

ORIGINAL RESEARCH

AI/ML

QTNet



Predicting Drug-Induced QT Prolongation With Artificial Intelligence-Enabled Electrocardiograms

Hao Zhang, MS,^{a,*} Constantine Tarabanis, MD,^{b,*} Neil Jethani, PhD,^{a,c} Mark Goldstein,^c Silas Smith, MD,^d Larry Chinitz, MD,^b Rajesh Ranganath, PhD,^{a,c} Yindalon Aphinyanaphongs, MD, PhD,^a Lior Jankelson, MD, PhD^b

ABSTRACT

BACKGROUND Prediction of drug-induced long QT syndrome (diLQTS) is of critical importance given its association with torsades de pointes. There is no reliable method for the outpatient prediction of diLQTS.

OBJECTIVES This study sought to evaluate the use of a convolutional neural network (CNN) applied to electrocardiograms (ECGs) to predict diLQTS in an outpatient population.

METHODS We identified all adult outpatients newly prescribed a QT-prolonging medication between January 1, 2003, and March 31, 2022, who had a 12-lead sinus ECG in the preceding 6 months. Using risk factor data and the ECG signal as inputs, the CNN QTNet was implemented in TensorFlow to predict diLQTS.

RESULTS Models were evaluated in a held-out test dataset of 44,386 patients (57% female) with a median age of 62 years. Compared with 3 other models relying on risk factors or ECG signal or baseline QTc alone, QTNet achieved the best ($P < 0.001$) performance with a mean area under the curve of 0.802 (95% CI: 0.786–0.818). In a survival analysis, QTNet also had the highest inverse probability of censorship-weighted area under the receiver-operating characteristic curve at day 2 (0.875; 95% CI: 0.848–0.904) and up to 6 months. In a subgroup analysis, QTNet performed best among males and patients ≤ 50 years or with baseline QTc < 450 ms. In an external validation cohort of solely suburban outpatient practices, QTNet similarly maintained the highest predictive performance.

CONCLUSIONS An ECG-based CNN can accurately predict diLQTS in the outpatient setting while maintaining its predictive performance over time. In the outpatient setting, our model could identify higher-risk individuals who would benefit from closer monitoring. (J Am Coll Cardiol EP 2024;10:956–966) © 2024 by the American College of Cardiology Foundation.

The QT interval measures the time from ventricular depolarization to completion of repolarization, corresponding to the interval between the beginning of the QRS to the end of the T-wave on the electrocardiogram (ECG).¹ The QTc interval is tightly conserved between 350 and 470 ms (or 450 ms in men), as QTc prolongation beyond 500 ms increases the risk of torsades de pointes (TdP),^{2,3} a form of polymorphic ventricular tachycardia. Congenital long QT syndrome (LQTS) is

From the ^aDepartment of Population Health, NYU Langone Health, New York University School of Medicine, New York, New York, USA; ^bLeon H. Charney Division of Cardiology, Cardiac Electrophysiology, NYU Langone Health, New York University School of Medicine, New York, New York, USA; ^cCourant Institute of Mathematical Sciences, New York University, New York, New York, USA; and the ^dRonald O. Perelman Department of Emergency Medicine, NYU Langone Health, New York, New York, USA. *Drs Zhang and Tarabanis contributed equally to this work.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the [Author Center](#).

Manuscript received August 24, 2023; revised manuscript received January 19, 2024, accepted January 31, 2024.

a heritable syndrome caused largely by mutations in 1 of 3 cardiac ion channels and their corresponding currents I_{Na} , I_{Kr} , and I_{Ks} .⁴ Drug-induced long QT syndrome (diLQTS) is an acquired form of LQTS that can be caused by various cardiac and noncardiac medications, most frequently by inhibition of the rapid component of the delayed rectifier potassium current (I_{Kr}).⁵

Given a 1% to 10% incidence of TdP among patients taking certain medications with a known QT-prolonging effect, predicting diLQTS is of important clinical significance.⁵ Yet, genetic and acquired factors including common comorbidities result in substantial heterogeneity in the susceptibility to diLQTS, and in turn, suboptimal tools for individual risk stratification.^{6,7} In the outpatient setting, in which continuous ECG and clinical monitoring is limited, the importance of predicting diLQTS is augmented. Reliance on a normal baseline QTc does not ensure low risk for diLQTS,^{7,8} as reduced repolarization reserve often remains subclinical, with overt diLQTS appearing only upon drug exposure.²

Given the recent advances in utilizing artificial intelligence (AI) to predict complex and occult clinical syndromes,⁹ we set to develop and test an ECG-based AI architecture for the prediction of diLQTS in the outpatient setting. The best-performing model was externally validated in a cohort of suburban outpatient practices.

METHODS

COHORT SELECTION. We identified all adult (≥ 18 years of age) patients newly prescribed a QT-prolonging medication in the outpatient setting who had a 12-lead (baseline) sinus ECG in the preceding 6 months at our institution between January 1, 2003, and March 31, 2022 (Figure 1A). For patients with multiple preceding ECGs, the one temporally closest to the time of prescription was selected as the baseline ECG (Figure 1B). A total of 38 cardiac and noncardiac QT-prolonging medications with either a known or possible risk for TdP, as classified by CredibleMeds,¹⁰ were selected (Supplemental Table 1). Approximately, 16% of the prescriptions were prescribed as needed (“PRN”). Dofetilide and sotalol were excluded because of the Food and Drug Administration-mandated inpatient loading.¹¹ This study was approved by the NYU Langone Health Institutional Review Board.

We included outpatient practices belonging to NYU Langone in New York City in our main cohort. Patients receiving prescriptions at NYU Long Island practices served as an external validation cohort and

were not included in the main cohort (Supplemental Table 2). The main cohort was divided into training, validation, and test datasets in a 3:1:1 ratio, and a single prescription corresponding to each patient was randomly selected from the test dataset to generate the held-out test dataset (Figure 1A, Supplemental Table 3). No patient was included in multiple datasets, and 1 prescription instance per patient was isolated at random to ensure that each patient was included only once in the held-out test dataset.

OUTCOMES. For each patient in the entire cohort, we obtained all follow-up ECGs occurring 6 months after the time of prescription, as the majority (70%) of patients were maintained on the newly prescribed QT-prolonging medication for < 6 months.

