USING ECG WAVEFORM DATA TO PREDICT CTRCD IN ICI PATIENTS

Dr. Abdullah Sarkar | Hamza Khan |
Dr. Rohit Jain | Prof. Jeff Chiang

UCLA Health

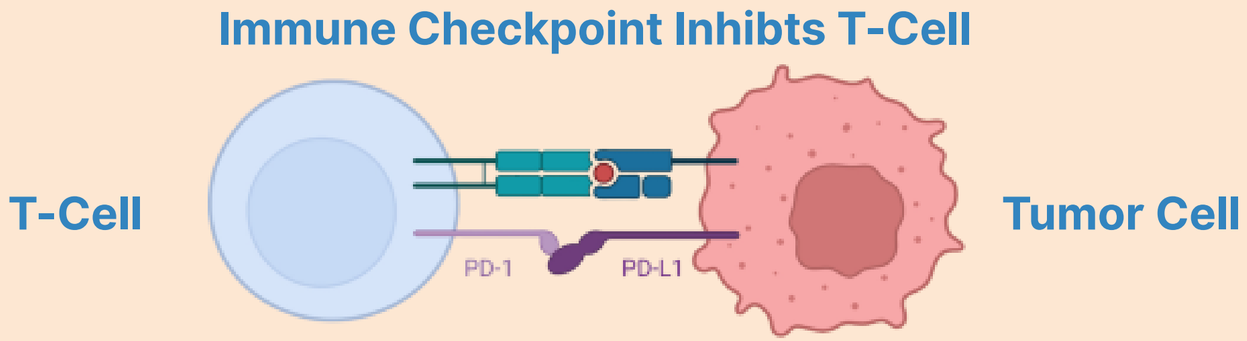
AI ECG

STUDY DESIGN

AI ECG

Our population of interest is cancer patients who received Immune Checkpoint Inhibitor (**ICI**) therapy and later presented Cancer Therapy Related Cardiac Dysfunction (**CTRCD**)

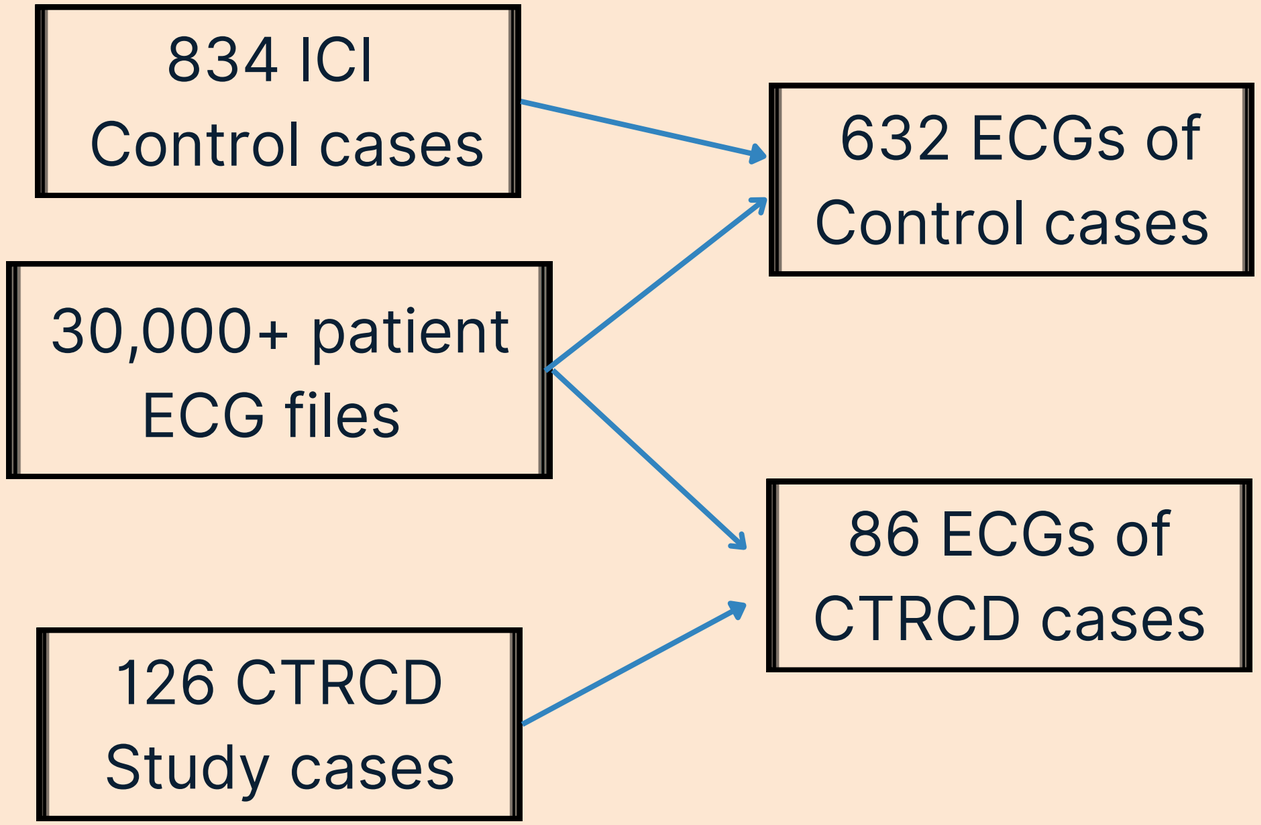
Using ECG waveform data, comorbidities, and patient demographics, we seek to train a model that predicts if an ICI patient is likely to develop CTRCD



