



Published in final edited form as:

Lancet Digit Health. 2024 January ; 6(1): e70–e78. doi:10.1016/S2589-7500(23)00220-0.

Electrocardiographic Deep Learning for Predicting Post-Procedural Mortality: A Model Development and Validation Study

David Ouyang, MD^{1,2}, John Theurer, BS¹, Nathan R. Stein, MD¹, J. Weston Hughes, BS³, Pierre Elias, MD^{4,5}, Bryan He, PhD³, Neal Yuan, MD¹, Grant Duffy, BS¹, Roopinder K. Sandhu, MD¹, Joseph Ebinger, MD¹, Patrick Botting, BS¹, Melvin Jujjavarapu, BS¹, Brian Claggett, PhD⁶, James E. Tooley, MD⁷, Tim Poterucha, MD⁴, Jonathan H. Chen, MD PhD⁸, Michael Nurok, MD⁹, Marco Perez, MD⁷, Adler Perotte, PhD⁴, James Y. Zou, PhD^{3,10}, Nancy R. Cook, PhD¹¹, Sumeet S. Chugh, MD^{1,2}, Susan Cheng, MD¹, Christine M. Albert, MD¹

¹.Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA

².Division of Artificial Intelligence in Medicine, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA

³.Department of Computer Science, Stanford University, Palo Alto, CA

⁴.Milstein Division of Cardiology, Department of Medicine, Columbia University Irving Medical Center, New York, NY

⁵.Department of Biomedical Informatics, Columbia University Irving Medical Center

⁶.Division of Cardiovascular Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA

⁷.Division of Cardiology, Department of Medicine, Stanford University, Palo Alto, CA

⁸.Division of Bioinformatics Research, Department of Medicine, Stanford University, Palo Alto, CA

⁹.Division of Anesthesia, Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, CA

¹⁰.Department of Biomedical Data Science, Stanford University, Palo Alto, CA

¹¹.Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA

Abstract

Background. Pre-operative risk assessments used in clinical practice are limited in their ability to identify risk for post-operative mortality. Deep learning analysis of electrocardiography identify hidden risk markers that can help prognosticate post-operative mortality. We aimed to develop

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

Author Contributions

DO, JWH, NS, JT, JE, RKS, NY, PB, EP, JHC, JE, MT retrieved, and quality controlled all ECG data, merged EMR data, and verified the raw data. DO and JT developed and trained the deep learning algorithms. JWH, PE, TP, JET, JHC, JT, AP, EAA, JYZ were responsible for external validation, MP, NY, MN, 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. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

a prognostic model that accurately predicts all-cause mortality in patients undergoing medical procedures and received pre-operative electrocardiographic diagnostic testing.

Methods. In a derivation cohort of 45,969 pre-operative patients, a deep learning algorithm was developed to leverage waveform signals to discriminate post-operative mortality. Model performance was assessed in a holdout test dataset and in two external hospital cohorts and compared with the Revised Cardiac Risk Index score. The primary outcome was post-procedural mortality across three healthcare systems. The secondary outcome was major adverse cardiovascular events.

Findings. In the held-out internal test cohort, the algorithm discriminates mortality with an AUC of 0.83 (95% CI 0.79-0.87) surpassing the discrimination of the RCRI score with an AUC of 0.67 (CI 0.61-0.72). The algorithm similarly discriminated risk for mortality in two independent healthcare systems with AUCs of 0.79 (0.75-0.83) and 0.75 (0.74-0.76) respectively. Patients determined to be high risk by the deep learning model had an unadjusted odds ratio (OR) of 8.83 (5.57- 13.20) for post-operative mortality as compared to an unadjusted OR of 2.08 (CI 0.77-3.50) for post-operative mortality for RCRI >2. The deep learning algorithm performed similarly for patients undergoing cardiac surgery - AUC of 0.85 (CI 0.77 – 0.92), non-cardiac surgery - AUC of 0.83 (0.79 – 0.88), and catheterization/endoscopy suite procedures - AUC of 0.76 (0.72 – 0.81).

Interpretation. A deep learning algorithm interpreting pre-operative ECGs can improve discrimination of post-operative mortality. The deep learning algorithm worked equally well for risk stratification of cardiac surgeries, non-cardiac surgeries, and catheterization lab procedures, and was validated in three independent healthcare systems. This algorithm can provide additional information to clinicians making the decision to perform medical procedures and risk-stratify risk of future complications.

INTRODUCTION

In the United States, over 20 million surgeries are performed annually.³ Pre-operative risk assessment for adverse procedural outcomes – the most serious of which include death – is routinely performed in clinical practice.^{4, 5} However, current approaches to predicting post-operative risk remain limited.⁶⁻⁸ Over the past three decades, expert guidelines and tools for facilitating pre-operative assessments have evolved to include biomarkers in addition to demographic and clinical data. However, even the most comprehensive risk scores based on recognized risk markers provide only modest ability to discriminate post-operative outcomes with areas under the curve (AUC) ranging from 0.57 to 0.75.^{8, 13-16}

Early identification of patients at high risk for post-procedural mortality can significantly guide patient care, consideration of alternative treatment pathways, and aid in shared decision and communication of risk to patients^{1, 2}. Post-operative biomarker levels have been used to further risk stratify for post-operative mortality with strong discrimination¹⁷⁻¹⁹, however the dependence on post-procedural information limits the opportunity to risk stratify pre-operatively. Novel methods of perioperative risk assessment are needed to achieve better discrimination across the heterogeneous population of pre-operative patients.

The emergence of deep learning analyses now offers the opportunity to capture previously unmeasured risk markers and simultaneously assess complex interactive relationships

from readily available clinical resources for risk prediction.^{20–22} One such ideal resource for perioperative risk discrimination is the 12-lead electrocardiogram (ECG). ECGs are inexpensive, non-invasive, and rapid diagnostic tests that are routinely obtained in the preoperative setting as per clinical guidelines.^{12, 23} Prior studies have applied deep learning analyses of ECG waveforms obtained in other settings to identify clinical traits and outcomes not previously associated with conventional ECG measures or even expert human ECG interpretations.^{5,24–28} Therefore, we hypothesized deep learning analyses applied to a single pre-operative ECG could reliably discriminate post-operative mortality and improve upon established clinical approaches to pre-operative assessment. To evaluate this hypothesis, we conducted a comprehensive study of an AI algorithm trained on perioperative ECGs and evaluated the performance on cohorts from three independent healthcare systems.

METHODS

Derivation Cohort:

All patients undergoing inpatient procedures at Cedars-Sinai Medical Center between January 1, 2015 and December 31, 2019 were included in the derivation cohort. During this time, 153,465 patients aged 18 years or older underwent 261,328 procedures in the operating room, catheterization laboratory, and endoscopy suite. Of this source population, 45,969 patients who had a complete ECG waveform image available for at least one 12-lead ECG performed within 30 days prior to the procedure date were included in the analysis. These patients contributed 112,794 ECGs preceding 59,975 procedures to the analyses (Figure 1).

