Deep learning analysis of electrocardiogram for risk prediction of drug-induced arrhythmias and diagnosis of long QT syndrome

Edi Prifti [1,2]*, Ahmad Fall [1], Giovanni Davogustto [3†], Alfredo Pulini [1,4†],
Isabelle Denjoy [5], Christian Funck-Brentano[6], Yasmin Khan[7],
Alexandre Durand-Salmon[7], Fabio Badilini[8], Quinn S. Wells[3,9], Antoine Leenhardt[5],
Jean-Daniel Zucker [1,2], Dan M. Roden [3,9,10], Fabrice Extramiana [5], and
Joe-Elie Salem [3,6,9]*

[1]IRD, Sorbonne University, UMMISCO, 32 Avenue Henri Varagnat, Bondy 93143, France; [2]Sorbonne University, INSERM, NutriOmics, 91 Boulevard de l'Hopital, Paris 75013, France; [3]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; [4]Faculty of Medicine, Université de Paris, Paris, France; [5]CNMR Maladies Cardiaques Héréditaires Rares, Hôpital Bichat, Paris, France; [6]Clinical Investigation Center Paris-Est, CIC-1901, INSERM, UNICO-GRECO Cardio-Oncology Program, Department of Pharmacology, Pitié-Salpêtrière University Hospital, Sorbonne Universite, 47 Boulevard de l'Hopital, Paris 7513, France; [7]Banook Group, Nancy, France; [8]AMPS LLC, New York, USA; [9]Department of Pharmacology, Vanderbilt University Medical Center, Nashville, TN, USA; and [10]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

See the editorial comment for this article 'Long QT syndrome, artificial intelligence, and common sense', by P. J. Schwartz and H. L. Tan, doi:10.1093/eurheartj/ehab611.

|	
Aims	Congenital long-QT syndromes (cLQTS) or drug-induced long-QT syndromes (diLQTS) can cause torsade de pointes (TdP), a life-threatening ventricular arrhythmia. The current strategy for the identification of drugs at the high risk of TdP relies on measuring the QT interval corrected for heart rate (QTc) on the electrocardiogram (ECG). However, QTc has a low positive predictive value.
Methods and results	We used convolutional neural network (CNN) models to quantify ECG alterations induced by sotalol, an I_{Kr} blocker associated with TdP, aiming to provide new tools (CNN models) to enhance the prediction of drug-induced TdP (diTdP) and diagnosis of cLQTS. Tested CNN models used single or multiple 10-s recordings/patient using 8 leads or single leads in various cohorts: 1029 healthy subjects before and after sotalol intake ($n = 14\,135$ ECGs); 487 cLQTS patients ($n = 1083$ ECGs: 560 type 1, 456 type 2, 67 type 3); and 48 patients with diTdP ($n = 1105$ ECGs, with 147 obtained within 48 h of a diTdP episode). CNN models outperformed models using QTc to identify exposure to sotalol [area under the receiver operating characteristic curve (ROC-AUC) = 0.98 vs. 0.72, $P \le 0.001$]. CNN models had higher ROC-AUC using multiple vs. single 10-s ECG ($P \le 0.001$). Performances were comparable for 8-lead vs. single-lead models. CNN models predicting sotalol exposure also accurately detected the presence and type of cLQTS vs. healthy controls, particularly for cLQT2 (AUC-ROC = 0.9) and were greatest shortly after a diTdP event and declining over time ($P \le 0.001$), after controlling for QTc and intake of culprit drugs. ECG segment analysis identified the J-T_{peak} interval as the best discriminator of sotalol intake.
Conclusion	CNN models applied to ECGs outperform QTc measurements to identify exposure to drugs altering the QT interval, congenital LQTS, and are greatest shortly after a diTdP episode.

---

* Corresponding authors. Email: edi.prifti@ird.fr (E.P.); Tel: +33 (0)1.42.17.85.35, Email: joe-elie.salem@aphp.fr (J.-E.S.)
† These authors contributed equally to the work.

Graphical Abstract
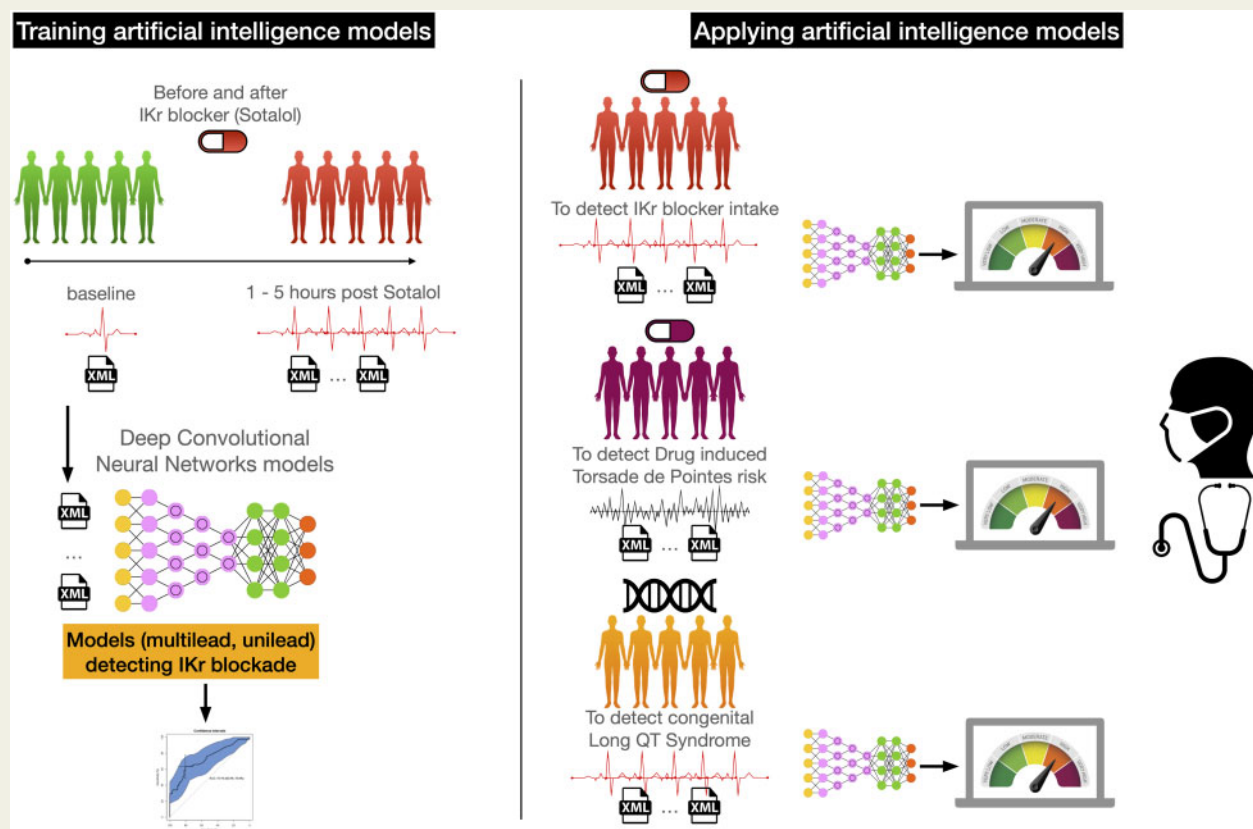
Convolutional neural network models applied to ECGs outperform QTc measurements to identify exposure to drugs blocking IKr, congenital long QT syndrome, and are greatest shortly after a drug-induced Torsade-de-Pointes episode.

Introduction

Torsades de pointes (TdP) is a distinctive form of life-threatening polymorphic ventricular arrhythmia associated with prolonged QT interval, corrected for heart rate (QTc), on the electrocardiogram (ECG).[1–3] TdP and QTc prolongation are favoured by congenital or drug-induced alterations in potassium and cardiac sodium channels.[4–6] There are three main forms of congenital long QT syndromes (cLQTS): type 1 and type 2 are caused by loss-of-function mutations in the potassium channels *KCNQ1* (cLQT1, I_{Ks} current) and *KCNH2* (cLQT2, I_{Kr} current), respectively, and type 3 is caused by mutations in *SCN5A* increasing the non-inactivating 'late' sodium current I_{NaL} (cLQT3).[6–8] Drug-induced LQTS (diLQTS) is the other main cause of TdP, with almost all culprit drugs blocking I_{Kr} and the most torsadogenic among them also activating I_{NaL}.[5] Over 100 cardiac or non-cardiac drugs are currently approved despite favouring TdP risk because these drugs are thought to have a favourable risk–benefit ratio in some patients.[9,10]

QTc, which reflects ventricular repolarization duration, is the time between the beginning of the QRS complex and the end of the T-wave.[11] QTc is prolonged in cLQTS and diLQTS and is a hallmark of TdP. Specific T-waveform patterns have been described for each sub-type of cLQTS and for diLQTS.[12–15] Current individual and population risk stratification strategies for TdP are almost exclusively based on the quantification of QTc.[4] Regulatory agencies require new drugs to undergo thorough QT studies, where the magnitude of drug-induced QTc prolongation is evaluated as a surrogate for TdP risk.[16] However, limiting ECG evaluation to the QTc is poorly predictive of TdP.[17] An unbiased and complex examination of the ECG data beyond simple QTc prolongation could provide relevant insight into identifying drugs and patients at risk of TdP.

Artificial intelligence is being increasingly applied to complex medical problems.[18] Techniques such as deep learning, including convolutional neural networks (CNN), are bringing a radical change in the field of pattern recognition, improving earlier models in learning tasks such as image classification, ECG analysis, and natural language

processing.[19–21] Herein, we tested if such models were able to learn the ECG footprint of sotalol, an I_{Kr} blocker drug inducing TdP, to develop a new tool using ECG to recognize beyond QTc, exposure to I_{Kr} blocker drugs, and improve the prediction of drug-induced TdP (diTdP) events and classification of cLQTS types, particularly cLQT2 (*Graphical Abstract*).

Methods

Study cohort datasets and QTc measurement

We studied ECGs form four cohorts (*Figure 1*). The 'Generepol cohort' (NCT00773201)[15,22] was conducted at Pitié-Salpétrière Clinical Investigation Center (start–end: 2008–12, Paris, France): ECGs from 990 healthy subjects were recorded before and 1, 2, 3, and 4 h after an 80-mg oral sotalol dose (sotT1, sotT2, sotT3, and sotT4). The 'Pharmacia's cohort' was an open-label, nonrandomized study involving healthy controls ($n = 39$, 28 males) receiving a fixed oral sotalol sequence administered on 3 successive days: 24-h baseline without sotalol (Day 0); 160 mg in all participants at 8:00 am Day 1; and 320 mg in 21 males at 8:00 am Day 2. The study was conducted at Pharmacia's Clinical Research Unit (start–end: 2002; Kalamazoo, MI, USA).[23,24] The 'cLQTS cohort' included 487 patients confirmed by genetic testing to have one of the three main cLQTS followed at the Arrhythmia Unit of Bichat Hospital (start–end:

1992–2018, Paris, France; 64% asymptomatic).[25] The 'diTdP cohort' included 48 patients prospectively enrolled and followed at Vanderbilt University Medical Center (start–end: 2002–19, Nashville, TN, USA) who had experienced at least one diTdP episode; acute cardiac ischaemia at the time of the event and genetically confirmed underlying cLQTS were exclusion criteria. All cohorts were approved by institutional review boards, and written informed consent was obtained from participants when appropriate.

