



Published in final edited form as:

J Cardiovasc Pharmacol Ther. 2021 July ; 26(4): 335–340. doi:10.1177/1074248421995348.

Prediction of Drug-Induced Long QT Syndrome Using Machine Learning Applied to Harmonized Electronic Health Record Data

Steven T. Simon, MD¹, Divneet Mandair, MD², Premanand Tiwari, MS³, Michael A. Rosenberg, MD^{1,3}

¹Division of Cardiology, University of Colorado School of Medicine, Aurora, CO, USA

²Department of Medicine, University of Colorado School of Medicine, Aurora, CO, USA

³Colorado Center for Personalized Medicine, University of Colorado School of Medicine, Aurora, CO, USA

Abstract

Background: Drug-induced QT prolongation is a potentially preventable cause of morbidity and mortality, however there are no widespread clinical tools utilized to predict which individuals are at greatest risk. Machine learning (ML) algorithms may provide a method for identifying these individuals, and could be automated to directly alert providers in real time.

Objective: This study applies ML techniques to electronic health record (EHR) data to identify an integrated risk-prediction model that can be deployed to predict risk of drug-induced QT prolongation.

Methods: We examined harmonized data from the UCHHealth EHR and identified inpatients who had received a medication known to prolong the QT interval. Using a binary outcome of the development of a QTc interval >500 ms within 24 hours of medication initiation or no ECG with a QTc interval >500 ms, we compared multiple machine learning methods by classification accuracy and performed calibration and rescaling of the final model.

Results: We identified 35,639 inpatients who received a known QT-prolonging medication and an ECG performed within 24 hours of administration. Of those, 4,558 patients developed a QTc > 500 ms and 31,081 patients did not. A deep neural network with random oversampling of controls was found to provide superior classification accuracy (F1 score 0.404; AUC 0.71) for the development of a long QT interval compared with other methods. The optimal cutpoint for prediction was determined and was reasonably accurate (sensitivity 71%; specificity 73%).

Corresponding Author: Steven T. Simon, Division of Cardiology, University of Colorado School of Medicine, 12631 E. 17th Avenue, Mail Stop B130 Aurora, CO 80045, USA. steven.simon@cuanschutz.edu.

Author Contributions

STS led the construction and writing of the manuscript. MR and PT constructed the model and performed the analysis. All authors provided feedback to the research methods and review of the manuscript.

Supplemental Material

Supplemental material for this article is available online.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Conclusions: We found that deep neural networks applied to EHR data provide reasonable prediction of which individuals are most susceptible to drug-induced QT prolongation. Future studies are needed to validate this model in novel EHRs and within the physician order entry system to assess the ability to improve patient safety.

Keywords

drug-induced long QT syndrome; machine learning; electronic health record

Introduction

Drug-induced long QT syndrome (diLQTS) is characterized by marked prolongation of the QT interval on an electrocardiogram (ECG) following drug administration. This prolongation of the QT interval increases the risk of developing Torsades de Pointes (TdP), a form of polymorphic ventricular tachycardia which can be fatal.¹ The 99th percentile for the QTc interval is 480 milliseconds (ms) in adult women and 470 ms in adult men² and the risk of TdP significantly increases when the QTc interval exceeds 500 ms.²⁻⁸ There have been numerous cardiac and noncardiac medications associated with QT prolongation and TdP, and it is the most common cause for withdrawal of medications on the market as well as failure to make it out of the drug development pipeline.⁹⁻¹¹ Recently, this condition has gained media notoriety given that at least 2 of the medications utilized early in the treatment of COVID-19, hydroxychloroquine and azithromycin, can cause QT prolongation.¹² Given that diLQTS is a preventable cause of morbidity and mortality, an accurate prediction model is needed to stratify an individual's risk of developing this condition. In most cases, alternative therapies could be considered for an individual identified as high risk.¹³ In cases where a medication may be unavoidable,¹⁴ knowing an individual's risk could be used to guide monitoring.¹⁵ A system for stratifying patients by risk of diLQTS^{16,17} would improve safety for the use of known QT-prolonging medications.