Severe QT prolongation was the primary outcome and was defined as a follow-up ECG with either QTc > 500 ms or a change in the QTc from baseline (Δ QTc) > 60 ms.¹ The QTc was calculated using the Bazett formula.¹² The task of predicting severe QT prolongation was treated as a survival analysis problem.^{13–15} For patients with a follow-up ECG exhibiting severe QT prolongation, the time of the first such ECG was used as the event time (Figure 1B). In the external validation cohort and held-out test dataset, we implemented a strict outcome criterion for the event time (ie, severe QT prolongation) in which the event had to occur while the drug was being prescribed. For prescription instances without any follow-up ECGs exhibiting QT prolongation, the time of the last ECG was used as the censorship time. Patients without any follow-up ECGs were considered censored at time zero (ie, at the time of prescription).

Two main reasons underlie the choice of predicting QT prolongation with a survival analysis approach. First, the resulting time-to-event prediction models simultaneously produce risk scores for developing severe QT prolongation for any given time (30, 60, 90 days, etc.) within the 6-month period, allowing for the evaluation of temporal patterns. Second, this approach addresses cases in which patients at high risk for QT prolongation could have developed it outside the defined 6-month period or in which patients were lost to follow-up. Using a survival analysis approach, both of these cases are not excluded but instead are treated as censored data points by constraining the model to place all of the (nonzero) risk for QT prolongation after the censoring time. This process can amount to high-risk prediction even for a patient with no observed QT prolongation, as long as

ABBREVIATIONS AND ACRONYMS

AI = artificial intelligence

AUROC = area under the receiver-operating characteristic curve

CNN = convolutional neural network

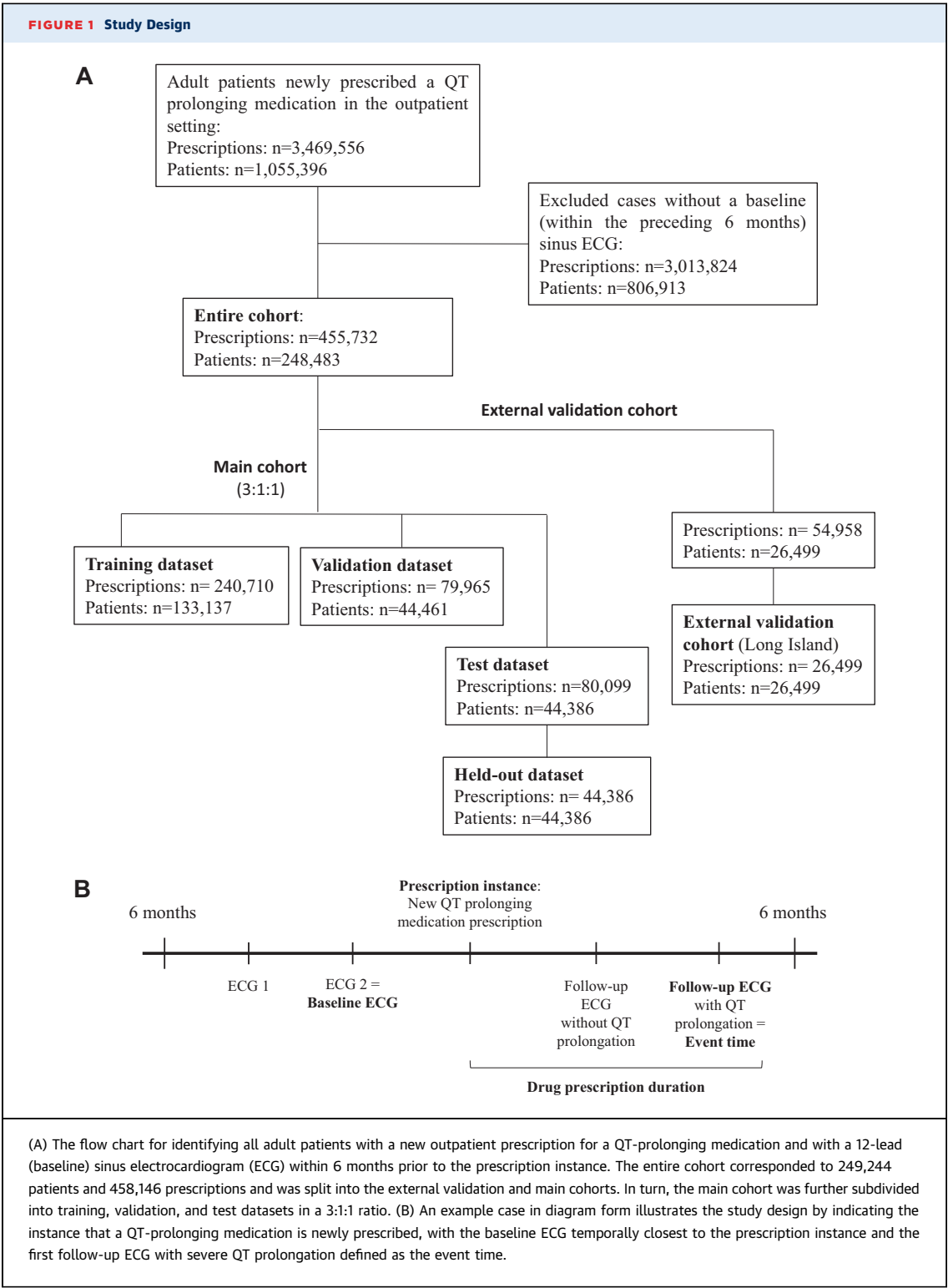
diLQTS = drug-induced long QT syndrome

Grad-CAM = gradient-weighted class activation mapping

ipcw-AUROC = inverse probability of censorship-weighted area under the receiver-operating characteristic curve

LQTS = long QT syndrome

TdP = torsades de pointes



they are similar enough to other patients with observed prolongation. Hence, adjusting for the probability of censorship allows for the estimation of predictive performance in the entire cohort.

MODEL INPUTS. The QTNet model takes in the baseline ECG time series (ie, signal) and all tabular data (except for Outcomes) listed in [Supplemental Table 3](#), which include demographic characteristics,

comorbidities, blood pressure, laboratory values, medications, and baseline QTc. The input tabular data (Supplemental Table 3) were selected based on known risk factors for QT prolongation identified in prior risk scores.¹⁶⁻¹⁹ Demographics, comorbidities, blood pressure, laboratory values, and medications were retrieved from NYU's electronic health record system (Epic Systems). Comorbidities were ascertained using codes from the International Classification of Diseases-10th Revision-Clinical Modification (Supplemental Table 4). All ECGs and their measurements were accessed through MUSE (GE Healthcare). Variables with missing values were imputed by multivariate imputation with chained equations using the Scikit-learn package.