Recordings from these patients were reviewed by two expert cardiologists and tracings with ventricular or junctional tachycardia during the 10-s acquisition were excluded from the analyses. In all cohorts, QTc was heart rate corrected with Fridericia's formula and details concerning the respective inter and intra-observer variability for QTc measurements in the cohorts are detailed elsewhere.[11,15,22–25]

Data preparation

Raw 10-s ECG data (sampling frequency: 250 and 500 Hz) were acquired with a variety of devices at the different centres. The 250-Hz signals were up-sampled to 500 Hz using a cubic interpolation. The ECG contained eight independent leads (LI, LII, V1–V6), allowing for the reconstruction of 12 leads (addition of LIII, aVF, aVL, aVR). ECGs were provided in .scp or .xml files depending on recording devices (General Electric MAC5500, Marquette MAC15/MACVU, M3700 System, PageWriter Touch/Trim/XL/TC, Mortara ELI200 and Cardionics Cardioplug devices). They were parsed using Biosig software, and Python xmltodict library, as appropriate.[26] The data were stored in Python dictionaries and converted onto

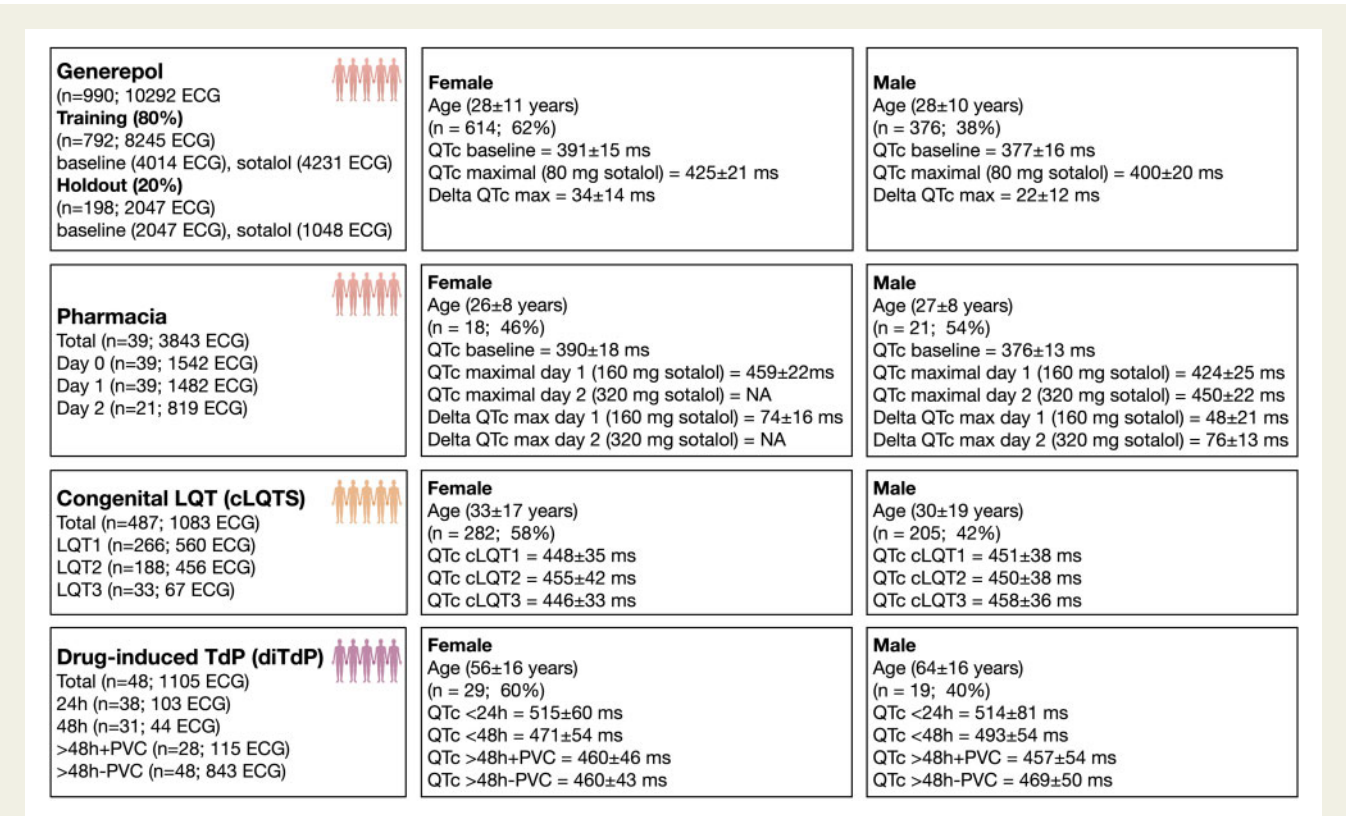| **Generepol** (n=990; 10292 ECG) **Training (80%)** (n=792; 8245 ECG) baseline (4014 ECG), sotalol (4231 ECG) **Holdout (20%)** (n=198; 2047 ECG) baseline (2047 ECG), sotalol (1048 ECG)	**Female** Age (28±11 years) (n = 614; 62%) QTc baseline = 391±15 ms QTc maximal (80 mg sotalol) = 425±21 ms Delta QTc max = 34±14 ms	**Male** Age (28±10 years) (n = 376; 38%) QTc baseline = 377±16 ms QTc maximal (80 mg sotalol) = 400±20 ms Delta QTc max = 22±12 ms
Pharmacia Total (n=39; 3843 ECG) Day 0 (n=39; 1542 ECG) Day 1 (n=39; 1482 ECG) Day 2 (n=21; 819 ECG)	**Female** Age (26±8 years) (n = 18; 46%) QTc baseline = 390±18 ms QTc maximal day 1 (160 mg sotalol) = 459±22ms QTc maximal day 2 (320 mg sotalol) = NA Delta QTc max day 1 (160 mg sotalol) = 74±16 ms Delta QTc max day 2 (320 mg sotalol) = NA	**Male** Age (27±8 years) (n = 21; 54%) QTc baseline = 376±13 ms QTc maximal day 1 (160 mg sotalol) = 424±25 ms QTc maximal day 2 (320 mg sotalol) = 450±22 ms Delta QTc max day 1 (160 mg sotalol) = 48±21 ms Delta QTc max day 2 (320 mg sotalol) = 76±13 ms
Congenital LQT (cLQTS) Total (n=487; 1083 ECG) LQT1 (n=266; 560 ECG) LQT2 (n=188; 456 ECG) LQT3 (n=33; 67 ECG)	**Female** Age (33±17 years) (n = 282; 58%) QTc cLQT1 = 448±35 ms QTc cLQT2 = 455±42 ms QTc cLQT3 = 446±33 ms	**Male** Age (30±19 years) (n = 205; 42%) QTc cLQT1 = 451±38 ms QTc cLQT2 = 450±38 ms QTc cLQT3 = 458±36 ms
Drug-induced TdP (diTdP) Total (n=48; 1105 ECG) 24h (n=38; 103 ECG) 48h (n=31; 44 ECG) >48h+PVC (n=28; 115 ECG) >48h-PVC (n=48; 843 ECG)	**Female** Age (56±16 years) (n = 29; 60%) QTc <24h = 515±60 ms QTc <48h = 471±54 ms QTc >48h+PVC = 460±46 ms QTc >48h-PVC = 460±43 ms	**Male** Age (64±16 years) (n = 19; 40%) QTc <24h = 514±81 ms QTc <48h = 493±54 ms QTc >48h+PVC = 457±54 ms QTc >48h-PVC = 469±50 ms

Figure 1 Experimental design and main characteristics of the study cohorts. Description of the main characteristics of the four study cohorts. The Generepol cohort was composed of healthy volunteers given a single 80-mg dose of oral sotalol. This dataset was used to train and test the models. The Pharmacia's cohort was composed of 39 healthy volunteers before (Day 0) and after a single 160-mg dose of oral sotalol (Day 1), followed in some men by a single 320 mg dose of oral sotalol (Day 2). This cohort was only used to test the models. The congenital long-QT syndrome (cLQTS) cohort was composed of congenital long-QT syndrome patients of Types 1, 2, and 3. The drug-induced torsade de pointes (diTdP) cohort included patients who experienced events of drug-induced torsade de pointes with no underlying identified congenital long-QT syndrome.

3D tensors (8 leads, 5000 time points for each lead, recordings) used to train and test the models. Standardization was performed at the whole ECG level, with each lead signal standardized by the mean of all other lead signals for models including all eight leads ('multilead') and at the lead level for 'unilead' models. No other transformations, including filtering, were used.

Sotalol-intake classification with the multilead and unilead models

We used either the eight leads concomitantly (LI, LII, V1–6; termed 'multilead') or each of the eight leads independently ('unilead') to train a CNN model to predict $Sot+$ (having received sotalol, as a surrogate for I_{Kr} blockade) and $Sot-$ classes (normal ECG before sotalol intake). The Generepol cohort (healthy volunteers before and after sotalol intake) was split into two sets: general training (80% for multilead models, 90% for unilead models) and holdout (20% for multilead models, 10% for unilead models). Ten times 10-fold cross-validation was performed in the general training set for parameter optimization. Each split was performed according to the subjects' IDs and, therefore, each training partition had distinct subjects from the testing split. Descriptions of how the multilead and unilead models were constructed are provided in the Supplementary material online, *Figures S1 and S2*. After cross-validation, each model was trained on the training set of the Generepol cohort. Then, models were tested on the holdout Generepol set (completely independent of the training set) and the three other study cohorts.

Voting vs. single electrocardiogram analysis

Performance indicators were computed using both 10-s single ECG signal analysis (ECG level in figures) and by averaging risk scores from multiple recordings acquired within minutes of a same timepoint (multiple 10-s recordings; patients' level in figures; i.e. the voting analysis) for a given patient and condition. The output provided by the models was a score ranging from 0 to 1 indicating a likelihood of being $Sot+$ (having ingested sotalol). To classify a patient into $Sot+$ or $Sot-$ classes, ECGs from the same patient were processed by the models and the patient was affected as being $Sot+$ (vs. $Sot-$) based on the mean classification score of the different 10-s ECG, on which a threshold of 0.5 was applied ($Sot+$ if score ≥0.5). The performance metrics of all tested models can be found in *Figures 2–4*, Supplementary material online, *Figures S3 and S4*, and Supplementary material online, *Table S1*.