ECG AND PATIENT DATA

AI ECG

- Patients' ECG waveform data is stored in XML files.
- XML files include:
  - IPPAT – primary identification for patients across databases
  - Patient Demographics
  - 2 sets of 8 leads: median beat and 10s waveform
  - Notes of abnormalities
- Supplement with other CSVs that detail patient comorbidities and treatments



Include patients that have had an ECG completed 4 years **before** or 1 month **after** begining ICI therapy

MODEL FEATURES

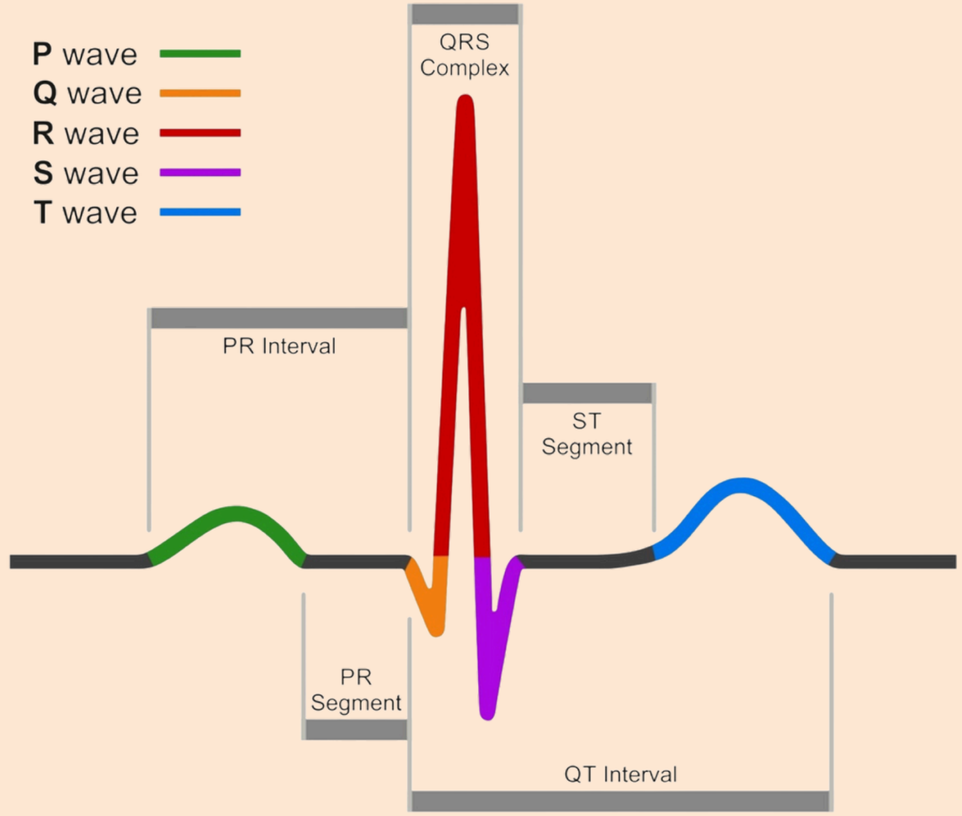
AI ECG

- **Waveform Data – 10s, 250 Hz recording from 8 leads:**
  - Lead I, Lead II, V1, V2, V3, V4, V5, V6
- **Measurements and Intervals:**
  - Atrial Rate, Ventricular Rate, Heart Rate, Baseline Ejection Fraction, Creatinine, QT Interval, QRS Duration, T Axis
- **Patient Demographics and Info:**
  - Patient Age, Gender, Race (*removed*), Age at first ICI, Lifetime count of ICIs
- **Clinical Covariates:**
  - Smoke, ACS before ICI, Arrythmia before ICI, CAD before ICI, HF before ICI, Cardiac Arrest before ICI, Stroke before ICI, Hypertension, Hyperlipidemia, Diabetes, ICI-Group

DATA PREPARATION

AI ECG

- XMLs had varying sample counts between ECG so we regularized all to 2500 sample count
- In some models, we increased our training samples by dividing the 10s waveform strip using a 2 second sliding window
- Decode the base64-encoded raw signal waveforms into signed 16-bit values
- Use Neurokit2 to isolate intervals of interest, including QT-interval
- Remove the *Race* feature because there were too many unknowns/missing values



MODEL STRUCTURES AND EVALUATION

AI ECG

01 200-Tree Random Forest

02 Two Tower: MLP and 1-D CNN

03 Augmented Two Tower

04 1-D CNN Waveform Isolation

05 1-D CNN with 2s Window

MODEL 1

AI ECG

200-Tree Random Forest

- All features excluding waveform – no leads
- Develop a baseline understanding of predictive power of non-waveform data
- K-5 Fold Cross Validation
- **ROC-AUC: 0.61 (±0.04)**
  - Only modestly better than guessing
  - Patient Demographics, numerical data, covariates and comorbidities don’t offer unexpectedly strong predictive power
- Strong class imbalance requires regularization

MODEL 2

AI ECG