Machine-learning methods can leverage big data sources to predict various cardiovascular and non-cardiovascular outcomes. Many standard machine-learning techniques, such as support vector machines,¹⁸ random forests,¹⁹ and boosted/ bagged decision trees,²⁰ have successfully predicted clinical outcomes; however, recently deep-learning models have demonstrated superior results.^{19,21-25} Deep learning methods are composed of multiple hierarchical layers of neural networks which allows improved learning potential and power but requires a larger amount of data. Data from the electronic health record (EHR) is increasingly a source for the development of prediction models using machine learning methods. Consolidation of EHR data into harmonized databases allows deep learning methods to be utilized for clinical applications and allows a model created in one EHR to be directly deployed in others.^{17,26-29} Furthermore, a model built directly off of EHR data can be directly inserted into a physician order entry (POE) system to alert a provider of increased risk in real time.

In this study we present a method for risk stratification using machine learning techniques applied to harmonized EHR data. We hypothesize that deep machine-learning methods can

reliably identify patients at risk for diLQTS. Ultimately, we aim to develop a system that can be integrated into EHR systems to alert providers of this risk.

Methods

Data Source and Study Population

We examined data from the UC Health hospital system, an integrated academic healthcare system that includes 3 main regional centers (North, Central, South) across the Colorado Front Range that share a single Epic instance. Data were gathered from 2003 through November, 2018, and were queried using Google BigQuery to create a dataset and conduct analyses directly on the Google cloud platform.

Inpatient encounters were selected in which the patient was administered a known QT-prolonging medication (Supplementary Table 1),³⁰ and had an ECG performed within 24 hours of medication administration.

Case Selection

The primary outcome was the presence or absence of a QTc interval of greater than 500 ms. Cases were identified as those encounters in which the QTc by ECG exceeded 500ms^{31,32} and controls were those in which no ECG had a QTc > 500 ms throughout the encounter, after having received a known QT-prolonging medication. These designations (case/control) were the labels used for training and testing the learning models. ECGs with conduction disease in the form of bundle branch block, intraventricular conduction disease, or ventricular pacing, defined by a QRS duration of greater than 120 ms, were excluded.

Model Development and Data Analysis

We used a common data model for EHR data, based on the Observational Health Data Sciences and Informatics (OHDSI) collaboration, which uses the Observation Medical Outcomes Partnership common data model (OMOP-CDM) which is a mapping of the raw EHR data to a harmonized dataset.²⁸ 6,458 variables were included in the model which consisted of medications, procedure codes, diagnosis codes, labs, and demographic data. There were no missing variables for continuous variables, and the remaining data was coded as a binary variable with no history of that variable being coded as a zero.

After label assignment, data was split into a training (80%) and testing set (20%). The training set was used to develop and compare deep learning models using 5-fold cross-validation, to allow comparison of models and tuning of hyperparameters.

Given the relative infrequency of the outcome, there was imbalance between the cases and controls. Re-sampling techniques can be used to improve this balance and avoid the over-assigning of new cases to the majority class.²⁶ Both oversampling techniques (increase frequency of cases relative to controls) and undersampling techniques (reduce frequency of controls used) were employed. These techniques included SMOTE,³³ ADASYN, random oversampling, and random undersampling. The random oversampling technique generated 21,100 additional samples and the random undersampling discarded 21,100 samples in order to match the majority and minority class.

Once an optimal resampling approach was selected, several classification algorithms (naïve Bayesian classification, regularized logistic regression, random forest classification, deep neural networks) were compared. The deep neural network model contained 6 hidden layers. The first hidden layer used 1024 neurons, and layers 2-6 had 512 neurons each. The activation function used was sigmoid for all of the neurons on the hidden layer and 50% dropout at each hidden layer. Models were selected based on discrimination using the F₁-score,³⁴ which measures performance on imbalanced datasets more effectively than alternative metrics such as classification accuracy and the receiver operating characteristic³⁵ (ROC), which was used for secondary comparison.

We also examined calibration of the optimal model for discrimination using calibration curves, binned by deciles of predicted probability, displaying frequency of events within each bin. To perform Platt rescaling, we performed logistic regression using 50% of the testing set (random sampling) with predicted probability from the optimal model on the actual events, and then used the predicted coefficients to calculate a “rescaled” risk probability on the other 50% of the testing set. We used Youden’s J method to identify an optimal cutoff for the rescaled prediction model,³⁶ and then examined sensitivity and specificity curves (as well as false positive and false negative) over various cutoffs, with attention to identifying an optimal cutoff to fit the clinical application.

All analyses were run on Google Cloud Platform, using 96 CPUs and 620 GB of RAM. Scripts were composed in Python (version 3) and were run on Jupyter Notebook with Tensorflow platform on the Google Cloud Platform. Machine learning packages included scikit-learn and keras.

