Enhanced QTc Interval Monitoring in the CSICU: Evaluating the Impact of Synthetic Data and Machine Learning Techniques

Hyunjun Choi[1], Debbie Lin Teodorescu[2,3], Trevor Mears[4], Gizem Bilgili[2], Xi Li[1], Jui-Hsuan Chang[1], Nicholas Matsumoto[1], Miguel E. Hernandez[1], Zhiping Paul Wang[1], Bernice Coleman[2,4], Jason H. Moore[1]

Introduction

Prolonged QTc has been recognized since the 1970s as a cardiac risk marker, with a QTc of 500 ms considered dangerously prolonged and requiring immediate action. The gold standard for monitoring remains manual measurement on 12-lead ECGs, necessitating costly, labor-intensive assessments. We have redefined QTc monitoring as a binary classification task to differentiate dangerously prolonged QTc (≥ 500 ms) from less dangerous cases. The challenge is that automated telemetry measurements achieve a 1.00 negative predictive value but only a 0.07 positive predictive value, resulting in high false alarm rates. To address these issues, synthetic data generation techniques were utilized, which are more efficient and less laborious compared to manual measurement of QT intervals on 12-lead ECGs. In this study, a total of six data-driven methods were used, which are based on the following techniques: Gaussian Copula, CTGAN, Bayesian Network, TVAE, RTVAE, and DDPM. Our primary objective was to benchmark classification models trained on three training dataset configurations: original, synthetic, and combined. We leveraged Aliro, an automated machine learning platform that optimizes model selection and hyperparameter tuning, to develop these classification models. We evaluated these models' ability to meet predefined PPV and NPV thresholds for QTc risk classification while investigating how each type of training data impacts model performance on the test dataset.

Methods

Study Population and Data Collection

In this Cedars-Sinai Medical Center IRB-approved study, we conducted a retrospective analysis of sequential sets of simultaneous 12-lead ECG and telemetry strips collected April 1, 2023–August 31, 2023, from adult patients aged 18 or older admitted to a tertiary-care cardiac surgical ICU. The dataset consists of 23 input features and one target feature. The training set contains 88 instances, while the test set contains 23 instances.

Study Design and Analysis.

As depicted in the bottom part of Figure 1, this figure outlines the model evaluation process, including synthetic data generation, model optimization, filtering, and evaluation of high-performing models.

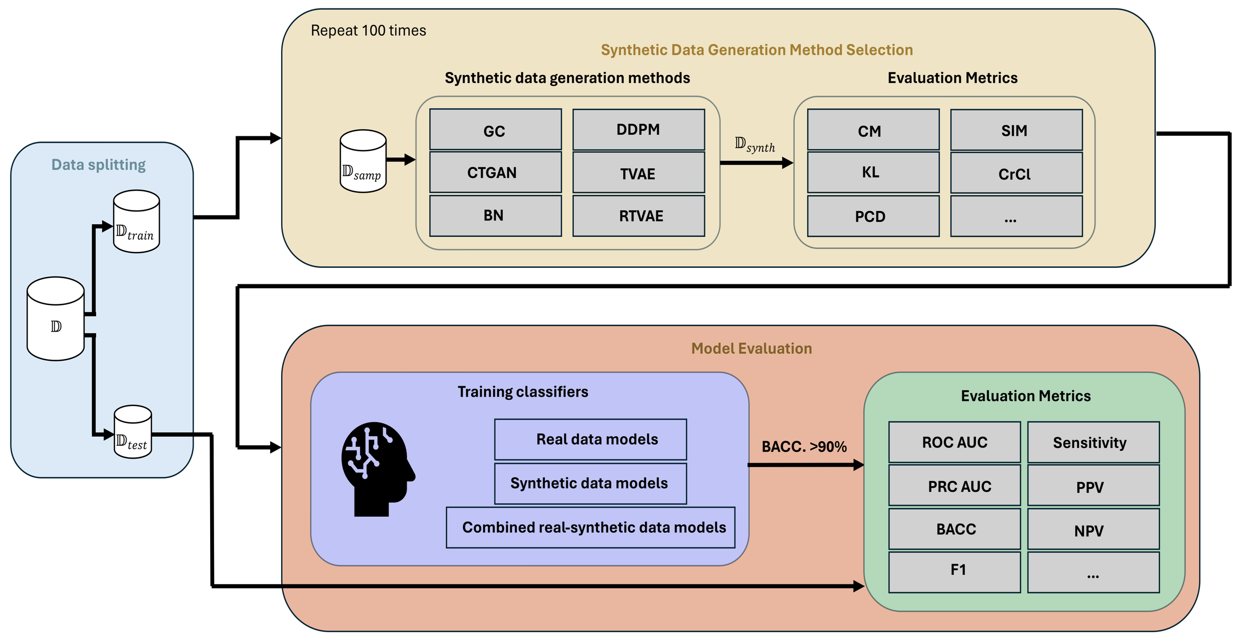

Figure 1: Synthetic Data Generation Method Selection and Model Evaluation Process.

