Evaluating Biomedical Features for Reactive Hypoglycaemia Prediction with Machine Learning

Frans Dany[(✉)] and Fifi Retiaty

Biomedical Research Center, The National Research and Innovation Agency (BRIN), Jakarta, Indonesia
fransdany1@gmail.com

Abstract. Reactive hypoglycaemia (RH) is a condition when post-prandial plasma glucose level drops, usually <70 mg/dl. Its exact cause is unknown and associated with risk of diabetes among non-diabetic individuals. Attempts are made to predict RH but met with difficulties due to varied data characteristics and inconsistent results among different studies. Hence, machine learning (ML) techniques serve as alternative to improve model accuracy on condition that more data are added. This study was aimed to evaluate ML algorithms in RH prediction based on limited data of biomedical features. Data of 1540 participants of biomedical cohort study of risk factors for non-communicable diseases in 2021 were selected. Binary RH was set to be target variable. Analysis was carried out using Orange software with five supervised ML algorithms, i.e., logistic regression, decision tree, random forest, support vector machine (SVM) and gradient boosting. The result showed that RH cases were found at 1.36%, one among these being diabetic and two had prediabetes. SVM gave the overall best performance with area under curve (AUC) of 0.733 compared to other algorithms. However, classifier evaluation metrics (F1, precision, recall) were much better if non-RH condition was selected as the target. RH prediction unexpectedly plunged their respective values to zero except for gradient boosting (F1 0.027, precision 0.060, recall 0.018), indicating large difference of samples between two categories of RH variable. As promising as it is, these results suggest that careful interpretation of ML-based modelling is still mandatory when discrepancy of sample size between classified groups is encountered.

Keywords: Reactive hypoglycaemia (RH) · machine learning (ML) · biomedical · cohort

1 Introduction

Reactive hypoglycaemia (RH) is a condition when post-prandial plasma glucose level drops, usually <70 mg/dl. This is detected when someone takes oral glucose tolerance test or 2-h postprandial glucose check after fasting within certain hours [1]. It is often triggered by anti-hypoglycaemic drugs particularly among elders on diabetes medication [2]. Although RH is frequently not severe and can be overcome quickly, it can be

troublesome and sometimes harmful if happens in a wrong time and place without any bystanders noticing. There are several factors which can flare up this condition such as hormonal factors, history of bypass gastric surgery and so forth but most often, exact or notable causes of RH is unknown [3]. In addition, RH is also associated with risk of diabetes among non-diabetic individuals due to increased insulin sensitivity during fasting state. The abnormal outburst of insulin secretion after one takes glucose load in oral glucose tolerance test (OGTT) may lead to pancreatic beta cell exhaustion in long terms which is then associated with insulin resistance [4]. Therefore, it is not surprising if abnormalities in some biomedical parameters may also be related to risk of RH [5].

Attempts are made to predict RH but met with difficulties due to varied data characteristics. The incidence of RH is relatively low in several studies, making the prediction harder using common statistical approach especially when certain assumptions are not met [6]. Machine learning offers a promising alternative in predictive modelling since the process of finding the most optimal pattern is carried out automatically based on our data characteristics by using several algorithms and not restricted to specific statistical assumptions [7]. Nevertheless, the predictive accuracy of ML-based model gets higher when more data are added progressively into the existing system [8]. This study was aimed to evaluate the ML-based modelling of several biomedical variables or features to predict RH with limited data taken from biomedical cohort study of risk factors for non-communicable diseases in Bogor 2021. TheThis study results of this analysis can be used to improve risk factor-based screening method for RH which can later become the basis for policies in health programs.