Results

Across the entire dataset, 35639 inpatients were identified who received a known QT-prolonging medication and had an ECG within 24 hours following administration. Of those, 4558 (12.8%) patients developed a QTc > 500 ms by ECG, while the QTc of the remaining 31081 (87.2%) patients remained < 500 ms for the duration of the encounter. The age and gender did not significantly differ between the groups, and the prevalence of cardiovascular risk factors for each group is displayed in Table 1.

We examined various re-sampling methods to address the substantial degree of imbalance in the dataset of cases compared to controls. We found that random oversampling had the overall best performance by F1 score and AUC (F1 0.404; AUC 0.71) as compared with other re-sampling methods including SMOTE, random undersampling, and ADASYN (Table 2).

Using random oversampling on the training data, we tested the accuracy of various classification models on the held-out dataset. A deep neural network provided superior classification for development of a long QT interval compared with other methods with an F1 score of 0.404 and an AUC of 0.71 (Figure 1). The F1 Score is a measure of the weighted average of precision and recall which takes into account false positives and false negatives

and is a more useful measurement with uneven distributions. Prediction accuracy for other methods is listed in Table 3.

In order to rescale predictions, we performed logistic regression of predicted risk of diLQTS on actual outcomes in a 50% split sample from the testing set, in which the model parameters were β_0 (intercept) -3.7 ± 0.1 , β_1 4.1 ± 0.2 , pseudo- R^2 0.15, $P < .0001$, and then applied the predicted probabilities from that model to the predictions in the other 50% sample for Platt rescaling. After recalibration, no subject had over 0.5 predicted probability of diLQTS, with overall better calibration than the original model (Figure 2). We then applied Youden's J method to identify the optimal cutoff probability for classification to both the original and resampled probabilities, and found that while rescaling had decreased the optimal cutoff from 0.43 to 0.14, there was no difference in the discrimination based on AUC (0.71 for both). When applied to the testing set, this cutpoint yielded a sensitivity of 71%, specificity 73%, false positive rate 28%, and false negative rate 30% (Figure 3).

Discussion

In this analysis, we examined the prediction of drug-induced QT prolongation using harmonized EHR data. We found that a deep neural network with random oversampling had the most accurate classification performance and overall performance of the model was reasonable.

A frequent challenge encountered in prediction analysis of healthcare data is that of an imbalanced class distribution. In this case, the minority outcome (development of a prolonged QTc interval), was much less frequent than the majority outcome (no significant increase in QTc interval). This imbalance can become a problem when applying machine learning methods as some assume a relatively equal balance between classes. This assumption leads to a tendency to favor the majority class when applied to an imbalanced dataset, which can decrease the accuracy in classifying minority occurrences.³⁷ Because this is a frequently encountered scenario, prior groups have studied various rebalancing strategies.^{37,38} Previous analysis has shown an improvement in prediction accuracy when applying rebalancing methods, as long as the dataset is sufficient in size.³⁹ Resampling techniques including random oversampling of the minority class or random undersampling of the majority class each have drawbacks. In the case of oversampling the minority class, this can lead to an overfit model, while in the case of undersampling the majority class, there is an associated loss of information.⁴⁰ In our dataset, rebalancing using over-sampling was likely the superior method given the increase in power obtained with this method. However, no one technique will be optimal for all datasets and it remains essential to evaluate multiple rebalancing methods for any given problem.

We found the prediction classification using a deep neural network to be reasonably accurate for classifying individuals at risk of developing diLQTS and superior to alternative classification methods. Deep learning methods combine multiple layers of neurons allowing for nonlinear relationships to be modeled.⁴¹ These methods have recently been found to have superior prediction capabilities and have been applied to the prediction of cardiovascular outcomes and cardiovascular imaging analysis.^{42,43} Deep learning techniques

tend to perform better than other techniques as the amount of data increases which is likely why it had superior performance when applied to our large dataset. One of the major disadvantages of using a deep learning model is the black-box nature of the method and therefore an inability to determine individual predictor variables for the outcome. Therefore, this analysis does not inform the identification or modification of risk factors for developing drug induced QT prolongation.

We developed our model directly from harmonized EHR data (OMOP common data model). Machine learning models are notoriously biased for the population from which they are derived, and using a common data model allows for direct application and validation in a separate EHR to examine bias. Once verified, this model can also be directly incorporated into the EHR as a clinical decision support tool. Based on a given risk probability for the development of a prolonged QT interval, patients can be assigned a risk score. This information is stored on the EHR backend and is hidden from users but can be incorporated into front-end logic in the form of a best-practice advisory (BPA). Identifying individuals at increased risk could sway providers to choose an alternative therapy not known to cause QT prolongation when possible, or to appropriately monitor patients who must receive a QT prolonging agent.

