Machine Learning Classification for Blood Glucose Performances Using Insulin Sensitivity and Respiratory Scores in Diabetic ICU Patients

Athirah Abdul Razak[1(✉)], Radiyati Umi Partan[2], Normy Norfiza Razak[1(✉)], Asma Abu-Samah[3], Norliyana Nor Hisham Shah[1], and Mohd Shahnaz Hasan[4]

[1] Universiti Tenaga Nasional, Kajang 43000, Selangor, Malaysia
Normy@uniten.edu.my
[2] Universitas Sriwijaya, Palembang, Jakarta 30128, Indonesia
[3] Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
[4] University Malaya Medical Centre, Lembah Pantai, 59100 Kuala Lumpur, Malaysia

Abstract. Diabetes Mellitus (DM) patients with acute respiratory failure in the Intensive Care Unit (ICU) are susceptible to hyperglycaemia with adverse outcome of mortality. Clinically, Partial Pressure of Oxygen over a Fraction of Inspired Oxygen (P/F) scores is use as an indicator for acute respiratory failure and studies have shown that Insulin Sensitivity (S_I) can be used as the glycaemic control biomarker for DM patients. Since the elevation of blood glucose in ICU patients is linked to the progression of the acute respiratory system, this preliminary study initiates the combination of S_I, P/F, and DM status as the main predictors for machine learning classification. This assessment was done to identify which classification models and predictors between insulin sensitivity (S_I), (P/F) scores, and diabetic status will give higher accuracy on Blood Glucose (BG) performance with 7 types of classifier models. In total, 5684 total inputs from 3 predictors extracted from 76 ICU patients were split into 80:20 ratio for training and test sets with five-fold cross-validations. BG performances using three predictors from training vs. test data show that the k-Nearest Neighbor and Neural Network classifiers showed that the highest accuracies achieved were 54.1% and 54.5%, respectively. The sensitivity and specificity evaluated for both model's robustness demonstrated the possibility of using k-Nearest Neighbor and Neural Network for future BG performance prediction. Based on the model's robustness increment result, 8% vs. 12% and 10% vs. 4% shows a possibility that S_I, P/F scores, and DM can be utilized together as an input to classify glycemic level using both classifier models with a larger dataset from respiratory failures patients.

Keywords: Classification models · Machine learning · Diabetes mellitus · Insulin Sensitivity · Respiratory score · Blood glucose performance

1 Introduction

Critical care patients with or without history of diabetes are commonly linked to hyperglycemia (Blood Glucose (BG) > 11.1 mmol/L) [1, 2] during their stay at the Intensive

Care Unit (ICU). This condition may occur when the cortisol and counterregulatory hormones increased with insulin resistance [3]. These patients, especially the diabetics, are additionally vulnerable to multiple organ failures, sepsis, or infection, and worst, a higher rate of mortality and morbidity [4]. The pandemic of Coronavirus Disease (COVID-19) cases raised concerns as 42.5% of critically ill patient's mortality had Diabetes Mellitus (DM) [5]. Lim et al. [5] showed that 9.8% of COVID-19 patients have a comorbidity of DM and from the study, 1.2% mortality rate was reported. To date, From The Desk of the Director General of Health Malaysia report as of 27 July 2021, 1,044,071 Malaysian have been detected positive with COVID-19 and the numbers of mortality are 8,201 (0.78%) in cumulative [6]. This report [7] illustrated that out of 125 mortality cases, 52% had DM. Moreover, COVID-19 also had similar symptoms as a Severe Acute Respiratory Syndrome (SARS) [8]. Additionally, according to the Malaysian Registry of Intensive Care quarterly report, 35.8% of the ICU admitted patients suffer from multiple organ failures in the first 24 h of admission. Respiratory failures represented one-third of the single organ failure by patient, with 21.9%. In the first 24 h of admissions, 15.7% patients were reported to have Acute Respiratory Distress Syndrome (ARDS) [9]. Few studies have also demonstrated that respiratory failures are relatively high in acute patients with diabetes [10–12]. From these studies, an interrelationship and correlation were observed between respiratory failure and diabetes. Logette et al., [13] showed evidence that linked elevated BG to COVID-19 patients. When the glycemic level of diabetic patients is uncontrolled, these patients are susceptible to various adverse outcomes such as multiple organ failures. Thus, if the early prediction of BG level in critically ill diabetic patients can be known using daily medical data available from the ICU charts to ease glycaemic control, the occurrence of respiratory failure might be improved in critically ill patients.

