Abstract No. 4CPS-131

care
education
research

Singapore General Hospital
SingHealth

Development Of An Augmented Intelligence Tool To Predict Risk of Uncontrolled Type 2 Diabetes For Personalized Pharmaceutical Care At the Outpatient Setting

WC Ong[1], REL Gan[1], CYL Lim[1], J Ng[1], GY Khee[1], YK Hwang, A Phua[2], S Lim[2], TL Chua[2], YM Bee[3], PS Lim[1]

(1) Division of Pharmacy, Singapore General Hospital; (2) Department of Data Science, Singapore General Hospital; (3) Department of Endocrinology, Singapore General Hospital

INTRODUCTION

Diabetes is a global health crisis, affecting 1 in 10 adults. In Singapore alone, over 400,000 patients currently have diabetes (DM), a number expected to reach 1 million by 2050 [1]. Diabetes is a chronic condition that requires ongoing management to prevent complications. However, comprehensive coordinated patient-centred care approach to proactively engage patients in achieving glycemic target remains a challenge in the outpatient setting.

Hence, the development of an innovative Augmented-Intelligence (AI) model designed to assess the risk of uncontrolled Type 2 diabetes mellitus (T2DM) allows early identification of high-risk population to provide proactive awareness, personalized pharmaceutical care, goal orientation with increased patient engagement and support for better self-management of T2DM.

OBJECTIVE

To develop an AI model that proactively identifies patients at risk of T2DM six months in advance, providing personalized pharmaceutical care support at the outpatient setting.

METHODS

This was a single-centered, retrospective study. The cohort was divided into controlled T2DM and uncontrolled T2DM based on the hemoglobin A1c (HbA1c) at T1 [6 months after the index visit (T0)]. Various data was obtained retrospectively at or nearest to the index visit (T0).

1552 eligible T2DM patients followed up at Singapore General Hospital (SGH) Specialized Outpatient Clinic for at least 1 year from 2021 to 2024.

Exclusion criteria
- Type 1 DM or drug-induced DM (n=11)
- Less than 1 year follow-up (n=388)
- Missing HBA1c & medication adherence (n=13)

1140 patients included in the analysis

Figure 1. Study sample flow chart

Outcome Definition
- Definition of uncontrolled T2DM is based on the American Diabetes Association (ADA) [2]
  - ➢Age <75: hemoglobin A1c (HbA1c) >7%
  - ➢Age ≥ 75: HbA1c >8%

Data Acquisition
- Demographics, physical measurements, clinical characteristics, diabetes medications, laboratory biomarkers, outpatient clinic attendances and clinical interventions at or nearest to T0 were obtained from the institution's electronic medical records.
- Past diabetes-related hospitalization in the past 1 year from T0 was captured.
- Medication adherence data was retrieved from an in-house database.
- Missing data in the physical measurements and biomarkers was assumed to be missing at random. No imputation method was performed.
- Data was de-identified prior to analysis.

Feature Selection
- Feature Importance Plot and Recursive Feature Elimination (RFE) were used.
- Top 22 features & medication adherence were selected for the initial comparison.
- Further refinement of the features was conducted to select only the most relevant ones for optimal model performance.

Statistical Analysis
- The outer train-test split was created following an 80:20 ratio, and within the training set, 10-fold cross validation was used.
- A set of diverse machine learning (ML) algorithm were chosen to compare and evaluate.

Model Evaluation
- Area under the curve (AUC), recall, precision, F1 score, accuracy, specificity, negative predictive value

RESULTS AND DISCUSSION

Table 1: Patient Demographics and clinical characteristics

Characteristics		n (%) unless indicated: Controlled T2DM (n=411)	Uncontrolled T2DM (n=729)	p-value*
Age (median, IQR), years		68.0 (59.0-77.0)	65.0 (57.0-70.0)	**<0.05**
Race, %	Chinese	306 (74.5)	510 (70.0)	**0.045**
	Malay	37 (9.0)	66 (9.1)	
	Indian	40 (9.7)	113 (15.5)	
	Others	28 (6.8)	40 (5.5)	
Male Gender		226 (55.0)	373 (51.2)	0.215
Current Smoker		72 (6.3)	23 (5.6)	0.453
Disease Duration (median, IQR), years^		13 (7-22)	16 (9-22)	**0.012**
Body Mass Index (BMI) (median, IQR), kg/m²^		25.5 (22.6-28.7)	26.3 (23.5-29.7)	**0.007**
HbA1c at T0 (median, IQR), %^		6.9 (6.3-7.7)	8.1 (7.4-9.2)	**<0.001**
Serum creatinine at T0 (median, IQR), umol/L^		106 (73-211)	92 (67-136)	**<0.05**
Serum triglyceride (TG) at T0 (median, IQR), mmol/L^		1.4 (1.0-1.9)	1.4 (1.0-2.1)	0.099
Serum low density lipid (LDL) at T0 (median, IQR), mmol/L^		2.0 (1.7-2.6)	2.3 (1.8-2.8)	**0.008**
Total number of DM medications (oral + injectables), %	0-1	123 (30.1)	76 (10.5)	**<0.001**
	2-3	224 (54.9)	402 (55.4)	
	≥4	61 (15.0)	247 (34.1)	
Medication non-adherence[a]		121 (29.4)	281 (38.5)	**0.002**