Prior investigators have developed risk scores for predicting diLQTS, but this is the first to our knowledge to utilize deep learning methods applied to EHR data.^{44,45} Prior studies have typically focused on a specific patient population, while this study examined all inpatient encounters regardless of reason for admission or hospital location. Other groups have used machine learning techniques in the prediction of long QT, but have focused on the drug molecules rather than individual factors.⁴⁶

Calibration measures how accurately the model's predicted probability matches the observed probability of the outcome. Prior to recalibration, our model showed generally poor calibration, with a pattern indicative of overfitting; however, after Platt rescaling, the model had better calibration, and showed satisfactory discrimination. Further evaluation of sensitivity/specificity curves identified other potential cutoffs for screening for diLQTS, depending on the overall goal and follow-up options (Figure 3). By choosing different cutpoints, we can improve the sensitivity of our prediction at the expense of specificity, or vice-versa, which can be an important clinical tool. Currently, patients who are initiating certain antiarrhythmic agents including Dofetilide and Sotalol are recommended to undergo direct observation in the hospital. When trying to determine who may not need such intensive monitoring, decreasing false negatives could be favored, and in our model, lowering the cutoff to 0.04 for the rescaled prediction yields a sensitivity of 0.94, with a decrease in specificity to 0.24 (Figure 3). Alternatively, when identifying outpatients initiating more benign agents who might benefit from more intensive monitoring, a higher cutpoint with low false positives may be preferred.

There are several limitations of this study. First, it was performed in a single hospital system and global applicability is unproven. Validation in novel institutions is necessary prior to widespread clinical use. Additionally, the retrospective and observational nature lead to many missed cases, as most patients will not have an ECG performed within 24 hours of

medication administration. Further, some variables that could potentially be useful for model development are excluded if they were not included in the harmonized dataset.

Conclusion

As we enter the age of precision medicine, we work toward the goal of identifying the right drug, for the right patient, at the right time. In this investigation, we applied ML models to individuals receiving known QT prolonging medications, and found that these methods can reasonably predict which patients are susceptible to develop significant drug induced QT prolongation. The combination of random oversampling and a deep neural network algorithm provided superior classification compared with other models. Our methodology and use of harmonized EHR data is an important step for developing prediction tools that can be applied to clinical practice. Future studies should examine whether the addition of common genetic variants can improve the prediction accuracy of diLQTS and whether clinical decision support tools can be developed to improve the safety surrounding the use of medications known to prolong the QT interval.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Data were made available through the Health Data Compass data warehouse, part of the Colorado Center for Personalized Medicine (<https://www.healthdatacompass.org>).

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grants from the NIH/NHLBI (K23 HL127296; R01 HL146824) and from the Google Cloud COVID-19 research credits program.

References

1. Roden DM. Clinical practice. Long-QT syndrome. *N Engl J Med*. 2008;358(2):169–176. [PubMed: 18184962]
2. Drew BJ, Ackerman MJ, Funk M, et al. 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(8):1047–1060. [PubMed: 20142454]
3. Moss AJ, Schwartz PJ, Crampton RS, et al. The long QT syndrome. Prospective longitudinal study of 328 families. *Circulation*. 1991;84(3):1136–1144. [PubMed: 1884444]
4. Fermini B, Fossa AA. The impact of drug-induced QT interval prolongation on drug discovery and development. *Nat Rev Drug Discov*. 2003;2(6):439–447. [PubMed: 12776219]
5. Noseworthy PA, Newton-Cheh C. Genetic determinants of sudden cardiac death. *Circulation*. 2008;118(18):1854–1863. [PubMed: 18955676]
6. Noseworthy PA, Peloso GM, Hwang SJ, et al. QT interval and long-term mortality risk in the Framingham heart study. *Ann Non-invasive Electrocardiol*. 2012;17(4):340–348.
7. Roden DM, Woosley RL, Primm RK. Incidence and clinical features of the quinidine-associated long QT syndrome: implications for patient care. *Am Heart J*. 1986;111(6):1088–1093. [PubMed: 3716982]
8. Woosley RL, Chen Y, Freiman JP, Gillis RA. Mechanism of the cardiotoxic actions of terfenadine. *JAMA*. 1993;269(12):1532–1536. [PubMed: 8445816]