Synthetic Data Generation Methods

We undertook a rigorous comparative analysis encompassing six distinct approaches. These metrics encompassed Pairwise Correlation Difference (PCD), Log Cluster Measure (CM), Kullback-Leibler (KL) Divergence, Cross-Classification (CrCl), SDV Quality Scoree, SDV Range Coverage (RC), and SDV Statistic Similarity (SIM), all designed to quantify the degree of concordance between the original and synthetically generated data. Figure 2 and Figure 3 present a detailed summary of the mean values and 95% confidence for the different synthetic data generation methods across the metrics.

- Gaussian Copula (GC), Bayesian Network (BN), Conditional tabular GAN (CTGAN), DDPM, RTVAE, TVAE.

Data Preprocessing

The dataset underwent a stratified partitioning process, with 80% (n = 88) allocated for model training and 20% (n = 23) for testing. Missing values in the training set were imputed using the scikit-learn multivariate feature imputer. Continuous variables were transformed using Z-score normalization. The same imputation and normalization methods were applied to the testing set using the parameters from the training set.

Model Training and Evaluation

Seven machine learning models were evaluated: Decision Tree Classifier (DTC), Gradient Boosting Classifier (GBC), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Random Forest Classifier (RFC), and XGBoost Classifier (XGB). Models were trained on three datasets: original, synthetic (500 and 1000 instances), and combined. Top-performing models were selected based on balanced accuracy and evaluated on the test set using Positive Predictive Value (PPV) and Negative Predictive Value (NPV).

Method	PCD↓	CM↓	KL↓	CrCL(1)	Quality Score(1)	RC(1)	SIM(1)
GC	4.079 (3.446, 4.712)	−2.481 (−3.056, −1.906)	200.391 (173.635, 227.147)	0.937 (0.739, 1.143)	0.854 (0.835, 0.874)	0.885 (0.822, 0.949)	0.975 (0.962, 0.988)
CTGAN	7.845 (7.055, 8.635)	−1.660 (−1.830, −1.491)	230.573 (206.453, 254.693)	0.872 (0.662, 1.082)	0.742 (0.713, 0.771)	0.934 (0.888, 0.979)	0.877 (0.841, 0.913)
BN	**2.610** (2.005, 3.214)	**−3.878** (−4.582, −3.175)	**163.039** (142.997, 183.081)	**1.041** (0.888, 1.194)	**0.898** (0.886, 0.910)	0.946 (0.901, 0.990)	**0.980** (0.972, 0.987)
TVAE	4.943 (3.969, 5.918)	−2.166 (−2.473, −1.860)	190.000 (170.881, 209.120)	0.893 (0.680, 1.107)	0.831 (0.813, 0.849)	0.749 (0.676, 0.823)	0.958 (0.944, 0.972)
RTVAE	6.593 (5.323, 7.862)	−1.789 (−2.039, −1.539)	422.037 (407.451, 436.623)	0.857 (0.665, 1.048)	0.674 (0.649, 0.700)	0.324 (0.249, 0.400)	0.968 (0.953, 0.983)
DDPM	8.155 (7.459, 8.850)	−1.619 (−1.751, −1.486)	428.754 (407.358, 450.151)	0.844 (0.668, 1.020)	0.620 (0.603, 0.638)	**0.992** (0.974, 1.011)	0.890 (0.867, 0.912)

Figure 2: Performance Metrics of Synthetic Data Generation Methods: Means and 95% Cis.

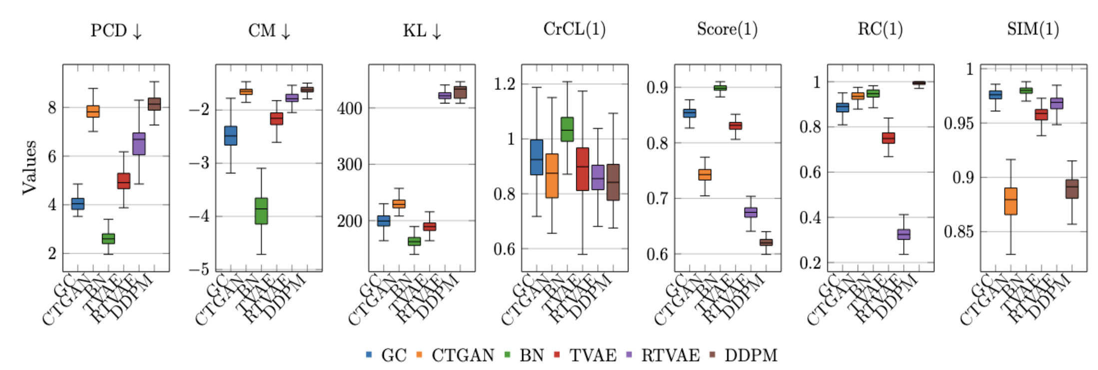

Figure 3: Box Plot of Evaluation Metrics for Various Synthetic Data Generation Models..

