding the matrix notation in (10), the first row of (10) states that

$$\begin{aligned} w_{s_{(k)}} - \lambda_1^* &= [Z^\top W Z x^*]_0 \\ &= \left(\sum_{S \in \mathcal{S}} w_S\right) x_0^* + \sum_{j=1}^p \left(\sum_{S \in \mathcal{S} : j \in S} w_S\right) x_j^* \\ &= \left(\sum_{S \in \mathcal{S}} w_S\right) v_{(k),\emptyset} \\ &\quad + \left(\sum_{S \in \mathcal{S} : 1 \in S} w_S\right) (v_{(k),N} - v_{(k),\emptyset}), \end{aligned}$$

where we have made use of the constraints from (9) and the symmetry of the weights. Thus,

$$\begin{aligned} \lambda_1^* = w_{s_{(k)}} &- \left(\sum_{S \in \mathcal{S}} w_S\right) v_{(k),\emptyset} \\ &- \left(\sum_{S \in \mathcal{S} : 1 \in S} w_S\right) (v_{(k),N} - v_{(k),\emptyset}). \end{aligned}$$

For row $\ell = 1, \ldots, p$, we have that

$$\begin{aligned} [Z^\top W Z x^*]_\ell = & \sum_{i=1}^{2^p} 1\{\ell \in s(i)\} w_{s(i)} x_0^* \\ & + \sum_{i=1}^{2^p} 1\{\ell \in s(i)\} w_{s(i)} \sum_{j=1}^p x_j^* 1\{j \in s(i)\}) \\ = & \left(\sum_{S : 1 \in S} w_S\right) x_0^* \\ & + \left(\sum_{S : 1,2 \in S} w_S\right) (v_{(k),N} - v_{(k),\emptyset}) \\ & + \left(\sum_{S : 1 \in S} w_S - \sum_{S : 1,2 \in S} w_S\right) x_\ell^*, \end{aligned}$$

using the symmetry of the weights. Thus, row ℓ of (10) is

$$\begin{aligned} & w_{s_{(k)}} 1\{\ell \in s_{(k)}\} - [G^\top \lambda^*]_\ell = [Z^\top W Z x^*]_\ell \\ \Rightarrow & w_{s_{(k)}} 1\{\ell \in s_{(k)}\} - \lambda_2^* = \\ & \left(\sum_{S : 1 \in S} w_S\right) x_0^* \\ & + \left(\sum_{S : 1,2 \in S} w_S\right) (v_{(k),N} - v_{(k),\emptyset}) \\ & + \left(\sum_{S : 1 \in S} w_S - \sum_{S : 1,2 \in S} w_S\right) x_\ell^*. \end{aligned} \tag{11}$$

Summing (11) across $\ell = 1, \ldots, p$ yields

$$\begin{aligned}\lambda_2^* = \frac{1}{p}\sum_{\ell=1}^{p}\Bigg[& w_{s_{(k)}} 1\{\ell \in s_{(k)}\} - \left(\sum_{S\,:\,1\in S} w_S\right) x_0^* \\ & - \left(\sum_{S\,:\,1,2\in S} w_S\right)(v_{(k),N} - v_{(k),\emptyset}) \\ & - \left(\sum_{S\,:\,1\in S} w_S - \sum_{S\,:\,1,2\in S} w_S\right) x_\ell^*\Bigg] \\ = \frac{1}{p} w_{s_{(k)}} |s_{(k)}| & - \left(\sum_{S\,:\,1\in S} w_S\right) v_{(k),\emptyset} \\ & - \left(\sum_{S\,:\,1,2\in S} w_S + \frac{p-1}{p}\right)(v_{(k),N} - v_{(k),\emptyset}),\end{aligned}$$

where we have again made use of the constraints and the symmetry of W, as well as the difference-of-weights result that $\sum_{S\,:\,1\in S} w_S - \sum_{S\,:\,1,2\in S} w_S = (p-1)$. Plugging this result into (11) and rearranging terms yields that, for each $\ell = 1, \ldots, p$,

$$x_\ell^* = \frac{w_{s_{(k)}}}{p-1}\left\{1\{\ell \in s_{(k)}\} - \frac{1}{p}|s_{(k)}|\right\} + \frac{1}{p}(v_{(k),N} - v_{(k),\emptyset}), \tag{12}$$

where we have again made use of the constraints, the symmetry of W, and the difference-of-weights result.

