An Electrocardiographic Deep Learning Algorithm to Predict Post-Operative Mortality and Complications

David Ouyang[1,2], John Theurer[1], Nathan R. Stein[1], J. Weston Hughes[3], Pierre Elias[4,5], Bryan He[3], Neal Yuan[1], Grant Duffy[1], Roopinder K. Sandhu[1], Joseph Ebinger[1], Patrick Botting[1], Eunice Park[6], Jessica T. Soto[7], James E. Tooley[7], Tim Poterucha[4], Jonathan H. Chen[8], Marco Perez[7], Michael Thompson[6], Geoffrey Tison[9], Adler Perotte[4], Euan A. Ashley[7], James Y. Zou[3,10], Nancy R. Cook[11], Sumeet S. Chugh[1,2], Susan Cheng[1] and Christine M. Albert[1]

1. Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center
2. Division of Artificial Intelligence in Medicine, Department of Medicine, Cedars-Sinai Medical Center
3. Department of Computer Science, Stanford University
4. Milstein Division of Cardiology, Department of Medicine, Columbia University Irving Medical Center
5. Department of Biomedical Informatics, Columbia University Irving Medical Center
6. Enterprise Information Services, Cedars-Sinai Medical Center
7. Department of Medicine, Division of Cardiology, Stanford University
8. Department of Medicine, Division of Bioinformatics Research, Stanford University
9. Department of Medicine, Division of Cardiology, University of California San Francisco,
10. Department of Biomedical Data Science, Stanford University
11. Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital

Correspondence: David.ouyang@cshs.org and Christine.albert@cshs.org

Key Points

- Standard pre-operative risk evaluation relies on few clinical features with limited ability to predict mortality and postoperative complications. A novel deep learning algorithm, PreOpNet, evaluates electrocardiogram (ECG) waveform signals to accurately identify patients at risk for post-operative mortality and cardiovascular complications.
- PreOpNet was validated and performs well in three separate independent healthcare systems and in clinical settings in which patients benefit most from pre-operative risk evaluation.
- Using a new model architecture with fewer parameters than previously designed deep learning models on ECGs, our model accurately predicts post-operative outcomes rapidly such that PreOpNet can be run quickly on a standard hospital workstation.

Summary Paragraph

Perioperative risk evaluation is an integral part of clinical decision-making prior to surgical intervention[1-3]. Numerous scoring systems have been developed to balance surgical risk with the potential benefits of intervention, however, these algorithms have subjective variables, can overestimate risk in low-risk patients, or be cumbersome to apply[4-7]. Electrocardiograms (ECGs) are inexpensive, non-invasive, and rapid diagnostic tests frequently obtained during preoperative evaluation[1,4], and deep learning algorithms applied to ECGs have been shown to identify subtle phenotypes[8,9] and prognosticate long term outcomes[10]. We used 149,348 ECGs from 52,606 patients undergoing inpatient procedures to train PreOpNet, a novel deep learning algorithm to predict post-operative mortality and cardiovascular events. By interpretation of electrocardiogram waveform signals from a single pre-operative 12-lead ECG performed within 30 days prior to procedure, PreOpNet predicted post-operative mortality with an AUC of 0.83 (95% CI 0.78 - 0.88) and major cardiovascular events with an AUC of 0.79 (95% CI 0.75 - 0.83), surpassing standard clinical risk scores and previously published deep learning architectures. At two separate healthcare systems, PreOpNet performed similarly well, predicting post-operative mortality with an AUC of AUC of 0.79 (95% CI 0.73 – 0.84) and 0.75 (95% CI 0.74 – 0.76) in separate external test datasets. This performance was achieved with a new model architecture less than 1/50th the size of previous models[11] and efficient enough to be run on a standard computer.

Introduction

Patients undergoing major surgery are at risk for perioperative cardiac complications, including myocardial infarction, heart failure, arrhythmias, cardiac arrest, and death[12,13]. To balance surgical risk with the potential benefits of intervention, pre-operative risk assessment aims to identify patients at increased risk of complications[1,14]. Numerous scoring systems have been developed which integrate clinical features, diagnoses, and functional status to assist in risk stratification[12,15,16]. However, these algorithms have subjective components, can overestimate risk in low-risk patients, or require tedious manual data entry of clinical information[2,6,17].

Electrocardiograms (ECGs) are inexpensive, non-invasive, and rapid diagnostic tests frequently obtained during preoperative evaluation[4]. Society guidelines recommend routine preoperative ECG testing in patients with cardiovascular comorbidities and patients undergoing intermediate-to-high risk surgeries[1,4]. Deep learning algorithms have been recently applied to medical imaging and data to achieve high precision interpretations as well as identify information beyond the interpretation of human experts[18-20]. Deep learning analysis of ECG waveforms have had potentially promising performance in prognosticating outcomes[10], identifying subclinical disease[8,21], and identifying systemic phenotypes not traditionally associated with ECGs[22,23]. Given prior performance in identifying occult arrhythmias[9], ventricular dysfunction[8], anemia[22], age[23], and other factors associated with perioperative risk, deep learning algorithm interpretation of pre-operative ECGs could potentially prognosticate postoperative outcomes with higher accuracy than traditional risk calculators.

To overcome current limitations in assessment of postoperative outcomes, we designed, trained, and validated PreOpNet, a deep learning algorithm to predict post-operative cardiovascular events and death based on waveform signals from a single pre-operative 12-lead ECG (Figure 1). To evaluate the performance of our deep learning model architecture, we compare the predictions of PreOpNet with conventional risk calculators as well as to a previously published deep learning architecture[8]. Additionally, PreOpNet's performance was also evaluated in two separate healthcare systems and in common clinical settings which benefit most from pre-operative risk evaluation.

Results

Deep learning model design

Inspired by prior literature regarding lightweight deep learning model architecture design and neural architecture search[24,25], we designed a novel neural network with atrous convolutions and residual connections with bottleneck layers for ECG interpretation. Optimizing for model runtime and minimizing model complexity, we designed our neural network architecture to start with an initial 1-dimensional atrous convolutional layer to downsample ECG waveforms prior to downstream convolutional layers[26]. The atrous convolution's dilation and step size was grid-searched by hyperparameter tuning for optimal AUC with all other hyperparameters held constant.

Downstream, we incorporated serial layers with an inverted residual structure where the input and output are bottleneck layers with an intermediate expansion layer[24]. In each set of expansion layers with bottleneck layers preceding and succeeding, the number of input channels gradually increased to allow for integration of information across ECG leads. From initial 12 input channels representing each ECG lead, the input channels for each layer increased to 32, 16, 24, 40, 80, 112, 192, then 320 input channels before average pooling prior to the final fully connected layer. For computational efficiency, each convolutional layer only relied on 1-dimensional convolutions while inflation of output channels allows for integration of information across leads. The number of modules paralleled the design of EfficientNet[25], although further attempts to prune the model showed only slight cost to performance with further reductions in model size and number of layers.

