Examining The Ability of Different Machine Learning Approaches to Predict Health Outcomes with a Digital Health Platform

Mansur Shomali, MD, CM[1], Abhimanyu Kumbara, MS[1], Junjie Luo, MS[2], Anand Iyer, PhD[1], Gordon Gao[2], PhD
[1]Welldoc, Columbia, Maryland,USA,[2]Center for Digital Health and Artificial Intelligence, Johns Hopkins Carey Business School, Baltimore, MD,USA

BACKGROUND

Monitoring glucose, along with **M**edication, **E**ducation, **D**iet, **A**ctivity, and **L**ab/vital data (MEDAL), aids in optimizing diabetes management. We have developed an AI-driven and personalized digital health platform that enables individuals to combine data from continuous glucose monitoring (CGM) devices and MEDAL engagement data to manage chronic conditions like diabetes. The combination of these data can be useful in informing health outcomes and developing accurate models that can be used to predict future events and patterns, such as glucose time in range (TIR)[1], an important measure now tracked via CGM. The development of efficient and accurate predictive models, like the one shared below, may provide new insights to inform treatment decisions and optimize self-management of chronic conditions, such as diabetes.

SPECIFIC AIMS/PURPOSE

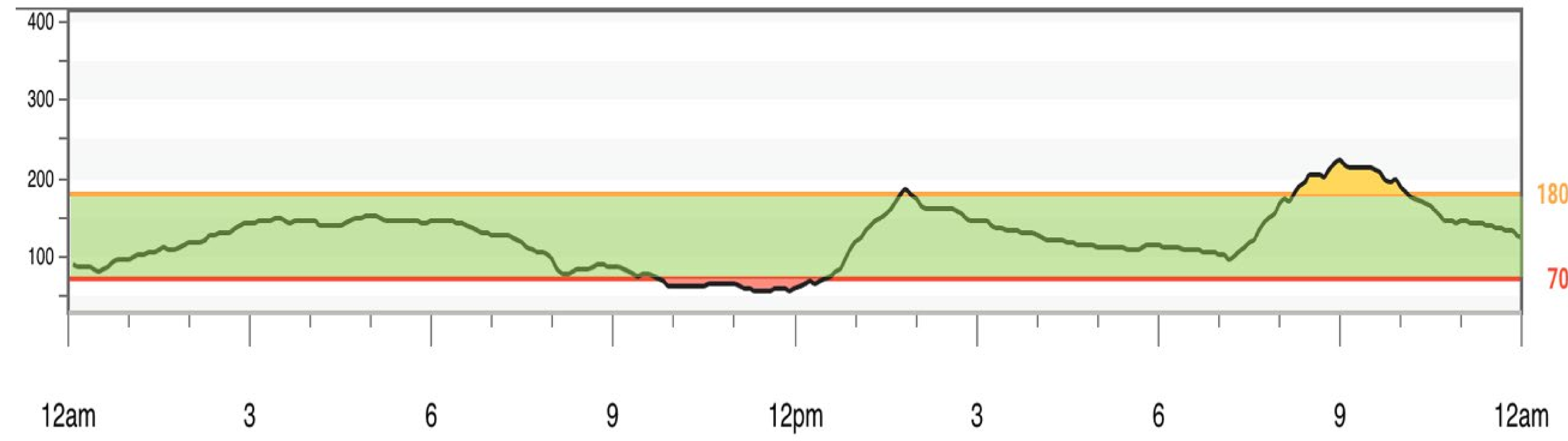
We have previously shown that combining dense CGM data with sparse MEDAL engagement data can accurately predict future MEDAL engagement[2] and future TIR.[1] In this new analysis, to better understand which ML model would be best suited to predict future TIR, we applied an ensemble of machine learning (ML) models and evaluated their respective performance. Additionally, we explored the importance of MEDAL feature engagements and their impact on the prediction outcome.

METHODS

We evaluated real-world CGM data and MEDAL engagement data from a digital health platform for 304 individuals with diabetes. The baseline period to train the models 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. The input features included various CGM outcome variables, all MEDAL feature engagements in the baseline period, and demographic information. TIR >=0.7 or <0.7 was defined as a binary outcome variable for prediction, since TIR >=0.7 is shown to be a clinically meaningful target for optimal diabetes management.[3] We implemented Light GBM (LGBM), Random Forest Classifier, Quadratic Discriminant Analysis, Naïve Bayes, and Logistic Regression models to predict whether TIR in the prediction period would be greater than 0.7 or not. SHAP (**SH**apley **A**dditive ex**P**lanations) values were evaluated to understand the importance of input features on the prediction outcome.