Note that for $k = 2, \ldots, 2^p - 1$, $v_{(k),N} = v_{(k),\emptyset} = 0$. Thus, for $k = 2, \ldots, 2^p - 1$, and $\ell = 1, \ldots, p$, if $\ell \in s_{(k)}$ then

$$x_\ell^* = \frac{w_{s_{(k)}}}{p-1}\left\{1 - \frac{1}{p}|s_{(k)}|\right\} = \frac{1}{p}\binom{p-1}{|s_{(k)}|-1}^{-1};$$

if $\ell \notin s_{(k)}$ then

$$x_\ell^* = \frac{w_{s_{(k)}}}{p-1}\left\{-\frac{1}{p}|s_{(k)}|\right\} = -\frac{1}{p}\binom{p-1}{|s_{(k)}|}^{-1}.$$

Also, (12) implies that if $k = 1$ then $x_\ell^* = -\frac{1}{p}$, while if $k = 2^p$ then $x_\ell^* = \frac{1}{p}$. Thus,

$$\begin{aligned}x_\ell^*(v_{(k)}) = & \frac{1}{p}1\{k \neq 1, k \neq 2^p\} \\ & \times \left[\binom{p-1}{|s_{(k)}|-1}^{-1} 1\{\ell \in s_{(k)}\} - \binom{p-1}{|s_{(k)}|}^{-1} 1\{\ell \notin s_{(k)}\}\right] \\ & + \frac{1}{p}(v_{(k),N} - v_{(k),\emptyset}),\end{aligned}$$

precisely what we wished to show. □

Combining the results of Lemma 1 and Lemma 2, we have that $x_j^*(v_{(k)}) = \psi_j(v_{(k)})$ for all one-hot vectors $v_{(k)}$, $k = 1, \ldots, 2^p$. Thus, the Shapley values are equivalent to the solution of the weighted least squares problem.

2.2. SHAP values versus SPVIM

Under certain conditions, the mean absolute SHAP value is related to the SPVIM value. Recall that for each feature subset $s \subseteq N$ and corresponding fitted models $\hat{f}_s$, the SHAP value for the jth feature at x is defined as

$$\sum_{s\in N\setminus\{j\}} \frac{1}{p}\binom{p-1}{|s|}^{-1}\{\hat{f}_{s\cup j}(x) - \hat{f}_s(x)\}.$$

Suppose there exists a factor $c > 0$ such that for all feature subsets s, the scaled norm between oracle prediction models $f_{0,s\cup j}$ and $f_{0,s}$ provides a lower bound on the difference between their predictiveness measures, i.e.,

$$\|f_{0,s\cup j} - f_{0,s}\|_1 \lesssim c\,(V(f_{0,s\cup j}, P_0) - V(f_{0,s}, P_0))\,. \tag{13}$$

Then it is easy to show that the mean absolute SHAP value for the oracle model implies large SPVIM values. The lower bound (13) holds if the predictiveness measure V is convex in its first argument, such as when V is the mean squared error.

3. Additional numerical results

In the main manuscript, we ran a 200-variable simulation with a continuous outcome. In Figure 1, we provide the estimated SPVIM value and mean absolute SHAP value for each sample size and feature considered in that analysis. The vertical bars denote the Monte-Carlo error based on 1000 replicates of the experiment for each sample size.

4. Additional details for predicting mortality of patients in the intensive care unit

In this section, we describe our analysis of data on patients' stays in the intensive care unit (ICU) (Silva et al., 2012) in more detail.

First, we computed the minimum, weighted mean, and maximum value of the 15 time-series variables presented in Table 1. The weighted mean corresponds to a linear regression fit to the time series. We then dropped any variable that had a proportion of missing values greater than or equal to 30%. This procedure resulted in a total of 37 features that we used to predict mortality: the summaries of the time-series variables along with all general descriptors measured at admission.

