Computer Methods and Programs in Biomedicine 262 (2025) 108657

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

Methods for estimating resting energy expenditure in intensive care patients: A comparative study of predictive equations with machine learning and deep learning approaches

Christopher Yew Shuen Ang [a], Mohd Basri Mat Nor [b], Nur Sazwi Nordin [b], Thant Zin Kyi [c], Ailin Razali [b], Yeong Shiong Chiew [a,*]

[a] School of Engineering, Monash University Malaysia, Selangor, Malaysia
[b] Kulliyyah of Medicine, International Islamic University Malaysia, Pahang, Malaysia
[c] Innure Biotechnologies Sdn Bhd, Petaling Jaya, Selangor, Malaysia

ARTICLE INFO

Keywords:
Resting energy expenditure
Machine learning
Deep learning
Predictive equations
Indirect calorimetry
Feature importance analysis

ABSTRACT

Background: Accurate estimation of resting energy expenditure (REE) is critical for guiding nutritional therapy in critically ill patients. While indirect calorimetry (IC) is the gold standard for REE measurement, it is not routinely feasible in clinical settings due to its complexity and cost. Predictive equations (PEs) offer a simpler alternative but are often inaccurate in critically ill populations. While recent advancements in machine learning (ML) and deep learning (DL) offer potential for improving REE estimation by capturing complex relationships between physiological variables, these approaches have not yet been widely applied or validated in critically ill populations.

Methodology: This prospective study compared the performance of nine commonly used PEs, including the Harris-Benedict (H-B1919), Penn State, and TAH equations, with ML models (XGBoost, Random Forest Regressor [RFR], Support Vector Regression), and DL models (Convolutional Neural Networks [CNN]) in estimating REE in critically ill patients. A dataset of 300 IC measurements from an intensive care unit (ICU) was used, with REE measured by both IC and PEs. The ML/DL models were trained using a combination of static (i.e., age, height, body weight) and dynamic (i.e., minute ventilation, body temperature) variables. A five-fold cross validation was performed to assess the model prediction performance using the root mean square error (RMSE) metric.

Results: Of the PEs analysed, H-B1919 yielded the lowest RMSE at 362 calories. However, the XGBoost and RFR models significantly outperformed all PEs, achieving RMSE values of 199 and 200 calories, respectively. The CNN model demonstrated the poorest performance among ML models, with an RMSE of 250 calories. The inclusion of additional categorical variables such as body mass index (BMI) and body temperature classes slightly reduced RMSE across ML and DL models. Despite data augmentation and imputation techniques, no significant improvements in model performance were observed.

Conclusion: ML models, particularly XGBoost and RFR, provide more accurate REE estimations than traditional PEs, highlighting their potential to better capture the complex, non-linear relationships between physiological variables and REE. These models offer a promising alternative for guiding nutritional therapy in clinical settings, though further validation on independent datasets and across diverse patient populations is warranted.

1. Introduction

Critical illness is accompanied by a hypermetabolic state which leads to the activation of various catabolic hormones. This response elevates energy expenditure and accelerates the catabolism of glycogen storage, fats and muscles, significantly increasing the risk of malnutrition in intensive care unit (ICU) patients [1]. Consequently, malnutrition is prevalent in the ICU, affecting approximately 40 %-80 % of patients [2]. In recent years, nutritional therapy in critical illness is highly recognised as an essential component in managing critically ill patients [3]. Studies have demonstrated that malnutrition in these patients is significantly associated with poorer clinical outcomes, including higher rates of

* Corresponding author.
E-mail address: chiew.yeong.shiong@monash.edu (Y.S. Chiew).

https://doi.org/10.1016/j.cmpb.2025.108657
Received 14 October 2024; Received in revised form 27 January 2025; Accepted 8 February 2025
Available online 9 February 2025

complications, increased hospital costs and mortality. As such, providing adequate nutritional support is essential for improving malnutrition-related outcomes.

The effectiveness of nutritional support depends on accurately estimating and meeting the energy and protein needs of critically ill patients by prescribing and delivering appropriate nutrients. Therefore, determining energy requirements is vital as prescribed targets guide nutrition delivery. However, assessing energy needs in critically ill patients is challenging as the effects of illness, injury and stress on Resting Energy Expenditure (REE) are often varied and unpredictable [4,5]. In clinical practice, caloric requirements for critically ill patients are determined using indirect calorimetry (IC), which is considered the gold standard method for assessing energy expenditure by assessing oxygen consumption (VO_2) and carbon dioxide production (VCO_2) [6–8]. However, the process of obtaining IC measurements can be costly and time-consuming, making it impractical for routine use in most ICU practice [8].

Apart from IC, predictive equations (PEs) are the most commonly used method in estimating REE due to their simplicity [9]. Studies have shown PEs incorporating "dynamic" variables and respiratory data have demonstrated better concordance with REEs measured by IC compared to PEs developed for healthy adults or those based solely on "static" variables [10,11]. However, majority of the PEs in use have been repeatedly proven inaccurate and unacceptable for clinical use compared to REE measurements from IC [10,12]. Most importantly, these commonly used equations are not generally validated in patients with higher nutritional risk, and their inaccuracies become more pronounced in populations with extreme weight, advanced age, severe illness, or malnutrition, potentially leading to significant underfeeding or overfeeding [3].

Hence, machine-learning (ML) methods have emerged as a promising alternative to traditional PEs for estimating REE in critically ill patients. ML models have the potential to capture complex, non-linear relationships between numerous static and dynamic clinical variables that affect REE, including patient-specific factors such as respiratory parameters, vital signs, and metabolic trends. Recent work on ML-based methods for REE prediction in acute kidney injury (AKI) patients demonstrated improved accuracy over traditional PEs (r-value of 0.69 vs 0.24) by incorporating dynamlic clinical variables like ventilator parameters and biochemical markers [13]. However, root mean squared error (RMSE) of the best model was still relatively high at 490 calories, while applicability of the developed models is limited to severe AKI patients. Other studies investigating the use of artificial neural networks [14,15] and various other ML models [16] have achieved higher prediction accuracies (RMSE $\leq$ 210 calories). However, these studies typically compare their models with a limited set of predictive equations (PEs) and may not be widely applicable across various patient populations, particularly critically ill patients.

Currently, there is an urgent need for a safe, accurate, and personalised nutrition platform to assist intensive care physicians in providing optimal nutrition to critically ill patients. Determining the optimal feeding regimen for a patient involves complex considerations, and errors can result in either underfeeding or overfeeding, both of which are associated with prolonged hospital stays [8]. Although indirect calorimetry is available, a single reading may not be sufficient to establish daily REE goals accurately. This study investigates the ability of ML methods of REE estimation in ICU patients and compare its performance of model-based PEs [10,17–20]. More specifically, the performance of the ML models developed in this study for REE estimation is evaluated using three different methodological approaches:

1. **Comparative analysis of prediction methods:** The performance of ML/DL models was systematically compared with PEs to determine whether data-driven approaches offer superior predictive accuracy in ICU patients.
2. **Impact of feature selection and dataset composition:** ML/DL models were trained using datasets with varying parameter compositions to evaluate the influence of different physiological and clinical variables on predictive accuracy.
3. **Data augmentation and imputation techniques:** Strategies such as data stratification, augmentation, and imputation were examined to determine their effectiveness in addressing the limitations imposed by small dataset sizes and improving model generalisability.

This comparison clearly highlights the potential advantages and limitations of each method, offering valuable insights into their accuracy and feasibility for clinical use.

2. Methodology

2.1. Study design

A prospective observational study was conducted from October 2022 to October 2024 in the Intensive Care Unit (ICU) at Sultan Ahmad Shah Medical Centre @ IIUM (SASMEC), Kuantan, Pahang. The study protocol was approved by the IIUM Research Ethics Committee (IREC) (ID No: IREC 2020–042). Virtual data of 269 patients were also retrieved from patient records and previous studies, which had been previously approved by IREC (ID No: IREC 2020–042, IREC 2020–168). Informed consent was obtained from the patients or their legal representatives.

Clinical characteristics and anthropometric measurements were recorded upon admission, and vital signs were recorded during the REE measurement using Indirect calorimetry method. These parameters will be used for the PEs as well as ML models. The recorded parameters include: 1) age (years); 2) height (cm); 3) actual body weight, ABW (kg); 4) body mass index, BMI (kg/m^2); 5) ideal body weight, IBW (kg); 6) minute ventilation, MV (L/min); 7) Apache 11 scores; 8) gender; 9) body surface area, BSA (m^2); 10) body temperature ($^{\circ}C$); 11) SOFA scores; 12) SAPS2 scores; and 13) NUTRIC scores.

2.2. Measurement of REE – indirect calorimetry

REE measurements were performed using Indirect Calorimetry (Cosmed, Quark RMR 2.0, Indirect Calorimetry Lab, Italy). A standard protocol for conducting the measurement was followed [5,21]. Prior to each measurement, the metabolic monitor will be allowed to warm up for 30 min. Gas and flowmeter calibrations will be performed by an experienced healthcare personal. The REE (REE_{Cosmed}) will be recorded after 30 minutes of the non-fasting state according to standard protocol and manufacturer instructions. The first five minutes will be discarded routinely. Steady state is defined as having $<$ 10 % variation of oxygen consumption (VO_2) and carbon dioxide production (VCO_2) for 25 minutes IC measurement; or $<$ 5 % variation for 5 minutes IC measurement. A respiratory quotient (RQ) less than 0.7 or more than 1.0 may suggest inadequate IC measurement.

The REE (REE_{GE}) was also measured using GE Carescape R860 ventilator metabolic module (GE HealthCare Technologies, Inc, Chicago, Illinois). The carescape GE R860 metabolic module has a sensor incorporated in the breathing circuit to measure the concentration of oxygen and carbon dioxide in the patient's breath. It uses this data as well as other parameters including patient height, weight, age and gender to calculate the REE.

