

## Deep learning models for the prediction of intraoperative hypotension

Solam Lee<sup>1,2</sup>, Hyung-Chul Lee<sup>3</sup>, Yu Seong Chu<sup>4</sup>, Seung Woo Song<sup>5</sup>, Gyo Jin Ahn<sup>6</sup>, Hunju Lee<sup>1</sup>, Sejung Yang<sup>4,\*</sup> and Sang Baek Koh<sup>1,\*</sup>

<sup>1</sup>Department of Preventive Medicine, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea, <sup>2</sup>Department of Dermatology, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea, <sup>3</sup>Department of Anaesthesiology and Pain Medicine, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Republic of Korea, <sup>4</sup>Department of Biomedical Engineering, Yonsei University, Wonju, Republic of Korea, <sup>5</sup>Department of Anaesthesiology, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea and <sup>6</sup>Department of Emergency Medicine, Yonsei University Wonju College of Medicine, Wonju, Republic of Korea

\*Corresponding authors. E-mails: [syang@yonsei.ac.kr](mailto:syang@yonsei.ac.kr), [kohhj@yonsei.ac.kr](mailto:kohhj@yonsei.ac.kr)

### Abstract

**Background:** Intraoperative hypotension is associated with a risk of postoperative organ dysfunction. In this study, we aimed to present deep learning algorithms for real-time predictions 5, 10, and 15 min before a hypotensive event.

**Methods:** In this retrospective observational study, deep learning algorithms were developed and validated using bio-signal waveforms acquired from patient monitoring of noncardiac surgery. The classification model was a binary classifier of a hypotensive event (MAP <65 mm Hg) or a non-hypotensive event by analysing biosignal waveforms. The regression model was developed to directly estimate the MAP. The primary outcome was area under the receiver operating characteristic (AUROC) curve and the mean absolute error (MAE).

**Results:** In total, 3301 patients were included. For invasive models, the multichannel model with an arterial pressure waveform, electrocardiography, photoplethysmography, and capnography showed greater AUROC than the arterial-pressure-only models (AUROC<sub>15-min</sub>, 0.897 [95% confidence interval {CI}: 0.894–0.900] vs 0.891 [95% CI: 0.888–0.894]) and lesser MAE (MAE<sub>15-min</sub>, 7.76 mm Hg [95% CI: 7.64–7.87 mm Hg] vs 8.12 mm Hg [95% CI: 8.02–8.21 mm Hg]). For the noninvasive models, the multichannel model showed greater AUROCs than that of the photoplethysmography-only models (AUROC<sub>15-min</sub>, 0.762 [95% CI: 0.756–0.767] vs 0.694 [95% CI: 0.686–0.702]) and lesser MAEs (MAE<sub>15-min</sub>, 11.68 mm Hg [95% CI: 11.57–11.80 mm Hg] vs 12.67 [95% CI: 12.56–12.79 mm Hg]).

**Conclusions:** Deep learning models can predict hypotensive events based on biosignals acquired using invasive and noninvasive patient monitoring. In addition, the model shows better performance when using combined rather than single signals.

**Keywords:** artificial intelligence; biosignals; deep learning; digital medicine; hypotension; intraoperative hypotension; perioperative medicine

### Editor's key points

- Intraoperative hypotension has been associated with risk of postoperative organ dysfunction.
- In this retrospective observational study of data from 3301 patients, the authors applied deep learning algorithms to develop real-time predictions before a hypotensive event using biosignal waveforms in noncardiac surgery.
- Deep learning models predicted hypotensive events based on biosignals acquired using routine invasive and noninvasive patient monitors.

Intraoperative hypotension is associated with poor patient outcomes, including post-surgical acute kidney injury, myocardial infarction, long-term patient outcome, and mortality.<sup>1–4</sup> Risk factors include male sex, old age, ASA physical status 4,<sup>5</sup> emergency surgery, and low pre-induction BP.<sup>6</sup> There are no universal criteria to define intraoperative hypotension,<sup>7,8</sup> and there is no consensus regarding whether intraoperative hypotension should be defined based on an absolute threshold (e.g. MAP of 65 mm Hg) or a relative difference from baseline MAP.<sup>7,9</sup>

Prolonged hypotension is associated with poor post-operative outcomes. Recent studies have shown that even for short duration, hypotensive events are associated with increased risk of postoperative complications.<sup>2</sup> Although the aforementioned risk factors can provide epidemiological information regarding the risk of intraoperative hypotension, they do not help reduce its incidence or duration. Actionable real-time prediction models have been developed recently by analysing parameters obtained during patient monitoring<sup>10,11</sup> to predict future hypotension. However, there are currently two clinical trials reporting conflicting results on whether their use leads to better patient outcome.<sup>12,13</sup>

These previous studies<sup>10,11</sup> used features extracted from arterial pressure waveforms as a single data source. However, haemodynamic changes are associated with alteration of physiological profiles, including the electrocardiogram and respiratory patterns.<sup>14,15</sup> Photoplethysmography and electrocardiography can be used to estimate BP, preload dependency, or arterial impedance.<sup>16,17</sup> End-tidal carbon dioxide is reduced in a hypotensive patient owing to reduced cardiac output, and is positively correlated with BP.<sup>18</sup>

We postulated that better prediction of intraoperative hypotension can be made by analysing not only the arterial pressure waveform, but also other biosignals routinely acquired during surgery, and that reliable predictions can be made without invasive arterial pressure monitoring. This study aimed to develop and validate deep learning algorithms to predict intraoperative hypotension by analysing multiple waveforms acquired during surgery through invasive or noninvasive patient monitoring.

## Methods

### Data source and study approval

This retrospective observational study adhered to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis statement of reporting.<sup>19</sup> The data set consisted of biosignal waveforms obtained from the VitalDB<sup>20</sup> database (<http://vitaldb.net/data-bank>), which included 6388 patients who underwent noncardiac surgical

procedures between June 2016 and August 2017 at Seoul National University Hospital, Seoul, South Korea, a tertiary referral centre (Fig. 1a). Data collection was approved by the institutional review board (IRB) of Seoul National University Hospital (H-1408-101-605) and registered at <https://clinicaltrials.gov/ct2/show/NCT02914444>. A waiver of study approval was granted by the IRB of Yonsei University Wonju Severance Christian Hospital because of the retrospective nature and the use of de-identified data (CR320318).