Embedding analyses

CNN models generate outputs, such as $Sot+$ or $Sot-$, by analysing raw input through a series of intermediate 'layers' termed embeddings.[27] A distinctive feature of CNN models is their ability to discover novel representations of complex data, and one way to access such knowledge is by extracting the embeddings (transformation of the input data by the neural network). In this study, ECGs were transformed in the CNN embeddings by deriving vectors of 512 values. To represent these complex datasets in two dimensions for human interpretation, a nonlinear dimension reduction technique was applied based on the t-SNE algorithm[28] (perplexity = 100, iteration = 1000) using the Rtsne package. ECG data (vectors of 512 values) were thus visualized and annotated as points on these maps. All dimensions of the embeddings were used to identify partitions with the k-means method with default parameters implemented in base R. Details concerning embedding analyses are in Supplementary material online, *Figure S5*.

Electrocardiogram segment occlusion analysis (interpretability)

We sought to identify which parts of the ECG signals were most useful in our classification models to classify an ECG as $Sot+$. To accomplish this goal, we iteratively dropped ('occluded') a predefined portion of the data (in this case, a segment of the ECG signal) and re-performed the prediction. Here, we used a window of 50 points (corresponding to 100 ms in 500 Hz recordings) that was iteratively moved across the signal to identify which parts of the ECG signal were the most useful for the classification of ECG as $Sot+$. Feature importance profile (FIP) was generated for each segment and provided us with a relevant score for identifying which ECG segments were more or less important for predicting $Sot+$. Details concerning occlusion methods are in Supplementary material online, *Figure S6*. We implemented the occlusion method in Python with Tensorflow-2.

Statistical analyses

Data are presented as count and frequencies, or median and interquartile range (IQR) for categorical and continuous variables, respectively. We used mixed-effects linear models to best describe the data and their relations while controlling for random effects such as patient ID (multiple recordings per patient). Models were compared using ANOVA and the best models were selected based on the Akaike information criterion. Accuracy, recall, precision, F1 score, and area under the receiver operating characteristic curve (ROC-AUC) were used to evaluate the different models generated. The Chi²-test was used for comparing proportions. Statistics and graphics were performed using R-packages (lme445, lmerTest, ggplot2, pROC). A $P \le 0.05$ was deemed statistically significant; all tests were two-tailed.

Results

Study population characteristics

The main characteristics of the four study cohorts are summarized in *Figure 1*.

The Generepol cohort contained 10 292 10-s ECG recordings from 990 healthy subjects (62% women, median [range] age 24 [18–60] years) in sinus rhythm before and 1, 2, 3, and 4 h after the administration of 80 mg sotalol (respectively, denoted as baseline and sotT1–sotT4). The median number of 10-s ECG/participant in this cohort was 15 [range: 12–18].

The Pharmacia's cohort contained 3843 10-s ECG recordings from 39 healthy subjects (46% women, median [range] age 25 [18–45] years) in sinus rhythm before and up to 12 h after the intake of 160 mg sotalol on Day 1 and 320 mg sotalol on Day 2. The median number of 10-s ECG/participant in this cohort was 114 [range: 42–117].

The cLQTS cohort included 487 participants (median [range] age 28 [0–84] years; confirmed by genetic testing) with 1083 10-s ECG recordings (median number of ECG/patient 3, IQR 6, longest follow-up 23 years). The three cLQTS types were represented, with 266 cLQT1 (62% women), 188 cLQT2 (54% women), and 33 cLQT3 (45% women) patients. A total of 213 participants (44%) had at least one recording performed while on beta-blocker, with 116, 88, and 9 (44%, 47%, and 27%) participants for cLQT1, cLQT2, and cLQT3, respectively. ECGs were in sinus rhythm, except for 8 (0.7%) with supra-ventricular arrhythmia and 2 (0.2%) with either atrial and/or ventricular pacing.
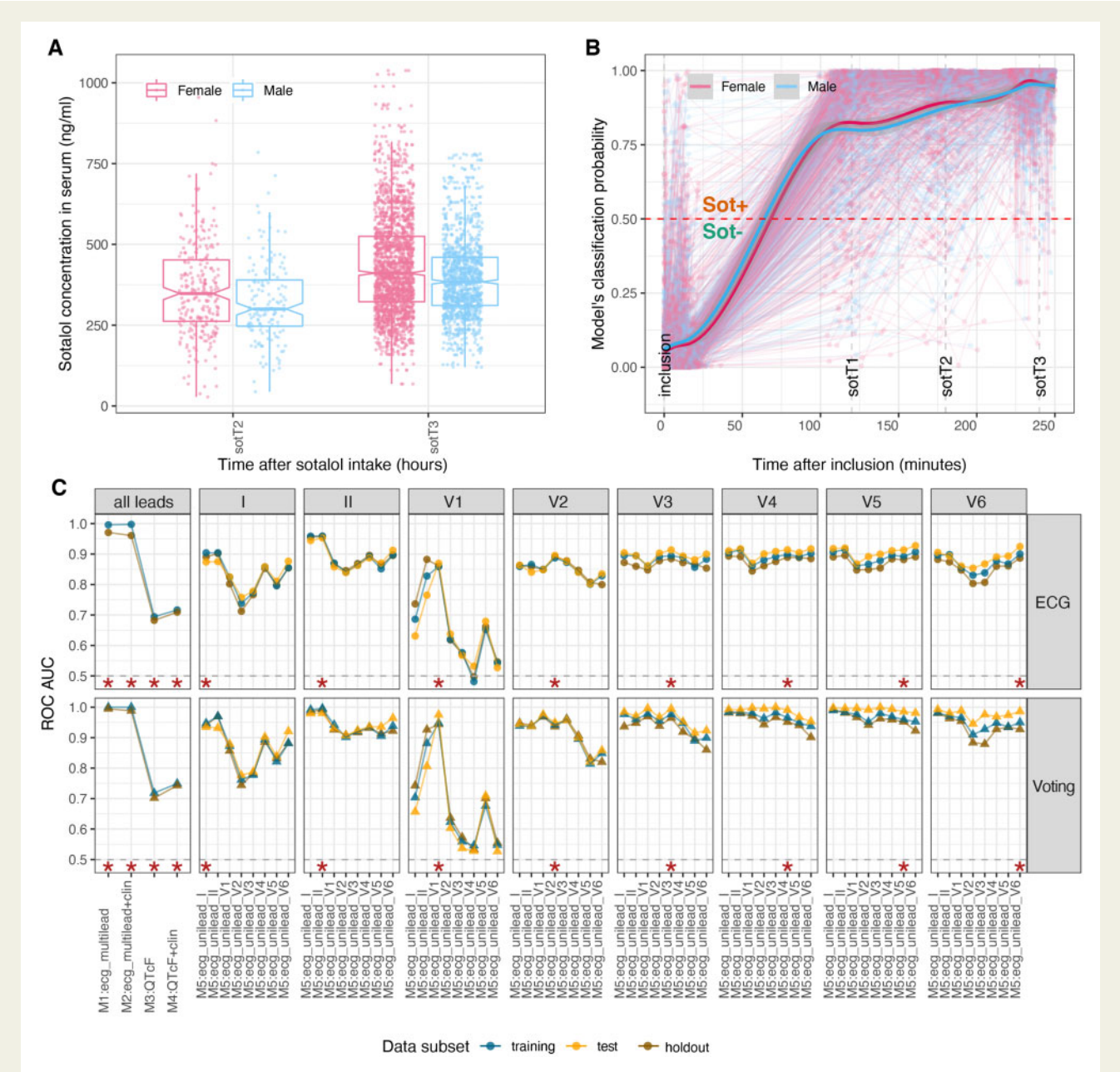
Figure 2 Classification performance of convolutional neural network and linear regression (QT) models in discriminating baseline electrocardiogram before sotalol from those after sotalol intake (SotT1, SotT2, SotT3) in Generepol. (A) Boxplots, illustrating the distribution of circulating sotalol concentration (ng/mL) in Generepol cohort two and three hours after 80mg oral sotalol intake. Data are displayed separated and coloured by gender. (B) Scatterplot illustrating the evolution of the M1: ecg_multilead classification score for the *Sot+* class (*y*-axis) across time from inclusion (*x*-axis) in the Generepol cohort. All points (averaged electrocardiograms) of a study participant are linked together as trajectories and are coloured by gender. Summarized loess (local regression) distribution of the data ± standard error is overlaid on top and grouped by gender. The red horizontal line corresponds to the *Sot+/Sot−* classification threshold (= 0.5). (C) Area under the receiver operating characteristic curve for the convolutional neural network multilead models (M1, M2), non- convolutional neural network standard QT-based linear regression models (M3, M4) as well as all convolutional neural network unilead M5 models in classifying each individual 10-s electrocardiogram recording (top) or using a voting strategy (in triplicates of 10-s electrocardiogram per study participant and time point, bottom). Multiple 10-s electrocardiograms recorded at each time point were assigned a *Sot+* classification score. When the risk score was ≥0.5, the electrocardiogram was classified as *Sot+*. With the voting approach, a mean *Sot+* classification score was computed. The same threshold was applied to predict the *Sot−/Sot+* class. Blue, orange, and brown colours, respectively, depict the training, test, and holdout subsets of the Generepol's cohort (see *Figure 1*). Each model tested on the same lead as trained is annotated by a red star. For the multilead models, all leads are used to train and test.
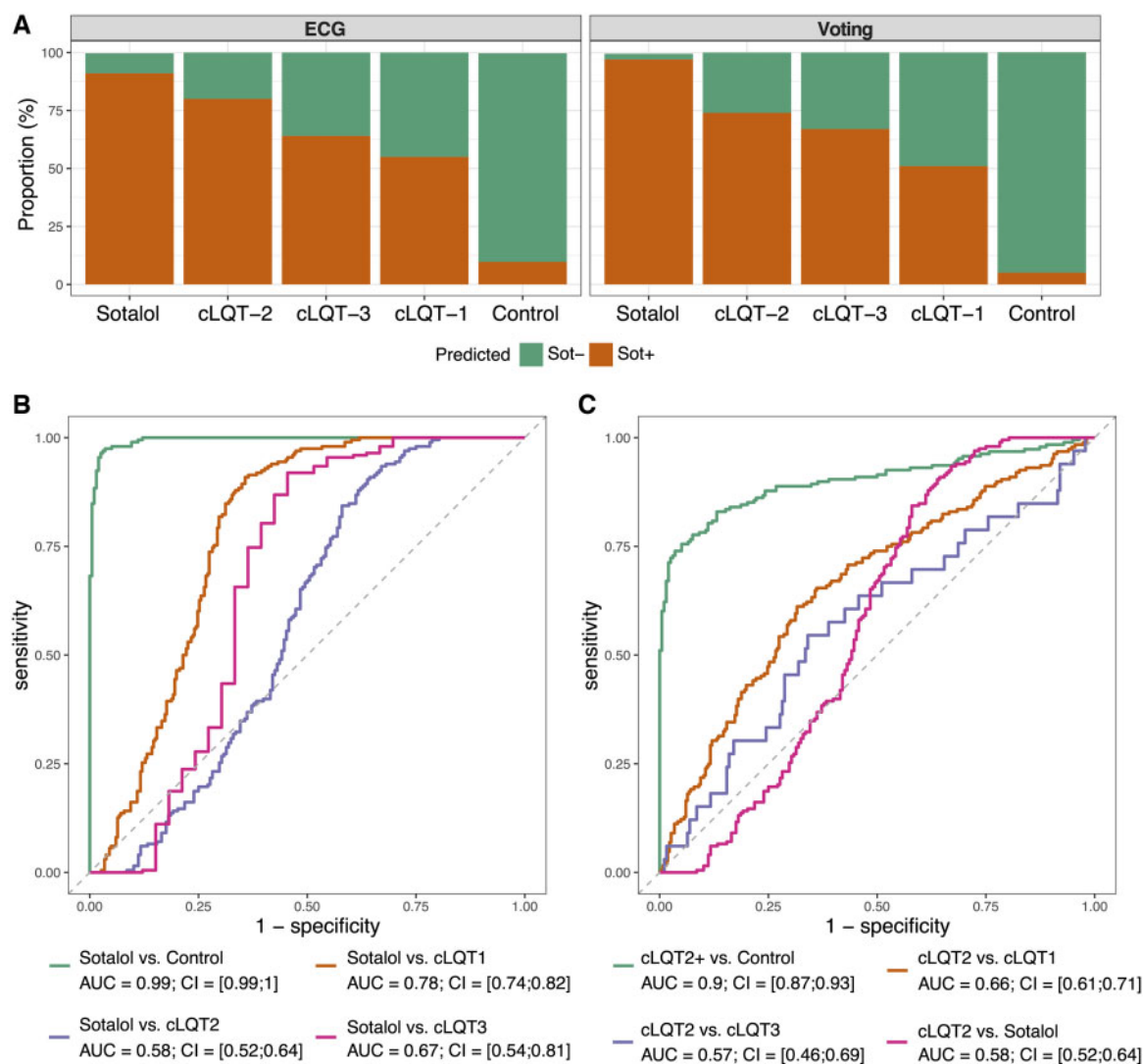
Figure 3 Convolutional neural network model performance in classifying study participants as Sot+/Sot- in Generepol holdout dataset and congenital long-QT syndrome cohort. (*A*) Left: Percentage of all electrocardiogram for study participants, which are classified as *Sot+* in the holdout Generepol dataset [healthy volunteers before (Control) and 1–3 h after sotalol intake (Sotalol)] as well as the cLQT1, cLQT2, and cLQT3 groups. Right: Similar to the left panel, with the exception that groups of electrocardiogram were classified as *Sot+* using the patient voting strategy instead of individual 10-s electrocardiogram. (*B*) Receiver operating characteristic curves indicating the separation between patients on sotalol (Sotalol) and each of the control, cLQT1, cLQT2, and cLQT3 groups. (*C*) Receiver operating characteristic curves indicating the separation between cLQT2 and cLQT1, cLQT3, sotalol exposed, and control groups.