9. Lasser KE, Allen PD, Woolhandler SJ, Himmelstein DU, Wolfe SM, Bor DH. Timing of new black box warnings and withdrawals for prescription medications. *JAMA*. 2002;287(17):2215–2220. [PubMed: 11980521]
10. Roden DM. Drug-induced prolongation of the QT interval. *N Engl J Med*. 2004;350(10):1013–1022. [PubMed: 14999113]
11. Sauer AJ, Newton-Cheh C. Clinical and genetic determinants of torsade de pointes risk. *Circulation*. 2012;125(13):1684–1694. [PubMed: 22474311]
12. Mercurio NJ, Yen CF, Shim DJ, et al. Risk of QT interval prolongation associated with use of hydroxychloroquine with or without concomitant azithromycin among hospitalized patients testing positive for coronavirus disease 2019 (COVID-19). *JAMA Cardiol*. 2020;5(9):1036–1041. [PubMed: 32936252]
13. Poluzzi E, Raschi E, Godman B, et al. Pro-arrhythmic potential of oral antihistamines (H1): combining adverse event reports with drug utilization data across Europe. *PLoS One*. 2015;10(3):e0119551. [PubMed: 25785934]
14. Calkins H, Reynolds MR, Spector P, et al. Treatment of atrial fibrillation with antiarrhythmic drugs or radiofrequency ablation: two systematic literature reviews and meta-analyses. *Circ Arrhythm Electrophysiol*. 2009;2(4):349–361. [PubMed: 19808490]
15. Poncet A, Gencer B, Blondon M, et al. Electrocardiographic screening for prolonged QT interval to reduce sudden cardiac death in psychiatric patients: a cost-effectiveness analysis. *PLoS One*. 2015;10(6):e0127213. [PubMed: 26070071]
16. Sorita A, Bos JM, Morlan BW, Tarrell RF, Ackerman MJ, Caraballo PJ. Impact of clinical decision support preventing the use of QT-prolonging medications for patients at risk for torsade de pointes. *J Am Med Inform Assoc*. 2015;22(e1):e21–e27. [PubMed: 25324555]
17. Tisdale JE, Jaynes HA, Kingery JR, et al. Effectiveness of a clinical decision support system for reducing the risk of QT interval prolongation in hospitalized patients. *Circ Cardiovasc Qual Outcomes*. 2014;7(3):381–390. [PubMed: 24803473]
18. Ramirez J, Monasterio V, Mincholé A, et al. Automatic SVM classification of sudden cardiac death and pump failure death from autonomic and repolarization ECG markers. *J Electrocardiol*. 2015;48(4):551–557. [PubMed: 25912974]
19. Beaulieu-Jones BK, Greene CS. Semi-supervised learning of the electronic health record for phenotype stratification. *J Biomed Inform*. 2016;64:168–178. [PubMed: 27744022]
20. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer-Verlag; 2009.
21. Stallkamp J, Schlipsing M, Salmen J, Igel C. Man vs. computer: benchmarking machine learning algorithms for traffic sign recognition. *Neural Netw*. 2012;32:323–332. [PubMed: 22394690]
22. Kooi T, Litjens G, van Ginneken B, et al. Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal*. 2017;35:303–312. [PubMed: 27497072]
23. Agarwalla S, Sarma KK. Machine learning based sample extraction for automatic speech recognition using dialectal Assamese speech. *Neural Netw*. 2016;78:97–111. [PubMed: 26783204]
24. Tanana M, Hallgren KA, Imel ZE, Atkins DC, Srikumar V. A comparison of natural language processing methods for automated coding of motivational interviewing. *J Subst Abuse Treat*. 2016;65:43–50. [PubMed: 26944234]
25. Bao W, Yue J, Rao Y. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS One*. 2017;12(7):e0180944. [PubMed: 28708865]
26. Rolland B, Reid S, Stelling D, et al. Toward rigorous data harmonization in cancer epidemiology research: one approach. *Am J Epidemiol*. 2015;182(12):1033–1038. [PubMed: 26589709]
27. Boffetta P, Bobak M, Borsch-Supan A, et al. The consortium on health and ageing: network of cohorts in Europe and the United States (CHANCES) project—design, population and data harmonization of a large-scale, international study. *Eur J Epidemiol*. 2014;29(12):929–936. [PubMed: 25504016]
28. FitzHenry F, Resnic FS, Robbins SL, et al. Creating a common data model for comparative effectiveness with the observational medical outcomes partnership. *Appl Clin Inform*. 2015;6(3):536–547. [PubMed: 26448797]