### Participants

VitalDB contains all information collected from monitoring devices used intraoperatively, including patient monitors, anaesthesia machine, electroencephalogram, infusion pumps, cardiac output monitor, and cerebral/peripheral oximeter. Only data on arterial pressure, electrocardiography (lead II), photoplethysmography, and capnography waveform, which are monitored in most surgeries, were used in this study to improve generalisability. Only patients for whom all four waveforms were available were included. Eligible patients were randomly divided into a training data set to develop models (60%), a validation data set to fine-tune parameters (10%), and a test data set to assess model performance (30%).

### Data quality control

The VitalDB data set is fully accessible and open data that can be used and validated by researchers. The arterial trace and other signals were extracted and recorded from the analogue port of the approved patient monitor (Tram module of Solar™ 8000 patient monitor; GE Healthcare, Wauwatosa, WI, USA). We performed a self-quality check by comparing BPs calculated from the recorded arterial pressure waveform and those extracted directly from the patient monitor. Mean absolute error (MAE) was 4.93 mm Hg for systolic BP (SBP), 4.36 mm Hg for diastolic BP (DBP), and 3.94 mm Hg for MAP. Thus, data quality was adequate for our study (MAE <5 mm Hg).

### Data preparation

Original waveform data obtained from the database were resampled at a rate of 100 Hz. Using peak detection algorithms, each cardiac cycle unit was extracted from the arterial pressure waveform. For removal of nuisance signals and artifacts, any segment in which cardiac cycle was exceptionally slow, fast, or undetectable was excluded. Segments with a MAP of <20 or >160 mm Hg were also excluded.

Hypotensive events were selected as rhythm segments of any length, with hypotension lasting >1 min. The 30 s input segments for prediction of hypotensive events were 30 s rhythms occurring 5, 10, and 15 min before each event. In contrast, non-hypotensive events were selected as rhythm segments of any length, in which non-hypotension lasted >20 min. The 30 s input segments were selected for 5, 10, and 15 min predictions within the event. Up to two non-hypotensive inputs were selected from one non-hypotensive segment to balance data set classes.

### Model building

Details of the deep learning method can be found in Supplementary Method S1. In deep learning, complex hierarchical representations are learned from input data with multiple

abstraction levels. Models were developed and validated to perform classification and regression tasks (Fig. 1a). The classification model was a binary classifier of a hypotensive event or a non-hypotensive event occurring after 5, 10, and 15 min, by analysing the present biosignal waveforms. A hypotensive event was defined as hypotension ( $\text{MAP} \leq 65 \text{ mm Hg}$ ) lasting  $>1$  min, whilst a non-hypotensive event was defined as non-hypotension ( $\text{MAP} > 65 \text{ mm Hg}$ ) stable for  $>20$  min.<sup>10,11</sup> The primary task of the regression model was to estimate MAP as a continuous value after 5, 10, and 15 min.

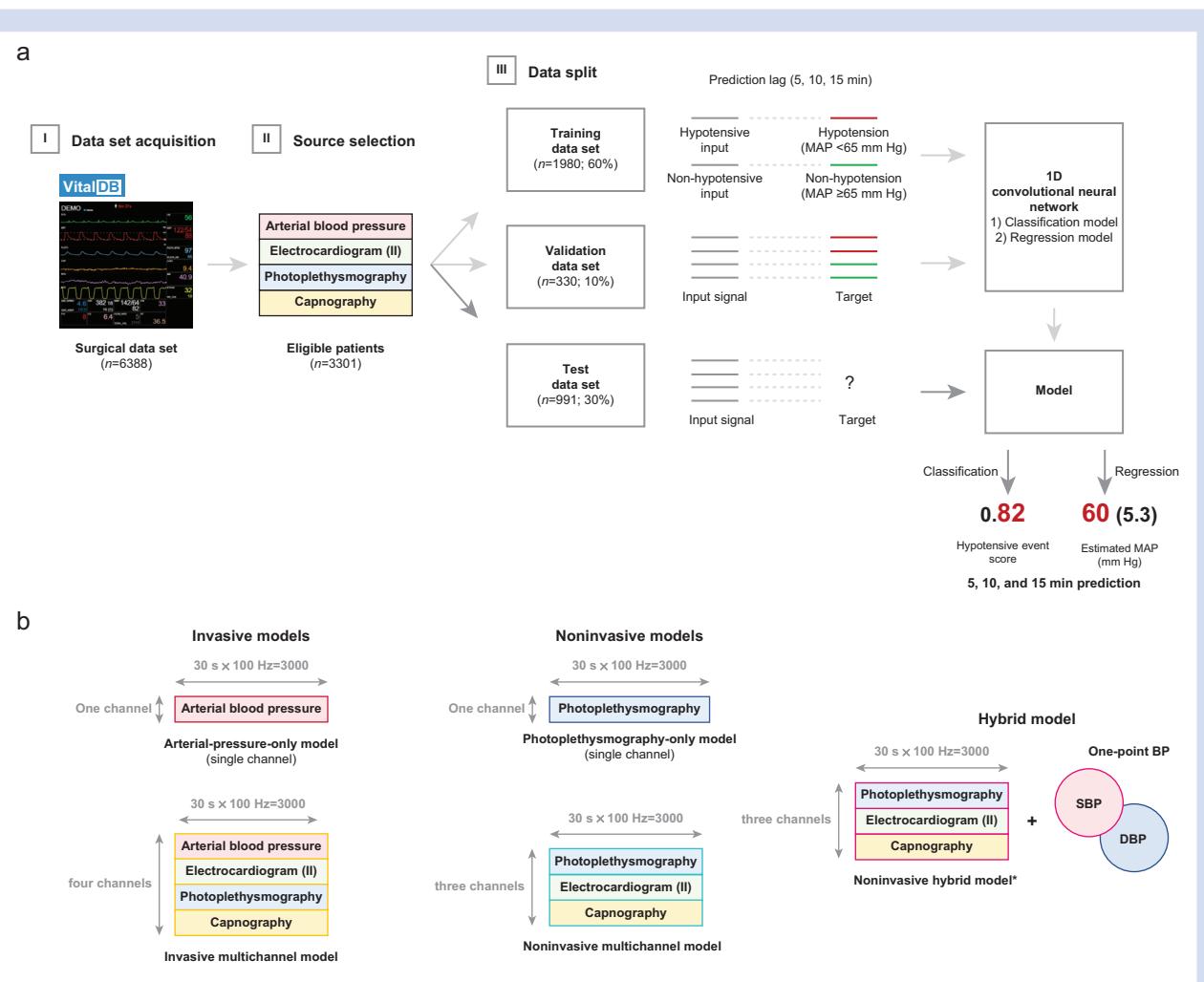
The invasive models included arterial pressure waveforms as input signals (Fig. 1b), whilst the noninvasive models did not. The invasive models consisted of a model, in which only the arterial pressure waveform was used as a single input signal, and a multichannel model, in which arterial pressure, electrocardiography, photoplethysmography, and capnography were used. The noninvasive model consisted of a model, in which the photoplethysmography waveform was used as a

single input signal, and a multichannel model, in which the other signals (except arterial pressure) were used.

