Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 6, December 2026, pp. 5348~5359
ISSN: 2302-9285, DOI: 10.11591/eei.v15i6.13971

Selecting the right machine learning for community blood glucose screening in an ageing Thai population

Pichit Boonkrong[1], Junwei Yang[2], Manaporn Chatchumni[3], Montira Muanjan[4]
[1]College of Biomedical Engineering, Rangsit University, Pathum Thani, Thailand
[2]Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand
[3]School of Nursing, Rangsit University, Pathum Thani, Thailand
[4]Faculty of Nursing, Prince of Songkla University, Hat Yai, Thailand

Article Info

Article history:

Received Jul 15, 2026
Revised Sep 8, 2026
Accepted Sep 12, 2026

Keywords:

Ageing society
Gradient boosting
Non-communicable diseases
Risk stratification
Shapley additive explanations interpretation

ABSTRACT

Community-based screening is central to managing non-communicable diseases (NCDs) in Thailand's ageing population, yet village protocols reduce continuous glucose readings to binary risk labels. Using records from 684 participants screened in January 2025 at a community hospital in Sing Buri Province, Thailand, this study asks whether regression or classification is the safer paradigm for capillary blood glucose screening using Dextrostix (DTX). Three algorithms, multiple linear regression (MLR), random forest (RF), and eXtreme gradient boosting (XGB), were benchmarked in both paradigms across four feature-selection strategies: univariate filtering, least absolute shrinkage and selection operator (LASSO) regularization, recursive feature elimination (RFE), and full-feature inclusion. On the 20% hold-out set (n=137), the best regression model (XGB-LASSO) reached only root mean square error (RMSE)=16.0284 and coefficient of determination (R^2)=0.2731, and surveillance error grid (SEG) and Bland-Altman analyses showed it was clinically unsafe, with a flattened, heteroscedastic error profile that overestimates low and underestimates high glucose. Classification proved the more reliable choice: the XGB-RFE model discriminated elevated from normal glucose at the referral threshold (accuracy=0.7226, F1=0.7146, and area under the receiver (AUC)=0.7796), and Shapley additive explanations (SHAP) analysis identified the dominant glycemic drivers, comorbid diabetes mellitus (DM) and blood pressure (BP), for transparent risk attribution. XGB-RFE gives nurses and volunteers a clear refer-or-reassure decision, supporting safer community glucose screening in low-resource settings.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Montira Muanjan
Faculty of Nursing, Prince of Songkla University
15 Karnjanavanich Road, Hat Yai, Songkhla 90110, Thailand
Email: 6810430002@psu.ac.th

1. INTRODUCTION

The global surge in non-communicable diseases (NCDs), particularly diabetes, represents a profound threat to public health sustainability and economic stability [1]–[3]. In many developing regions, the transition from infectious to chronic disease burdens has outpaced the capacity of traditional healthcare infrastructures to respond. Since these conditions are often asymptomatic in their early stages, reactive treatment models lead to late-stage complications, including cardiovascular failure and renal disease [3], [4]. There is an urgent need to shift toward proactive, data-driven screening that intercepts disease progression before irreversible damage occurs, increasingly through the integration of digital health frameworks.

The challenge is acute in Thailand, now entering a fully aged society with a metabolic-disorder prevalence that strains community-based primary care. Screening relies heavily on village health volunteers (VHVs), yet current protocols reduce readings to binary "*at-risk*" or "*normal*" labels [5]–[7]. This is operationally convenient but conceptually limited: it discards the continuous glucose spectrum and the physiological drift that precedes diagnosis. Point-of-care glucometry remains central to community screening, providing rapid measurements in good agreement with laboratory testing despite some variability [7]–[9], and machine learning (ML) has been applied to glucose prediction and monitoring with encouraging accuracy [10], [11]. Whether such models should target a continuous value (regression) or a diagnostic category (classification), however, is rarely examined directly. Table 1 contrasts the regression and classification paradigms across output type, use case, evaluation, interpretability, and clinical suitability for community glucose screening.

Table 1. Key considerations for ML approaches of blood glucose measurement using Dextrostix (DTX) for community-based diabetes screening interventions for early detection

Aspects	Regression	Classification
Output type	Continuous value: glucose concentration (mg/dL); glycated hemoglobin (HbA1c) (%).	Discrete categories for glucose levels (e.g., normal, risk, and diabetic).
Use case	Quantifying the exact glycemic level for monitoring, forecasting, and dose- or severity-calibrated feedback [12].	Assigning a patient to a predefined diagnostic group for triage and referral decisions [13].
Algorithms	Multiple linear regression (MLR), support vector machine (SVM), random forest (RF), eXtreme gradient boosting (XGB), light gradient boosting machine (LightGBM), and long short-term memory (LSTM).	Logistic regression (binary logistic regression (BLR) and ordinal logistic regression (OLR)), RF, XGB, LightGBM/categorical boosting (CatBoost), SVM, and artificial neural network (ANN).
Evaluation	Root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and R^2; clinical safety assessed with the Clarke error grid (CEG)/surveillance error grid (SEG) [14].	Accuracy, precision, sensitivity/recall, specificity, F1-score, and area under the receiver operating characteristic curve (ROC-AUC) [15].
Clinical interpretation	Reveals the magnitude and direction of glycemic drift, enabling severity-graded, personalized intervention; SHAP quantifies each feature's push on the predicted value [16].	Yields an actionable risk tier aligned with diagnostic thresholds; Shapley additive explanations (SHAP) explains class membership for transparent referral decisions [17].
Class imbalance/ data demands	No class balancing needed, but sensitive to outliers and skewed glucose distributions; benefits from larger, finely measured samples.	Requires explicit imbalance handling (class weight, synthetic minority over-sampling technique (SMOTE)/adaptive synthetic sampling (ADASYN)) when the diabetic group is small [18].
Suitability	Higher granularity for digital-therapeutics feedback, but continuous errors may fall in unsafe zones, raising deployment risk.	Robust, threshold-aligned labels that map directly onto VHVs triage; generally safer for point-of-care decisions.

Regression suits continuous glucose prediction; classification suits risk stratification or diagnosis. Choice depends on application and data structure. In noisy community data, regression may yield directionally misleading estimates, whereas categorical risk-tier flags may be safer for point-of-care deployment [19]. Whether regression's added complexity and clinical risk are justified, or whether classification remains more appropriate for VHV-led screening, has not been directly tested in the Thai community-health literature, which favors classification. This weighting is deliberate, since screening is decision-driven. Classification into referral tiers enables efficient triage, resource allocation, and clear communication, with feature interpretation offering clinical transparency for each decision [17], [20]. Regression still requires thresholding for referrals, reintroducing arbitrary cut-points and decision-point errors. Classification better aligns with community screening priorities, where throughput, interpretability, and consistency are essential [9], [10]. However, reporting practice is also uneven: clinical-safety evaluation by error-grid and Bland-Altman analysis and model transparency by SHAP-based attribution are rarely applied together to community screening models [10], [11].