Results

Figure 3 illustrates a heatmap depicting the performance of each model on the test set. The results offer valuable insights into the potential of synthetic data for enhancing QTc interval monitoring in Cardiac Surgery Intensive Care Units (CSICU) and improving sudden cardiac death prevention.

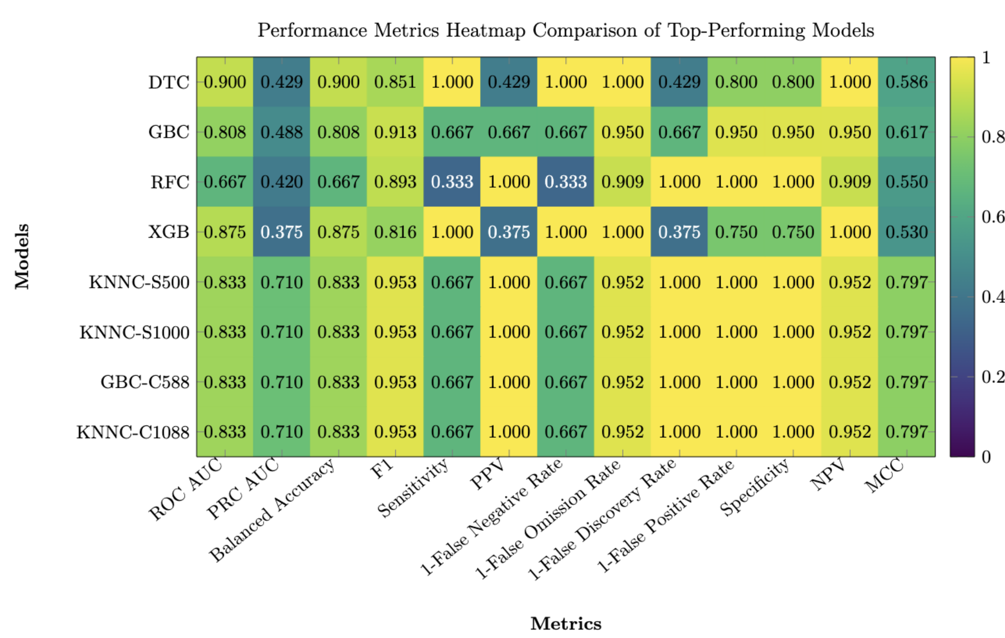

Figure 4: Performance Metrics of Synthetic Data Generation Methods: Means and 95% Cis.

Discussion

- The challenge of QTc monitoring involves highly imbalanced datasets that are costly to acquire.
- Accurate classification is critical, as under classification increases the risk of sudden cardiac death, and overclassification can unnecessarily interfere with other treatments.
- Machine learning offers powerful classification capabilities but faces practical obstacles in clinical use: reliance on large datasets, imprecise methods for estimating sample sizes, and broad feature sets that increase data collection costs.
- The synthetic data approach's resulting models will require further validation and refinement with larger sets of data from patients before clinical use.
- Nevertheless, this study is more efficient than it would have been initially because there are confirmed to be classifiers likely to be effective, improving participation by stakeholders; and the highly informative feature set has been better delineated.

Conclusion

- The investigation revealed that machine learning models trained on datasets that include synthetic data generated by the Bayesian Network-based synthetic data generation method demonstrate promising performance on the test dataset, meeting the predefined criteria for positive predictive value (PPV) and negative predictive value (NPV).
- These observations indicate that employing synthetic data created by the BN-based method shows significant potential for increasing model performance in the QTc risk classification task.
- Additionally, this method is advantageous because, compared to the high costs required for data collection using the manual 12-lead ECG measurement method, it allows for relatively simple synthetic data generation using existing data.
- Follow up investigations should verify these discoveries by conducting extensive studies and prospective trials in various clinical environments.
- The incorporation of synthetic data into therapeutic decision-making procedures necessitates a meticulous examination of ethical and practical considerations.
- By thoroughly addressing these several aspects, we can guarantee the reliability as well as effectiveness of our approach in practical clinical applications.

Author Affiliation

1 Department of Computational Biomedicine, Cedars-Sinai Medical Center, West Hollywood, CA
2 Smidt Heart Institute, Cedars-Sinai Medical Center, West Hollywood, CA
3 Pulse Heart Institute, Tacoma, WA
4 Nursing Research and Quality Department, Brawerman Nursing Institute, Cedars Sinai Medical Center, West Hollywood, CA

Acknowledgements

This work was supported by NIH grants AG066833.