* Bolded P-values are ≤0.05 and hence were considered as statistically significant.
^ Missing value: disease duration (n=154), BMI (n=96), HbA1c at T0 (n=182), serum creatinine (n=182), serum TG (n=632), serum LDL (n=651)
[a] Medical adherence was assessed using two self-reported questions. Patients will be classified as adherent only if they answered 'all the time' and 'no' to the two questions, respectively.
1) "In the last one month, how often did you take your medications as prescribed by the doctor?" (All the time; Nearly all the time; Most of the time; About half the time; Less than half the time) [3] and
2) "At times do you forget to take your prescription medications?" (Yes; No) [4].

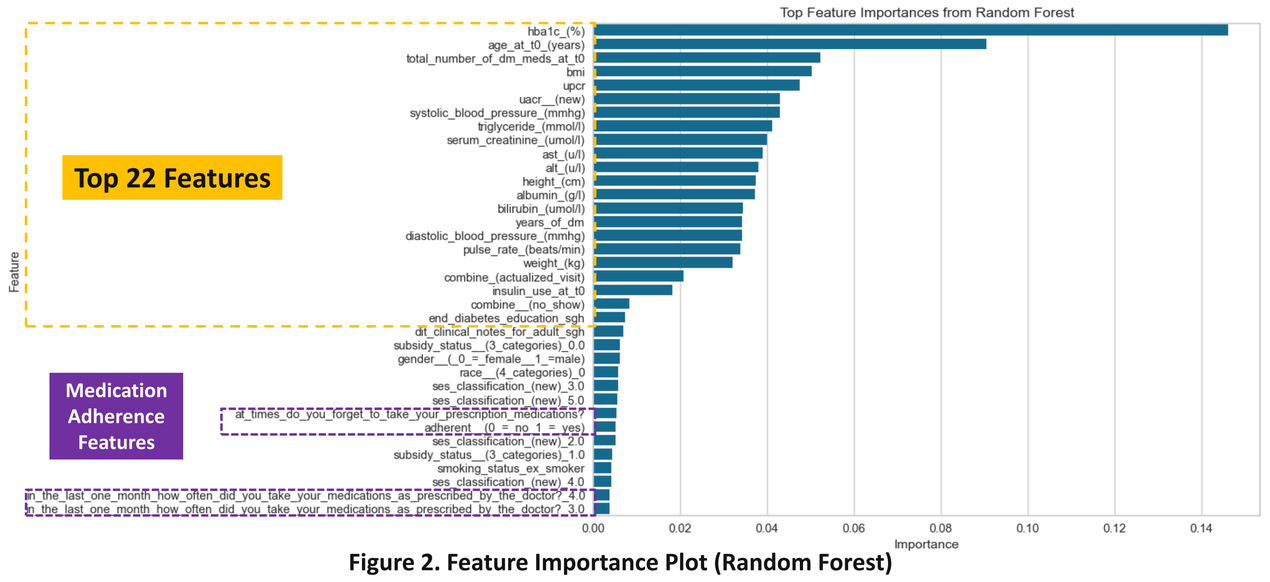

Figure 2. Feature Importance Plot (Random Forest)

Table 2. Comparison of the 3 model performance based on the top 22 features & medication adherence#

Model	AUC 10 Fold CV Mean	Median Test AUC 95% CI (10,000 bootstraps)	Median Test RECALL 95% CI (10,000 bootstraps)	Median Test F1 95% CI (10,000 bootstraps)	Median Test Prec. 95% CI (10,000 bootstraps)	Median Test Accuracy 95% CI (10,000 bootstraps)	Median Test NPV 95% CI (10,000 bootstraps)	Median Test Specificity 95% CI (10,000 bootstraps)
Random Forest	0.814	0.865 [0.812-0.912]	0.898 [0.846-0.945]	0.845 [0.799-0.887]	0.799 [0.735-0.859]	0.789 [0.737-0.842]	0.767 [0.655, 0.869]	0.597 [0.489-0.703]
Logistic Regression	0.772	0.803 [0.740-0.860]	0.844 [0.781-0.900]	0.815 [0.765-0.860]	0.789 [0.722-0.850]	0.754 [0.697-0.807]	0.681 [0.571-0.788]	0.589 [0.493-0.704]
Gradient Boosting	0.815	0.848 [0.791-0.898]	0.891 [0.837, 0.841]	0.847 [0.800-0.889]	0.808 [0.744-0.868]	0.794 [0.741-0.846]	0.763 [0.652-0.859]	0.622 [0.513-0.730]

\# Table result is slightly different from abstracts due to update in dataset.