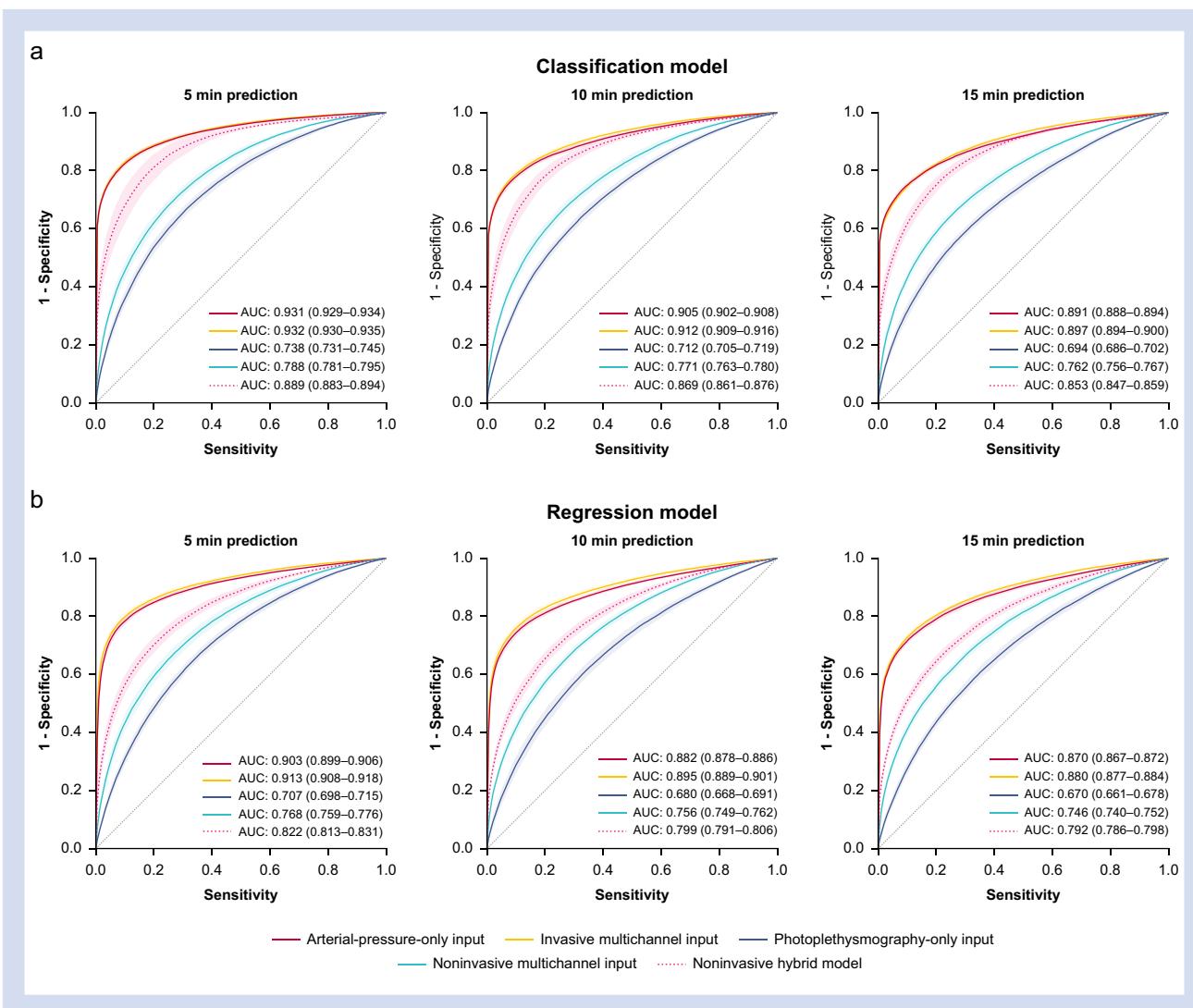
The noninvasive model was expected to have a lower prediction accuracy because of the absence of an arterial pressure waveform. Therefore, we developed a hybrid model, in which one-point SBP and DBP were provided as adjunct information in the noninvasive model (Fig. 1b). This was based on a logistic regression model for the classification task and a linear regression model for the regression task, which analyses the final output calculated from the noninvasive multichannel model and the one-point SBP and DBP calculated from the arterial pressure waveform (although the waveform itself was not used as an input signal).

## Outcome measurement

The primary outcome was the area under the receiver operating characteristic (AUROC) curve for the classification models and the MAE for the regression models. The secondary



**Fig 1.** Study design and models. (a) The study used data from VitalDB, which consists of high-fidelity intraoperative biosignals acquired from 6388 surgical patients. In total, 3301 eligible patients were assigned to the training, validation, and test data sets. The one-dimensional convolutional neural network (CNN) was trained and validated for performing the classification and regression tasks for hypotension prediction. (b) Invasive models, noninvasive models, and hybrid model. DBP, diastolic BP; SBP, systolic BP.



**Fig 2.** Receiver operating characteristic curves. The receiver operating characteristic curve shows the performance of the deep neural networks based on the input signals and prediction lags for intraoperative hypotensive events. The line and shade show the point estimates and 95% confidence intervals for the applied thresholds for the model output. (a) Classification model. (B) Regression model based on binary classification of the estimated MAP by applying variable thresholds. AUC, area under the curve.

outcomes included the sensitivity, specificity, positive predictive value, and negative predictive value with the following cut-off points: (i) a cut-off point with a sensitivity of 90% to predict a future hypotensive event, and (ii) an optimal cut-off point calculated using Youden's index. The intra-class correlation coefficient was calculated amongst the observed MAP and the estimated MAP from the regression models.