2   Materials and Methods

This was a cross-sectional study using data collected during the Biomedical Cohort Study of Risk Factors for Non-Communicable Diseases which took place in Bogor, 2021. Ethical approval (reference number: LB.02.01/2/KE.169/2021 was obtained from the Ethical Committee of NIHRD (Balitbangkes), Ministry of Health, Indonesia and the research team took prior written informed consent from the willing participants. Up to 1545 respondents participated in the study, yet 1540 managed to follow the entire process of biomedical data collection and these were selected for final analysis. Binary RH was set to be target variable and its classification followed the criteria of hypoglycaemia in Consensus of PERKENI or Indonesian Association of Endocrinologists 2021 [9]. Meanwhile, age group (<45 vs ≥ 45 years old), sex, level of lipid profile (total cholesterol, low density lipoprotein/LDL, high-density lipoprotein/HDL and triglyceride), categorical variable of estimated glomerular filtration rate (eGFR) according to CKD-EPI formulation (≥ 90 vs < 90 ml/min/1.73 m^2), and triglyceride-glucose index (TyGi) and fasting plasma glucose were selected as biomedical features. The calculation of TyGi the was based on the formula proposed by Simental et al. [10]. Analysis was carried out using Orange software version 3.32 with five known supervised ML algorithms which included logistic regression, decision tree, random forest, support vector machine (SVM) and gradient boosting. These five algorithms were chosen due to their robust performance in terms of labelled classification and their practical availability in Orange software. The Test and Score widget was picked to determine model accuracy estimation

Table 1. Characteristics of all participants in this study.

Variables	n	Proportion
Age		
<45 years old	462	30.0
≥45 years old	1078	70.0
Gender		
Male	466	30.3
Female	1074	69.7
eGFR group		
≥90 ml/min/1.73 m^2	306	19.9
<90 ml/min/1.73 m^2	1234	80.1
Reactive hypoglycaemia (RH)		
No	1519	98.62
Yes	21	1.38
Total cholesterol (mg/dl)	202.82 ± 37.87*	
HDL (mg/dl)	50.88 ± 13.15*	
LDL (mg/dl)	125.17 ± 30.30*	
Triglycerides (mg/dl)	130.25 ± 101.60*	
Fasting plasma glucose (mg/dl)	105.15 ± 40.65*	
2-h OGTT (mg/dl)	139.25 ± 61.43*	
TyG index (TyGi)	8.61 ± 0.69*	

*Values are displayed as means ± standard deviation and calculated from 1540 respondents

by using stratified random sampling approach, 80:20 ratio of training to test proportion and 100 repetitions. Besides area under curve (AUC) and classification accuracy, values for other classifier evaluation metrics such as F1, precision and recall were also displayed. The Explain Model widget was also added to help visualization of potentially significant predictors from the most optimal algorithm.

3 Results

Most participants in this study were already above 45 years old and dominated by female. Those who experienced reactive hypoglycaemia (RH) only less than 2% (Table 1). RH respondents were predominantly male without having prediabetes nor diabetes mellitus at least for the time point we performed the biomedical data collection (Table 2).

The analysis of ML-based modelling shows that support vector machine gave the highest overall performance based on the value of area under curve (AUC) and its value did not change when RH is switched to non-RH condition. The situation is pretty much the same with CA, however the opposite was noticeable for other classification evaluation metrics, F1, precision and recall.

Table 2. Characteristics of respondents with reactive hypoglycaemia (RH).

Variables	n	Proportion
Age		
<45 years old	10	47.6
≥45 years old	11	52.4
Gender		
Male	14	66.7
Female	7	33.3
eGFR group		
≥90 ml/min/1.73 m^2	7	33.3
<90 ml/min/1.73 m^2	14	66.7
Hyperglycaemic status		
Non-preDM/DM	18	85.7
Prediabetes (PreDM)	2	9.5
Diabetes mellitus (DM)	1	4.8

Table 3. Results of the Test and Score evaluation metrics for each algorithm and their comparison between reactive hypoglycaemia (RH) and non-RH group.

Algorithms	AUC	CA	F1	Precision	Recall	Classification
Decision Tree	0.500	0.987	0.000	0.000	0.000	RH
SVM	0.733	0.987	0.000	0.000	0.000	RH
Random Forest	0.587	0.987	0.000	0.000	0.000	RH
Logistic Regression	0.714	0.986	0.000	0.000	0.000	RH
Gradient Boosting	0.695	0.984	0.027	0.060	0.018	RH
Decision Tree	0.500	0.987	0.993	0.987	1.000	Non-RH
SVM	0.733	0.987	0.993	0.987	1.000	Non-RH
Random Forest	0.587	0.987	0.993	0.987	1.000	Non-RH
Logistic Regression	0.714	0.986	0.993	0.987	0.999	Non-RH
Gradient Boosting	0.695	0.984	0.992	0.987	0.996	Non-RH