29. Johnson KW, Torres Soto J, Glicksberg BS, et al. Artificial intelligence in cardiology. *J Am Coll Cardiol*. 2018;71(23):2668–2679. [PubMed: 29880128]
30. Woosley RL, Black K, Heise CW, Romero K. [CredibleMeds.org](https://www.crediblemeds.org/): what does it offer? *Trends Cardiovasc Med*. 2018;28(2):94–99. [PubMed: 28801207]
31. Gibbs C, Thalamus J, Heldal K, Holla OL, Haugaa KH, Hysing J. Predictors of mortality in high-risk patients with QT prolongation in a community hospital. *Europace*. 2017;20(F1):f99–f107.
32. Trinkley KE, Page RL 2nd, Lien H, Yamanouye K, Tisdale JE. QT interval prolongation and the risk of torsades de pointes: essentials for clinicians. *Curr Med Res Opin*. 2013;29(12):1719–1726. [PubMed: 24020938]
33. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321–357.
34. Chai KE, Anthony S, Coiera E, Magrabi F. Using statistical text classification to identify health information technology incidents. *J Am Med Inform Assoc*. 2013;20(5):980–985. [PubMed: 23666777]
35. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning (ICML)*; June 25–29, 2006; Pittsburgh, PA: Carnegie Mellon University; 2006:233–240.
36. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J*. 2005;47(4):458–472. [PubMed: 16161804]
37. Zhao Y, Wong ZS, Tsui KL. A framework of rebalancing imbalanced healthcare data for rare events' classification: a case of look-alike sound-alike mix-up incident detection. *J Healthc Eng*. 2018;2018:6275435. [PubMed: 29951182]
38. Bates J, Fodeh SJ, Brandt CA, Womack JA. Classification of radiology reports for falls in an HIV study cohort. *J Am Med Inform Assoc*. 2016;23(e1):e113–e117. [PubMed: 26567329]
39. Xue JH, Hall P. Why does rebalancing class-unbalanced data improve AUC for linear discriminant analysis? *IEEE Trans Pattern Anal Mach Intell*. 2015;37(5):1109–1112. [PubMed: 26353332]
40. Tang Y, Zhang YQ, Chawla NV, Krasser S. SVMs modeling for highly imbalanced classification. *IEEE Trans Syst Man Cybern B Cybern*. 2009;39(1):281–288. [PubMed: 19068445]
41. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444. [PubMed: 26017442]
42. Betancur J, Commandeur F, Motlagh M, et al. Deep learning for prediction of obstructive disease from fast myocardial perfusion SPECT a multicenter study. *JACC Cardiovasc Imaging*. 2018;11(11):1654–1663. [PubMed: 29550305]
43. Zhang J, Gajjala S, Agrawal P, et al. Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. *Circulation*. 2018;138(16):1623–1635. [PubMed: 30354459]
44. 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(4):315–325. [PubMed: 23541006]
45. 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: a pilot study. *Circulation*. 2017;135(14):1300–1310. [PubMed: 28213480]
46. Lancaster MC, Sobie EA. Improved prediction of drug-induced torsades de pointes through simulations of dynamics and machine learning algorithms. *Clin Pharmacol Ther*. 2016;100(4):371–379. [PubMed: 26950176]