MODEL ARCHITECTURE. For the QTNet model, we implemented a convolutional neural network (CNN) to learn a concise 1-dimensional representation of the ECG time series. This representation was fused with the tabular data (Supplemental Table 3) and then fed through a fully connected neural network with a softmax output layer to generate the probability of each binned time-interval: [0, 2], [2, 7], [7, 14], [14, 21], [21, 30], [30, 45], [45, 60], [60, 75], [75, 90], [90, 105], [105, 120], [120,135], [135, 150], [150,165], and [165, 180] days. The CNN architecture was based on the current state of the art for arrhythmia detection²⁰ and is a 34-layer ResNet CNN consisting of 16 residual connections (Supplemental Figure 1). The input to the network was a $8 \times 2,500$ matrix, representing the 8 measured leads (lead III and the augmented leads are arithmetically computed) by 10-second duration sampled at 250 Hz (500 Hz ECGs were downsampled to 250 Hz).

MODEL TRAINING AND HYPERPARAMETER TUNING. The aforementioned data inputs were used to train the model in predicting the event time distribution using an Adam optimizer for 100 epochs to minimize the negative log-likelihood survival loss for a discrete-time interval prediction. After each epoch, we evaluated the models on the validation dataset. We used a learning rate scheduler that multiplies the learning rate by 0.8 after 10 epochs of no validation loss improvement. Early stopping was triggered after the validation loss ceased to improve for 20 epochs. We repeated this process to train each model to predict the censorship time instead of using the negative log-likelihood censorship loss for a discrete-time interval prediction.

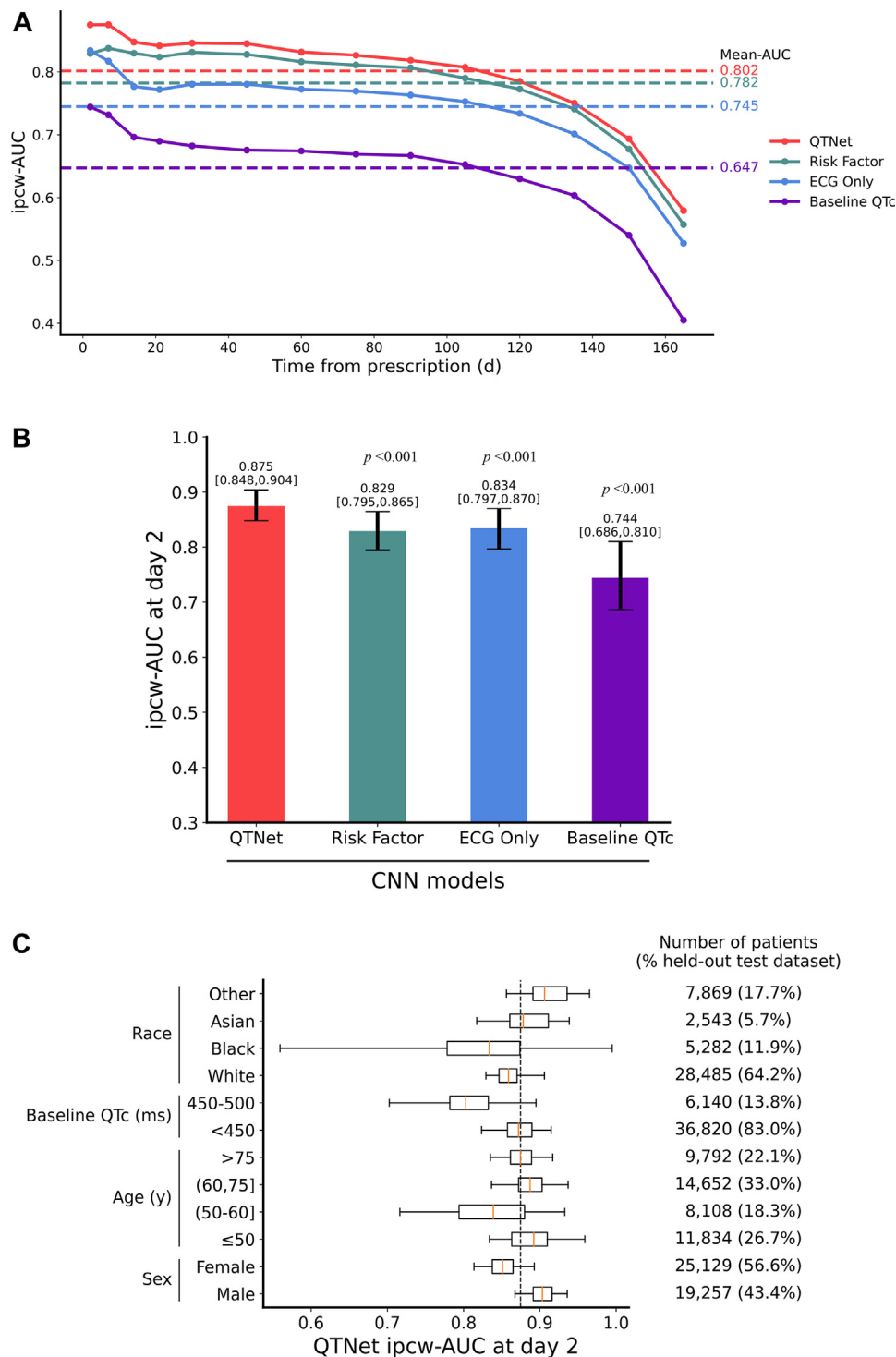
We used the validation dataset to select the best network architecture and hyperparameter configuration. We selected the model with the time-averaged inverse probability of censorship-weighted area

under the receiver-operating characteristic curve (ipcw-AUROC) across times (mean AUROC). We examined the effect of varying the number and dimensionality of the fully connected layers, considering 1, 2, or 3 layers of dimension 100 or 1,000. We considered the effect of temporal dimensionality reduction by modifying the stride lengths to reduce the input 2,500-dimensional vector to either a 10-dimensional vector or an 80-dimensional vector. We also tuned the batch size (32, 64, 128) and learning rate (10-5, 10-4, 10-3). We implemented only the fully connected portion of the neural network for the models involving exclusively tabular data and performed hyperparameter tuning accordingly.