Respiratory failure can be assessed with P/F scores of oxygenation index based on the Partial Pressure of Oxygen (PaO_2) and Fraction of Inspired Oxygen (FiO_2) [14]. One of the standard procedures in diagnosing diabetes is to measure insulin resistance and through euglycaemic clamp [15], but this method can be time-consuming for ICU use [16]. Therefore, Insulin Sensitivity (S_I) estimation is used to replace insulin resistance for glycemic control as S_I reflects the inversed parameter of insulin resistance. Moreover, S_I has been suggested as biomarker in several various glycemic control studies [17–19]. In achieving the targeted BG performances, S_I was used in stochastic targeted studies [20–23] for glycaemic control, predicting the BG within 5 to 95%. Since S_I has been used for glycaemic control, we hypothesized that it can predict the BG performance by exploiting medical data available in the ICU charts. Additionally, P/F scores can give an early prediction for glycaemic control by stratifying the BG in time of the target bands using classification technique.

A study [24] has shown that there might be an association between insulin resistance and lung dysfunction. Therefore, there is a need to examine if respiratory P/F score and S_I information can be used together as strong factors in predicting ICU glycemic control, especially for diabetic patients. Thus, this paper's objective is to assess classification of ICU patients' BG performance by classifying three different BG bands using P/F score, S_I, and diabetes mellitus (DM) status as predictors. In order to identify which model has the best performance accuracy, seven different classifiers are compared: Decision Tree, k-

Nearest Neighbor (kNN), Support Vector Machine (SVM), Naïve Bayes, Discriminant, Ensemble and Neural Network model. This preliminary study was done to identify which classifier models are more feasible towards glycaemic level prediction if P/F scores and DM status are included in model-based glycaemic control.

2 Materials and Method

2.1 Study Population

Retrospective patients' data for the case study in total were 76 (47 DM and 29 NDM) with 10704 h of length of stay and collected from University Malaya Medical Centre (UMMC) in 2018. These patients' dataset has examined the following factors: P/F score (in k/Pa), S_I (continuous in L/mU.min), diabetes mellitus (DM) status (binary, with 1 for DM and 0 for Non-DM), and the BG performance (three discrete classes). Instead of using a daily worst representative P/F score as usually done for Sequential Organ Failure Assessment (SOFA) score to decide on respiratory failure, P/F scores were extracted directly from the clinical chart whenever they were available. FiO_2 data were assumed similar to the previous value whenever the value of PO_2 and BG are available. On average, patients have three recorded P/F scores daily. Using BG measurements, provided nutrition and insulin, hourly S_I (L/mU.min) were fitted first with Intensive Care Insulin Nutrition Glucose (ICING) model [25] through integral fitting process method [26]. Instead of using a constant value to represent individual insulin resistance, S_I was estimated from the fitting process method to represent the patient's hourly varying metabolism.

Patients demographics are shown in Table 1. Patients' age and Acute Physiology and Chronic Health Evaluation (APACHE II) scores, S_I and P/F scores are presented in the average and standard deviation ($\pm$SD). Rank sum p-value test was used to determine the differences in age demographics (years old), the total of ICU stay (hours) distribution, S_I and P/F score. A P-value less than 0.05 is considered significantly different. The distribution of age and the total hours of ICU stay show no significant difference. Patients with diabetes status were presented by 11 and 7 female patients in the training and test datasets, respectively. In total, 1902 rows of input variables were divided randomly into training and testing data, with a ratio of 80:20. Three of the attribute variables, that is, S_I, DM, and P/F scores, were used as the predictors also known as features, and BG performance with three different classes was selected as the desired output to represent BG in time of the target bands. The data were divided into two partitions, where 4269 data points (1423 rows $\times$ 3 input variables) of the data were used for training and 1437 (479 rows $\times$ 3 input variables) for testing. Five-fold k cross-validation was used in training data to validate the accuracy of the model during the learning process.