### Statistical analysis

Descriptive statistics were used to describe patient characteristics and expressed as means (standard deviation) or absolute numbers (proportion) as appropriate. Characteristics of the study populations were presented based on the experiment with a median AUROC value for the invasive multichannel model. All statistics were reported with the point estimates and 95% confidence intervals. Uncertainty was

calculated from 16 repeated experiments with data sets individually separated using a pseudo-number generator. Python 3.7.0 (Python Software Foundation, Wilmington, DE, USA) and R version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria) were used for signal pre-processing, model development and validation, visualisation, and statistical analysis.  $P<0.05$  was considered statistically significant.

### Sensitivity analysis

In the original analysis, the cut-off point for defining hypotension was MAP <65 mm Hg. Although this criterion has been commonly used previously,<sup>10–13</sup> it would not be a legitimate target for critically ill or older patients.<sup>21</sup> Therefore, we performed sensitivity analyses, where hypotension was defined as MAP <55 mm Hg. The other settings were preserved as in the original analysis.

**Table 1** Performance of the classification models. AUROC, area under the receiver operating characteristic; NPV, negative predictive value; PPV, positive predictive value.

Model	Value (95% confidence interval)					
	5 min prediction		10 min prediction		15 min prediction	
Operating point	Sensitivity=specificity	Sensitivity of 90%	Sensitivity=specificity	Sensitivity of 90%	Sensitivity=specificity	Sensitivity of 90%
Arterial-pressure-only model						
AUROC	0.931 (0.929–0.934)	90.0 (90.0 –90.0)	0.905 (0.902–0.908)	90.0 (90.0 –90.0)	0.891 (0.888–0.894)	90.0 (90.0 –90.0)
Sensitivity (%)	85.6 (85.3–86.0)	90.0 (90.0 –90.0)	82.9 (82.6–83.2)	90.0 (90.0 –90.0)	81.2 (80.9–81.5)	90.0 (90.0 –90.0)
Specificity (%)	85.6 (85.3–85.9)	75.8 (74.7 –77.0)	82.9 (82.6–83.2)	63.5 (62.2 –64.9)	81.1 (80.9–81.4)	61.9 (60.8 –63.0)
PPV (%)	78.3 (77.7–78.9)	69.4 (68.2 –70.6)	74.1 (73.5–74.7)	59.4 (58.4 –60.4)	70.8 (70.2–71.4)	54.7 (53.6 –55.8)
NPV (%)	90.7 (90.5–91.0)	92.6 (92.4 –92.8)	89.1 (88.8–89.5)	91.5 (91.2 –91.8)	88.5 (88.2–88.8)	91.2 (90.9 –91.4)
Invasive multichannel model						
AUROC	0.932 (0.930–0.935)	90.0 (90.0 –90.0)	0.912 (0.909–0.916)	90.0 (90.0 –90.0)	0.897 (0.894–0.900)	90.0 (90.0 –90.0)
Sensitivity (%)	85.8 (85.5–86.1)	90.0 (90.0 –90.0)	83.5 (83.1–83.9)	90.0 (90.0 –90.0)	81.4 (81.1–81.7)	90.0 (90.0 –90.0)
Specificity (%)	85.8 (85.5–86.1)	76.6 (75.6 –77.6)	83.4 (83.1–83.8)	67.8 (66.6 –69.0)	81.4 (81.1–81.6)	58.0 (56.7 –59.3)
PPV (%)	78.6 (77.9–79.2)	70.1 (68.9 –71.3)	74.9 (74.1–75.6)	62.3 (61.2 –63.5)	71.1 (70.3–71.8)	57.1 (56.0 –58.2)
NPV (%)	90.9 (90.6–91.1)	92.6 (92.4 –92.8)	89.5 (89.3–89.8)	91.5 (91.2 –91.8)	88.6 (88.4–88.9)	91.2 (90.9 –91.4)
Photoplethysmography-only model						
AUROC	0.738 (0.731–0.745)	90.0 (90.0 –90.0)	0.712 (0.705–0.719)	90.0 (90.0 –90.0)	0.694 (0.686–0.702)	90.0 (90.0 –90.0)
Sensitivity (%)	67.8 (67.3–68.2)	90.0 (90.0 –90.0)	65.7 (65.2–66.1)	90.0 (90.0 –90.0)	64.2 (63.8–64.7)	90.0 (90.0 –90.0)
Specificity (%)	67.7 (67.3–68.2)	33.8 (32.6 –35.0)	65.6 (65.1–66.9)	29.4 (28.3 –30.5)	64.2 (63.7–64.6)	25.5 (24.7 –26.4)
PPV (%)	56.6 (55.4–56.8)	45.3 (44.5 –46.1)	53.0 (52.1–53.8)	43.0 (42.2 –43.7)	50.2 (49.4–51.3)	40.5 (39.7 –41.3)
NPV (%)	77.5 (76.9–78.1)	84.7 (84.2 –85.2)	76.4 (75.8–76.9)	83.3 (82.7 –83.4)	76.1 (75.6–76.6)	81.9 (81.5 –82.4)
Noninvasive multichannel model						
AUROC	0.788 (0.781–0.795)	90.0 (90.0 –90.0)	0.771 (0.763–0.780)	90.0 (90.0 –90.0)	0.762 (0.756–0.767)	90.0 (90.0 –90.0)
Sensitivity (%)	71.5 (71.0–72.0)	90.0 (90.0 –90.0)	70.0 (69.5–70.5)	90.0 (90.0 –90.0)	69.5 (69.1–69.9)	35.9 (34.9 –37.0)
Specificity (%)	71.5 (71.0–72.0)	42.9 (41.9 –44.0)	70.0 (69.5–70.4)	38.3 (37.0 –40.0)	69.5 (69.1–69.9)	44.2 (43.3 –45.0)
PPV (%)	60.4 (59.6–61.1)	49.0 (48.1 –49.8)	57.9 (57.2–58.7)	46.3 (45.6 –47.0)	56.1 (55.4–57.9)	86.4 (86.0 –86.8)
NPV (%)	80.5 (80.0–81.0)	87.6 (87.3 –87.9)	79.8 (79.2–80.3)	86.6 (86.1 –87.1)	80.2 (79.7–80.7)	90.8 (90.5 –91.0)
Noninvasive hybrid model						
AUROC	0.889 (0.883–0.894)	90.0 (90.0 –90.0)	0.869 (0.861–0.876)	90.0 (90.0 –90.0)	0.853 (0.847–0.859)	90.0 (90.0 –90.0)
Sensitivity (%)	80.7 (79.6–81.8)	90.0 (90.0 –90.0)	79.0 (78.4–79.7)	58.4 (57.0 –60.0)	77.5 (77.0–78.1)	55.3 (54.2 –56.4)
Specificity (%)	80.7 (79.6–81.7)	66.7 (64.8 –68.7)	79.0 (79.3–79.7)	56.1 (55.1 –57.1)	66.0 (65.2–66.8)	53.2 (52.3 –54.0)
PPV (%)	71.8 (70.4–73.1)	62.3 (61.0 –63.6)	69.0 (68.0–70.0)	90.8 (90.5 –91.1)	86.0 (85.4–86.5)	90.8 (90.5 –91.0)

## Data and code availability

The data analysed for the development and validation of the models in this study are available in VitalDB (<https://vitaldb.net>). The code for data quality control, signal pre-processing, model development and validation, and model explanation is available in a repository at <https://doi.org/10.17632/wdpxsyrg2s.2>.