MODEL EVALUATION. We used both the held-out test dataset derived from the main cohort and the external validation cohort to evaluate our models. The ability to predict the primary outcome was measured using the following metrics: the ipcw-AUROC at each time interval, as well as the mean AUROC. The ipcw-AUROC at the aforementioned time points was calculated using the censorship model trained on the full set of features (Supplemental Table 3). For comparison, we have trained 3 additional fully connected neural networks: 1) risk factor; 2) baseline ECG; and 3) baseline QTc. The risk factor model was trained on the same set of tabular inputs as QTNet except for the 2 ECG features, while the baseline ECG and baseline QTc models were trained using only the baseline ECG signal and measured scalar baseline QTc, respectively. The baseline ECG model demonstrates the predictive power with just the ECG signal, while the risk factor model serves as a reference point to show the additional predictive value provided by ECG signal in QTNet. For QTNet, we also performed subgroup analyses stratified by race, sex, age, and baseline QT prolongation status, comparing mean AUROC and ipcw-AUROC at day 2. For all evaluation metrics, we adopted the bootstrap method to calculate 95% CIs and permutation testing to calculate *P* values when comparing models. Separately, gradient-weighted class activation mapping (Grad-CAM) was used for the heatmap visualization,²¹ highlighting which portions of the ECG waveform the model focuses on when predicting QT prolongation.

STATISTICS. Values are presented as median (Q1-Q3) for continuous variables and as count (percentage) for categorical variables, unless otherwise specified. Analysis of variance was used for comparisons of continuous variables among >2 groups, while the chi-square test was used for comparisons of categorical variables. Statistical analysis was performed using

FIGURE 2 Model Performance Over Time



(A) The inverse probability of censorship-weighted area under the receiver-operating characteristic curve (ipcw-AUROC) plotted over time for each convolutional neural network (CNN) model applied to the held-out test dataset. QTNet outperforms the discriminative performance of each other model when estimating the risk of severe QT prolongation within a given time period. The dashed lines represent the mean area under the receiver-operating characteristic curve (AUROC) for each model. (B) The ipcw-AUROC at day 2 plotted as a bar chart with 95% CIs for each CNN model applied to the held-out test dataset, with QTNet outperforming all other models with statistically significant margins. (C) The ipcw-AUROC at day 2 for the QTNet model in subgroups of race, baseline QTc, age, and sex presented as box (Q1-Q3) and whisker (95% CI) plots. The dotted line represents the QTNet ipcw-AUROC at day 2 for the whole held-out test dataset. ECG = electrocardiogram.

the software Scikit-learn (version 1.02). Values of $P < 0.05$ were considered statistically significant.

RESULTS

The entire cohort consisted of 400,774 new QT-prolonging medication outpatient prescriptions with a 12-lead sinus ECG in the preceding 6 months, corresponding to 221,984 patients (Figure 1A). The held-out test dataset and external validation cohort were used to evaluate model performance. Demographic characteristics, comorbidities, baseline ECG features, laboratory values, blood pressure, and medications at the time of prescription were collected (Supplemental Table 2). The held-out test dataset consisted of 44,386 unique patients, who had a median age of 62 years (Q1-Q3: 49-74 years), were predominantly female (56.6%) and White (78.4%), and counted hypertension (57.0%), thyroid disorders (21.0%), and diabetes (23.2%) among the most common comorbidities. Patients had normal median potassium (4.3 mmol/L [Q1-Q3: 4.0-4.6 mmol/L]) and magnesium (2.0 mmol/L [Q1-Q3: 1.8-2.1 mmol/L]) levels at the time of prescription, with 9.3% of patients in the held-out test dataset on a loop diuretic. Approximately 6.7% of patients were already on at least 1 other medication with a known or possible risk of diLQTS when prescribed a new QT-prolonging medication. Both the median baseline QTc (433 ms [Q1-Q3: 416-452 ms]) and QRS (88 ms [Q1-Q3: 80-96 ms]) were normal. The baseline QTc was normal (QTc <450 ms for males, <470 ms for females)²² for 36,871 (82.9%) patients in the held-out test dataset. A total of 21.2% of prescription instances in the held-out test dataset had at least 1 follow-up ECG within 6 months following the prescription. Among them, severe QT prolongation was observed in 360 (0.8%) patients, of which 188 (52.2%) had a normal baseline QTc.

The time-dependent ipcw-AUROC is depicted in Figure 2A, showing QTNNet outperforming the discriminative performance of the other 3 models when estimating the risk of severe QT prolongation within any given time period <180 days. Focusing on QT prolongation prediction by day 2 (Figure 2B), QTNNet (ipcw-AUROC: 0.875; 95% CI: 0.848-0.904) outperformed each of the risk factor (ipcw-AUROC: 0.829; 95% CI: 0.795-0.865; $P < 0.001$), baseline ECG (ipcw-AUROC: 0.834; 95% CI: 0.797-0.870; $P < 0.001$), and baseline QTc (0.744; 95% CI: 0.686-0.810; $P < 0.001$) models. The QTNNet ipcw-AUROC at day 2 for subgroups of race, baseline QTc, age, and sex is presented as a box-and-whisker plot in Figure 2C. The best performance in each subgroup is noted among patients with a baseline QTc <450 ms (ipcw-AUROC:

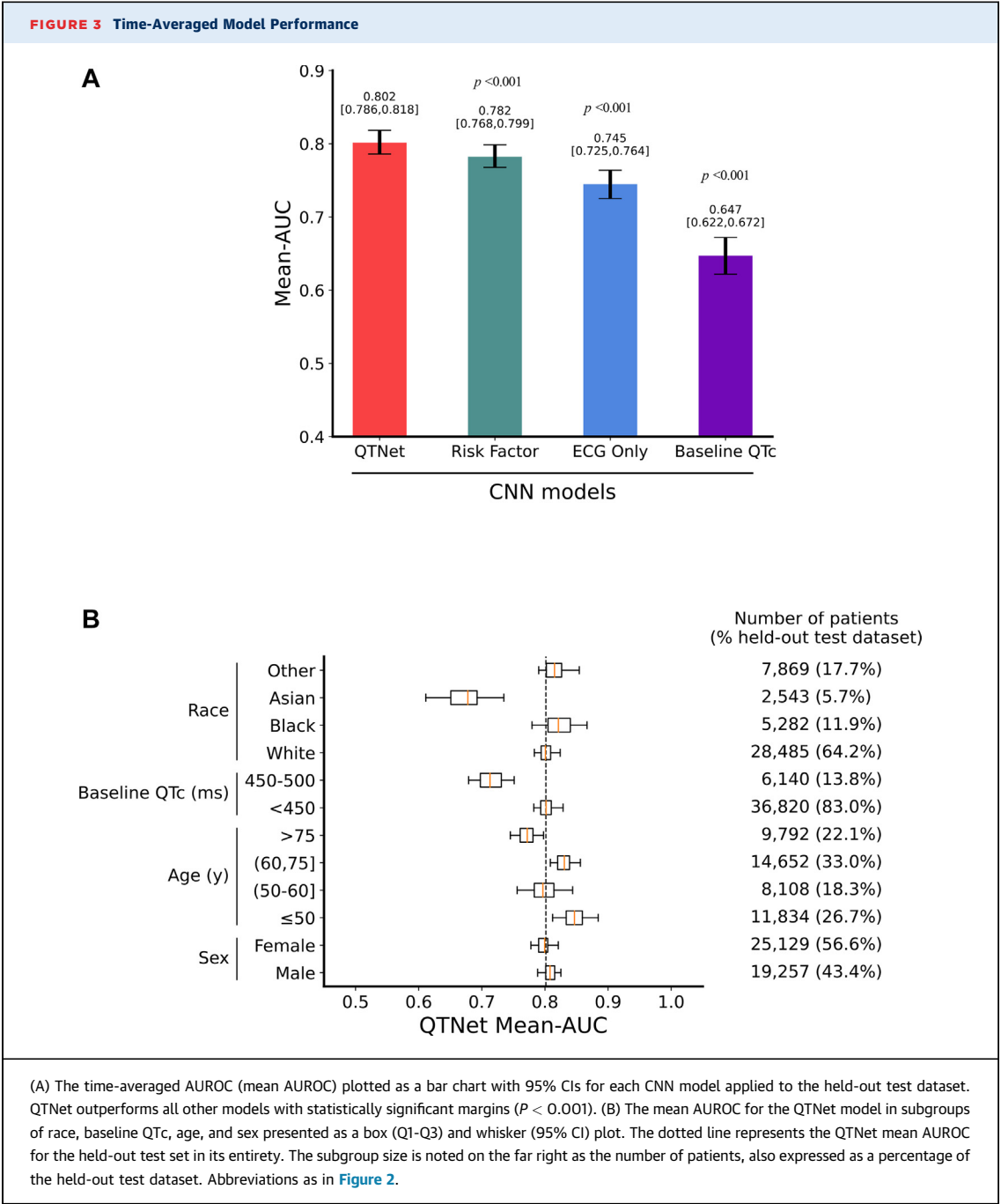
0.872; 95% CI: 0.821-0.916), age ≤ 50 years (ipcw-AUROC: 0.892; 95% CI: 0.829-0.959), other race (ipcw-AUROC: 0.906; 95% CI: 0.846-0.966), and males (ipcw-AUROC: 0.903; 95% CI: 0.866-0.936).

As depicted in Figures 3A and 3B, QTNNet achieved the best performance (mean AUROC: 0.802; 95% CI: 0.786-0.818), outperforming predictions based on the risk factor (mean AUROC: 0.782; 95% CI: 0.768-0.799), baseline ECG (mean AUROC: 0.745; 95% CI: 0.725-0.764), and baseline QTc (mean AUROC: 0.647; 95% CI: 0.622-0.672) models with $P < 0.001$ for all comparisons between models. The QTNNet mean AUROC for subgroups stratified by race, sex, age and baseline QTc can be found in Figure 3C. The best performance in each subgroup is noted among patients with a baseline QTc <450 ms (mean AUROC: 0.802; 95% CI: 0.781-0.829), age ≤ 50 years (mean AUROC: 0.847; 95% CI: 0.812-0.885), identifying as Black (mean AUROC: 0.821; 95% CI: 0.778-0.868), and males (mean AUROC: 0.808; 95% CI: 0.787-0.826). Separately, a Grad-CAM heatmap with all leads overlaid indicates that the model “focuses” most at the beginning of the T-wave and not the QRS complex (Figure 4).

The external validation cohort consisted of 26,499 unique patients, with a median age of 64 years (Q1-Q3: 52-75 years), and who were predominately female (59.6%) and White (87.7%) (Supplemental Table 2). The cohort was evaluated using models developed from the training dataset and performance metrics are presented in Figure 5. Similar to the previous results, QTNNet maintained high discriminative performance when estimating the risk of severe QT prolongation over time (Figure 5A). At day 2, QTNNet had an ipcw-AUROC (0.873; 95% CI: 0.767-0.942) equivalent to that of the baseline ECG (0.882; 95% CI: 0.800-0.952; $P = 0.674$), but still outperformed risk factor (0.828; 95% CI: 0.694-0.917; $P < 0.001$) and baseline QTc (0.813; 95% CI: 0.713-0.887; $P = 0.011$) models (Figure 5B). The QTNNet again achieved the best mean AUROC among the 4 models (QTNNet: 0.804; 95% CI: 0.770-0.836; risk factor: 0.786; 95% CI: 0.750-0.819; baseline ECG: 0.776; 95% CI: 0.743-0.807; baseline QTc: 0.699; 95% CI: 0.657-0.739) with significant P values for all comparisons between models.

DISCUSSION

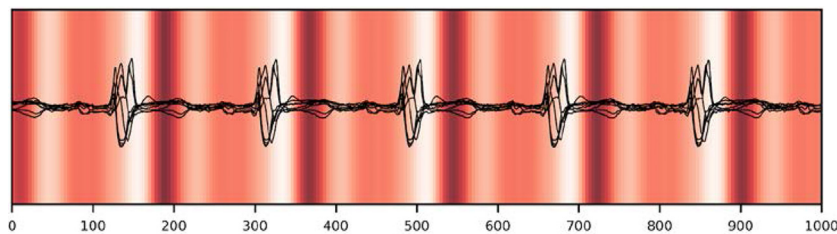
In recent years, there has been a proliferation in the successful application of deep neural networks to the ECG, including for the QT interval.^{8,23-25} For example, when applied to smartphone-enabled mobile ECGs, a CNN accurately calculated the QTc



of a standard 12-lead ECG and exhibited high specificity for correctly measuring severely prolonged (≥ 500 ms) QTc.²⁴ A model trained to recognize ECG patterns associated with exposure to the potent IKr blocker sotalol exhibited high accuracy in identifying congenital LQTS type 2.⁸ Recently, our group developed a novel CNN model for prediction of the projected sinus QTc in patients with atrial fibrillation.²⁵

Accurately predicting diLQTS has remained a challenge, with significant potential clinical impact considering the preventable nature of diLQTS and TdP. In addition, diLQTS is a common cause for premature termination of clinical studies evaluating new medications and market withdrawal of existing medications.²⁶ Yet, models to date have been largely based on the compilation of known risk factors to acutely (<24-48 hours) predict diLQTS in hospitalized

FIGURE 4 Gradient-Weighted Class Activation Mapping



Gradient-weighted class activation mapping was used for the heatmap visualization, highlighting which portions of the electrocardiogram waveform that the model focuses on when predicting QTc prolongation. Above is an example with all leads overlaid in which a darker red color indicates parts of the electrocardiogram waveform of greatest contribution to the prediction task.