The diTdP cohort included 48 participants (60% women; median [range] age at the time of the first ECG 60 [18–85] years) with 1105 10-s ECG recordings (median number of ECG/patient 31). The median follow-up was 4 years [range: 0–17]. Sixty-six percent of the 10-s ECG ($n = 733/1105$) were recorded while patients were on I_{Kr} blocker drugs with known risk for TdP,[9] with amiodarone (29/48), sotalol (12/48), dofetilide (9/48), fluconazole (7/48), and hydroxychloroquine (4/48) being the most prevalent.[4,15,29] Some patients took multiple drugs with TdP known risk (one drug: 69%, two drugs: 24%, three drugs: 5%). Recordings from these patients were classified into four categories using the combination of delay between ECG intake

and the diTdP event, associated with the presence/absence of premature ventricular contractions (PVC): <24 h, 24–48 h, >48 h + PVC and >48 h - PVC. Of these 1105 ECG recordings, 930 were obtained in sinus rhythm (84%), 171 (15%) in supraventricular arrhythmia, and 4 in junctional rhythm. A total of 162 (15%) and 183 (17%) 10-s ECG had at least one ventricular and/or atrial paced complex. At least one PVC was seen in 143 (13%) ECGs.

QTc evaluation

Serial QTc surveillance is the method cardiologists use to evaluate TdP risk in clinical practice.[16] When QTc is >480 ms or is increased
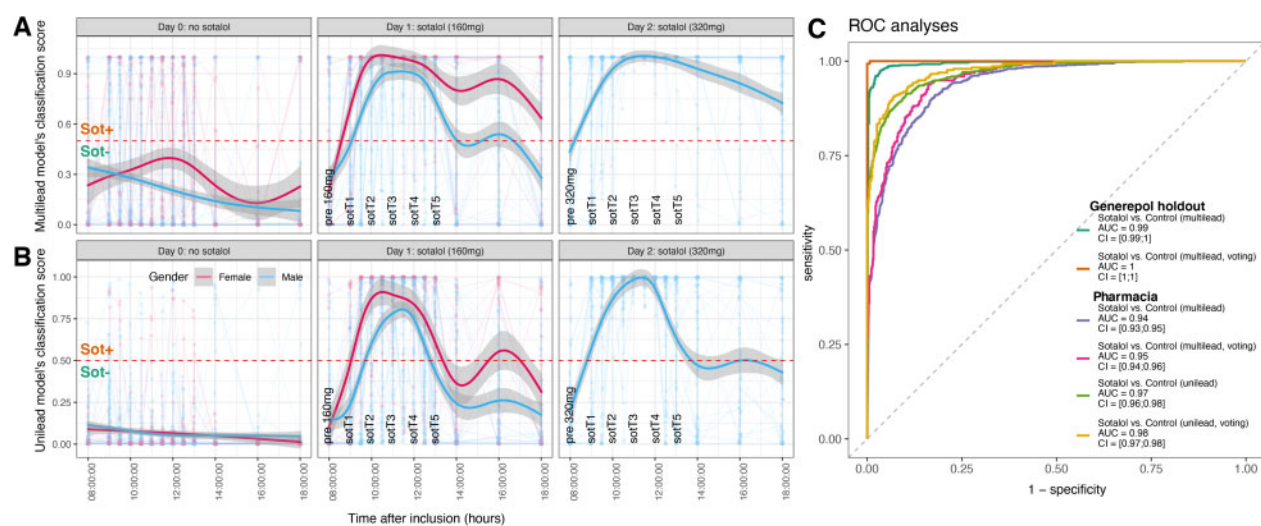
Figure 4 Convolutional neural network model performance in classifying study participants as Sot+/Sot- in Pharmacia's cohort. Scatterplot illustrating the evolution of the M1: ecg_multilead (A) and M5: ecg_unilead_II (B) classification score for the *Sot+* class (*y*-axis) across time from inclusion (*x*-axis) in Pharmacia's cohort. All points, single electrocardiograms of a study participant, are linked together as trajectories and are coloured by gender. Summarized loess (local regression) distribution of the data ± standard error is overlaid on top and grouped by gender. The horizontal dotted line corresponds to the *Sot+/Sot−* classification threshold. (*C*) Receiver operating characteristic curves indicating the separation between subjects on sotalol (Sotalol; 2–4 h post 160 mg intake on Day 1 and post 320 mg intake on Day 2) versus before (Day 0 and before intake of Day 1) in Pharmacia's study.

by ≥60 ms after drug intake compared to baseline, patients are considered at potential TdP risk.[16] In Generepol, the mean QTc at baseline was 14 ms higher in women vs. men (391 ± 15 vs. 377 ± 16 ms; $P < 2e-16$) as is well-recognized.[8,30] Maximal QTc prolongation after sotalol was more pronounced in women vs. men (34 ± 14 vs. 23 ± 12 ms; $P < 2e-16$). Similar results were obtained in the Pharmacia's study when comparing QTc before and after sotalol intake (*Figures 1 and 5*).

In cLQTS, no difference in QTc was detected among the three types of cLQTS on the first ECG available for each patient (449 ± 36, 453 ± 40, and 452 ± 35 ms for cLQT1, cLQT2, and cLQT3, respectively, $n = 483$; *Figure 5*). The mean QTc in the cLQTS cohort was 65 ms greater than the pre-sotalol values from Generepol (451 ± 38 vs. 386 ± 18 ms, $P < 2e-16$).

In the diTdP cohort, QTc values were higher within 24-h of diTdP events (501 ± 70 ms) vs. within 24–48 h (478 ± 45 ms; $P < 0.02$), or vs. >48 h with and without PVC (455 ± 50 ms, $P < 2.14e-8$ and 459 ± 45 ms, $P < 8.5e-12$, respectively; *Figure 5*). The mean QTc in the diTdP cohort was 86 ms longer than the pre-sotalol values from Generepol (469 ± 63 vs. 386 ± 18 ms, $P < 2e-16$).

Convolutional neural network models and sotalol intake on electrocardiogram

To learn the sotalol footprint as a proxy of drug-induced I_{Kr} blockade on ECG, we trained different CNN models (M) on a subset of Generepol. The first model used all leads (LI–II, V1–V6) from raw ECG data (M1: ecg_multilead). A second model used clinical information (age, sex, and serum potassium) in addition to the ECG data

(M2: ecg_multilead +clin). In this study, we first focused on 10-s ECG recordings at baseline before sotalol, and 1, 2, and 3 h after sotalol intake.