Study population characteristics

We trained PreOpNet with 119,179 ECGs from 42,129 patients, with a mean of 2.8 ECGs per patient (Supplementary Table 1). In our sample of 340,703 inpatient procedures from an academic medical center, there were 2,465 (0.8%) deaths and 3,471 (1.1%) post-procedural major cardiovascular events that occurred among the study sample during 323,897 inpatient hospitalizations. The mean age of the patient at time of procedure was 64.6 (SD 16.0) years. Forty-four percent of the participants were women, and 19.5% had pre-existing coronary artery disease. 78,021 procedures (22.8%) had a pre-operative ECG within 30 days prior to procedure, and a total of 149,348 preoperative ECGs were matched to these procedures (Supplementary Figure 2).

Evaluation of model performance

Using information from a single 12 lead ECG, PreOpNet predicted same-hospitalization postoperative mortality with an AUC of 0.83 (95% CI 0.78 - 0.88). In comparison, AUC for the clinical risk calculator, the revised cardiac risk index (RCRI) score, was 0.58 (95% CI 0.51 – 0.65) for postoperative mortality. For the combined endpoint major adverse cardiovascular events, the PreOpNet had an AUC of 0.79 (95% CI 0.75 - 0.83) and the RCRI score had an AUC of 0.57 (95% CI 0.52 – 0.62). The addition of clinical features in the RCRI score to PreOpNet did not improve the model performance (AUC 0.81; 95% CI 0.76 – 0.86 and 0.79; 95% CI 0.75 - 0.83 for postoperative mortality and major cardiovascular events respectively).

For the RCRI score, a threshold of 2 is a common key decision-making boundary suggesting elevated risk[4]. At this threshold, which encompassed 15% of patients in our data, the RCRI score was associated with an odds ratio of 2.60 (95% CI 1.15 - 4.54) for post-operative mortality and an odds ratio of 2.17 (1.27 - 3.31) for major cardiovascular events in our test dataset. In comparison, patients above a similar percentile of risk (85th percentile) in PreOpNet's prediction had an odds ratio of 8.77 (95% CI 4.82 – 14.98) for post-operative mortality and 7.34 (4.93 - 10.56) for major cardiovascular events as compared to those below this level of risk (Table 1). Local Interpretable Model-agnostic Explanations[29] highlighted the QRS complexes as the most relevant features used for model decision making, and frequently highlighted precordial premature contractions and intraventricular block on the precordial leads (Figure 2).

Generalization to other healthcare systems

To assess the cross-healthcare-system reliability of the model, PreOpNet's prediction of post-operative mortality was additionally tested, without any tuning, on external test datasets of 162,540 ECGs from 101,375 patients from Stanford Healthcare (SHC) and 9,028 ECGs from 9,028 patients from Columbia University Medical Center (CUMC). On the external test dataset from SHC with a mortality rate of 1.3%, PreOpNet predicted same-hospitalization postoperative mortality with an AUC of 0.75 (95% CI 0.74 - 0.76). The external test dataset from CUMC had a mortality rate of 1.6% and PreOpNet predicted mortality with an AUC of 0.79 (95% CI 0.73 - 0.84). The 85th percentile of risk predicted by PreOpNet had an odds ratio of 4.81 (95% CI 4.20 – 5.43) at SHC and an odds ratio of 6.15 (5.44 – 6.92) at CUMC for post-operative mortality.

Model performance in various clinical settings

Our initial training set included both elective and emergent procedures. Preoperative risk stratification is most frequently applied in the elective procedural setting, and for patients undergoing elective procedures, PreOpNet had an AUC of 0.82 (95% CI 0.69 – 0.95) for post-operative mortality and an AUC of 0.78 (95% CI 0.67 – 0.88) for major cardiovascular events. Given the potential utility of PreOpNet in urgent or emergent settings, we performed a secondary analysis on non-elective procedures that occurred in patients admitted through the emergency room. PreOpNet predicted death with an AUC of 0.72 (95% CI 0.69 – 0.75) and major cardiovascular events with an AUC of 0.69 (95% CI 0.66 – 0.71).

Given the low incidence of complications after low-risk surgical procedures, electrocardiograms often are not obtained in low-risk patients undergoing low risk procedures. We performed a secondary analysis of the performance in patients most likely to undergo ECG screening during pre-operative evaluation represented by patients either with known cardiovascular disease or undergoing intermediate or high-risk surgery. Without additional subset-specific model training, PreOpNet predicted death with an AUC of 0.76 (95% CI 0.72-0.80) and major cardiovascular events with an AUC of 0.73 (95% CI 0.70-0.76) compared to an AUC of 0.63 (95% CI 0.53-0.73) and 0.64 (95% CI 0.57-0.70) respectively for the RCRI score. For patients undergoing cardiovascular procedures, PreOpNet predicted death with an AUC of 0.77 (95% CI 0.74-0.80) and major cardiovascular events with an AUC of 0.71 (95% CI 0.69-0.74)

Comparison with other ECG deep learning models

To compare the performance of our architecture with other deep learning models, we implemented a recently described model for ECG waveforms[8] to perform the same task and trained this benchmark model with same training data with labels of postoperative outcomes. The previously published deep learning architecture had an AUC of 0.68 (95% CI 0.65-0.70) for predicting death and an AUC of 0.57 (95% CI 0.55 – 0.59) for predicting major cardiovascular events.

In addition to superior performance in predicting mortality and complications, PreOpNet is a highly efficient deep learning model architecture with fewer parameters and requiring less computational power to train, up to 100x smaller than other previously published architectures (Supplemental Table 3). The improvement in speed-up and computational efficiency allows for the model to be run solely on a standard CPU in 0.078 seconds at inference time.

Discussion

In a large cohort of patients undergoing inpatient procedures, we trained and validated a novel electrocardiogram based deep learning algorithm, PreOpNet, which predicted post-operative death and major adverse events from a single pre-operative ECG. When compared to a standard risk assessment tool and another deep learning algorithms applied to ECGs, PreOpNet was superior in predicting post-operative mortality and major cardiovascular events. This predictive accuracy was re-affirmed in external test datasets from two different healthcare systems with diverse patient populations. PreOpNet, to our knowledge is the first purpose built deep learning architecture for predicting perioperative outcomes, and its predictive performance was achieved despite requiring less computational processing than previously published neural architectures for interpreting ECGs.