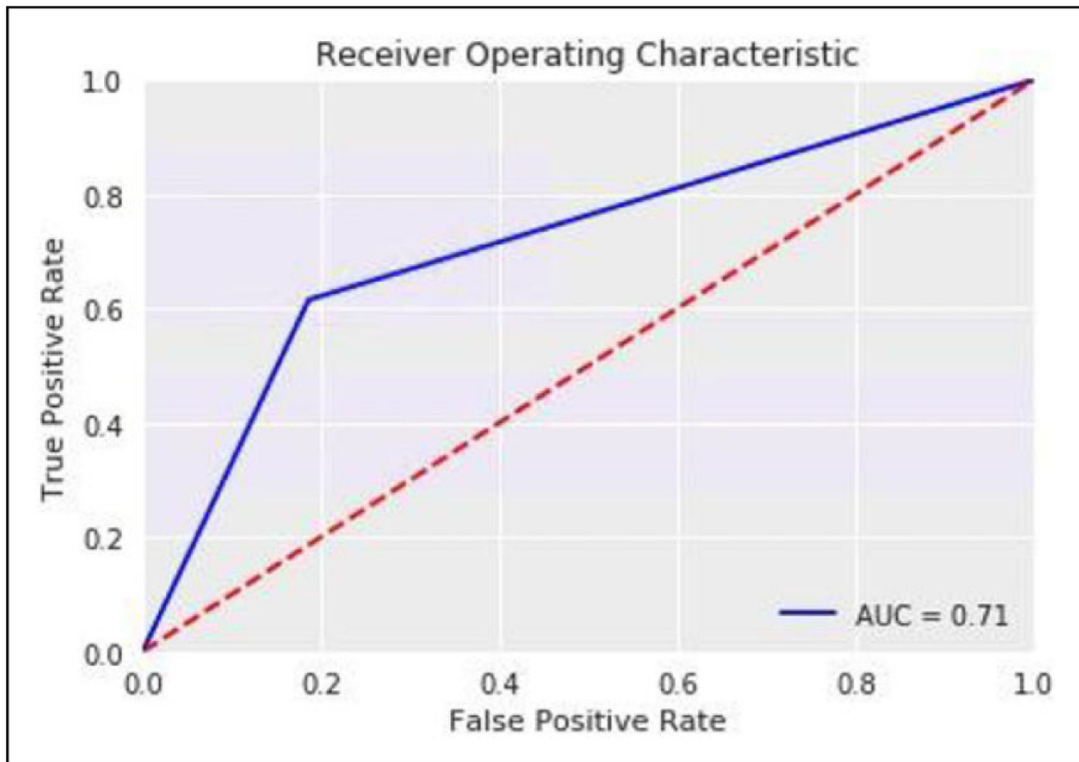


Figure 1.
Receiver operating characteristic for deep neural network.

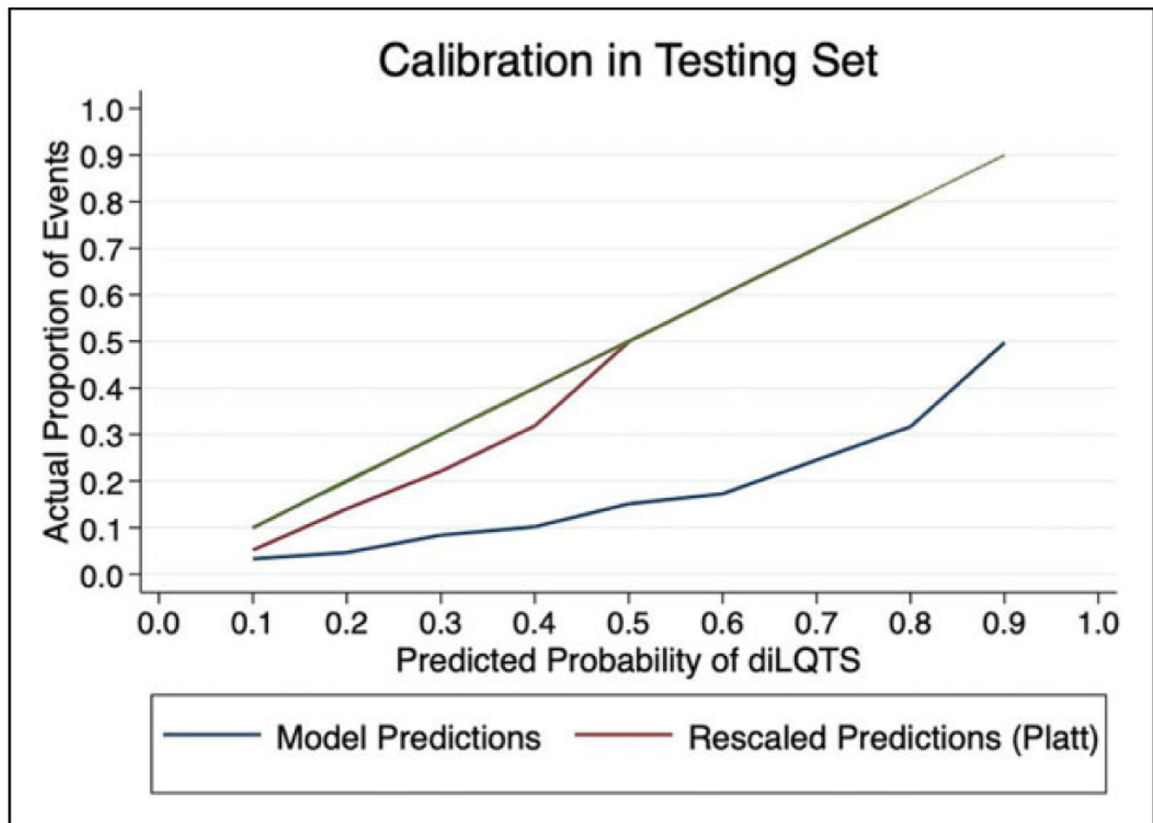


Figure 2.
Original model predictions and predictions following rescaling.