The models provided an output score indicating a likelihood of sotalol intake in the [0–1] range. A score of 0 predicted the absence of sotalol intake whereas 1 corresponded to the highest probability for sotalol exposure. The mean predicted score at baseline was low (0.06) but increased rapidly for ECG recorded at one (SotT1, 0.80) and two (SotT2, 0.88) and peaked at 3 h after sotalol (SotT3, 0.95) (*Figure 2*); there were no sex differences. Notably, this increase in model score predictions tracked the increase in sotalol blood concentration (*Figure 2*). The output score was then converted into a binary variable based on a threshold (*Sot-* if model-derived score <0.5, *Sot+* if ≥0.5). Performance indicators [ROC-AUC, accuracy, precision, recall (all ranging within [0–1]), and F1 score (ranging within [0–0.5])] were evaluated for each model in 10-s ECG recording individually or on the mean of multiple 10-s ECG of the same participant at a given time point ('voting strategy'), in the training, cross-validation, and holdout sets (*Figures 2 and 3* and Supplementary material online, *Figures S3 and S4*). The mean cross-validation ROC-AUC of M1: ecg_multilead for discriminating the ECG of patients before vs. after sotalol intake was 0.948 when computed on single 10-s ECG and 0.98 with the voting approach. Similarly, for M2: ecg_multilead + clin, the mean test accuracy was 0.948 (ECG) and 0.98 with voting (*Figure 2* and Supplementary material online, *Figure S3*). No difference was observed between M1 and M2. This indicated that the information contained in age, sex, and serum potassium was likely embedded in the ECG footprint captured by the CNN model.
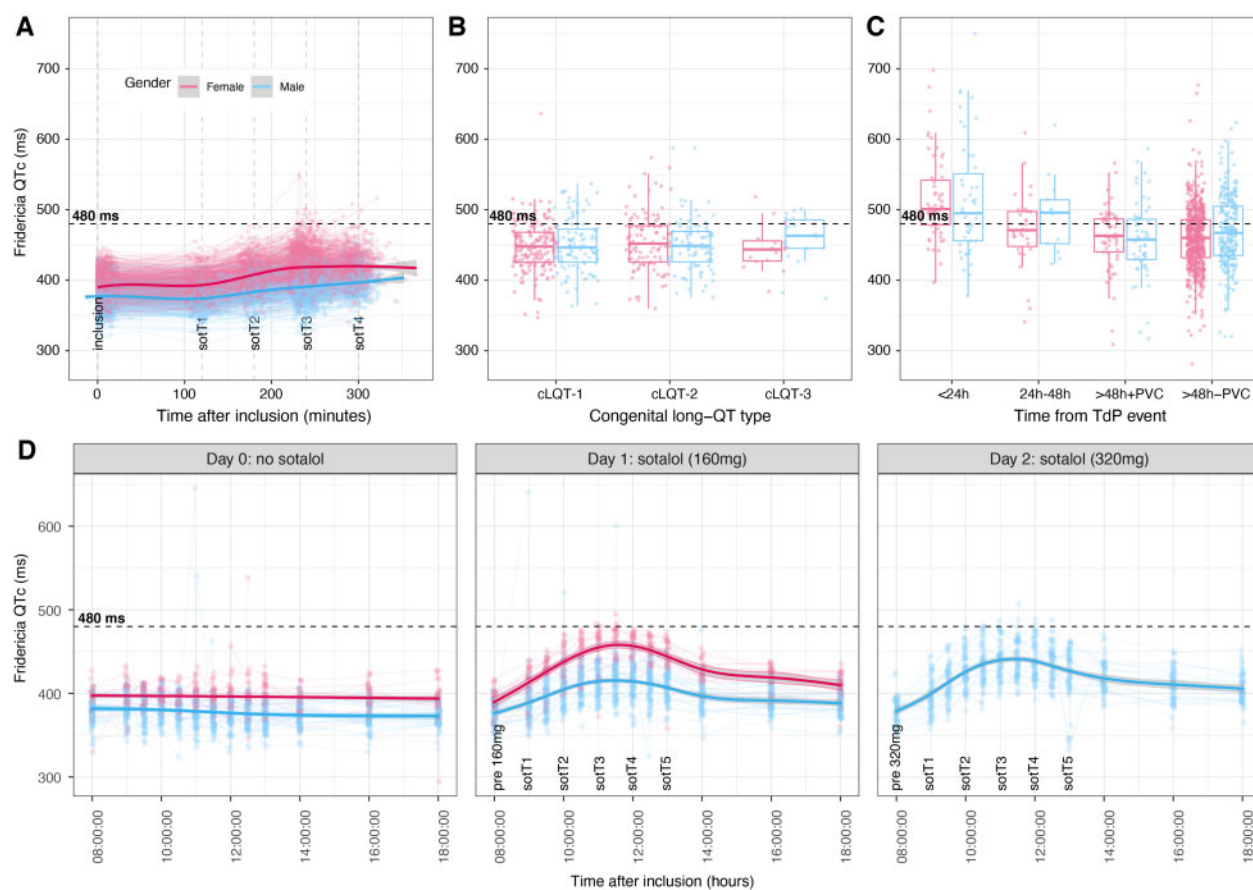
Figure 5 QTc distribution across study cohorts. The distribution of QTc values following the sotalol-induced QTc prolongation in the Generepol (A) and Pharmacia's cohort (D). The X-axis represents the time (min) following sotalol administration and lines link electrocardiogram recordings from the same participant over the duration of the protocol. Summarized loess (local regression) distribution of the data ± standard error is overlaid on top and grouped by gender (males in blue and females in red). (B) Boxplots of the estimated QTc values in the congenital long-QT syndrome cohort by subtypes. (C) Boxplots of QTc values across the drug-induced torsade de pointes cohort grouped by time to torsade de pointes event and presence or not of premature ventricular contractions. Electrocardiograms are grouped and coloured by gender and the black horizontal lines indicate the 480-ms at-risk QTc threshold (A–D).

Therefore, the M1: ecg_multilead model was deemed sufficient to be used thereafter. Its precision (voting), recall, and F1 score were very high (0.955, 0.927, and 0.470, respectively).

For comparison with current practice, we also tested the performance of QTc (M3: QTcF) alone, and with the same additional clinical information as above (M4: QTcF + clin) to discriminate on the presence/absence of sotalol intake. The linear regression model based on QTc alone (M3: QTcF) displayed a lower ROC-AUC of 0.695 (10 s ECG) and 0.720 (voting) vs. M1: ecg_multilead (ROC-AUC: 0.948 and 0.98, respectively; *P* < 1.5e-141). After integration of clinical data to QTc (M4: QTcF + clin), model performance increased significantly (*P* < 3.3e-16) to 0.717 (ECG) and 0.750 (voting) vs. M3: QTcF (*Figure 2* and Supplementary material online, *Figure S3*). Overall, QTc models were less effective than CNN models, even after integration of relevant clinical covariates. All four models (M1–4) displayed significantly higher ROC-AUC with the voting vs. individual 10-s ECG strategy (*P* < 1.2e-20 for M1: ecg_multilead, Supplementary material online, *Table S1*). This demonstrates the importance of having longer

recordings of at least 30 s (mainly triplicates of 10-s ECG in our study). Results were similar in the holdout set (*Figures 2 and 3* and Supplementary material online, *Figure S3*). All performance indicators for all these models are in Supplementary material online, *Figures S3 and S4* and Supplementary material online, *Table S1*.

Thereafter, we tested the hypothesis that the ECG footprint for sotalol exposure could also be detected by the analysis of single leads. For this, we trained eight different models—one for each lead (LI, LII, V1–V6; see 'Methods'). Their performances were comparable to the multilead models (*Figure 2* and Supplementary material online, *Figure S4*). The best scores were obtained with the model trained and tested on lead LII [M5: ecg_unilead_LII; ROC-AUC = 0.958 (10 s ECG) and 0.992 (voting) in the holdout set]. When this model trained on one lead was tested on the rest of the leads, it performed well, with mean holdout AUC-ROC of 0.883 (10-s ECG) and 0.96 (voting). However, while the mean recall was high 0.913 (10-s ECG) and 0.963 (voting), the precision was lower 0.597 (10-s ECG) and 0.605 (voting). Similar results were obtained with other unilead

models, except for the one trained on V1, which did not generalize well on the other leads (Supplementary material online, *Figure S4* and Supplementary material online, *Table S1*).

Finally, we validated M1: ecg_multilead and M5: ecg_unilead_LII models (trained in the training subset of Generepol) in the Pharmacia's cohort, an independent dataset of healthy controls before and after sotalol intake. Both M1 and M5 models performed very well to discriminate sotalol intake using ECGs (ROC-AUC 0.94–0.98 depending on the models, 10-s ECG vs. voting; *Figure 4*).

Convolutional neural network models and congenital long-QT syndromes types

Since diLQTS and cLQTS are both characterized by prolonged QTc, we hypothesized that the models (M1: ecg_multilead) trained to recognize the sotalol EGG footprint would also be able to discriminate ECG from cLQTS subjects compared to Generepol baseline data, particularly for cLQT2, which shares the same pathophysiological mechanism of I_{Kr} blockade with sotalol-induced LQTS. We used M1: ecg_multilead trained on a subset of Generepol (80%) and applied it to evaluate its potential in discriminating ECG from the heathy volunteers before and after sotalol intake (20% holdout from Generepol, never used for training) and cLQTS patients. The model prediction results confirmed our hypothesis (*Figure 3*). First, we showed that the vast majority of ECGs before and after sotalol intake (95%, 97%, voting, respectively) from the holdout Generepol cohort were correctly classified as *Sot+* and *Sot−*, respectively (*Figure 3*). Second, most ECGs from cLQTS (66%) were classified as *Sot+*. cLQT2 displayed the strongest proportion (80%, 74%) of *Sot+*, followed by cLQT3 (64%, 67%) and cLQT1 (55%, 51%) at the individual 10-s ECG level and after voting, respectively. *Figure 3* displays AUC-ROC results comparing ECGs from healthy participants on sotalol vs. their baseline ECG before sotalol (controls), cLQT1, cLQT2, and cLQT3. M1: ecg_multilead was highly efficient (AUC-ROC = 0.9) in discriminating cLQT2 from healthy controls contrasting with low AUC-ROC (0.58) of ECG analysis from cLQT2 vs. healthy subjects having received sotalol. These results indicate that M1: ecg_multilead could not discriminate well between these latter two groups, supporting the hypothesis of shared ECG footprints alterations between cLQT2 and sotalol intake (I_{Kr} blockade). Notably, M1: ecg_multilead moderately separated cLQT2 from cLQT1 and cLQT3 (*Figure 3C*). The mean M1: ecg_multilead ECG-derived score in cLQT2 was 0.53, significantly higher than cLQT1 (0.34, $P < 7.4e-7$) and cLQT3 (0.43, $P < 0.14$), after adjustment for beta-blockers (accounting for significant interaction between beta-blockers intake and cLQT2, effect size = 0.19, $P < 7.3e-6$; but not for other cLQTS types). Of note, age and sex were not significantly associated with M1: ecg_multilead score in cLQTS.

