

RESEARCH

Open Access



Diagnosis of chronic fatigue syndrome using beat-to-beat autonomic measurements

Sławomir Kujawski^{1*} , Hanna Tabisz¹, Karl J. Morten², Aleksandra Modlińska¹, Joanna Słomko¹ and Paweł Zalewski^{1,3}

Abstract

Background An artificial intelligence (AI) pipeline was used to differentiate patients suffering from Chronic Fatigue Syndrome (CFS) from healthy controls (HC) based on high-frequency, large-scale data obtained using beat-to-beat measurement of the autonomic nervous system (ANS) and cardiovascular function.

Methods This prospective, case-control study included a cohort of 112 CFS patients and 61 HCs examined. Heart rate (HR), high-frequency R-to-R interval (HF RRI), diastolic blood pressure (dBp), stroke volume (SV), and SV index (SV/FFM) were measured using the Task Force Monitor. A novel sequential learning approach was applied: first, a Transformer model was trained, followed by an XGBoost classifier that learned from the errors of the Transformer. Matthews correlation coefficient (MCC), accuracy, and Area Under the Receiver Operating Characteristic Curve (ROC AUC) were assessed. Model classifications were explained globally.

Results The applied classifier achieved a subject-level accuracy of 0.89, an MCC of 0.79, and an AUC of 1.00. Lower values of beat-to-beat difference in HR and raw HF RRI (indicating reduced cardiac vagal tone) and higher values of dBp difference (more beat-to-beat increases, indicating higher sympathetic vascular tone) were related to being more likely classified as CFS patients. Low values of SV difference and low values of SV/FFM (both indicating less effective cardiac hemodynamics) were related to being more likely classified as CFS patients.

Conclusions The AI-driven classifier demonstrates remarkable proficiency in distinguishing between patients with CFS and HC. By leveraging this automated pipeline, beat-to-beat measurements of the ANS can significantly enhance the objective assessment of CFS diagnosis.

Keywords ME/CFS, Explainable artificial intelligence, Autonomic nervous system

Background

Chronic Fatigue Syndrome (CFS) is defined by unexplained, persistent fatigue and a range of symptoms, none of which are pathognomonic. These include objectively measurable dysfunctions such as cognitive and autonomic nervous system (ANS) disturbances, and post-exertional malaise (PEM), which, despite its specificity, along with other symptoms, relies mainly on subjective patient reports due to the lack of reliable biomarkers [1–5]. This absence of quantifiable clinical or biological markers makes CFS diagnosis challenging, often resulting in long delays for patients [6].

*Correspondence:

Sławomir Kujawski
skujawski@cm.umk.pl

¹Department of Exercise Physiology and Functional Anatomy, Ludwik Rydygier Collegium Medicum in Bydgoszcz Nicolaus Copernicus University in Toruń, Świętojańska 20, Bydgoszcz 85-077, Poland

²The Nuffield Department of Women's and Reproductive Health, The Women Centre, The University of Oxford, The John Radcliffe Hospital, Headley Way, Headington, Oxford, UK

³Laboratory of Centre for Preclinical Research, Department of Experimental and Clinical Physiology, Warsaw Medical University, 1b Banacha Street, Warsaw 02-097, Poland



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

The ANS function, which regulates cardiac muscle, can be assessed non-invasively, and impairment has been documented in CFS patients [7]. In a previous study, we identified autonomic subgroups among CFS patients: those with sympathetic dysautonomia exhibited more severe disease, lower quality of life, and greater autonomic symptoms, while those with balanced ANS function had better outcomes [8].

In this study, beat-to-beat cardiovascular and autonomic parameters were recorded at rest, producing high-frequency time-series data from 112 CFS patients and 61 healthy controls out of an initial 1638 screened subjects [9]. Handling such large datasets requires advanced analytics. Previously, we used a machine learning ensemble for blood-based Raman spectroscopy diagnosis [10]. Here, we designed an automatic AI pipeline combining a Transformer neural network and XGBoost classifiers to distinguish parameter dynamics between CFS and controls. XGBoost, an effective method for tabular data, works by sequentially training decision trees to correct predecessor errors [11, 12]. Our novel sequential learning applies the Transformer first, then trains XGBoost on its prediction errors to improve performance.

To enhance model transparency and address “black box” concerns, we employed explainable artificial intelligence (XAI) [13]. The primary goal is to develop a fast, cost-effective, and accurate AI-based diagnostic tool for clinicians to identify CFS patients efficiently.