Furthermore, patients with both REE_{Cosmed} and REE_{GE} values (N = 127), the mean absolute differences in the REE values (abs[REE_{Cosmed} – REE_{GE}]) and the standard deviation of this absolute difference were calculated. This analysis was performed to determine the expected differences in REE values between the two gold-standard methods for the given patient cohort. Among the 269 patients in the cohort, a single IC measurement was obtained for each of the 238 patients, while the remaining 31 patients underwent two IC measurements per patient.

2.3. Model-based predictive equations (PEs)

REE was estimated using the predictive equations detailed in Table 1. These equations were developed from cohort analysis and were able to show reasonable prediction accuracy [10].

The REE values of the patients were determined using these PEs are compared to values determined using gold-standard measurement methods (REE_{Cosmed} and REE_{GE}) using the root mean square error (RMSE) metric:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \quad (1)$$

Where N is the number of samples, y_i is the actual value for the i-th observation, $\hat{y}_i$ is the predicted value for the i-th observation. The RMSE is the square root of the average of the squared differences between the actual and predicted values of REE. It provides a measure of how well a model's predictions match the actual data, with lower RMSE values indicating better model performance. The calculation of RMSE is performed with respect to the REE_{Cosmed} values whenever available, else REE_{GE} values are used.

2.4. Machine learning approaches

In this study, 3 different ML models and one deep learning (DL) model were chosen for REE prediction model development: 1) Extreme Gradient Boosting (XGBoost) [31,32]; 2) Random Forest Regressor (RFR) [32,33]; 3) Support Vector Regression (SVR) [34]; and 4) the DL model, Convolutional Neural Network (CNN).

Hyperparameter tuning of the XGB, RFR, and SVR models was performed using GridSearchCV (Scikit-learn) using a train-test split of 90:10. The tuned hyperparameters were used to train each model in a 5-fold cross-validation, where the data was repeatedly shuffled into training and non-repeating testing datasets, with an 80:20 train-test split ratio. A 5-fold cross-validation was also performed with the CNN model. In each fold, the RMSE (Equation 1) of each model was recorded.

The architecture of the CNN model developed in this study is presented in Fig. 1, where layer-specific hyperparameters are detailed in Table 2. The CNN model was trained with 100 epochs and a mini-batch size of 16.

2.4.1. Model training and performance analysis

All 4 models (ML and DL) were trained using combinations of input parameters consisting: 1) age (years); 2) height (cm); 3) actual body weight, ABW (kg); 4) body mass index, BMI (kg/m^2); 5) ideal body weight, IBW (kg); 6) minute ventilation, MV (L/min); 7) Apache 11 scores; 8) gender hotcode (male = 1 and female = 0); 9) body surface area, BSA (m^2); 10) body temperature (°C); 11) SOFA scores; 12) SAPS2 scores; and 13) NUTRIC scores.

BMI was also further split into 4 separate classes: underweight (BMI_category_Underweight, BMI < 18.5), normal weight (BMI_category_Normal weight, 18.5 ≤ BMI < 24.9), overweight (BMI_category_Overweight, 24.9 ≤ BMI < 29.9), and obese (BMI_category_Obese, BMI > 29.9). Body temperature was also further split into 3 classes: hypothermia (Temp_category_Hypothermia, < 36 °C), normal (Temp_category_Normal, ≥ 36 °C and ≤ 37.5 °C), and fever (Temp_category_Fever, > 37.5 °C).

All data (20 input variables and one output variable of IC) were scaled using StandardScaler (Scikit-learn) to ensure data standardisation, leading to improved model convergence and performance. In cases where variables used in model training include missing values, these patients with missing values were removed from the dataset. All analyses were performed on a workstation running an Intel Xeon W-1250P CPU (4.1 GHz), 32GB of ram, and a NVIDIA GeForce RTX2080 Super GPU.

Table 1
Model-based equations used to estimate resting energy expenditure (REE).

No.	Name	Equation	Results of original study/ Meta-analyses
1	Harris and Benedict (H-B1919) [17,20]	M: (wt × 13.7516) + (ht × 5.0033) − (age × 6.755) + 66.473 F: (wt × 9.5634) + (ht × 1.8496) − (age × 4.6756) + 655.0955	M: Derived from N = 136 patients; CL: ± 210.5.0 calories F: Derived from N = 169 patients; CL: ± 201.0 calories
2	Harris and Benedict (H-B1984) [20]	M: 13.397 × wt + 4.799 × ht − 5.677 × age + 88.362 F: 9.247 × wt + 3.098 × ht − 4.33 × age + 477.593	M: Derived from N = 168 patients; CL: ± 213.0 calories F: Derived from N = 103 patients; CL: ± 211.9 calories
3	Penn State [18,19,22]	0.96 x MSJ + 167 x Tmax + 31 x MV - 6212	Derived from N = 169 patients [23,24] In an analysis of 47 subjects (28 M, 19 F), the absolute difference of REE values relative to IC was 183 ± 131 calories [22]. Mean of differences: -207.57 ± 453.93 calories [25]
4	Modified Penn State [18,26]	0.71 x MSJ + 85 x Tmax + 64 x MV - 3085	Developed using N = 51 patients (R^2 = 0.66, $F_{47,3}$ = 30.1) [26] Difference in mean REE (calories) relative to IC measurements based on phases of critical illness [10]: 1) Acute: -151 (1630 ± 395 vs 1781 ± 483) 2) Late: -236 (1648 ± 372 vs 1884 ± 508) 3) Chronic: -205 (1651 ± 309 vs 1856 ± 445)
5	Ireton-Jones 1997 [19,27]	M: 1784 + (5 × wt) – (11 × age) + 244 F: 1784 + (5 × wt) – (11 × age)	Developed using N = 200 patients. Validated using N = 100 patients where the mean prediction error was 8 calories/day. [27] Mean of differences: 26.81 ± 461.16 calories [25]
6	TAH equation [11]	891.6 x ht + 9.0 x wt + 39.7 x MV − 5.6 x (Age) − 354	Derived from N = 294 patients R^2 = 0.442, RMSE = 348.3 calories, SEE = 325.6 calories
7	Mifflin-St Jeor equation (MSJ) [18,28]	MSJ, M: (10 × wt) + (6.25 × ht) - (5 × age) + 5 MSJ, F: (10 × wt) + (6.25 × ht) - (5 × age) - 161	Derived from N = 498 healthy subjects (251 M, 247 F), aged 19–78 years (45 ± 14 years). Mean of differences: −438.39 ± 431.97 calories [25]
8	Swinamer equation [29]	945 x BSA – 64 x age + 108 x Tmax + 24.2 x RR + 817 x Vt – 4349 Where BSA, body surface area (m^2) = 0.007184x ($ht^{0.725}$) x ($wt^{0.425}$)	Derived from N = 108 subjects RMSE: 11 – 13 % based on REE (calories) values of the following stratified groups [26]: 1) Non-obese young patients (N = 52): 2051 ± 342 2) Obese young patients (N = 47): 2596 ± 401 3) Non-obese elderly patients (N = 52): 1709 ± 320 4) Obese elderly patients (N = 51): 2138 ± 363
9	Weight-based equation [19,30]	Wt x 25	Mean of differences: 47.08 ± 557.74 calories [25]

*M: male, F: female, wt: weight (kg), ht: height (m), MV: minute ventilation (L/min), RR: respiratory rate (breaths/min), Tmax: maximum body temperature, Vt: tidal volume (L), MSJ: Mifflin-St Jeor equation (Equation 7), CL: confidence limits, SEE: standard error of estimate.

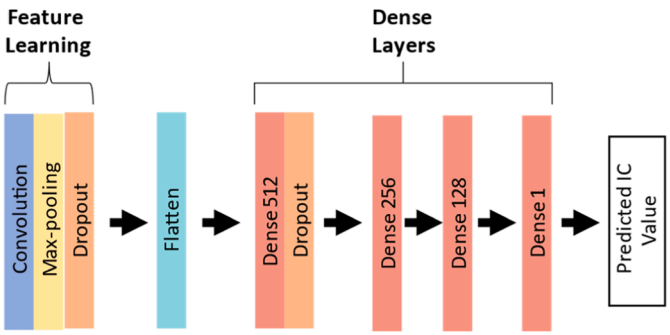

Fig. 1. CNN model architecture.

Table 2
Architecture and hyperparameters of the CNN model for IC prediction.

Stack	Layers	Type	DCNN
Feature Learning	1	Convolution-1	64 filters, kernel size: 2, activation: ReLU
	2	Max Pooling-1	Pool size: 2
	3	Dropout-1	0.1
	-	Flatten	-
	4	Dense	512 output neurons, activation: ReLU
	5	Dropout-2	0.1
	6	Dense	256 output neurons, activation: ReLU
	7	Dense	128 output neurons, activation: ReLU
	8	Dense	1 output neurons, activation: ReLU
	-	Optimiser	Adam: (1) Learning rate: 0.001, (2) Momentum term β1: 0.9, and (3) Momentum term β2: 0.009.
	-	Loss function	Mean Squared Error

2.4.2. Feature importances analysis

Feature importance quantifies the contribution of each input variable to the decision-making process of a ML model [35]. The feature importance of each input variable was determined for the RFR and XGBoost models to determine the contribution of each input variable to the model-predicted REE. Note, these plots are not presented for the SVR and CNN models as the model architectures do not allow for the extraction and ranking of each input variable.

