

Predicting Cardiovascular Diseases Risk in Thai Population by Machine Learning

Gumpanart Veerakul¹; Kitsuchart Pasupa²; Yod Pinroj³; Lertlak Chaothawee⁴; Pradit Somprakit⁵; Pongtorn Kietdumrongwong⁶; Somkiat Tonphu⁶; Warut Chaiwong⁶, Saowaluck Yasri⁶



Gumpanart Veerakul

¹ Cardiovascular Disease Prevention, Bangkok Heart Hospital, Bangkok, Thailand

² School of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

³ The Bangkok Neuroscience Center, Bangkok International Hospital, Bangkok, Thailand

⁴ Cardiac Imaging Center, Bangkok Heart Hospital, Bangkok, Thailand

⁵ Senior Chief Information Officer Group 1; Managing Director of Greenline Synergy Co., Ltd. Bangkok Heart hospital, Bangkok, Thailand

⁶ BDMS Health Research Center, Bangkok Dusit Medical Services PCL., Bangkok, Thailand

* Address Correspondence to author:
Gumpanart Veerakul, MD
Cardiovascular Disease Prevention,
Bangkok Heart Hospital,
2 Soi Soonvijai 7, New Petchburi Rd.,
Bangkok 10310, Thailand
email: gumprevention@gmail.com

Received: May 27, 2025

Revision received: May 29, 2025

Accepted after revision: September 8, 2025

BKK Med J 2025;21(2): 133-144.

DOI: 10.31524/bkkmedj.2025.21.006

www.bangkokmedjournal.com

Abstract

OBJECTIVES: To establish a clinical data lake for artificial intelligence (AI) and develop a machine learning model to predict cardiovascular disease (CVD) risk.

MATERIALS AND METHODS: Following IRB approval, de-identified clinical data from 2.9 million patients (2010–2019) across eight Bangkok Dusit Medical Services (BDMS) hospitals were collected in compliance with the Personal Data Protection Act (PDPA). Two datasets were constructed: BDMS-CVD-large (n = 9,072), comprising 3-year clinical records with 20 SHAP-selected features plus age and sex; and BDMS-CVD-Small (n = 107), incorporating coronary artery calcium scores (CACS) and time-from-test. XGBoost models were trained using 5-fold cross-validation, grid search, and repeated across 10 random splits.

RESULTS: The BDMS-CVD-Large model achieved strong performance (F_1 -score: Macro 0.93 ± 0.008 ; Weighted 0.97 ± 0.003), with age, HDL, and LDL as key predictors. Including CACS improved the F1-score (0.92 ± 0.032 vs. 0.87 ± 0.031), confirming its value. Limitations included potential occult CVD, exclusion of over 40% of cases due to incomplete data, and missing longitudinal data in many patients.

CONCLUSION: This study demonstrates the feasibility of machine learning (ML) based CVD prediction using large-scale clinical data under PDPA compliance. Prospective validation over 5–10 years is warranted, and integrating CACS may enhance future predictive accuracy.

Keywords: cvd risk prediction, Thais, supervised machine learning

Data from Thailand Health Statistics 2022 indicated that cardiovascular diseases (CVD), leading by stroke and coronary artery disease (CAD), remained the second major causes of death after neoplasm.¹ From 2018-2022, the mortality rate continuously rose from 47.1 to 58 per 10⁶ for stroke and 31.8 to 35.1 per 10⁶ for CAD, which reflected an ineffective CVD prevention program. In contrast, CVD mortality in US and Europe, had been declined and major contribution, 44-76%, came from controlling CVD risk factors, not treatment.^{2,3} Therefore, the western preventive CVD guidelines consistently recommended to screen asymptomatic cases at age of 40 or beyond, by applying their own systemic risk score, i.e., the SCORE⁴ or atherosclerotic cardiovascular disease (ASCVD) score⁵ to reduce CVD risk in future.

For Thai population, the CVD risk score (modified Coronary Risk Chart, available at <https://www.caprecvdrisk.com>), an only 10-yr prospective study, had been proven for better prediction of future CVD events in asymptomatic Thais,⁶ comparing with other scores, i.e., Thai CV⁷ and WHO score⁸. In addition, very low predictive CVD risk of Thai CV score was reported by Na Witayanun and colleagues⁹ who analyzed 2,020 ST segment elevation myocardial infarction Thais, and found that only 5.2% among heart attack victims were a very high group. These predictive variations came from various factors including percent loss of studied population, defining different clinical outcome, risk factor definitions and methodologies, i.e. prospective or retrospective study.⁶⁻⁹

Currently, the large data analysis allowed us to explore the more complex relationship of old and new risk factors which might improve CVD prediction. Recently, big data analysis by machine learning (ML) had been introduced with more promising CVD prediction.¹⁰⁻¹² However, no ML had never been used in Thai population, so we initiated this pilot project to study the feasibility and limitations of ML for CVD prediction in Thais by using clinical data from eight Bangkok Dusit Medical Services (BDMS) chain hospitals. The two main objectives were to create a necessary data lake to support artificial intelligence and to develop a predictive model for predicting future CVD risk by using machine learning.

Material and Methods

After IRB approval, demographic, laboratory, radiological results, treatment, clinical diagnosis (disease registry) and

vital signs were anonymously retrieved between January 1, 2010 to December 31, 2019 including 18,505,467 service episodes of 2,969,401 cases, from eight BDMS chain hospitals (appendix 1).

These data had been safely handling and secured according to the Joint Commission International, ISO27001 for obtaining consent, data storage, extraction for research and monitoring according to Personal Data Protection Act BE2562, by Data Protection Officer (DPO) under Privacy Policy and record all data assembly and activity, so-called RoPA (The Record of Processing Activities) in separate server and back-up system, SSL/TLS (Secure Socket Layers / Transport Layer Security), see Figure 1 and 2.