Convolutional neural network models and drug-induced torsade de pointes events

We evaluated M1: ecg_multilead model to predict the risk of diTdP events in the diTdP cohort. We quantified the association between M1: ecg_multilead score and the TdP footprint on ECG from patients who had had a diTdP event. The TdP footprint was coded as a four-class variable combining the delay from the diTdP event (<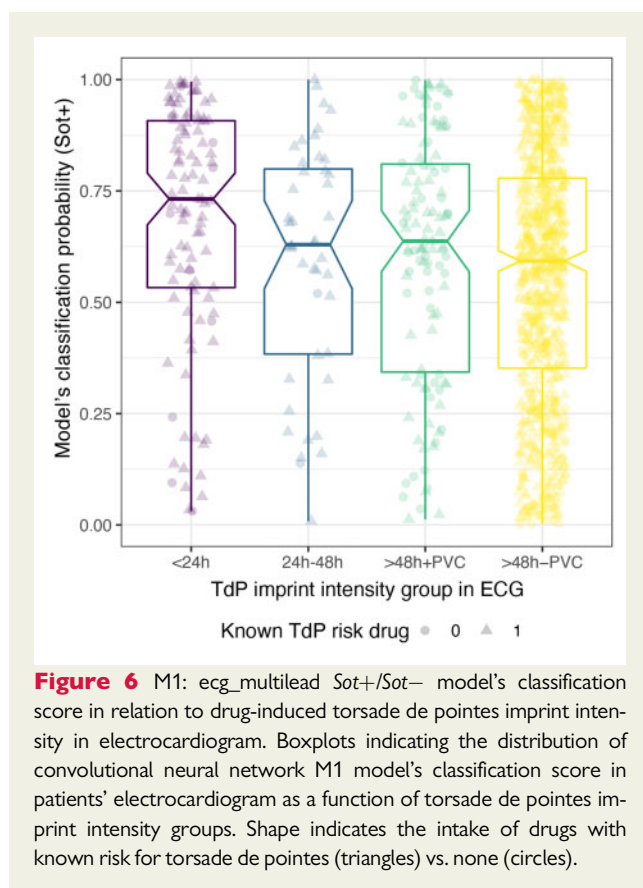24, 24–48, and >48 h) and the existence or absence of PVCs in the >48-h subgroup. Using a mixed linear model, we showed that TdP footprint was associated with the M1: ecg_multilead score (highest within 24 h from diTdP vs. >48 h from diTdP without PVC (mean: 0.68 vs. 0.56, $P < 0.0018$; *Figure 6*) after adjusting for a significant association with QTc ($P < 1.87e-10$) and intake of drugs with a known risk for TdP ($P < 3.17e-7$).

Convolutional neural network and novel representation of electrocardiogram data

The complex representation of an ECG, learned by the layers of the M1: ecg_multilead CNN model, is contextual to the presence or absence of the sotalol footprint. We extracted these representations (embeddings) of all the ECGs of the studied cohorts by accessing the output of the last convolutional layers (see Supplementary material online, *Methods*). When annotating all ECGs from Generepol as a function of the M1: ecg_multilead predicted risk score (*Figure 7A*), we noticed a gradient pattern corresponding closely to the time between ECG acquisition and sotalol intake (*Figure 7B*). This demonstrated the relevance of what the model 'learned' from the ECG data in recognizing sotalol exposure. In cLQTS, most of cLQT2 ECG were located in the high-level score zone of the t-SNE map (top part of the map), indicating ECG features resembling those of sotalol-induced I_{Kr} blockade as seen previously. This contrasts with those from cLQT1 and cLQT3, which were uniformly distributed in the t-SNE map (*Figure 7C*). In the diTdP cohort, most ECGs were located near the average to high-risk zones of the t-SNE map, being particularly high when recorded within 24 h of the diTdP (*Figure 7D*), at a time when residual I_{Kr} blockade was most likely to be present. Taken together, these results indicate that the classification accuracy in recognizing the sotalol footprint also extends to CNN M1 model-identified embeddings, which condense clinically relevant information. Such novel representations of the data open perspectives for novel TdP risk stratification of ECG and patients (Supplementary material online, *Figure S5*).

Interpretability analyses of convolutional neural network

Figure 8 displays the results of the 'occlusion analysis' designed to identify ECG sub-segments (i.e. features) most important for the models. In lead II, we found that the standardized FIP changed with increased sotalol blood concentration (maximum at 3 h in Generepol). Initially, at inclusion (before sotalol intake), the FIP was highly negative over the QRS and positive, although with low amplitude, on the P-wave offset and T-wave onset and offset. These features are used by the model to recognize normal ECG complexes without a sotalol footprint—the QRS complex indicating a regularly occurring attribute used to calibrate the data input. One hour after sotalol intake, the FIP distribution started to change. The FIP intensity of the QRS decreased and the importance of the signal after the T-wave and before P-wave onset increased. This region corresponds to the RR time, that is, cardiac heart rate. Indeed, sotalol has beta-blocking properties known to slow the sinus rate, which were captured by the model. Two hours after sotalol, the FIP increased in the first part of the T-wave

Figure 6 M1: ecg_multilead *Sot+/Sot−* model's classification score in relation to drug-induced torsade de pointes imprint intensity in electrocardiogram. Boxplots indicating the distribution of convolutional neural network M1 model's classification score in patients' electrocardiogram as a function of torsade de pointes imprint intensity groups. Shape indicates the intake of drugs with known risk for torsade de pointes (triangles) vs. none (circles).

(corresponding to the J-Tpeak interval), which reached maximum intensity 3 h after sotalol. At that time, I_{Kr} blockade was active and strongly apparent on ECG. We performed the same experiment in unilead models trained on V2 and V3 and FIP behaved similarly (*Figure 8* and Supplementary material online, *Figure S6*).

Discussion

QTc prolongation, although imperfect, has been shown to be associated with TdP and is currently used in clinical practice as a surrogate for evaluating the risk of TdP.[31] Here, we propose a new approach to improve TdP risk prediction. We hypothesized that it would be possible to use cutting edge artificial intelligence models to learn the footprint of drugs at the high risk of TdP in healthy volunteers. We then used these models to quantify a novel risk score in other participants exposed to these drugs or in patients with cLQTS. The main finding of our study is that training deep CNN models using raw digital ECG data allows for an automated and comprehensive TdP risk stratification that complements QTc measurement. The CNN was trained to recognize ECG alterations induced by sotalol as a model of I_{Kr} blockade, the major mechanism by which drugs cause QTc prolongation, and predispose to TdP.[15] The CNN models accurately detected ECG associated with the intake of drugs at risk of TdP and discriminated the presence and type of cLQTS, being particularly accurate for cLQT2. Moreover, these models improved the prediction of diTdP event, even after controlling for QTc and intake of drugs at

known risk of TdP. Analyses of the CNN models highlighted specific interpretable ECG features, particularly the J-Tpeak interval to recognize the sotalol-induced ECG footprint. Models based on a single lead performed in general as well as those using eight leads, except for V1.

Because TdP is a relatively rare event, we first used a population of healthy volunteers exposed to sotalol so we could generate enough labelled data for the CNN model to be robust. The rationale for using a cohort exposed to sotalol is that this drug is known to prolong ventricular repolarization through I_{Kr} inhibition, that rarely but dose dependently can lead to TdP.[32,33] The CNN models developed here were able to accurately classify if a patient was or not exposed to sotalol, regardless of the time after drug intake. Furthermore, multiple acquisitions taken together with a voting approach improved the classification. This demonstrated the presence of rich information contained within the ECGs, exceeding the sole measurement of QTc including with relevant clinical information. Classification from ECG features learned in the CNN models could become a useful approach in compliance ascertainment and drug adjustment, eventually more practical, less costly, and faster than standard blood analysis.

Similar molecular and physiological mechanisms to sotalol action are known to be involved in cLQT2 patients with *KCNH2* mutations, which also lead to decreased I_{Kr} current.[15] Here, we demonstrated that the similarities of the sotalol ECG footprint with cLQT2 allowed to accurately classify 80% of the ECG from cLQT2 patients. This result has potential clinical applications such as screening incoming patients for cLQTS and discrimination of types, with very low cost, before using more expensive genetic tests or scarce expert ECG repolarization evaluation. Although QTc is prolonged in all cLQTS, the ECG waveforms carry specificities including T-wave morphology abnormalities that are specific to each type of cLQTS.[34] However, the models developed herein were not trained to distinguish the different cLQTS groups, particularly cLQT1 and 3, for which more data are needed.

When applied to an independent study cohort of patients who experienced diTdP events, our CNN model-derived scores were higher within 24 h of the diTdP events vs. ECGs from same individuals >24 h (and even more 48 h) after or before the event. These results indicate that such models could be helpful to diagnose patients who experienced an out-of-hospital TdP event or even risk stratify patients with continuous surveillance for emerging diTdP events.

To the best of our knowledge, this is the first study that successfully deploys the original approach of learning drug footprints to predict drug-induced heart pathology risk based on ECG. A prior study was able to correlate drug concentrations on ECG using CNN.[35] The authors analysed 10-s ECG recordings of 42 patients receiving dofetilide, another I_{Kr} blocker antiarrhythmic drug, or placebo. In their experiments, they used the data from two distinct prospective randomized controlled trials available in the PhysioNet repository[36] and found that their CNN model was superior to QTc alone in predicting plasma dofetilide concentration. However, the database used in their study was relatively small (dozens of patients) and they did not use cross-validation in training, with the well-known risk of overfitting. Furthermore, they could not assess the capacity of their artificial intelligence model to detect an arrhythmic risk, or cLQTS, and interpretability of their findings was not performed (*Figure 8*) as done in the present study.