Two-Tower: MLP + 1-D CNN

- Waveform tower: a 1-D CNN over the 10 s, 8-lead signals
- Tabular tower: an MLP over numerical + categorical features
- Fusion & classifier: concatenation → dense → sigmoid
- Stratified 5-fold CV, with class weights to help the imbalanced CTRCD class
- **ROC-AUC: 0.64 ± 0.063**
  - significant learning within the waveforms
  - high variance suggests model isn't stable between folds
  - Model still needs hyperparameter tuning or more regularization and augmentations

MODEL 3

AI ECG

Augmented Two-Tower

- Same Two-Tower approach with some changes to prevent overfitting:
  - Gaussian Noise, Random Time-Shift , Random cropping, time-warp, lead-drop out
- Stratified 5-fold CV, with class weights to help the imbalanced CTRCD class
- **ROC-AUC: 0.66 ± 0.043**
  - Improvement with augmentation but still need further tuning or structural change to make novel predictive gains

MODEL 4

AI ECG

1-D CNN

- Dial back to a simple 1-D CNN approach with just the waveform
  - Seeking to understand the predictive power of the 10s waveform strip
- **ROC-AUC: 0.63 ± 0.06**
  - Significantly better than the random forest trial
  - Likely suffering to overfitting because of the small positive class – only 86 CTRCD cases

MODEL 5

AI ECG

1-D CNN with 2s non-overlapping windows

- Divide the ECG into 5 windows of 2 seconds each to bloat training samples
- Run on simple 1-D CNN to gauge any novel predictive power
- Did not complete this model

NEXT STEPS

AI ECG

- We were not able to produce a model with sufficient predictive power within the given time
- Made signficant strides between model iterations
- In future models:
  - develop the 2 second window: overlapping vs non-overlapping
  - More hyperparamter tuning
  - Trial and error for new structures until something sticks

AREAS OF STRUGGLE

AI ECG

- Project progress was stalled primarily in early phases:
  - Access to data
  - Access to computing platforms and packages
  - Adjusting to ULEAD firewalls
  - Finalizing study design
- Model structures struggled to perform with 85 cases – more prone to overfitting

ACKNOWLEDGEMENTS

Special Thank You to
Dr Stein-Merlob
Prof Chiang
Kevin Muy

THANK YOU

AI ECG