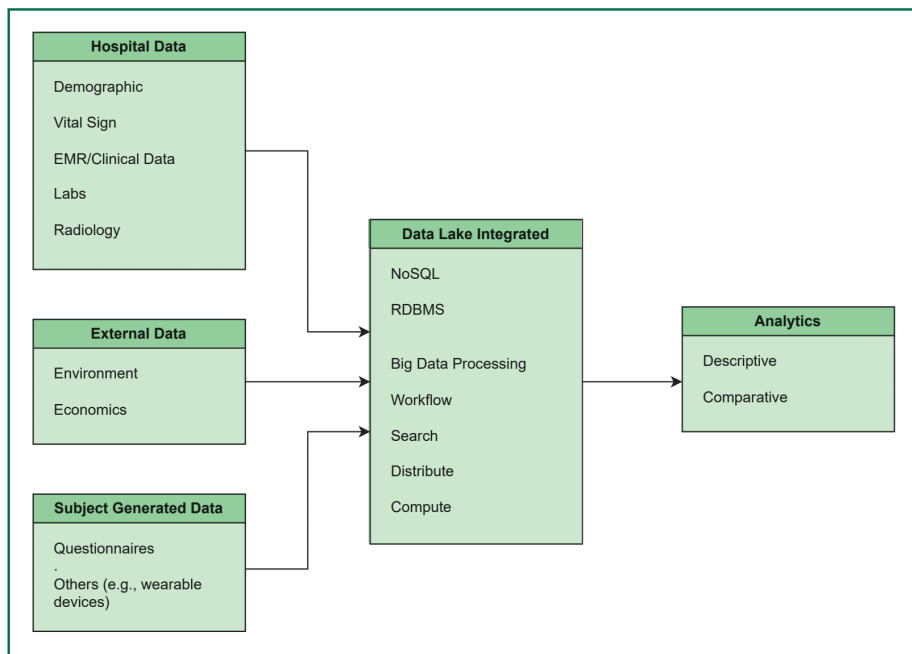


Figure 1: Flow and details of retrieval data

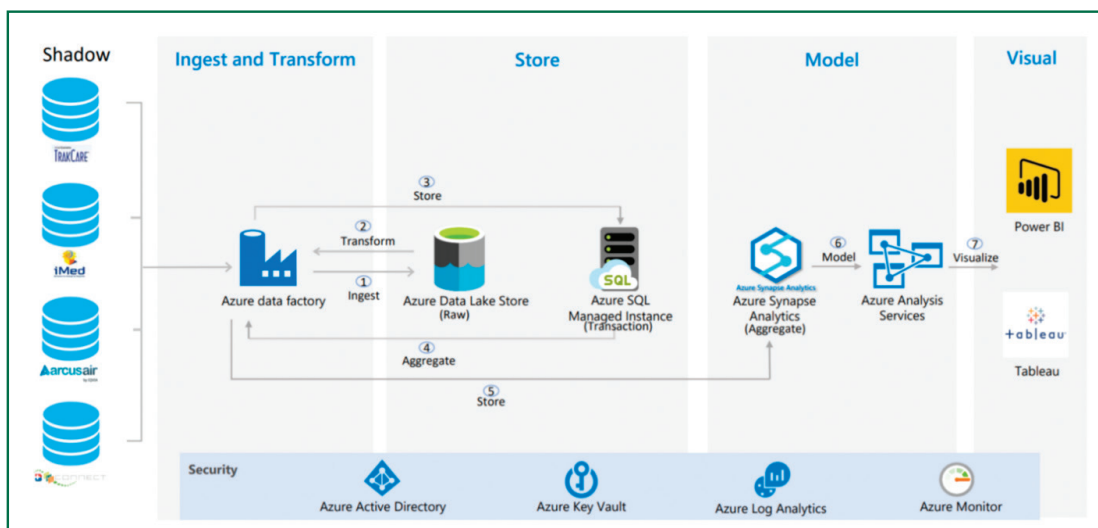


Figure 2: Multiple steps of data retrieval, transform, store and building model

Currently the calcium score (CACS) is widely accepted for individual risk assessment¹³⁻¹⁵, however, its predictive role in Thais remains unclear.¹⁶ To address this, we constructed two datasets: one with common measures excluding the calcium score and another incorporating it.

The details are as follows.

BDMS-CVD-Large dataset – data without calcium score

The data were cleaned through a series of preprocessing steps. First, similar data were merged to ensure consistency. Then, outliers were removed by excluding percentile values

below 0.5% and above 99.5% for each feature, in order to minimize model deviations (see Figure 3 for an example).

Next, filtering was applied due to the large volume of test data. The objective was to develop a generalized model with minimal complexity. To achieve this, significant and discriminative features were prioritized, identified through the mean absolute SHapley Additive exPlanations (SHAP) value for each feature, see Figure 4, obtained from Extreme Gradient Boosting (XGBoost) to classify patients, averaged across 10 models. Subsequently, only 20 features with the highest SHAP values, with age and gender were considered in the later analysis.

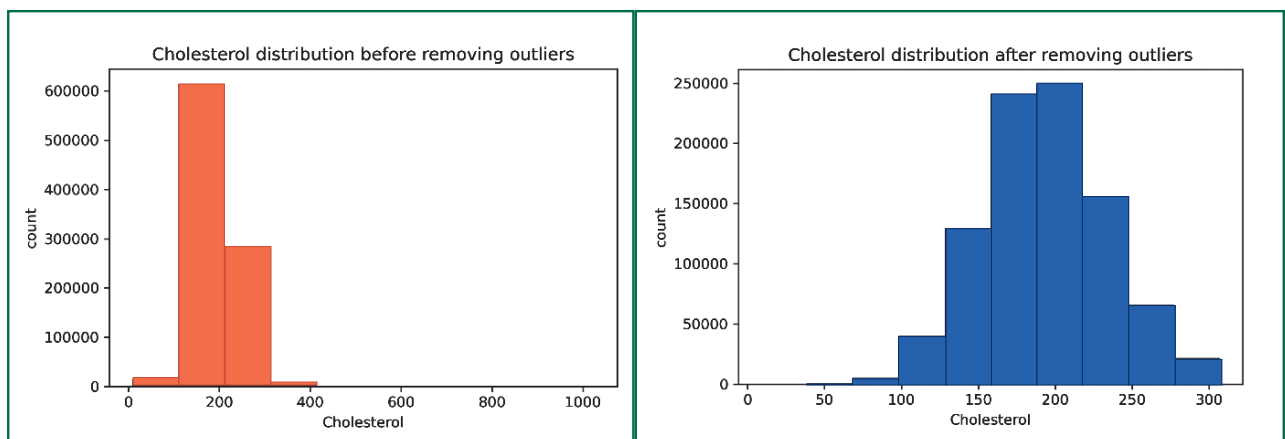


Figure 3: Sample histogram before (left) and after (right) cleaning by removing outliers of data (cholesterol level)

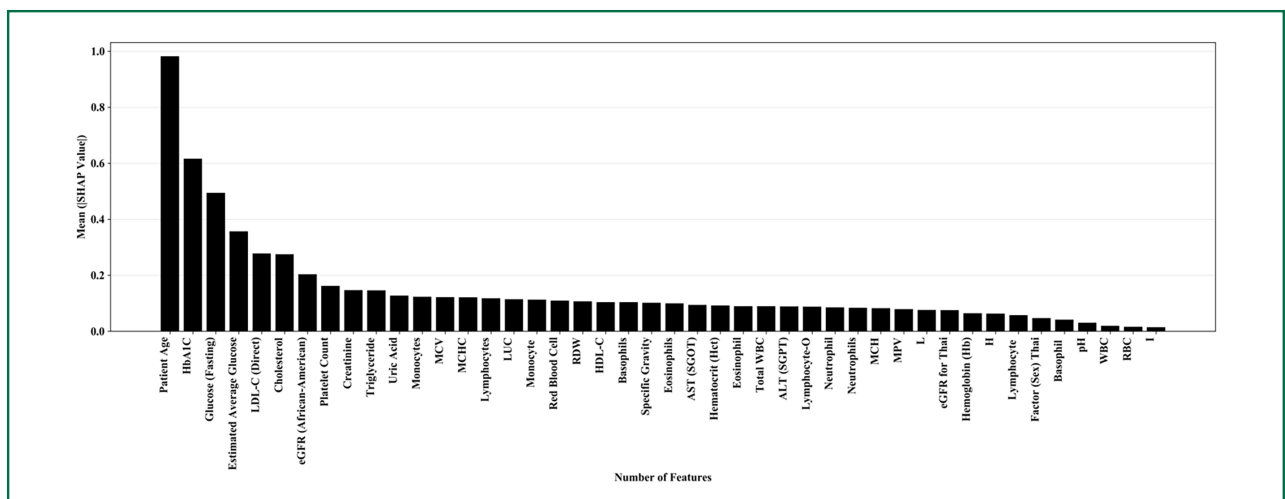


Figure 4: Mean absolute SHAP (SHapley Additive exPlanations) values of all parameters from XGBoost, averaged across 10 models

Then, data extraction was conducted to build the predictive model by selecting Thai cases with clinical and laboratory data spanning over 3 years, resulting in a total of 220,121 individuals. For each individual, the average test results for each year were calculated in cases where multiple test results were available. Subsequently, a design matrix was created, incorporating test results from the past two years (see sample in Figure 5, e.g., BMI).

