Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200150

Emerging Concepts and Applied Machine Learning Research in Patients with Drug-Induced Repolarization Disorders

Mina BJELOGRLIC[a,1], Arnaud ROBERT[a], Arnaud MIRIBEL, Mehdi NAMDAR[b], Baris GENCER[b,c], Christian LOVIS[a] and François GIRARDIN[d]

[a] *Division of Medical Information Sciences, University Hospitals of Geneva and University of Geneva, Geneva, Switzerland*

[b] *Cardiology Division, Department of Medicine, University Hospitals of Geneva and University of Geneva, Geneva, Switzerland*

[c] *Brigham and Women's Hospitals, TIMI study group, Harvard Medical School, Boston, United States of America*

[d] *Division of Clinical Pharmacology and Toxicology; Medical Direction, University Hospitals of Geneva and University of Geneva, Geneva, Switzerland*

Abstract. The paper presents a review of current research to develop predictive models for automated detection of drug-induced repolarization disorders and shows a feasibility study for developing machine learning tools trained on massive multimodal datasets of narrative, textual and electrocardiographic records. The goal is to reduce drug-induced long QT and associated complications (Torsades-de-Pointes, sudden cardiac death), by identifying prescription patterns with pro-arrhythmic propensity using a validated electronic application for the detection of adverse drug events with data mining and natural language processing; and to compute individual-based predictive scores in order to further identify clinical conditions, concomitant diseases, or other variables that correlate with higher risk of pro-arrhythmic situations.

Keywords. Machine Learning, Clinical Decision Support System, Pharmacovigilance, Adverse Drug Events, Electrocardiography, Long QT, Torsades-de-Pointes, Repolarization Disorders, Analytic-Decision Modelling.

1. Introduction

Drug-induced ventricular arrhythmia is a major cause of iatrogenic morbidity and mortality [1], particularly in circumstances, where a wide range of psychotropic and illicit drugs are consumed (e.g. people who inject drugs, living in detention or with mental illness). Electrocardiographic (ECG) screening is a widely available method for detecting repolarization abnormalities. Long QT (LQT), defined as a QT interval duration of >470 ms for men and >480 ms for women, is the main surrogate marker for Torsades-de-Pointes (TdP), a pathognomonic, drug-induced polymorphic ventricular arrhythmia [2]. The probability to develop a TdP increases with the extent of the

[1]Mina Bjelogrlic, PhD, Division of Medical Information Sciences, University Hospitals of Geneva and University of Geneva, Geneva, Switzerland. E-mail address: mina.bjelogrlic@hcuge.ch.

corrected QT interval (QTc) prolongation. Most TdP are transient and unsustained, but can degenerate to malignant ventricular fibrillation (VF) and sudden cardiac death (SCD).

Machine Learning (ML) based techniques are expected to support clinicians by allowing more accurate detection of abnormalities that would be missed using traditional diagnostic criteria (QT duration, axis, visual T-wave morphology). In the past decade, algorithms for pattern recognition and classification using learning techniques have become interesting in health applications for two main reasons; firstly, the digital information stored in Electronic Health Records (EHRs) is massively increasing [3], and there are available open databases, such as the MIT-BIH Arrhythmia Database [4]; secondly, the performance of ML algorithms have reached a point where either the task is better achieved by automatic systems, or further hidden underlying physiological phenomena can be re-addressed from a new analytical perspective. Recent research groups have tackled the problem of ECG-based heartbeat classification in order to detect arrhythmias [5]. ECG-based ML approaches include several steps after the acquisition of the ECG signals; preprocessing, segmentation, feature extraction, selection, and classification with learning algorithms [6].

2. Scope

Virtually all medications inducing LQT act by blocking the outward IKr current, which is mediated by the potassium channel encoded by the KCNH2 gene [7].The measured QT value is further corrected according to a calculation based on heart rate (most frequently Bazett and Fridericia formulas), i.e. the QTc. Long QTc interval is a sensitive marker for TdP, but it is not specific since drug-induced QTc. More importantly, the lack of LQT specificity causes the discontinuation of valuable patients' treatment and new drug development [8]. Separate analysis of J-Tpeak and Tpeak-Tend intervals helps to identify medications which prolong the QTc interval with counterbalancing inward current block and limited risk of arrhythmias. Automated measurement methodology and algorithm for the evaluation of the early repolarization J-Tpeak and late repolarization Tpeak-Tend intervals were proposed to assess drug-induced proarrhythmia propensity, using a two-step T-wave delineation method [9]. A wide range of illicit substances and medications are associated with repolarization abnormalities. In psychiatry, risk factors for developing drug-induced TdP include chronic hepatitis C infection, HIV, electrolyte abnormalities, renal or hepatic dysfunction, pre-existing coronary heart disease, treatment with more than one QT-prolonging drug, older age, female sex, and genetic predisposition [10,11]. Abnormal electrocardiographic (ECG) recordings are reported in 27.3% of psychiatric inpatients at hospital admission and repolarization disorders were the most prevalent abnormal findings (11.8%).