The accuracy was much better for non-RH group with F1 and precision value approaching 1 and perfect recall score in 3 algorithms (Table 3) while potential biomedical features from SVM model plot is shown in Fig. 1.
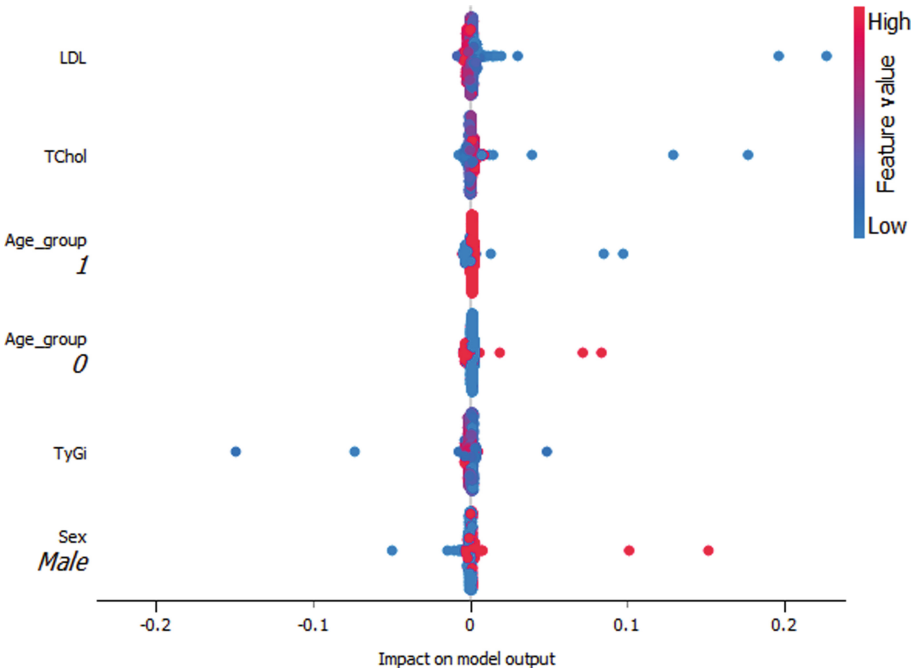
Fig. 1. Support vector machine (SVM) model plot of potential biomedical features for reactive hypoglycaemia (RH) prediction.

4 Discussion

In this study, we showed that the proportion of reactive hypoglycaemia (RH) among cohort respondents was relatively low, however most cases were found among those without prediabetes or diabetes mellitus (DM). As shown in Table 2, most RH cases were among men and those with decreased kidney function and as expected, this result was partly conflicting with other studies in terms of risk factor prediction for RH [6, 11]. Statistical modelling for a given outcome can be tricky as certain statistical assumptions should be met and prediction may depend on sample size, number of variables, existence of outliers and other data characteristics [12]. There are still inconsistent results of significant predictors for RH among several studies [13, 14]. Modelling via machine learning (ML) provides promising alternative when it comes to outcome prediction as it is more flexible in performing algorithms by taking into account data structure besides other properties [15].

Since the outcome variable was labelled with two categories, supervised learning method was more appropriate if the main purpose is to seek model classification [16]. Several known algorithms for this include decision tree, random forest, support vector machine (SVM), logistic regression which is frequently used in statistical software, and gradient boosting. To perform this one may need to learn essential coding of programming languages like Python, R or others. Another way that is more practical especially for beginners is to utilize built-in software and Orange becomes a great alternative since

it's free, user-friendly and visually pleasing [17]. Nevertheless, predicting RH in this study was still proven to be difficult given the relatively low value of area under curve (AUC) across five ML algorithms. Support vector machine (SVM) gave the highest yield at 0.733 and this value was still deemed poor when it comes to predictive accuracy [18]. SVM involves position adjustment of an imaginary line referred as hyperplane that segregates two classified groups at the most optimal results it can get. This algorithm also solves several issues not only in classification but also regression or non-linear problems, which may explain its high performance [19].