BMI _t	BMI _{t-1}	BMI _{t-2}
30.10	30.1	30.74
...	...	...
24.86	24.93	24.86

Figure 5: Sample of extracted and selected data (BMI) with three consecutive years

The developed predictive supervised ML model is a classification task, necessitating an output variable that corresponds to the input data from the design matrix. This output variable is categorized into two types: positive cases, defined as individuals who have had three consecutive years of negative diagnostic results—specifically, no identification of coronary artery disease (ICD codes I20-I25) or cerebrovascular diseases (ICD codes I60-I69), see Table 1—and subsequently develop a positive diagnosis within the next five years; and negative cases, referring to individuals who have had three consecutive years of negative diagnostic results and remain negative within the next five years.

The dataset under consideration exhibited a significant class imbalance, with a substantially higher number of negative cases (non-diseased individuals) compared to positive cases (diseased individuals). To address this imbalance and preserve as much information from the positive cases as possible, a systematic approach was implemented.

Initially, individuals with a positive diagnosis were selected, focusing on those confirmed to have the disease. Subsequently, patients within this group who had more than 40% missing data in the selected 20 features were excluded to ensure data quality and reliability. Following this, features that still exhibited over 20% missing data, even after the prior exclusions, so they were removed from the dataset. To handle the remaining missing values, the k-means algorithm with $k = 5$ was employed for data imputation.

After processing the positive cases, attention shifted to the negative cases. Data from this group were selected, similar to the previous steps, patients with more than 40% missing data were excluded. The features chosen during the positive case processing were then applied to the negative cases to maintain consistency across the dataset. Missing values within this group were also addressed using the k-means algorithm with $k = 5$, ensuring uniformity in data imputation methods. Finally, the datasets of positive and negative cases were combined, resulting in a comprehensive dataset comprising 9,072 individuals, as detailed in Table 2.

Table 1: The distribution of the top 10 CVD cases identified using ICD code I20-25 and I60-69 in the dataset.

#	ICD Descriptions	Number
1	Atherosclerotic heart disease	190,465
2	Sequelae of stroke, not specified as hemorrhage or infarction	31,198
3	Sequelae of cerebral infarction	28,322
4	Chronic ischemic heart disease, unspecified	16,859
5	Cerebral infarction due to thrombosis of cerebral arteries	12,400
6	Stroke, not specified as hemorrhage or infarction	10,568
7	Cerebral infarction, unspecified	9,813
8	Sequelae of intracerebral hemorrhage	5,832
9	Sequelae of other and unspecified cerebrovascular diseases	5,721
10	Ischemic cardiomyopathy	3,890

Table 2: The dataset used to create the model in this study comprises laboratory data collected over three consecutive years for each case.

Description	Details
Total selected cases	9,072
Total variables	27
% +CVD	9.5%
Laboratory data available for 3 consecutive years	Monocytes Lymphocytes Platelet count Glucose Cholesterol LDL-C Triglyceride HDL-C
Demographic data	Age Gender

BDMS-CVD-Small dataset – data with calcium score

We incorporate CACS data from the Thai population to enhance risk prediction models. The dataset consists of 6,350 individuals who underwent screening. We then remove outliers based on the data preprocessing criteria established in previous experiments.

Next, we integrated the newly processed data into the previous dataset. This resulted in a refined dataset containing individuals with laboratory test results spanning over three years and calcium scores recorded within the study period and up to two years prior. The final dataset included 107 individuals, comprising 55 diagnosed cases projected to develop the disease within 1-5 years and 52 non-disease cases. Additionally, we introduced a new feature representing the time of calcium score measurement, indicating the number of years elapsed since the test was conducted relative to the present year.

Experimental Framework

XGBoost was used as the study model due to its high performance with complex datasets and its capacity for interpretability. The dataset was divided into a training set and a test set in a 60:40 ratio, ensuring that each subset maintained the original proportion of positive and negative cases. Model training involved selecting the most appropriate model through five-fold cross-validation combined with grid search, using the F1-score as the evaluation metric, which is suitable for imbalanced data. The model's performance was then assessed using the test set. To mitigate sampling bias, the experiment was conducted ten times, each with different random splits of the training and test sets. The development was carried out using Scikit-learn.¹⁷ The experimental framework is illustrated in Figure 6.

Results

BDMS-CVD-Large dataset

We evaluated the model's performance using a range of metrics, including Precision, Recall, and F_1 -score, on the test set, as presented in Table 3. To provide a comprehensive assessment, we report both the Macro Average and Weighted Average F_1 -scores.

The Macro Average F_1 -score, calculated as the unweighted mean of the F1-scores for all classes, reflects the model's ability to perform equally well across all classes. By treating each class with equal importance, this metric is particularly valuable for evaluating performance on underrepresented (minority) classes. The model achieved a Macro Average F_1 -score of 0.93 ± 0.008 , demonstrating its capacity to maintain balanced performance across both majority and minority classes.

The Weighted Average F1-score, on the other hand, incorporates the relative size (support) of each class, thereby reflecting the model's overall performance weighted by the prevalence of each class in the dataset. While this metric highlights the model's strong performance on the majority classes, it may somewhat overshadow performance on the minority class. Despite this, the Weighted Average F_1 -score of 0.97 ± 0.003 indicates the model's overall robustness.

We further analyzed the Receiver Operating Characteristic (ROC) curves for all 10 experimental runs, as illustrated in Figure 7. The results demonstrate the model's effectiveness in performing the task. Additionally, the optimal parameters of the XGBoost algorithm that yielded the highest F_1 -score are detailed in Table 4.