This gap was addressed through a direct comparison of regression and classification models applied to capillary blood glucose prediction in 684 participants from a Thai primary care unit, using identical low-cost predictors. Both paradigms were assessed for statistical accuracy, clinical safety (via SEG analysis), and interpretability (via SHAP). The implications of this comparison may extend beyond the present cohort, as the choice of predictive paradigm could influence whether digital tools reduce or amplify referral error in volunteer-delivered NCD screening, particularly in the context of ageing populations.

2. COMPUTATIONAL FRAMEWORK

The computational framework predicts capillary blood glucose (DTX) among 684 participants from a Thai community hospital using readily obtainable screening variables. As visualized in Figure 1, analysis progresses from a MLR baseline to ensemble learners, namely and XGB, enabling both interpretable linear relationships and complex non-linear dependencies between demographic, anthropometric, and lifestyle

predictors and the DTX outcome to be captured [15], [21]. Python integrated with the scikit-learn library served as the primary programming and ML framework in this study. The pipeline integrates feature selection and parallel regression and classification models towards possible implementation. In Figure 1, EDA denotes exploratory data analysis.

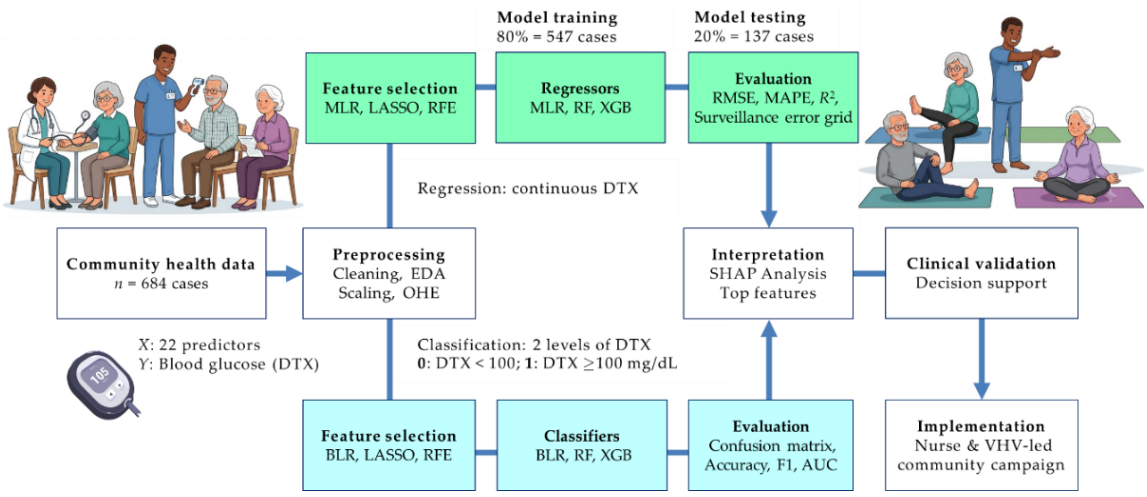


Figure 1. Overview of the computational framework for blood glucose prediction

2.1. Community recruitment and data acquisition

This section describes the recruitment of 684 adults from a Thai community hospital NCD screening program. Eligible participants required complete DTX readings alongside demographic, anthropometric, and lifestyle variables, all routinely captured to support feasible, nurse-led implementation in rural settings.

2.1.1. Study population

Participants were drawn from NCD screening surveys conducted in January 2025 at Hua Pa Subdistrict Health Promotion Hospital, Phrom Buri District, Sing Buri Province, Thailand. Community-dwelling adults aged ≥30 years were eligible if they had a complete capillary glucose reading together with complete demographic, anthropometric, and lifestyle fields; records with a missing DTX target, incomplete predictor fields, or duplication were excluded, yielding a final analytic sample of 684 participants (aged 30-93 years). This study adhered to the Declaration of Helsinki, received ethics approval from the Sing Buri Provincial Public Health Office (SEC 03/2568), and obtained written informed consent from all participants.

2.1.2. Variable collection

Each participant contributed a structured profile spanning demographic indicators (age, sex), an anthropometric measure (body mass index (BMI)), vital signs, lifestyle behaviors, family history of NCDs, and recorded comorbidities. Lifestyle variables were operationalized from the screening questionnaire as follows: taste preference (sweet, oily, salty, and bland) was captured as self-reported habitual dietary preference and coded as separate binary indicators; regular exercise was self-reported as at least one session per week (yes/no); and alcohol consumption and smoking were self-reported current use (yes/no). Family history of NCDs and comorbidities were recorded as present or absent from self-report, confirmed against clinical records where available. The continuous target variable was the capillary blood glucose level (mg/dL); the remaining variables formed the predictor matrix. This emphasis on variables that are inexpensive and routinely captured during community screening ensures the framework is feasible for rural, nurse-led implementation [22].

2.2. Data preprocessing

Raw NCD records were transformed into a machine-readable matrix before modelling. Continuous predictors were rescaled by min–max normalization so that disparate measurement units did not bias distance- and gradient-based learners, $x'_j = (x_j - x_{min})/(x_{max} - x_{min})$ where x_j is the raw value of feature j and x'_j its normalized value on [0, 1]. Categorical attributes were converted to binary indicator vectors by one-hot encoding (OHE). The processed data were partitioned into an 80% training set (547 records) and a 20%

independent test set (137 records), and all preprocessing and hyperparameter tuning were nested within a 5-fold cross-validation (CV) loop applied to the training partition to prevent data leakage.

2.3. Feature selection techniques

Feature selection reduced dimensionality among the NCD predictors. Pairwise collinearity was first screened with the variance inflation factor (VIF), $\mathrm{VIF}_j = 1/(1-R_j^2)$ where R_j^2 is the coefficient of determination obtained by regressing feature j on the remaining predictors; variables exceeding the conventional VIF>10 threshold was removed. Statistically significant predictors (p<0.05) were then identified through MLR or BLR, while a high-dimensional sparse subset was obtained by least absolute shrinkage and selection operator (LASSO) regularization defined as $\hat{\beta} = \mathrm{argmin}_\beta \left[\frac{1}{2n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^T\beta) - \lambda\sum_{j=1}^{p}|\beta_j|\right]$ with regularization strength λ shrinking uninformative coefficients to zero. Finally, recursive feature elimination (RFE) was applied to select the most informative features by recursively removing the least important ones based on model coefficient weights. For the loss function L and regularization penalty R, the weight is computed by $W = \mathrm{argmin}_W\, \mathrm{L}(X, Y{:}\, W) + \lambda \mathrm{R}(W)$ and the ranking can be made by $c_i = f(W_i)$. Together these procedures produced distinct input configurations: all features, MLR/BLR-significant, LASSO-selected, and RFE-confirmed, used to cross-validate the ML models [15], [23].

