Characterizing Electrocardiograms of Severe Hypoglycemia using K-Means based Clustering

Sejal Mistry, BA, Samir Abdelrahman, MS, PhD, Candace Reno, PhD, Danielle Groat, PhD, Ramkiran Gouripeddi, MBBS, MS, Simon Fisher, MD-PhD, Julio Facelli, PhD
University of Utah, Salt Lake City, UT, USA

Introduction
Severe hypoglycemia is an acute condition that, if uncorrected, can lead to death. Hypoglycemia has been shown to cause abnormalities in electrocardiogram (ECG) parameters. Studies suggest that severe hypoglycemia-associated fatality is mediated by cardiac arrhythmias. Unsupervised learning methods have successfully categorized cardiomyopathies with human-engineered features of ECG data, but have not been tested in the context of hypoglycemia. K-means is a widely used machine learning method that extracts data subgroups. We propose to identify appropriate k-means based clustering methods in order to characterize ECG changes in severe-hypoglycemia induced rats.

Methods
91 Sprague-Dawley rats (37 diabetic, 54 non-diabetic) were administered insulin beginning at euglycemia (>15 mg/dL glucose). Once severe hypoglycemia (<15 mg/dL glucose) was achieved, rats were administered insulin for either 3 hours or until death. ECGs were measured in millivolts (mV) and collected every millisecond (ms) while glucose measurements were collected every 15 minutes. ECGs were filtered using a finite impulse response (FIR) bandpass, standard scaler, and absolute scaler. ECGs were partitioned 7.5 minutes after every glucose reading into segments of 10 seconds, 1 second, 300 ms, and 200 ms. This resulted in 1,534 samples for analysis. Time segments were de-identified of all experimental groups before analysis. K-means clustering (k-means) was tested using the elbow method for 2-30 clusters and evaluated using the silhouette coefficient (SC). K-means for longitudinal data (kml) was tested for 2-26 clusters and evaluated using the Calinski Harabatz score (CH). K-means for longitudinal data using shape-respecting distance (kmlShape) was tested for 2-10 clusters and evaluated using SC and CH. Python 3.7 was used for data preparation while k-means analyses were conducted with R-packages.

Results
The best performing method was kmlShape. Neither the k-means nor kml identified useful clusters. kmlShape used 300 ms segments and identified five clusters with a SC of 0.0500 and CH of 30.721. The colors indicate clusters of ECG that kmlShape identified while the grey represents all ECG segments used for analysis (Figure 1). Cluster 1 and 5 demonstrate irregular rhythm (no discernable P, QRS, or T wave). Cluster 2, 3, and 4 demonstrate regular rhythms, with Cluster 2 having the fastest HR and Cluster 3 and 4 having slower HR. Notably, Cluster 2 contained the greatest proportion of samples from euglycemic conditions while Clusters 3 and 4 contained the greatest proportions of samples from hypoglycemic conditions. Cluster 3 additionally contained more samples from diabetic rats while Cluster 4 contained more samples from non-diabetic rats.

Conclusion
These findings suggest that kmlShape may be an appropriate method to cluster minimally processed ECG data. The patterns of varying HR in Clusters 2, 3, and 4 support previous findings linking bradyarrhythmia (slow HR) to sudden hypoglycemia-associated arrhythmias. Studies of severe hypoglycemia are unethical in humans, thus this study demonstrates a novel method to classify ECG waves in previously unexplored conditions. Specifically, these findings provide a foundational understanding to cardiac electric activity in low glycemic states, and have implications for management of severe hypoglycemia and for severe-hypoglycemia induced cardiac arrhythmias. Future work entails increasing the number and types of samples, conducting a beat-by-beat kmlShape analysis, and investigation of more sophisticated clustering methods.

References