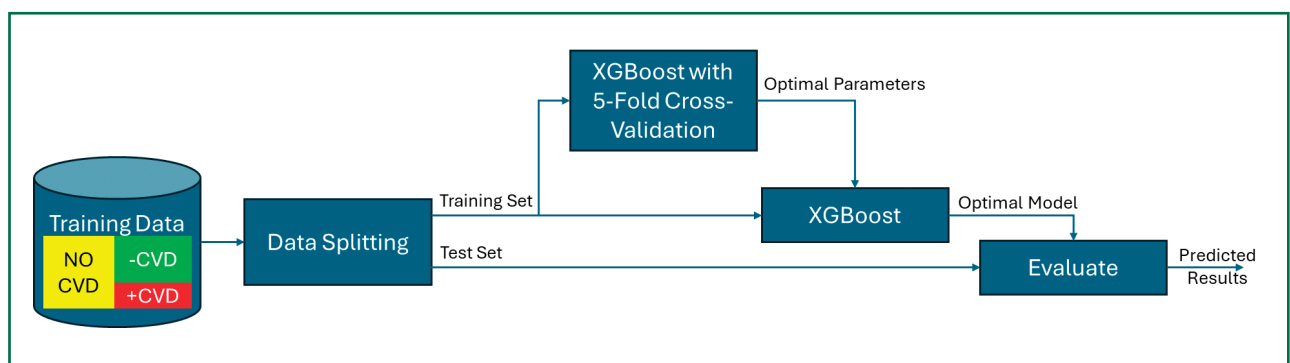


Figure 6: Experimental Framework -- Data Splitting, Parameter Tuning, Training, and Testing. The experiments were conducted 10 times with different random splits.

Table 3: Model performances on the test set, averaged across 10 runs on BDMS-CVD-Large dataset.

	Precision	Recall	F_1 -score	Support
Negative Class	0.98 ± 0.002	0.99 ± 0.001	0.99 ± 0.001	3599 ± 15.95
Positive Class	0.96 ± 0.012	0.80 ± 0.020	0.87 ± 0.014	354 ± 15.95
Macro Average	0.97 ± 0.006	0.90 ± 0.010	0.93 ± 0.008	3593
Weighted Average	0.98 ± 0.002	0.98 ± 0.002	0.97 ± 0.003	3593

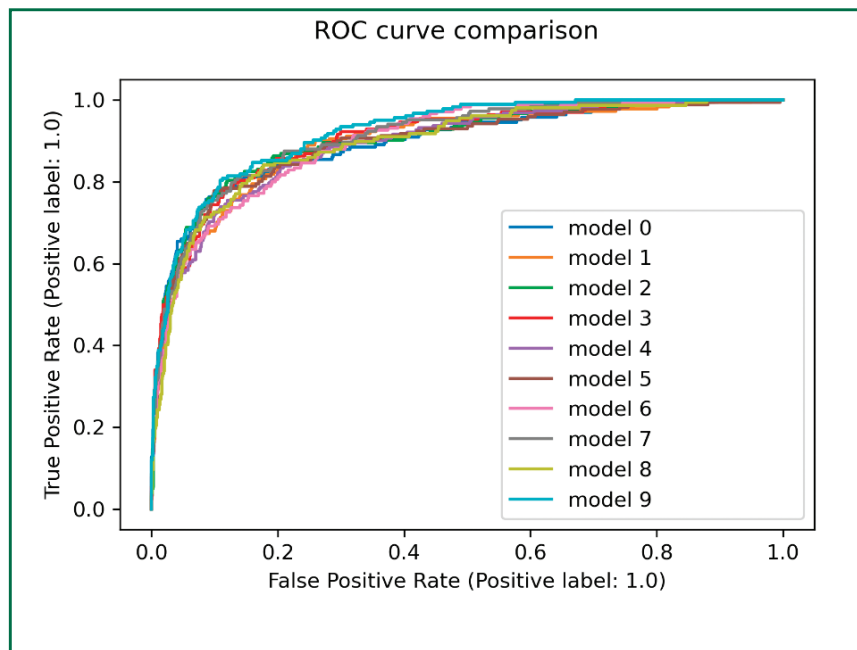


Figure 7: ROC curves of ten experiments

Additionally, we demonstrated the importance of the features used in the model by computing SHAP values, as shown in Figure 8. It can be observed that age is the most significant feature for prediction, followed by HDL and LDL.

BDMS-CVD-Small dataset

We then conducted experiments following the same framework as the previous study to compare the performance of models with and without calcium score information. The results are presented in Table 5. The findings indicate that incorporating calcium scores improves the overall model performance, achieving an F_1 -score of 0.92 ± 0.032 , compared to an F_1 -score of 0.87 ± 0.031 when calcium score data is excluded.

We analyzed feature importance using SHAP values for models with and without calcium scores, as shown in Figure 9. The results indicate that the calcium score is the most important feature in models incorporating it, whereas age is the most significant factor in models that do not use the calcium score.

Table 4: The optimal parameters of XGBoost.

Parameters	Value
Max Depth	6
Number of Estimators	1000
Subsample	0.7
Subsample Ratio of Columns by Tree	0.5
Learning Rate	0.1

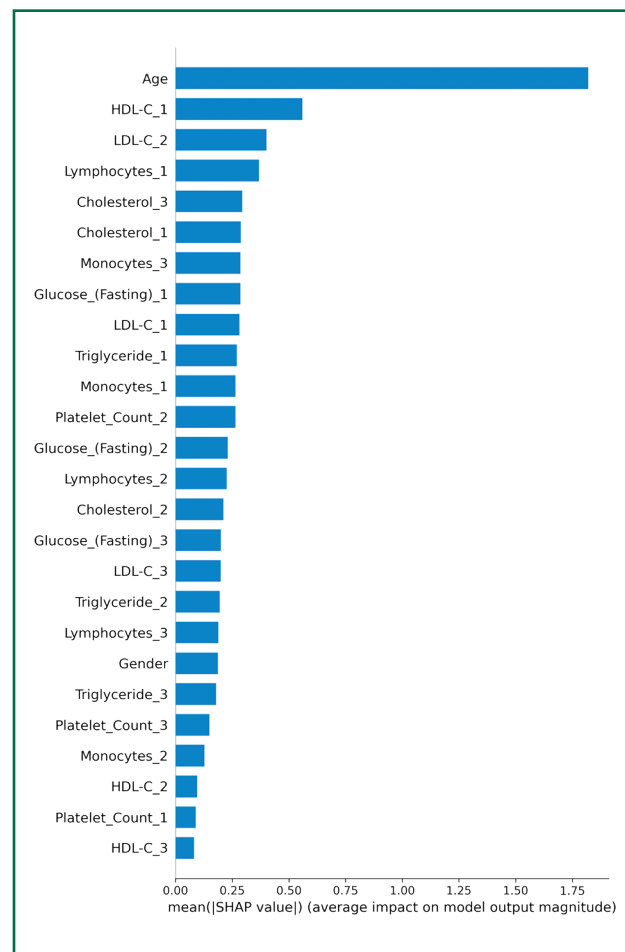


Figure 8: Order of selected parameters by SHAP from maximal to minimal significance

Table 5: Model performance on the test set, averaged over 10 runs, comparing results with and without calcium score on BDMS-CVD-Small dataset.