From this derivation cohort, patients were randomly split 8:1:1 into a subset of 36,839 patients (contributing 90,633 ECGs) for training, 4,549 patients (contributing 11,217 ECGs) for internal validation, and the remaining 4,581 patients (contributing 10,944 ECGs) for final algorithm test analyses. All ECG waveform data were acquired through the clinical enterprise data warehouse at a sampling rate of 500 Hz and extracted as 10 second, 12×5000 matrices of amplitude values. ECGs with missing leads were excluded from analyses. Associated clinical data for each patient were also obtained from the electronic health record. The study was approved by the Cedars-Sinai Medical Center, Stanford University, and Columbia University Institutional Review Boards. Exemption from informed consent was obtained given the use of retrospective, deidentified data for the study. The TRIPOD reporting guidelines were used.

Clinical and Outcome Assessments: Patient demographic, clinical, and outcomes data were assessed from the electronic health record at the time of each procedure. From these data, the pre-operative clinical characteristics needed for calculating the revised cardiac risk index (RCRI)¹ were identified, including: 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 For the main analysis outcome was death during the hospitalization or during readmissions within 30 days. Procedures were subsetted by type, including procedures performed in operating room by cardiac surgeons (cardiac surgeries), procedures performed in the operating room by other surgeons (non-cardiac surgeries, for

which the most common are listed in Supplementary Table 1), and procedures performed the catheterization lab or endoscopy suite. Procedural complications were identified using relevant post-operative diagnoses that were present for the first time after procedure date or at discharge. Outcomes were adjudicated up to 30 days after date of procedure, with multiple procedures having independent outcomes windows based on procedure date. If the same outcome fell within the 30-day window for multiple procedures, the outcome was attributed to each of the procedures. Diagnoses were encoded by International Classification of Diseases (ICD)-9 or ICD-10 codes, and MACE was identified using previously validated criteria from the EHR.²⁹⁻³¹

Electrocardiogram Assessments: We trained and validated a deep learning algorithm based on waveform signals from a single pre-operative 12-lead ECG, termed PreOpNet, on the outcome of post-operative mortality (Figure 1). The input of the model was a 12-lead ECG obtained within 30 days prior to an operative procedure, and the output were the hospitalization level outcomes following that procedure. Patients with multiple procedures were treated independently during model training, with each ECG paired with the most proximal subsequent procedure. To maximize sample size, similar to prior studies of deep learning using electrocardiography,³² all ECGs in the window of interest were utilized in the training set as training examples. In the internal validation and test cohorts, the analysis was limited to a single ECG most proximal to the procedure of interest for an individual patient to mimic how the model would be applied in clinical practice. When used, clinical features were input into the last fully connected layer prior to model output. Models were trained using the PyTorch deep learning library.

Deep Learning Algorithm Development

Based on prior literature regarding lightweight deep learning model architecture design and neural architecture search, PreOpNet was designed to analyze 12-lead ECG waveform data starting with atrous convolutions followed by subsequent multi-channel 1D convolutions. The number of layers paralleled the design of EfficientNet, and to optimize model runtime and minimize model complexity, the number of layers were limited to less than 1/10th the size of previously described architectures.⁵⁻⁷ After initial atrous layers, PreOpNet incorporated convolutional layers with an inverted residual structure where the input and output are bottleneck layers with an intermediate expansion layer.³³ 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. The deep learning model had input only of the 12-lead ECG waveform data, and comparison non-deep learning models included clinical data including age and sex as well as structured ECG information.

The model was 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. The atrous convolution's dilation and step size was grid-searched by hyperparameter tuning for optimal AUC with all other hyperparameters held constant (Supplemental Figure 1). We note that our new model architecture is less than 1/50th

the size of previous models³⁵ and efficient enough to be run on a standard computer (Supplementary Table 2). Local Interpretable Model-agnostic Explanations³⁶ was used to identify and visualize relevant features in the ECG used for model decision making.

Statistical Analyses within the Internal Test Cohort:

Evaluation of Model Performance and Comparison to an Established Risk Calculator:

Following algorithm development in the training and internal validation datasets, the ability of PreOpNet to discriminate the primary outcome of post-operative mortality in the held-out internal test dataset was assessed using the area under the curve (AUC) of the receiver operating characteristic curve (ROC). To understand the models performance in key patient populations, independent analyses were performed limited to patients undergoing cardiac surgery, non-cardiac surgery, interventional endoscopy suite or catheterization laboratory procedures, and patients with known cardiovascular disease or undergoing intermediate or high-risk surgeries. The performance of PreOpNet were compared in the held-out internal test dataset with an established risk calculator (RCRI score)⁸, conventional ECG measures and interpretations, and alternate algorithms based on ECG measurement data.

The association of PreOpNet with perioperative outcomes was compared with that of an RCRI score >2, a commonly used threshold to identify high perioperative risk associated with a MACE rate greater than 8% in meta-analyses. In the training set, 15% of the cohort had an RCRI >2; thus a similar threshold for high perioperative risk (top 15%) was set for PreOpNet. Odds ratios, sensitivity, and specificity for post-operative mortality and MACE were estimated at this threshold.³⁷ Secondary analyses with MACE as the outcome were also performed. The continuous and categorical net reclassification index (NRI) for mortality and MACE rate associated with the addition of PreOpNet to an RCRI score >2 was calculated. 10,000 bootstrapped samples were used to obtain 95th confidence intervals for each estimate. To discern the potentially most informative features of the ECG waveform, in the context of comparing performance to that of the RCRI, 0.5% of the waveform for 1000 samples per study were iteratively randomly perturbed to identify which changes most impacted model performance.³⁶ Implementation timings were evaluated using Python's timeit module.

External Validation of PreOpNet for Discrimination of Post-Operative Mortality

To assess algorithm performance in other hospital settings, the PreOpNet algorithm was applied without any additional further fine tuning or training to patients from two separate external healthcare systems. To maximize rigor of external validation, patient data was not transferred across institutions: rather, investigators from each external institution independently collected outcomes information and ran model inference on their datasets to report summary statistics. The Stanford Healthcare (SHC) cohort included 101,375 patients contributing 162,540 pre-operative ECGs from May 1, 2007 to June 30, 2018. All clinical characteristics and outcomes data were provided through the Stanford Research Data Repository Observational Medical Outcomes Partnership (OMOP) common data model. ECG waveform data was provided through 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. The Columbia University Medical Center (CUMC) cohort included 9,028 patients contributing 9,028 pre-operative ECGs from January 1, 2020 to March 31, 2020. At CUMC, clinical characteristics and outcomes data were obtained from the clinical enterprise data warehouse and ECG waveform data was obtained from Muse (GE Healthcare) data management system. For each of the two external cohorts, the AUC for post-operative mortality from analyses of a single pre-operative ECG was calculated. Procedures were linked to the most proximal preceding ECG performed within 30 days prior and post-operative mortality was assessed as mortality within 30 days after the procedure or during the hospitalization. The most common procedures are summarized in Supplementary Table 1.