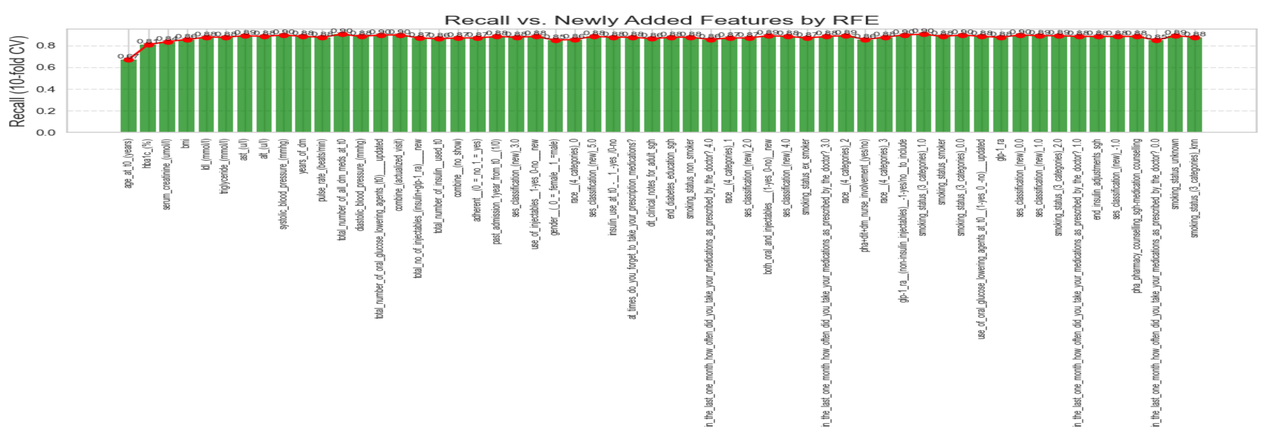

Figure 3. RFE Plot for Recall (Random Forest)

- The final seven features to be included for potential mobile application: age, body mass index, smoking status, baseline HbA1c at T0, serum creatinine, triglycerides, and low-density lipoprotein.

Table 3. Comparison of model performance based on 7 features for a mobile application-based AI model

Model	AUC 10 Fold CV Mean	Median Test AUC 95% CI (10,000 bootstraps)	Median Test RECALL 95% CI (10,000 bootstraps)	Median Test F1 95% CI (10,000 bootstraps)	Median Test Prec. 95% CI (10,000 bootstraps)	Median Test Accuracy 95% CI (10,000 bootstraps)	Median Test NPV 95% CI (10,000 bootstraps)	Median Test Specificity 95% CI (10,000 bootstraps)
Random Forest	0.809	0.836 [0.776, 0.890]	0.898 [0.846, 0.944]	0.854 [0.808, 0.893]	0.814 [0.752, 0.872]	0.803 [0.750, 0.851]	0.776 [0.671, 0.872]	0.634 [0.530, 0.737]
XGBoost	0.777	0.792 [0.728, 0.851]	0.856 [0.796, 0.911]	0.814 [0.764, 0.858]	0.776 [0.711, 0.838]	0.750 [0.693, 0.803]	0.688 [0.569, 0.795]	0.561 [0.452, 0.667]
LightGBM	0.787	0.807 [0.744, 0.864]	0.864 [0.806, 0.917]	0.829 [0.780, 0.873]	0.798 [0.733, 0.860]	0.772 [0.715, 0.825]	0.714 [0.606, 0.818]	0.610 [0.505, 0.716]

CONCLUSION

- Random Forest model demonstrated high performance in predicting uncontrolled T2DM.
- Machine learning techniques are promising to build accurate models to forecast disease outcomes and provide large-scale personalized person-centered pharmacy care.

REFERENCES

1. S Seah & C Yang. War against diabetes: Doctors seeing rise in patients below 40 due to lifestyle habits, early screening. Channel NewsAsia.
2. NA ElSAyed, et al. Glycemic Targets: Standards of Care in Diabetes 2023. Diabetes Care 2023;46(Suppl 1):S97-S110,
3. Gehi AK. Self-reported medication adherence and cardiovascular events in patients with stable coronary heart disease: The heart and soul study. Arch Intern Med 2007; 167: 1798-1803.
4. Rolfson DB, et al. Validity and reliability of the Edmonton Frail Scale. Age Ageing 2006; 35: 526-529.

PATIENTS. AT THE HEART OF ALL WE DO.®

SingHealth DukeNUS ACADEMIC MEDICAL CENTRE

Singapore General Hospital · KK Women's and Children's Hospital · Sengkang Health · National Cancer Centre Singapore · National Dental Centre Singapore · National Heart Centre Singapore · National Neuroscience Institute · Singapore National Eye Centre · Polyclinics SingHealth · Bright Vision Hospital