Machine Learning Prediction of Blood Potassium at Different Time Cutoffs

J. Bergquist, D. Dade, B. Zenger, R. MacLeod, X. Ye, R. Ranjan, T. Tasdizen, B. Steinberg

Introduction: Recent studies have developed ECG machine learning (ECG-ML) tools to classify abnormal serum potassium, enabling rapid, sensitive, and automated monitoring that would otherwise require blood tests. Because serum potassium and ECG morphology changes exhibit a physiologically well-understood connection, we are motivated to interrogate ML potassium classification tasks with respect to the time between the ECG and potassium readings.

Methods: Using a dataset of over 350,000 matched clinical 12 lead ECGs and blood potassium measurements, we trained a convolutional neural network to classify abnormal (serum potassium above 5 mEq/L) vs normal (serum potassium between 4 and 5 mEq/L) from the ECG alone. We compared training with ECGs and potassium measurements filtered to be within 1 hour, 30 minutes, and 15 minutes of each other. We explored scenarios that both leveraged all available data at each time cutoff as well as restricted data to match training set sizes across the time cutoffs. For each case, we trained five separate instances of our neural network to account for variability.

Results: Table 1 summarizes the results. Truncating the training sets to the same size as the 15 minute cutoff results in comparable accuracy and area under the receiver operator curve (ACU) for all time cutoffs.

Table 1. Network performance in the testing datasets across training scenarios. Weighted accuracy and area under the receiver operator curve (AUC) are shown as ± one standard deviation.

Time Cutoff (seconds)	Training Set Size	Test Weighted Accuracy	Test AUC
900	148,146	72.5% ±1.1%	0.814 ±0.0032
1800	244,818	72.4% ±4.3%	0.825 ±0.0082
1800	148,146	72.6% ±2.2%	0.823 ±0.0051
3600	344,693	76.3% ±0.5%	0.850 ±0.0030
3600	148,146	72.8% ±3.7%	0.830 ±0.0052

Discussion: Contrary to expectation, the theoretically more difficult task of potassium classification within 1 hour of ECG produced comparable results to prediction within 15 minutes when training set size was controlled. With expanded training set size, the 1 hour classification performance improved, as would be expected from other ML studies. Our future studies will continue to explore the performance of ML potassium predictions through investigations of failure cases, identification of biases, and explainability analyses.