RESULTS

Derivation (CSMC) Cohort Characteristics

The CSMC cohort included 45,969 patients who underwent 59,975 inpatient procedures between years 2013 and 2019 (Figure 1). The mean age at the time of pre-operative ECG was 65.1 ± 15.9 years, 45% were women, and 22% had pre-existing coronary artery disease as shown in Table 1. There were 1,065 (1.8%) subsequent deaths and 1,730 (2.9%) post-procedural MACE events during the hospitalization in the cohort. Clinical characteristics and index hospitalization outcomes are also shown in Table 1 for the training, internal validation, and internal test subsets of the cohort. Compared with those patients who were excluded based on not having ECG data within 30 days of their procedure, patients with pre-operative ECGs were more likely to be older, male, and with more cardiovascular risk factors (Supplementary Table 3).

Discrimination of Post-Procedural Mortality at CSMC

For the outcome of mortality, the PreOpNet algorithm was developed in the derivation cohort using training, validation, and test datasets. In the held-out test dataset, the algorithm was then shown to discriminate mortality with an AUC of 0.83 (95% CI 0.79-0.87) (Table 2). By contrast, the conventional RCRI score discriminated post-operative mortality with an AUC of 0.67 (0.61-0.72). In the 15% of patients who had an RCRI score of 2 or greater, the unadjusted odds ratio for post-operative mortality was of 2.08 (95% CI, 0.77-3.50). In comparison, patients in the top 15% of the PreOpNet algorithm had an adjusted odds ratio of 9.17 (95% CI, 5.85-13.82) for post-operative mortality. The addition of the components of the RCRI score to the PreOpNet algorithm trained only on 12-lead ECG waveforms did not significantly improve model performance (AUC 0.83; 95% CI, 0.77-0.89) in the test dataset. There were no significant differences in model performance in patient subsets by age, sex, or race. At the pre-specified calibrated threshold of risk comparable to a RCRI score ≥ 2 , the PreOpNet algorithm demonstrated a specificity of 0.87 (0.86-0.88) and sensitivity of 0.57 (0.48-0.68) for post-operative mortality (Supplementary Table 5). In comparison, the RCRI score ≥ 2 had slightly higher specificity of 0.94 (0.93-0.94) but much lower sensitivity of 0.12 (0.05-0.19). At this threshold, the positive predictive value was slightly higher for PreOpNet as compared to the RCRI score. (Table 2).

Discrimination of Post-Procedural Mortality in Important Clinical Scenarios

PreOpNet performed well both in patients undergoing major surgical procedures in the operating room as well as patients undergoing procedures in the catheterization laboratory or endoscopy suite. For patients with surgeries in the operating room, PreOpNet discriminated post-operative mortality with an AUC of 0.84 (0.76 – 0.92), compared to an AUC of 0.70 (0.61 – 0.78) for the RCRI score. For patients with procedures in the catheterization lab or endoscopy suite, PreOpNet discriminated post-operative mortality with an AUC of 0.83 (0.78 – 0.98), compared to an AUC of 0.66 (0.60 – 0.72) for the RCRI score. PreOpNet performed similarly in discriminating mortality in patients undergoing cardiovascular surgery with an AUC of 0.85 (0.77 – 0.92) and patients undergoing non-cardiac surgery with an AUC of 0.83 (0.79 – 0.88). In patients undergoing cardiac surgery, the RCRI score discriminated post-operative mortality with an AUC of 0.62 (0.52–0.72), and in patients undergoing non-cardiac surgery, the RCRI score discriminated post-operative mortality with an AUC of 0.70 (0.63-0.77).

Given that ECGs are often not obtained in low-risk patients undergoing low-risk procedures, a secondary analysis was performed in patients most likely to be considered at least moderate-risk – patients either with known cardiovascular disease or those undergoing elective intermediate-risk or high-risk surgery. Without additional subset-specific fine-tuning, the PreOpNet algorithm discriminated post-operative mortality in this subset with an AUC of 0.80 (0.71-0.88). In clinical practice, pre-operative risk assessment most commonly occurs in the elective procedural setting. Thus, secondary analyses were performed limited to those patients in the CSMC test cohort who were undergoing elective procedures (3,691 patients contributing 5,165 ECGs). In this setting, the PreOpNet algorithm discriminated post-operative mortality with an AUC of 0.80 (0.67-0.92).

Discrimination of Post-Procedural Mortality in External Validation Cohorts

To assess external validity of the PreOpNet algorithm, the discriminatory ability was evaluated in two external health system cohorts. The external test evaluation cohorts included 101,375 patients in the Stanford Healthcare (SHC) system contributing 162,540 ECGs and 9,028 patients in the Columbia University Medical Center (CUMC) system contributing 9,028 ECGs. In the SHC cohort, the post-operative mortality rate was 1.3% and PreOpNet discriminated this outcome with an AUC of 0.75 (0.74-0.76). In the CUMC cohort, the post-operative mortality rate was 1.6% and the algorithm discriminated this outcome with an AUC of 0.79 (0.75-0.83) (Table 2). The PreOpNet algorithm pre-specified high risk group (>15%) had an unadjusted odds ratio of 5.88 (5.00-7.00) in SHC and 6.20 (3.87-10.41) in CUMC for mortality. Results from analyses of specificity, sensitivity, and positive and negative predictive value were similar in the external validation cohorts when compared with results observed in the CSMC cohort (Table 2).

Major Cardiovascular Events (MACE) within the Internal Test Dataset.

Secondary analyses were performed within the internal test dataset that utilized a combined secondary MACE outcome that included non-fatal major cardiovascular events and post-operative mortality. For this secondary outcome, the PreOpNet algorithm discriminated events in the held-out internal test dataset with an AUC of 0.77 (0.73-0.80) whereas the

RCRI score had an AUC of 0.63 (0.59-0.68). Patients with an RCRI score of 2 or greater had an unadjusted odds ratio of 1.67 (0.77-2.68) for MACE when compared with those with an RCRI score <2. By contrast, patients identified by the PreOpNet algorithm to be high risk had an unadjusted odds ratio 5.38 (3.75-7.49) for MACE (Table 2). The PreOpNet algorithm demonstrated a specificity of 0.88 (0.88-0.89) and sensitivity of 0.41 (0.33-0.49) while the RCRI score again had higher specificity at 0.94 (0.93-0.94) but much lower sensitivity at 0.10 (0.05-0.15). The PPV for MACE was higher for PreOpNet as compared with the RCRI with similar negative predictive values (Table 2).

Clinical Risk Assessment and Reclassification for MACE

The ability of the PreOpNet algorithm to reclassify risk in the internal hold-out test dataset was also evaluated. When compared with the RCRI score, application of the PreOpNet algorithm led to significant improvement in the continuous net reclassification index (NRI 0.53; 95% CI, 0.38 to 0.68). In categorical analyses using the pre-specified threshold of risk (top 15th percentile PreOpNet prediction), 981 (82.4%) of 1190 patients originally classified as high-risk for MACE by the RCRI score (RCRI ≥ 2) were identified as low-risk by PreOpNet (Supplementary Table 4). Of these high-to-low reclassified risk patients, 33 (3.4%) patients experienced MACE. By contrast, of the 4,739 patients classified as low-risk by RCRI (RCRI <2), the PreOpNet algorithm reclassified 327 (6.9%) to be high risk; of these patients, 26 (8.0%) patients experienced MACE. Despite a fair amount of reclassification, the categorical NRI at this cut-point was not significant (NRI 0.06; 95% CI -0.04 to 0.18) for MACE.