2.4. Regression and classification models

The framework employed the same three learners in both paradigms, spanning interpretable statistics to ensemble learning. MLR (regression) and BLR (classification) served as the baselines, assigning explicit coefficients to each feature [12], [14]; RF combined predictions from multiple decision trees to improve robustness and reduce overfitting [13], [15]; and XGB built trees sequentially with gradient descent and regularization to capture intricate physiological dependencies. For classification, the continuous DTX target was dichotomized at 100 mg/dL (normal vs. elevated), matching the VHV's binary referral decision, as both prediabetic and diabetic cases warrant referral and capillary readings are non-diagnostic [4], [5]. This cut-point follows the American Diabetes Association (ADA) impaired-fasting glucose threshold and Thailand's NCD screening trigger for confirmatory venous testing. RF and XGB hyperparameters were tuned by GridSearchCV with 5-fold CV on the training partition (test_size=0.2, random_state=42), with class_weight=balanced retained in classification as a safeguard, although at 41.23% elevated cases the task was near-balanced rather than skewed, and macro-averaged F1 was reported for the same reason; the search grids and selected values are reported in Table 2. L_1 penalization ($\lambda \cdot \Sigma|\beta|$) was applied to both tasks: regression was minimized using prediction error and L_1 penalty with λ=0.3793 (selected via LASSO CV), while classification was minimized using log-loss and L_1 penalty over C ∈ {0.01, 0.1, 1.0, 10.0} (reported as $\lambda = 1/C$). Regression performance was quantified using RMSE, relative absolute error (RAE), MAPE, and coefficient of determination (R^2). Classification performance was evaluated using accuracy, sensitivity, and macro-averaged F1-score, while discriminative ability was assessed using the ROC-AUC [17], [24].

Table 2. Hyperparameter tuning for regression and classification models

Model	Hyperparameter	Search grid	Regression	Classification
RF	n_estimators	{100, 200, 300}	300	200
	max_depth	{None, 5, 10, 15}	5	5
	min_samples_split	{2, 5}	5	2
	max_features	{sqrt(22), log2(22)}	sqrt(22)	sqrt(22)
XGB	n_estimators	{100, 200, 300}	200	100
	max_depth	{3, 5, 7}	3	3
	learning_rate	{0.01, 0.05, 0.1}	0.01	0.05
	subsample	{0.7, 0.9, 1.0}	1	1
	colsample_bytree	{0.7, 0.9, 1.0}	0.7	0.9

2.5. Surveillance error grid

Beyond conventional error metrics (RMSE, MAPE, and R^2), predictive accuracy for glucose regression should be evaluated for clinical risk using the SEG, developed by the Diabetes Technology Society from a survey rating the risk associated with paired reference ($\mathrm{G_{ref}}$) and predicted ($\mathrm{G_{pred}}$) glucose values [12], [14]. Unlike simple percent-error thresholds, SEG assigns a continuous risk score $R(\mathrm{G_{ref}}, \mathrm{G_{ref}})$ by fitting a bivariate polynomial surface to clinician rated risk over a log-transformed, rotated coordinate system:

$$u = \frac{\log_{10}\mathrm{G_{ref}} + \log_{10}\mathrm{G_{pred}}}{\sqrt{2}},\ v = \frac{\log_{10}\mathrm{G_{ref}} - \log_{10}\mathrm{G_{pred}}}{\sqrt{2}},\ R(u,v) = \sum_{ij} a_{ij}\, u^i v^j$$

where coefficients a_{ij} are least-squares fits to the survey data. The resulting score maps each prediction into one of five risk zones: A (none), B (slight), C (moderate), D (great), and E (extreme), with the percentage of predictions falling in zones A and B serving as the primary safety benchmark (typically requiring ≥98%–99% for clinical acceptability).

2.6. Shapley additive explanations analysis and clinical interpretation

Model transparency was achieved through SHAP, which decomposes each prediction into additive feature contributions based on cooperative game theory [15], [20]. For a prediction $\hat{y}(x)$, the contribution ϕ_j of feature j is computed as the average marginal effect over all feature subsets $S \subseteq F\{j\}$, i.e., $\phi_j = \Sigma_{S \subseteq F j} |S|! \, (|F| - |S| - 1)! \, [f(S \cup j) - f(S)] / |F|!$ where $f(S)$ denotes the model output conditioned on subset S. Global importance is quantified by mean absolute SHAP values, $\frac{1}{n} \Sigma_{i=1}^{n} \left| \phi_j^{(i)} \right|$, ranking predictors by their average impact on DTX. Local explanations decompose individual predictions to identify which factors elevate or lower a given participant's estimated glucose, providing interpretable, patient-specific guidance for nurses and VHVs.

3. NUMERICAL RESULTS

This section reports the empirical performance of the 12 regression and 12 classification models for DTX prediction. Regression accuracy is first benchmarked across feature-selection strategies, then examined for clinical safety through SEG and Bland-Altman analyses, which expose its unsuitability. The parallel classification track is subsequently evaluated to identify the optimal combination, whose predictions are interpreted with SHAP to reveal the dominant glycemic drivers and their community-level implications.

3.1. Cohort characteristics

The study population presents the profile of an aging Thai community with elevated cardiometabolic exposure. Table 3 summarizes the continuous features and the DTX target, and Table 4 reports the categorical features, both stratified by sex to expose physiological and behavioral disparities. Descriptive inspection of DTX indicates a right-skewed distribution typical of community glucose screening, with a tail of participants in the prediabetic and diabetic range that motivates continuous modeling rather than a single binary cut-point. Sex-based differences are visible, with slightly higher BMI and family history of diabetes among females alongside near-universal lifestyle exposures. Values in Table 4 are *n* (%) of participants coded "*present*=1" for each binary indicator.

Table 3. Descriptive statistics of continuous features and the DTX target (n=684)

Feature	Description	Mean	SD	Min	Med	Max	Clinical note
Y	Blood glucose (mg/dL)	100.65	17.26	52	97	241	Serum sugar level; glycemic status
X_1	Age (years)	59.15	13.64	30	59.5	93	Key for risk stratification
X_2	BMI (kg/m²)	24.72	5.17	16.72	23.85	56.31	Weight management/metabolic risk
X_3	Systolic (mmHg)	130.48	14.57	83	130	204	Peak arterial pressure
X_4	Diastolic (mmHg)	77.64	10.69	62	78	120	Minimum arterial pressure

Note: SD, standard deviation; Min, minimum; Med, median; Max, maximum. Y denotes the outcome and X_1–X_4 denote feature codes.