## Results

### Participants and data set

Of the 6388 patients for whom all four biosignal waveforms for arterial BP, electrocardiography, photoplethysmography, and capnography were available, 3301 were eligible for the study (Fig. 1a). The remaining were excluded because of the absence of any of the waveforms (mainly arterial pressure waveform).

Amongst the included patients, 1980 (60%) were assigned to the training data set, 330 (10%) were assigned to the validation data set, and 991 (30%) were assigned to the test data set. Patient and clinical characteristics, and data-set composition for hypotensive events and non-hypotensive events are summarised in [Supplementary Table S1](#).

### Classification model

The receiver operating characteristic curves for the classification models are presented in [Fig. 2a](#). For the invasive models, the multichannel model with all four waveforms showed greater but not statistically significant AUROCs than that of the arterial-pressure-only model for the 5, 10, and 15 min predictions. For the noninvasive models, the multichannel model, in which electrocardiography, photoplethysmography, and capnography were used, significantly outperformed the photoplethysmography-only model for the 5, 10, and 15 min predictions. The other metrics, with the optimal operating threshold balancing the sensitivity and specificity, and the threshold of high sensitivity are shown in [Table 1](#).

### Regression model

[Table 2](#) and [Fig. 3](#) summarise the MAEs and the other metrics of the regression models. For the invasive models, the multichannel model showed significantly lower MAEs than the arterial-pressure-only model. For the noninvasive models, the multichannel model showed significantly lower MAEs than the photoplethysmography-only model. [Fig. 2b](#) shows the ROC curve of the regression models based on binary classification of the estimated MAP as either hypotensive or non-hypotensive by applying variable thresholds. The AUROCs of the regression models were slightly lower than those of the classification models. However, for the invasive and noninvasive models, multichannel models, such as the classification models, showed consistently better performance than the mono-channel models.

### Hybrid model

We anticipated that the invasive models using arterial pressure waveform would outperform the noninvasive models. Therefore, we investigated the improvement in performance using the noninvasive models in combination with the parameters obtained without continuous intra-arterial monitoring ([Fig. 1d](#)). The hybrid model showed significantly higher performances than the noninvasive multichannel model when determining 5, 10, and 15 min hypotensive events for the classification task (AUROCs; [Table 1](#) and [Fig. 2](#)) and the regression task (MAEs; [Table 2](#)). However, they were still outperformed by the invasive models.

### Model visualisation and explanation

[Supplementary Figure S2](#) shows an example of model implementation for a haemodynamically unstable surgical patient. A surge of outputs was observed 15 min before the onset of each hypotensive event. The outputs remained high during the hypotensive event; however, the event did not last 15 min. This discrepancy may have resulted from interventions to normalise the BP (e.g. infusion of a vasoactive agent). Although the output distributions differed across models, they showed similarity in the pattern of alarm activation, whose output value was higher than its operating threshold.

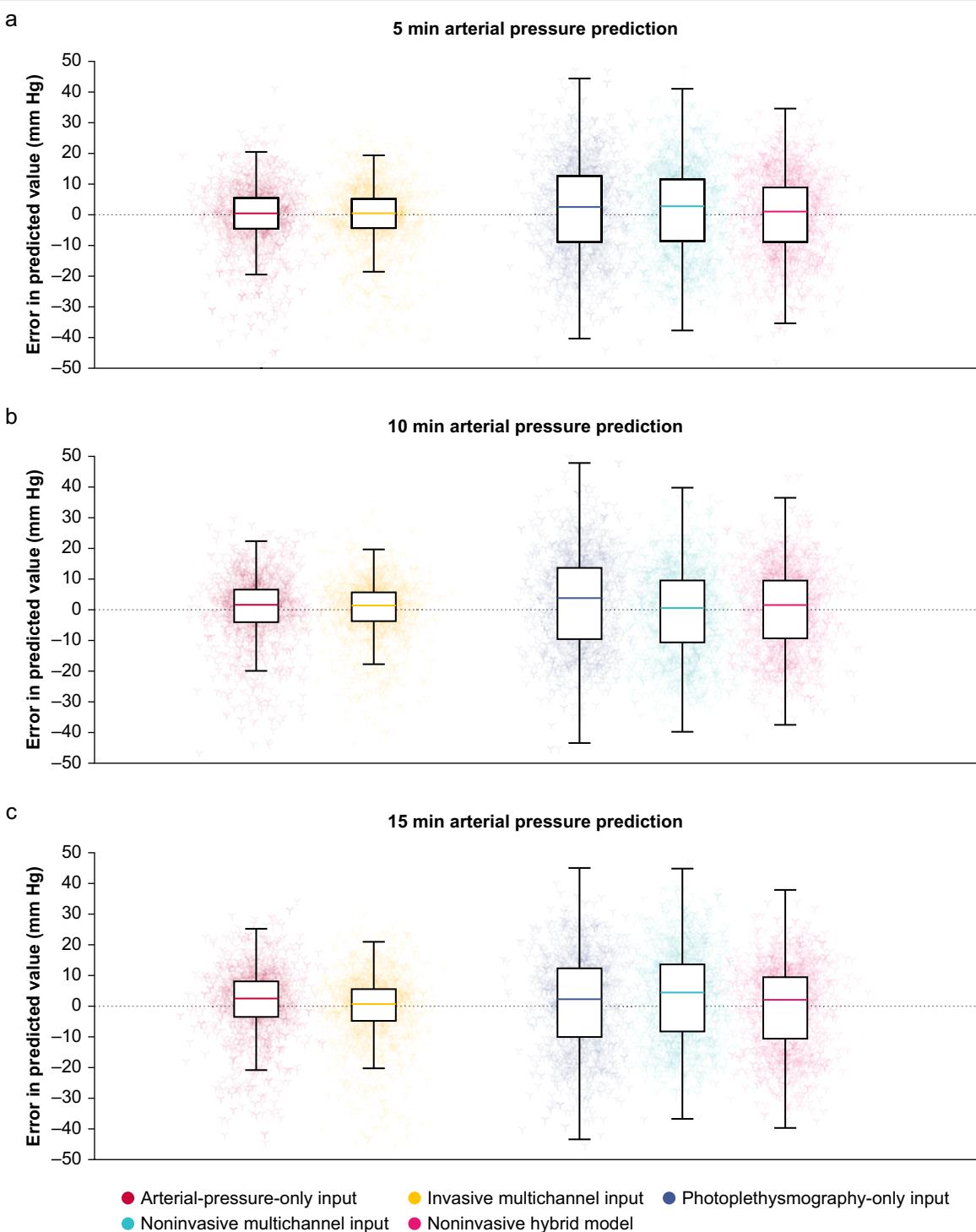
[Fig. 4](#) shows a sensitivity map generated using Grad-CAM<sup>22</sup> to visualise the waveform regions used for decision-making in a hypotensive event. In the invasive multichannel models ([Fig. 4a](#)), although the arterial pressure waveform is the primary source for prediction, the model also detects changes in other signals for better predictions (e.g. breathing pattern alterations, beat-to-beat variability, and electrocardiographic noise). The importance of such changes for decision-making was more evident in the noninvasive models ([Fig. 4b](#)).