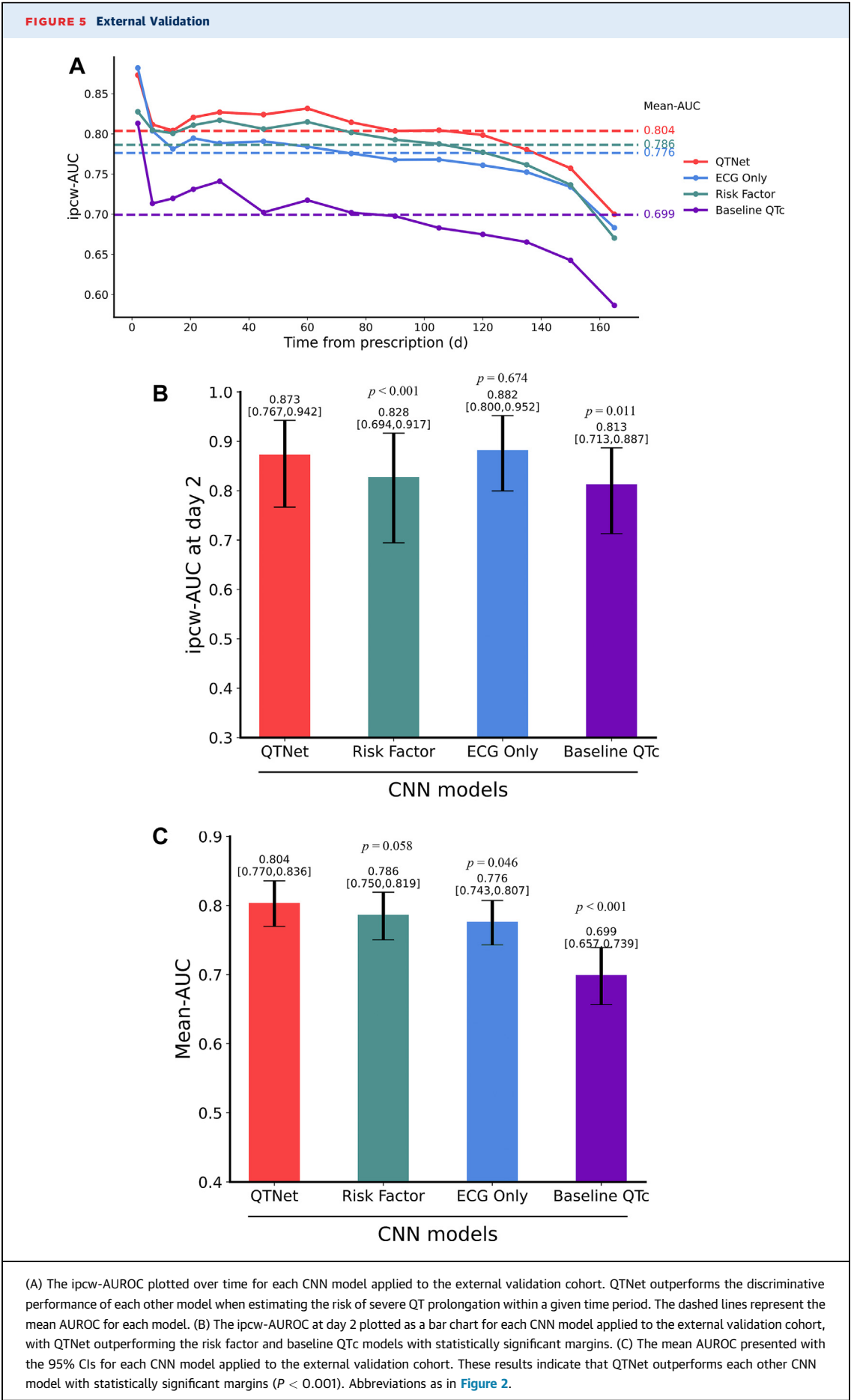
patients.^{16-19,27-30} To our knowledge, we present the first ECG-based AI model for time-weighted diLQTS risk prediction in nonhospitalized patients (**Central Illustration**).

In the inpatient population, diLQTS can be potentially monitored more reliably given the routine acquisition of ECGs or the availability of continuous telemetry monitoring.¹ On the contrary, in the outpatient setting, such ECG monitoring can be resource-intensive and impractical. As confirmed in our study, the predictive utility of the baseline QTc is limited, inadequately capturing subclinical reduction in repolarization reserve.^{1,7,8} Our AI model addresses these issues through its derivation from an outpatient cohort and its time-weighted predictive architecture over 6 months (**Figures 2A and 5A**). To ensure that the model is predicting diLQTS specifically, we mandated that severe QT prolongation in the held-out test dataset and external validation cohort had to occur within the drug prescription period. Additionally, median electrolyte levels were within normal limits both at the time of the prescription instance and when severe QT prolongation occurred among diLQTS cases (**Supplemental Table 5**). The QTNet model outperformed all reference models over the entire 6-month period, while also maintaining a predictive accuracy greater than or equal to its mean AUROC for at least the initial 3 months.