2.2 Classification Learners

The seven classifiers model used in the study were supervised machine learning techniques [27]. The seven classifiers model used in training the dataset were Decision Tree (Model 1), Discriminant (Model 2), Naïve Bayes (Model 3), Support Vector Machine (Model 4), k-Nearest Neighbor (Model 5), Ensemble (Model 6) and lastly Neural Network (Model 7). Classification is often being used as the prediction for medical diagnosis

Table 1. Patients demographic for train and test data.

Demographics	Total	Train data	Test data	*P*-value
Number of patients (%)	76 (100%)	61 (80.2%)	15 (19.8%)	–
Total ICU stay (hours)	10704	8016	2688	Not significant
Mean (±SD) of age (years old)	59 (±14)	58 (±15)	61 (±9.7)	Not significant
Gender (%) • Male • Female	56 (73.7%) 20 (26.3%)	47 (77%) 14 (23%)	9 (60%) 6 (40%)	–
Ethnicity • Malay • Chinese • Indian • Foreigner	26 (34.2%) 30 (39.4%) 19 (25%) 1 (1.4%)	22 (35.4%) 24 (39.3%) 14 (22.9%) 1 (2.4%)	4 (26.7%) 6 (40%) 5 (33.3%) –	–
Diabetes status (%)	47 (61.8%)	38 (62.2%)	9 (40%)	–
Respiratory failures (%)	16 (21%)	15 (24.5%)	1 (6.6%)	–
Mean (±SD) of APACHE II Scores	18 (±7)	17 (±7)	20 (±7)	0.293
Mean (±SD) of PaO_2/FiO_2 Scores	299 (±130)	299 (±132)	300 (±132)	0.8879
Mean (±SD) of S_I	$3.9e^{-4}$ $(\pm 5.7e^{-4})$	$3.6e^{-4}$ $(\pm 4.7e^{-4})$	$4.9e^{-4}$ $(\pm 8.4e^{-4})$	0.8879
Number of data input (rows)	5684 (1902)	4247 (1423)	1437 (479)	–

[28], and showed efficiency in using small data for decision-making, [29]. Haque *et al.* showed an observation between 8 different machine learning algorithm performance for diabetes neuropathy diagnosis and Random Forest had outperformed among all those classifiers techniques [30]. Meanwhile, Singh *et al.* [28] showed that out of 8 classifiers used, the SVM classifier had the best accuracy to detect the relationship between diabetes and hypertension. From the previous studies, there are various types of classifiers that had classified patients with diabetes status. However, our focus of classification in this study is to observe BG performance with 7 types of classifiers using S_I, P/F, and DM as input predictors within diabetic ICU patients. P/F was included as one of the features to represent the respiratory score of the patients with 7 types of classifiers. BG performances as the desired output were classified into three BG level target ranges:

- Class 1: BG ≤ 7.7 mmol/L
- Class 2: BG = 7.8 to 11.1 mmol/L
- Class 3: BG ≥ 11.2 mmol/L

Figure 1 shows the framework for the usage of difference classifiers using MATLAB version 2021a.

Model accuracy results was demonstrated in Eq. 1. The confusion matrix represented will be used to illustrate the model with the highest performance. From the confusion matrix, the cohort train and test data results of True Positive (TP), True Negative (TN),

Fig. 1. Framework for classification learner

False Positive (FP), and False Negative (FN) were classified. To further test the classifier model robustness, sensitivity and specificity are calculated as represented in Eqs. 2 and 3, respectively. The nearer the value of sensitivity to 1, the more robust the classifier.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \times 100\% \tag{3}$$

3   Results and Discussion

The accuracies of seven classifiers based on the 3 input predictors for train data are shown in Table 2. The test data accuracy was examined only with the three predictors as it is important to find out how all the three predictors especially P/F can influence the BG performance for critically ill DM patients. We hypothesized that P/F score, DM, and S_I could be used in classifying BG performances. When the three predictors of S_I-P/F-DM were used, the training Model 3, 4 and 7 showed the highest accuracy with results of 55.4%, 53.5% and 54.1% respectively. Meanwhile, when only two predictors of S_I-P/F were used, the training model results of 3 and 7 accuracies were observed and had decreased by 1.4%, and 0.6% respectively. The result for S_I-DM predictors shows that Models 5 and 7 had increased accuracy from 53.0% to 55% and 54.1% to 55.3% respectively.