### Sensitivity analysis

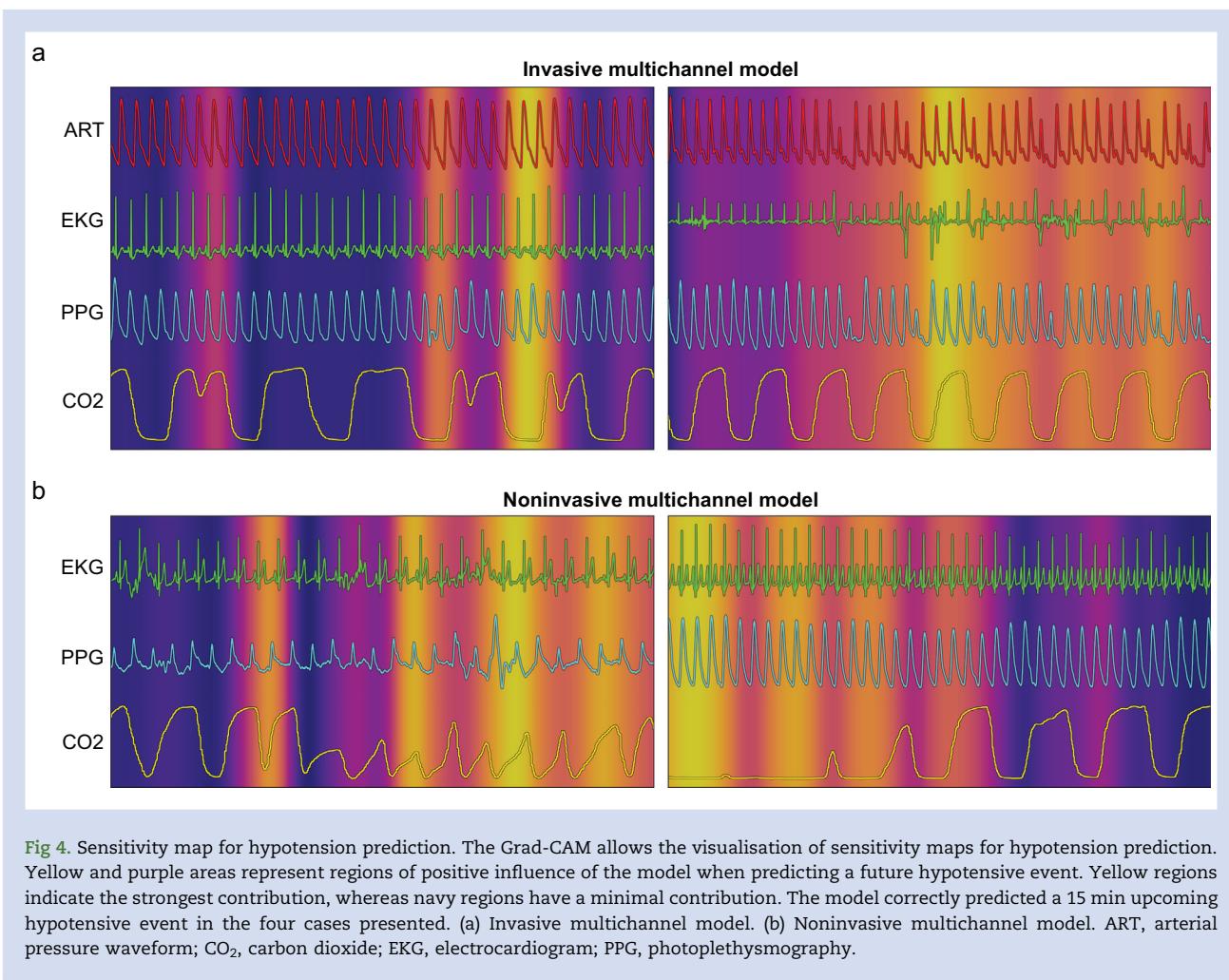
A sensitivity analysis for hypotension defined as MAP <55 mm Hg was performed. The data-set composition and