When we look more closely at the evaluation metrics, AUC and classification accuracy (CA) were quite high for RH and non-RH but the opposite applied for other parameters, F1, precision and recall. AUC is calculated mainly based on true positive rate and false positive rate [20]. From the value of AUC from SVM, it is notable that some overlaps in classification still remained at certain extent and prediction score did not match with the actual label of our target variable. Meanwhile, CA is determined by sum of true positive and true negative divided by all 4 categories in a confusion matrix. Since the number of true negative (non-RH group) approached near perfection, it is understandable that this condition could 'overpower' limited performance in true positive prediction, making its value so high [21]. On the other hand, precision is the ratio of true positives divided by the sum of the true positives and false positives. Since the number of precision was zero, it can be implied that the value of RH prediction score for true positives was practically null. Recall is another term to denote sensitivity which is expressed by the ratio of true positives divided by the sum of the true positives and false negatives. Again, this value reached null for the same practical reason as that for precision metrics. F1 score is basically the harmonic mean of the precision and recall and will automatically give zero as both previously described metrics were null [21, 22].

As shown from Fig. 1, low density lipoprotein (LDL) and total cholesterol (TChol) seemed to be the most significant predictors for our model, but given its low feature value (blue), its prediction power for RH was weak. Surprisingly, those aged below 45 years and men were more prone to RH and triglyceride-glucose index (TyGi) as proxy for insulin resistance was not a considerable determinant, suggesting that other unknown risk factors, such as genetic or protein markers might contribute to RH rather than the biomedical features provided in our dataset [6, 23, 24]. Another interesting thing is that gradient boosting was the only algorithm which gave non-zero value although it was still extremely low. Gradient boosting iteratively builds a "stronger" model from a collection of "the weaker ones" by learning downsizes from each weaker model and this learning might give hit at some spots [26]. However, its performance was still not satisfactory as their overall prediction score across data points was low for RH. These findings suggest that precision, recall and automatically, F1 score, are more important metrics to detect imbalance classification since they only focus on the performance score for true positive as their numerator [22]. Discrepancy of sample size between two labels in our RH target variable supports the above assumption and this issue may be addressed by using several measures such as redefining classification in our target variable, exploring other learning methods besides adding other biomedical features in our model [21, 27].

To summarize our preliminary analysis, support vector machine (SVM) gave fairly good in overall predictive performance for reactive hypoglycaemia (RH). But, the evaluation metrics were of much higher value in non-RH group compared to RH, indicating large difference of sample size between two categories of RH variable. Although machine learning (ML)-based modelling looks very promising, these results suggest that ML-based modelling must be carefully interpreted when limited data and discrepancy of sample size between classified groups are encountered. One way to improve this modelling is to follow up the data specific in RH group from the baseline to the most current observation period and more related determinants for RH are added in the analysis. Other machine learning algorithms including unsupervised and deep learning methods, such as convolutional and recurrent neural networks, can be implemented to investigate the hidden patterns between variables that were not detected otherwise by the supervised ones.

References

1. AgiostratidouG, AnhaltH, BallD, BlondeL, GourgariE, HarrimanKN, et al. Standardizing clinically meaningful outcome measures beyond HbA1c for type 1 diabetes: A consensus report of the American Association of Clinical Endocrinologists, the American Association of Diabetes Educators, the American Diabetes Association, the Endocrine Society, JDRF International, The Leona M. and Harry B. Helmsley Charitable Trust, the Pediatric Endocrine Society, and the T1D Exchange. Diabetes Care. 2017;40(12):1622–30.
2. AbdelhafizAH, Rodríguez-MañasL, MorleyJE, SinclairAJ. Hypoglycemia in older people - A less well recognized risk factor for frailty. Aging Dis. 2015;6(2):156–67.
3. BrunJF, FedouC, MercierJ. Postprandial reactive hypoglycemia. Diabetes Metab. 2000;26(5):337–51.
4. KittahNE, VellaA. Pathogenesis and management of hypoglycemia. Eur J Endocrinol. 2017;177(1):R37–47.
5. AhmadI, ZelnickLR, BatacchiZ, RobinsonN, DigheA, Manski-NankervisJAE, et al. Hypoglycemia in people with type 2 diabetes and CKD. Clin J Am Soc Nephrol. 2019;14(6):844–53.
6. HallM, WalickaM, PanczykM, TraczykI. Metabolic parameters in patients with suspected reactive hypoglycemia. J Pers Med. 2021;11(4).
7. SeoW, LeeYBin, LeeS, JinSM, ParkSM. A machine-learning approach to predict postprandial hypoglycemia. BMC Med Inform Decis Mak. 2019;19(1):1–13.
8. KokolP, KokolM, ZagoranskiS. Machine learning on small size samples: A synthetic knowledge synthesis. Sci Prog. 2022;105(1):1–16.
9. Perkumpulan Endokrinologi Indonesia (PERKENI). Pedoman Pengelolaan dan Pencegahan Diabetes Melitus Tipe 2 Dewasa di Indonesia 2020. (2020). PB PERKENI. PERKENI. 2020;46.
10. Simental-MendíaLE, Guerrero-RomeroF. The correct formula for the triglycerides and glucose index. Eur J Pediatr. 2020;179(7):1171.

11. PrévostMS, Rabasa-LhoretR, TalboMK, YardleyJE, CurryEG, BrazeauAS. Gender Differences in Strategies to Prevent Physical Activity-Related Hypoglycemia in Patients With Type 1 Diabetes: A BETTER Study. Diabetes Care. 2022;45(3):e51–3.
12. NimonKF. Statistical assumptions of substantive analyses across the general linear model: A mini-review. Front Psychol. 2012;3(AUG):1–5.
13. LvX, FangK, HaoW, HanY, YangN, YuQ. Identification of reactive hypoglycemia with different basic bmi and its causes by prolonged oral glucose tolerance test. Diabetes, Metab Syndr Obes Targets Ther. 2020;13:4717–26.
14. VuL, KefayatiS, IdéT, PavuluriV, JacksonG, LattsL, et al. Predicting Nocturnal Hypoglycemia from Continuous Glucose Monitoring Data with Extended Prediction Horizon. AMIA . Annu Symp proceedings AMIA Symp. 2019;2019:874–82.
15. BzdokD, AltmanN, KrzywinskiM. Points of Significance: Statistics versus machine learning. Nat Methods. 2018;15(4):233–4.
16. SaravananR, SujathaP. Algorithms : A Perspective of Supervised Learning Approaches in Data Classification. 2018 Second Int Conf Intell Comput Control Syst. 2018;(Iciccs):945–9.
17. NaikA, SamantL. Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime. Procedia Comput Sci. 2016;85(Cms):662–8.
18. BowersAJ, ZhouX. Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education Outcomes. J Educ Students Placed Risk. 2019;24(1):20–46.
19. BhavsarH, PanchalMH. A Review on Support Vector Machine for Data Classification. Int J Adv Res Comput Eng Technol. 2012;1(10):2278–1323.
20. Hajian-TilakiK. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Casp J Intern Med. 2013;4(2):627–35.
21. KaurH, PannuHS, MalhiAK. A systematic review on imbalanced data challenges in machine learning: Applications and solutions. ACM Comput Surv. 2019;52(4).
22. YacoubyR, AxmanD. Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models. 2020;79–91.
23. ItariuBK, ZeydaM, PragerG, StulnigTM. Insulin-like growth factor 1 predicts post-load hypoglycemia following bariatric surgery: A prospective cohort study. PLoS One. 2014;9(4):1–9.
24. 23. 2014_Insulin-Like Growth Factor 1 Predicts Post-Load Hypoglycemia following Bariatric Surgery.pdf.
25. TorekovSS, IepsenE, ChristiansenM, LinnebergA, PedersenO, HolstJJ, et al. KCNQ1 long QT syndrome patients have hyperinsulinemia and symptomatic hypoglycemia. Diabetes. 2014;63(4):1315–25.
26. BentéjacC, CsörgőA, Martínez-MuñozG. A Comparative Analysis of XGBoost. 2019;1–20.
27. LiuT, FanW, WuC. A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. Artif Intell Med. 2019;101:101723.