At first, amongst all classifiers, the Kernel Naive Bayes (Model 3), Fine Gaussian SVM (Model 4), and Narrow Neural Network (Model 7) classifiers give the highest performances compared to other model functions during the training. However, after the trained classifiers were tested, the accuracy shows that Models 2, 5 and 7 had better results using three predictors with 53.2%, 54.1%, and 54.5%, respectively. Despite having low prediction performance result during training, using all the three predictors indicate that P/F is feasible to be included in giving early prediction BG performance for patients with DM.

Figure 2 shows the tabulation of sensitivity and specificity of BG performance results for both data in k-Nearest Neighbour (Model 5) and Neural Network (Model 7) classifiers to compare model robustness with three predictors. These two classifiers were presented and further tested for model robustness as the results were the top two highest. For both training and test data, Model 5 had sensitivity of 0.62 and 0.54, respectively. Meanwhile, specificity for Model 5 shows 0.66 and 0.76 based on the test set, respectively. Then,

Table 2. Models accuracies.

Model	Train (Accuracy %)				Test (Accuracy %)
	SI-P/F-DM	SI-P/F	SI-DM	P/F-DM	
Model 1	52.4	52.1	54.1	50.7	51.5
Model 2	52.8	51.5	51.6	51.2	53.2
Model 3	**55.4**	**54.0**	54.8	51.2	52.6
Model 4	53.5	53.5	54.8	50.5	49.6
Model 5	53.0	51.1	**55.0**	50.8	54.1
Model 6	52.9	53.0	54.6	**51.4**	52.8
Model 7	54.1	53.5	55.3	50.9	**54.5**
Mean (±SD)	53.4 (±0.9)	52.6 (±1.0)	54.3 (±1.1)	50.9 (±0.2)	52.6 (±1.5)

Model 7 shows the sensitivity of 0.65 and 0.53 for train and test data. The specificity for Model 7 of train and test data are 0.63 and 0.67, respectively. Model 5 has better ability to stratify BG performance while Model 7 is more sensitive and robust classifying the predictor.
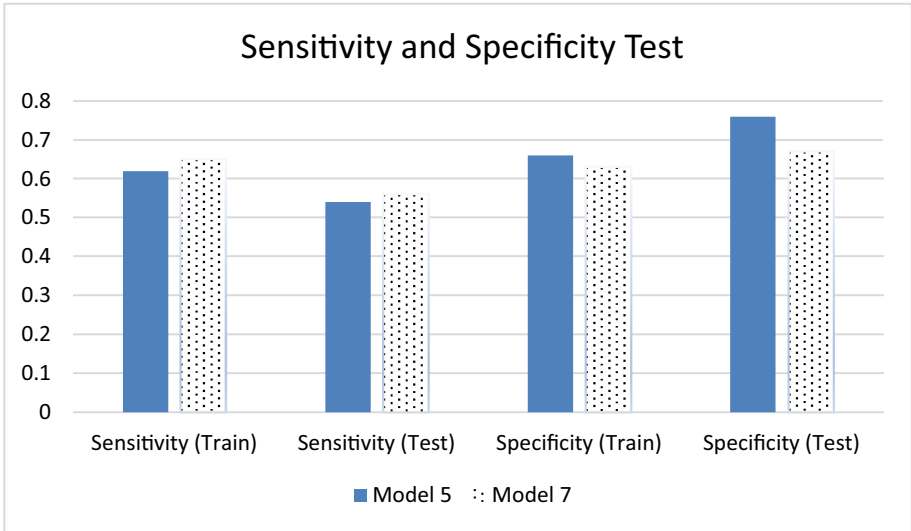

Fig. 2. Sensitivity and specificity for training and test data of k-Nearest Neighbour (Model 5) and Neural Network (Model 7).