3. Methodology

In the University Hospitals of Geneva (HUG), since 2017, all ECG devices (n=220) have been replaced by the same electrocardiographic devices (Philipps TC-70) and all records are automatically converted into a DICOM® [12] file (>8'000 ECGs per month performed in all care settings at HUG) and stored in the Picture Archiving Communication System (PACS) [13]. Computerized ECG health records with structured metadata in XML format are available to be extracted from the hospital DPI (Dossier

Patient Intégré) [14] secured servers for research purposes. All records can be analyzed with socio-demographic, clinical, laboratory, genetic, and drug prescription data at the individual patient level, using an agreed protocol and database structure. The latter cardiac information combined with electronic prescription systems and natural language processing (NLP), allow ECG feature extraction and drug imputability assessment, based on interactions, dose variations, and chronological relationship with drug exposition. Moreover, it stores all measurements from the recordings, meaning it includes boundary detection, epoch extraction, and characteristic measurements for each lead. In the proposed methodology (see **Figure 1**), electrographic, structured, and unstructured text data is collected and stored in a multimodal database, which will first pass a quality control, and then be used to extract and select relevant features to build models for RD prediction.

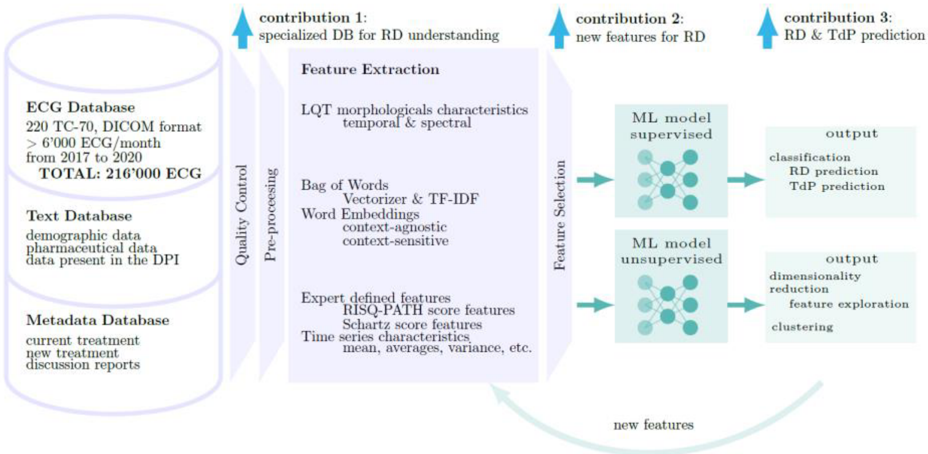

Figure 1. Research Methodology. Database (DB), Repolarization Disorders (RD), LQT (Long QT), Torsades de Pointe (TdP), Machine Learning (ML).

A feature set proposed in the literature include ECG-intervals, morphological (e.g. amplitude), Vector Cardiographs (VCG) [15], Temporal VCG (TVCG) [16], different transforms, including temporal polynomial fitting with Hermite Basis decompositions [17], and different versions of time-frequency representations, such as Discrete Wavelet Transforms (DWT) [18]. The classification, according to the resulting feature vectors is done using classifiers such as linear discriminant (LD), support vector machines (SVM) [19], Random Forests [20], and ensemble classifiers with the advantage of bagging both temporal and morphological feature-based classifiers [21]. Long-short-term memory (LSTM) networks, combined with a wavelet sequence (WS) [22], and algorithms with deep learning (DL) architectures [23] have been investigated. For inter-patient classification convolutional neural networks (CNN) and recurrent neural networks (RNN) give very promising results [24].

4. Feasibility study

There are no a priori exclusion criteria. Nonetheless, ECG of insufficient quality (not present QT measurements) or with inverted electrodes will not be included. With a minimum 70'000 ECG recorded and saved in the hospital's database (PACS), the estimated amount of available ECG over the target period (three years) is 200'000 EC. **Figure 2** shows estimations, over a four week period, of QT density distributions depending on age and sex.

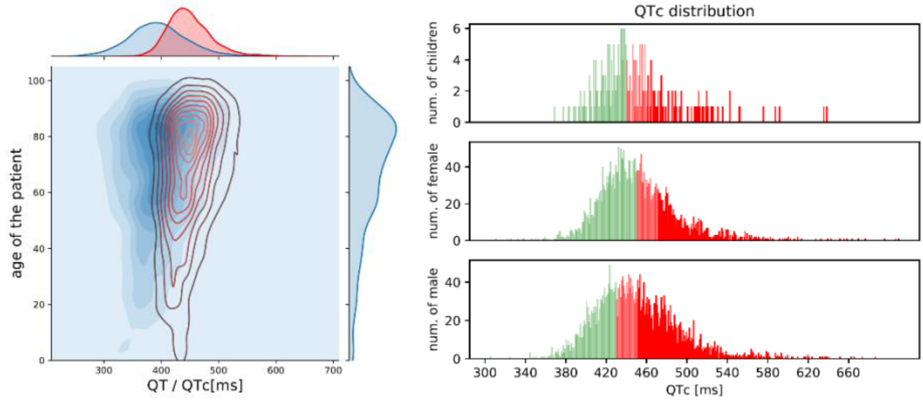

Figure 2. Left: density distributions estimation (based on a four-week period in 2019) of patient age and QT/QTc (Bazett's formula, in red) values of patients at HUG. Right: QTc (Bazett's formula) distributions over a four-week period in 2019 at HUG. For children (under the age of 15 years), women (above 15 years old, female sex), and men (above 15 years old, male sex). From green to dark red; normal, intermediate, and long QTc [ms].

5. Discussion

Approximately 30% of the inpatients are taking QT prolonging drugs, out of which 50% will have additional QT prolonging drugs [25]. In our preliminary analysis of ECG structured data (early 2019), RD prevalence and QT interval variability increased with patient age and the number of medications. The proposed research will exploit large databases to answer the central question of the influence of drugs on RD, particularly LQT and related severe complications. This question is of major importance since QT prolongation may lead to iatrogenic life-threatening arrhythmia, whose indicators could be detected to prevent consequences by instant electronic alert systems. ML-based approach could bring about innovation opportunities and new dimensions to comprehend iatrogenicity. Applied ML will influence decision-making processes (at individual and more global levels, such as health authorities), education, and training (including general practitioners, cardiologists, and medicine students). Additionally, CDSS can potentially decrease the number of misdiagnosed patients, the extra costs of non-suitable hospitalizations (40% of patients oriented to specialists for LQTS are misdiagnosed) [26], and prevent patients from severe cardiac events, such as arrhythmias or, possibly, sudden cardiac death.

References

[1] Schwartz PJ, Woosley RL. Predicting the Unpredictable. J Am Coll Cardiol 2016 Apr;67(13):1639–1650. [doi: 10.1016/j.jacc.2015.12.063]