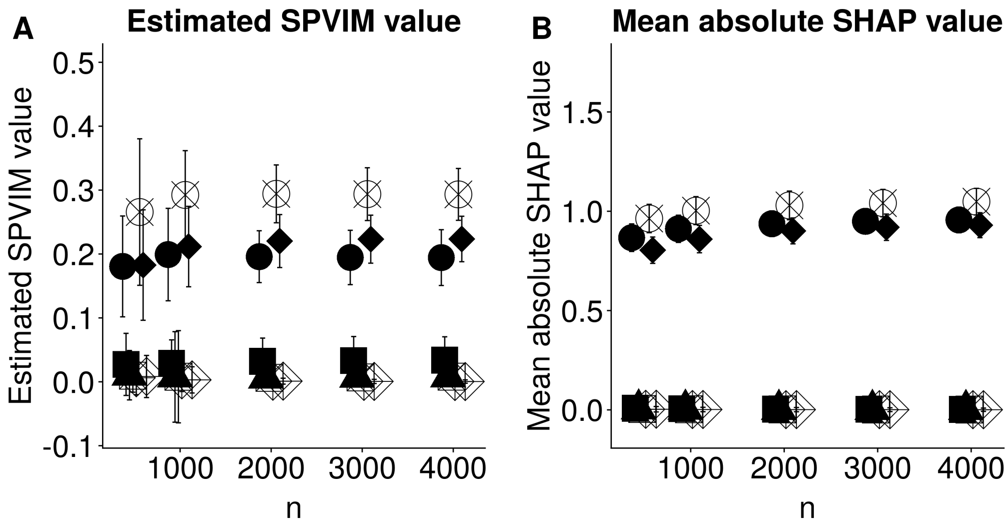

Figure 1. Average estimated SPVIM values (left) and mean absolute SHAP values (right) with Monte-Carlo error bars for the two hundred-variable simulation conducted in the main manuscript. Filled and crossed circles and filled and crossed diamonds denote X_1, X_3, X_5, and X_6 and X_{14}, respectively; filled squares, filled triangles, and crossed squares denote X_{11}, X_{12}, and X_{13}, respectively. The true SPVIM values are approximately $(0.192, 0.291, 0.228, 0.037, 0.01, 0.01, 0)$ for the non-noise features (1, 3, 5, 11, 12, 13, 14) and zero for X_6.

Variable group	Variable name	Summary measure	Included in analysis[1]
Glasgow Coma Scale (GCS)	GCS	min, weighted mean[2], max	Included
Metabolic panel	HCO3 (serum bicarbonate)	min, weighted mean, max	Included
	BUN (blood urea nitrogen)	min, weighted mean, max	Included
	Na (serum sodium)	min, weighted mean, max	Included
	K (serum potassium)	min, weighted mean, max	Included
	Glucose	min, weighted mean, max	Included
Systolic arterial blood pressure (SysABP)	SysABP	min, weighted mean, max	Not included
Complete blood count test	White blood cell count (WBC)	min, weighted mean, max	Included
	Hematocrit (HCT)	min, weighted mean, max	Included
Temperature (Temp)	Temp	min, weighted mean, max	Included
Lactate	Lactate	min, weighted mean, max	Not included
Heart rate (HR)	HR	min, weighted mean, max	Included
Respiration	Respiration rate (RespRate)	min, weighted mean, max	Not included
	Mechanical ventilation (MechVent)	min, weighted mean, max	Not included
	O2 (oxygen)	ratio of FiO2, PaO2	Not included
Urine output	Urine	min, weighted mean, max	Included
General descriptors	Gender	identity[3]	Included
	Height	identity	Not included
	Weight	identity	Included
	Age	identity	Included
	ICU admission type	identity	Included

Table 1. Available features in the MIMIC-II database, along with summary measures computed and an indicator of whether or not the feature was included in the analysis. Impossible values (e.g., ≤ 0 for many variables) were dropped. Summary measures (minimum value, weighted mean, and maximum value) were computed for all time-series variables. Any variable with proportion missing > 0.3 was not included in the analysis, leading to a final analysis dataset with 37 variables.

[1]Features with a proportion of missing values > 0.3 were dropped from the analysis.

[2]Estimated response at mean measurement time from a linear regression of response on time.

[3]All general descriptors were measured a single time, at admission.

References

I Silva, G Moody, DJ Scott, LA Celi, and RG Mark. Predicting in-hospital mortality of icu patients: The physionet/computing in cardiology challenge 2012. In *Computing in Cardiology (CinC), 2012*. IEEE, 2012.

AW van der Vaart. *Asymptotic Statistics*, volume 3. Cambridge University Press, 2000.

BD Williamson, PB Gilbert, N Simon, and M Carone. A unified approach for inference on algorithm-agnostic variable importance. *arXiv*, 2020. https://arxiv.org/abs/2004.03683.