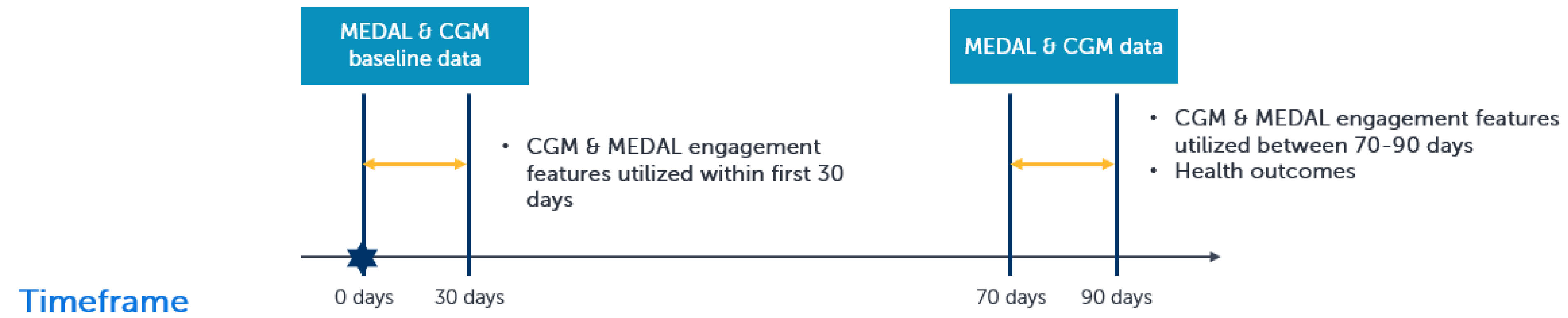
Figure 1: Screenshots of the Digital Health Solution

Figure 2: Sample CGM Trace



In this typical example of CGM data, glucose is plotted on the vertical axis every 5 minutes over 24 hours. The target range (glucose between 70 and 180 mg/dL) is highlighted in green. Below range values are shaded red and above range values are shaded yellow.