3.2. Feature selection outcomes

Screening multicollinearity among the 22 features, the VIF ranged between 1.0261 and 4.6269, so all 22 were retained for modelling. MLR and BLR significance testing retained a compact subset of predictors at $p<0.05$. LASSO preserved the broad predictor set as a high-dimensional benchmark, and RFE identified a smaller all-relevant feature core. Feature selection produced three subsets per paradigm including, MLR-significant, LASSO, and RFE for regression, and their BLR-significant, LASSO, and RFE counterparts for classification. The two paradigms selected markedly different predictors: regression retained continuous glycemic correlates via signed coefficients while the classification filters weighted comorbidity (como) and lifestyle indicators more heavily, confirming that optimal inputs are paradigm-specific rather than transferable between the continuous and categorical tasks. These differences are summarized in Table 5, which presents the coefficients and feature-selection results for the 22 predictor variables in the regression and classification approaches.

Table 4. Prevalence of selected categorical features (n=684)

Feature	Description	All (%)	Male (%)	Female (%)	Clinical note
X_5	Sex	684 (100)	307 (44.88)	377 (55.12)	Slight female predominance
X_6	Family history: diabetes	198 (28.95)	72 (23.45)	126 (33.42)	Intergenerational glycemic risk
X_7	Family history: hypertension (HT)	314 (45.91)	130 (42.35)	184 (48.81)	Shared haemodynamic background
X_8	Family history: heart disease	11 (1.61)	3(0.98)	8(2.12)	Positive for coronary artery disease (CAD) in parent
X_9	Family history: stroke	15 (2.19)	4(1.30)	11 (2.92)	Positive for stroke in parent
X_{10}	Family history: kidney disease	18 (2.63)	7(2.28)	11 (2.92)	Positive for chronic kidney disease (CKD) in parent
X_{11}	Como: diabetes	122(17.84)	44 (14.33)	78 (20.69)	Strongest single predictor
X_{12}	Como: HT	263 (38.45)	115 (37.46)	148 (39.26)	Co-occurs with elevated systolic blood pressure (BP)
X_{13}	Como: heart disease	29 (4.24)	17 (5.54)	12 (3.18)	CAD
X_{14}	Como: stroke	2 (0.29)	2 (0.65)	0 (0.00)	Acute cerebrovascular accident (CVA)
X_{15}	Como: kidney disease	3 (0.44)	0 (0.00)	3 (0.80)	Acute kidney injury (AKI) secondary to dehydration
X_{16}	Taste preference: sweet	296 (43.27)	146 (47.56)	150 (39.79)	Reflects high-sucrose Thai diet
X_{17}	Taste preference: oily	217 (31.73)	100 (32.57)	117 (31.03)	Fried, high-fat or fast food
X_{18}	Taste preference: salty	238 (34.80)	118 (38.44)	120 (31.83)	Admits to high-sodium diet
X_{19}	Taste preference: bland	322 (47.08)	131 (42.67)	191 (50.66)	avoids strong flavors/ seasonings
X_{20}	Drinking alcohol	560 (81.87)	211 (68.73)	349 (92.57)	Social history, no withdrawal sign
X_{21}	Smoking	588 (85.96)	219 (71.34)	369 (97.88)	Sex-asymmetric lifestyle factor
X_{22}	Exercise: >1 time/week	342 (50.00)	154 (50.16)	188 (49.87)	Inverse association with glucose

Note: X_5–X_{22} denote feature codes.

Table 5. Coefficients from MLR, LASSO regularization and RFE across the 22 predictor variables

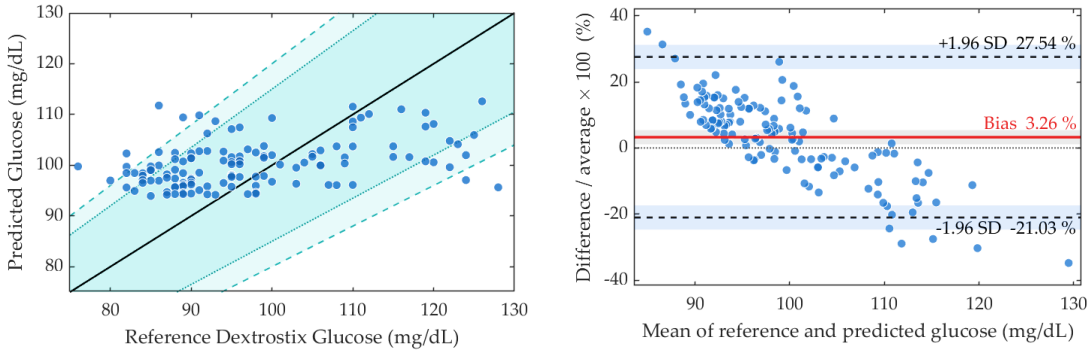
	Regression			Classification		
Feature	MLR	LASSO	RFE	BLR	LASSO	RFE
X_1	**1.5770**	<0.001	**1**	**-0.2547**	**0.2131**	**1**
X_2	**2.3939**	**1.2167**	**1**	**-0.2585**	**0.2301**	**1**
X_3	**3.5033**	**2.4916**	**1**	**-0.3621**	**0.1963**	**1**
X_4	**1.7962**	**0.1495**	**1**	**-0.2314**	**0.1348**	**1**
X_5	-0.9935	**-0.7434**	**1**	0.1490	**0.1923**	**1**
X_6	1.3351	**0.3775**	**1**	-0.1421	**0.0806**	4
X_7	-1.3421	**-1.0344**	**1**	0.0973	**0.1035**	2
X_8	-0.1030	<0.001	11	-0.0049	**0.0476**	8
X_9	**-1.7131**	**-0.9329**	8	0.2906	**0.2265**	**1**
X_{10}	-0.6471	<0.001	9	0.1467	**0.0736**	6
X_{11}	**4.6756**	**3.5486**	**1**	**-0.4961**	**0.3945**	**1**
X_{12}	**2.4558**	**1.0575**	4	**-0.2456**	**0.0862**	**1**
X_{13}	0.3577	<0.001	7	-0.0747	**0.0366**	10
X_{14}	-0.5867	**-0.1438**	10	-0.0158	**0.0468**	7
X_{15}	-0.2737	< 0.001	12	0.0321	**0.0346**	9
X_{16}	1.4022	**0.2491**	5	-0.1669	<0.001	12
X_{17}	0.6643	<0.001	**1**	-0.1229	**0.0759**	5
X_{18}	0.4202	<0.001	2	-0.0989	**0.0221**	11
X_{19}	-1.1595	<0.001	**1**	**0.1982**	**0.0755**	**1**
X_{20}	-0.7763	**-0.4963**	3	0.1341	**0.2323**	**1**
X_{21}	0.5719	**0.5383**	6	-0.0980	**0.2575**	**1**
X_{22}	-0.4704	**-0.1800**	**1**	0.0665	**0.0910**	3
	$\|F_1\|$=7	$\|F_2\|$=15	$\|F_3\|$=10	$\|F_4\|$=8	$\|F_5\|$=21	$\|F_6\|$=11