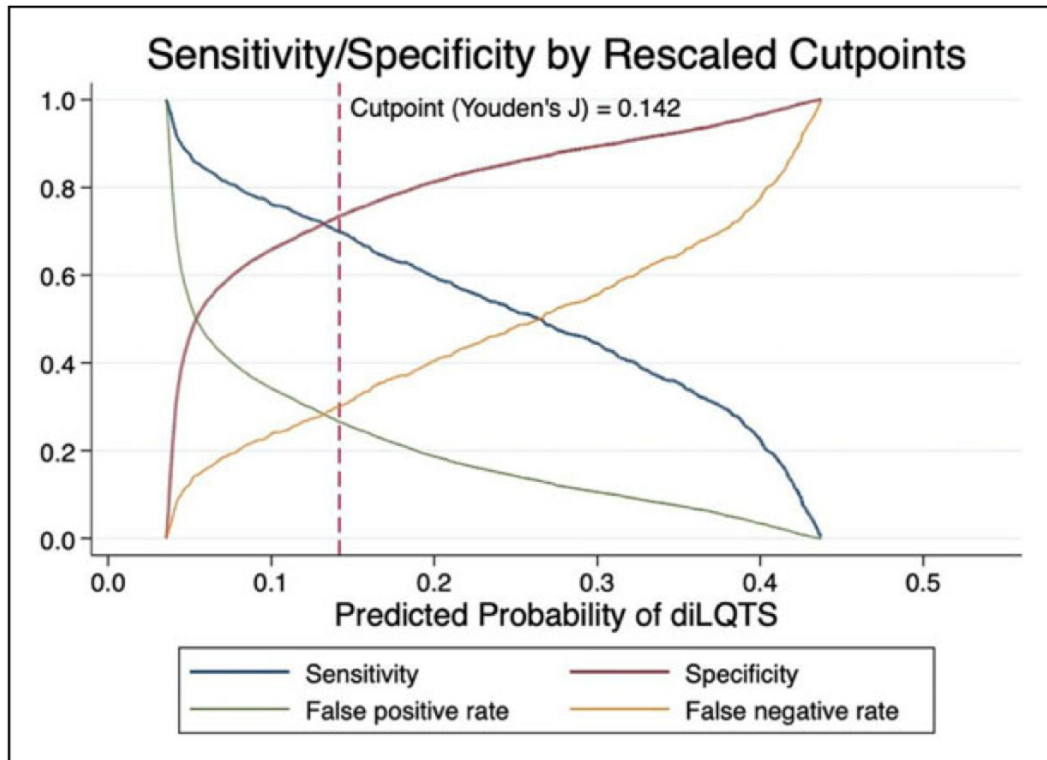


Figure 3.
Model performance at varying cutpoints.

Table 1.UCHealth Population Baseline Demographics.^a

	Control	Case
Number (%)	31081	4558
	87.22%	12.78%
Age (Mean ± SD)	54.34, 18.81	58.78, 17.45
Female sex (%)	53.3%	52.04%
Hypertension (%)	2.77%	2.41%
Coronary artery disease (%)	1.07%	1.38%
Heart failure (%)	0.3%	0.68%
Diabetes mellitus (%)	0.9%	0.61%
Obesity (%)	0.53%	0.54%
Chronic kidney disease (%)	0.31%	0.41%

^aDiagnoses based on presence of diagnosis code (ICD-10; ICD-9) for each. Hypertension: I10x; 401.x, Coronary artery disease: I25.1; 414.01 Heart failure: I50.9, 428.0, Type II Diabetes Mellitus: E11.9, 250.00, Obesity: E66.9, 278.0, Chronic kidney disease: N18.9, 585.9.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Various Resampling Techniques Using Deep Neural Network.

Sampling method	F-Score	AUC
SMOTE	0.385	0.68
RandomUnderSampler	0.126	0.53
ADASYN	0.35	0.68
RandomOverSampler	0.404	0.71

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Accuracy of ML Methods.

ML method	F1-score	AUC
Random Forest	0.385	0.69
Logistic Regression	0.364	0.65
Naïve Bayes	0.383	0.67
Deep Neural network	0.404	0.71

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript