Published in final edited form as:

Circulation. 2026 January 06; 153(1): 35–46. doi:10.1161/CIRCULATIONAHA.125.077494.

Deep Learning–Based Continuous QT Monitoring to Identify High-Risk Prolongation Events After Class III Antiarrhythmic Initiation

Rayan A. Ansari, BS[*],
Sabyasachi Bandyopadhyay, PhD[*],
Rishi K. Trivedi, MD, PhD,
Kelly A. Brennan, MS,
Xichong Liu, MD,
Prasanth Ganesan, PhD,
J. Weston Hughes, PhD,
Alexander C. Perino, MD,
Euan A. Ashley, MD, PhD,
Paul J. Wang, MD,
Todd Coleman, PhD,
Marco V. Perez, MD,
David Ouyang, MD,
Sanjiv M. Narayan, MD, PhD,
Albert J. Rogers, MD, MBA

Department of Medicine (R.A.A., S.B., K.A.B., X.L., P.G., A.C.P., E.A.A., P.J.W., M.V.P., S.M.N., A.J.R.), Cardiovascular Institute (S.B., K.A.B., X.L., P.G., A.C.P., E.A.A., P.J.W., M.V.P., S.M.N., A.J.R.), Department of Biomedical Data Science (R.A.A.), School of Medicine, Department of Bioengineering (T.C.), Stanford University, CA. Department of Biomedical Informatics (J.W.H.), Milstein Division of Cardiology (J.W.H.), Department of Medicine (J.W.H., X.X.), Columbia University Irving Medical Center, New York, NY. Department of Cardiology, Cedars-Sinai Medical Center, Los Angeles, CA (R.T., D.O.). Division of Research, Kaiser Permanente Northern California, Pleasanton, CA (D.O.).

Abstract

Correspondence to: Albert J. Rogers, MD, MBA, Department of Medicine, Cardiovascular Institute, Stanford University, 870 Quarry Rd Ext, Stanford, CA 94305. rogersaj@stanford.edu.
[*]R.A. Ansari and S. Bandyopadhyay contributed equally.

This work was presented as an abstract at AHA Scientific Sessions, New Orleans, LA, November 7–10, 2025.

Disclosures

R.A. reports intellectual property rights from Stanford University. S.B. reports consulting fees from Linus Health Inc. and intellectual property rights from Stanford University. E.A.A. reports consulting fees from Apple Inc. M.V.P. reports consulting fees from Apple Inc., Boston Scientific, Biotronik Inc., Bristol Myers Squibb, QALY, and Johnson & Johnson, and has an equity interest in QALY. D.O. reports consulting fees or equity in Ultromics, InVision, EchoIQ, and Pfizer. S.M.N. reports consulting compensation from Abbott Inc., UpToDate, and LifeSignals.ai, and intellectual property rights from the University of California Regents and Stanford University. A.J.R. reports consulting fees or equity in WearLinq Inc., YorLabs Inc., and EBR Systems, and intellectual property rights from Stanford University. The other authors report no conflicts.

BACKGROUND: Drug-induced QT prolongation after successful inpatient loading of class III antiarrhythmics may occur during routine outpatient care. Insertable cardiac monitors offer continuous signals but are limited by single-lead configuration. We hypothesized that a spatially aware deep learning system (3DRECON-QT) can reconstruct spatial information from a single lead vector to quantify QT/QTc and identify high-risk prolongation.

METHODS: We developed 3DRECON-QT using a multitask encoder–decoder that ingests a 10-s single-lead signal, reconstructs 12 leads, and predicts QT/QTc. The model was developed using 12-lead ECGs with clinician-adjudicated QT/RR from a large health system and tested in an external center with different ECG hardware. Continuous monitoring performance was assessed in a public dofetilide-loading data set with serial ECGs. In a real-world cohort of outpatients on dofetilide or sotalol presenting to the hospital or emergency room for any reason, rates of ventricular arrhythmias and QT prolongation were assessed. Device validation was tested in patients with insertable cardiac monitor recordings paired with clinical 12-lead ECGs.

RESULTS: 3DRECON-QT classified prolonged QTc from single-lead signals with area under the receiver operating characteristics curve, 0.942 (mean absolute error, 17.5 ms) in the internal test set and 0.943 (mean absolute error, 21.1 ms) externally. During continuous dofetilide monitoring, predictions correlated with ground truth (r, 0.851; mean absolute error, 17.8 ms; area under the receiver operating characteristics curve, 0.936 for prolonged QTc, 0.816 for ≥15% QTc rise). QTc prediction from true insertable cardiac monitor recordings showed r=0.824 and mean absolute error, 17.5 ms. In outpatients on class III antiarrhythmics (n=1676), 16.5% had high-risk QTc prolongation. Ventricular arrhythmia events were 3.97% versus 0.86% without prolongation (adjusted odds ratio, 4.24 [95% CI, 1.81–9.90]). 3DRECON-QT detected these events with area under the receiver operating characteristics curve 0.94 (F1 score, 0.60).

CONCLUSIONS: A single-lead, deep-learning approach can achieve guideline-level measurement accuracy, enable continuous QTc surveillance from nonstandard ECG vectors, and identify clinically meaningful outpatient QTc prolongation associated with a >4-fold increase in serious ventricular arrhythmias. This strategy may enhance safety monitoring after class III antiarrhythmic initiation and support targeted intervention.

Keywords

anti-arrhythmia agents; arrhythmias, cardiac; deep learning; drug-related side effects and adverse reactions; electrocardiography; monitoring, ambulatory; Torsades de Pointes

---

The QT interval is a critical electrocardiographic marker for assessing the risk of life-threatening arrhythmias, including torsades de pointes and sudden cardiac death.[1,2] However, its measurement requires a 12-lead ECG, which is impractical for continuous, long-term monitoring in ambulatory care settings.[3]

Patients initiated on QT-prolonging drugs such as dofetilide or sotalol are monitored as inpatients, but QT prolongation may occur between periodic 12-lead ECGs, and may fluctuate beyond discharge in patients with electrolyte fluctuations such as those with renal dysfunction or heart failure.[4] Insertable cardiac monitors (ICMs) provide a platform for continuous rhythm monitoring, yet their nonstandard, single-lead configuration precludes direct QT interval measurement according to society guidelines.[5]

We hypothesized that 3DRECON-QT, a spatially informed, multitask architecture applied to single-lead ECG signals, can accurately reconstruct the 12-lead ECG to identify clinically relevant QTc prolongation and reveal hidden burdens of arrhythmia risk. We have previously shown that machine learning of intracardiac electrograms can identify patients at risk for sudden cardiac arrest and suggests cellular phenotypes associated with that risk.[6] We have also presented a deep learning model of the 12-lead ECG alone to identify patients with wall motion abnormalities.[7] This study extends these principles to reconstruct a multilead ECG from a nonstandard, narrowly spaced ECG signal (as obtained from an ICM or chest-worn patch monitor) to quantitatively measure QT interval in a broad group of patients. Further, in a patient population with class III antiarrhythmic (dofetilide or sotalol) exposure, we assess the incidence QT prolongation and the risk of significant arrhythmia events after successful load of the medications.

To test this, we aimed to: (1) develop and validate 3DRECON-QT using large and diverse internal and external ECG data sets, respectively; (2) assess its ability to detect drug-induced QT prolongation during continuous monitoring; and (3) quantify the real-world clinical risk associated with outpatient QT prolongation and validate the ability of 3DRECON-QT to identify high-risk events in this real-life cohort.

METHODS

Study Populations and Data Sets

Model development was performed at Stanford University and tested at an independent external site (Cedars-Sinai Medical Center). The study was approved by the institutional review boards of both centers. A waiver of consent was granted for retrospective analysis. Across these sites and a public database, 5 distinct patient cohorts were established. These comprised the Stanford Model Development Cohort, Cedars-Sinai External Validation Cohort, Continuous QT Monitoring Cohort (ECG Effects of Ranolazine, Dofetilide, Verapamil, and Quinidine PhysioNet Dataset), Stanford Real-World Class III Antiarrhythmic Exposure Cohort, and the Stanford ICM Recording Cohort. The data that support the findings of this study are available from the corresponding author upon reasonable request. A dataset overview and model input for each analysis is available in Table S1.

Stanford Model Development Cohort: We analyzed 794 813 12-lead ECG data obtained from 258 205 unique patients in the Stanford Hospital ECG database (iECG, Phillips, Amsterdam, The Netherlands). The data collection period extended from 2005 to 2018. Patients were randomly divided into training (80%), validation (10%), and testing (10%) cohorts to ensure all ECGs from any single patient were assigned exclusively to one data partition. The test set (10%) included 78 245 12-lead ECG recordings from 25 821 patients not previously seen by the model during training. Exclusion criteria included records in legacy formats before system migration, poor signal quality (>70% of leads 0-valued or flatlined), incomplete or absent machine annotations, and missing QT or RR interval measurements. All rhythms, including atrial and ventricular paced ECGs, were

included. The consort diagram for the development and validation cohorts is provided in Figure 1.

Cedars-Sinai External Validation Cohort: ECGs (N=1 048 575) from a geographically distinct health system and alternate ECG acquisition equipment (MUSE-GE system) were collected to test model generalization between January 1, 1900, and December 31, 2023. Exactly 100 000 (from 65 291 unique patients) ECG recordings were selected from all consecutive ECGs with complete lead data and calculated measures. Patient characteristics of the Stanford Model Development and the Cedars-Sinai External Validation cohorts are provided in Table 1 (by patient) and in Table S2 (by ECG recording).

Continuous QT Monitoring Cohort (ECGRDVQ PhysioNet Data Set): A total of 1056 serial 12-lead ECGs from 22 healthy patients undergoing dofetilide loading, were each monitored for 24 hours after drug initiation.8 This was used specifically for the simulated continuous monitoring analysis. Device information for the ECGRDVQ data set was not specified.

Stanford Real-World Class III Antiarrhythmic Exposure Cohort: A separate retrospective clinical-encounter cohort of Stanford patients treated with dofetilide or sotalol, included 72 919 ECGs from 1676 unique patients across 2083 outpatient encounters (Stanford, 2008–2024). This cohort was used to assess the real-world clinical problem and validate 3DRECON-QT in the target population.

Stanford ICM Recording Cohort: The Medtronic CareLink database was accessed for all patients with an ICM implant (LINQ I and II, Medtronic, Minneapolis, MN) between 2016 and 2024. Only patients whose final transmission occurred during the period were accessible through the archives of older studies. Transmissions from a total of 387 patients were available. The timestamps from the presenting rhythm were compared with 12-lead ECGs acquired during routine clinical care. To ensure comparable physiological states between the 12-lead ECG and the ICM recordings, we included pairs obtained within ±7 days and with a heart rate difference |Δ heart rate |≤20 beats per minute. Twenty-six patients met these inclusion criteria.

ECG Recording and Preprocessing

All ECG recordings used in the Stanford Model Development Cohort and the Real-World Class III Antiarrhythmic Cohort were acquired using the Philips iECG system. Cedars-Sinai external testing recordings were obtained through the MUSE-GE system. We extracted 10-s, 12-lead ECG segments and standardized all signals to a 500-Hz sampling rate. A 0-phase, fifth order Butterworth bandpass filter (0.5–40 Hz) was applied to remove baseline wander and high-frequency noise.

In all analyses, the input to 3DRECON-QT is a single-lead, narrowly spaced chest wall signal. For model development, external testing, continuous monitoring, and real-world class III analyses, the input vector was a derived ICM signal calculated voltage differential between precordial leads V3 and V2. True ICM recordings were the input for the Stanford ICM Recording analysis. Twelve-lead ECGs were used as the target for reconstruction and

the basis for gold standard QT labels. A summary of model input and 12-lead ECG use is provided in Table S1.

We standardized each recording by subtracting the record-wide mean and dividing by its SD, so that all leads entered the network with zero mean and unit variance. We excluded recordings containing >70% 0-valued samples or any lead with both 0 mean and 0 SD. In addition, recordings lacking cardiologist-interpreted QT or RR annotations were removed. The CONSORT diagram for the development and external test sets is provided in Figure 2.

QT Annotation

Ground truth QT and RR interval measurements for our development and external validation cohorts were extracted from the clinical ECG systems after adjudication by cardiologists during routine clinical care. QTc values were calculated from the measured QT intervals and RR intervals using the Bazett formula.[9] Across both data sets, a prolonged QT was defined according to the corrected QT interval and the QRS duration according to the dofetilide Food and Drug Administration drug insert labeling.[10] For a narrow QRS (≤100 ms), a prolonged QTc was defined as >500 ms. For a wide QRS (>100 ms), a QTc >550 ms was determined to be prolonged.[11] Based on a mixed-model pooled variance calculation (Detailed Methods), the interobserver variance (caused by reader bias) is 47.1 ms^2 (SD, 6.86 ms) whereas the within-observer variance (caused by physiological variability, repeat measures, noise, etc) is 2392.7 ms^2. This results in a variance ratio of 0.02, or a proportion of variance of 0.02, suggesting minimal effect of interobserver labeling on QT assessment and clinically robust QT labels (Figure S1).

Development of 3DRECON-QT Model

The 3DRECON-QT model is a deep learning architecture designed to reconstruct a full 12-lead ECG and estimate the QT interval from a single-lead input. It employs an encoder and decoder framework with a multitask learning objective, encouraging the network to learn physiologically and spatially meaningful latent features that jointly inform both waveform reconstruction and QT prediction (Figure 2).

A 10-s single-lead ECG signal with 500-Hz sampling is processed through a squeeze-and-excitation ResNeXt encoder, which extracts a compact latent representation of cardiac electrical activity.[12,13] Two task-specific “heads” operate in parallel: a reconstruction head, which generates the 12-lead ECG by conditioning on the spatial coordinates (θ and ϕ) of each lead,[14] and a QT prediction head, which applies a transformer decoder to the shared latent representation to estimate the QT interval.[15] Model optimization used a weighted composite loss across the 2 tasks, stochastic gradient descent with cosine annealing, and early stopping based on validation performance.

Implementation was performed in Python (PyTorch v2.3.1).[16] Detailed architectural, training, and hyperparameter details are provided in the Supplemental Methods.

Continuous QT Monitoring Analysis (ECGRDVQ)

To assess the performance of 3DRECON-QT to detect changes in QT interval during continuous monitoring, we validated the model in the Continuous QT Monitoring Cohort (ECGRDVQ PhysioNet Dataset), which investigates the pharmacokinetic profile of dofetilide taken with oral ingestion during safety studies of the medication.[8] We assessed 3DRECON-QT for detecting drug-induced QT prolongation using 24-hour ECGs from the ECGRDVQ dofetilide-loading data set (simulating ambulatory monitoring). The input signal is the same as described in the "ECG Recording and Preprocessing" section. We examined the temporal trends of aggregated mean predicted and ground truth QT/QTc. We performed classification analysis for prolonged QT interval, using the thresholds as described above for dofetilide loading and, secondarily, a ≥15% increase in QTc from baseline.[17]

Real-World Class III Antiarrhythmic Exposure Analysis

To assess the clinical impact of an enhanced monitoring strategy in a high-risk patient population, we conducted a retrospective encounter-based analysis using the Stanford Real-World Class III Antiarrhythmic Exposure Cohort, a data set comprising 1676 unique patients receiving outpatient dofetilide or sotalol treatment at Stanford between 2008 and 2024 at the time of a hospital or emergency room encounter for any reason. We categorized the primary diagnosis from the encounter using the Clinical Classifications Software Refined for *International Classification of Diseases, Tenth Revision,* clinical modification codes to express which types of clinical encounters were associated with prolonged QT events on the first ECG of each encounter.[18] We then established a case-control framework for serious arrhythmic events (torsades de pointes, ventricular fibrillation, and sudden cardiac death), calculating adjusted odds ratios (ORs) using the Fisher exact test. Last, we ensured that QT assessments in this high-risk population were appropriately classified using 3DRECON-QT performance on this clinical data set using classification analysis and receiver operator characteristics.

Stanford ICM Recording Analysis

We evaluated 3DRECON-QT using Medtronic LINQ ICM signals paired with a clinical 12-lead ECG obtained within ±7 days. Analyses were performed at the patient level (one paired QTc value per patient). For each patient, 3DRECON-QT generated predictions from the real ICM recordings, with the paired 12-lead ECG providing the reference QTc value. We repeated the QTc analysis using the derived ICM signal from the paired 12-lead ECG. Agreement with the reference QTc across patients was assessed using Pearson correlation coefficient and the mean absolute error (MAE) between the prediction and ground truth. We also compared waveform morphology between the real and simulated ICM signals and between the corresponding, model-predicted 12-lead reconstructions. R-peaks were detected using the Neurokit2 Python package, after which individual beats were segmented, temporally aligned to the R-peak, and normalized before overlay.[19]

Statistical Analysis

Statistical analysis was performed in Python (v3.8) using SciPy and Scikit-learn.[20] Descriptive statistics were reported as mean±SD or median [interquartile range]. For QT

prediction performance, agreement between predicted and ground truth values was assessed using Bland-Altman analysis, Pearson correlation coefficient (r), and linear regression to calculate R^2, and MAE. Classification performance for prolonged QTc detection was evaluated using sensitivity, specificity, positive predictive value, negative predictive value, F1 score, area under the receiver operating characteristics curve (AUROC), and area under the precision-recall curve (AUPRC). CIs (95% CI) were generated using nonparametric bootstraps (1000 iterations, random sampling with replacement at the ECG level). Sensitivity analysis used patient-level clustered bootstrap with bias-corrected and accelerated intervals to account for multiple ECGs per patient. Association between QTc prolongation and clinical events was assessed using the Fisher exact test for univariate analysis and multivariable logistic regression to calculate adjusted ORs controlling for age, heart failure, atrial fibrillation, drug class, and laboratory values (serum creatinine, magnesium, and potassium). P values <0.05 were considered statistically significant.

RESULTS

Participants

There were no significant differences between patient characteristics across the internal splits of the Stanford Model Development Cohort (aggregate demographics: mean age, 61.9 years; 52.2% male; race or ethnicity, White 57.6%, Asian 13.4%, Black 5.7%, Hispanic or Latino 12.7%). The Cedars-Sinai External Validation Cohort contained older individuals with a higher proportion of Black individuals compared with the Stanford Model Development Cohort (Table 1).

3DRECON-QT Model Performance

Performance on the Development and External Validation Cohorts: Prolonged-QT classification using derived ICM recordings achieved an AUROC of 0.942 (95% CI, 0.941–0.943), AUPRC of 0.632 (95% CI, 0.626–0.639), and an accuracy of 0.957 (95% CI, 0.955–0.957) in the test partition of the Stanford Model Development Cohort. In the Cedars-Sinai External Validation Cohort, performance was similarly high, with an AUROC of 0.943 (95% CI, 0.941–0.945), AUPRC of 0.781 (95% CI, 0.774–0.789), and an accuracy of 0.925 (95% CI, 0.923–0.926) (Figure 3A).

The median ground truth QTc was 411 ms (interquartile range, 391–441 ms). In the test split of the Stanford Model Development Cohort, the model-predicted median QTc was 418 ms (interquartile range, 401–441 ms). The Bland-Altman analysis for QTc revealed a bias of –5.46±29.38 ms (Figure 3C).

To qualify 12-lead ECG reconstructions systematically, we analyzed performance across the Stanford Model Development test set, achieving mean Pearson correlation of 0.70 across all 12 leads (Table S3). Saliency analysis of the output signals shows activation across leads at clinically expected time points, including QRS onset and T wave offset along with accurate 12-lead reconstructions (Figure S2). To evaluate the contribution of the spatial encoding to QT prediction, we compared 3DRECON-QT with an ablated variant without the theta (spatial angle) encoder. The model with spatial encoding achieved QTc MAE of 25.9 ms and

Pearson r of 0.73, whereas the ablated model without spatial encodings showed markedly degraded performance (QTc MAE, 39.23 ms; Pearson r, 0.04), despite similar training loss convergence (Figure S3).

A linear regression of the predicted QT (uncorrected) resulted in an R^2 value of 0.747 (95% CI, 0.742–0.752) and a Pearson correlation coefficient of 0.864 (95% CI, 0.861–0.867; Figure 3B). This represented a MAE of 17.52 ms (95% CI, 17.38–17.64) for the predicted QT interval in the test split of the Stanford Model Development Cohort.

In the Cedars-Sinai External Validation Cohort, the MAE for QT prediction was 21.14 ms (95% CI, 20.99–21.29). The prolonged QTc classification revealed an AUROC of 0.943 (95% CI, 0.941–0.945) and an AUPRC of 0.781 (95% CI, 0.774–0.789). See Table 2 for full test characteristics.

A sensitivity analysis using patient-level clustered bootstraps using bias-corrected and accelerated intervals is presented in Table S4 and showed minimal change from standard bootstrapping.

Performance in Continuous QT Monitoring Cohort: In patients from the Continuous QT Monitoring Cohort (ECGRDVQ PhysioNet Dataset), 3DRECON-QT demonstrated a statistically significant correlation between predicted QT intervals and ground truth values (Pearson r=0.851; P<0.001; Figure 4). For serial QT measurements over the duration of recording, the MAE was 17.8 ms for QT interval. For detecting prolonged QTc using the absolute risk threshold (see Methods, “QT Annotation”), the model achieved an AUROC of 0.936 (95% CI, 0.900–0.967) and AUPRC of 0.395. For the secondary task of detecting a ≥15% increase in QTc from baseline, the model achieved an AUROC of 0.816 (95% CI, 0.790–0.843) and AUPRC of 0.538.

Validation of QTc and 12-Lead Morphology From Medtronic LINQ ICM Recordings: In the Stanford ICM Recording Cohort, 26 patients met criteria for time Δ (±7 days) between the ICM recording and the clinical reference 12-lead ECG. Figure 5A shows beat-wise morphological comparison between an actual ICM recording and simulated ICM from the 12-lead ECG. 3DRECON-QT predicted QTc values from the ICM recording with a Pearson correlation of r=0.824 (95% CI, 0.67–0.92) and mean absolute error of 17.5 ms (95% CI, 12.4–23.2) with the gold standard QTc (Figure 5B). A comparison for the same patient for the ICM and 12-lead reconstructions is shown in Figure 5C. Using derived ICM signals from the reference 12-lead ECG for the same cohort had a Pearson r=0.937 (95% CI, 0.885–0.970) and mean absolute error of 18.6 ms (95% CI, 14.1–23.6) with the gold standard QTc measurement. The difference in mean absolute error between the 2 methods was –1.1 ms (95% CI, –5.5 to 3.5).

Clinical Impact in a High-Risk Outpatient Cohort

Baseline characteristics for patients in the Stanford Real-World Class III Antiarrhythmic Cohort are provided in Table S5. In this cohort, 1 in 6 patients (16.5%, 277 of 1676) had at least one documented episode of high-risk QTc prolongation after their initial inpatient loading on presentation to medical care for any indication. These events occurred during

outpatient visits for a wide range of primary diagnoses, not just arrhythmia management follow-up. Figure 6A shows that approximately one-third of admissions with prolonged QT were for noncardiac causes, whereas one-third of cases occurred during a primary visit for arrhythmia evaluation, and one-third for non-arrhythmic cardiac evaluation.

The overall incidence of ventricular arrhythmic event–related hospitalization (torsades de points, ventricular fibrillation, and sudden cardiac death) in the cohort was 1.37%. This was increased for patients with prolonged QTc compared with those without (3.97% versus 0.86%; univariate OR, 4.78; $P<0.05$). After adjusting for age, heart failure, atrial fibrillation, drug class (sotalol/dofetilide), and laboratory values including creatinine, potassium, and magnesium, the risk of event remained 4-fold (OR, 4.24; 95% CI, 1.81–9.90; $P<0.05$).

For patients with an active outpatient dofetilide or sotalol prescription, 3DRECON-QT accurately identified prolonged QTc based on the derived ICM signal with an AUROC of 0.94 and an F1 score of 0.60 (Figure 6B).

DISCUSSION

3DRECON-QT accurately predicts QT/QTc within clinical tolerances (mean error, 18–21 ms) with AUROC 0.94 for classifying prolonged QT from a single-lead, nonstandard ECG signal. The system was robust to patient populations across 2 health systems with varying demographics. Further, when applied to a continuous monitoring scenario during pharmacological therapy, the model predicted QTc changes subsequently validated by observed serial measurements. Moreover, the model revealed QT prolongation in patients evaluated for noncardiac conditions, which could easily be missed in routine practice.

In our real-world analysis, 1 in 6 patients receiving class III antiarrhythmic medications presented from the ambulatory setting with dangerous QTc prolongation for encounters related to both cardiac and noncardiac causes. These episodes confer a >4-fold increase in serious ventricular arrhythmia events despite previous successful in-hospital drug loading. Overall event rates of ventricular arrhythmia in patients with and without prolonged QTc were 0.86% and 3.97%, respectively, which is consistent with rates reported in randomized trials[21,22] and systematic review.[4] The causal relationship between QTc prolongation and arrhythmia events was not adjudicated in this study. Out-of-hospital arrests that did not reach our center would be underreported by this analysis. This analysis underscores the unmet clinical need of safe ambulatory monitoring of these medications.

Previous evidence that successful loading at a given dose of antiarrhythmic dose does not guarantee safety of that dose at future encounters.[23] Current guidelines require a 12-lead ECG for QT assessment at least every 3 to 6 months for patients on class III antiarrhythmic agents.[5,24] However, a recent statement from the Heart Rhythm Society highlights that although single-lead wearable devices can provide serial QT interval measurements, their accuracy and clinical utility remain inferior to standard 12-lead ECGs.[25] Our data show how 3DRECON-QT may address this gap by providing near real-time QT safety monitoring from existing device infrastructure without additional leads, patient visits, or patient compliance burden.[26] Cloud-based artificial intelligence pipelines are already in

use (such as AccuRhythm AI, Medtronic, Minneapolis, MN) to filter out false-positive events in rhythm detection upstream of the provider portal, which provides a vantage point for immediate implementation. Last, this paradigm could possibly shorten or obviate mandatory hospital initiation of these medications for low-risk patients through hybrid inpatient-outpatient safety protocols and by providing objective triggers for dose reduction, electrolyte assessment, or drug discontinuation.[27]

The multitask encoder in 3DRECON-QT learns latent representations that preserve ventricular repolarization vectors, explaining high-fidelity QT inference despite limited spatial input. Previous deep-learning approaches have estimated QT from lead I or smartwatch tracings but lacked external validation.28 Further, the more fixed morphology of the lead I recording is less generalizable to insertable or chest-worn recorders that have variable and nonstandard recording vectors. To achieve long-term, continuous monitoring, a more flexible model is required. In addition, the training data involve the largest QT assessment study to date with strict patient-level splits, which contributed to its robust performance in multi-institution external testing.

Limitations of this study include its retrospective design and the fact that QT labels are derived from clinician-reviewed ECGs during routine clinical care rather than those specifically ordered for QT assessment. Simultaneous ICM and 12-lead ECG recordings are not available because of the proprietary nature of the ICM recordings. Our method of training 3DRECON by deriving vectors approximating the ICM allows us to overcome this limitation. Physiological and positional factors such as circadian variation, posture, and rate dependency are known to influence QT measurements and may affect the relationship between the ICM and 12-lead ECG signals. The broad inclusion of patient states in the development and external testing sites supports the generalizability of the model. The high correlation of the QTc predictions from the true ICM recordings in our data set supports the use of the V3-V2 vector surrogate for training and the generalizability of the 3DRECON-QT network. Prospective simultaneous collection of 12-lead signals with digital ICM recordings would provide further clinical validation and the opportunity to improve model performance through fine tuning. Twelve-lead reconstructions are only validated for measurement of the QT interval. Last, the clinical impact for patients receiving other QT-prolonging agents such as antipsychotics, antiemetics, antiepileptics, and others is not assessed in this study and may have a different benefit threshold compared with class III antiarrhythmics to deploy this monitoring strategy.

Future work may involve deployment in patient populations with preexisting ICMs or chest-worn patches or consumer devices to detect safety concerns. In our high-risk cohort, precision exceeding 75% indicates a modest false-positive burden, with fewer than 2 notifications needed to identify one episode of prolonged QTc event. Notifications for prolongation events may be resolved by a 12-lead ECG and medication review. Conversely, the sensitivity of >80% from a single measurement may offer a substantial improvement in safety over current sporadic clinic-based monitoring. Although continuous monitoring introduces additional challenges such as motion artifacts, future deployment may incorporate noise-filtering strategies. Studying appropriate patient selection is critical to increase drug safety without creating undue health care use.

Conclusions

3DRECON-QT demonstrates that deep-learning reconstruction of latent spatial information from single-lead signals can meet guideline-level QT accuracy. This model may address a substantial hidden burden of high-risk prolongation after class III antiarrhythmic initiation and other QT-prolonging drugs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

A.J.R., R.A., and S.B. conceived of the study. S.B., J.W.H., R.A., and A.J.R. were involved in acquiring the Stanford data set in a format appropriate for this study and preprocessing. R.T. and D.O. performed data acquisition, preparation, and analysis at Cedars Sinai Medical Center. S.B., R.A., A.J.R., S.M.N., M.P., A.P., P.G., X.L., K.B., and T.C. performed data analysis and interpretation. Clinical expertise was provided by P.J.W., E.A., A.P., M.P., and S.M.N. Article editing was provided by S.M.N., T.C., and M.P. All authors reviewed the article and approved the final version.

Sources of Funding

A.J.R. is supported by the National Institutes of Health (NIH; No. K23 HL166977). J.W.H. is supported by the National Science Foundation (No. DGE-1656518). M.V.P. is supported by funding from NIH and Apple Inc. S.M.N. reports grant funding from NIH (Nos. R01 HL83359, R01 HL149134, R01 HL1662260, and T32 HL166155). D.O. reports research support from NIH grant Nos. R00HL157421, R01HL173526, and R01HL173487, and Alexion AstraZeneca Rare Disease.

Nonstandard Abbreviations and Acronyms

AT/AF	atrial tachycardia or atrial fibrillation
AUPRC	area under the precision-recall curve
AUROC	area under the receiver operating characteristics curve
ICM	insertable cardiac monitor
MAE	mean absolute error

REFERENCES

1. Drew BJ, Ackerman MJ, Funk M, Gibler WB, Kligfield P, Menon V, Philippides GJ, Roden DM, Zareba W; American Heart Association Acute Cardiac Care Committee of the Council on Clinical Cardiology, the Council on Cardiovascular Nursing, and the American College of Cardiology Foundation. Prevention of torsade de pointes in hospital settings: a scientific statement from the American Heart Association and the American College of Cardiology Foundation. Circulation. 2010;121:1047–1060. doi: 10.1161/CIRCULATIONAHA.109.192704 [PubMed: 20142454]
2. Sandau KE, Funk M, Auerbach A, Barsness GW, Blum K, Cvach M, Lampert R, May JL, McDaniel GM, Perez MV, et al. ; American Heart Association Council on Cardiovascular and Stroke Nursing; Council on Clinical Cardiology; and Council on Cardiovascular Disease in the Young. Update to practice standards for electrocardiographic monitoring in hospital settings: a scientific statement from the American Heart Association. Circulation. 2017;136:e273–e344. doi: 10.1161/CIR.0000000000000527 [PubMed: 28974521]
3. Steinberg JS, Varma N, Cygankiewicz I, Aziz P, Balsam P, Baranchuk A, Cantillon DJ, Dilaveris P, Dubner SJ, El-Sherif N, et al. 2017 ISHNEHRS expert consensus statement on ambulatory

ECG and external cardiac monitoring/telemetry. Heart Rhythm. 2017;14:e55–e96. doi: 10.1016/j.hrthm.2017.03.038 [PubMed: 28495301]

4. Tisdale JE, Chung MK, Campbell KB, Hammadah M, Joglar JA, Leclerc J, Rajagopalan B; American Heart Association Clinical Pharmacology Committee of the Council on Clinical Cardiology and Council on Cardiovascular and Stroke Nursing. Drug-induced arrhythmias: a scientific statement from the American Heart Association. Circulation. 2020;142:e214–e233. doi: 10.1161/CIR.0000000000000905 [PubMed: 32929996]
5. Joglar JA, Chung MK, Armbruster AL, Benjamin EJ, Chyou JY, Cronin EM, Deswal A, Eckhardt LL, Goldberger ZD, Gopinathannair R, et al. ; Peer Review Committee Members. 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation. 2024;149:e1–e156. doi: 10.1161/CIR.0000000000001193 [PubMed: 38033089]
6. Rogers AJ, Selvalingam A, Alhusseini MI, Krummen DE, Corrado C, Abuzaid F, Baykaner T, Meyer C, Clopton P, Giles W, et al. . Machine learned cellular phenotypes in cardiomyopathy predict sudden death. Circ Res. 2021;128:172–184. doi: 10.1161/CIRCRESAHA.120.317345 [PubMed: 33167779]
7. Rogers AJ, Bhatia NK, Bandyopadhyay S, Tooley J, Ansari R, Thakkar V, Xu J, Soto JT, Tung JS, Alhusseini MI, et al. Identification of cardiac wall motion abnormalities in diverse populations by deep learning of the electrocardiogram. NPJ Digit Med. 2025;8:21. doi: 10.1038/s41746-024-01407-y [PubMed: 39799179]
8. Johannesen L, Vicente J, Mason JW, Sanabria C, Waite-Labott K, Hong M, Guo P, Lin J, Sørensen JS, Galeotti L, et al. Differentiating drug-induced multichannel block on the electrocardiogram: randomized study of dofetilide, quinidine, ranolazine, and verapamil. Clin Pharmacol Ther. 2014;96:549–558. doi: 10.1038/clpt.2014.155 [PubMed: 25054430]
9. Bazett HC. An analysis of the time-relations of electrocardiograms. Ann Noninvasive Electrocardiol. 1997;2:177–194. doi: 10.1111/j.1542-474x.1997.tb00325.x
10. Tikosyn (dofetilide) [package insert]. Pfizer Inc.; 2014. Accessed June 5, 2025. https://www.accessdata.fda.gov/drugsatfda_docs/label/2013/020931s007lbl.pdf
11. Pedersen OD, Bagger H, Keller N, Marchant B, Køber L, Torp-Pedersen C. Efficacy of dofetilide in the treatment of atrial fibrillation-flutter in patients with reduced left ventricular function: a Danish investigation of arrhythmia and mortality on dofetilide (diamond) substudy. Circulation. 2001;104:292–296. doi: 10.1161/01.cir.104.3.292 [PubMed: 11457747]
12. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE; 2018:7132–7140.
13. Xie S, Girshick R, Dollar P, Tu Z, He K. Aggregated residual transformations for deep neural networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2017:1492–1500.
14. Chen J, Zheng X, Yu H, Chen DZ, Wu J. Electrocardio panorama: synthesizing new ECG views with self-supervision. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence.- IJCAI Organization; 2021:3597–3605.
15. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need [published online June 12, 2017]. arXiv [csCL]. 2017; doi: 10.48550/arXi.1706.03762
16. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, et al. PyTorch: an imperative style, high-performance deep learning library [published online December 3, 2019]. arXiv [csLG]. 2019; doi: 10.48550/arXi.1912.01703
17. Alam R, Aguirre AD, Stultz CM. QTNet: deep learning for estimating QT intervals using a single lead ECG. Annu Int Conf IEEE Eng Med Biol Soc. 2023;2023:1–4. doi: 10.1109/EMBC40787.2023.10341204
18. Agency for Healthcare Research and Quality, Healthcare Cost and Utilization Project (HCUP). Clinical Classifications Software Refined (CCSR) for ICD-10-CM Diagnoses. Accessed June 5, 2025. https://www.hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp

19. Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SHA. NeuroKit2: a Python toolbox for neurophysiological signal processing. Behav Res Methods. 2021;53:1689–1696. doi: 10.3758/s13428-020-01516-y [PubMed: 33528817]

20. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. ; SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17:261–272. doi: 10.1038/s41592-019-0686-2 [PubMed: 32015543]

21. Køber L, Bloch Thomsen PE, Møller M, Torp-Pedersen C, Carlsen J, Sandøe E, Egstrup K, Agner E, Videbaek J, Marchant B, et al. ; Danish Investigations of Arrhythmia and Mortality on Dofetilide (DIAMOND) Study Group. Effect of dofetilide in patients with recent myocardial infarction and left-ventricular dysfunction: a randomised trial. Lancet. 2000;356:2052–2058. doi: 10.1016/s0140-6736(00)03402-4 [PubMed: 11145491]

22. Torp-Pedersen C, Møller M, Bloch-Thomsen PE, Køber L, Sandøe E, Egstrup K, Agner E, Carlsen J, Videbaek J, Marchant B, et al. Dofetilide in patients with congestive heart failure and left ventricular dysfunction. Danish Investigations of Arrhythmia and Mortality on Dofetilide Study Group. N Engl J Med. 1999;341:857–865. doi: 10.1056/NEJM199909163411201 [PubMed: 10486417]

23. Cho JH, Youn SJ, Moore JC, Kyriakakis R, Vekstein C, Militello M, Poe SM, Wolski K, Tchou PJ, Varma N, et al. Safety of oral dofetilide reloading for treatment of atrial arrhythmias. Circ Arrhythm Electrophysiol. 2017;10:1–9. doi: 10.1161/circep.117.005333

24. Rautaharju PM, Surawicz B, Gettes LS, Bailey JJ, Childers R, Deal BJ, Gorgels A, Hancock EW, Josephson M, Kligfield P, et al. ; American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part IV: the ST segment, T and U waves, and the QT interval: a scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society. Endorsed by the International Society for Computerized Electrocardiology. J Am Coll Cardiol. 2009;53:982–991. doi: 10.1016/j.jacc.2008.12.014 [PubMed: 19281931]

25. Wan EY, Ghanbari H, Akoum N, Itzhak Attia Z, Asirvatham SJ, Chung EH, Dagher L, Al-Khatib SM, Stuart Mendenhall G, McManus DD, et al. HRS white paper on clinical utilization of digital health technology. Cardiovasc Digit Health J. 2021;2:196–211. doi: 10.1016/j.cvdhj.2021.07.001 [PubMed: 35265910]

26. Giudicessi JR, Schram M, Bos JM, Galloway CD, Shreibati JB, Johnson PW, Carter RE, Disrud LW, Kleiman R, Attia ZI, et al. Artificial intelligence-enabled assessment of the heart rate corrected QT interval using a mobile electrocardiogram device. Circulation. 2021;143:1274–1286. doi: 10.1161/CIRCULATIONAHA.120.050231 [PubMed: 33517677]

27. Tisdale J, Wroblewski HA, Kingery JR, Trujillo TN, Overholser BR, Kovacs RJ. A risk score to predict QT interval prolongation in hospitalized patients. J Am Coll Cardiol. 2011;57:E112. doi: 10.1016/s0735-1097(11)60112-5

28. Alam R, Aguirre A, Stultz CM. Detecting QT prolongation from a single-lead ECG with deep learning. PLOS Digit Health. 2024;3:e0000539. doi: 10.1371/journal.pdig.0000539 [PubMed: 38917157]

Clinical Perspective

What Is New?

- A spatially aware deep learning model (3DRECON-QT) can reconstruct 12-lead electrocardiographic information from a single-lead, insertable cardiac monitor signal and accurately predict QT/QTc intervals.
- External validation across health systems and hardware platforms showed robust generalization to identify prolonged QT.
- In a real-world cohort of patients on class III antiarrhythmic drugs, 3DRECON-QT can identify QT prolongation that is frequent and associated with a 4-fold higher risk of ventricular arrhythmias.

What Are the Clinical Implications?

- Continuous ambulatory QT monitoring may be achieved using existing cardiac monitor infrastructure without new hardware or significant patient effort, enabling enhanced drug safety surveillance.
- Ambulatory surveillance may detect high-risk QT prolongation events after discharge from class III antiarrhythmic initiation to guide early intervention and dose adjustment.
- This approach may provide the groundwork to shorten or obviate inpatient drug initiation for selected patients and support broader remote safety monitoring across other QT-prolonging medications.