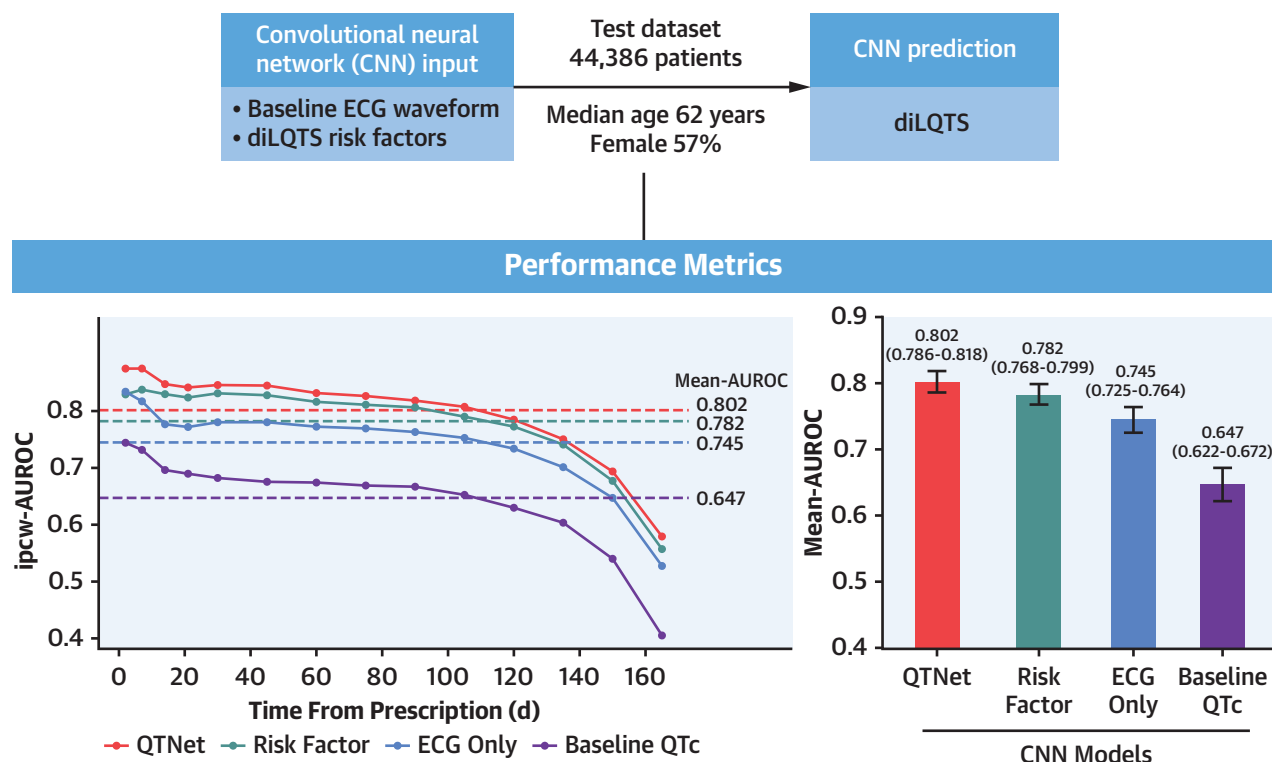
Compared with the risk factor model, which relied on traditional clinical risk factors utilized in prior predictive models,^{18,19,28} QTNet achieved superior performance by combining established predictors for diLQTS with the raw ECG signal. This incremental improvement achieved by the addition of the ECG signal (**Figures 2, 3, and 5**) suggests that the CNN captures “concealed” risk signal embedded in the ECG. This can be explained by the AI model’s ability to capture known risk factors that are inaccurately or

insufficiently reported in the electronic health record, or by capturing unknown risk features embedded in the signal, different from the recognized diLQTS clinical risk markers.³¹ In turn, this could account for the ECG-only model outperforming the risk factor model in the external validation cohort and achieving equivalent performance with QTNet (**Figures 5A and 5B**). Using Grad-CAM analysis to interrogate the ECG-based prediction signal, we found that the neural network is “focusing” on the first half of the T-wave and the PR interval while sparing the QRS complex (**Figure 4**). These findings demonstrate the potential of CNNs to serve as hypothesis-generating tools by identifying previously undescribed ECG features contributing to a prediction task.

Importantly, the QTNet model maintained the best predictive performance among all models in an external validation cohort (**Figure 5**). Unlike the main cohort used for training QTNet, our external validation cohort consisted of NYU Long Island outpatient practices (**Figure 1A**). Consequently, although the model was trained on predominantly urban outpatient practices, it was also externally validated on solely suburban outpatient locations exhibiting similar predictive performance. This supports QTNet’s generalizability to outpatient practices of varying geographic location. To further inform QTNet’s generalizability from a pharmacotherapy standpoint, we calculated the percentage breakdown of the top medication classes among diLQTS cases (**Supplemental Table 6**). Notably, QTNet achieved a similar predictive performance despite different medication class makeups in the held-out test dataset and external validation cohort. The subgroup analysis also speaks to the model’s generalizability to patients with varying demographic characteristics. QTNet exhibited higher predictive performance in groups with male sex and normal



CENTRAL ILLUSTRATION Predicting Time Dependent Drug Induced QT Prolongation Using Artificial Intelligence



Zhang H, et al. J Am Coll Cardiol EP. 2024;10(5):956-966.

We identified all adult patients with a new outpatient prescription for a QT-prolonging medication and with a 12-lead sinus electrocardiogram (ECG) within 6 months prior to the prescription instance. Using risk factor data and the ECG signal as inputs, we designed the convolutional neural network (CNN) QNet to predict drug-induced long QT syndrome (diLQTS) in the outpatient setting. Compared with 3 other models relying on risk factors or ECG signal or baseline QTc alone, QNet achieved the best ($P < 0.001$) performance with a mean area under the receiver-operating characteristic curve (AUROC) of 0.802 (95% CI: 0.786-0.818), as well as the highest inverse probability of censorship-weighted area under the receiver-operating characteristic curve (ipcw-AUROC) in a survival analysis up to 6 months.

baseline QTc (Figures 2C and 3B). Interestingly, the association of age with predictive performance was bimodal, with marked precision in the youngest age group (≤ 50 years of age). In a real-world setting, the present model could identify higher-risk individuals who would benefit from more rigorous monitoring in an outpatient clinical setting or in a trial assessing the safety of a novel drug.

STUDY LIMITATIONS. Limitations of this study include its single center and retrospective design. The single-center design limitation is partially mitigated by the use of the external validation cohort included in the study. Separately, the future clinical adoption of this AI model necessitates that a hospital or clinic utilize an electronic health record and digital ECG database to derive risk factors and the raw ECG signal as inputs for the CNN, respectively. Future studies will focus on the prospective application of QNet in

an external population to further characterize its predictive performance.

CONCLUSIONS

We report QNet, a CNN model that includes the ECG signal input for the outpatient prediction of diLQTS. We demonstrate a superior performance over previously described prediction models, with time-dependent metrics. In an external validation cohort of solely suburban outpatient practices, QNet maintained its predictive performance. Prospective studies are warranted to test the applicability of QNet, particularly as it relates to ECG monitoring frequency and modification of drug therapy after administration of a QT-prolonging medication.