Perioperative risk stratification is actionable clinical prognostication that is used to guide medical decision making. Each year, many clinical encounters are performed for perioperative risk stratification, however risk scores have significant variability in predicted risk and utilization is often limited by tedious manual entry of many clinical variables[2,6,17]. Previous risk calculators were derived using less sensitive metrics of post-operative outcomes and have been shown to perform poorly in patient cohorts with different clinical characteristics and higher disease burden[17]. Preoperative ECGs are frequently obtained for risk stratification, even for intermediate risk procedures; thus, an ECG-guided algorithm could be readily incorporated into existing clinical workflows[4]. Thus, a prognostic deep learning model utilizing preoperative ECGs may result in significant improvement in both the accuracy and utilization of pre-operative risk stratification methods.

Deep learning algorithms applied to electrocardiograms have been shown to detect occult arrhythmias[9], ventricular dysfunction[27,28], anemia[22], age[23], and long-term mortality[10,29]. Given the ability for deep learning algorithms to identity variables relevant to perioperative risk as well as screen for subclinical disease, we applied this prognostic ability in the setting of peri-operative evaluation. In this study, we designed a new deep learning architecture inspired by prior non-healthcare computer vision models and show that deep learning algorithm interpretation of pre-operative ECGs can accurately prognosticate postoperative outcomes and augment information from traditional risk calculators. While, previous work has shown that deep learning models of

ECGs can predict longer-term mortality[10], this is one of the first applications of deep learning of ECGs to impact a more immediate actionable clinical outcome.

A particular strength of our deep learning model is the relatively fewer parameters and high computational efficiency compared to previously published neural network architectures for interpreting ECG waveforms[8,11]. Being more computational efficient than previously published models, PreOpNet can be deployed in standard hospital workstations without graphical processing units (GPUs), which are less common in clinical settings. Further work remains to advance the integration of artificial intelligence in medicine, including efforts optimize deep learning model architectures for healthcare applications. To promote further development in the field of artificial intelligence in medicine, we release the full code of our algorithm and data-processing workflow.

The strengths of our investigation include training PreOpNet with the large cohort of patients undergoing inpatient procedures across a decade, validation across three large, diverse medical centers, the use of state-of-the-art deep learning architectures on ECGs, and the prediction of important outcomes at a critical junction. Nevertheless, our findings should be confirmed in other cohorts, and ideally in a prospective manner[28,30]. PreOpNet was validated in study populations centered around three large academic medical centers situated in urban metropolitan areas and may perform differently in other settings. Additionally, our outcomes were derived from the electronic health record, and therefore, our study primarily focused on outcomes during the same hospitalization. To minimize heterogeneity assessing cardiovascular outcomes from the EHR, external validation was focused only on the outcome of mortality. Given the relatively limited prevalence of post-operative complications, we had limited statistical power to perform additional secondary analyses of subgroups.

Our results represent an important step towards improving perioperative risk evaluation. Recognizing that human experts have only limited intuition for prognosticating post-procedural outcomes, risk calculators are integrated into society guidelines to frame the question of perioperative risk stratification[1,4]. The application of more advanced and easier to use algorithms can overcome often tedious data entry and small cohort sizes of earlier validation studies. While we validated PreOpNet across multiple clinical scenarios, further studies are warranted to determine the prospective validity of deep learning algorithms to prognosticate procedural risk.

Acknowledgements

D.O. is supported by NIH K99 HL157421-01. J.W.H. and B.H. is supported by the NSF Graduate Research Fellowship. NY is supported by NIH 5T32HL116273-07. J.H.C. is supported by NIH R56LM013365, NSF SPO 181514, and Gordon and Betty Moore Foundation Grant GBMF8040. J.Y.Z. is supported by NSF CCF 1763191, NIH R21 MD012867-01, NIH P30AG059307 and by a Chan-Zuckerberg Biohub Fellowship. NRC and CMA are supported by HL 091069 and HL11669.

Author Contributions

DO, JWH, NS, JT, JE, RKS, NY, PB, EP, JHC, JE, MT retrieved and quality controlled all ECG data and merged electronic medical record data. JWH, PE, TP, JET, JHC, JT, AP, EAA, JYZ were responsible for external validation. DO, JT, JWH, BH, GD, JTS, PE, NC, SC developed and trained the deep learning algorithms, performed statistical tests, and created all the figures. DO, MP, SSC, EAA, SC, CMA performed clinical evaluation of model performance. DO, NS, JT, SC, CMA wrote the manuscript with critical review and feedback by all authors.

Methods

Data sources and study population

Using the electronic health record system, we identified 340,703 inpatient procedures performed in 178,954 patients age 18 or greater at Cedars-Sinai Medical Center between January 1, 2015 and December 31, 2019. All operating room, catheterization laboratory, and endoscopy suite procedures as well as procedures with anesthesia sedation were considered in the analysis. Procedures were then linked to all ECGs performed anytime within a 30-day window prior to the procedure (Figure 1), and this yielded 149,348 ECGs from 78,021 procedures performed in 52,606 patients which were included in the analyses (Supplemental Figure 2). Patients were randomly split 8:1:1 into training, validation, and test cohorts. Pre-operative clinical characteristics and post-operative outcomes including death and major cardiovascular events (post-operative myocardial infraction, cardiac arrest, and in-hospital mortality) were extracted from the electronic health system and linked to each procedure. The institutional review board of the respective institutions approved the study protocol.

ECG waveform preprocessing and analysis

The ECG waveform data was exported from the data management system, normalized by mean and standard deviation, and paired with surgical outcomes. All ECGs were acquired at a sampling rate of 500 Hz and 10 second, 12 lead ECGs were extracted as 12x5000 matrix of amplitude values. Data was linked by medical record number, and pre-operative ECGs were identified based on ECGs within a 30-day window between date of ECG and date of procedure. ECGs with missing leads were excluded from the study cohort. To maximize training sample size and similar to prior studies of deep learning on ECGs, we utilized all ECGs in the window of interest for training and each ECG-outcome pair was used as a unit of observation for model training. Clinical features of existing coronary artery disease, congestive heart failure, stroke or transient ischemia attack, pre-operative insulin use, creatinine greater than 2mg/dL, and elevated risk surgery as defined by American College of Cardiology and American Heart Association guidelines[4] were identified within the electronic medical record were used to calculate the revised cardiac risk index[1].