**Table 2** Performance of the regression models. AUROC, area under the receiver operating characteristic.

Model	5 min prediction	10 min prediction	15 min prediction
<b>Arterial-pressure-only model</b>			
AUROC	0.903 (0.899–0.906)	0.882 (0.878–0.886)	0.870 (0.867–0.872)
Mean absolute error (mm Hg)	7.4 (7.4–7.5)	7.84 (7.73–8.0)	8.1 (8.0–8.2)
Intra-class correlation	0.863 (0.860–0.865)	0.841 (0.836–0.847)	0.830 (0.826–0.834)
<b>Invasive multichannel model</b>			
AUROC	0.913 (0.908–0.918)	0.895 (0.889–0.901)	0.880 (0.877–0.884)
Mean absolute error (mm Hg)	7.1 (7.0–7.2)	7.5 (7.3–7.6)	7.8 (7.6–7.9)
Intra-class correlation	0.877 (0.873–0.880)	0.856 (0.851–0.862)	0.844 (0.839–0.850)
<b>Photoplethysmography-only model</b>			
AUROC	0.707 (0.698–0.715)	0.680 (0.668–0.691)	0.670 (0.661–0.678)
Mean absolute error (mm Hg)	12.3 (12.2–12.4)	12.7 (12.6–12.8)	12.7 (12.6–12.8)
Intra-class correlation	0.502 (0.489–0.516)	0.446 (0.429–0.464)	0.432 (0.418–0.447)
<b>Noninvasive multichannel model</b>			
AUROC	0.768 (0.759–0.776)	0.756 (0.749–0.762)	0.746 (0.740–0.752)
Mean absolute error (mm Hg)	11.5 (11.4–11.6)	11.5 (11.4–11.6)	11.7 (11.6–11.8)
Intra-class correlation	0.595 (0.582–0.608)	0.581 (0.570–0.593)	0.561 (0.550–0.572)
<b>Noninvasive hybrid model</b>			
AUROC	0.822 (0.813–0.831)	0.799 (0.791–0.806)	0.792 (0.786–0.798)
Mean absolute error (mm Hg)	10.4 (10.3–10.5)	10.7 (10.7–10.8)	10.79 (10.7–10.85)
Intra-class correlation	0.701 (0.695–0.707)	0.670 (0.665–0.675)	0.658 (0.652–0.664)



**Fig 3.** Error in the regression models. In the box plots, the boundaries of the box indicate the 25th and 75th percentiles of the error, a coloured line within the box marks the median, a red line within the box marks the mean, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the box indicate the 10th and 90th percentiles. The symbols behind the box plot indicate the errors calculated from each individual data point. (a) Five-minute arterial pressure prediction. (b) Ten-minute arterial pressure. (c) Fifteen-minute arterial pressure.



performance metrics are summarised in [Supplementary Table S2](#). There was prominent class imbalance between the hypotensive event and the non-hypotensive events owing to the strict definition for hypotension. However, the AUROCs of the classification models and the MAEs of the regression models showed minimal differences compared with those of the original analysis.

## Discussion

This study shows that a deep learning model can predict 5, 10, and 15 min hypotensive events based on biosignals acquired using routine invasive and noninvasive patient monitoring. The model showed adequate performance across diverse patients and types of surgery. Improved model performance was observed when using combined rather than single signals.

Several patient characteristics and clinical features are associated with a higher risk of hypotension during surgery.<sup>5</sup> Although identification of these risk factors would help perioperative risk assessment, this cannot prevent or manage hypotensive events occurring during surgery. Although short hypotensive events can cause post-surgical cardiac and renal failure,<sup>1–4</sup> maintaining normal BP can reduce postoperative organ dysfunction.<sup>23</sup> Few studies have used haemodynamic parameters extracted from patient monitoring, such as HR

variability or arterial stiffness.<sup>24–26</sup> Nevertheless, most had limited predictive ability in real time. With our real-time prediction model for future hypotensive events, proactive interventions can be undertaken before the events take place.

Deep learning algorithms are increasingly used for analysing biosignal waveforms to predict medical conditions, such as arrhythmia, valvular heart diseases, hyperkalaemia, and anaemia.<sup>27–30</sup> In our model, prediction of medical conditions was based on detection of changes and signs in biosignals caused by disease, physiological condition, or compensatory mechanisms. Unsupervised learning can benefit from detection of clinically imperceptible and subtle changes that cannot be recognised by human visualisation. Our models are characterised by extraction of unsupervised features from multiple and heterogeneous biosignals. Although there are machine-learning-based approaches for predicting intraoperative hypotension, most are limited by relying only on the arterial pressure waveform. In our study, the AUROCs were consistently greater in multichannel models than in single-channel models. This could be attributed to additional predictive information extracted from other signals, even if not directly indicative of arterial BP. This advantage was evident in the noninvasive models, for which no direct information was available.

As expected, the invasive models showed greater performance than the noninvasive models. However, continuous arterial pressure monitoring requires invasive arterial catheterisation. Although there is a lack of epidemiological studies or standard indications of arterial catheterisation for surgical patients, it is performed selectively for high-risk patients owing to its invasive nature. Therefore, when invasive monitoring is not indicated, the noninvasive models may be adopted as a continuous screening adjunct to conventional patient monitoring. The noninvasive models showed considerable performance gain when one-point SBP and DBP were provided (the hybrid model). Although it did not match the performance of the invasive models, a combination with noninvasive BP monitoring may offer more reliable prediction for patients without arterial catheterisation.

We presented both classification and the regression models. Although the AUROCs of the classification models were greater than those of the regression models, the difference was not statistically significant. We believe that the regression models are more clinically applicable owing to the advantage of directly providing a numerical value for MAP. Preparations for intervention may vary depending on the severity of the predicted hypotension. Although our study used an absolute threshold of 65 mm Hg (55 mm Hg for the sensitivity analysis), there is no universal definition for intraoperative hypotension. Whether use of index-based guidance leads to better patient care remains controversial. Therefore, rather than relying on a binary classification, it would be preferable to interpret estimated MAP according to the clinical context for predicting intraoperative hypotension.

Our study has some limitations, such as a single data source and lack of external validation. No additional data set, including multiple waveforms of surgical patients, was available. Nevertheless, we included >3000 patients who underwent general, gynaecological, thoracic, and urogenital surgeries to achieve generalisability. We only included patients who underwent arterial pressure monitoring from the tertiary referral centre that cares for patients with the highest severity in Korea. Therefore, there is a selection bias that could limit generalisability of the findings.

We analysed retrospective data, which did not include detailed records for acute surgery or anaesthesiological interventions (e.g. sudden blood loss, vascular clamp, or vasoactive infusion). Therefore, the model provides an unreliable estimate with the input signals acquired during impending or ongoing hypotensive events, even if they are not the focus of interest of our model. Invasive arterial pressure, considered the gold standard in our study, could be inaccurate, owing to underdamping and resonance.<sup>31</sup> Therefore, the MAE of the regression models may be partially attributable to measurement errors. Further, the hybrid models used one-point SBP and DBP, calculated from the arterial pressure waveform. However, noninvasive BP is not always well calibrated with invasive BP.<sup>32</sup> Therefore, performance could be overestimated compared with that of one-point BP measured using an oscillometric device.

## Authors' contributions

Study conception/design: SL, H-CL, SBK, SY

Data collection: H-CL

Data analysis: SL, H-CL, YSC

Data interpretation: all authors

Statistics: SL

Drafting of article: SL

Critical revision of article: all authors

SY and SBK had full access to all the data in the study and had final responsibility for the decision to submit for publication.