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 waveforms⁴¹ 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 2). The improvement in speed-up and computational efficiency allows for the model to be run solely on a standard CPU at inference time. This allows for a web interface deployment for ready access by clinicians (Supplementary Figure 2). In a series of repeated runtime experiments, the PreOpNet application accessed on a standard clinical workstation (Windows 10 64-bit OS, 3.8 GHz processor) was able to intake image data from 50 de novo ECGs and output post-operative risk estimates within 0.032 ± 0.004 seconds per ECG for the local software installation and 0.041 ± 0.006 seconds per ECG for the web application accessed by a mobile phone.

DISCUSSION

In a large cohort of patients undergoing inpatient procedures, a deep learning algorithm utilizing the waveforms of a single pre-operative 12-lead ECG identified risk for post-operative death for cardiac surgeries, non-cardiac surgeries, and catheterization lab interventions. Compared with a widely used standard peri-operative risk assessment tool and alternative ECG assessment tools, PreOpNet was able to more effectively identify high-risk patients who went on to experience postoperative mortality. Further, the accuracy of PreOpNet for discriminating post-operative mortality was re-affirmed in two external healthcare system cohorts with diverse patient populations. Additionally, in secondary analyses, PreOpNet identified high-risk patients who went on to experience MACE within the internal test dataset in the CSMC patient population. To our knowledge, PreOpNet is the first deep learning architecture designed to aid clinicians in discriminating post-operative outcomes.

Over the last several years, the 12-lead ECG has been the subject of deep learning algorithm development for potential clinical applications, with promising results. Previously applied to improve detection of occult cardiovascular traits (e.g. atrial fibrillation,³² hypertrophic cardiomyopathy³⁹, myocardial infarction¹³, cardiac amyloidosis⁴⁰, and ventricular dysfunction^{41, 42}), recent deep learning ECG models have been shown to also identify non-cardiovascular specific traits such as liver disease,¹¹ anemia,²⁷ age,²⁸ and long-term mortality.^{24, 43} These latter results highlight the potential of deep learning methods to extract broad as well as specific novel information from ECG waveforms. Therefore, to optimize performance for the task of pre-operative risk assessment, our algorithm used readily available ECG waveform data collected from large diverse real-world patient cohorts. The resulting PreOpNet algorithm demonstrated ability not only to discriminate post-operative mortality, but to do so while outperforming a conventional clinical risk score and human-interpretable features from the ECG. As indicated in other clinical fields where artificial intelligence has been shown to leverage latent features from a medical image to refine diagnosis or prognostication,^{22, 44, 45} our results indicate the potential value of the PreOpNet algorithm to augment clinical decision making for pre-operative risk assessment.

Several limitations of this preliminary study merit consideration. Many ambulatory procedures for clinically assessed low-risk patients do not involve acquiring a pre-operative ECG. Therefore, this algorithm may not be applicable to such lower risk patients. Additionally, RCRI is most applicable and designed to be evaluated in patients undergoing non-cardiac surgery, so the most direct comparison is in this setting with an AUC of 0.83 for PreOpNet vs. 0.70 for RCRI. Other comparisons only accentuate the difference in performance. Also, all analyses were performed on retrospective cohorts. For each site, investigators at that site performed their own cohort selection as well as running inference, with only code and model weights shared across institutions. In our analyses, we notice drops in model performance in external validation, likely due to changes in cohort make-up and procedural definition. For example, the CUMC cohort was only collected from 3 months at the beginning of the COVID19 pandemic and the SHC ECG data was extracted from a separate ECG data storage system, resulting in differences in cohort selection that could change model performance. Additional prospective validation studies are needed in large

and diverse external cohorts – particularly of the exploratory secondary MACE endpoint– to precisely evaluate PreOpNet’s performance in discriminating events.^{42, 46} Notwithstanding these limitations, the current study offers several strengths including the ability to leverage internal training, validation, and test datasets within a large derivation cohort of patients undergoing inpatient procedures over a decade of time. The algorithm was also able to be externally validated for post-procedural mortality in two large, diverse medical centers.

In summary, our findings demonstrate how a novel deep learning algorithm, applied to a single pre-operative ECG, can improve discrimination of post-operative adverse outcomes while running efficiently on a standard clinical workstation. Recognizing that clinicians have limited time for making clinical assessments and decisions around potential post-procedural outcomes, conventional risk calculators using easily accessible information have been recommended by professional society practice guidelines to aid in peri-operative risk stratification.^{12, 23} The opportunity to implement potentially more informative and easier to use prediction algorithms, in a manner that integrates with existing clinical workflows, offers a potential path towards improving post-operative outcomes. These promising results warrant further studies to determine the prospective validity of deep learning algorithms for prognosticating post-procedural risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Disclosures

DO reports support from NIH NHLBI R00HL157421, Alexion, as well as consulting/Honoraria for lectures from EchoIQ, Ultramix, Pfizer, InVision, Korean Society of Echo, and Japanese Society of Echo. JWH reports funding from the NSF (DGE-1656518). PE reports research support from NIH T32HL007854-21 and serve on the American College of Cardiology Innovation Council. BC reports consulting from Alnylam, Cardurion, Corvia, CVRX, Cytokinetics, Intellia, Rocket. TP reports stock in Abbot and Baxter as well as research support from AHA, Eido, Pfizer, Edwards, and New York Academy of Medicine. JC reports research funding from NIH/NIA (AI17812101) and NIH/NIDA (UG1DA015815-CTN-0136), AIMI-HAI partnership, Doris Duke Charitable Foundation (20211260), Google (SPO136094), AHA, and NCATS/CTSA (R56LM013365), is a co-founder of Reaction Explorer LLC, and receives consulting fees from Sutton Pierce and Younker Hyde MacFarlane PLLC. MP reports research support from the NIH (HL136390), patent for ECG based cardiac arrhythmia detection ((WO2014205310A3)), and consulting fees from Apple Inc., Biotronik, Boston Scientific, QALY Inc., Johnson and Johnson, Bristol Myers Squibb. CMA reports support from NIH (HL165840), consulting fees from Medtronic, Novartis, Illumina, Medtronic, participation in advisory boards for Boston Scientific, Medtronic, and Element Science, and is board of trustee for Heart Rhythm Society.

Data Sharing

All code and analytical methods applied for the algorithm and analyses are available at <https://github.com/ecg-net/PreOpNet>. The patient data is not available given the potentially identifiable nature of the associated data.