3.3. Regression model performance

Table 6 reports regression performance across the feature-set and learner combinations on the 20% hold-out set, quantified consistently with RMSE, RAE, MAPE, and R^2. Non-linear ensembles were expected to outperform the linear baseline, with the best configuration achieving the lowest error and highest R^2. XGB with the 15-feature LASSO subset gave the best overall fit (RMSE=16.0284, R^2=0.2731), narrowly ahead of MLR with RFE (RMSE=16.1605, R^2=0.2611); RAE and MAPE followed the same ordering. With all models capped at ~27% explained variance and comparable errors, the performance ceiling appears data-bound, so predictions require clinical corroboration and are unsuitable as standalone decision support. The SEG and Bland-Altman analyses jointly quantify the clinical safety of continuous DTX prediction for the best regression model, and together show it is not fit for this task (Figure 2). In the clinical risk plot in Figure 2(a), reference and predicted glucose are compared against graded low-risk zones (±15% and ±20%). On the hold-out set, the best regression model placed 85.40% of predictions in SEG zone A and 13.87% in zone B (99.27% in A and B), with one prediction (0.73%) in zone D and none in zones C or E. Figure 2(b) presents the Bland-Altman plot of percentage differences, showing proportional bias and widening limits of agreement.

Table 6. Regression performance for blood glucose prediction on the hold-out set (n=137)

Regressor	Feature set	|F |	RMSE	RAE	MAPE	R^2
MLR	All predictor variables (ALL)	22	16.1833	0.843	11.0939	0.259
MLR	MLR (statistically significant (sig.))	7	16.1187	0.8542	11.2456	0.2649
MLR	LASSO	15	16.3066	0.8493	11.146	0.2477
MLR	RFE	10	16.1605	**0.8418**	**11.0624**	0.2611
RF	ALL	22	16.5711	0.88	11.4971	0.2231
RF	MLR (sig.)	7	16.4369	0.8665	11.3549	0.2356
RF	LASSO	15	16.7479	0.8882	11.6182	0.2064
RF	RFE	10	16.479	0.8697	11.3747	0.2317
XGB	ALL	22	16.2845	0.8516	11.1507	0.2497
XGB	MLR (sig.)	7	16.5849	0.8627	11.2847	0.2218
XGB	LASSO	15	**16.0284**	0.8621	11.3381	**0.2731**
XGB	RFE	10	16.4395	0.8616	11.2965	0.2354



(a) (b)

Figure 2. Clinical validation of the optimal regression on the testing set (n=137): (a) SEG mapping reference against predicted glucose onto risk zones A (none) to E (extreme), with graded ±15% and ±20% bands and (b) Bland-Altman plot of percentage difference showing proportional bias and widening limits of agreement

3.4. Classification model performance

Because regression fails to track true fluctuations across the full range, it offers no safety margin and is unsuitable for community use, making classification the more defensible choice. The binary screening results are given in Table 7.

Table 7. Classification performance for the normal-elevated DTX glucose task

Classifier	Feature set	$\|\mathcal{F}\|$	Accuracy	Sensitivity	F1	AUC
BLR	ALL	22	0.6569	0.6516	0.642	0.7099
BLR	BLR (sig.)	8	0.635	0.639	0.6245	0.7096
BLR	LASSO	21	0.6496	0.6459	0.6355	0.7094
BLR	RFE	11	0.6715	0.6629	0.655	0.7087
RF	ALL	22	0.6715	0.681	0.6628	0.7417
RF	BLR (sig.)	8	0.6277	0.6198	0.6115	0.7131
RF	LASSO	21	0.6496	0.6549	0.6395	0.7514
RF	RFE	11	0.6788	0.6822	0.6678	0.7514
XGB	ALL	22	0.7226	0.7343	0.7146	0.7734
XGB	BLR (sig.)	8	0.708	0.7275	0.7023	0.7659
XGB	LASSO	21	0.6861	0.6743	0.668	0.7509
XGB	RFE	11	**0.7226**	**0.7343**	**0.7146**	**0.7796**

Elevated cases comprised 41.23% of the cohort, a near-balanced task. Across the twelve classifiers of different feature-set combinations, discrimination was moderate, with test-set AUC ranging from 0.71 to 0.78. XGB was the strongest family, its RFE configuration (eleven predictors) achieving the best overall performance (F1=0.7146, AUC=0.7796). RF was intermediate, peaking at AUC=0.7514 (LASSO and RFE), while the logistic baseline BLR was weakest, with AUCs clustered near 0.71; even its best configuration (ALL, AUC=0.7099) fell well below XGB-RFE, and XGB outperformed BLR across every feature set. This gap confirms that the elevated-glucose decision boundary is governed by non-linear predictor interactions the logistic baseline cannot capture, and that reducing to eleven RFE-selected predictors preserved discriminative power without loss. The XGB-RFE classifier therefore gives VHVs a dependable, low-cost screening aid,

flagging elevated capillary glucose at the referral threshold from only routinely collected variables. Its clear normal-versus-elevated output maps directly onto the existing refer-or-reassure workflow, making it well suited for VHV-led DTX glucose screening in resource-limited settings. Figure 3 compares the leading XGB-RFE classifier with the baseline BLR classifier: Figure 3(a) presents their confusion matrices, including missed referrals and false alarms, and Figure 3(b) presents their receiver operating characteristic (ROC) curves and corresponding AUC values.