Dataset	Class	Precision	Recall	F1-score	Support
BDMS-CVD dataset	Negative	0.87 ± 0.073	0.85 ± 0.068	0.86 ± 0.043	18.4 ± 1.96
	Positive	0.89 ± 0.045	0.89 ± 0.056	0.89 ± 0.022	22.5 ± 1.96
	Macro Average	0.88 ± 0.033	0.88 ± 0.030	0.87 ± 0.031	41
	Weighted Average	0.88 ± 0.028	0.88 ± 0.030	0.88 ± 0.029	41
BDMS-CVD-CACS dataset	Negative	0.91 ± 0.041	0.91 ± 0.077	0.91 ± 0.041	18.5 ± 1.96
	Positive	0.93 ± 0.056	0.93 ± 0.036	0.93 ± 0.025	22.5 ± 1.96
	Macro Average	0.92 ± 0.030	0.92 ± 0.034	0.92 ± 0.032	41
	Weighted Average	0.93 ± 0.030	0.92 ± 0.030	0.92 ± 0.030	41

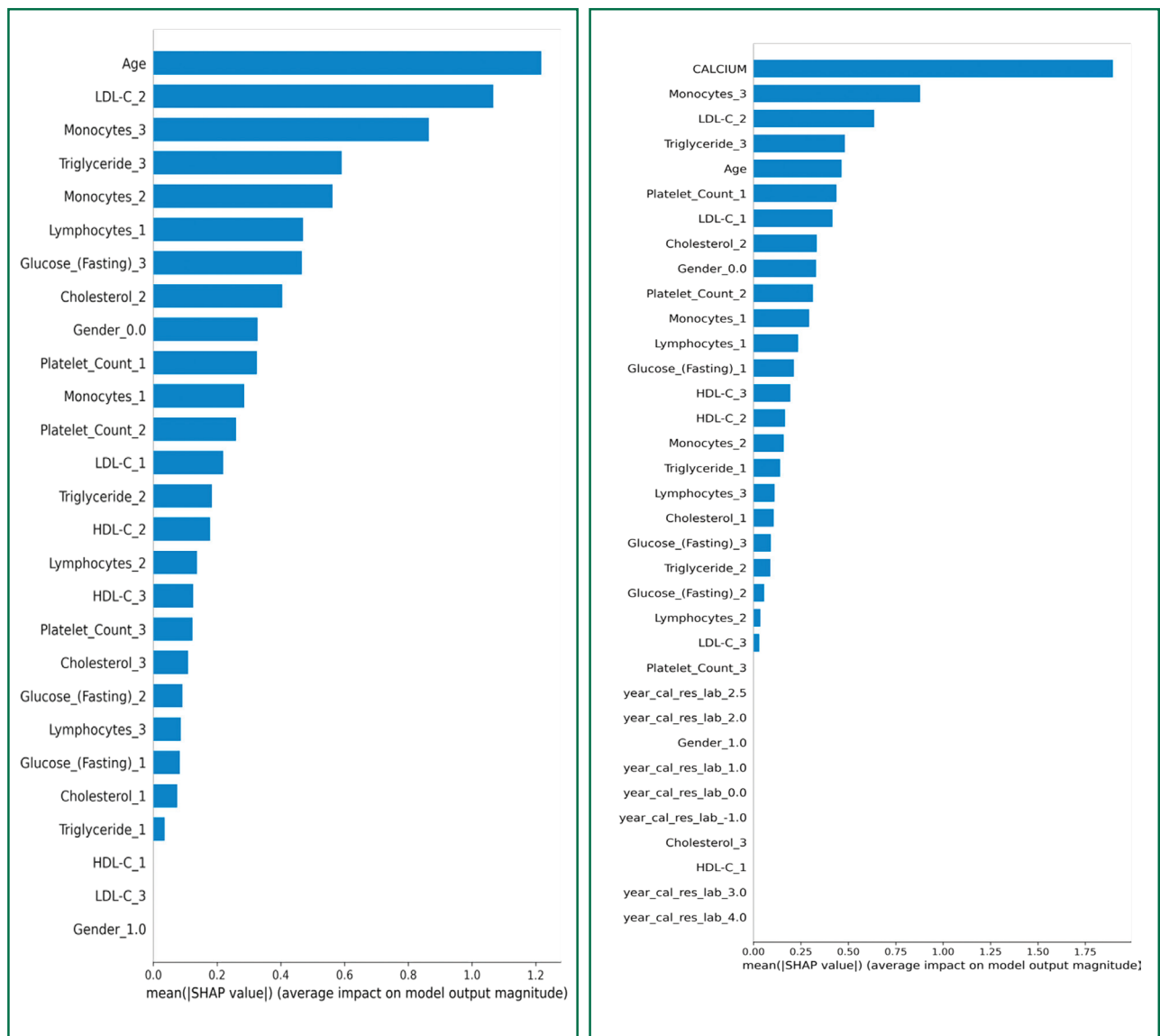


Figure 9: Order of selected parameters by SHAP from maximal to minimal significance on BDMS-CVD-Small, comparing results without (left) and with calcium scores (right)

We visualized the model's decision-making process by constructing a decision tree, allowing us to understand the criteria used for classification on the models with and without calcium score information, as shown in Figures 10 and 11, respectively.

Discussion

CVD situation in Thailand

CVD remained the second cause of death in Thai population and the trend was getting worse. From 2018-2022, with no effective prevention, the mortality rate from stroke and CAD in Thais had risen by 23.2% and 10.5%, respectively.¹ Western CVD prevention guidelines encouraged to stratify future CVD risk in individuals beginning at 40 years for prevention, by their native score, i.e., the SCORE⁴ for European, and ASCVD score⁵ for US citizen. However, these available systemic risk scores for CVD prediction have several limitations when applying to different population. Most of them are based on the known CVD risk factors in specific studied population, using various clinical outcomes, i.e. fatal or non-fatal CVD and different methodologies, which could provide both under and over CVD prediction.^{9,18}

Can machine learning (ML) improve CVD risk prediction in Thais?

Using a large data set by ML allowed us to see more complex interaction of both old and new risk factors in non-linear fashion, beyond conventional analysis, i.e.,

log-regression and might improve CVD prediction.¹⁰⁻¹² Weng and colleagues prospectively compared four machine-learning algorithms (random forest, logistic regression, gradient boosting machines, neural networks) by using routine clinical data of 378,256 patients from UK family practices, with no cardiovascular disease at the outset, to predict both fatal and non-fatal CVD outcome.¹⁰ After 12 months, 24,970 incident cardiovascular events (6.6%) were observed and it was neural network that provided the best prediction with sensitivity of 67.5%, positive predictive value of 18.4%, specificity of 70.7% and negative predictive value of 95.7%. They concluded that ML significantly improves accuracy of cardiovascular risk prediction, increasing the number of patients identified who could be benefit from preventive treatment, while avoiding unnecessary treatment of others. Since there was no ML prediction CVD for Thais, we initiated this retrospective pilot project to explore the feasibility and limitations of using ML techniques, specifically XGBoost, due to its clear interpretability process.

The roles of non-conventional risk factors

In this study, ten variables that were available for three consecutive years, including age, HDL-cholesterol, LDL-cholesterol, lymphocytes, total cholesterol, monocytes, glucose (fasting), triglyceride, and platelet count, were utilized with ML model, XGBoost, see Table 2. The CVD prediction model was created, based on negative CVD cases to become positive CVD cases in the next five years. From our understanding, this is the first ML model created for Thais, to predict future CVD within the next five years.