Figure 3 shows that BG bands in Class 1,2 and 3 positively predicted in training data for Model 5 are 24, 680 and 50 respectively. Meanwhile, in test results for Model 5 are 33, 208 and 18. As for Model 7, 80, 582 and 108 of the predicted Class 1, 2 and 3 are true in train data. The test results that classified true for Class 1, 2 and 3 are 44, 192,

and 25. Our findings show that k-Nearest Neighbor (k-NN) had better classification for larger data such as Class 2. In comparison to k-NN, Neural Network classifier had higher accuracy in classifying Class 1 and 3 meaning that Neural Network is more feasible for data with small volumes. From the specificity values, there are increment of 10% and 4% for Model 5 than Model 7.
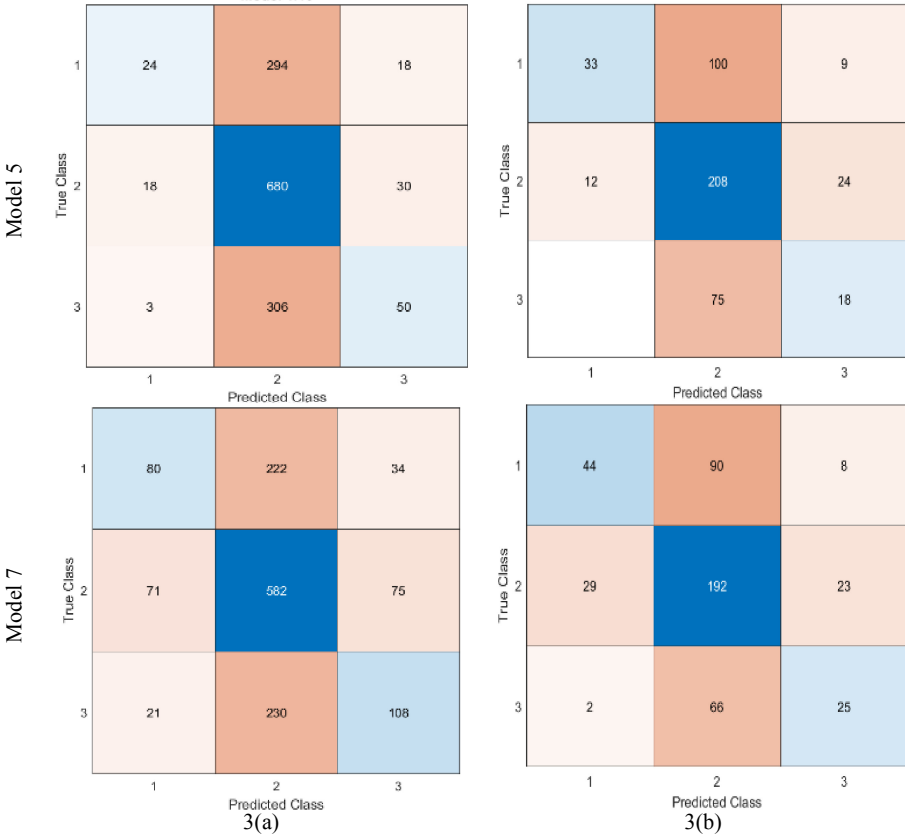

Fig. 3. Confusion matrix of the (a) training data (b) test data

Based on the classification results, TP predictions for both data partitions need to be calculated by two stages, one for model accuracy and followed by model sensitivity for robustness. This is because there are more than two classes to be classified. From the overall mean of classifier models, there is an increment of 4% from train to test data showing that the combination of S_I, P/F and DM status shows positive potential for classifying the BG within the time band. Moreover, sensitivity and specificity showed the highest test data accuracies for Model 5 and Model 7. This indicates an improvement in classifier robustness whereby the results are increased by 8% vs 12% and 10% vs 4% from train to test data. There are several possibilities that contribute to low prediction. First, a study by Razak et al. [31] suggested increasing the value of S_I to a physiologically

relevant one that represented the human biomarker. Our study showed that the value of S_I used most of the time had 0 value which is not acceptable. Thus, improving S_I to a physiologically relevant value range of $1e^{-3}$ to $1e^{-5}$ may also help improve the classifiers features. Second, the PaO2 and FiO2 data collected from ICU charts whenever both data are available. Thus, in this study, we assumed the data of FiO_2 to complete the PO_2 data. However, since S_I represented the hourly varying parameter of ICING model, the data collected for P/F score can be interpolated to hourly as well for future study to complement all the features used.