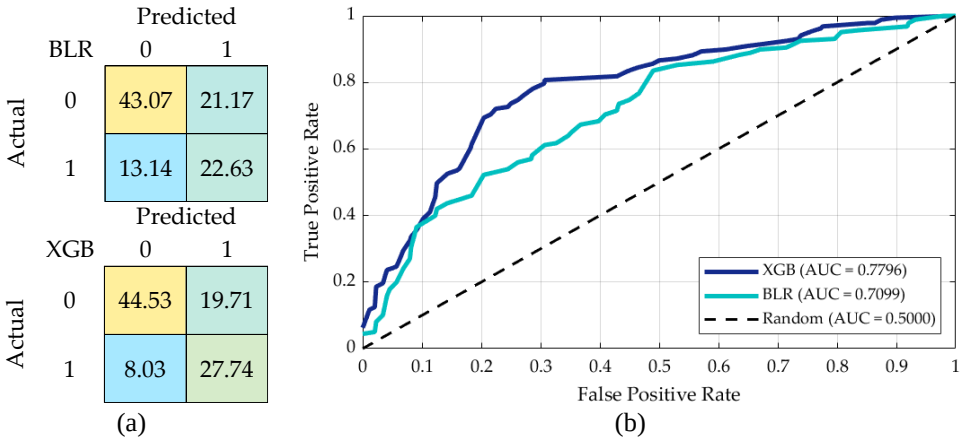


Figure 3. Binary classification performance: (a) confusion matrices for the leading XGB-RFE and baseline BLR classifiers, with off-diagonal cells indicating missed referrals and false alarms and (b) ROC curves with corresponding AUC

3.5. Shapley additive explanations feature importance

Global SHAP rankings identify the dominant determinants of DTX blood glucose in this cohort. Figure 4 presents the SHAP interpretation of the optimal XGB-RFE classifier with AUC=0.7796. Figure 4(a) ranks feature by mean absolute SHAP value, from highest to lowest impact on the classification outcome. Contrary to the initial expectation that BMI and age would dominate, the leading contributors are comorbid diabetes mellitus (DM) (Como: DM, 0.2553), systolic pressure (0.2160), and diastolic pressure (0.1506), followed only then by age (0.1275) and BMI (0.1272); the top five are shown in teal and the remaining features in navy blue. This ordering indicates that prevalent diabetes and BP status, rather than anthropometric measures, drive the model most strongly across all 684 samples. Figure 4(b) provides a granular, per-patient view in the same ranking order: each dot is one participant, colored from navy (low feature value) to teal (high), with horizontal position showing whether a feature pushes the prediction toward the elevated or normal class. Notably, comorbid DM was the single strongest determinant of predicted glucose, marking prevalent diabetes as the dominant driver of elevated-class assignment.

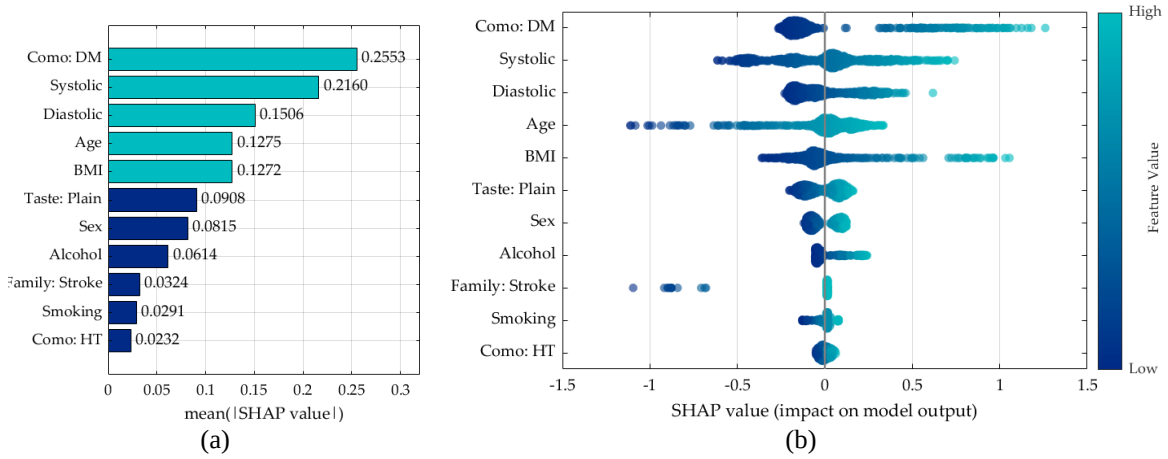


Figure 4. SHAP interpretation of the optimal XGB-RFE classifier: (a) global feature importance ranked by mean absolute SHAP value and (b) beeswarm summary for impact of features on model output

4. DISCUSSION

This section interprets key findings in clinical and community contexts, compares regression and classification models, and aligns top predictors with metabolic risk evidence. It then discusses implications for digital therapeutics and nurse-led care before addressing study strengths, limitations, and future directions.

4.1. Principal findings

This study reframes community glucose screening as a choice between two modeling paradigms and tests them on identical, low-cost inputs. For continuous DTX prediction, the non-linear advantage was marginal: only XGB paired with LASSO features edged the linear baseline, while RF actually trailed it, and every model plateaued near 27% explained variance. Such a ceiling is common in community cardiometabolic ML and signals that much of the physiological signal lies in determinants this routine dataset never captures [25]. As prediction error increased with glucose concentration, flattening the fitted model's slope toward the cohort mean and pulling low readings upward and high readings downward. A participant with a true capillary value near 60 mg/dL would be nudged into an apparently safe range, while someone near 200 mg/dL would be underestimated toward normal errors at exactly the hypo- and hyperglycemic extremes where a screening decision is most consequential [21]. The Bland-Altman analysis expresses this same behavior as proportional bias with widening limits of agreement, confirming that the continuous estimate carries no dependable safety margin [11], [26]. While regression theoretically preserves the glycemic drift magnitude that binary classification discards, its graded signal proved practically unreliable at the extremes for safe referral, so XGB-RFE classification instead emerges as the fit-for-purpose paradigm for VHV-led screening.

4.2. Key predictors and clinical relevance

In the present study, SHAP ranked comorbid DM and systolic and diastolic BP ahead of BMI and age. This finding is relevant to Thailand’s community screening context, where diabetes and hypertension are addressed within national NCD services [5], [6]. Related evidence from South Korean and Japanese cohorts identified age, BP, BMI, and fasting blood glucose as important predictors of incident hypertension [27]. The practical reading is direct: a volunteer should prioritize the participant who already carries a diabetes or hypertension diagnosis over one flagged by a high BMI alone. BMI still matters as an actionable lifestyle target, age reflects accumulated vascular exposure that warrants closer monitoring, and a family history of diabetes points to intergenerational risk that supports early counseling, e.g., prompting a volunteer to advise the adult children of a diabetic patient during the same visit [4], [17], [19]. Turning opaque data into clear patient-specific evidence, these rankings empower nurses to explain complex results, advancing health literacy instead of obscuring it.