Figure 3: Baseline Period and Outcome Period Timeframe

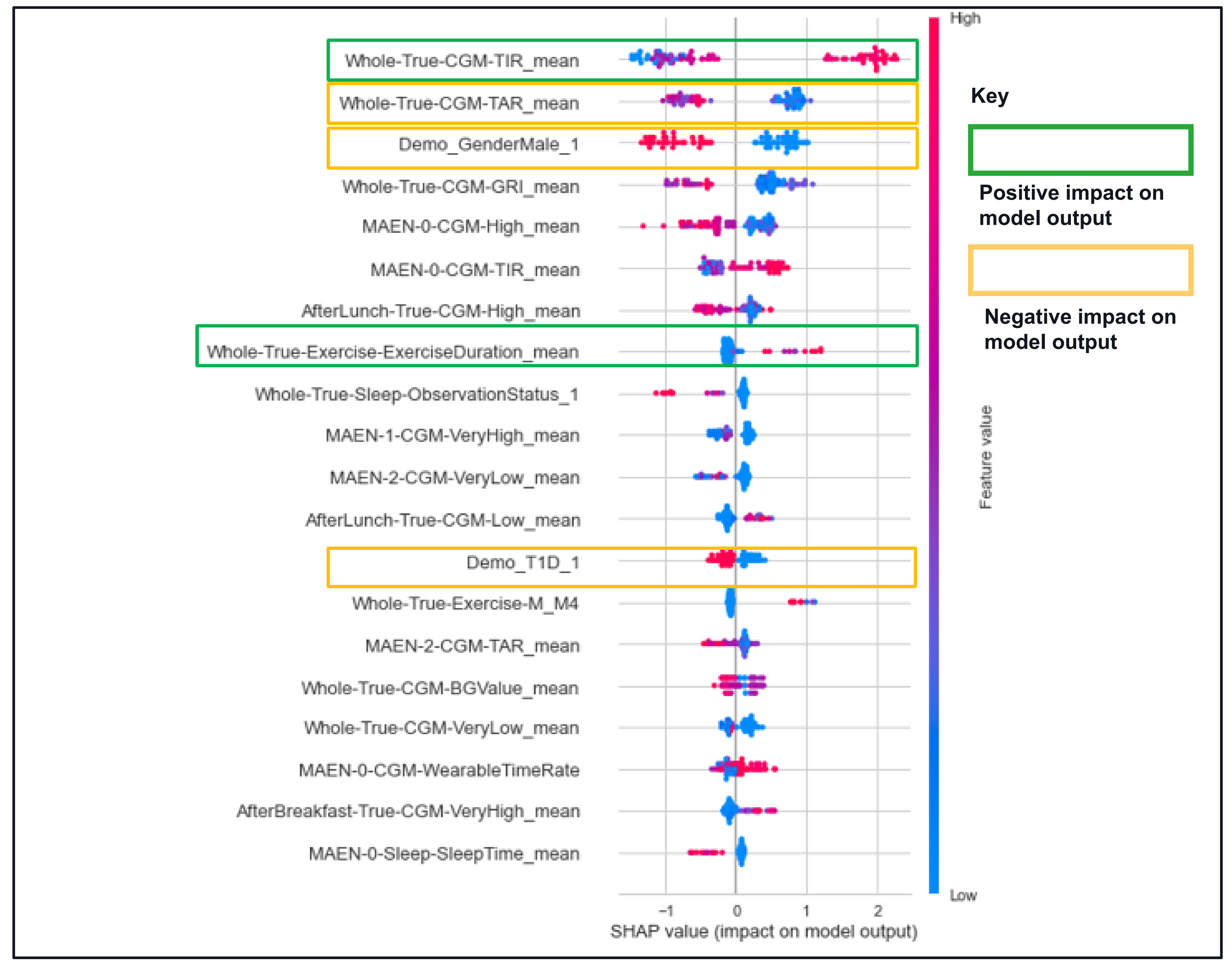


RESULTS

Figure 4: Prediction Model Output Statistics for Each ML Model

Model Name	Sample Size (n)	Accuracy	AUC	Precision	Recall	F1 Score
Light Gradient Boosting Machine	304	0.7970	0.8813	0.8248	0.7500	0.7812
Random Forest Classifier	304	0.7688	0.8289	0.7983	0.7600	0.7607
Quadratic Discriminant Analysis	304	0.5608	0.5541	0.5435	0.9427	0.6857
Naive Bayes	304	0.5286	0.6287	0.5598	0.2527	0.3405
Logistic Regression	304	0.6167	0.5982	0.6521	0.5967	0.6135

Figure 5: LGBM Model Feature Importance Results



- Among the five different models that were implemented, the Light GBM model had the best prediction accuracy of 0.80 and AUC of 0.88.
- The Random Forest Classifier model had the next best performance with an accuracy of 0.77 and AUC of 0.82.
- The 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.
- High baseline time above range (TAR), male gender, and Type 1 diabetes negatively impacted TIR, meaning they reduced the probability of TIR being over 0.7.

CONCLUSIONS

- Machine learning models that combine CGM data along with MEDAL engagement data can help to predict future health outcomes such as TIR.
- The various model output statistics show that decision tree-based models such as Light GBM and Random Forest classifiers could be better suited for predicting health outcome variables such as TIR above or below clinically meaningful thresholds.
- Feature importance analysis can help us to understand which baseline CGM and MEDAL features are more impactful on predicting future health outcomes.
- Future analyses can help strengthen predictive modeling, including feature importance analysis, to better understand how utilization of specific digital health features corresponds to predicting future health outcomes.
- Continued research in predictive modeling, leveraging the power of real-time devices, like CGM in combination with digital health will be instrumental in guiding the next generation of AI and personalized digital solutions.

REFERENCES

[1]Junjie Luo, Abhimanyu Kumbara, Anand Iyer, Mansur Shomali, Gordon Gao. Using Early Engagement Data from a Digital Health Solution to Predict Future Health Outcomes. Presented at INFORMS Healthcare 2023, Toronto, ON Canada, July 26-28, 2023.

[2]Junjie Luo, Abhimanyu Kumbara, Anand Iyer, Mansur Shomali, Gordon Gao. Using Early Engagement Data from a Digital Health Solution to Predict Future Engagement Patterns. Presented at Conference on Health IT and Analytics (CHITA), Washington DC, May 5-6, 2023.

[3]Battelino T, Danne T, Bergenstal RM, et al. Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range. Diabetes Care. 2019 Aug;42(8):1593-1603.