Developing a Personalized Clinical Decision Support System for Statin Therapy for Primary Prevention using OMOP-CDM and Deep Learning Techniques

Su Min Kim[1], Ju-Hyeon Kim[2], Yunjin Yum[3], Eunbeen Jo[1], Jose Moon[1], Jong-Ho Kim[2,4], Yong Hyun Kim[5], Eung Ju Kim[6], Hyung Joon Joo[1,2,4]

[1]Department of Medical Informatics, College of Medicine, Korea University, Seoul, Republic of Korea; [2]Department of Cardiology, College of Medicine, Korea University, Seoul, Republic of Korea; [3]Department of Biostatistics, Korea University College of Medicine, Seoul, Republic of Korea; [4]Korea University Research Institute for Medical Bigdata Science, College of Medicine, Korea University, Seoul, Republic of Korea; [5]Division of Cardiology, Department of Internal Medicine, Korea University Ansan Hospital, Seoul, Republic of Korea; [6]Division of Cardiology, Department of Internal Medicine, Korea University Guro Hospital, Seoul, Republic of Korea

Background

Dyslipidemia is considered a major risk factor for cardiovascular disease (CVD). Statins are the cornerstone of dyslipidemia treatment. However, their clinical use may be hindered by side effects like myalgia and an increased risk of diabetes[1]. Current guidelines suggest specific target low-density lipoprotein cholesterol (LDL-C) levels for statin therapy based on patient characteristics but may not account for individual variability in statin response or tolerability[2]. This highlights the need for personalized approaches in statin therapy. This study aimed to develop a deep learning algorithm utilizing OMOP-CDM data to recommend the optimal statin therapy to achieve target LDL-C levels based on individual cardiovascular risk in primary prevention (Figure 1).

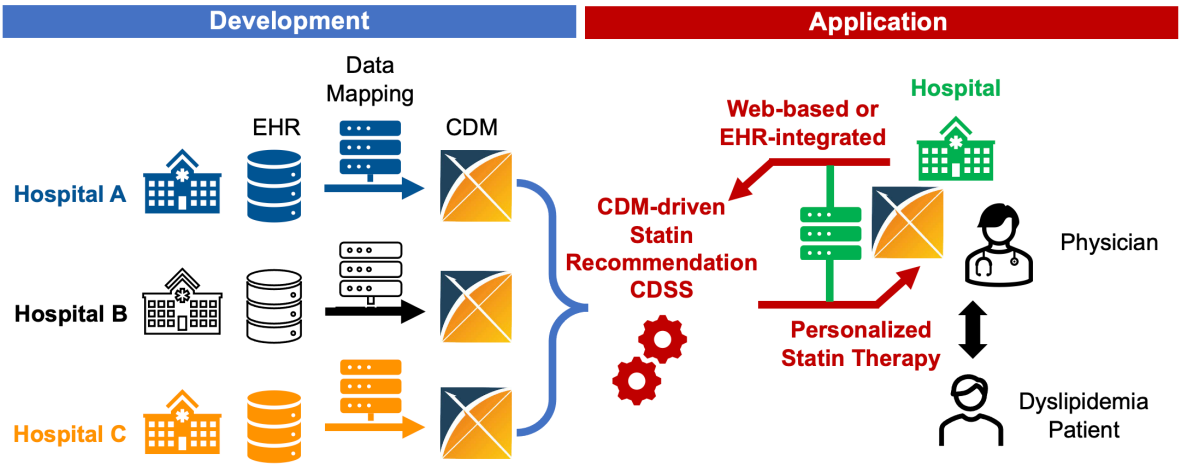

Figure 1. Development and application of CDM-driven statin recommendation CDSS for dyslipidemia patients

Methods

We extracted data from patients initiating statin therapy for the first time between 2003 and 2022 from the electronic health record databases of three tertiary university hospitals. Exclusions were patients with overt atherosclerotic cardiovascular disease, those without direct LDL-C measurements, and those unlikely to adhere to statin therapy. A total of 22,447 patient datasets including demographic, clinical information, statin prescription data, diagnostic information, and laboratory data were used. The extracted data were mapped to the OMOP-CDM. This includes mapping local codes to standard OMOP-CDM concepts, and storing them in the appropriate tables (Person, Observation, Drug_exposure, Drug_strength, Condition_occurrence, and Measurement tables). a

total of 18,073 patients from two hospitals and 4,374 patients from another hospital were allocated to the development dataset and the external validation dataset, respectively. Multiple predictive models were developed using the development dataset employing multiple linear regression, KNN, SVM, Random Forest, XGB, and Multi-Layer Perceptron (MLP) with 5-fold cross-validation.