4.3. Implications for community care

Pairing an interpretable prediction with an accessible digital channel offers a realistic path from a one-off screening to sustained follow-up. The MICArdiac trial protocol describes an artificial intelligence (AI)-supported mobile health intervention designed to improve BP management [23]. A nationwide cross-sectional study in Thailand highlights opportunities to improve BP control and the quality of continuous hypertension care [6]. The model’s deployability rests on its simple output, a normal-versus-elevated flag at the 100 mg/dL threshold, which maps directly onto the volunteer's existing refer-or-reassure decision, uses only routine screening variables, and requires no additional equipment or visits during monthly village surveys, e.g., the same tablet form that records BP and history could return the flag immediately, so a borderline reading is referred for confirmatory venous testing on the spot rather than lost to follow-up. Discrete labels also sit more safely in electronic records and decision-support tools than continuous estimates that carry the regression model's demonstrated risk. By delivering such prompts through Thailand's widely used LINE Official Account and reinforcing them with internet of things (IoT)-linked home monitoring (e.g., via automated referral-appointment reminders to flagged patients), care can transition from isolated snapshots to longitudinal management. Embedded within the national VHV program and universal-coverage NCD screening, this kind of low-cost triage aid offers a scalable route to earlier detection that supports frontline judgment instead of replacing it [5], [7]. Framed as precision public health, this is stratification of finite volunteer effort toward those most likely to need confirmatory testing. The same pipeline transfers directly to other routine screening targets, such as hypertension or dyslipidemia staging, where the referral decision is likewise categorical.

4.4. Strengths and limitations

The main strength is feasibility: every predictor is captured in routine community screening, and the parallel regression-classification design, validated with SEG and Bland-Altman analysis, tests clinical safety rather than statistical accuracy alone. The findings hold within defined boundaries. The single-site, cross-sectional design supports screening-level inference rather than causal claims, capillary DTX readings are

noisier than plasma glucose, and self-reported lifestyle variables together with a high local exposure prevalence mean transportability warrants confirmation elsewhere. Metrics come from a single fixed-seed 80/20 partition, so repeated-seed resampling would add variance estimates alongside external validation. The modest explained variance, typical of community NCD data, indicates that further gain lies in richer inputs rather than more complex models. Future work should extend the binary scheme to three tiers (normal, prediabetes, and diabetes), validate prospectively across Thai provinces, test low-cost non-invasive sensors, and confirm that these predictions improve glycemic outcomes through digital therapeutics.

5. CONCLUSION

This study reframed community glucose screening as a choice between paradigms and evaluated both on identical, low-cost inputs from an aging Thai cohort, and the evidence points one way: for VHV-led screening, classification is the safer and more deployable choice than regression. Non-linear ensembles improved continuous prediction only marginally, and the SEG and Bland-Altman analyses showed that regression estimates carried unacceptable clinical risk, flattening toward the mean so that low glucose was overestimated and high glucose underestimated, failing in exactly the ranges where accuracy is critical. Beyond this specific result, the paired SEG and Bland-Altman safety check and SHAP attribution constitute a reusable evaluation protocol for any community screening model that outputs a clinical number. The XGB-RFE classifier instead delivered a dependable normal-versus-elevated decision at the 100 mg/dL referral threshold, and SHAP interpretation identified comorbid DM, BP, age, and BMI as its leading glycemic drivers, turning an opaque prediction into patient-specific evidence a nurse or volunteer can act on and explain. As discrete labels integrate more safely into electronic screening and digital channels such as LINE and IoT monitoring than continuous estimates do, this classifier offers a practical foundation for sustainable, precision-stratified community care. Prospective multi-site trials remain the necessary next step to confirm that digital-therapeutic delivery built on this foundation translates into measurable glycemic improvement.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the participants of the Hua Pa Subdistrict Health Promotion Hospital for their time and cooperation. We also thank the local healthcare staff and community leaders for their invaluable assistance in data collection.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Pichit Boonkrong	✓	✓	✓	✓	✓			✓	✓	✓	✓		✓	
Junwei Yang		✓	✓					✓		✓	✓			
Manaporn Chatchumni				✓		✓				✓	✓			
Montira Muanjan	✓			✓		✓	✓			✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

ETHICAL APPROVAL

The research received approval from the Human Research Ethics Committee of Sing Buri Province, Thailand (approval no. SEC 03/2568; approved on January 24, 2025). Following written informed consent, all participant data was handled confidentially for research purposes in strict accordance with the Declaration of Helsinki and The Belmont Report.

DATA AVAILABILITY

The de-identified dataset, together with the preprocessing and feature-selection scripts, the SEG evaluation code, and the SHAP visualization notebooks, are available on request from the corresponding author, [MM].

REFERENCES

[1] World Health Organization (WHO), "Noncommunicable diseases." 2026. [Online]. Available: https://www.emro.who.int/noncommunicable-diseases/, (Accessed: Jan. 09, 2026).

[2] K. M. Agboyibor *et al.*, "Prevalence, time trends and associated factors of adult overweight and obesity in 36 countries in the WHO African region from 2003 to 2022: a study of 54 WHO STEPS surveys representing 156 million adults," *BMJ Global Health*, vol. 11, no. 1, p. e019988, Jan. 2026, doi: 10.1136/bmjgh-2025-019988.

[3] World Health Organization (WHO), "From Strategy to Actions: Prioritizing NCD prevention and control Regional workshop to advance NCD prevention and control in the WHO South-East Asia Region," Jan. 21, 2026. [Online]. Available: https://www.who.int/publications/i/item/sea-ncd-117, (Accessed: Oct. 30, 2026).

[4] American Diabetes Association Professional Practice Committee, "2. Diagnosis and Classification of Diabetes: Standards of Care in Diabetes—2024," *Diabetes Care*, vol. 47, suppl. 1, pp. S20–S42, Jan. 2024, doi: 10.2337/dc24-S002.

[5] N. Rajatanavin *et al.*, "Effective coverage of diabetes and hypertension: an analysis of Thailand's national insurance database 2016–2019," *BMJ Open*, vol. 12, no. 12, p. e066289, Dec. 2022, doi: 10.1136/bmjopen-2022-066289.

[6] B. Sakboonyarat and R. Rangsin, "Characteristics and clinical outcomes of people with hypertension receiving continuous care in Thailand: a cross-sectional study," *The Lancet Regional Health - Southeast Asia*, vol. 21, p. 100319, Feb. 2024, doi: 10.1016/j.lansea.2023.100319.

[7] C. Khanthavudh, A. Grealish, V. Tzouvara, and M. Leamy, "Supporting healthcare in rural communities in Thailand: An exploratory qualitative study to understand the role and current mental health practices of village health volunteers," *PLOS ONE*, vol. 20, no. 3, p. e0320559, Mar. 2025, doi: 10.1371/journal.pone.0320559.

[8] L. A. Suplotova and O. O. Alieva, "Evolution of blood glucose self-monitoring technology," *Diabetes Mellitus*, vol. 26, no. 6, pp. 566–574, Nov. 2023, doi: 10.14341/DM13063.

[9] K. Elman, J. Wainstein, M. Boaz, D. Jakubowicz, and Y. Bar-Dayan, "Random blood glucose screening at a public health station encouraged high risk subjects to make lifestyle changes," *International Journal of Clinical Practice*, vol. 71, no. 8, p. e12984, Aug. 2017, doi: 10.1111/ijcp.12984.