In summary, this study was done to classify the BG performance bands using S_I, P/F, and DM. P/F scores can be used to predict the classes of BG, this input predictor can be used as a non-invasive and time-consuming method to predict glycemic levels. In the future, if these attributes were selected for further research, P/F predictors can be partitioned earlier based on the different ranges of SOFA score or partitioning the S_I based on days of stays. Additionally, data collected for this study was before the occurrence of the pandemic and only 16 out 76 patients had pneumonia cause respiratory failures. In the future, it is interesting to study on how P/F can influence sample size that involved COVID-19 patients as well, as COVID-19 are categorized in the severe acute respiratory system.

4   Conclusion

This preliminary work showed that k-Nearest Neighbor and Neural Network models are feasible to classify BG performance. Although the model's accuracy is only 54.1% and 54.5%, P/F scores can be used to make an early prediction of glycemic level through classification. Based on the test result using S_I-P/F-DM as predictors, the accuracy was increased for both models. Similarly, model robustness test using sensitivity, and specificity demonstrated a total increase of 20% and 14% for both models. This increment shows that there is a likely possible relationship between P/F scores and diabetes status and may influenced BG performances. In future, if we have data set from COVID-19 patients, we can make early prediction for BG performance with the train model for those two classifiers.

1. References

1. McCowen, K.C., Malhotra, A., Bistrian, B.R.: Stress-induced hyperglycemia. Crit. Care Clin. **17**, 107–124 (2001)
2. Dungan, K.M., Braithwaite, S.S., Preiser, J.C.: Stress hyperglycaemia (2009)
3. Preiser, J.-C. (ed.): The Stress Response of Critical Illness: Metabolic and Hormonal Aspects. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-27687-8

4. Marik, P.E., Bellomo, R.: Stress hyperglycemia: an essential survival response! Crit Care **17**, 305 (2013)
5. Lim, B., et al.: Clinical characteristics and risk factors for severe COVID-19 infections in Malaysia: a nationwide observational study. Lancet Reg. Health - West. Pac. **4**, 100055 (2020)
6. D.G Of Health: From the Desk of the Director-General of Health Malaysia Kenyataan Akhbar KPK 31 Disember 2020 – Situasi Semasa Jangkitan Penyakit di Malaysia (2021). https://kpkesihatan.com/2020/03/11/kenyataan-akhbar-kpk-11-mac-2020-situasi-semasa-jangkitan-penyakit-coronavirus-2019-covid-19-di-malaysia/
7. D.G Of Health: From the Desk of the Director-General of Kenyataan Akhbar KPK 13 Julai 2021 – Situasi Semasa Jangkitan Penyakit Coronavirus (2021)
8. Zheng, X.y., Guan, W.j., Zhong, N.s.: Clinical characteristics of COVID-19 in developing countries of western pacific: low case-fatality rate unraveled. Lancet Reg. Health - West. Pac. **6**, 100073 (2021)
9. Ling, T.L., Har, L.C., Nor, M.R.M., Ismail, N.I., Ismail, W.N.W.: Malaysian Registry of Intensive Care Report (2016)
10. Edriss, H., Selvan, K., Sigler, M., Nugent, K.: Glucose levels in patients with acute respiratory failure requiring mechanical ventilation. J. Intensive Care Med. **32**, 578–584 (2017)
11. Ardigo, D., Valtuena, S., Zavaroni, I., Baroni, M.C., Delsignore, R.: Pulmonary complications in diabetes mellitus: the role of glycemic control. Curr. Drug Targets Inflamm. Allergy **3**, 455–458 (2004)
12. Abu-Samah, A., Razak, A.A., Razak, N.N., Suhaimi, F.M., Jamaludin, U.: The correlation of model-based insulin sensitivity and respiratory P/F score. In: Ibrahim, F., Usman, J., Ahmad, M.Y., Hamzah, N. (eds.) ICIBEL 2019. IP, vol. 81, pp. 54–62. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-65092-6_6
13. Logette, E., et al.: A machine-generated view of the role of blood glucose levels in the severity of COVID-19. Front. Public Health **9**, 1–53 (2021). https://doi.org/10.3389/fpubh.2021.695139
14. Villar, J., et al.: Assessment of PaO2/FiO2 for stratification of patients with moderate and severe acute respiratory distress syndrome. BMJ Open **5**, e006812 (2015)
15. Holzinger, U., Kitzberger, R., Fuhrmann, V., Funk, G.C., Madl, C., Ratheiser, K.: Correlation of calculated indices of insulin resistance (QUICKI and HOMA) with the euglycaemic hyperinsulinaemic clamp technique for evaluating insulin resistance in critically ill patients. Eur. J. Anaesthesiol. **24**, 966–970 (2007)
16. Muniyappa, R., Madan, R.: Assessing insulin sensitivity and resistance in humans. In: Endotext, pp. 1–21 (2000)
17. Abu-Samah, A., et al.: Model-based glycemic control in a Malaysian intensive care unit: performance and safety study. Med. Devices Evid. Res. **12**, 215–226 (2019)
18. Chase, J.G., et al.: Insulin sensitivity, its variability and glycemic outcome: a model-based analysis of the difficulty in achieving tight glycemic control in critical care. In: IFAC (2011)
19. Blakemore, A., et al.: Model-based insulin sensitivity as a sepsis diagnostic in critical care. J. Diabetes Sci. Technol. **2**, 468–477 (2008)
20. Stewart, K.W., et al.: Safety, efficacy and clinical generalization of the STAR protocol: a retrospective analysis. Ann. Intensive Care **6**(1), 1 (2016). https://doi.org/10.1186/s13613-016-0125-9
21. Abu-Samah, A., et al.: Model-based insulin-nutrition administration for glycemic control in Malaysian critical care: first pilot trial. In: Ibrahim, F., Usman, J., Ahmad, M.Y., Hamzah, N., Teh, S.J. (eds.) ICIBEL 2017. IP, vol. 67, pp. 189–196. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7554-4_33
22. Evans, A., et al.: Stochastic targeted (STAR) glycemic control: design, safety, and performance. J. Diabetes Sci. Technol. **6**, 102–115 (2012)