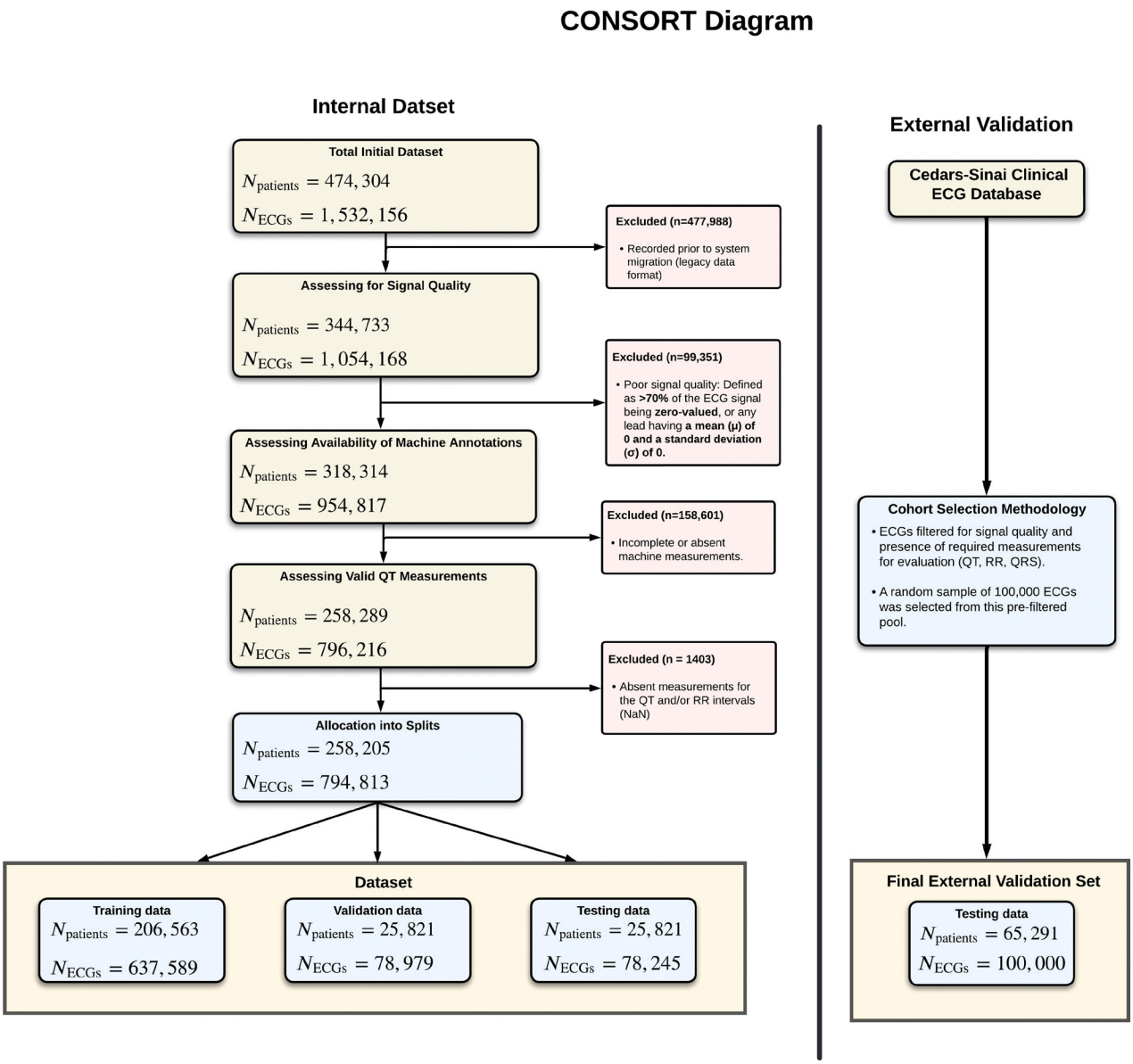

Figure 1. Consort diagram showing data flow for 12-lead ECGs in the development (Stanford) and external testing cohorts (Cedars-Sinai).

QT indicates QT interval; and RR, R–R interval.

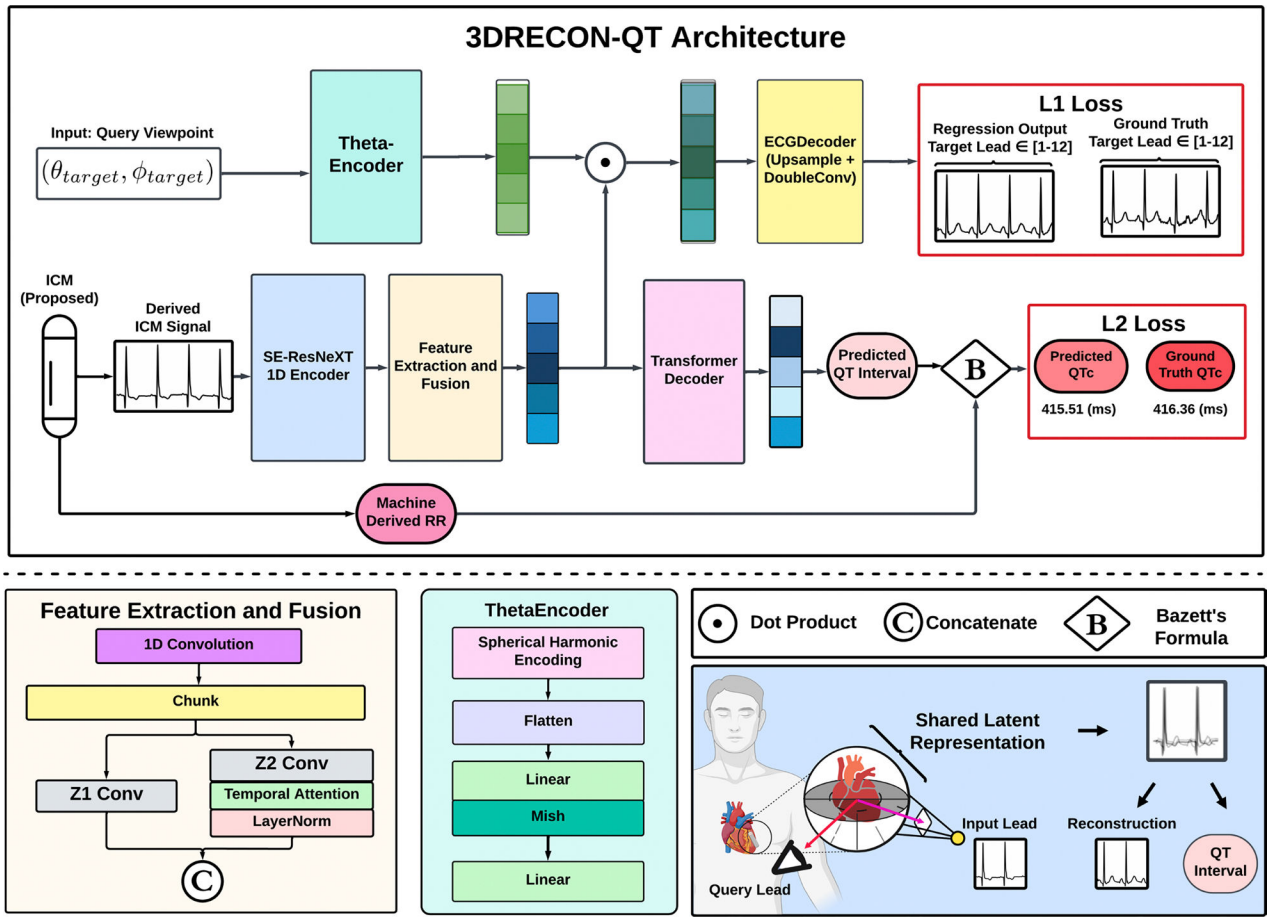

Figure 2. Overview of the 3DRECON-QT architecture for QT interval prediction from a single ICM recording.

The model encodes spatial representation of the query lead (ThetaEncoder) and fuses it with the encoded ICM signal (SE-ResNext) and uses a transformer decoder for QT interval prediction (L2 loss) and an ECGDecoder for reconstruction (L1 loss). 1D indicates 1-dimensional; B, Bazett's formula; Conv, convolution; ICM, insertable cardiac monitor; QTc, corrected QT interval; RR, R–R interval; SE-ResNeXt, Squeeze-and-Excitation ResNeXt network; ThetaEncoder, spherical angular encoder; and Z1 Conv/Z2 Conv, first and second convolutional blocks in the Z-stack feature extractor.