[10] D. C. Klonoff *et al.*, "The Diabetes Technology Society Error Grid and Trend Accuracy Matrix for Glucose Monitors," *Journal of Diabetes Science and Technology*, vol. 18, no. 6, pp. 1346–1361, Nov. 2024, doi: 10.1177/19322968241275701.

[11] E. H. Weissler *et al.*, "The role of machine learning in clinical research: transforming the future of evidence generation," *Trials*, vol. 22, no. 1, p. 537, Aug. 2021, doi: 10.1186/s13063-021-05489-x.

[12] H. Khadem, M. R. Eissa, H. Nemat, O. Alrezj, and M. Benaissa, "Classification before regression for improving the accuracy of glucose quantification using absorption spectroscopy," *Talanta*, vol. 211, p. 120740, May 2020, doi: 10.1016/j.talanta.2020.120740.

[13] B. Barman, H. K. Choudhury, and B. Jajodia, "Interpretable machine learning for diabetes risk prediction: a large-scale analysis of Indian national survey data," *Discover Public Health*, vol. 22, no. 1, p. 832, Dec. 2025, doi: 10.1186/s12982-025-01236-8.

[14] S. Ghimire, T. Celik, M. Gerdes, and C. W. Omlin, "Deep learning for blood glucose level prediction: How well do models generalize across different data sets?," *PLOS ONE*, vol. 19, no. 9, p. e0310801, Sep. 2024, doi: 10.1371/journal.pone.0310801.

[15] P. Boonkrong, S. Shakya, W. Kraunamkam, and T. Simmachan, "Dietary patterns and psoriasis severity in Thai patients: a machine learning approach for small sample data," *Scientific Reports*, vol. 15, no. 1, p. 33088, Sep. 2025, doi: 10.1038/s41598-025-17657-z.

[16] P. Dash, K. Farzia, D. Priyadarshini, and S. Nayak, "Ensemble neural network modelling for stratified HbA1c prediction: Integrating past glucose measurement as a predictor of glycaemic control," *Bioinformation*, vol. 21, no. 10, pp. 3941–3946, Oct. 2025, doi: 10.6026/973206300213941.

[17] R. P. Bhatta, "Diabetes Prediction Using Random Forest and XGBoost Machine Learning Algorithm," *Journal of Engineering Technology and Planning*, vol. 6, no. 1, pp. 88–103, Dec. 2025, doi: 10.3126/joetp.v6i1.87829.

[18] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[19] P. Achenbach *et al.*, "A classification and regression tree analysis identifies subgroups of childhood type 1 diabetes," *eBioMedicine*, vol. 82, p. 104118, Aug. 2022, doi: 10.1016/j.ebiom.2022.104118.

[20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.

[21] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.

[22] N. Turnbull, L. K. Nghiep, A. Butsorn, A. Khotprom, and K. Tudpor, "Machine learning models identify micronutrient intake as predictors of undiagnosed hypertension among rural community-dwelling older adults in Thailand: a cross-sectional study," *Frontiers in Nutrition*, vol. 11, p. 1411363, Jul. 2024, doi: 10.3389/fnut.2024.1411363.

[23] L. Laranjo, E. T. O'Hagan, H. Klimis, S. Marschner, and C. K. Chow, "Effect of a mobile health intervention optimised with artificial intelligence on blood pressure and other cardiovascular risk factors in adults with high blood pressure: rationale and design of My Intelligent Cardiac Assistant (MICArdiac) randomised controlled trial," *Open Heart*, vol. 13, no. 1, p. e003819, Apr. 2026, doi: 10.1136/openhrt-2025-003819.

[24] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, Mar.

2024, doi: 10.1038/s41598-024-56706-x.
[25] I. Naskinova, M. Kolev, D. Karova, and M. Milev, "Machine Learning-Based Blood Pressure Prediction Using Cardiovascular Disease Data: A Comprehensive Comparative Study," *Electronics*, vol. 15, no. 2, p. 312, Jan. 2026, doi: 10.3390/electronics15020312.
[26] E. Bisong, N. Jibril, P. Premnath, E. Buligwa, G. Oboh, and A. Chukwuma, "Predicting high blood pressure using machine learning models in low- and middle-income countries," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, p. 234, Aug. 2024, doi: 10.1186/s12911-024-02634-9.
[27] S. H. Hwang *et al.*, "Machine Learning–Based Prediction for Incident Hypertension Based on Regular Health Checkup Data: Derivation and Validation in 2 Independent Nationwide Cohorts in South Korea and Japan," *Journal of Medical Internet Research*, vol. 26, p. e52794, Nov. 2024, doi: 10.2196/52794.

BIOGRAPHIES OF AUTHORS



Pichit Boonkrong is an assistant professor of Applied Mathematics at the College of Biomedical Engineering, Rangsit University, and serves as senior advisor to the Standing Committee on Education, Parliament of Thailand. With over a decade of teaching experience, he emphasizes conceptual understanding and critical thinking in mathematics and statistics education, particularly in multidisciplinary applications. His research interests encompass mathematical and statistical modeling, data science, and machine learning to address real-world challenges in medical and healthcare contexts. He can be contacted at email: pichit.bk@rsu.ac.th.



Junwei Yang is a Ph.D. student in data analytics at the Department of Electrical Engineering, Chulalongkorn University. His research focuses on biomedical data analytics and smart healthcare applications. He possesses technical competence in signal processing, transformer-based models, survival analysis, medical language models, and medical image processing for diagnosis. He applies symbolic signal processing, electroencephalography (EEG) source localization, and International Classification of Diseases, 10th Revision (ICD-10) analysis to address complex biomedical challenges through data-driven healthcare innovation. He can be contacted at email: 6873005121@student.chula.ac.th.



Manaporn Chatchumni is an associate professor of nursing and director of the Health Welfare Office at Rangsit University. Her research program encompasses postoperative pain management, nursing education, critical care, end-of-life care, and gerontological nursing. Her scholarly work further addresses care science applications to optimize patient outcomes across diverse clinical settings. Committed to advancing nursing practice, she enhances pain management and nursing competency through evidence-based education. She can be contacted at email: manaporn@rsu.ac.th.



Montira Muanjan is a Ph.D. student at the Faculty of Nursing, Prince of Songkla University, Thailand. Her research focuses on digital health, nursing education, and healthcare innovation to improve elderly care and healthcare systems. Her previous studies have investigated early trauma detection, nurse burnout, and the association between adolescent social media addiction and depression. She integrates ambient assisted living (AAL) and digital health into nursing education, bridging practice, technology, and care to support healthy ageing. She can be contacted at email: 6810430002@psu.ac.th.