ACKNOWLEDGMENTS The computational requirements for this work were supported in part by the NYU

Langone High Performance Computing (HPC) Core's resources and personnel.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

The authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Lior Jankelson, Leon H. Charney Division of Cardiology, NYU Langone Health, 560 First Avenue, New York, New York 10016, USA. E-mail: lior.jankelson@nyulangone.org. @ExtraStim. OR Mr Hao Zhang, NYU Langone Health, 560 First Avenue, New York, New York 10016, USA. E-mail: hao.zhang@nyulangone.org.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: An ECG-based CNN exhibited the highest accuracy in predicting diLQTS in the outpatient setting over 6 months. Predictive performance was maintained in an external validation cohort and subgroups with varying demographic characteristics.

TRANSLATIONAL OUTLOOK: In a real-world setting, the present model could identify higher-risk individuals who would benefit from more rigorous QTc monitoring in the outpatient setting.

REFERENCES

- Drew BJ, Ackerman MJ, Funk M, et al. Prevention of torsade de pointes in hospital settings. *J Am Coll Cardiol*. 2010;55:934-947.
- Roden DM. Drug-induced prolongation of the QT interval. *N Engl J Med*. 2004;350:1013-1022.
- Straus SMJM, Kors JA, De Bruin ML, et al. Prolonged QTc interval and risk of sudden cardiac death in a population of older adults. *J Am Coll Cardiol*. 2006;47:362-367.
- Roden DM. Long QT syndrome: reduced repolarization reserve and the genetic link. *J Intern Med*. 2006;259:59-69.
- Schwartz PJ, Woosley RL. Predicting the unpredictable. *J Am Coll Cardiol*. 2016;67:1639-1650.
- Roden DM. Taking the "Idio" out of "Idiosyncratic": predicting Torsades de Pointes. *Pacing Clin Electrophysiol*. 1998;21:1029-1034.
- Roden DM. Predicting drug-induced QT prolongation and torsades de pointes. *J Physiol*. 2016;594:2459-2468.
- Prifti E, Fall A, Davogusto G, et al. Deep learning analysis of electrocardiogram for risk prediction of drug-induced arrhythmias and diagnosis of long QT syndrome. *Eur Heart J*. 2021;42:3948-3961.
- Jethani N, Puli A, Zhang H, et al. New-onset diabetes assessment using artificial intelligence-enhanced electrocardiography. Preprint. *arXiv*. Posted online May 5, 2022. <https://doi.org/10.48550/arxiv.2205.02900>
- Woosley R, Heise C, Gallo T, Tate J, Woosley D, Romero K. QT drugs list. AZCERT, Inc., Tucson, AZ. Accessed October 26, 2023. <http://www.CredibleMeds.org>
- Dar T, Murtaza G, Yarlagadda B, et al. Dofetilide initiation and implications of deviation from the standard protocol - a real world experience. *J Atr Fibrillation*. 2019;12:2265.
- Bazett H. An analysis of the time-relations of electrocardiograms. *Heart*. 1920;7:353-370.
- Blanche P, Dartigues J-F, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013;32:5381-5397.
- Gerds TA, Kattan MW, Schumacher M, Yu C. Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring. *Stat Med*. 2013;32:2173-2184.
- Han X, Goldstein M, Puli A, Wies T, Perotte AJ, Ranganath R. Inverse-weighted survival games. *Adv Neural Inf Process Syst*. 2021;34:2160-2172.
- Haugaa KH, Bos JM, Tarrell RF, Morlan BW, Caraballo PJ, Ackerman MJ. Institution-wide QT alert system identifies patients with a high risk of mortality. *Mayo Clin Proc*. 2013;88:315-325.
- Tisdale JE, Jaynes HA, Kingery JR, et al. Development and validation of a risk score to predict QT interval prolongation in hospitalized patients. *Circ Cardiovasc Qual Outcomes*. 2013;6:479-487.
- Vandael E, Vandenberg B, Vandenberghe J, Vandenberg B, Willems R, Foulon V. A smart algorithm for the prevention and risk management of QTc prolongation based on the optimized RISQ-PATH model. *Br J Clin Pharmacol*. 2018;84:2824-2835.
- Hincapie-Castillo JM, Staley B, Henriksen C, Saidi A, Lipori GP, Winterstein AG. Development of a predictive model for drug-associated QT prolongation in the inpatient setting using electronic health record data. *Am J Health Syst Pharm*. 2019;76:1059-1070.
- Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med*. 2019;25:65-69.
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis*. 2019;128:336-359.
- Goldenberg I, Moss AJ, Zareba W. QT interval: how to measure it and what is "normal." *J Cardiovasc Electrophysiol*. 2006;17:333-336.
- Galloway CD, Valys AV, Shreibati JB, et al. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. *JAMA Cardiol*. 2019;4:428-436.
- Giudicessi JR, Schram M, Bos JM, et al. Artificial intelligence-enabled assessment of the heart rate corrected QT interval using a mobile electrocardiogram device. *Circulation*. 2021;143:1274-1286.
- Tarabani C, Ronan R, Shokr M, Chinitz L, Jankelson L. Development of an AI-driven QT correction algorithm for patients in atrial fibrillation. *J Am Coll Cardiol EP*. 2023;9:246-254.
- Fermini B, Fossa AA. The impact of drug-induced QT interval prolongation on drug discovery and development. *Nat Rev Drug Discov*. 2003;2:439-447.
- Strauss DG, Vicente J, Johannesen L, et al. Common genetic variant risk score is associated with drug-induced QT prolongation and torsade de pointes risk. *Circulation*. 2017;135:1300-1310.
- Simon ST, Mandair D, Tiwari P, Rosenberg MA. Prediction of drug-induced long QT syndrome using machine learning applied to harmonized electronic health record data. *J Cardiovasc Pharmacol Ther*. 2021;26:335-340.
- Simon ST, Trinkley KE, Malone DC, Rosenberg MA. Interpretable machine learning prediction of drug-induced QT prolongation: electronic health record analysis. *J Med Internet Res*. 2022;24:e42163.
- Van Laere S, Muyllle KM, Dupont AG, Cornu P. Machine learning techniques outperform conventional statistical methods in the prediction of high risk QTc prolongation related to a drug-drug interaction. *J Med Syst*. 2022;46:100.
- Khurshid S, Friedman S, Reeder C, et al. ECG-based deep learning and clinical risk factors to predict atrial fibrillation. *Circulation*. 2022;145:122-133.

KEY WORDS artificial intelligence, deep neural networks, drug-induced long QT syndrome, electrocardiogram deep learning, prolonged QT

APPENDIX For supplemental tables and a figure, please see the online version of this paper.