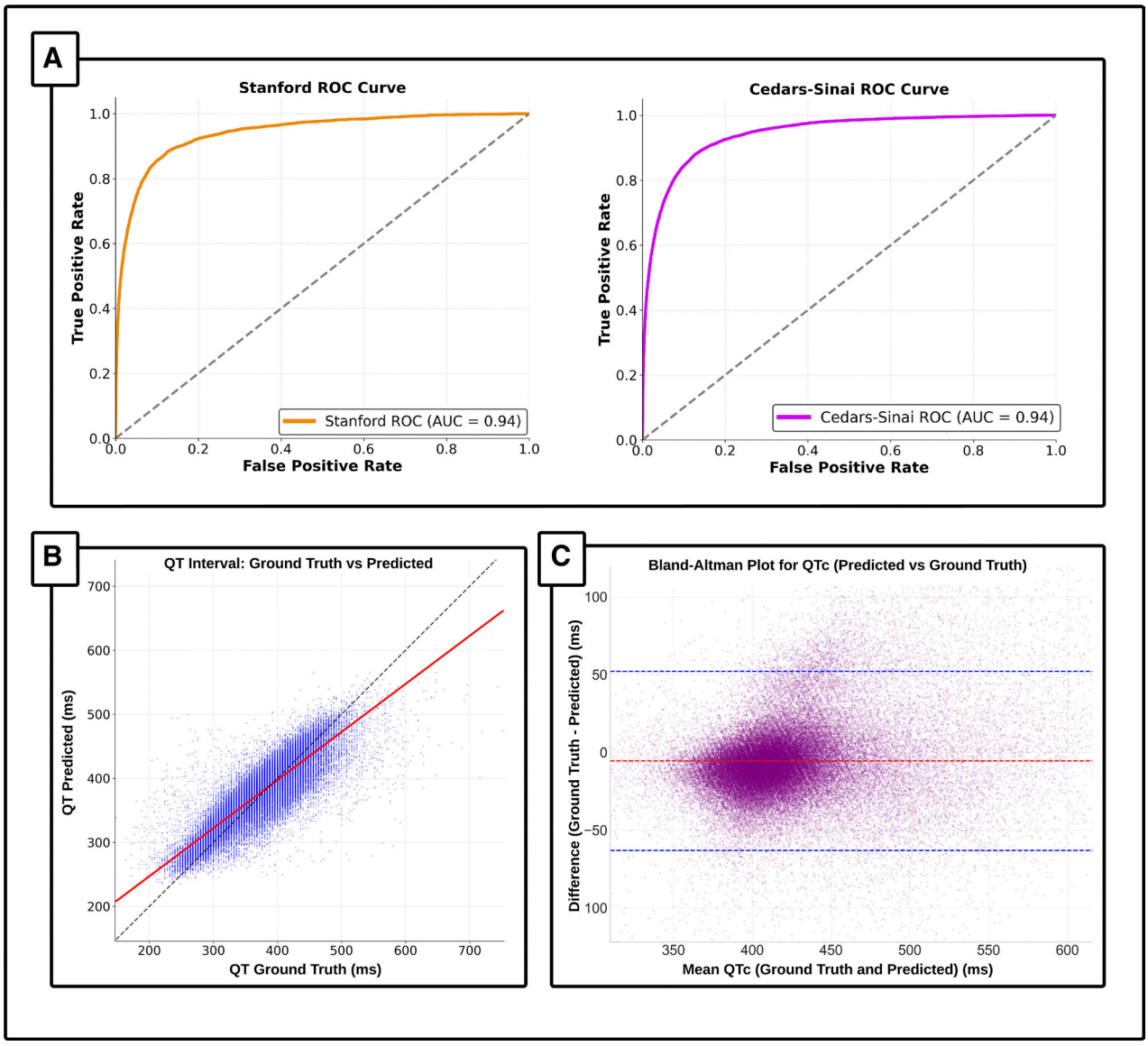

Figure 3. Model performance for QTc prediction and classification of prolonged QT.

A, Receiver operating characteristics curves for classification of prolonged QTc from derived ICM recordings in the development (Stanford) and external testing (Cedars-Sinai) cohorts. **B**, Scatter plot comparing ground truth QT (clinical 12-lead ECG) and model-predicted QT (from derived ICM signal) in the development cohort. **C**, Bland-Altman analysis of QTc residuals (all samples from the test set) with bias (red dashed line) and limits of agreement (1.96 * SD, blue dashed lines). AUC indicates area under the curve; QTc, corrected QT interval; and ROC, receiver operating characteristics.

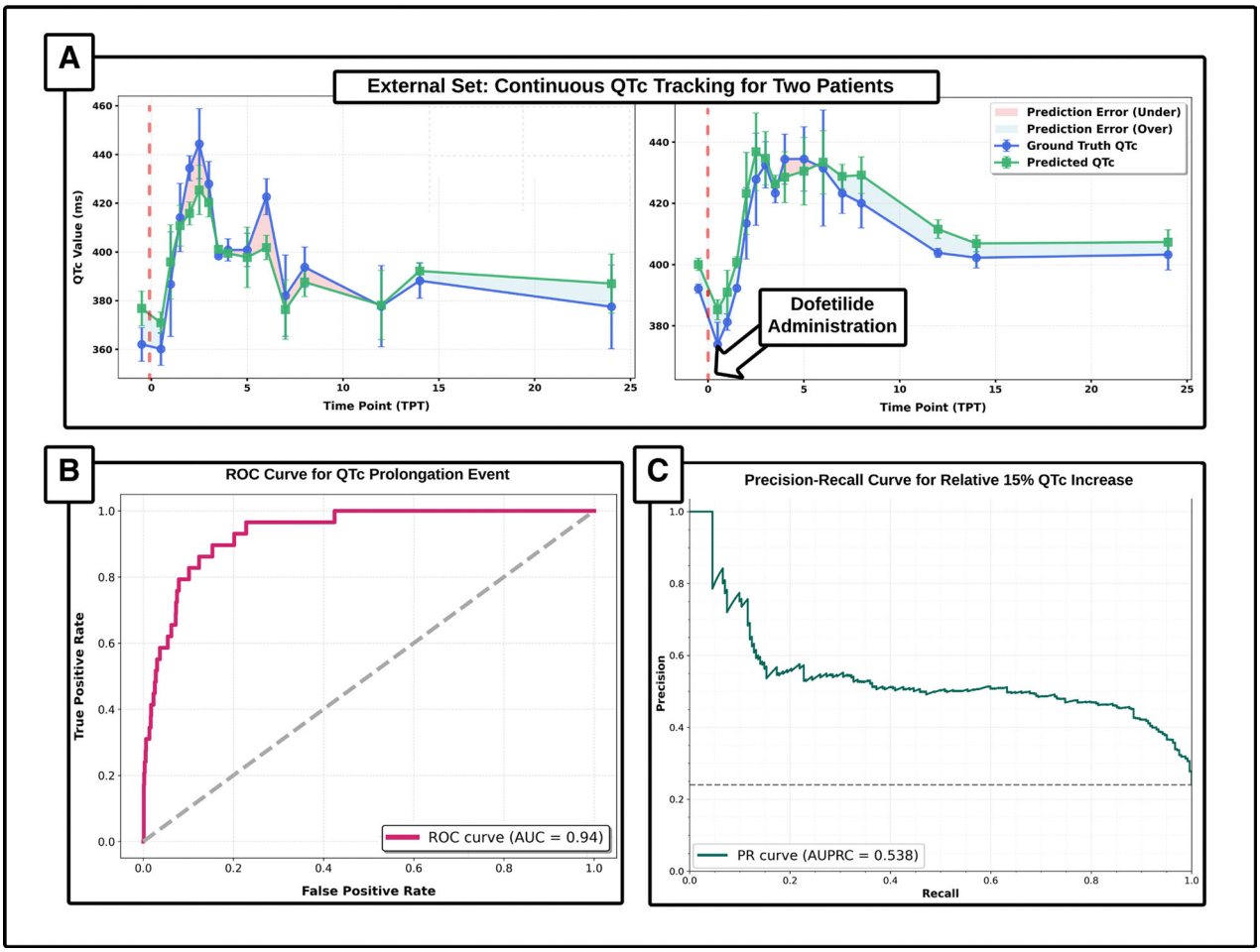

Figure 4. Continuous QTc tracking and performance evaluation.

A, Continuous monitoring of QTc over time in patients receiving a single dose of dofetilide, including response to drug administration, with comparison of predicted and ground truth values (continuous 12-lead ECG). **B**, Receiver operating characteristics curve evaluating model performance for classification of QTc prolongation events. **C**, Precision-recall curve assessing detection of relative QTc increases from baseline. AUC indicates area under the curve; QTc, corrected QT interval; and ROC, receiver operating characteristics.