Other studies have focused on CNN modelling of other cardiac diseases using multilead ECG input. For the detection of anterior
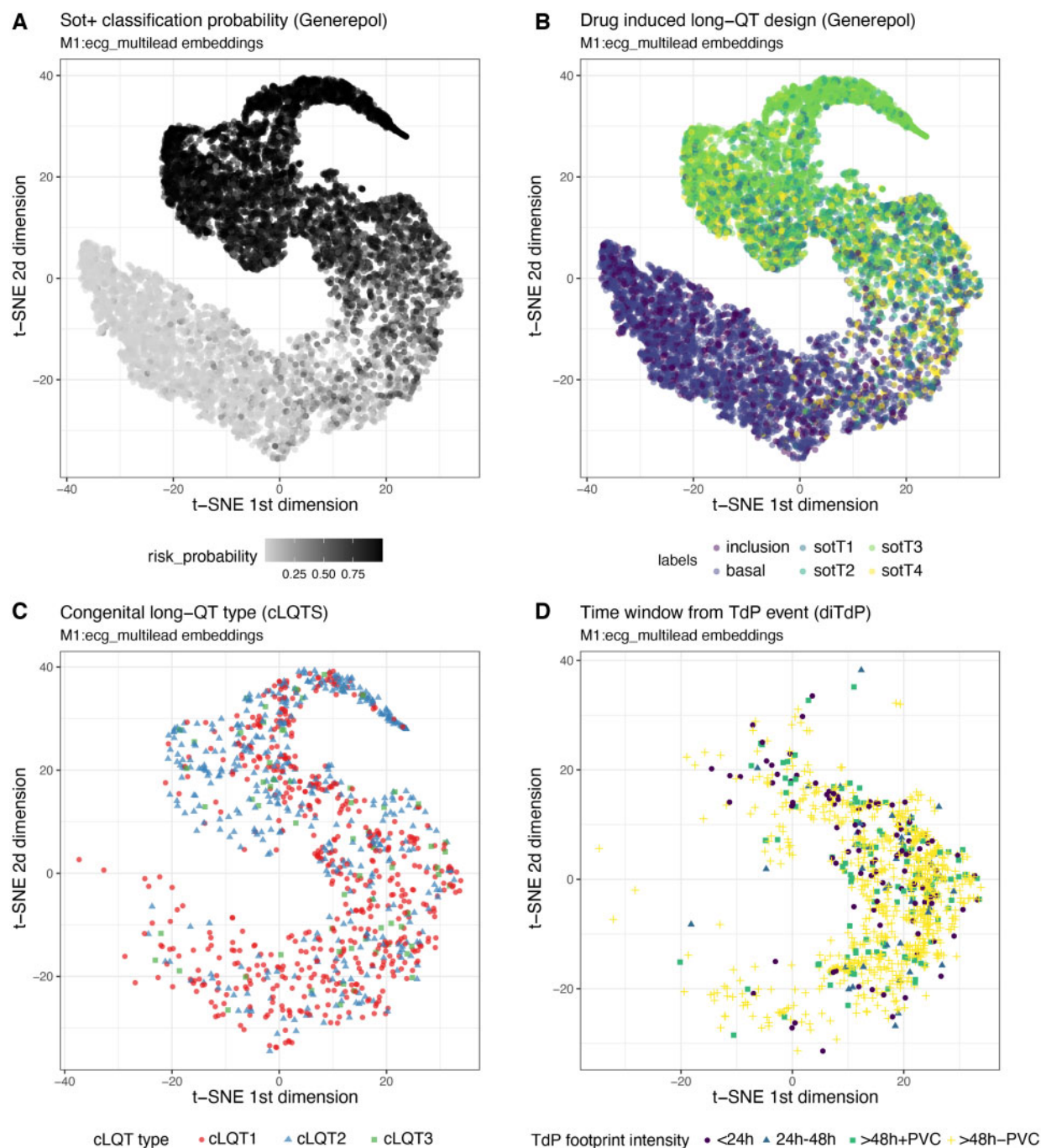
Figure 7 Risk prediction model's deep embeddings reveal clinically relevant data structure. All panels illustrate two dimensions after a t-SNE transformation of the 512 dimensions of the multilead M1 model embeddings (see 'Methods'). (*A*) t-SNE map where each point represents an electrocardiogram from the Generepol cohort. Greyscale indicates the M1: ecg_multilead classification score ranging in [0:1]. (*B*) Same t-SNE map as (*A*) where electrocardiograms are coloured by the experimental setup, from inclusion before and 1, 2, 3, or 4 h after sotalol intake. (*C*) Same t-SNE map with electrocardiogram from the congenital long-QT syndrome cohort. Electrocardiograms are annotated by congenital long-QT syndrome (cLQTS) type. (*D*) Same t-SNE map as (*A*) with electrocardiogram from the drug-induced torsade de pointes (diTdP) cohort of patients having experienced at least one drug-induced torsade de pointes event. Electrocardiograms are coloured by the four groups of torsade de pointes intensity footprint (timeframe from the drug-induced torsade de pointes event and presence/absence of PVCs on the electrocardiogram).
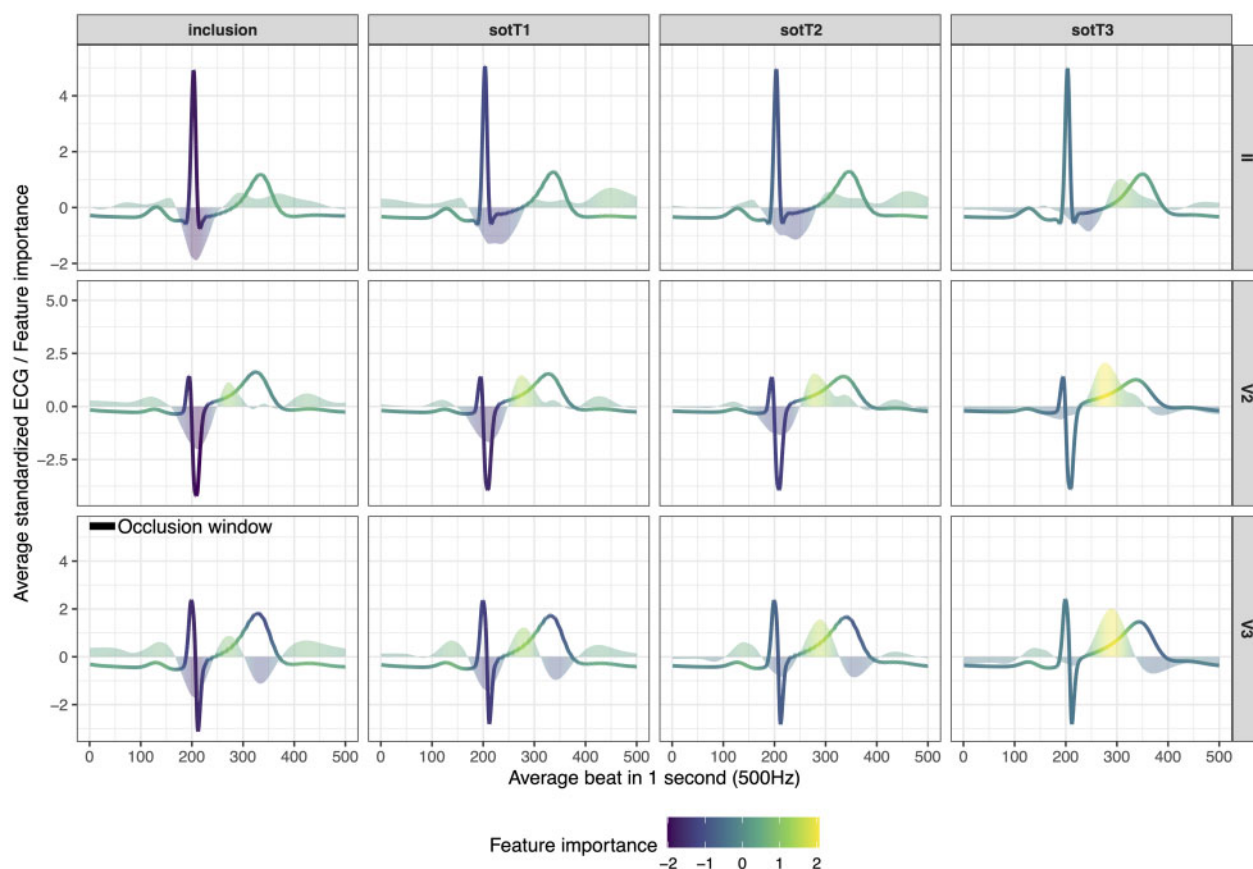
Figure 8 Interpretability of the sotalol footprint on the electrocardiogram signal. This figure displays an averaged signal of the standardized electrocardiogram for each segmented beat for leads LII, V2, and V3. All signals from the same time points were analysed together. Similarly, the standardized feature importance profile is summarized and laid behind the electrocardiogram profile. Colours for both the electrocardiogram and FIP indicate intensity of the feature importance profile.

myocardial infarction,[37] Liu et al. used a 4-lead approach that led to accuracies >90% with a five-fold cross-validation. Tison et al.[38] created a CNN-hidden Markov model that took 12-lead input in order to detect pulmonary arterial hypertension, hypertrophic cardiomyopathy, cardiac amyloid, and mitral valve prolapse. The ROC-AUC was in the 77–94% range for the four conditions. Similar technology was also used by Attia et al.,[39] who applied a CNN model on a large database (>97 000 patients) to detect left ventricular dysfunction. They used a large holdout set (>52 000 patients) and achieved an overall accuracy of 86%. Moreover, a subset of the patients, which were erroneously classified as having ventricular dysfunction, later developed a low ejection fraction, suggesting that the model was able to detect features of this condition before it became clinically diagnosed. Unfortunately, the healthy volunteers from Generepol misclassified by our model as taking sotalol before any intake were not followed, precluding any evaluation of their subsequent risk for TdP and sudden death.

We introduced CNN models trained with data obtained from one lead only. They were as accurate as the multilead model not only when classifying holdout data from the same leads but also from leads on which they were not trained. This is an unexpected result and

indicates that the sotalol footprint is detected by all leads and in similar ways, with the exception of lead V1. Moreover, ECG data were recorded with different acquisition devices and some ECGs, recorded in 250 Hz, were upsized using interpolation techniques. Still, the results were robust, regardless of the recording device. This paves the way to clinical applications where the patients or physicians could record a single electrode ECG, which could then be sent to a centralized server and analysed by the CNN models, with the goal of stratifying the risk for the patient to develop a TdP.

We also explored the CNN models to understand how the decision process was made and what was the model looking for in the ECG to provide a prediction. The occlusion-based interpretability algorithm uncovered the sotalol ECG footprint, which changed with time as blood sotalol concentration increased. The analysis of the footprint was consistent with existing knowledge on how sotalol influences cardiomyocyte action potential, mainly through blockade of I_{Kr} and beta-adrenergic receptor blockade. This approach opens novel avenues of research and applications in the context of drug monitoring for the pharmaceutical industry but also plays an important role in the acceptability of artificial intelligence in clinics. Providing an explanation for the prediction process is increasingly requested

when not mandatory,[40,41] especially for 'black boxes' such as deep neural networks, which train millions of parameters. J-Tpeak features emerged as the main attribute allowing for discrimination of sotalol intake. This is concordant with the emerging literature on the importance of this specific segment when predicting for diTdP beyond QTc.[42]

Lastly, we demonstrated that besides risk prediction, the CNN models learn clinically relevant knowledge. A post hoc analysis of the network's deep embeddings grouped ECGs from the studied cohorts according to their clinical relevance (*Figure 7*). These embeddings can be used to automatically stratify ECG, and ultimately patients, in novel classes that are yet to be characterized. However, identifying the best embeddings can be challenging since the number of model architectures to explore can be very large. More research and training data are needed in the context of translational clinical applications of CNN models for the diagnosis of the different types of cLQTS and prediction of diTdP.

Supplementary material

Supplementary material is available at *European Heart Journal* online.

Funding

Conflict of interest: A.L. reports consultancy and travel grants from SANOFI. A.L., A.P., J.-E.S., E.P., J.-D.Z., C.F.-B., and F.E. report a patent pending for detecting the risk of torsade de pointes (EP19305730.4, WO/2020/245322). G.D. reports a postdoctoral fellowship funding (18SFRN34110369). Other authors have nothing to disclose.