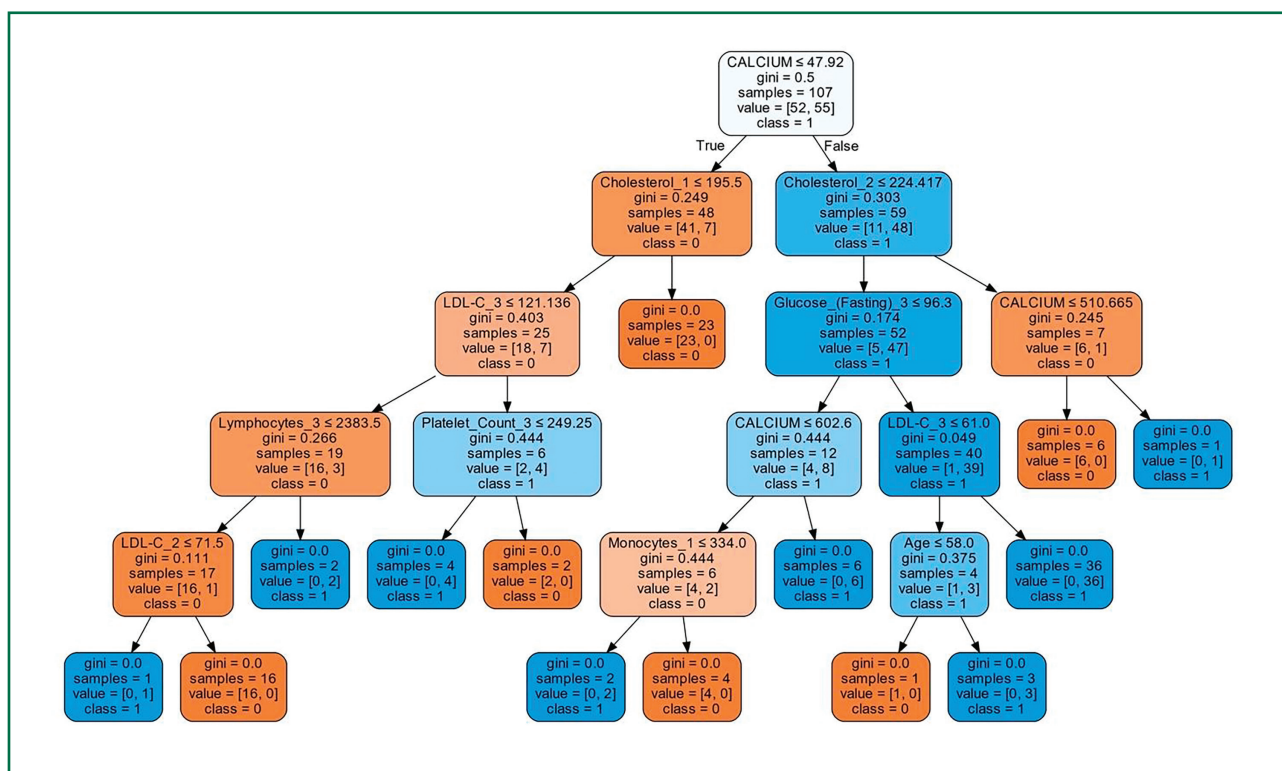


Figure 10: The binary decision tree model with calcium score information

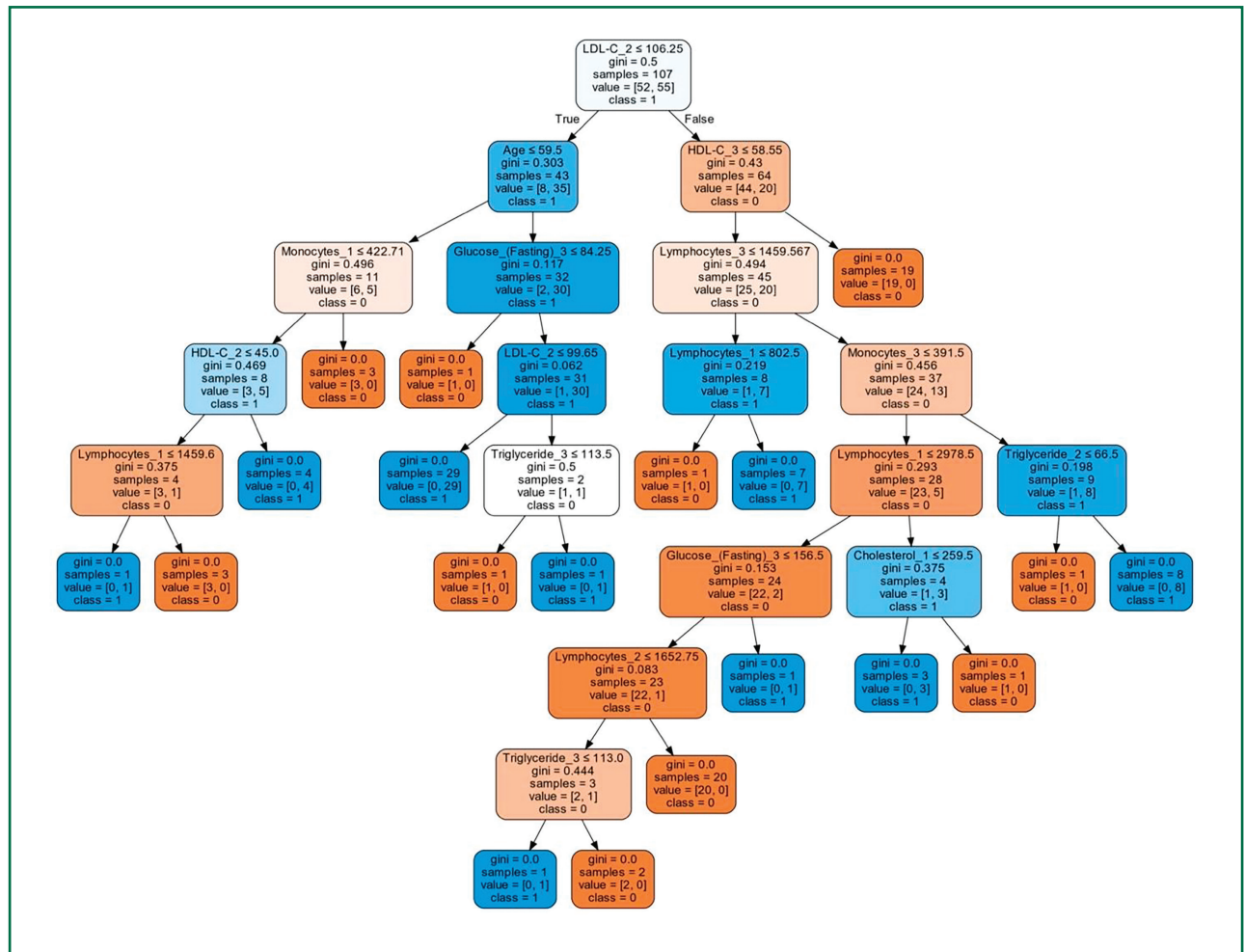


Figure 11: The binary decision tree model without calcium score information

By machine selection, based on SHAP values, both circulating platelets, monocytes and lymphocytes were included for prediction. Although all of them were not considered as the traditional CVD risk factors, all of them had played pivotal roles for both atherogenesis and regression.¹⁹⁻²⁵