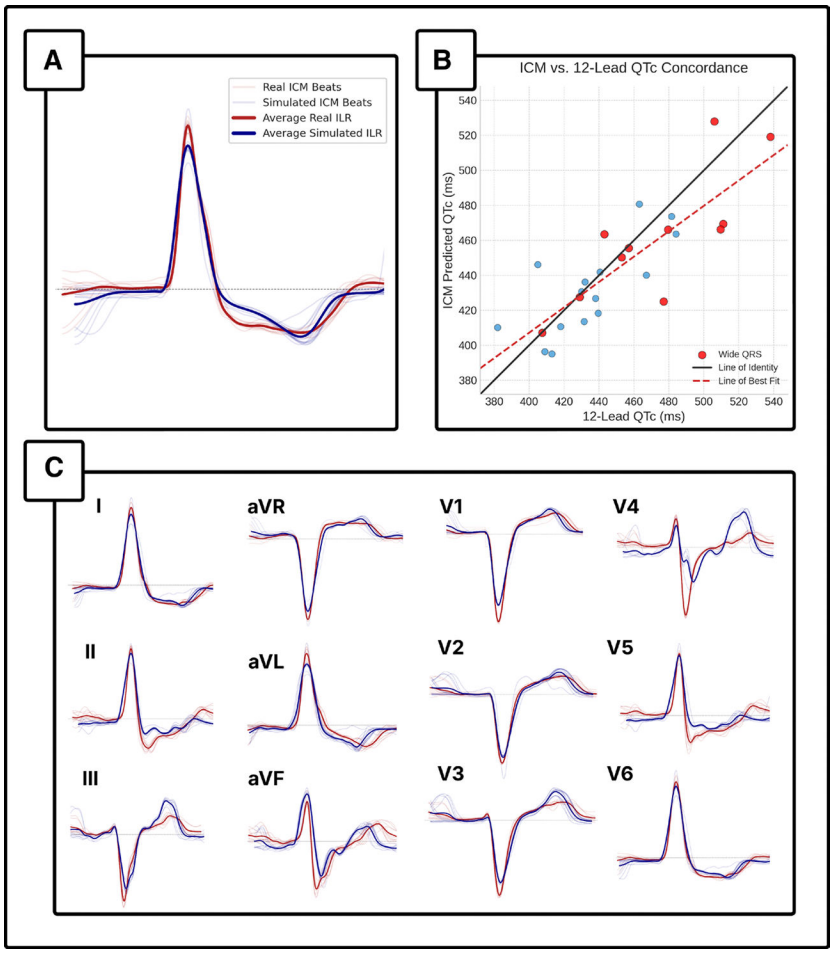

Figure 5. Validation of QTc and 12-lead morphology from Medtronic LINQ ICM recordings.

A, Beat-wise morphological comparison between real ICM recordings (red) and simulated ICM signals derived from the 12-lead ECG (blue) from a patient in ICM cohort. Individual beats are beat-detected and aligned to account for differences in heart rates. Average beats are styled with solid lines. **B**, Concordance between QTc values predicted from ICM recordings and gold-standard 12-lead QTc, with regression line (dashed red) and line of equality (solid black). Patients with wide QRS are labeled in red circles and those with narrow QRS in blue circles. **C**, Lead-wise comparison of 12 standard ECG leads reconstructed from the real ICM (red) and the derived ICM (blue) with average beats styled in solid lines. ICM indicates insertable cardiac monitor; and QTc, corrected QT interval.

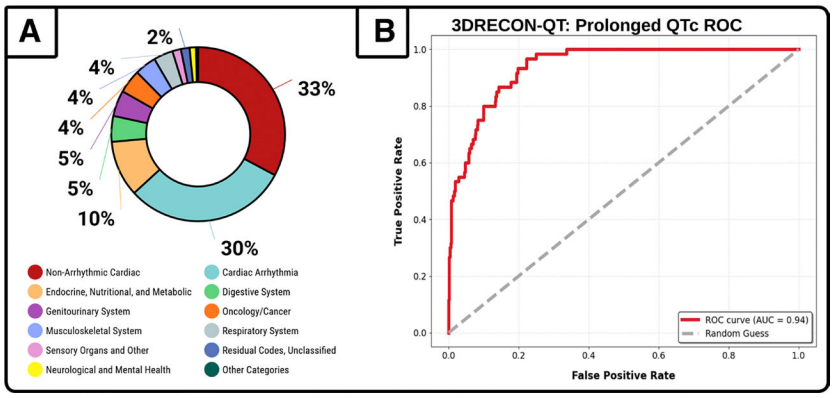

Figure 6. Clinical impact of prolonged QTc in a high-risk outpatient cohort.

A, Distribution of primary visit diagnoses among utpatients with documented QTc prolongation while on class III antiarrhythmic therapy (dofetilide or sotalol), demonstrating that prolonged QTc events occurred across a wide range of cardiac and noncardiac clinical encounters. **B**, Receiver operating characteristic curve showing model performance of 3DRECON-QT for detection of prolonged QTc in this high-risk cohort. AUC indicates area under the curve; QTc, corrected QT interval; and ROC, receiver operating characteristics.

Table 1.

Patient Characteristics of the Development Cohort and External Testing Cohort

Variable	Train (n=206 563)	Validation (n=25 821)	Test (n=25 821)	External (n=65 291)
Age	58.3±18.0	58.4±17.9	58.4±18.0	63.6±18.2
Male	101 151 (49.0%)	12 649 (49.0%)	12 533 (48.5%)	34 794 (53.3%)
Race				
Asian	28 181 (13.6%)	3600 (13.9%)	3510 (13.6%)	4543 (7.0%)
Black	9454 (4.6%)	1170 (4.5%)	1224 (4.7%)	9843 (15.1%)
Native American	608 (0.3%)	64 (0.2%)	76 (0.3%)	186 (0.3%)
Pacific Islander	2508 (1.2%)	286 (1.1%)	296 (1.1%)	206 (0.3%)
White	114 455 (55.4%)	14 329 (55.5%)	14 290 (55.3%)	44 450 (68.1%)
Other*	35 937 (17.4%)	4462 (17.3%)	4529 (17.5%)	4413 (6.8%)
Unknown	15 420 (7.5%)	1910 (7.4%)	1896 (7.3%)	1650 (2.5%)
Ethnicity				
Hispanic/Latino	26 720 (12.9%)	3249 (12.6%)	3351 (13.0%)	9203 (14.1%)
Non-Hispanic	164 032 (79.4%)	20 631 (79.9%)	20 536 (79.5%)	54 339 (83.2%)
Unknown	15 811 (7.7%)	1941 (7.5%)	1934 (7.5%)	1749 (2.7%)
AT/AF	30 851 (14.9%)	3862 (15.0%)	3792 (14.7%)	1813 (2.8%)
CAD	31 257 (15.1%)	3963 (15.3%)	3824 (14.8%)	16 138 (24.7%)
HTN	86 411 (41.8%)	10 943 (42.4%)	10 958 (42.4%)	30 391 (46.5%)
HF	23 354 (11.3%)	2918 (11.3%)	2862 (11.1%)	13 620 (20.9%)

AT/AF indicates atrial tachycardia or atrial fibrillation; CAD, coronary artery disease; HF, heart failure; and HTN, hypertension.

*Other includes patient-reported race entries that did not fit predefined EHR categories, including multiracial identities and free-text responses that could not be mapped to standard classifications.

Table 2.

Test Characteristics in the Holdout Internal Test Set and the External Test Set

Metric	Stanford	Stanford 95% CI	Cedars	Cedars 95% CI
Regression				
MAE	17.521	17.381–17.641	21.135	20.985–21.288
R^2	0.747	0.742–0.752	0.738	0.732–0.744
Pearson *r*	0.864	0.861–0.867	0.859	0.856–0.863
Classification				
Sensitivity	0.597	0.591–0.605	0.542	0.534–0.550
Specificity	0.977	0.976–0.978	0.988	0.987–0.989
Precision	0.606	0.589–0.614	0.883	0.877–0.890
N PV	0.977	0.976–0.977	0.929	0.927–0.930
Accuracy	0.957	0.955–0.957	0.925	0.923–0.926
F1 score	0.601	0.593–0.610	0.672	0.665–0.679
AUC (bootstrap)				
A UROC	0.942	0.941–0.943	0.943	0.941–0.945
A UPRC	0.632	0.626–0.639	0.781	0.774–0.789

AUC indicates area under the curve; AUPRC, area under the precision-recall curve; AUROC, area under the receiver operating characteristics curve; MAE, mean absolute error; and NPV, negative predictive value.