## Acknowledgements

The authors thank the funder of the study, which had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

## Declarations of interest

The authors declare that they have no conflicts of interest.

## Funding

MD–PhD/Medical Scientist Training Program through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare, Republic of Korea.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bja.2020.12.035>.

## Reference

- van Waes JA, van Klei WA, Wijeyesundara DN, van Wolfswinkel L, Lindsay TF, Beattie WS. Association between intraoperative hypotension and myocardial injury after vascular surgery. *Anesthesiology* 2016; **124**: 35–44
- Walsh M, Devereaux PJ, Garg AX, et al. Relationship between intraoperative mean arterial pressure and clinical outcomes after noncardiac surgery: toward an empirical definition of hypotension. *Anesthesiology* 2013; **119**: 507–15
- Devereaux PJ, Chan MT, Alonso-Coello P, et al. Association between postoperative troponin levels and 30-day mortality amongst patients undergoing noncardiac surgery. *JAMA* 2012; **307**: 2295–304
- An R, Pang QY, Liu HL. Association of intra-operative hypotension with acute kidney injury, myocardial injury and mortality in non-cardiac surgery: a meta-analysis. *Int J Clin Pract* 2019; **73**, e13394
- Mayhew D, Mendonca V, Murthy BVS. A review of ASA physical status—historical perspectives and modern developments. *Anesthesia* 2019; **74**: 373–9
- Südfeld S, Brechnitz S, Wagner JY, et al. Post-induction hypotension and early intraoperative hypotension associated with general anaesthesia. *Br J Anaesth* 2017; **119**: 57–64
- Vernooij LM, van Klei WA, Machina M, Pasma W, Beattie WS, Peelen LM. Different methods of modelling intraoperative hypotension and their association with postoperative complications in patients undergoing non-cardiac surgery. *Br J Anaesth* 2018; **120**: 1080–9
- Bijker JB, van Klei WA, Kappen TH, van Wolfswinkel L, Moons KG, Kalkman CJ. Incidence of intraoperative hypotension as a function of the chosen definition: literature definitions applied to a retrospective cohort using automated data collection. *Anesthesiology* 2007; **107**: 213–20
- Saugel B, Reuter DA, Reese PC. Intraoperative mean arterial pressure targets: can databases give us a universally

- valid “magic number” or does physiology still apply for the individual patient? *Anesthesiology* 2017; **127**: 725–6
10. Hatib F, Jian Z, Buddi S, et al. Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. *Anesthesiology* 2018; **129**: 663–74
  11. Davies SJ, Vistisen ST, Jian Z, Hatib F, Scheeren TWL. Ability of an arterial waveform analysis-derived hypotension prediction index to predict future hypotensive events in surgical patients. *Anesth Analg* 2020; **130**: 352–9
  12. Wijnberge M, Geerts BF, Hol L, et al. Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the HYPE randomized clinical trial. *JAMA* 2020; **323**: 1052–60
  13. Maheshwari K, Shimada T, Yang D, et al. Hypotension prediction index for prevention of hypotension during moderate- to high-risk noncardiac surgery. *Anesthesiology* 2020; **133**: 1214–22
  14. Simjanoska M, Gjoreski M, Gams M, Madevska Bogdanova A. Non-invasive blood pressure estimation from ECG using machine learning techniques. *Sensors (Basel)* 2018; **18**: 1160
  15. Magder S. Clinical usefulness of respiratory variations in arterial pressure. *Am J Respir Crit Care Med* 2004; **169**: 151–5
  16. Hosanee M, Chan G, Welykholowa K, et al. Cuffless single-site photoplethysmography for blood pressure monitoring. *J Clin Med* 2020; **9**: 723
  17. Tusman G, Bohm SH, Suarez-Sipmann F. Advanced uses of pulse oximetry for monitoring mechanically ventilated patients. *Anesth Analg* 2017; **124**: 62–71
  18. Kheng CP, Rahman NH. The use of end-tidal carbon dioxide monitoring in patients with hypotension in the emergency department. *Int J Emerg Med* 2012; **5**: 31
  19. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015; **350**: g7594
  20. Lee H-C, Jung C-W. Vital Recorder—a free research tool for automatic recording of high-resolution time-synchronised physiological data from multiple anaesthesia devices. *Sci Rep* 2018; **8**: 1527
  21. Lamontagne F, Richards-Belle A, Thomas K, et al. Effect of reduced exposure to vasopressors on 90-day mortality in older critically ill patients with vasodilatory hypotension: a randomized clinical trial. *JAMA* 2020; **323**: 938–49
  22. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision*; 2017. p. 618–26. Venice
  23. Futier E, Lefrant JY, Guinot PG, et al. Effect of individualized vs standard blood pressure management strategies on postoperative organ dysfunction among high-risk patients undergoing major surgery: a randomized clinical trial. *JAMA* 2017; **318**: 1346–57
  24. Hanss R, Bein B, Weseloh H, et al. Heart rate variability predicts severe hypotension after spinal anesthesia. *Anesthesiology* 2006; **104**: 537–45
  25. Kendale S, Kulkarni P, Rosenberg AD, Wang J. Supervised machine-learning predictive analytics for prediction of postinduction hypotension. *Anesthesiology* 2018; **129**: 675–88
  26. Alecu C, Cuignet-Royer E, Mertes PM, et al. Pre-existing arterial stiffness can predict hypotension during induction of anaesthesia in the elderly. *Br J Anaesth* 2010; **105**: 583–8
  27. Galloway CD, Valys AV, Shreibati JB, et al. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. *JAMA Cardiol* 2019; **4**: 428–36
  28. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 2019; **394**: 861–7
  29. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019; **25**: 65–9
  30. Kwon J-M, Cho Y, Jeon K-H, et al. A deep learning algorithm to detect anaemia with ECGs: a retrospective, multicentre study. *Lancet Digit Health* 2020; **2**: e358–67
  31. Romagnoli S, Ricci Z, Quattrone D, et al. Accuracy of invasive arterial pressure monitoring in cardiovascular patients: an observational study. *Crit Care* 2014; **18**: 644
  32. Seidlerová J, Tůmová P, Rokyta R, Hromadka M. Factors influencing the accuracy of non-invasive blood pressure measurements in patients admitted for cardiogenic shock. *BMC Cardiovasc Disord* 2019; **19**: 150

Handling editor: Hugh C. Hemmings Jr