Feature importance values are calculated as normalised scores ranging between 0 and 1, where the sum of all feature importance values across the feature set equals 1. For the RFR model, feature importance is derived from the mean decrease in impurity (MDI), also known as Gini importance to identify which variables contribute to predicting REE values [35,36]. It represents the average reduction in node impurity (variance for regression tasks) caused by a given feature across all trees in the ensemble. Features that contribute more to reducing variance have higher importance scores. For the XGBoost models, feature importance is calculated based on the Gain metric, which represents the average improvement in model performance when a feature is used for splitting [37]. Features that contribute more to enhancing model accuracy receive higher Gain values, indicating their relative importance in prediction. The feature importances were calculated using inbuilt python (3.10, Python Software Foundation, Wilmington, DE, USA) functions (Scikit-learn, *plot_importance*).

2.5. Data augmentation and stratification

Data for 269 ICU patients from the clinical cohort were collected and used for this study. Data augmentation was also performed to increase the dataset size. This was achieved by several methods listed below and illustrated in Fig. 2:

1. For a patient with both REE_{Cosmed} and REE_{GE} values, the patient is duplicated to form 2 independent patients (augmentation via duplication), one of which with the REE_{Cosmed} value, and another one with the REE_{GE} value. This augmented dataset is then further extended via:
   i. The dataset is doubled in size by synthesising data via GaussianCopulaSynthesizer (Synthetic Data Vault).
   ii. Missing values of variables within the original dataset ($N = 300$ IC measurements) are imputed using GradientBoostingRegressor (Scikit-learn, IterativeImputer).

The augmented patient dataset is also split into 2 groups: (a) patients with REE_{Cosmed} values, and (b) patients with REE_{GE} values. This stratification is performed to investigate if eliminating the inherent measurement differences between the two gold standard IC methods (Cosmed and GE) will improve model performance.

In cases where patients have both REE_{Cosmed} and REE_{GE} values, but are not stratified, REE_{Cosmed} values are used. Patient demographics, anthropometric measures and vital signs are summarised in Table 3 for all patients as well as the stratified cohorts. A summary of all datasets used in this study is detailed in Table 4. The implementation of 5-fold cross-validation for each dataset and model, as outlined in Table 4, ensures that 20 % of the dataset is allocated for quantitative model performance evaluation through REE prediction.

A pairwise Mann-Whitney U-test was also used to compare the statistical distributions of the: 1) original dataset (Dataset 1 from Table 4, $N = 300$), 2) stratified REE_{Cosmed} dataset (Dataset 17 from Table 4, $N =$

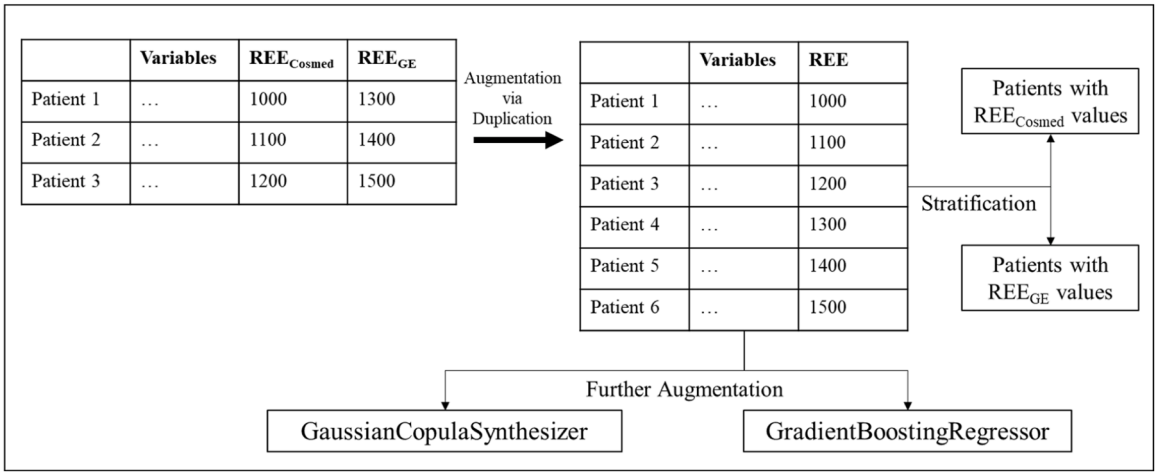

Fig. 2. Data augmentation and stratification flowchart.

Table 3
Patient demographics, anthropometric measures and vital signs.

Variable	All REE Measurements (N = 300)	Stratified REE_{Cosmed} (N = 176) [a]	Stratified REE_{GE} (N = 130) [a]
Gender (Male/ Female)	161/139	93/83	67/63
Age (Years)	62 [54 – 70]	63 [53 – 71]	63 [53 – 71]
Height (cm)	161 [155 – 169]	161 [155 – 168]	161 [155 – 170]
Actual body weight (kg)	68 [58 – 77]	70 [60 – 80]	70 [60 – 80]
BMI (kg/m^2)	27 [23 – 31]	26 [23 – 30]	26 [23 – 31]
Ideal body weight (kg)	57 [50 – 65]	57 [50 – 65]	57 [50 – 65]
Tmax (∘C)	36.7 [36.4 – 36.8]	36.7 [36.5 – 37.0]	36.7 [36.5 – 36.8]
Minute ventilation (L/ min)	7.3 [6.4 – 8.6]	7 [6.4 – 8.0]	7.2 [6.7 – 8.0]
Body surface area (m^2)	1.71 [1.58 – 1.87]	1.72 [1.60 – 1.87]	1.72 [1.60 – 1.88]
Respiratory rate (breaths/min)	18 [15 – 19]	18 [16 – 19]	18 [16 – 19]
Tidal volume, V_T (L)	0.42 [0.39 – 0.46]	0.42 [0.39 – 0.45]	0.42 [0.39 – 0.45]
APACHE 11	17 [13 – 22]	17 [12 – 21]	17 [14 – 21]
SOFA	7 [4 – 9]	6 [4 – 9]	6 [4 – 9]
SAPS2	39 [29 – 50]	38 [28 – 50]	39 [28 – 49]
NUTRIC	4 [3 – 6]	4 [2 – 5]	4 [2 – 5]

[a] Considering only patients with complete data for all variables

Table 4
Summary of all datasets used in this study.

No.	Input data	Dataset size	Remarks
1	Age, Height, ABW, BMI, IBW, MV	300	
2	Age, Height, ABW, BMI, IBW, MV, Apache 11	300	
3	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender	300	
4	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA	300	
5	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp	259	
6	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA	242	
7	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, NUTRIC	242	
8	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, SAPS2	179	
9	All*	179	
10	All*	306	AVD
11	All*, 4 BMI classes, 3 Body temp classes	306	AVD
12	All*	612	AVD, GCS
13	All*, 4 BMI classes, 3 Body temp classes	612	AVD, GCS
14	All*	427	AVD, GBR
15	All*, 4 BMI classes, 3 Body temp classes	427	AVD, GBR
16	All*	176	Stratified (REE_{Cosmed} patients)
17	All*, 4 BMI classes, 3 Body temp classes	176	Stratified (REE_{Cosmed} patients)
18	All*	130	Stratified (REE_{GE} patients)
19	All*, 4 BMI classes, 3 Body temp classes	130	Stratified (REE_{GE} patients)

* All: Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, SOFA, SAPS2, NUTRIC

AVD: Augmentation via duplication, GCS: GaussianCopulaSynthesizer, GBR: GradientBoostingRegressor.

176), 3) stratified REE_{GE} dataset (Dataset 19 from Table 4, N = 130), and 4) the augmented dataset (Dataset 11 from Table 4, N = 306). An Empirical Cumulative Distribution Functions (ECDF) plot was also used to visualise the statistical distributions of the four datasets.

3. Results

In an analysis of patients with both REE_{Cosmed} and REE_{GE} values (N = 127), the mean percentage difference in REE values is -1.63 ± 9.62 % with respect to the REE_{Cosmed} values. The mean ± standard deviation REE of the original patient cohort (N = 300, Table 3) is 1438 ± 425 calories. The mean ± standard deviation REE of the augmented patient cohort (N = 306) is 1444 ± 413 calories; whereas REE_{Cosmed} (N = 176) and REE_{GE} (N = 130) of the stratified patients are 1480 ± 387 calories and 1452 ± 386 calories, respectively. The statistical distributions of the four groups are visualised using an ECDF plot as shown in Fig. 3. A pairwise Mann-Whitney U-test revealed no statistically significant differences between the groups ($P > 0.05$). The RMSE values determined using PEs relative to REE_{Cosmed} or REE_{GE} values are presented in Table 5. The Harris and Benedict equation (H-B1919) has the lowest RMSE of 362 calories and will be used as the benchmark value for the developed ML and DL models.

All models (3 ML and 1 DL model) trained with datasets augmented via duplication (N = 306) have the best prediction performances. This RMSE values for each model trained with this dataset is summarised in Table 6. It is important to note that all the models in Table 6 have significantly better performance than the H-B1919 model, with a maximum RMSE improvement/reduction of 163 calories (Model 4 - XGBoost, Table 6).

Between the four models, the XGBoost model has the best performance, achieving the lowest average RMSE in a 5-fold cross validation, followed by the RFR model (Table 6). Table 7 presents the 5-fold cross validation results of the XGBoost model using differing input data (input variables, dataset size, augmentation or stratification techniques). Similar tables for the RFR, SVR and CNN models are attached in the Appendix (Tables A1, A2 and A3, respectively).

Scatter plots for the 1st fold of the cross validation are presented in Fig. 4 for the best RFR, XGBoost, SVR and CNN models as per Table 6. The feature importances plot across 5 cross-validation folds are presented for the RFR and XGBoost models in Fig. 5. From the plots, it is observed that variables such as BSA, ABW, Age, MV, SAPS2, IBW and SOFA are the most important factors affecting model performance.

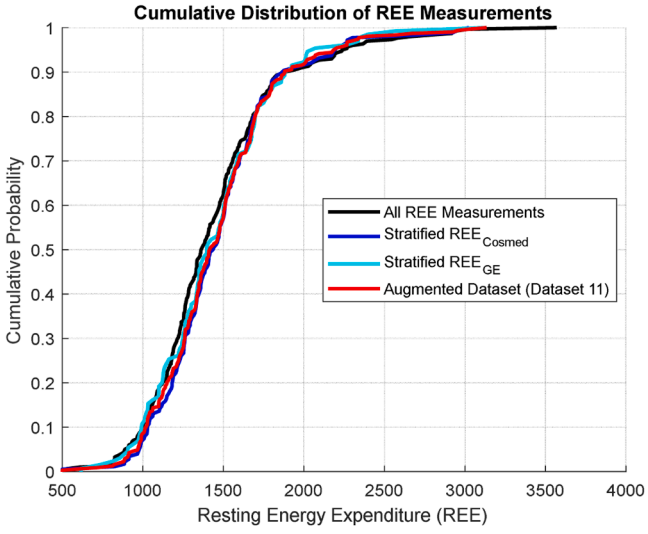

Fig. 3. The ECDF plot showing the statistical distribution of the: 1) original dataset (Black), 2) stratified REE_{Cosmed} dataset (Blue), 3) stratified REE_{GE} dataset (Cyan), and 4) the augmented dataset – Dataset 11 from Table 4 (Red).

Table 5
RMSE values determined using PEs relative to REE_{Cosmed} or REE_{GE} values for the original patient cohort ($N = 300$ IC measurements).

	H-B1919	H-B1984	Penn state	Modified penn state	Ireton-jones	TAH	MSJ	Swinamer	Weight base
RMSE (calories)	362	457	2202	1109	449	435	379	1515	577

Table 6
5-fold cross validation results for the best performing RFR, XGBoost, SVR and CNN models. All models were trained with datasets augmented via duplication. RMSE values (calories) are presented for each k-fold along with the mean ± standard deviation (stdev) RMSE value across all 5 folds.

Model	Input data	Dataset size	ML/DL model	K-fold					Mean ± Stdev
				1	2	3	4	5	
1	All*	306	RFR	188	189	251	212	170	202 ± 31
2	All*, 4 BMI classes, 3 Body temp classes	306	RFR	186	196	250	212	155	200 ± 35
3	All*	306	XGBoost	210	192	235	225	170	207 ± 26
4	All*, 4 BMI classes, 3 Body temp classes	306	XGBoost	183	182	226	234	169	199 ± 29
5	All*	306	SVR	205	239	294	241	171	230 ± 46
6	All*, 4 BMI classes, 3 Body temp classes	306	SVR	196	231	286	219	175	221 ± 42
7	All*	306	CNN	237	243	300	252	216	250 ± 31
8	All*, 4 BMI classes, 3 Body temp classes	306	CNN	248	228	270	249	205	240 ± 25

* All: Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, SOFA, SAPS2, NUTRIC.

Table 7
5-fold cross validation results for XGBoost model. The RMSE values (calories) are presented for each k-fold along with the mean ± standard deviation (stdev) RMSE value across all 5 folds.

Model	Input data	Dataset size	Remarks	K-fold					Mean ± Stdev
				1	2	3	4	5	
1	Age, Height, ABW, BMI, IBW, MV	300		299	341	335	383	477	367 ± 68
2	Age, Height, ABW, BMI, IBW, MV, Apache 11	300		308	327	351	376	472	367 ± 64
3	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender	300		321	327	328	402	496	375 ± 76
4	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA	300		307	320	331	348	470	355 ± 66
5	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp	259		314	367	398	288	348	343 ± 43
6	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA	242		394	338	307	342	353	347 ± 31
7	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, NUTRIC	242		420	322	315	334	360	350 ± 43
8	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, SAPS2	179		272	251	226	305	320	275 ± 38
9	All*	179		263	244	221	301	315	269 ± 39
10	All*	306	AVD	210	192	235	225	170	207 ± 26
11	All*, 4 BMI classes, 3 Body temp classes	306	AVD	183	182	226	234	169	199 ± 29
12	All*	612	AVD, GCS	238	211	261	269	233	242 ± 24
13	All*, 4 BMI classes, 3 Body temp classes	612	AVD, GCS	240	204	254	266	231	239 ± 24
14	All*	427	AVD, GBR	358	375	297	266	317	322 ± 44
15	All*, 4 BMI classes, 3 Body temp classes	427	AVD, GBR	345	371	274	262	329	316 ± 47
16	All*	176	Stratified (REE_{Cosmed} patients)	248	249	245	285	307	267 ± 27
17	All*, 4 BMI classes, 3 Body temp classes	176	Stratified (REE_{Cosmed} patients)	250	241	239	277	304	262 ± 28
18	All*	130	Stratified (REE_{GE} patients)	271	250	233	343	292	277 ± 43
19	All*, 4 BMI classes, 3 Body temp classes	130	Stratified (REE_{GE} patients)	271	234	232	331	255	264 ± 40

* All: Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, SOFA, SAPS2, NUTRIC.
AVD: Augmentation via duplication, GCS: GaussianCopulaSynthesizer, GBR: GradientBoostingRegressor.

4. Discussion

In an analysis of patients with both REE_{Cosmed} and REE_{GE} values ($N = 127$), the mean percentage difference in REE values is -1.63 ± 9.62 % with respect to REE_{Cosmed}, suggesting an inherent variability or systematic discrepancy between the two gold-standard methods of

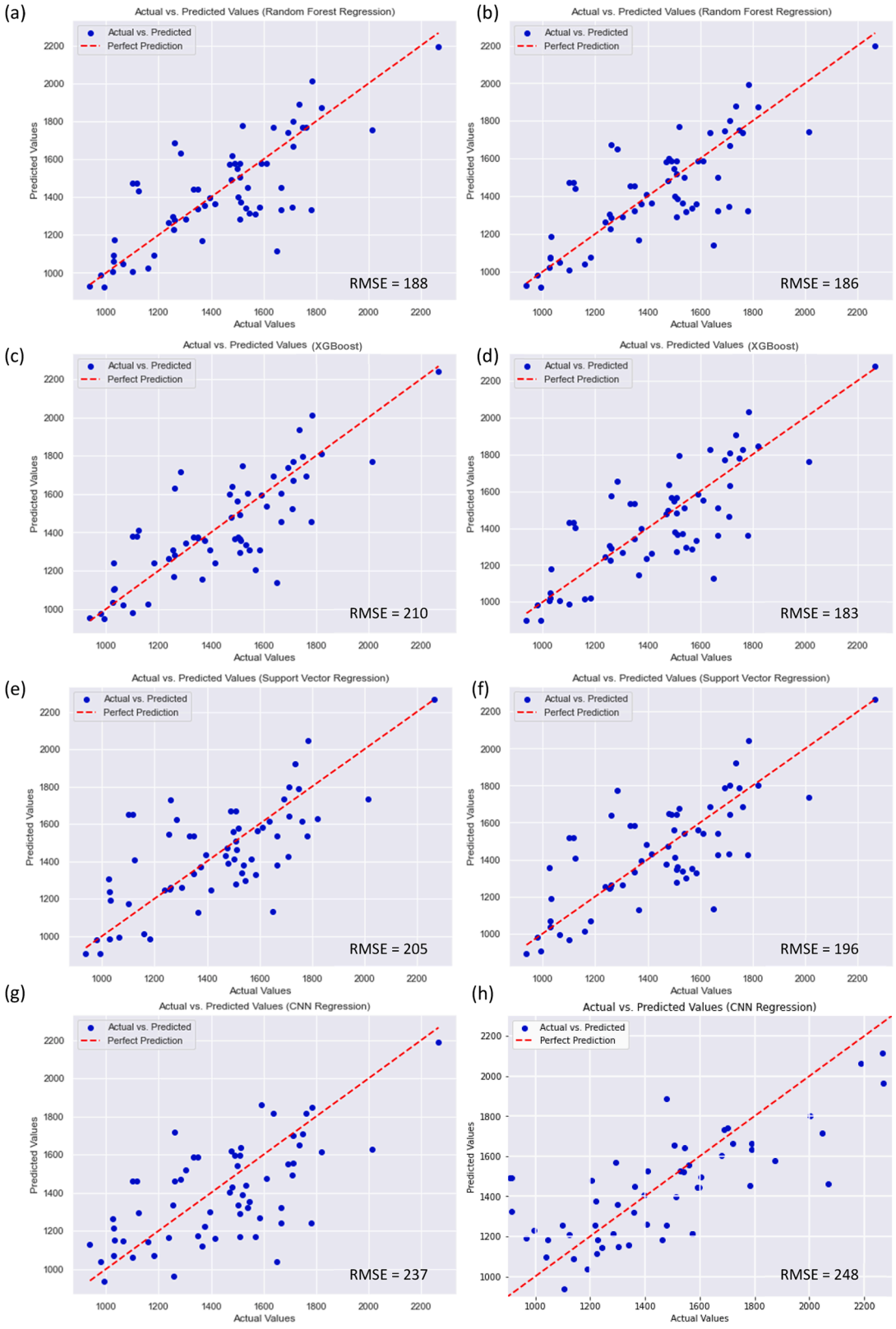

Fig. 4. Figures (a) to (g) represent scatter plots of REE values for the 1st fold of the cross validation for Models 1–8 in Table 6. The 5-fold cross validation of all ML/DL models were performed with a random state of 42, resulting in a mean ± standard deviation measured REE of 1438 ± 273 calories for all models.

measuring resting energy expenditure. The ECDF plot in Fig. 3 visualise the statistical distributions of the: 1) original dataset (N = 300), 2) stratified REE_{Cosmed} dataset (N = 176), 3) stratified REE_{GE} dataset (N = 130), and 4) the augmented dataset (Dataset 11 from Table 4, N = 306). The ECDF plot and a pairwise Mann-Whitney U-test revealed no statistically significant differences between the groups ($P > 0.05$). These findings indicate that the original, stratified, and augmented datasets are representative of the analysed patient population.

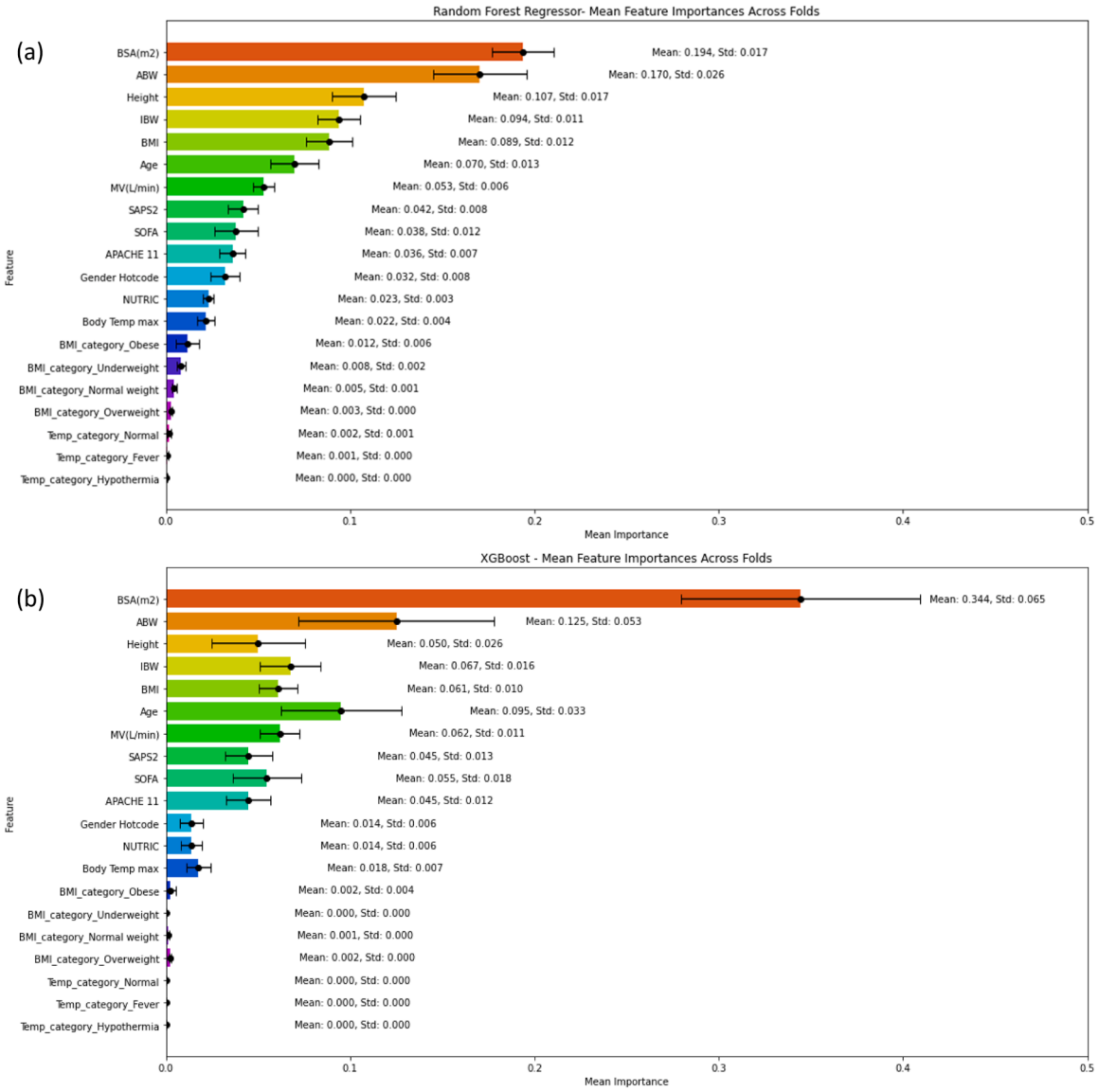

Fig. 5. Mean and standard deviation of the feature importances across 5 cross-validation folds for: (a) best RFR model (Table 6 – Model 2), and (b) the best XGBoost model (Table 6 – Model 4).

Of the 9 PEs analysed in this study, the Harris and Benedict equation (H-B1919) yielded the lowest RMSE at 362 calories (Table 5). This result stands despite other PEs demonstrating better predictive performance in their original studies and/or meta-analyses (Table 1). The discrepancy in performance observed in this study is likely due to differences in patient demographics, as the critically ill population analysed here may have characteristics that diverge from those used to develop the original equations. In contrast, the best ML/DL models (Table 6) significantly outperformed the H-B1919 model, with the XGBoost and RFR achieving the lowest RMSE of 199 and 200 calories respectively when trained using the "All*" data (Table 6). The CNN model had the poorest performance among of the 4 model architectures, with a RMSE of 250 calories, indicating that while CNN outperforms PEs, it may not be the most suitable architecture for regression-based problems. The addition of 4 BMI and 3 body temperature classes also results in marginally lower RSME values for all ML/DL models.

As observed in Tables 7, A1, A2 and A3, increasing the training dataset size through Augmentation via Duplication significantly improves ML/DL model performance. A notable reduction in mean RMSE over 5 cross-validation folds was observed (XGBoost: ~23 %, RFR: ~29 %, SVR: ~18 %, CNN: ~14 %) when comparing models trained on augmented versus non-augmented datasets. However, data augmentation (GaussianCopulaSynthesizer) and imputation techniques (GradientBoostingRegressor) do not enhance RMSE scores across all ML/DL models despite the increase in training dataset size, likely due to the limited size of the original patient dataset. Overall, models trained using mixed datasets, particularly dataset 11 (Table 4) demonstrate superior prediction performance compared to stratified datasets, likely due to the relatively larger sample size ($N = 306$ vs $N = 130 – 176$).

The feature importances across 5 cross-validation folds for the RFR and XGBoost models are presented in Fig. 5, effectively highlighting the contribution of each feature to the models' predictions. Interestingly, variables such as BMI and height are ranked high in RFR models, but low in the XGBoost models. This finding indicates that the ranking of feature importances not only between each k-fold of the same model (using identical hyperparameters) but also between the RFR and XGBoost models themselves. As such, the ranking of feature importances depends on both the training dataset composition and model architecture. The RFR model has a smaller standard deviation in comparison to the XGBoost model for most except for BMI, BMI_category_Normal weight, BMI_category_Obese, BMI_category_Underweight, Gender Hotcode, and Temp_category_Normal. These features were found to have lower

importance and lower variation in the order of feature importances between K-folds of the RFR model. Based on the importance plots, it is also observed that additional categorical variables related to BMI and temperature do not add much predictive power for the models, which could suggest their omission in future modelling attempts [38].

The feature importance plots also reveal that body composition variables are the most predictive variables in the models, followed by age, MV and critical illness scores such as APACHE II, SOFA and SAPS2 scores. This result is also expected as age, height and ABW are also the main input parameters for the PEs used in this study. Critical illness scores are not used in traditional PEs to assess REE. However, these scores reflect the clinical acuity and overall disease burden, which may be important determinants of energy expenditure [39,40]. In particular, these critical illness scores (APACHE 11 and SOFA) are shown to be important features for ML model decision making, consistently ranking above 10 out of 20 in mean feature importance scores for both the RFR and XGBoost models (Fig. 5). From the feature importance plots, BSA (m^2) demonstrates the highest feature importance scores among all the parameters, suggesting that it plays a pivotal role in predicting REE. While the Swinamer equation incorporates BSA as an input, the relatively high RMSE of 1515 calories (Table 5) may be attributed to its reliance on a simplified linear relationship between BSA and REE, as well as potential discrepancies arising from its development based on different patient demographics [29].

Ventilatory variables such as MV are important predictors of REE in the RFR and XGBoost models, as the increase in REE with carbon dioxide production (VCO_2) is a major factor behind elevated MV needs in long-term mechanically ventilated patients [41]. This observation aligns with findings from previous studies, which have indicated a strong correlation between respiratory parameters and metabolic rate [11,26,42,43]. While PEs such as the Penn State, modified Penn State and TAH equation incorporate MV as an input parameter, their RMSE values remain relatively high, at 2202, 1109 and 435 calories, respectively (Table 6). This is likely due to the fact that these equations are highly generalised and fail to account for the heterogeneity in patient demographics, disease states, treatments, and responses to ventilation.

In this study, BMI was included for training purpose along with patient's height and weight. While height and weight are raw inputs, BMI encapsulates their non-linear relationship (weight divided by height squared), enabling the model to leverage this interaction without learning it independently. Additionally, BMI as a derived feature may interact uniquely with other variables, potentially revealing patterns not captured by height and weight alone. Although the inclusion of BMI as an input parameter may raise concerns about multicollinearity within the feature set, its significance is evident from the feature importance plots in Fig. 5. In these plots, BMI is ranked the 5th most important feature with a mean ± standard deviation importance scores of 0.089 ± 0.012 for the RFR model and 0.061 ± 0.010 for the XGBoost model, respectively.

PEs and ML or DL models demonstrate comparable levels of clinical practicality, as both allow for the straightforward input of clinically measured variables. As such, the implementation of both approaches can be done using single-board computer-based devices [44] or tablets via on-device, edge- or cloud-computing [45,46]. Both approaches offer significant advantages over IC measurements in terms of cost-effectiveness, ease of use and accessibility, making them more practical for routine clinical application. In contrast, the acquisition of IC measurements is both resource intensive and time-consuming, making it impractical for routine use in most ICU practice [8]. While PEs are computationally inexpensive, the overall results indicate that ML/DL methods provide superior REE predictions compared to PEs by more accurately capturing the complex, and potentially non-linear, relationships between static and dynamic physiological variables and REE. In contrast, PEs may be overly generalised and less tailored to the specific characteristics and demographic variability of the critically ill patient population [25]. A study by Arabi *et al.* (PermiT trial) found that a 25 % variation in the calculated REE did not result in significant differences in clinical outcomes [47]. In the context of this study, 25 % of the mean REE of the augmented patient cohort equates to 361 calories, providing an acceptable margin of error for clinical application of the ML/DL models where the RMSE ranges from 199 – 250 (Table 6).

Despite the non-linear nature of both RFR and XGBoost models, both models align with clinical needs for explicability and interpretability through feature importance analysis and extracted rules, and can be considered as "white-box" models [32,48]. This interpretability, combined with their strong predictive performance, positions them as ideal candidates for widespread adoption in clinical practice, particularly in settings where transparency and model explainability are crucial. Additionally, the capacity of ML/DL models for non-linear, multi-variable processing, offers significant potential for more accurate and individualised predictions of REE, improving the precision of nutritional support. Specifically, the models developed in this study account for both static and dynamic physiological variables, a factor often overlooked by PEs. These models represent a promising alternative to PEs and IC methods, addressing the need for accurate and clinically viable tools for guiding nutrition therapy in critical care.

In terms of limitations, while model performance was validated using five-fold cross-validation, further validation is needed on independent, real-world patient datasets. Future work should aim to further evaluate the out-of-sample accuracy of these models in clinical trials to validate their clinical applicability. Additionally, model performance must be evaluated in diverse patient populations, including those from different geographical and ethnic backgrounds, to ensure that the models are generalisable and applicable across global settings. Specifically, further studies involving Asian cohorts are warranted, as metabolic responses and nutritional requirements may differ between ethnic groups [10]. The availability of larger datasets in the future could enable more precise patient-specific REE predictions, allowing ML/DL models to better account for both intra- and inter-patient variability in REE. This advancement could enhance model adaptability, improving personalised metabolic assessments and ultimately optimising clinical decision-making for individualised patient care.

Moreover, a small training dataset size represents a potential limitation, as it may fail to encompass the diversity of patient populations, potentially resulting in bias within ML/DL models and reducing the generalisability of their predictions. Additionally, the utility of data augmentation and imputation techniques, such as GaussianCopulaSynthesizer and GradientBoostingRegressor, may be constrained by the limited dataset size, increasing the risk of introducing artificial patterns that do not accurately reflect real-world variability and further amplifying model bias [49]. These challenges highlight the importance of continued development and rigorous validation of ML and DL models using larger and more diverse datasets to improve their reliability and clinical applicability [50]. Furthermore, REE varies significantly across different phases of critical illness, underscoring the need for accurate phase identification in critically ill patients [10,11]. The developed ML/DL models could be adapted to detect trends and transitions in patient anthropometrics and ventilatory parameters by incorporating continuous or sequential physiological data. This approach may enable real-time prediction of metabolic shifts, allowing clinicians to make timely adjustments to nutritional therapy based on the dynamic needs of critically ill patients.

In $N = 31$ patients, a maximum of 2 IC measurements were made using the same device (Cosmed: $N = 1$; GE: $N = 30$). For the Cosmed measurement ($N = 1$), there is no difference between the two IC measurements made, while the mean ± standard deviation absolute difference between the IC measurements for GE ($N = 30$) is 197 ± 177 calories. Due to the limited sample size of intra-patient IC measurements, it would not be possible to accurately determine the intra-patient device measurement uncertainty of the REE measurements. Previous studies have shown no significant difference in device-calculated REE [51], while a coefficient of variation (CV) of up to 10.9 % was found in

REE measurements of different gas analysis systems [52,53]. Therefore, future studies investigating the intra-patient and inter-device variations in REE measurements is warranted.

5. Conclusion

In this study, the performance of 3 different ML models including RFR, XGBoost, SVR and one DL model - CNN against traditional PEs for REE prediction in critically ill patients were investigated. The accuracy of both approaches was quantitatively assessed by comparing their predictions against gold standard IC measurements. The best performing model was found to be the XGBoost model, followed closely by the RFR model. In comparison, the Harris and Benedict equation (H-B1919) had the highest performance out of all 9 PEs. However, it exhibited approximately 82 % higher RMSE than the XGBoost model, indicating substantially lower predictive accuracy. While PEs are clinically practical, these findings indicate that ML/DL models can more accurately capture the complex, non-linear relationships between physiological variables and REE, making them better suited for the critically ill patient population. The results of this study show that the use of datasets including all input variables and augmented via AVD yielded the best ML/DL model performances. The integration of both static and dynamic variables, along with explainability tools like feature importance analysis, could bridge the gap between model complexity and clinical usability. Future clinical trials and studies involving diverse patient cohorts will be crucial for validating the clinical efficacy and generalisability of these models. Overall, this study establishes a foundation for integrating ML/DL methodologies into clinical practice, offering a pathway for more accurate and targeted nutritional support strategies in critical care settings.

Ethics Statement

A prospective observational study was conducted from October 2022 to October 2024 in the Intensive Care Unit (ICU) at Sultan Ahmad Shah Medical Centre @ IIUM (SASMEC), Kuantan, Pahang. The study protocol was approved by the IIUM Research Ethics Committee (IREC) (ID No: IREC 2020-042). Virtual data of 269 patients were also retrieved from patient records and previous studies, which had been previously approved by IREC (ID No: IREC 2020-042, IREC 2020-168). Informed consent was obtained from the patients or their legal representatives.

CRediT authorship contribution statement

Christopher Yew Shuen Ang: Writing – original draft, Methodology, Formal analysis. **Mohd Basri Mat Nor:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization. **Nur Sazwi Nordin:** Writing – review & editing, Validation, Conceptualization. **Thant Zin Kyi:** Writing – review & editing. **Ailin Razali:** Writing – review & editing. **Yeong Shiong Chiew:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

Funding Support: Ministry of Higher Education Malaysia, Prototype Development Research Grants Scheme (PRGS) [Ref Code: PRGS/1/2022/SKK07/UIAM/01/1].

Acknowledgements

Funding Support: Ministry of Higher Education Malaysia, Prototype Development Research Grants Scheme (PRGS) [Ref Code: PRGS/1/2022/SKK07/UIAM/01/1]. The authors wish to thank all the personnel from the Intensive Care Unit, Sultan Ahmad Shah Medical Centre @ IIUM (SASMEC), Department of Anaesthesiology, International Islamic University Malaysia, and Aiology Sdn. Bhd. for their contributions and support of this research.

Appendix

Table A1
5-fold cross validation results for RFR model. The RMSE values (calories) are presented for each k-fold along with the mean ± standard deviation (stdev) RMSE value across all 5 folds.

Model	Input data	Dataset size	Remarks	K-fold					Mean ± Stdev
				1	2	3	4	5	
1	Age, Height, ABW, BMI, IBW, MV	300		313	351	345	373	449	366 ± 51
2	Age, Height, ABW, BMI, IBW, MV, Apache 11	300		316	344	359	382	443	369 ± 48
3	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender	300		309	343	346	369	435	360 ± 47
4	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA	300		315	332	336	368	448	360 ± 53
5	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp	259		342	380	408	300	356	357 ± 41
6	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA	242		395	322	298	392	349	351 ± 42
7	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, NUTRIC	242		397	329	300	383	351	352 ± 39
8	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, SAPS2	179		322	241	234	322	314	286 ± 45
9	All*	179		324	228	235	321	321	286 ± 50
10	All*	306	AVD	188	189	251	212	170	202 ± 31

(continued on next page)

Table A1 (*continued*)

Model	Input data	Dataset size	Remarks	K-fold					Mean ± Stdev
				1	2	3	4	5	
11	All*, 4 BMI classes, 3 Body temp classes	306	AVD	186	196	250	212	155	200 ± 35
12	All*	612	AVD, GCS	246	216	276	263	231	246 ± 24
13	All*, 4 BMI classes, 3 Body temp classes	612	AVD, GCS	249	216	273	265	237	248 ± 23
14	All*	427	AVD, GBR	325	382	323	233	305	314 ± 53
15	All*, 4 BMI classes, 3 Body temp classes	427	AVD, GBR	312	370	328	228	301	308 ± 52
16	All*	176	Stratified (REE_{Cosmed} patients)	256	276	244	293	335	281 ± 36
17	All*, 4 BMI classes, 3 Body temp classes	176	Stratified (REE_{Cosmed} patients)	264	270	229	303	323	278 ± 36
18	All*	130	Stratified (REE_{GE} patients)	232	247	263	328	277	269 ± 37
19	All*, 4 BMI classes, 3 Body temp classes	130	Stratified (REE_{GE} patients)	245	247	259	345	266	272 ± 42

* All: Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, SOFA, SAPS2, NUTRIC.
AVD: Augmentation via duplication, GCS: GaussianCopulaSynthesizer, GBR: GradientBoostingRegressor.

Table A2
5-fold cross validation results for SVR model. The RMSE values (calories) are presented for each k-fold along with the mean ± standard deviation (stdev) RMSE value across all 5 folds.

Model	Input data	Dataset size	Remarks	K-fold					Mean ± Stdev
				1	2	3	4	5	
1	Age, Height, ABW, BMI, IBW, MV	300		316	333	349	392	424	363 ± 45
2	Age, Height, ABW, BMI, IBW, MV, Apache 11	300		320	335	354	381	418	361 ± 39
3	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender	300		298	348	354	383	389	354 ± 36
4	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA	300		305	337	336	380	414	354 ± 42
5	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp	259		352	377	395	282	341	349 ± 43
6	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA	242		404	333	285	344	340	341 ± 42
7	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, NUTRIC	242		412	340	285	345	344	345 ± 45
8	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, SAPS2	179		315	239	218	302	315	278 ± 46
9	All*	179		316	239	225	304	320	281 ± 45
10	All*	306	AVD	205	239	294	241	171	230 ± 46
11	All*, 4 BMI classes, 3 Body temp classes	306	AVD	196	231	286	219	175	221 ± 42
12	All*	612	AVD, GCS	257	220	284	269	233	253 ± 26
13	All*, 4 BMI classes, 3 Body temp classes	612	AVD, GCS	259	225	285	271	237	255 ± 24
14	All*	427	AVD, GBR	368	390	280	236	326	320 ± 63
15	All*, 4 BMI classes, 3 Body temp classes	427	AVD, GBR	349	368	266	226	368	315 ± 66
16	All*	176	Stratified (REE_{Cosmed} patients)	273	245	241	268	301	266 ± 24
17	All*, 4 BMI classes, 3 Body temp classes	176	Stratified (REE_{Cosmed} patients)	279	252	249	273	352	281 ± 42
18	All*	130	Stratified (REE_{GE} patients)	218	212	204	311	261	241 ± 45
19	All*, 4 BMI classes, 3 Body temp classes	130	Stratified (REE_{GE} patients)	247	211	275	360	228	264 ± 59

* All: Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, SOFA, SAPS2, NUTRIC.
AVD: Augmentation via duplication, GCS: GaussianCopulaSynthesizer, GBR: GradientBoostingRegressor.

Table A3
5-fold cross validation results for CNN model. The RMSE values (calories) are presented for each k-fold along with the mean ± standard deviation (stdev) RMSE value across all 5 folds.

Model	Input data	Dataset size	Remarks	K-fold					Mean ± Stdev
				1	2	3	4	5	
1	Age, Height, ABW, BMI, IBW, MV	300		326	347	369	369	433	369 ± 40
2	Age, Height, ABW, BMI, IBW, MV, Apache 11	300		364	361	355	363	419	372 ± 26
3	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender	300		307	370	338	352	406	355 ± 37
4	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA	300		318	347	330	337	393	345 ± 29
5	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp	259		344	394	399	311	355	361 ± 36
6	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA	242		431	389	318	375	331	369 ± 46
7	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, NUTRIC	242		429	386	304	380	339	368 ± 48
8	Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, SAPS2	179		288	290	241	300	310	286 ± 26
9	All*	179		286	290	251	307	315	290 ± 25
10	All*	306	AVD	237	243	300	252	216	250 ± 31
11	All*, 4 BMI classes, 3 Body temp classes	306	AVD	248	228	270	249	205	240 ± 25
12	All*	612	AVD, GCS	266	233	284	278	233	259 ± 24
13	All*, 4 BMI classes, 3 Body temp classes	612	AVD, GCS	267	228	312	282	249	268 ± 32
14	All*	427	AVD, GBR	364	399	333	303	340	348 ± 36
15	All*, 4 BMI classes, 3 Body temp classes	427	AVD, GBR	350	378	272	243	397	328 ± 67
16	All*	176	Stratified (REE_{Cosmed} patients)	259	288	244	314	297	280 ± 28
17	All*, 4 BMI classes, 3 Body temp classes	176	Stratified (REE_{Cosmed} patients)	286	308	249	303	286	286 ± 23
18	All*	130	Stratified (REE_{GE} patients)	232	307	307	339	290	295 ± 39
19	All*, 4 BMI classes, 3 Body temp classes	130	Stratified (REE_{GE} patients)	263	309	245	328	287	286 ± 34

* All: Age, Height, ABW, BMI, IBW, MV, Apache 11, Gender, BSA, Body temp, SOFA, SOFA, SAPS2, NUTRIC.
AVD: Augmentation via duplication, GCS: GaussianCopulaSynthesizer, GBR: GradientBoostingRegressor.

References

[1] P. Singer, Preserving the quality of life: nutrition in the ICU (in eng), Crit. Care 23 (Suppl 1) (Jun 14 2019) 139, https://doi.org/10.1186/s13054-019-2415-8.
[2] T. Cederholm, et al., GLIM criteria for the diagnosis of malnutrition - A consensus report from the global clinical nutrition community (in eng), Clin. Nutr. 38 (1) (Feb 2019) 1–9, https://doi.org/10.1016/j.clnu.2018.08.002.
[3] K.J. Lambell, O.A. Tatucu-Babet, L.A. Chapple, D. Gantner, E.J. Ridley, Nutrition therapy in critical illness: a review of the literature for clinicians (in eng), Crit. Care 24 (1) (Feb 4 2020) 35, https://doi.org/10.1186/s13054-020-2739-4.
[4] A.B. Guttormsen, C. Pichard, Determining energy requirements in the ICU (in eng), Curr. Opin. Clin. Nutr. Metab. Care 17 (2) (Mar 2014) 171–176, https://doi.org/10.1097/mco.0000000000000028.
[5] K.M. Schlein, S.P. Coulter, Best practices for determining resting energy expenditure in critically ill adults (in eng), Nutr. Clin. Pract. 29 (1) (Feb 2014) 44–55, https://doi.org/10.1177/0884533613515002.
[6] P. Singer, et al., ESPEN guideline on clinical nutrition in the intensive care unit (in eng), Clin. Nutr. 38 (1) (Feb 2019) 48–79, https://doi.org/10.1016/j.clnu.2018.08.037.
[7] S.A. McClave, et al., Guidelines for the provision and assessment of Nutrition support therapy in the adult critically ill patient: Society of Critical Care Medicine (SCCM) and American Society for Parenteral and Enteral Nutrition (A.S.P.E.N.) (in eng), JPEN. J. Parenter. Enteral. Nutr. 40 (2) (Feb 2016) 159–211, https://doi.org/10.1177/0148607115621863.
[8] H. Moonen, K.J.H. Beckers, A.R.H. van Zanten, Energy expenditure and indirect calorimetry in critical illness and convalescence: current evidence and practical considerations (in eng), J. Intensive Care 9 (1) (Jan 12 2021) 8, https://doi.org/10.1186/s40560-021-00524-0.
[9] M. Cordoza, L.N. Chan, E. Bridges, H. Thompson, Methods for estimating energy expenditure in critically ill adults (in eng), AACN. Adv. Crit. Care 31 (3) (Sep 15 2020) 254–264, https://doi.org/10.4037/aacnacc2020110.
[10] P.C. Tah, et al., A single-center prospective observational study comparing resting energy expenditure in different phases of critical illness: indirect calorimetry versus predictive equations, Crit. Care Med. 48 (5) (2020) [Online]. Available: https://journals.lww.com/ccmjournal/fulltext/2020/05000/a_single_center_prospective_observational_study.32.aspx.
[11] P.C. Tah, et al., Do we need different predictive equations for the acute and late phases of critical illness? A prospective observational study with repeated indirect calorimetry measurements (in eng), Eur. J. Clin. Nutr. 76 (4) (Apr 2022) 527–534, https://doi.org/10.1038/s41430-021-00999-y.
[12] T.d.O. Fernandes, C.M. Avesani, M.A. Kamimura, D.T. Aoike, L. Cuppari, Estimating resting energy expenditure of patients on dialysis: development and validation of a predictive equation, Nutrition 67-68 (2019) 110527, https://doi.org/10.1016/j.nut.2019.06.008, 2019/11/01/.
[13] D. Ponce, C.R. de Goes, L.G.M. de Andrade, Proposal of a new equation for estimating resting energy expenditure of acute kidney injury patients on dialysis: a machine learning approach, Nutr. Metab. 17 (1) (2020) 96, https://doi.org/10.1186/s12986-020-00519-y, 2020/11/17.
[14] V. De Cosmi, et al., Prediction of resting energy expenditure in children: may artificial neural networks improve our accuracy? J. Clin. Med. 9 (4) (2020) 1026 [Online]. Available: https://www.mdpi.com/2077-0383/9/4/1026.
[15] E. Disse, et al., An artificial neural network to predict resting energy expenditure in obesity, Clin. Nutr. 37 (5) (2018) 1661–1669, https://doi.org/10.1016/j.clnu.2017.07.017, 2018/10/01/.
[16] A. Bailey, M. Eltawil, S. Gohel, L. Byham-Gray, Machine learning models using non-linear techniques improve the prediction of resting energy expenditure in individuals receiving hemodialysis, Ann. Med. 55 (2) (2023) 2238182, https://doi.org/10.1080/07853890.2023.2238182, 2023/12/12.
[17] S.C. Luy, O.A. Dampil, Comparison of the Harris-Benedict equation, bioelectrical impedance analysis, and indirect calorimetry for measurement of basal metabolic rate among adult obese Filipino patients with prediabetes or type 2 Diabetes mellitus (in eng), J. ASEAN. Fed. Endocr. Soc. 33 (2) (2018) 152–159, https://doi.org/10.15605/jafes.033.02.07.