23. Fisk, L.M., Le Compte, A.J., Shaw, G.M., Penning, S., Desaive, T., Chase, J.G.: STAR development and protocol comparison. IEEE Trans. Biomed. Eng. **59**, 3357–3364 (2012)
24. Sagun, G., Gedik, C., Ekiz, E., Karagoz, E., Takir, M., Oguz, A.: The relation between insulin resistance and lung function: a cross sectional study. BMC Pulm. Med. **15**, 1–8 (2015)
25. Lin, J., et al.: A physiological intensive control insulin-nutrition-glucose (ICING) model validated in critically ill patients. Comput. Methods Programs Biomed. **102**, 192–205 (2011)
26. Hann, C.E., et al.: Integral-based parameter identification for long-term dynamic verification of a glucose-insulin system model. Comput. Methods Programs Biomed. **77**, 259–270 (2005)
27. Brink, H., Richards, J., Fetherolf, M.: Real-World Machine Learning. Manning Publications Co., Greenwich (2016)
28. Singh, N., Singh, P., Bhagat, D.: A rule extraction approach from support vector machines for diagnosing hypertension among diabetics. Expert Syst. Appl. **130**, 188–205 (2019)
29. Yahyaoui, A., Yumuşak, N.: Decision support system based on the support vector machines and the adaptive support. Biomed. Res. **29**, 1474–1480 (2018)
30. Haque, F., et al.: Performance analysis of conventional machine learning algorithms for diabetic sensorimotor polyneuropathy severity classification. Diagnostics **11**, 801 (2021)
31. Razak, A.A., Abu-Samah, A., Razak, N.N., Baharudin, S., Suhaimi, F.M., Jamaludin, U.: Endogenous glucose production variation assessment for Malaysian ICU patients based on diabetic status. In: Ibrahim, F., Usman, J., Ahmad, M.Y., Hamzah, N. (eds.) ICIBEL 2019. IP, vol. 81, pp. 129–136. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-65092-6_15