[2] Beach SR, Celano CM, Sugrue AM, Adams C, Ackerman MJ, Noseworthy PA, Huffman JC. QT Prolongation, Torsades de Pointes, and Psychotropic Medications: A 5-Year Update. Psychosomatics 2018 Mar;59(2):105–122. [doi: 10.1016/j.psym.2017.10.009]
[3] Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE J Biomed Health Inform 2018 Sep;22(5):1589–1604. [doi: 10/gddkw8]
[4] Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng Med Biol Mag 2001 Jun;20(3):45–50. [doi: 10/c5zzg6]
[5] Luz EJ da S, Schwartz WR, Cámara-Chávez G, Menotti D. ECG-based heartbeat classification for arrhythmia detection: A survey. Comput Methods Programs Biomed 2016 Apr;127:144–164. [doi: 10/f8fq6x]
[6] Llamedo M, Martínez JP. Heartbeat Classification Using Feature Selection Driven by Database Generalization Criteria. IEEE Trans Biomed Eng 2011 Mar;58(3):616–625. [doi: 10/d32zzt]
[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 May 7;37(18):1456–1464. [doi: 10.1093/eurheartj/ehv695]
[8] Haverkamp W. The potential for QT prolongation and proarrhythmia by non-antiarrhythmic drugs: clinical and regulatory implications. Report on a Policy Conference of the European Society of Cardiology. Eur Heart J 2000 Aug 1;21(15):1216–1231. [doi: 10.1053/euhj.2000.2249]
[9] Johannesen L, Vicente J, Hosseini M, Strauss DG. Automated Algorithm for J-Tpeak and Tpeak-Tend Assessment of Drug-Induced Proarrhythmia Risk. Rasmusson RL, editor. PLOS ONE 2016 Dec 30;11(12):e0166925. [doi: 10/f9mn7f]
[10] Girardin F, Sztajzel J. Cardiac adverse reactions associated with psychotropic drugs. Dialogues Clin Neurosci 2007 Mar;9(1):92–95.
[11] Girardin F, Gex-Fabry M, Berney P, Shah D, Gaspoz J-M, Dayer P. Drug-Induced Long QT in Adult Psychiatric Inpatients: The 5-Year Cross-Sectional ECG Screening Outcome in Psychiatry Study. Am J Psychiatry 2013 Dec;170(12):1468–1476. [doi: 10.1176/appi.ajp.2013.12060860]
[12] DICOM specification website [Internet]. Available from: https://www.dicomstandard.org/
[13] Rosset C, Rosset A, Ratib O. General Consumer Communication Tools for Improved Image Management and Communication in Medicine. J Digit Imaging 2005 Dec;18(4):270–279. [doi: 10.1007/s10278-005-6703-2]
[14] Lovis C, Baud RH, Planche P. Power of expression in the electronic patient record: structured data or narrative text? Int J Med Inf 2000;58–59:101–110. [doi: 10/fnmvqb]
[15] Soria ML, Martínez J. Analysis of multidomain features for ECG classification. 2009 36th Annu Comput Cardiol Conf CinC 2009. p. 561–564.
[16] Garcia G, Moreira G, Menotti D, Luz E. Inter-Patient ECG Heartbeat Classification with Temporal VCG Optimized by PSO. Sci Rep 2017 Sep 5;7(1):10543. [doi: 10/gbwznc]
[17] Li BN, Dong MC, Vai MI. Modelling cardiovascular physiological signals using adaptive Hermite and wavelet basis functions. IET Signal Process 2010;4(5):588. [doi: 10/bmr8fv]
[18] Daamouche A, Hamami L, Alajlan N, Melgani F. A wavelet optimization approach for ECG signal classification. Biomed Signal Process Control 2012 Jul;7(4):342–349. [doi: 10/dmp2db]
[19] Kanaan L, Merheb D, Kallas M, Francis C, Amoud H, Honeine P. PCA and KPCA of ECG signals with binary SVM classification. 2011 IEEE Workshop Signal Process Syst SiPS [Internet] Beirut, Lebanon: IEEE; 2011 [cited 2019 Mar 4]. p. 344–348. [doi: 10/fdppnv]
[20] Li T, Zhou M. ECG Classification Using Wavelet Packet Entropy and Random Forests. Entropy 2016 Aug 5;18(8):285. [doi: 10/f822bk]
[21] Mondéjar-Guerra V, Novo J, Rouco J, Penedo MG, Ortega M. Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers. Biomed Signal Process Control 2019 Jan;47:41–48. [doi: 10/gfxjrr]
[22] Yildirim Ö. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Comput Biol Med 2018 May;96:189–202. [doi: 10/gc8srm]
[23] Pyakillya B, Kazachenko N, Mikhailovsky N. Deep Learning for ECG Classification. J Phys Conf Ser 2017 Oct;913:012004. [doi: 10/gfxjrk]
[24] Guo L, Sim G, Matuszewski B. Inter-Patient ECG Classification with Convolutional and Recurrent Neural Networks. ArXiv181004121 Eess Q-Bio 2018 Sep 27;11.
[25] Zürcher DJ-P, Schlaepfer J, Waeber G, Pasquier M. Le QT long acquis. Rev Médicale Suisse 2013;9:1538–42.
[26] Taggart NW, Haglund CM, Tester DJ, Ackerman MJ. Diagnostic miscues in congenital long-QT syndrome. Circulation 2007 May 22;115(20):2613–2620. PMID:17502575