Platelet roles involved for both pro-thrombotic and pro-inflammatory process in heart attack cases. Decreases in the total and relative number of circulating lymphocytes have been observed during heart attack.¹⁹ Larger platelet particles contained more granules releasing prothrombotic factors (i.e. P-selectin, serotonin, ADP, and b-thrombo-globulin) and thromboxane A2 (TxA2) expressing more adhesion factors, i.e., GP Ib, IIb, IIIa.²⁰ Systematic review and meta-analysis had shown an elevated mean platelet volume (MPV) was statistically significant associated with acute myocardial infarction mortality after myocardial infarction, and restenosis following coronary angioplasty.²¹

The transformed monocytes (macrophage) are involved in both atheromatous plaque progression (M1-type) and regression (M2-type).²² Monocytes subtypes have been shown to be associated with future CVD events²³, stroke²⁴ and heart

attack.²⁵ These non-traditional risk factors needed further studies in our population for potentially predict future CVD risk.

How to apply CACS?

The CACS has been recommended by 2018 ACC/AHA lipid guideline to guide statin therapy when the risk is unclear²⁶ while the prognostic role of CACS in Thais was still uncertain¹⁹. However, our recent study compared the prevalence of CACS in 1,764 CVD Thais vs 2,340 non-CVD Thais, supported the role of CACS prediction in Thais, the higher CACS detected, the more likelihood of having CVD²⁷. Thus, it was not surprised that ML also raised the CACS as the most reliable CVD predictor in Thais.

However, its implication was not quite simple. First, in addition to age and gender difference, racial variation had been previously reported^{28,29} so we better used the CACS chart from Thais, not western data set. By comparing CACS of the non-CVD Thais with asymptomatic western and Koreans, the non-CVD Thais trended to have lower CACS than western but higher than Koreans, Table 6²⁷.

Table 6: Racial difference in asymptomatic US, Koreans and the non-CVD Thais.²⁷

Variables	US adults n (%)	Korean adults n (%)	Thai adults n (%)	p
n	7,533	15,128	2,340	
Age (Mean ± SD)	56.2 ± 7.7	56.2 ± 6.9	58.2 ± 12.1	< 0.001 ^a
Male	4,298 (57.1)	10,398 (68.7)	1,369 (58.5)	< 0.001 ^b
Coronary artery calcium score (CACS)				
0	1,040 (13.8)	7,175 (47.4)	936 (40.0)	< 0.001 ^b
> 0 – 100	3,998 (53.1)	5,688 (37.6)	751 (32.1)	
> 100 - 400	1,784 (23.7)	1,748 (11.6)	349 (14.9)	
> 400	711 (9.4)	517 (3.4)	304 (13.0)	

a: Student t-test, b: Parson's chi square test

Second, more importantly, the relative high prevalence of zero CAC, 14.9%, in the CVD Thais²⁷ did not warrant a safety marker for the low CVD risk as coded in western studies, the “power of zero”.³⁰

Therefore, to apply CACS, we should carefully combine the Thai data and use the percentile rank of age range, in separated gender, which had been discussed in details²⁷ and it was beyond the scope of this study.

How to implement ML without CACS?

Generally, CVD risk prediction would be divided by three levels, high, intermediate and low risk, according to machine calculated risk of $\geq 20\%$, $>10-19\%$ and $< 9\%$ respectively. Each level required different further investigation and interventions, intensity varied from healthy life-style modification (low risk, including healthy diet, regular exercise, stop smoking, stress reduction etc.) to medication (intermediate and high CVD risk) and further investigation including calcium score in low CVD risk case, to ensure individual low risk and exercise stress test (EST) in high CVD risk or cardiac imaging i.e. CT scan in cases of positive EST. Ideally, this ML model required prospective field test in Thai population before clinical use.

Limitations

Several limitations existed in this study. First, occult CVD, which had been reported in 3.1%⁶, could not be truly excluded by hospital bases population. Second, incomplete data entry in service hospitals excluded over 40% of case numbers for enrollment. Third, individual data required average parameters of three consecutive years for accuracy and this precluded several patients who had no regular complete check-up.

Conclusions

Despite having the above limitations, our study was the first one to show the feasibility of CVD prediction by Machine learning from BDMS large clinical data, under multi-task team and PDPA protection. However, verification of predictive accuracy required prospective study, at least 5 or 10 years before clinical implementation. It would be interesting to combine CAC, as an individual risk assessment, for future CVD prediction with this ML model.

References

1. Division of Health Statistics Bureau of Health Policy and Planning. Public Health Statistics AD 2018-2022, Bangkok: Office of the Permanent Secretary, Ministry of Public Health 2022. ISSN 0857-3093. ((Accessed January 10, 2024, at <https://spd.moph.go.th/wp-content/uploads/2022/11/Hstatis-tic64.pdf>)
2. Centers for Disease Control and Prevention (CDC). Decline in deaths from heart disease and stroke, 1900-1999. *MMWR Morb Mortal Wkly Rep* 1999;48(30):649-56.
3. Ford ES, Ajani UA, Croft JB, et al. Explaining the decrease in U.S. deaths from coronary disease, 1980-2000. *N Engl J Med* 2007;356(23):2388-98. doi: 10.1056/NEJMsa053935.
4. Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC Guideline on cardiovascular disease prevention in clinical practice. *Eur Heart J* 2021;42(34):3227-37. doi:10.1093/eurheartj/ehab484.
5. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* 2019;140(11):e596-e646. doi: 10.1161/CIR.0000000000000678.
6. Veerakul G, Khajornyai A, Wongkasia S. et al. Predicting and preventing cardiovascular events in asymptomatic patients: A 10-year prospective study. *BKK Med J* 2017;13: 1-12. doi: 10.31524/bkkmedj.2017.02.001.
7. The Rama-EGAT study group. Development of novel CVD score for Thai population. [Internet]. [updated 30 July 2017]. (Accessed January 10, 2024, at <http://www.thaiheart.org/EGAT/PDF2013.html>).