References

1. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY and Cohen ME. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg*. 2013;217:833–42 e1–3. [PubMed: 24055383]

2. Glance LG, Lustik SJ, Hannan EL, Osler TM, Mukamel DB, Qian F and Dick AW. The Surgical Mortality Probability Model: derivation and validation of a simple risk prediction rule for noncardiac surgery. *Ann Surg.* 2012;255:696–702. [PubMed: 22418007]
3. Siontis KC, Noseworthy PA, Attia ZI and Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol.* 2021.
4. Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF, Sugarbaker DJ, Donaldson MC, Poss R, Ho KK, Ludwig LE, Pedan A and Goldman L. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation.* 1999;100:1043–1049. [PubMed: 10477528]
5. Lindenauer PK, Pekow P, Wang K, Mamidi DK, Gutierrez B and Benjamin EM. Perioperative beta-blocker therapy and mortality after major noncardiac surgery. *N Engl J Med.* 2005;353:349–361. [PubMed: 16049209]
6. Fronczek J, Polok K, Devereaux PJ, Górka J, Archbold RA, Biccard B, Duceppe E, Le Manach Y, Sessler DI, Duchińska M and Szczeklik W. 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.* 2019;123:421–429. [PubMed: 31256916]
7. Davis C, Tait G, Carroll J, Wijeyesundera DN and Beattie WS. 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.* 2013;60:855–863. [PubMed: 23813289]
8. Ford MK, Beattie WS and Wijeyesundera DN. Systematic review: prediction of perioperative cardiac complications and mortality by the revised cardiac risk index. *Ann Intern Med.* 2010;152:26–35. [PubMed: 20048269]
9. Gupta PK, Gupta H, Sundaram A, Kaushik M, Fang X, Miller WJ, Esterbrooks DJ, Hunter CB, Pipinos II, Johanning JM, Lynch TG, Forse RA, Mohiuddin SM and Mooss AN. Development and validation of a risk calculator for prediction of cardiac risk after surgery. *Circulation.* 2011;124:381–387. [PubMed: 21730309]
10. Glance LG, Faden E, Dutton RP, Lustik SJ, Li Y, Eaton MP and Dick AW. 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.* 2018;129:889–900. [PubMed: 30001221]
11. Fleisher Lee A, Fleischmann Kirsten E, Auerbach Andrew D, Barnason Susan A, Beckman Joshua A, Bozkurt B, Davila-Roman Victor G, Gerhard-Herman Marie D, Holly Thomas A, Kane Garvan C, Marine Joseph E, Nelson MT, Spencer Crystal C, Thompson A, Ting Henry H, Uretsky Barry F and Wijeyesundera Duminda N. 2014 ACC/AHA Guideline on Perioperative Cardiovascular Evaluation and Management of Patients Undergoing Noncardiac Surgery. *Circulation.* 2014;130:e278–e333. [PubMed: 25085961]
12. Committee on S, Practice P, Apfelbaum JL, Connis RT, Nickinovich DG, American Society of Anesthesiologists Task Force on Preanesthesia E, Pasternak LR, Arens JF, Caplan RA, Connis RT, Fleisher LA, Flowerdew R, Gold BS, Mayhew JF, Nickinovich DG, Rice LJ, Roizen MF and Twersky RS. Practice advisory for preanesthesia evaluation: an updated report by the American Society of Anesthesiologists Task Force on Preanesthesia Evaluation. *Anesthesiology.* 2012;116:522–538. [PubMed: 22273990]
13. Sheth T, Chan M, Butler C, Chow B, Tandon V, Nagele P, Mitha A, Mrkobrada M, Szczeklik W, Faridah Y, Biccard B, Stewart LK, Heels-Ansdell D, Devereaux PJ, Coronary Computed Tomographic A and Vascular Events in Noncardiac Surgery Patients Cohort Evaluation Study I. Prognostic capabilities of coronary computed tomographic angiography before non-cardiac surgery: prospective cohort study. *BMJ.* 2015;350:h1907. [PubMed: 25902738]
14. Rao JY, Yeriswamy MC, Santhosh MJ, Shetty GG, Varghese K, Patil CB and Iyengar SS. A look into Lee's score: peri-operative cardiovascular risk assessment in non-cardiac surgeries-usefulness of revised cardiac risk index. *Indian Heart J.* 2012;64:134–8. [PubMed: 22572486]
15. Park SJ, Choi JH, Cho SJ, Chang SA, Choi JO, Lee SC, Park SW, Oh JK, Kim DK and Jeon ES. Comparison of transthoracic echocardiography with N-terminal pro-brain natriuretic Peptide as a tool for risk stratification of patients undergoing major noncardiac surgery. *Korean Circ J.* 2011;41:505–11. [PubMed: 22022325]

16. Choi JH, Cho DK, Song YB, Hahn JY, Choi S, Gwon HC, Kim DK, Lee SH, Oh JK and Jeon ES. Preoperative NT-proBNP and CRP predict perioperative major cardiovascular events in non-cardiac surgery. *Heart*. 2010;96:56–62. [PubMed: 19861299]
17. Machado MN, Rodrigues FB, Nakazone MA, Martin DF, Sabbag ATR, Grigolo IH, Silva-Junior OL, Maia LN and Jaffe AS. Prediction of Death After Noncardiac Surgery: Potential Advantage of Using High-Sensitivity Troponin T as a Continuous Variable. *J Am Heart Assoc*. 2021;10:e018008. [PubMed: 33660524]
18. Marston N, Brenes J, Garcia S, Kuskowski M, Adabag S, Santilli S and McFalls EO. Peak postoperative troponin levels outperform preoperative cardiac risk indices as predictors of long-term mortality after vascular surgery Troponins and postoperative outcomes. *J Crit Care*. 2012;27:66–72. [PubMed: 21798697]
19. Vascular Events In Noncardiac Surgery Patients Cohort Evaluation Study I, Devereaux PJ, Chan MT, Alonso-Coello P, Walsh M, Berwanger O, Villar JC, Wang CY, Garutti RI, Jacka MJ, Sigamani A, Srinathan S, Biccard BM, Chow CK, Abraham V, Tiboni M, Pettit S, Szczeklik W, Lurati Buse G, Botto F, Guyatt G, Heels-Ansdell D, Sessler DI, Thorlund K, Garg AX, Mrkobrada M, Thomas S, Rodseth RN, Pearse RM, Thabane L, McQueen MJ, VanHelder T, Bhandari M, Bosch J, Kurz A, Polanczyk C, Malaga G, Nagele P, Le Manach Y, Leuwer M and Yusuf S. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. *JAMA*. 2012;307:2295–304. [PubMed: 22706835]
20. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA, Harrington RA, Liang DH, Ashley EA and Zou JY. Video-based AI for beat-to-beat assessment of cardiac function. *Nature*. 2020;580:252–256. [PubMed: 32269341]
21. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, Back T, Chesus M, Corrado GC, Darzi A, Etemadi M, Garcia-Vicente F, Gilbert FJ, Halling-Brown M, Hassabis D, Jansen S, Karthikesalingam A, Kelly CJ, King D, Ledsam JR, Melnick D, Mostofi H, Peng L, Reicher JJ, Romera-Paredes B, Sidebottom R, Suleyman M, Tse D, Young KC, De Fauw J and Shetty S. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577:89–94. [PubMed: 31894144]
22. Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, Peng L and Webster DR. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018;2:158–164. [PubMed: 31015713]
23. Fleisher LA, Fleischmann KE, Auerbach AD, Barnason SA, Beckman JA, Bozkurt B, Davila-Roman VG, Gerhard-Herman MD, Holly TA, Kane GC, Marine JE, Nelson MT, Spencer CC, Thompson A, Ting HH, Uretsky BF, Wijeyesundera DN, American College of C and American Heart A. 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*. 2014;64:e77–137. [PubMed: 25091544]
24. Raghunath S, Ulloa Cerna AE, Jing L, vanMaanen DP, Stough J, Hartzel DN, Leader JB, Kirchner HL, Stumpe MC, Hafez A, Nemani A, Carbonati T, Johnson KW, Young K, Good CW, Pfeifer JM, Patel AA, Delisle BP, Alsaid A, Beer D, Haggerty CM and Fornwalt BK. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network. *Nat Med*. 2020.
25. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA, Munger TM, Asirvatham SJ, Scott CG, Carter RE and Friedman PA. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med*. 2019;25:70–74. [PubMed: 30617318]
26. Galloway CD, Valys AV, Shreibati JB, Treiman DL, Petterson FL, Gundotra VP, Albert DE, Attia ZI, Carter RE, Asirvatham SJ, Ackerman MJ, Noseworthy PA, Dillon JJ and Friedman PA. Development and Validation of a Deep-Learning Model to Screen for Hyperkalemia From the Electrocardiogram. *JAMA Cardiol*. 2019;4:428–436. [PubMed: 30942845]
27. Kwon J-M, Cho Y, Jeon K-H, Cho S, Kim K-H, Baek SD, Jeung S, Park J and Oh B-H. A deep learning algorithm to detect anaemia with ECGs: a retrospective, multicentre study. *Lancet Digit Health*. 2020;2:e358–e367. [PubMed: 33328095]
28. Attia ZI, Friedman PA, Noseworthy PA, Lopez-Jimenez F, Ladewig DJ, Satam G, Pellikka PA, Munger TM, Asirvatham SJ, Scott CG, Carter RE and Kapa S. Age and Sex Estimation

Using Artificial Intelligence From Standard 12-Lead ECGs. *Circ Arrhythm Electrophysiol.* 2019;12:e007284. [PubMed: 31450977]

29. Suchard MA, Schuemie MJ, Krumholz HM, You SC, Chen R, Pratt N, Reich CG, Duke J, Madigan D, Hripcsak G and Ryan PB. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *Lancet.* 2019;394:1816–1826. [PubMed: 31668726]
30. Singh S, Fouayzi H, Anzuoni K, Goldman L, Min JY, Griffin M, Grijalva CG, Morrow JA, Whitmore CC, Leonard CE, Selvan M, Nair V, Zhou Y, Toh S, Petrone A, Williams J, Fazio-Eynullayeva E, Swain R, Tyler Coyle D and Andrade S. Diagnostic Algorithms for Cardiovascular Death in Administrative Claims Databases: A Systematic Review. *Drug Saf.* 2019;42:515–527. [PubMed: 30471046]

Research In Context

Evidence before this study

Early identification of patients at high risk for post-procedural mortality can significantly guide patient care, consideration of alternative treatment pathways, and aid in shared decision and communication of risk to patients. We conducted a comprehensive search of society guidelines and primary research articles on PubMed to inform our understanding of the field. A formal meta-analysis was not performed but relevant literature was reviewed in English language publications. Post-operative biomarker levels have been used to further risk stratify for post-operative mortality with strong discrimination. However, even the most comprehensive risk scores based on recognized risk markers provide only modest ability to discriminate post-operative outcomes. All available literature was reviewed between 2021 and 2023 prior to publication of the manuscript.

Added value of this study

Deep learning approaches applied to pre-operative electrocardiograms (ECGs) can offer additional information that could improve discrimination of post-operative mortality and improve upon current pre-operative risk stratification tools. PreOpNet is a deep learning algorithm that leverages ECG waveform signals to augment discrimination of post-operative mortality, and its performance was evaluated in three independent healthcare systems.

Implications of all the available evidence

ECGs contain significant additional information that can assist in pre-operative risk-stratification and deep learning evaluation of ECGs can better discriminate post-operative mortality than the Revised Cardiac Risk Index.

Perspectives

COMPETENCY IN MEDICAL KNOWLEDGE:

Deep learning artificial intelligence analysis can accurately detect post-operative mortality and complications based on the standard 12-lead ECG.

TRANSLATIONAL OUTLOOK:

Implementation studies are needed to assess the cost-effectiveness of deep learning technology to identify patients undergoing surgery who benefit from referral for further risk stratification and/or intervention.

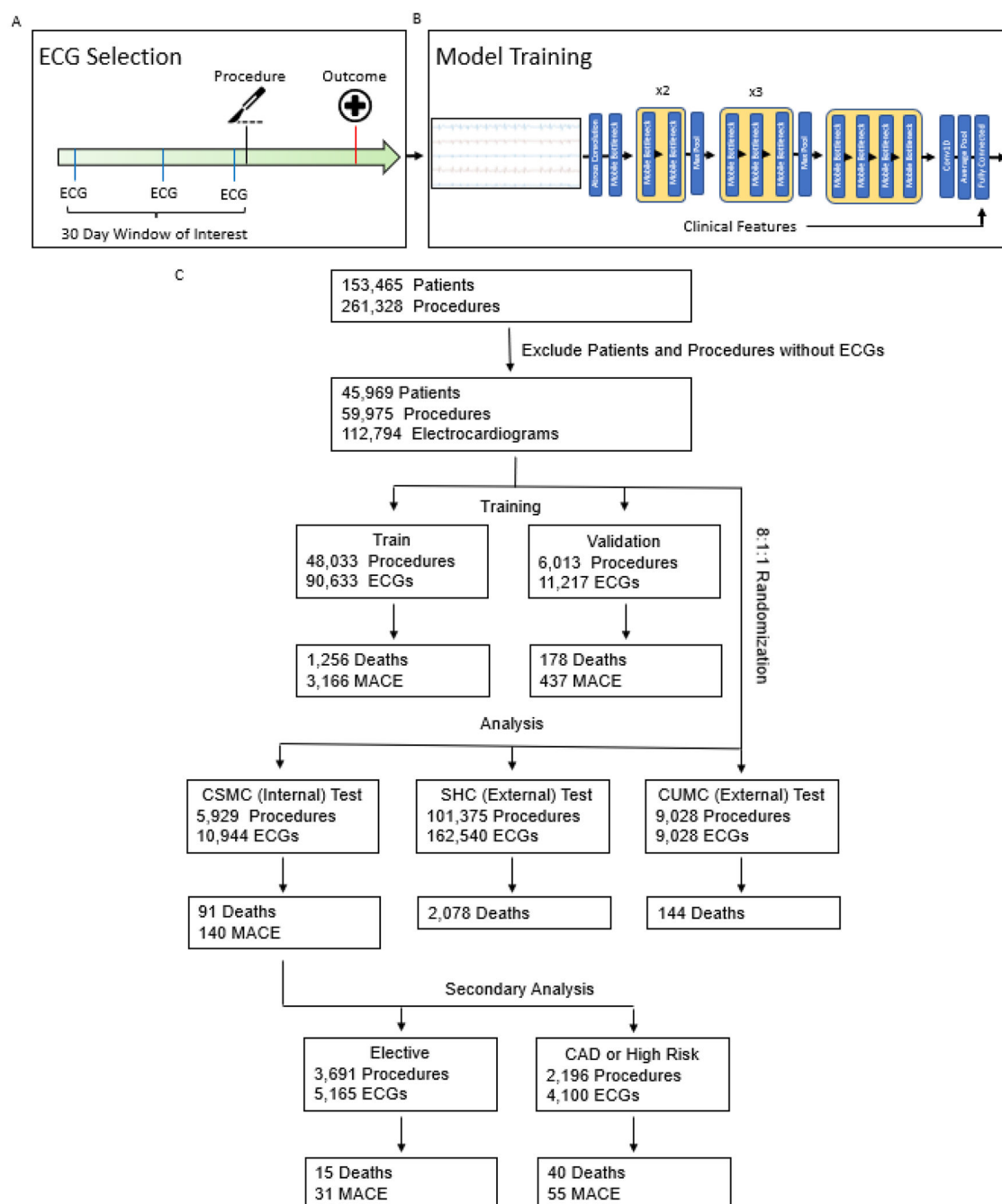


Figure 1. CSMC site cohort sampling.

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 discriminate post-operative mortality and complications, with input of the nearest 12-lead ECG. C. Consort Diagram. ECG = electrocardiogram. MACE = Major adverse cardiovascular events. CSMC = Cedars-Sinai Medical Center, SHC = Stanford Healthcare, and CUMC = Columbia University Medical Center.

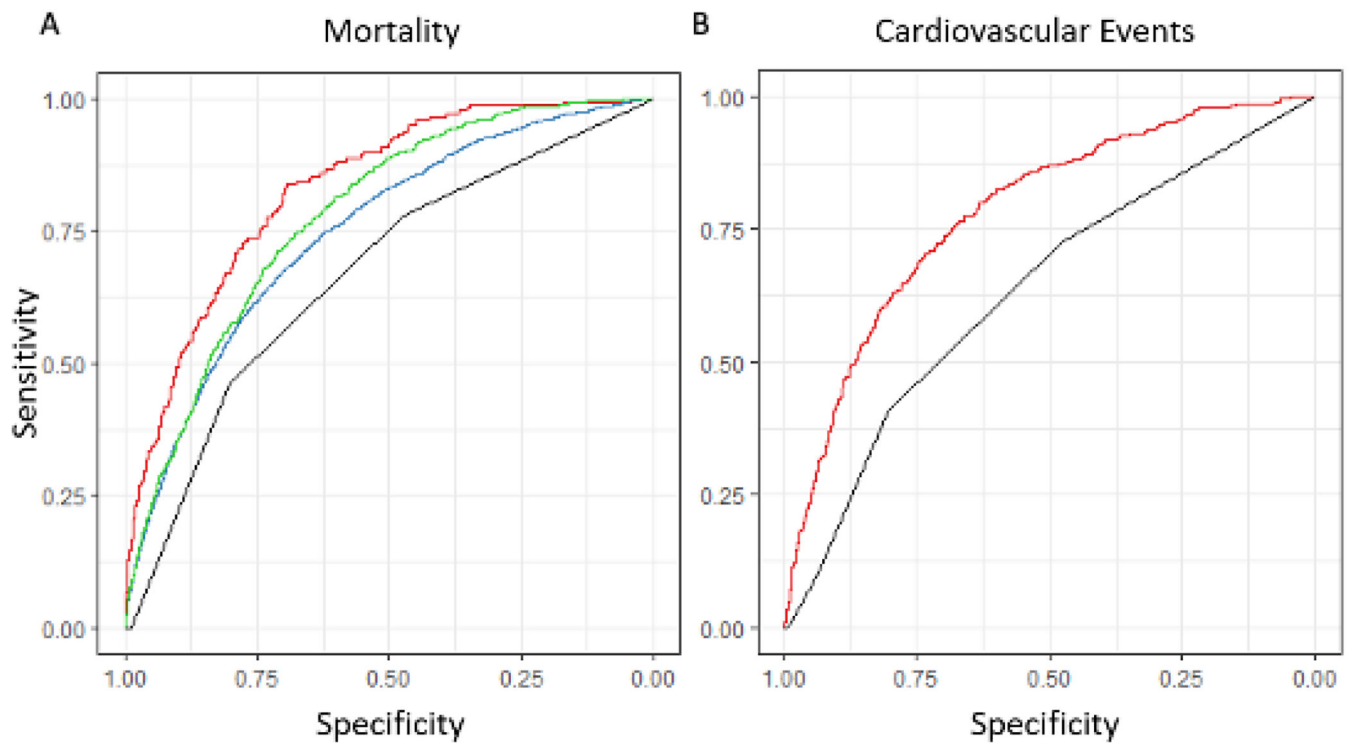


Figure 2. PreOpNet workflow and results.

Performance of PreOpNet at Cedars-Sinai Medical Center (CSMC, red), Stanford Healthcare (SHC, green), and Columbia University Medical Center (CUMC, blue), and Revised Cardiac Risk Index in discriminating (A) post-operative mortality and (B) major adverse cardiovascular events. E. Rate of mortality and complications by percentile of risk identified by PreOpNet. ECG = electrocardiogram. RCRI = Revised Cardiovascular Risk Index.

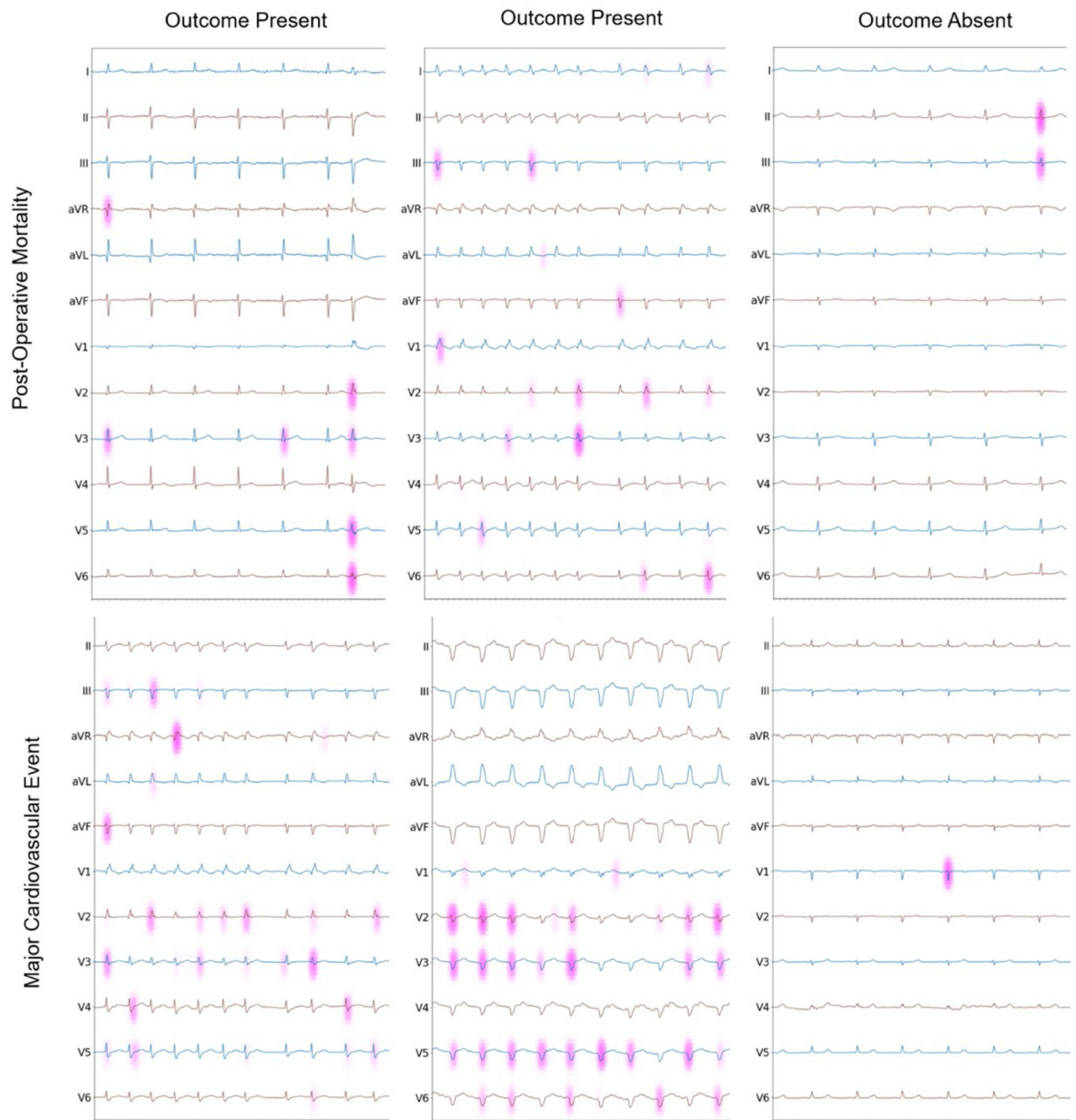


Figure 3. Interpretability Analysis of PreOpNet.

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

Table 1.

Characteristics and Outcomes of the Derivation Cohort

	Training Subcohort	Validation Subcohort	Test Subcohort
No. of Patients	36,839	4,549	4,581
No. of Procedures	48,033	6,013	5,929
No. of ECGs	90,633	11,217	10,944
Demographic and Clinical Characteristics			
Age, years (SD)	65.2 (15.8)	65.0 (16.4)	64.6 (16.1)
Female, n (%)	21744 (45.3)	2706 (45.0)	2641 (44.6)
Caucasian, n (%)	34412 (71.6)	4255 (70.8)	4143 (69.9)
Black	5323 (11.1)	662 (11.0)	739 (12.5)
Asian	3298 (6.9)	406 (6.7)	401 (6.8)
Other	4998 (9.9)	652 (10.5)	685 (11.6)
Heart Failure, n (%)	8689 (18.1)	1192 (19.8)	1119 (18.9)
Diabetes Mellitus, n (%)	7880 (16.4)	998 (16.6)	1061 (17.9)
Hypertension, n (%)	16638 (34.6)	2187 (36.4)	2063 (34.8)
Coronary Artery Disease, n (%)	10421 (21.7)	1293 (21.5)	1305 (22.0)
Stroke, n (%)	2697 (5.6)	352 (5.9)	343 (5.8)
Renal Disease, n (%)	4475 (9.3)	640 (10.6)	568 (9.6)
Procedure Types			
Cardiovascular, n (%)	19840 (41.3)	2595 (43.2)	2447 (41.3)
Intraperitoneal; intrathoracic; suprainguinal vascular, n (%)	8679 (18.1)	1074 (17.9)	1090 (18.4)
Insulin use prior to admission, n (%)	3022 (6.3)	366 (6.1)	374 (6.3)
Creatinine > 2 mg/dL, n (%)	4408 (9.2)	637 (10.6)	557 (9.4)
RCRI > 2, n (%)	2784 (5.8)	327 (5.4)	378 (6.4)
Post-Operative Outcomes			
Death during admission, n (%)	865 (1.8)	109 (1.8)	91 (1.5)
Cardiovascular death, n (%)	21 (0.0)	7 (0.1)	4 (0.1)
Major cardiovascular events, n (%)	1400 (2.9)	190 (3.2)	140 (2.4)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	Training Subcohort	Validation Subcohort	Test Subcohort
Cardiac arrest, n (%)	182 (0.4)	27 (0.4)	18 (0.3)
Myocardial infarction, n (%)	203 (0.4)	30 (0.5)	20 (0.3)
Heart block, n (%)	129 (0.3)	22 (0.4)	9 (0.2)
Pulmonary edema, n (%)	40 (0.1)	4 (0.1)	4 (0.1)

Table 2.

Discrimination of Post-Operative Outcomes by the PreOpNet or RCRI Algorithms.

Outcome	Model (Cohort)	AUC	Specificity	Sensitivity	PPV	NPV	Odds Ratio*
Death	PreOpNet (CSMC)	0.83 (0.79-0.87)	0.87 (0.86-0.88)	0.57 (0.48-0.68)	0.06 (0.05-0.08)	0.99 (0.99-0.99)	9.17 (5.85-13.82)
	PreOpNet (CUMC)	0.79 (0.75-0.83)	0.86 (0.80-0.92)	0.49 (0.48-0.50)	0.03 (0.02-0.03)	1.00 (0.99-1.00)	6.20 (3.87-10.41)
	PreOpNet (SHC)	0.75 (0.74-0.76)	0.93 (0.91-0.94)	0.32 (0.32-0.32)	0.02 (0.02-0.02)	1.00 (1.00-1.00)	5.88 (5.00-7.00)
	RCRI	0.67 (0.61-0.72)	0.94 (0.93-0.94)	0.12 (0.05-0.19)	0.03 (0.01-0.05)	0.99 (0.98-0.99)	2.08 (0.77-3.50)
MACE	PreOpNet (CSMC)	0.77 (0.73-0.80)	0.88 (0.88-0.89)	0.41 (0.33-0.49)	0.08 (0.06-0.10)	0.98 (0.98-0.99)	5.38 (3.75-7.49)
	RCRI	0.63 (0.59-0.68)	0.94 (0.93-0.94)	0.10 (0.05-0.15)	0.04 (0.02-0.06)	0.98 (0.97-0.98)	1.67 (0.77-2.68)

AUC = Area under the curve. MACE = Major adverse cardiovascular events. RCRI = Revised Cardiovascular Risk Index.

CSMC denotes results for the internal held-out test dataset from the derivation cohort, and SHC and CUMC denote results for the external validation cohorts.

* Unadjusted odds ratios for the outcome of interest (death or MACE) were calculated for patients in the $\geq 85^{\text{th}}$ percentile of risk compared to those in the $< 85^{\text{th}}$ percentile of risk, as determined by either PreOpNet or the RCRI (corresponding to an RCRI score ≥ 2).