[18] J.A. Doley, M. Sandberg, Prolonged mechanical ventilation and nutritional support regimens. Diet and Nutrition in Critical Care, Springer, New York, NY, USA, 2014, pp. 1–15. New York, NY.
[19] R.N. Walker, R.A. Heuberger, Predictive equations for energy needs for the critically ill, Respir. Care 54 (4) (2009) 509 [Online]. Available: http://rc.rcjournal.com/content/54/4/509.abstract.
[20] A.M. Roza, H.M. Shizgal, The Harris Benedict equation reevaluated: resting energy requirements and the body cell mass (in eng), Am. J. Clin. Nutr. 40 (1) (Jul 1984) 168–182, https://doi.org/10.1093/ajcn/40.1.168.
[21] P. Singer, J. Singer, Clinical guide for the use of metabolic carts: indirect calorimetry–No longer the orphan of energy estimation (in eng), Nutr. Clin. Pract. 31 (1) (Feb 2016) 30–38, https://doi.org/10.1177/0884533615622536.
[22] D. Frankenfield, J.S. Smith, R.N. Cooney, Validation of 2 approaches to predicting resting metabolic rate in critically ill patients (in eng), JPEN. J. Parenter. Enteral. Nutr. 28 (4) (Jul-Aug 2004) 259–264, https://doi.org/10.1177/0148607104028004259.
[23] D. Frankenfield, M. Hise, A. Malone, M. Russell, E. Gradwell, C. Compher, Prediction of resting metabolic rate in critically ill adult patients: results of a systematic review of the evidence, J. Am. Diet Assoc. 107 (9) (2007) 1552–1561, https://doi.org/10.1016/j.jada.2007.06.010.
[24] D.C. Frankenfield, J.S. Smith Jr., R.N. Cooney, S.A. Blosser, G.Y. Sarson, Relative association of fever and injury with hypermetabolism in critically ill patients (in eng), Injury 28 (9-10) (Nov-Dec 1997) 617–621, https://doi.org/10.1016/s0020-1383(97)00117-4.
[25] O. Zusman, I. Kagan, I. Bendavid, M. Theilla, J. Cohen, P. Singer, Predictive equations versus measured energy expenditure by indirect calorimetry: A retrospective validation, Clin. Nutr. 38 (3) (2019) 1206–1210, https://doi.org/10.1016/j.clnu.2018.04.020, 2019/06/01/.
[26] D.C. Frankenfield, A. Coleman, S. Alam, R.N. Cooney, Analysis of estimation methods for resting metabolic rate in critically ill adults (in eng), JPEN. J. Parenter. Enteral. Nutr. 33 (1) (Jan-Feb 2009) 27–36, https://doi.org/10.1177/0148607108322399.
[27] C. Ireton-Jones, J.D. Jones, Improved equations for predicting energy expenditure in patients: the Ireton-Jones equations (in eng), Nutr. Clin. Pract. 17 (1) (Feb 2002) 29–31, https://doi.org/10.1177/011542650201700129.
[28] M.D. Mifflin, S.T. St Jeor, L.A. Hill, B.J. Scott, S.A. Daugherty, Y.O. Koh, A new predictive equation for resting energy expenditure in healthy individuals (in eng), Am. J. Clin. Nutr. 51 (2) (Feb 1990) 241–247, https://doi.org/10.1093/ajcn/51.2.241.
[29] D.L. Swinamer, M.G. Grace, S.M. Hamilton, R.L. Jones, P. Roberts, E.G. King, Predictive equation for assessing energy expenditure in mechanically ventilated critically ill patients (in eng), Crit. Care Med. 18 (6) (Jun 1990) 657–661, https://doi.org/10.1097/00003246-199006000-00017.
[30] K.G. Kreymann, et al., ESPEN Guidelines on Enteral Nutrition: intensive care (in eng), Clin. Nutr. 25 (2) (Apr 2006) 210–223, https://doi.org/10.1016/j.clnu.2006.01.021.
[31] R. Guo, Z. Zhao, T. Wang, G. Liu, J. Zhao, D. Gao, Degradation state recognition of piston pump based on ICEEMDAN and XGBoost, Appl. Sci. 10 (2020) 6593, https://doi.org/10.3390/app10186593, 09/21.
[32] A.A. Khan, O. Chaudhari, R. Chandra, A review of ensemble learning and data augmentation models for class imbalanced problems: combination, implementation and evaluation, Expert Syst. Appl. 244 (2024) 122778, https://doi.org/10.1016/j.eswa.2023.122778, 2024/06/15/.
[33] V. Rodriguez-Galiano, M. Sánchez Castillo, J. Dash, P. Atkinson, J. Ojeda-Zujar, Modelling interannual variation in the spring and autumn land surface phenology of the European forest, Biogeosciences. 13 (2016) 3305–3317, https://doi.org/10.5194/bg-13-3305-2016, 06/06.
[34] L. Rosenbaum, A. Dörr, M. Bauer, F. Boeckler, A. Zell, Inferring multi-target QSAR models with taxonomy-based multi-task learning, J. Cheminform. 5 (2013) 33, https://doi.org/10.1186/1758-2946-5-33, 07/11.
[35] S.-W. Hwang, et al., Feature importance measures from random forest regressor using near-infrared spectra for predicting carbonization characteristics of kraft lignin-derived hydrochar, J. Wood Sci. 69 (1) (2023) 1, https://doi.org/10.1186/s10086-022-02073-y, 2023/01/05.
[36] G. Louppe, L. Wehenkel, A. Sutera, P. Geurts, Understanding variable importances in forests of randomized trees (in English), Adv. Neural Inf. Process Syst. 26 (2013) [Online]. Available: https://papers.nips.cc/paper_files/paper/2013/hash/e3796ae838835da0b6f6ea37bcf8bcb7-Abstract.html.
[37] M. Oppelt, et al., ADABase: A multimodal dataset for cognitive load estimation, Sensors 1 (12/28) (2022), https://doi.org/10.3390/s23010340.
[38] Z.K. Pourtaheri, S.H. Zahiri, Ensemble classifiers with improved overfitting, in: 2016 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC), 2016, pp. 93–97, https://doi.org/10.1109/CSIEC.2016.7482130, 9-11 March 2016.
[39] S. Shahi, H. Paneru, R. Ojha, R. Karn, R. Rajbhandari, B.P. Gajurel, SOFA and APACHE II scoring systems for predicting outcome of neurological patients admitted in a tertiary hospital intensive care unit (in eng), Ann. Med. Surg. (Lond) 86 (4) (Apr 2024) 1895–1900, https://doi.org/10.1097/ms9.0000000000001734.
[40] H. Mtaweh, et al., Systematic review of factors associated with energy expenditure in the critically ill, Clin. Nutr. ESPEN 33 (2019) 111–124, https://doi.org/10.1016/j.clnesp.2019.06.009, 2019/10/01/.
[41] R. Kiiski, J. Takala, Hypermetabolism and efficiency of CO2 removal in acute respiratory failure (in eng), Chest 105 (4) (Apr 1994) 1198–1203, https://doi.org/10.1378/chest.105.4.1198.
[42] C. Faisy, E. Guerot, J.L. Diehl, J. Labrousse, J.Y. Fagon, Assessment of resting energy expenditure in mechanically ventilated patients (in eng), Am. J. Clin. Nutr. 78 (2) (Aug 2003) 241–249, https://doi.org/10.1093/ajcn/78.2.241.
[43] L.S. Brandi, L. Santini, R. Bertolini, P. Malacarne, S. Casagli, A.M. Baraglia, Energy expenditure and severity of injury and illness indices in multiple trauma patients (in eng), Crit. Care Med. 27 (12) (Dec 1999) 2684–2689, https://doi.org/10.1097/00003246-199912000-00013.
[44] Q.A. Ng, et al., CAREDAQ: data acquisition device for mechanical ventilation waveform monitoring (in eng), HardwareX. 12 (Oct 2022) e00358, https://doi.org/10.1016/j.ohx.2022.e00358.
[45] B. Varghese, N. Wang, S. Barbhuiya, P. Kilpatrick, D.S. Nikolopoulos, Challenges and opportunities in edge computing, in: 2016 IEEE International Conference on Smart Cloud (SmartCloud), 2016, pp. 20–26, https://doi.org/10.1109/SmartCloud.2016.18, 18-20 Nov. 2016.
[46] M.M. John, H.H. Olsson, J. Bosch, AI on the edge: architectural alternatives, in: 2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 2020, pp. 21–28, https://doi.org/10.1109/SEAA51224.2020.00015, 26-28 Aug. 2020.
[47] Y.M. Arabi, et al., Permissive underfeeding or standard enteral feeding in critically ill adults (in eng), N. Engl. J. Med. 372 (25) (Jun 18 2015) 2398–2408, https://doi.org/10.1056/NEJMoa1502826.
[48] O. Loyola-González, Black-Box vs. White-Box: understanding their advantages and weaknesses from a practical point of view, IEEe Access. 7 (2019) 154096–154113, https://doi.org/10.1109/ACCESS.2019.2949286.
[49] S. Jäger, A. Allhorn, F. Bießmann, A benchmark for data imputation methods (in English), Front. Big Data 4 (2021) 693674, https://doi.org/10.3389/fdata.2021.693674.
[50] A. Arora, et al., The value of standards for health datasets in artificial intelligence-based applications (in eng), Nat. Med. 29 (11) (Nov 2023) 2929–2938, https://doi.org/10.1038/s41591-023-02608-w.
[51] W.A. Welch, S.J. Strath, A.M. Swartz, Congruent validity and reliability of two metabolic systems to measure resting metabolic rate (in eng), Int. J. Sports Med. 36 (5) (May 2015) 414–418, https://doi.org/10.1055/s-0034-1398575.
[52] J.A. Cooper, et al., Assessing validity and reliability of resting metabolic rate in six gas analysis systems (in eng), J. Am. Diet. Assoc. 109 (1) (Jan 2009) 128–132, https://doi.org/10.1016/j.jada.2008.10.004.
[53] S. Kaviani, et al., Determining the accuracy and reliability of indirect calorimeters utilizing the methanol combustion technique (in eng), Nutr. Clin. Pract. 33 (2) (Apr 2018) 206–216, https://doi.org/10.1002/ncp.10070.