Outcome extraction and clinical features selection

The outcome of analysis included mortality during the index hospitalization and a composite outcome of major adverse cardiac events, which included in-hospital post-operative myocardial infraction, cardiac arrest, and mortality. The labels for model training were on the procedure level, such that the same patient could have independent procedures with and without adverse outcomes, and ECGs were mapped to the proximal subsequent procedure. In the test cohorts, the analysis was limited to standard clinical practice – a single ECG most proximal to the procedure of interest for an individual patient, procedure, and hospitalization. Patient demographics, co-morbidities and surgical outcomes were extracted from the electronic data warehouse. Diagnoses were encoded by International Classification of Diseases (ICD)-10 codes. Co-morbidities were determined by diagnoses that were already identified on previous hospitalizations or encounters. Procedural complications were identified by diagnoses that were present for the first time on discharge diagnoses list from that hospital admission. The last available creatinine was obtained from the laboratory results and used to calculate the RCRI score. Insulin use was derived from the list of home medications on admission ascertained during medication reconciliation.

AI model design and training

We designed a novel convolutional neural network, PreOpNet, for ECG interpretation and integration with clinical risk factors to predict the co-primary outcomes of post-operative mortality and composite major adverse cardiac events. The model was trained to predict outcomes with the input of one 12-lead ECG obtained within 30 days prior to the procedure. If the same patient had multiple pre-operative ECGs, each were considered independent cases during model training, but in the test datasets, a single ECG nearest to the procedure was used for inference to mimic how the model would be utilized in clinical practice. When used, clinical features were input into the last fully connected layer prior to model output. Models were trained independently on the primary outcomes using the PyTorch deep learning library.

Optimizing for model runtime and minimizing model complexity, we designed our neural network architecture with initial atrous convolutions and subsequent multi-channel 1D convolutions to less than $1/10^{th}$ the size and 10x the runtime of previously described architectures[12,24,26]. Model layer, channel, width, and depth scaling parameters were inspired by prior literature on neural architecture search[31,32]. The model initialized with random weights and trained with a loss function of binary cross entropy for 100 epochs using an ADAM optimizer with an initial learning rate between 5e-3 and 1e-4. Early stopping was performed based on validation dataset's area under the receiver operating curve. We used Local Interpretable Model-agnostic Explanations[29] to identify and visualize relevant features in the ECG used for model decision making. Additionally, we calculated revised cardiac risk index (RCRI) scores for each patient and procedure to benchmark algorithm performance with conventional risk calculators on the held-out test dataset[1].

External healthcare system test dataset

To assess the reliability of PreOpNet across healthcare systems, our model was additionally tested, without additional fine tuning or further training, on external test datasets from Stanford Healthcare (SHC) and Columbia University Medical Center (CUMC). At SHC, clinical features and outcomes were extracted using the Stanford Research Data Repository Observational Medical Outcomes Partnership (OMOP) common data model. Post-operative mortality was assessed as mortality within 7 days after the procedure in patients with missing discharge date. ECG waveform data was extracted from the TraceMaster (Philips Healthcare) data management system and preprocessed with a low pass filter to further correct for wandering baselines and normalization of waveform data. At CUMC, clinical features and outcomes were extracted from the electronic data warehouse and ECG waveform data was extracted from Muse (GE Healthcare) data management system. Procedures requiring anesthesia or typically performed in the operating room or catherization lab were linked to the most proximal ECG performed within 30 days prior.

Statistical analysis

The performance of the PreOpNet model in predicting the primary outcomes was mathematically assessed by the area under the curve (AUC) of the receiver operating characteristic curve (ROC). After model derivation and training, primary and secondary analysis utilizing a single preoperative ECG were performed on trained models using the held-out test cohort, which was never seen during model training. Secondary sensitivity analyses were limited to procedures performed in patients with known cardiovascular disease, intermediate or high-risk surgeries, and elective procedures. Odds ratio were calculated based on RCRI score >= 2 and PreOpNet output at a similar threshold of risk by proportion of patients deemed at risk. We used two-sided

95% confidence intervals with 10,000 bootstrapped samples and obtaining the 95th percentile ranges for each outcome. Local Interpretable Model-agnostic Explanations was used with 1000 samples per study to identify relevant features in the ECG waveform by iteratively randomly perturbing 0.5% of the waveform and identifying which changes most impacted model performance. Statistical analysis was performed in R and Python.

Data and Code Availability

The code is available at https://github.com/ecg-net/PreOpNet.

Tables and Figures

Figure 1. PreOpNet workflow and results. A. ECGs within 30 days of an inpatient procedure was selected for the study and paired with post-operative outcomes. B. A novel light-weight model architecture was trained to predict post-operative mortality and complications, with input of the nearest 12-lead ECG. Performance of PreOpNet at Cedars-Sinai Medical Center (CSMC), Stanford Healthcare (SHC), and Columbia University Medical Center (CUMC), and standard clinical risk calculator in predicting (C) post-operative mortality and (D) major adverse cardiovascular events. E. Rate of mortality and complications by percentile of risk predicted by PreOpNet. ECG = electrocardiogram. RCRI = Revised Cardiovascular Risk Index.