Results

The study findings revealed that a significant proportion of initial statin prescriptions did not achieve the desired low-density lipoprotein cholesterol (LDL-C) target level, highlighting the limitations of empirical prescription practices. Feature importance analysis identified diabetes mellitus, SCORE2/SCORE2-OP, baseline LDL-C level, Hba1c level, and age as the critical features for achieving target LDL-C levels with statin therapy (Figure 2).

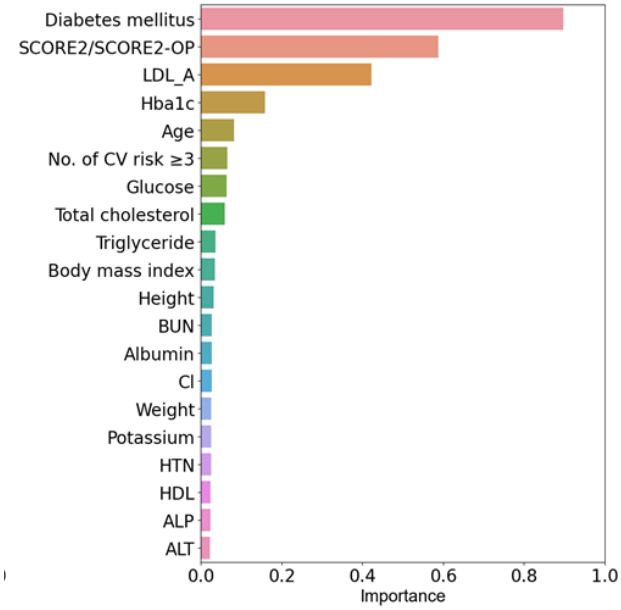

Figure 2. Feature importance

demonstrated the highest predictive performance. In the 5-fold cross-validation, the XGB model exhibited an Area Under the Receiver Operating Characteristic (AUROC) of 0.84, accuracy of 0.86, precision of 0.80, recall of 0.52, and F1-score of 0.63 (Table 1).

Table 1. Feature importance and model performance

Dataset	Model	Accuracy	Precision	Recall	F1-score	AUROC
5-fold cross validation	MLR	0.82 ± 0.01	0.65 ± 0.02	0.47 ± 0.02	0.54 ± 0.02	0.79 ± 0.00
	KNN	0.79 ± 0.01	0.54 ± 0.03	0.40 ± 0.03	0.46 ± 0.03	0.73 ± 0.02
	SVM	0.83 ± 0.01	0.70 ± 0.01	0.46 ± 0.02	0.55 ± 0.01	0.79 ± 0.01
	Random Forest	0.85 ± 0.01	0.80 ± 0.02	0.49 ± 0.03	0.60 ± 0.03	0.82 ± 0.01
	XGBoost	0.86 ± 0.01	0.80 ± 0.02	0.52 ± 0.03	0.63 ± 0.02	0.84 ± 0.01
	MLP	0.85 ± 0.01	0.75 ± 0.01	0.48 ± 0.03	0.59 ± 0.02	0.82 ± 0.01
External validation	MLR	0.79	0.67	0.39	0.49	0.78
	KNN	0.78	0.60	0.44	0.51	0.73
	SVM	0.80	0.69	0.39	0.50	0.74
	Random Forest	0.82	0.79	0.42	0.55	0.82
	XGBoost	0.82	0.71	0.48	0.57	0.81
	MLP	0.84	0.80	0.49	0.60	0.83

In the external validation dataset, the XGB model achieved an AUROC of 0.81, accuracy of 0.82, precision of 0.71, recall of 0.48, and F1-score of 0.57. By implementing the XGB model in clinical

practice, the probability of attaining the target LDL-C level through the initial statin prescription could be increased to a range of 71% to 80%.

Conclusion

The results highlight the potential of leveraging standardized OMOP-CDM data and deep learning techniques for personalized statin therapy in dyslipidemia. The scalable CDSS platform developed in this study can be readily implemented, considering the increasing adoption of the OMOP-CDM schema and vocabulary in healthcare institutions. Future work will focus on prospective validation of the algorithm's effectiveness and enhancing scalability by upgrading the platform based on HL7/FHIR.

Reference

1. Thompson PD, Panza G, Zaleski A, Taylor B. Statin-Associated Side Effects. J Am Coll Cardiol. 2016 May 24;67(20):2395-2410.
2. Visseren FLJ, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, Benetos A, Biffi A, Boavida JM, Capodanno D, Cosyns B, Crawford C, Davos CH, Desormais I, Di Angelantonio E, Franco OH, Halvorsen S, Hobbs FDR, Hollander M, Jankowska EA, Michal M, Sacco S, Sattar N, Tokgozoglu L, Tonstad S, Tsioufis KP, van Dis I, van Gelder IC, Wanner C, Williams B; ESC National Cardiac Societies; ESC Scientific Document Group. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J. 2021 Sep 7;42(34):3227-3337.