Data availability

Data and models used herein are available from the corresponding authors, upon reasonable request.

References

1. Dessertenne F. Ventricular tachycardia with 2 variable opposing foci. *Arch Mal Coeur Vaiss* 1966;**59**:263–272.
2. Rosso R, Hochstadt A, Viskin D, Chorin E, Schwartz AL, Tovia-Brodie O, Laish-Farkash A, Havakuk O, Gepstein L, Banai S, Viskin S. Polymorphic ventricular tachycardia, ischaemic ventricular fibrillation, and torsade de pointes: importance of the QT and the coupling interval in the differential diagnosis. *Eur Heart J* 2021;**42**:3965–3975.
3. Stramba-Badiale M, Karnad DR, Goulene KM, Panicker GK, Dagradi F, Spazzolini C, Kothari S, Lokhandwala YY, Schwartz PJ. For neonatal ECG screening there is no reason to relinquish old Bazett's correction. *Eur Heart J* 2018;**39**:2888–2895.
4. Roden DM. Drug-induced prolongation of the QT interval. *N Engl J Med* 2004;**350**:1013–1022.
5. Yang T, Chun YW, Stroud DM, Mosley JD, Knollmann BC, Hong C, Roden DM. Screening for acute IKr block is insufficient to detect torsades de pointes liability: role of late sodium current. *Circulation* 2014;**130**:224–234.
6. Schwartz PJ, Ackerman MJ, Antzelevitch C, Bezzina CR, Borggrefe M, Cuneo BF, Wilde AAM. Inherited cardiac arrhythmias. *Nat Rev Dis Primers* 2020;**6**:58.
7. Itoh H, Crotti L, Aiba T, Spazzolini C, Denjoy I, Fressart V, Hayashi K, Nakajima T, Ohno S, Makiyama T, Wu J, Hasegawa K, Mastantuono E, Dagradi F, Pedrazzini M, Yamagishi M, Berthet M, Murakami Y, Shimizu W, Guicheney P, Schwartz PJ, Horie M. The genetics underlying acquired long QT syndrome: impact for genetic screening. *Eur Heart J* 2016;**37**:1456–1464.
8. Salem JE, Yang T, Moslehi JJ, Waintraub X, Gandjbakhch E, Bachelot A, Hidden-Lucet F, Hulot JS, Knollmann BC, Lebrun-Vignes B, Funck-Brentano C, Glazer AM, Roden DM. Androgenic effects on ventricular repolarization: a translational study from the international pharmacovigilance database to iPSC-cardiomyocytes. *Circulation* 2019;**140**:1070–1080.
9. Schwartz PJ, Woosley RL. Predicting the unpredictable: drug-induced QT prolongation and torsades de pointes. *J Am Coll Cardiol* 2016;**67**:1639–1650.
10. Salem JE, Nguyen LS, Moslehi JJ, Ederhy S, Lebrun-Vignes B, Roden DM, Funck-Brentano C, Gougis P. Anticancer drug-induced life-threatening ventricular arrhythmias: a World Health Organization pharmacovigilance study. *Eur Heart J* 2021;**42**:3915–3928.
11. Saque V, Vaglio M, Funck-Brentano C, Kilani M, Bourron O, Hartemann A, Badilini F, Salem JE. Fast, accurate and easy-to-teach QT interval assessment: the triplicate concatenation method. *Arch Cardiovasc Dis* 2017;**110**:475–481.
12. Moss AJ, Zareba W, Benhorin J, Locati EH, Hall WJ, Robinson JL, Schwartz PJ, Towbin JA, Vincent GM, Lehmann MH, Keating MT, MacCluer JW, Timothy KW. ECG T-wave patterns in genetically distinct forms of the hereditary long QT syndrome. *Circulation* 1995;**92**:2929–2934.
13. Schwartz PJ, Ackerman MJ. The long QT syndrome: a transatlantic clinical approach to diagnosis and therapy. *Eur Heart J* 2013;**34**:3109–3116.
14. Salem JE, Bretagne M, Lebrun-Vignes B, Waintraub X, Gandjbakhch E, Hidden-Lucet F, Gougis P, Bachelot A, Funck-Brentano C; French Network of Regional Pharmacovigilance Centres. Clinical characterization of men with long QT syndrome and torsades de pointes associated with hypogonadism: a review and pharmacovigilance study. *Arch Cardiovasc Dis* 2019;**112**:699–712.
15. Salem JE, Germain M, Hulot JS, Voiriot P, Lebourgeois B, Waldura J, Tregouet DA, Charbit B, Funck-Brentano C. GENomE wide analysis of sotalol-induced IKr inhibition during ventricular REPOLarization. "Generepol Study": lack of common variants with large effect sizes. *PLoS One* 2017;**12**:e0181875.
16. U.S. Department of Health and Human Services Food and Drug Administration. *E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs—Questions and Answers (R3)*. Guidance for Industry; 2017. https://www.fda.gov/files/drugs/published/E14-Clinical-Evaluation-of-QT-QTc-Interval-Prolongation-and-Proarrhythmic-Potential-for-Non-Antiarrhythmic-Drugs-Questions-and-Answers-%28R3%29-Guidance-for-Industry.pdf (11 August 2021).
17. 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.
18. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med* 2016;**375**:1216–1219.
19. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;**521**:436–444.
20. Hinton G. Deep learning—a technology with the potential to transform health care. *JAMA* 2018;**320**:1101–1102.
21. Bos JM, Attia ZI, Albert DE, Noseworthy PA, Friedman PA, Ackerman MJ. Use of artificial intelligence and deep neural networks in evaluation of patients with electrocardiographically concealed long QT Syndrome from the surface 12-lead electrocardiogram. *JAMA Cardiol* 2021;**6**:532–538.
22. Salem JE, Dureau P, Bachelot A, Germain M, Voiriot P, Lebourgeois B, Tregouet DA, Hulot JS, Funck-Brentano C. Association of oral contraceptives with drug-induced QT interval prolongation in healthy nonmenopausal women. *JAMA Cardiol* 2018;**3**:877–882.
23. Extramiana F, Badilini F, Sarapa N, Leenhardt A, Maison-Blanche P. Contrasting time- and rate-based approaches for the assessment of drug-induced QT changes. *J Clin Pharmacol* 2007;**47**:1129–1137.
24. Sarapa N, Morganroth J, Couderc JP, Francom SF, Darpo B, Fleishaker JC, McEnroe JD, Chen WT, Zareba W, Moss AJ. Electrocardiographic identification of drug-induced QT prolongation: assessment by different recording and measurement methods. *Ann Noninvasive Electrocardiol* 2004;**9**:48–57.
25. Extramiana F, Badilini F, Denjoy I, Vaglio M, Green CL, Kligfield P, Leenhardt A, Maison-Blanche P. Sex influences on ventricular repolarization duration in normal subjects and in type 1, 2 and 3 long QT syndrome patients: different effect in acquired and congenital type 2 LQTS. *J Electrocardiol* 2020;**62**:148–154.
26. Baillet S, Friston K, Oostenveld R. Academic software applications for electromagnetic brain mapping using MEG and EEG. *Comput Intell Neurosci* 2011;**2011**:972050.
27. Su C, Tong J, Zhu Y, Cui P, Wang F. Network embedding in biomedical data science. *Brief Bioinform* 2020;**21**:182–197.
28. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008;**9**:2579–2605.
29. Nguyen LS, Dolladille C, Drici MD, Fenioux C, Alexandre J, Mira JP, Moslehi JJ, Roden DM, Funck-Brentano C, Salem JE. Cardiovascular toxicities associated

with hydroxychloroquine and azithromycin: an analysis of the World Health Organization Pharmacovigilance Database. *Circulation* 2020;**142**:303–305.

30. Salem JE, Alexandre J, Bachelot A, Funck-Brentano C. Influence of steroid hormones on ventricular repolarization. *Pharmacol Ther* 2016;**167**:38–47.

31. Hondeghem LM. Drug-induced QT prolongation and torsades de pointes: an all-exclusive relationship or time for an amicable separation? *Drug Saf* 2018;**41**: 11–17.

32. Funck-Brentano C. Pharmacokinetic and pharmacodynamic profiles of d-sotalol and d,l-sotalol. *Eur Heart J* 1993;**14**:30–35.

33. Haverkamp W, Breithardt G, Camm AJ, Janse MJ, Rosen MR, Antzelevitch C, Escande D, Franz M, Malik M, Moss A, Shah R. The potential for QT prolongation and pro-arrhythmia by non-anti-arrhythmic drugs: clinical and regulatory implications. Report on a Policy Conference of the European Society of Cardiology. *Cardiovasc Res* 2000;**47**:219–233.

34. Porta-Sanchez A, Spillane DR, Harris L, Xue J, Dorsey P, Care M, Chauhan V, Gollob MH, Spears DA. T-wave morphology analysis in congenital long QT syndrome discriminates patients from healthy individuals. *JACC Clin Electrophysiol* 2017;**3**:374–381.

35. Attia ZI, Sugrue A, Asirvatham SJ, Ackerman MJ, Kapa S, Friedman PA, Noseworthy PA. Noninvasive assessment of dofetilide plasma concentration using a deep learning (neural network) analysis of the surface electrocardiogram: a proof of concept study. *PLoS One* 2018;**13**:e0201059.

36. Moody GB, Mark RG, Goldberger AL. PhysioNet: a research resource for studies of complex physiologic and biomedical signals. *Comput Cardiol* 2000;**27**: 179–182.

37. Liu W, Zhang M, Zhang Y, Liao Y, Huang Q, Chang S, Wang H, He J. Real-time multilead convolutional neural network for myocardial infarction detection. *IEEE J Biomed Health Informatics* 2018;**22**:1434–1444.

38. Tison GH, Zhang J, Delling FN, Deo RC. Automated and interpretable patient ECG profiles for disease detection, tracking, and discovery. *Circ Cardiovasc Qual Outcomes* 2019;**12**:e005289.

39. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA, Munger TM, Asirvatham SJ, Scott CG, Carter RE, Friedman PA. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med* 2019;**25**:70–74.

40. Martens D, Vanthienen J, Verbeke W, Baesens B. Performance of classification models from a user perspective. *Decis Support Syst* 2011;**51**:782–793.

41. Goodman B, Flaxman S. Union regulations on algorithmic decision-making and a "right to explanation". *AI Mag* 2017;**38**:50–57.

42. Vicente J, Zusterzeel R, Johannesen L, Mason J, Sager P, Patel V, Matta MK, Li Z, Liu J, Garnett C, Stockbridge N, Zineh I, Strauss DG. Mechanistic model-informed proarrhythmic risk assessment of drugs: review of the "CiPA" initiative and design of a prospective clinical validation study. *Clin Pharmacol Ther* 2018; **103**:54–66.