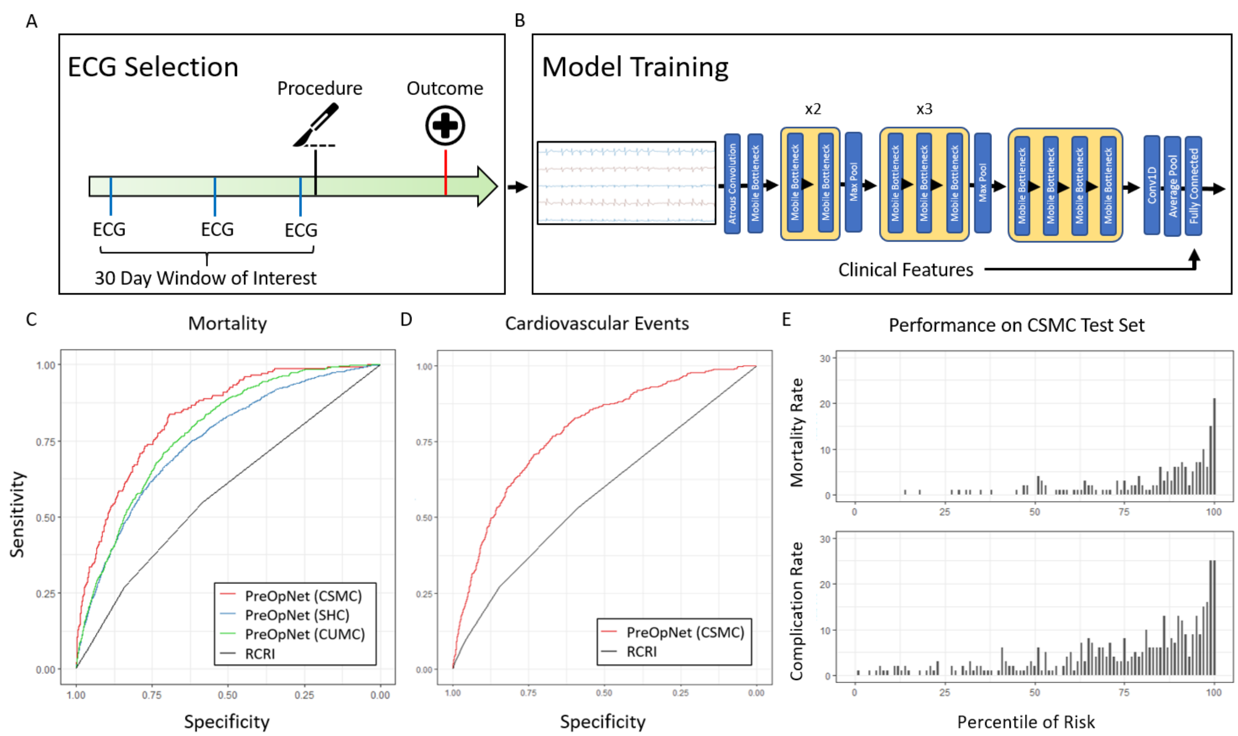


Figure 2. Interpretability Analysis of PreOpNet. Select electrocardiograms before procedures with positive and negative outcomes highlighting most relevant features as determined by interpretability analysis. A) Prediction of Mortality, B) Prediction of Major Adverse Cardiovascular Events

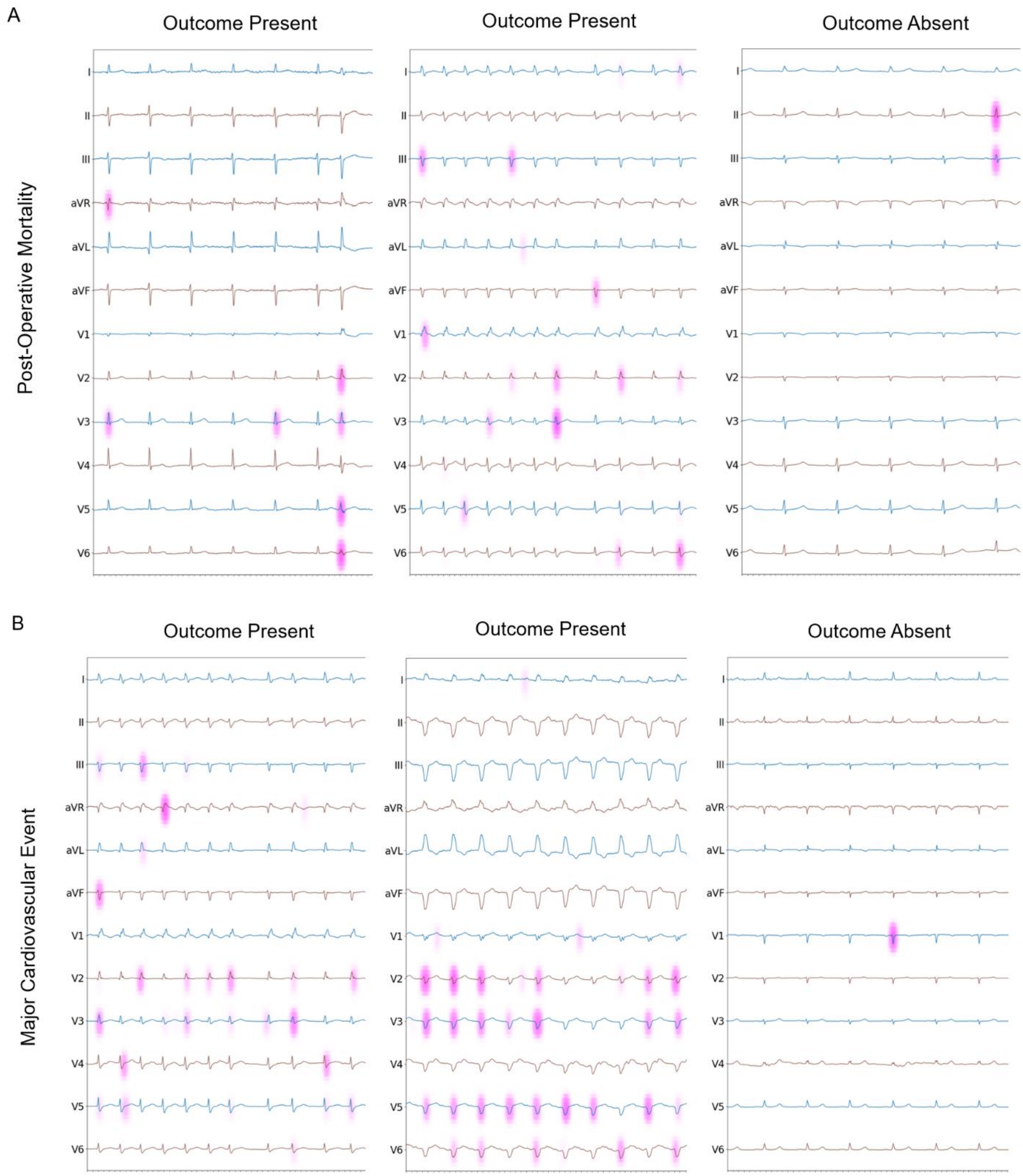


Table 1. Model Performance. Comparison of PreOpNet model performance with RCRI. AUC = Area under the curve. MACE = Major adverse cardiovascular events. RCRI = Revised Cardiovascular Risk Index. Odds ratio were calculated for the 85th percentile of risk, corresponding to a RCRI score of 2.

Outcome	Model	AUC	F1 Score	Odds Ratio for Event
Mortality	PreOpNet (CSMC)	0.83 (0.78 - 0.88)	0.16 (0.11 - 0.22)	8.77 (4.82 - 14.98)
	PreOpNet (CUMC)	0.79 (0.73 - 0.84)	0.08 (0.08 - 0.09)	6.15 (5.44 - 6.92)
	PreOpNet (SHC)	0.75 (0.74 - 0.76)	0.04 (0.04 - 0.04)	4.81 (4.20 - 5.43)
	RCRI	0.58 (0.51 - 0.65)	0.07 (0.04 - 0.12)	2.60 (1.15 - 4.54)
MACE	PreOpNet (CSMC)	0.79 (0.75 - 0.83)	0.25 (0.19 - 0.31)	7.34 (4.93 - 10.56)
	RCRI	0.57 (0.52 - 0.62)	0.12 (0.08 - 0.16)	2.17 (1.27 - 3.31)

Supplemental Table 1. Baseline Characteristics and Outcomes of the Study Participants.

	Total	Training	Validation	Test
Number of Patients	52,606	42,129	5,224	5,253
Number of Procedures	78,021	62,370	7,834	7,817
Number of ECGs	149,348	119,179	15,052	15,117
Demographics				
Age, years (SD)	64.6 (16.0)	64.6 (16.0)	64.9 (16.0)	64.5 (16.0)
Female, n (%)	34,432 (44.2 %)	27,649 (44.4 %)	3,337 (42.7 %)	3,446 (44.1 %)
Heart Failure, n (%)	15,562 (19.9 %)	12,512 (20.1 %)	1,500 (19.1 %)	1,550 (19.8 %)
Diabetes Mellitus, n (%)	13,369 (17.1 %)	10,767 (17.3 %)	1,268 (16.2 %)	1,334 (17.1 %)
Hypertension, n (%)	26,567 (34.1 %)	21,354 (34.2 %)	2,577 (32.9 %)	2,636 (33.7 %)
Coronary Artery Disease, n (%)	15,206 (19.5 %)	12,209 (19.6 %)	1,532 (19.6 %)	1,465 (18.7 %)
Stroke, n (%)	4,440 (5.7 %)	3,581 (5.7 %)	411 (5.2 %)	448 (5.7 %)
Renal Disease, n (%)	9,729 (12.5 %)	7,809 (12.5 %)	969 (12.4 %)	951 (12.2 %)
Procedure Type				
Cardiovascular, n (%)	33,419 (42.8 %)	26,707 (42.8 %)	3,377 (43.1 %)	3,335 (42.7 %)
Intraperitoneal; intrathoracic; suprainguinal vascular, n (%)	10,069 (12.9 %)	8,081 (13.0 %)	1,013 (12.9 %)	975 (12.5 %)
Insulin use prior to admission, n (%)	5,625 (7.2 %)	4,506 (7.2 %)	556 (7.1 %)	563 (7.2 %)
Creatinine > 2 mg/dL, n (%)	8,804 (11.3 %)	7,016 (11.2 %)	879 (11.2 %)	909 (11.6 %)
RCRI > 2, n (%)	4,369 (5.6 %)	3,500 (5.6 %)	450 (5.7 %)	419 (5.4 %)
Outcomes				
Death during admission, n (%)	1,613 (2.1 %)	1,256 (2.0 %)	178 (2.3 %)	179 (2.3 %)
Any major cardiovascular complication, n (%)	4,007 (5.1 %)	3,166 (5.1 %)	437 (5.6 %)	404 (5.2 %)
Cardiac Arrest, n (%)	1,138 (1.5 %)	907 (1.5 %)	139 (1.8 %)	92 (1.2 %)
Myocardial Infarction, n (%)	1,607 (2.1 %)	1,276 (2.0 %)	160 (2.0 %)	171 (2.2 %)
Heart Block, n (%)	56 (0.1 %)	44 (0.1 %)	5 (0.1 %)	7 (0.1 %)
Pulmonary Edema, n (%)	135 (0.2 %)	111 (0.2 %)	16 (0.2 %)	8 (0.1 %)

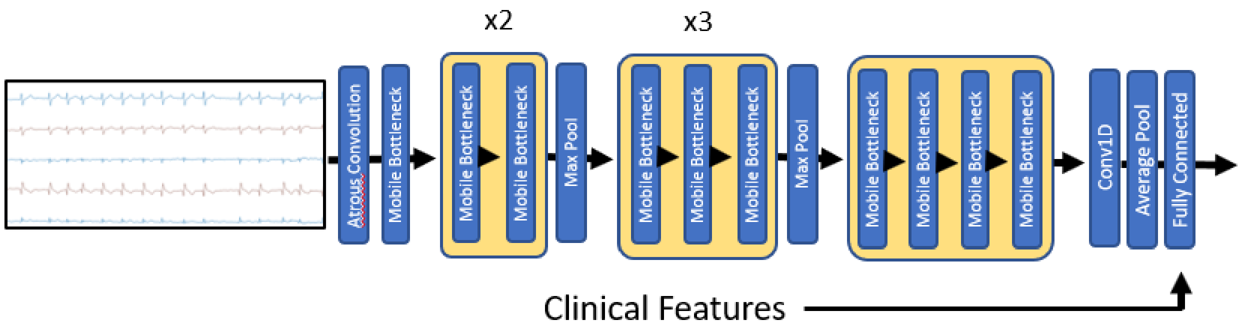
Supplemental Table 2. Baseline Characteristics at Procedure-Level by Outcome.

	Cases with Mortality	Cases without Mortality	Cases with MACE	Cases without MACE	Elective Procedures	Emergent Procedures
Number of Procedures	1,613	76,408	4,007	74,014	44,846	24,160
Number of ECGs	6,069	143,279	17,319	132,029	63,159	63,997
Demographics						
Age, years (SD)	65.5 (16.0)	64.6 (16.0)	65.5 (15.6)	64.6 (16.0)	64.3 (15.1)	66.6 (17.1)
Female, n (%)	593 (36.9 %)	33,839 (44.3 %)	1,399 (35.0 %)	33,033 (44.7 %)	20,279 (45.2 %)	10,544 (43.6 %)
Heart Failure, n (%)	477 (29.6 %)	15,085 (19.7 %)	1,112 (27.8 %)	14,450 (19.5 %)	8,242 (18.4 %)	5,071 (21.0 %)
Diabetes Mellitus, n (%)	290 (18.0 %)	13,079 (17.1 %)	733 (18.3 %)	12,636 (17.1 %)	7,326 (16.3 %)	4,878 (20.2 %)
Hypertension, n (%)	510 (31.6 %)	26,057 (34.1 %)	1,326 (33.1 %)	25,241 (34.1 %)	15,827 (35.3 %)	8,733 (36.1 %)
Coronary Artery Disease, n (%)	277 (17.2 %)	14,929 (19.5 %)	841 (21.0 %)	14,365 (19.4 %)	9,259 (20.6 %)	4,736 (19.6 %)
Stroke, n (%)	98 (6.1 %)	4,342 (5.7 %)	240 (6.0 %)	4,200 (5.7 %)	2,191 (4.9 %)	1,787 (7.4 %)
Renal Disease, n (%)	298 (18.5 %)	9,431 (12.3 %)	680 (17.0 %)	9,049 (12.2 %)	4,513 (10.1 %)	4,149 (17.2 %)
Procedure Type and Clincial Risk Characteristics						
Cardiovascular, n (%)	599 (37.1 %)	32,820 (43.0 %)	2,442 (60.9 %)	30,977 (41.9 %)	20,546 (45.8 %)	8,807 (36.5 %)
Intraperitoneal; intrathoracic; suprainguinal vascular, n (%)	216 (13.4 %)	9,853 (12.9 %)	385 (9.6 %)	9,684 (13.1 %)	7,026 (15.7 %)	1,806 (7.5 %)
Insulin use prior to admission, n (%)	157 (9.7 %)	5,468 (7.2 %)	301 (7.5 %)	5,324 (7.2 %)	2,922 (6.5 %)	1,996 (8.3 %)
Creatinine > 2 mg/dL, n (%)	549 (34.0 %)	8,255 (10.8 %)	1,025 (25.6 %)	7,779 (10.5 %)	2,775 (6.2 %)	4,374 (18.1 %)
RCRI > 2, n (%)	163 (10.1 %)	4,206 (5.5 %)	363 (9.1 %)	4,006 (5.4 %)	1,997 (4.5 %)	1,791 (7.4 %)
Outcomes						
Death during admission, n (%)	1,613 (100.0 %)	0 (0.0 %)	1,613 (40.3 %)	0 (0.0 %)	158 (0.4 %)	819 (3.4 %)
Any major cardiovascular complication, n (%)	1,613 (100.0 %)	2,394 (3.1 %)	4,007 (100.0 %)	0 (0.0 %)	206 (0.5 %)	2,694 (11.2 %)
Cardiac Arrest, n (%)	156 (9.7 %)	982 (1.3 %)	1,138 (28.4 %)	0 (0.0 %)	31 (0.1 %)	778 (3.2 %)
Myocardial Infarction, n (%)	106 (6.6 %)	1,501 (2.0 %)	1,607 (40.1 %)	0 (0.0 %)	21 (0.0 %)	1,296 (5.4 %)
Heart Block, n (%)	6 (0.4 %)	50 (0.1 %)	56 (1.4 %)	0 (0.0 %)	1 (0.0 %)	48 (0.2 %)
Pulmonary Edema, n (%)	7 (0.4 %)	128 (0.2 %)	135 (3.4 %)	0 (0.0 %)	2 (0.0 %)	118 (0.5 %)

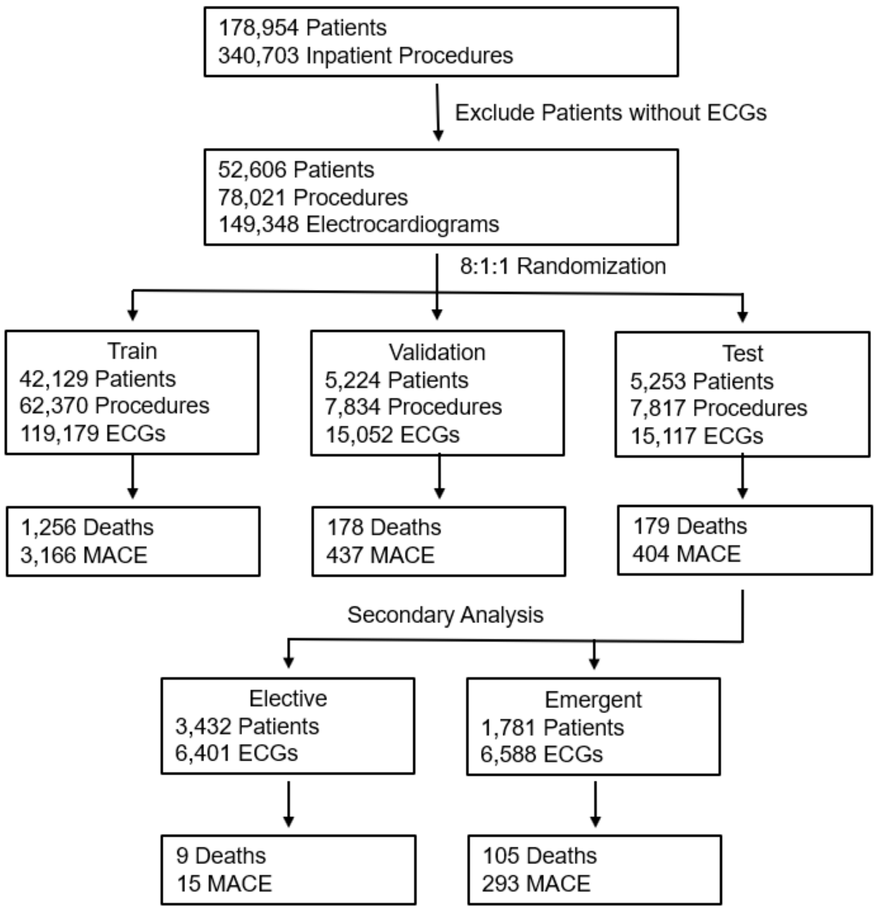
Supplemental Table 3. Performance and computational complexity of deep learning models. FLOPs = Floating point operations

	FLOPs*
PreOpNet	57,913,936
Attia et al.	95,773,024
Kwon et al.	199,355, 858
Goto et al.	3,155,739,318

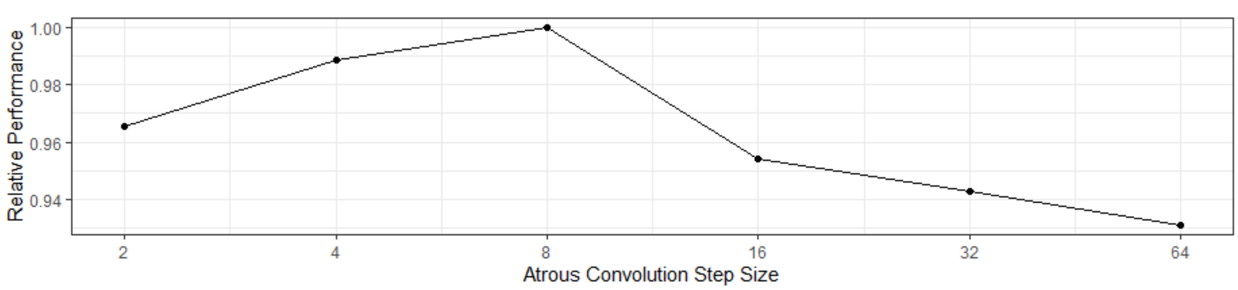
Supplemental Figure 1: Model Architecture

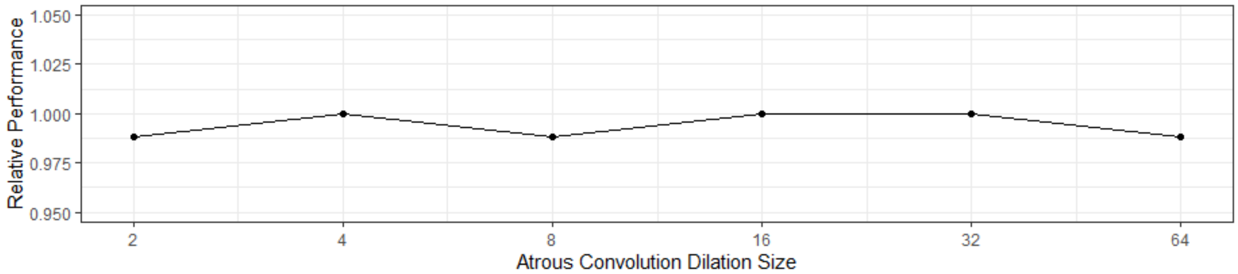


Supplementary Figure 2. Cohort selection procedure and flow diagram. ECG = electrocardiogram. MACE = Major adverse cardiovascular events.



Supplemental Figure 3: Hyperparameter Sweep





References

1 Committee on, S. *et al.* Practice advisory for preanesthesia evaluation: an updated report by the American Society of Anesthesiologists Task Force on Preanesthesia Evaluation. *Anesthesiology* **116**, 522-538, doi:10.1097/ALN.0b013e31823c1067 (2012).

2 Ford, M. K., Beattie, W. S. & Wijeysundera, D. N. Systematic review: prediction of perioperative cardiac complications and mortality by the revised cardiac risk index. *Ann. Intern. Med.* **152**, 26-35, doi:10.7326/0003-4819-152-1-201001050-00007 (2010).

3 Cohn, S. L. & Fernandez Ros, N. Comparison of 4 Cardiac Risk Calculators in Predicting Postoperative Cardiac Complications After Noncardiac Operations. *Am. J. Cardiol.* **121**, 125-130, doi:10.1016/j.amjcard.2017.09.031 (2018).

4 Fleisher, L. A. *et al.* 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines. *J. Am. Coll. Cardiol.* **64**, e77-137, doi:10.1016/j.jacc.2014.07.944 (2014).

5 Cohen, M. E. *et al.* Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. *J. Am. Coll. Surg.* **217**, 336-346.e331, doi:10.1016/j.jamcollsurg.2013.02.027 (2013).

6 Fronczek, J. *et al.* External validation of the Revised Cardiac Risk Index and National Surgical Quality Improvement Program Myocardial Infarction and Cardiac Arrest calculator in noncardiac vascular surgery. *Br. J. Anaesth.* **123**, 421-429, doi:10.1016/j.bja.2019.05.029 (2019).

7 Biccard, B. M. & Rodseth, R. N. Utility of clinical risk predictors for preoperative cardiovascular risk prediction. *Br J Anaesth* **107**, 133-143, doi:10.1093/bja/aer194 (2011).

8 Attia, Z. I. *et al.* Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat. Med.* **25**, 70-74, doi:10.1038/s41591-018-0240-2 (2019).

9 Attia, Z. I. *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* **394**, 861-867, doi:10.1016/S0140-6736(19)31721-0 (2019).

10 Raghunath, S. *et al.* Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat. Med.*, doi:10.1038/s41591-020-0870-z (2020).

11 Goto, S. *et al.* Artificial intelligence-enabled, fully automated detection of cardiac amyloidosis using electrocardiograms and echocardiograms. *bioRxiv*, doi:10.1101/2020.07.02.20141028 (2020).

12 Lee, T. H. *et al.* Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation* **100**, 1043-1049, doi:10.1161/01.cir.100.10.1043 (1999).

13 Lindenauer, P. K. *et al.* Perioperative beta-blocker therapy and mortality after major noncardiac surgery. *N. Engl. J. Med.* **353**, 349-361, doi:10.1056/NEJMoa041895 (2005).

14 Fleisher Lee, A. *et al.* 2014 ACC/AHA Guideline on Perioperative Cardiovascular Evaluation and Management of Patients Undergoing Noncardiac Surgery. *Circulation* **130**, e278-e333, doi:10.1161/CIR.0000000000000106 (2014).

15 Gupta, P. K. *et al.* Development and validation of a risk calculator for prediction of cardiac risk after surgery. *Circulation* **124**, 381-387, doi:10.1161/CIRCULATIONAHA.110.015701 (2011).

16 Glance, L. G. *et al.* Impact of the Choice of Risk Model for Identifying Low-risk Patients Using the 2014 American College of Cardiology/American Heart Association Perioperative Guidelines. *Anesthesiology* **129**, 889-900, doi:10.1097/ALN.0000000000002341 (2018).

17 Davis, C., Tait, G., Carroll, J., Wijeysundera, D. N. & Beattie, W. S. The Revised Cardiac Risk Index in the new millennium: a single-centre prospective cohort re-evaluation of the original variables in 9,519 consecutive elective surgical patients. *Can. J. Anaesth.* **60**, 855-863, doi:10.1007/s12630-013-9988-5 (2013).

18 Ouyang, D. *et al.* Video-based AI for beat-to-beat assessment of cardiac function. *Nature* **580**, 252-256, doi:10.1038/s41586-020-2145-8 (2020).

19 McKinney, S. M. *et al.* International evaluation of an AI system for breast cancer screening. *Nature* **577**, 89-94, doi:10.1038/s41586-019-1799-6 (2020).

20 Poplin, R. *et al.* Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* **2**, 158-164, doi:10.1038/s41551-018-0195-0 (2018).

21 Galloway, C. D. *et al.* Development and Validation of a Deep-Learning Model to Screen for Hyperkalemia From the Electrocardiogram. *JAMA Cardiol* **4**, 428-436, doi:10.1001/jamacardio.2019.0640 (2019).

22 Kwon, J.-M. *et al.* A deep learning algorithm to detect anaemia with ECGs: a retrospective, multicentre study. *Lancet Digit Health* **2**, e358-e367, doi:10.1016/S2589-7500(20)30108-4 (2020).

23 Attia, Z. I. *et al.* Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs. *Circ. Arrhythm. Electrophysiol.* **12**, e007284, doi:10.1161/CIRCEP.119.007284 (2019).

24 Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *arXiv [cs.CV]* (2018).

25 Tan, M. & Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv [cs.LG]* (2019).

26 Chen, L.-C. P., George; Schroff, Florian; Adam, Hartwig. Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv [cs.CV]*.

27 Attia, Z. I. *et al.* Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. *J. Cardiovasc. Electrophysiol.* **30**, 668-674, doi:10.1111/jce.13889 (2019).

28 Yao, X. *et al.* Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nat Med*, doi:10.1038/s41591-021-01335-4 (2021).

29 Siontis, K. C., Noseworthy, P. A., Attia, Z. I. & Friedman, P. A. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol*, doi:10.1038/s41569-020-00503-2 (2021).

30 Wu, E. *et al.* How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nat Med* **27**, 582-584, doi:10.1038/s41591-021-01312-x (2021).