Examining The Ability of Different ML Approaches to Predict Health Outcomes with Digital Health Platform

Mansur Shomali, MD[1], Junjie Luo, MS, Abhimanyu Kumbara, MS, MBA, Anand Iyer, PhD, MBA, and Gordon Gao, PhD[2]
[1] Welldoc Inc. [2] Johns Hopkins Carey Business School, Center for Digital Health and Artificial Intelligence (CDHAI)

Introduction

Monitoring glucose data, along with connecting it to other important health factors like medication, education, diet, activity, and lab data (MEDAL) is critical to optimal glucose management. Continuous glucose monitoring (CGM) sensors provide dense glucose data to accurately monitor glucose levels. Digital health platforms like Welldoc®, delivers artificial intelligence (AI) driven, personalized digital coaching and insights. This enables individuals with diabetes to connect glucose data from CGM devices with MEDAL engagement data to support their overall health. Additionally, real-time data sources like CGM in combination with AI driven digital health solutions provides the power to predict health outcomes. We hypothesized that by combining the dense CGM data with the MEDAL data from the platform, we could accurately predict future glucose time in range (TIR). We also evaluated an ensemble of machine learning models to determine an appropriate model based on performance.

Methods

We evaluated real-world CGM data and MEDAL engagement data from a digital health platform for 304 individuals with type 1 and type 2 diabetes. The baseline period to train the model was defined as first 30 days from when the first MEDAL and CGM reading was recorded. The prediction period was set to 70-90 days from baseline period. An individual had to have at least 10 days of CGM data in the prediction period, with at least 70% sensor wear time in those 10 days. The input features included various CGM outcome variables, all MEDAL features in the baseline period, as well as demographic information. The TIR >=0.7 or <0.7 was defined as a binary outcome variable for prediction, as TIR >=0.7 is shown to be a clinically meaningful target for optimal diabetes management.

We evaluated Light Gradient Boosting Machine (GBM), Random Forest Classifier, Quadratic Discriminant Analysis, Naïve Bayes, and Logistic Regression models to predict whether TIR would be over 0.7 or not in the prediction periods. SHAP values were evaluated to understand feature importance among the input features.

Results

Among the five different models that were implemented, the Light GBM model had the best prediction accuracy of 0.80, and an AUC of 0.88. Random Forest Classifier had the next best model performance with accuracy of 0.77 and AUC of 0.82. Interestingly, Quadratic Discriminant Analysis model had the best recall score of 0.94. As for feature importance, a high baseline TIR and long exercise duration were shown to increase the probability of future TIR being over 0.7. Factors contributing to a reduced probability of future TIR over 0.7 include high baseline time above range (TAR), gender, and Type 1 Diabetes, as their SHAP values showed negative impact on the model output.

Conclusion

Combining CGM data with MEDAL engagement data through a digital health platform does provide the ability to accurately predict TIR, with a clinically meaningful threshold of over 0.7. Based on assessment of various models, decision tree based models like Light GBM and Random Forest classifier were determined to be better suited for prediction of TIR over a certain threshold. Evaluation of the features within the digital health solution that aligned with high TIR prediction probability provides important information towards optimizing digital health for use with real-time data sources and sensors, like CGM. In future, we may want to further expand on this analysis by segmenting the baseline population based on health and MEDAL engagement characteristics and evaluate model performance along with feature importance for those segments to understand micro behaviors.