Pre-Surgical Risk Stratification using Deep Learning on 12-lead ECGs for Non-Cardiac Populations

Carl Harris, Anway Pimpalkar, Ataes Aggarwal, Patrick Yang, Xiaojian Chen, Casey Overby Taylor, Joseph Greenstein, Robert D. Stevens

Department of Biomedical Engineering ● Johns Hopkins University ● Baltimore, MD

Precision Care Medicine ● Design Day 2024

Background

- Recent studies highlight that about 3% of non-cardiac surgeries lead to major adverse cardiovascular and cerebrovascular events (MACCE), translating to **1.51 million events in the U.S.** over a decade, with non-fatal heart attacks and strokes being the most common (**Fig. 1**).
- **Pre-surgical risk stratification** is vital to decisions around care allocation, informed consent of the patient, and perioperative treatment strategies. Despite this, existing method suffer from poor discriminative capacity.
- **Electrocardiogram (ECG)**, a non-invasive method to record the electrical activity of the heart, has historically been **underutilized** in predicting post-surgical complications, despite routine use in clinical settings.
- ECGs provide a cost-effective and accessible **solution** for monitoring various physiological markers in patients.

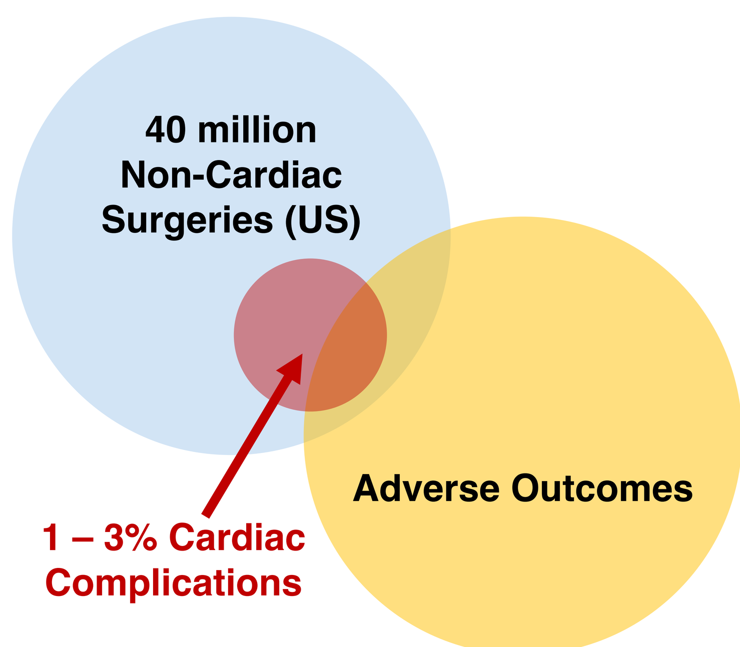

Fig 1. Understanding the problem.

Project Aims

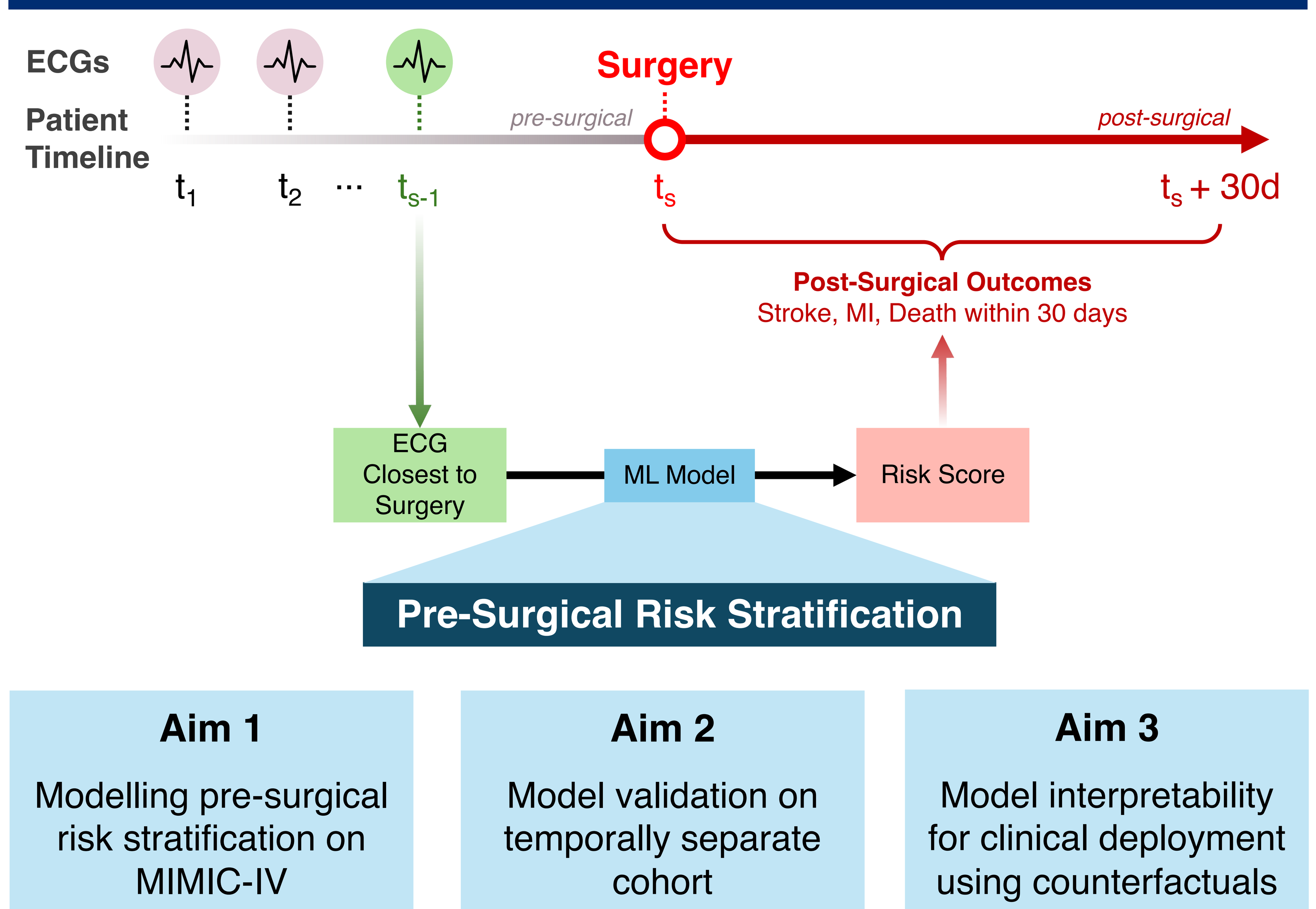

Fig 2. Tasks of interest described as a patient timeline.

- Our objective is to develop a **deep learning model** that enhances current techniques for processing ECG signals to predict adverse surgical outcomes, including myocardial infarction (MI), stroke, and death.
- For model validation, we are testing its performance on a **temporally distinct cohort**.
- To elucidate the model's methodology in assessing various ECGs and generating a risk score, we plan to employ a **counterfactual explanation** approach.

Methodology

Approach

- We develop our model in the MIMIC-IV dataset, identifying patients with pre-surgical ECG undergoing major cardiac (n=29,527) or noncardiac (n=54,870) surgery (**Fig. 3**).
- Our primary **outcomes** of interest are post-operative MI, in-hospital mortality (IHM), and a composite of stroke, MI, or death within 30 days.
- We develop a **waveform model** that uses a Convolutional Neural Network (CNN) to predict outcomes using only waveforms (**Fig. 4L**).
- We also use a **fusion model**, which combines the CNN with relevant demographic (age, sex), admission type, clinical (RCRI) variables, and Elixhauser components (**Fig. 4R**).

Interpretability

- Interpretable, explainable models are critical to clinical adoption. Existing approaches to interpretability in CNN-based ECG classifiers rely on saliency maps, which shows "where" the model is looking, but not "what" it's looking at.
- **Counterfactual** explanatory models generate plausible modifications that allow for visualization of morphological differences relevant to the classification verdict[6,7].

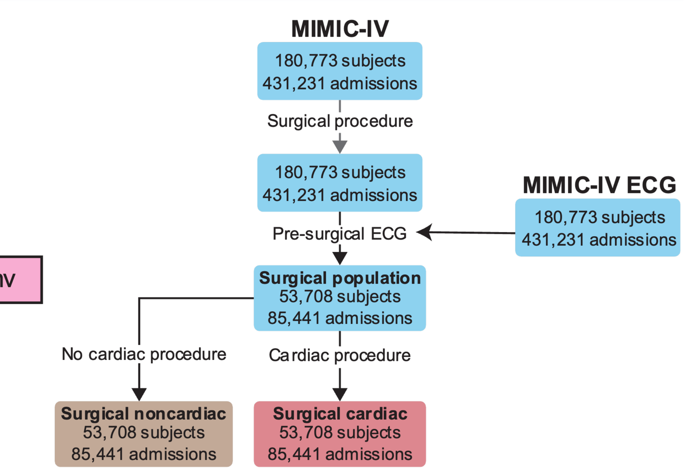

Fig 3. Population selection criteria.

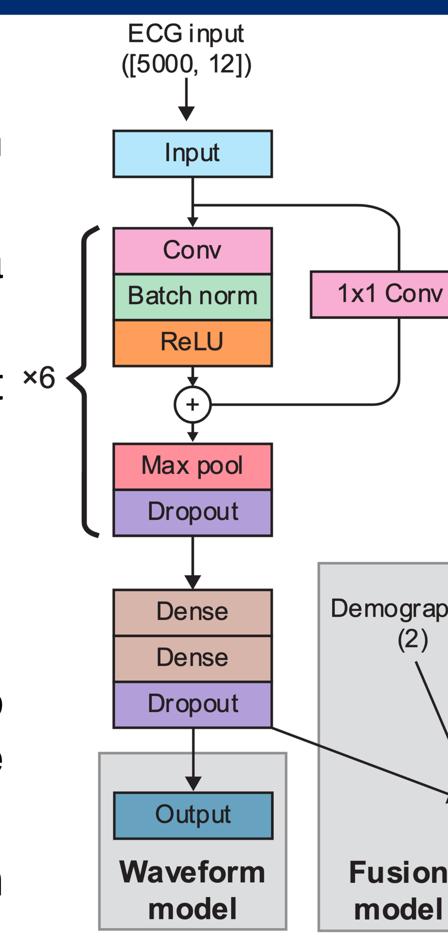
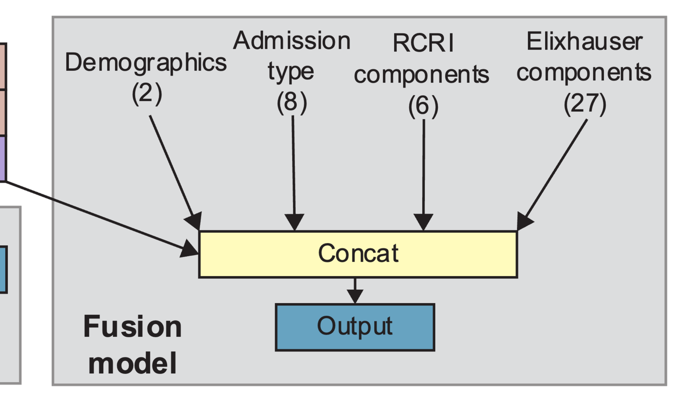

Fig 4. Model architecture. **L.** Convolutional Neural Network (CNN); **R.** Fusion variables.

Results

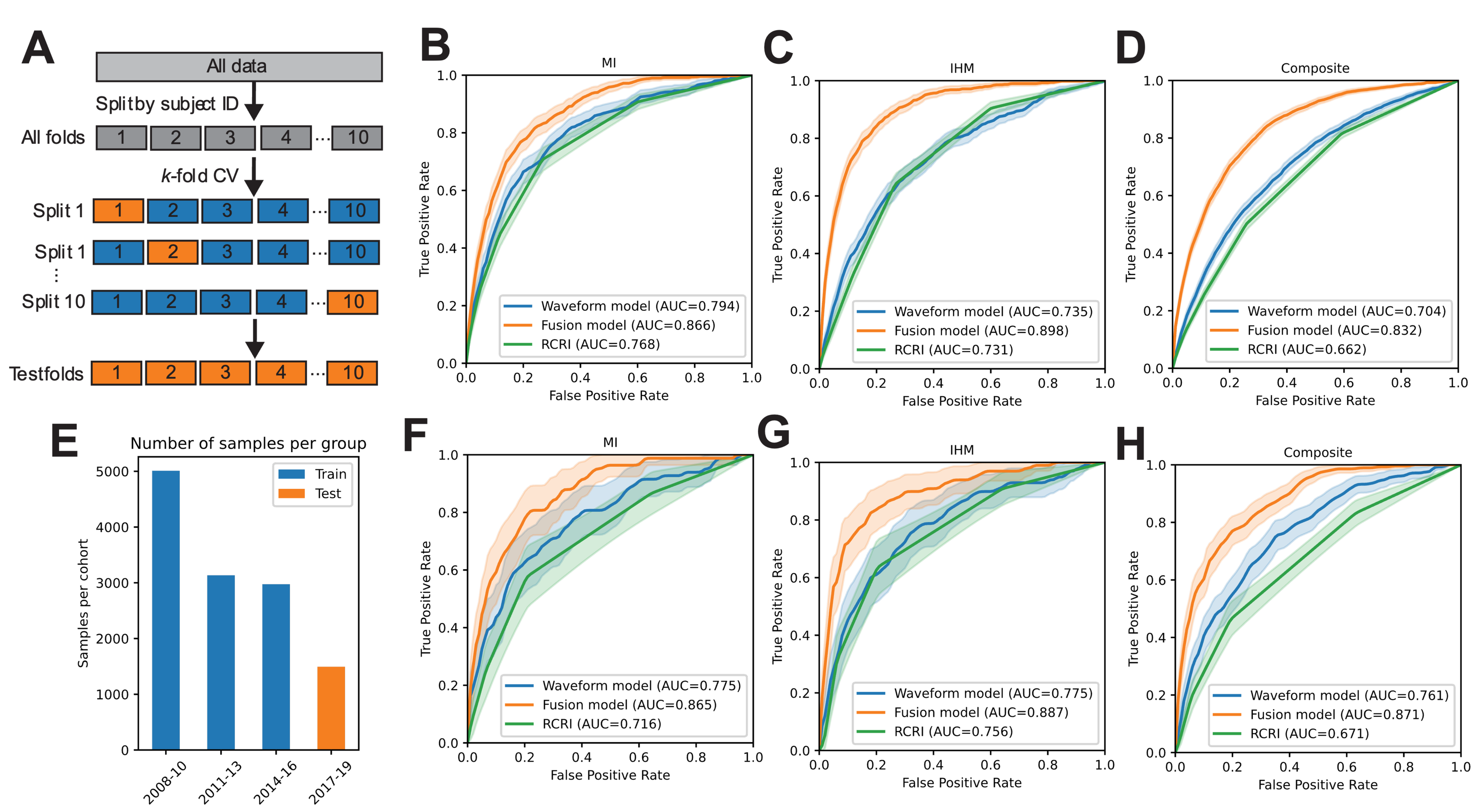

Fig 5. Results of k-fold validation and temporal stratification, done independently.

- **Fig. 5A** depicts the k-fold validation scheme for the main results, where data are split into 10 folds (by patient ID).
- Test folds are combined to evaluate the final performance of the model, as shown by the ROC curves (with AU-ROC values inset) in predicting post-operative MI (**Fig. 5B**), IHM (**Fig. 5C**), and composite outcome (**Fig. 5D**).
- In MIMIC-IV, our **fusion model significantly outperforms both the waveform-only model and RCRI** ($p < 0.001$) in the noncardiac surgical population.
- **Fig. 5E** shows the number of samples in the temporal stratification analysis. Blue bars are used to train the model (corresponding to patients admitted between 2008 and 2016), and samples corresponding to the orange bar (2017-19) are used to evaluate it. The corresponding performance is shown in panels **Fig. 5F, 5G and 5H**.

Interpretability

- Our trained model, given an ECG input, produces a risk score for a given set of outcomes.
- Our **counterfactual model** (**Fig. 6**) intervenes on this input waveform by **introducing minimal, physiologically plausible modifications** to the underlying morphology such that it **elicits a different user-defined risk score** corresponding to a particular outcome.

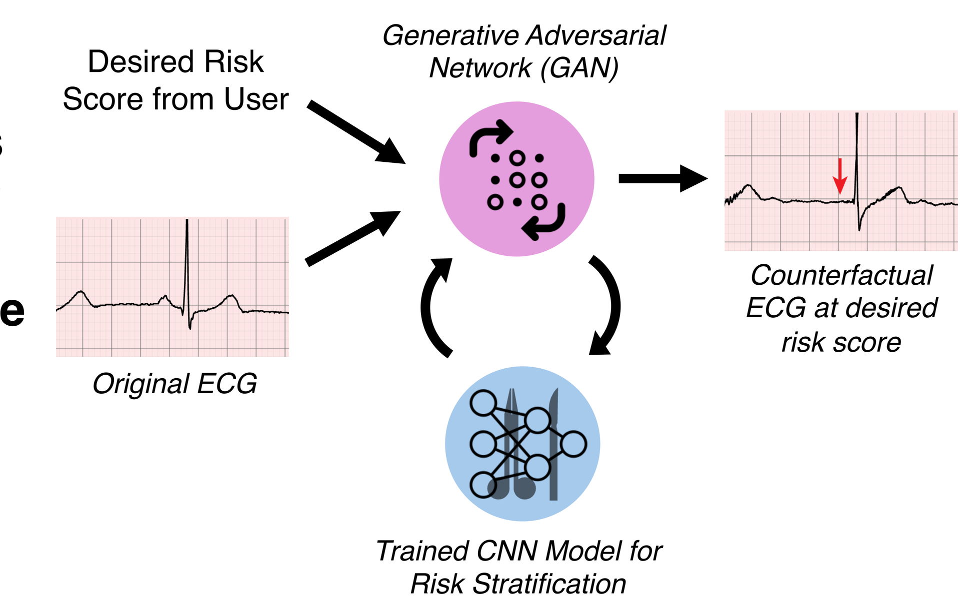

Fig 6. Process of generating counterfactuals.

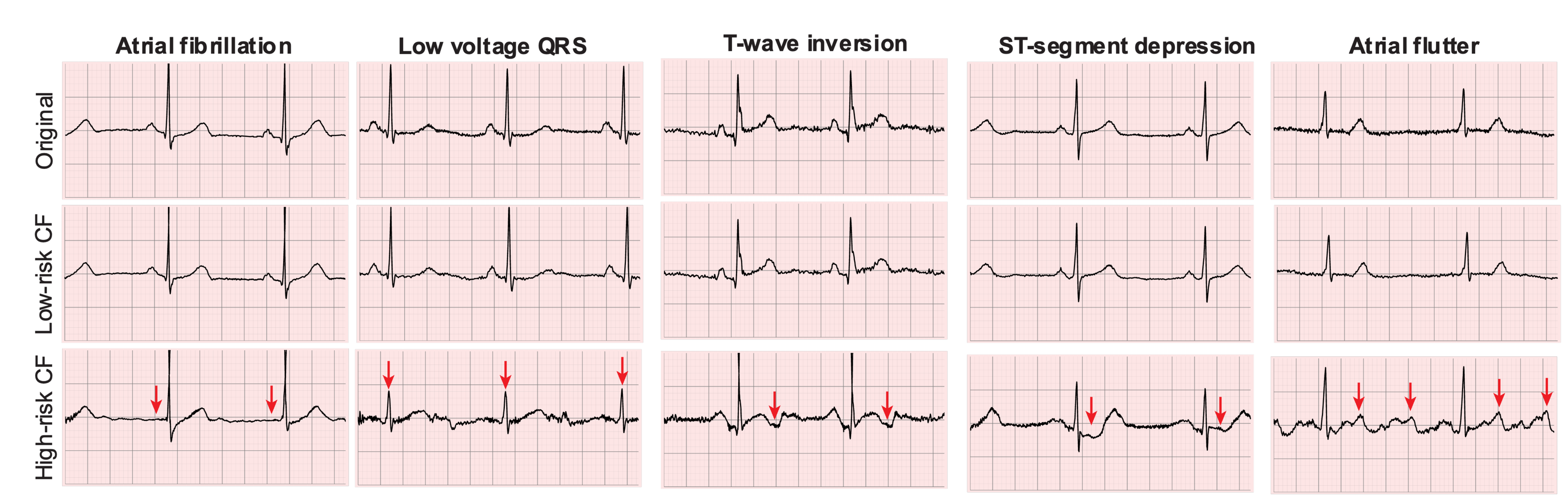

Fig 7. Examples of generated counterfactuals to validate the efficacy of the method, with red arrows indicating the changes induced by the generator.

Next Steps

- External validation on a separate dataset to eliminate geographical population biases and establish robustness.
- Hyperparameter tuning for the developed models.

Acknowledgements
This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 2139757, awarded to CH.

Scan for bibliography