8. The Rama-EGAT study group. Development of novel CVD score for Thai population. [Internet]. [updated 30 July 2017]. (Accessed January 10,2024, at <http://www.thaiheart.org/WorldHealthOrganization>. Prevention of cardiovascular disease: pocket guidelines for assessment and management of cardiovascular risk: (WHO/ISH cardiovascular risk prediction charts for the African Region). World Health Organization 2007. (Accessed January 10,2024, at <https://iris.who.int/handle/10665/43787>).
9. Na-witayanan J, Kanoksilp A, Kehasukcharoen W. et al. Incidence of low cardiovascular disease risk in patients with ST elevation myocardial infarction in Central Chest Institute. Thesis for interventional cardiology fellow board examination 2017.
10. Weng S, Resp J, Kai J, et al. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One* 2017;12(4):e0174944. doi: 10.1371/journal.pone.0174944.
11. Zhao J, Feng QP, Wu P. et al. Learning from longitudinal data in electronic health record and genetic data to improve cardiovascular event prediction. *Sci Rep* 2019;9(1):717. doi: 10.1038/s41598-018-36745-x.
12. Alaa AM, Bolton T, Angelantonio ED. et al. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PLoS One* 2019;14(5):e0213653. doi: 10.1371/journal.pone.0213653.
13. Erbel R, Möhlenkamp S, Moebus S, et al. Coronary risk stratification, discrimination, and reclassification improvement based on quantification of subclinical coronary atherosclerosis: The Heinz Nixdorf Recall study. *J Am Coll Cardiol* 2010;56(17):1397-406. doi: 10.1016/j.jacc.2010.06.030.
14. Budoff MJ, Möhlenkamp S, McClelland R, et al. A comparison of outcomes with coronary artery calcium scanning in unselected populations: The Multi-Ethnic Study of Atherosclerosis (MESA) and Heinz Nixdorf RECALL study (HNR). *J Cardiovasc Comput Tomogr* 2013;7(3):182-91. doi: 10.1016/j.jcct.2013.05.009.
15. Hermann DM, Gronewold J, Lehmann N, et al. Coronary artery calcification is an independent stroke predictor in the general population. *Stroke* 2013 Apr;44(4):1008-13. doi: 10.1161/STROKEAHA.111.678078.
16. Chaothawee L, Veerakul G, Poonsawas P. et al. Low or Even Zero CAC may not be Warranted as Safe in Thai Officers: A 10-year prospective cohort study. *BKK Med J* 2020;16(1):1-8. doi: 10.31524/bkkmedj.2020.11.001
17. Pedregosa F, Varoquaux G, Gramfort A, et al. Machine learning in Python. *J Mach Learn Res* 2011;12:2825-30.
18. Veerakul G, Khajornyai A, Wongkasai S, et al. Predicting and Preventing Cardiovascular Events in Asymptomatic Patients: A 10-Year Perspective Study. *BKK Med J* 2017;13(1):1-12. doi: 10.31524/bkkmedj.2017.02.001
19. Thomson SP, Gibbons RJ, Smars PA, et al. Incremental value of the leukocyte differential and the rapid creatine kinase-MB isoenzyme for the early diagnosis of myocardial infarction. *Ann Intern Med* 1995;122(5):335-41. doi: 10.7326/0003-4819-122-5-199503010-00003.
20. Korniluk A, Koper-Lenkiewicz OM, Kemon H et al. Mean Platelet Volume (MPV): New Perspectives for an Old Marker in the Course and Prognosis of Inflammatory Conditions. *Mediators Inflamm* 2019;2019:9213074. doi: 10.1155/2019/9213074.
21. Chu SG, Becker RC, Berger PB, et al. Mean platelet volume as a predictor of cardiovascular risk: A systematic review and meta-analysis. *J Thromb Haemost* 2010;8(1):148-56. doi: 10.1111/j.1538-7836.2009.03584.x.
22. Williams H, Mack CD, Li SCH, et al. Nature versus Number: Monocytes in Cardiovascular Disease. *Int J Mol Sci* 2021;22(17):9119. doi: 10.3390/ijms22179119.
23. Rogacev KS, Cremers B, Zawada AM, et al. CD14++CD16+ monocytes independently predict cardiovascular events: A cohort study of 951 patients referred for elective coronary angiography. *J Am Coll Cardiol* 2012;60(16):1512-20. doi: 10.1016/j.jacc.2012.07.019.
24. Urra, X, Villamor N, Amaro S. et al. Monocyte subtypes predict clinical course and prognosis in human stroke. *J Cereb Blood Flow Metab* 2009;29(5):994-1002. doi: 10.1038/jcbfm.2009.25.
25. Tapp LD, Shantsila E, Wrigley BJ, et al. The CD14++CD16+ monocyte subset and monocyte-platelet interactions in patients with ST-elevation myocardial infarction. *J Thromb Haemost* 2012;10(7):1231-41. doi: 10.1111/j.1538-7836.2011.04603.x.
26. Grundy S M, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: Executive Summary A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* 2019;139(25):e1046–e1081. doi: 10.1161/CIR.0000000000000624
27. Veerakul G, Chaothawee L, Chaiwong W, et al. Prevalence of calcified coronary plaque burden in thais. *BKK Med J* 2024;20(2):115-124. doi: 10.31524/bkkmedj.2024.21.010
28. Bild DE, Detrano R, Peterson D, et al. Ethnic differences in coronary calcification: the multi-ethnic study of atherosclerosis (MESA). *Circulation* 2005;111(10):1313-20. doi: 10.1161/01.CIR.0000157730.94423.4B.
29. Detrano R, Guerci AD, Carr JJ, et al. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *N Engl J Med* 2008;358(13):1336-45. doi: 10.1056/NEJMoa072100.
30. Valenti V, Ó Hartaigh B, Heo R, et al. A 15-year warranty period for asymptomatic individuals without coronary artery calcium: A prospective follow-up of 9,715 individuals. *JACC Cardiovasc Imaging* 2015;8(8):900-9. doi:10.1016/j.jcmg.2015.01.02

Appendix 1

Eight Hospitals in Bangkok Dusit Medical Service Contributed:

- | | |
|---------------------------------|----------------------------------|
| 1. Bangkok Hospital Headquarter | 5. Bangkok Hospital Chiang Rai |
| 2. Bangkok Hospital Pattaya | 6. Samitivej Sukhumvit Hospital |
| 3. Bangkok Hospital Hua Hin | 7. Samitivej Srinakarin Hospital |
| 4. Bangkok Hospital Chiang Mai | 8. Samitivej Chinatown Hospital |

Appendix 2

ICD code

- | | |
|---|--|
| I21 – Acute myocardial infarction | I25.110 – Atherosclerotic heart disease of native coronary artery with unstable angina pectoris |
| I21.0 – ST elevation (STEMI) myocardial infarction of anterior wall | I25.111 – Atherosclerotic heart disease of native coronary artery with unstable angina pectoris, with documented spasm |
| I21.1 – ST elevation (STEMI) myocardial infarction of inferior wall | I25.118 – Atherosclerotic heart disease of native coronary artery with other forms of angina pectoris |
| I21.2 – ST elevation (STEMI) myocardial infarction of other sites | I25.119 – Atherosclerotic heart disease of native coronary artery with unspecified angina pectoris |
| I21.3 – ST elevation (STEMI) myocardial infarction of unspecified site | I63.9 CVA |
| I21.4 – Non-ST elevation (NSTEMI) myocardial infarction | I63.9 Stroke |
| I21.9 – Acute myocardial infarction, unspecified | I65.23 Carotid Artery Occlusion, Bilateral |
| I25 – Chronic ischemic heart disease | I65.23 Carotid Artery Stenosis, Bilateral |
| I25.1 – Atherosclerotic heart disease of native coronary artery | I65.29 Carotid Artery Occlusion |
| I25.10 – Atherosclerotic heart disease of native coronary artery, without angina pectoris | I65.29 Carotid Artery Stenosis |
| I25.11 – Atherosclerotic heart disease of native coronary artery with angina pectoris | I67.2 Cerebral Atherosclerosis |
| | I67.9 Ischaemic Cerebrovascular Disease infarction of other sites |