Materials and methods

Participants and study design

This prospective, case-control study included patients with CFS examined at Collegium Medicum Bydgoszcz, Poland (between February 2013 and January 2020). Therefore, no CFS cases related to COVID-19 were included in the cohort. Most patient data were obtained from baseline measurements in our previous intervention-based studies on CFS: an Individualised Activity Programme (IAP), and Whole-Body Cryotherapy and Static Stretching (WBC + SS) [14, 15]. The research was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee, Ludwik Rydygier Memorial Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University, Toruń (KB 332/2013 and KB 660/2017). All study participants gave written, informed consent for their data to be used for research purposes. CFS patients included in the studies met the Fukuda diagnostic criteria for CFS, which require unexplained, persistent, or relapsing fatigue lasting six or more months, accompanied by at least four out of eight additional symptoms such as post-exertional malaise, impaired memory or concentration, unrefreshing sleep, muscle/joint pain, tender lymph nodes, sore throat, and new headaches. The inclusion criteria for CFS in this

study were as follows: (1) age between 25 and 65 years; (2) fatigue for more than 6 months of unknown origin; and (3) at least four additional symptoms, including malaise after exertion, impaired memory and/or concentration, headache, unrefreshing sleep, tender lymph nodes (cervical or axillary), sore throat, and muscle or joint pain. The exclusion criteria, which included the existence of a medical condition that could cause chronic fatigue (such as autoimmune disease, psychosocial stress, or cardiovascular disease), were based on Fukuda’s criteria for diagnosing CFS [16]. The pre-test health condition assessment comprised a clinical examination as well as a basic psychological and neurological evaluation. Doctors experienced in diagnosing CFS validated the inclusion and exclusion criteria, ensured that thorough physical examinations were performed, and ruled out chronic comorbidities that could explain the primary symptoms. Figure S1 shows the study flowchart. Patients for the IAP study were recruited via advertisements on local and national media, with initial screening of over 1400 volunteers leading to 69 individuals meeting inclusion criteria following exclusion of neurological, neurodegenerative, psychiatric, and immunologic disorders possibly explaining primary symptoms. A total of 53 patients ultimately participated in baseline measures and the intervention.

The WBC+SS study recruited 250 individuals self-identifying as fatigued from January to July 2018. Patients fulfilling Fukuda criteria were eligible if aged 25 to 65 years with fatigue lasting more than six months, with fatigue severity quantified by Fatigue Severity Scale scores and presence of associated symptoms. Inclusion involved thorough pre-test assessments, including basic neurological, psychiatric, and physical examinations by specialists experienced in ME/CFS, who confirmed diagnostic criteria and excluded other organic causes. Anxiety and depression screening was performed using the Hospital Anxiety and Depression Scale (HADS) and Beck Depression Inventory (BDI-II) to exclude major psychiatric illness at baseline. The healthy control group for the current analysis was recruited separately with detailed exclusion of comorbidities, lifestyle factors, and medication use to ensure comparability.

Data acquisition

Body composition analysis

A multi-frequency bioelectrical impedance analyzer (Tanita MC-180MA Body Composition Analyzer, Tanita UK Ltd.) was used to assess body composition, applying a ‘normal’ body type algorithm. To reduce interference from the skin’s lipid layer, hands and feet were cleaned before measurement. Participants stood on the scale with feet electrodes and held hand grips with arms extended. Measured parameters included height, weight, fat-free mass (FFM), and fat-free mass index (FFMI, calculated

as FFM divided by height squared) [17]. Missing demographic data were imputed using subgroup medians (CFS or healthy control, male or female).

Autonomic and cardiovascular measurements

Task Force Monitor (TFM, CNSystems, Medizintechnik, Graz, Austria) is a specialized high-tech device used for cardiovascular and autonomic measures. The primary application of TFM is in the automated and calculated analysis of heart rate (ECG), oscillometric blood pressure readings (oscBP, contBP), and oscillometric heart rate monitoring [18]. At rest, all parameters were automatically measured five minutes after the signals stabilized. TFM parameters were described previously in detail [19]. Each patient was measured in a supine position at complete rest. The median number of heartbeats recorded during measurement was 166. The following parameters were assessed: Heart rate [HR], systolic blood pressure [sBP], diastolic blood pressure [dBP], mean blood pressure [mBP], stroke volume [SV], stroke index [SI], breathing frequency [BF], autonomic parameters: low-frequency (LF), high-frequency (HF) of HR variability [HRV], as well as signed differences in SV, sBP, dBP based on the change of these values between subsequent heartbeats. These variables are denoted as “differences”. The stroke volume index (SV/FFM) was calculated as SV divided by FFMI.

Data preprocessing and feature engineering

All data are presented as mean \pm SD or SE. Parametric test assumptions were verified using the Shapiro-Wilk test for normality, Levene’s test for homogeneity of variances, and visual inspection of histograms. Interaction between group (CFS vs. healthy controls [HC]) and sex (male vs. female) was examined via two-way ANOVA. Prior to this, the Shapiro-Wilk and Levene’s tests confirmed normality and equal variances; residual plots further validated these assumptions.

Raw, beat-to-beat cardiovascular data collected via the Task Force Monitor (TFM) underwent preprocessing before use in classifiers. Outliers (>3 SD from the mean) were replaced with the preceding row’s value using the `dplyr` package [20]. Ectopic beats were manually removed. To reduce redundancy, highly correlated variables were excluded, leaving 16 TFM parameters for further analysis. Pandas and NumPy with Python 3.9 facilitated preprocessing and analysis [21–23].

Data were divided into training and testing sets, with 20% of subjects from each group allocated to the test subset. Both sets were standardized via scaling. The training set contained 36,624 rows; the test set 13,035 rows.

An exhaustive search over all 65,535 subsets of the 16 features was performed using `itertools`. Within each fold of stratified 5-fold cross-validation, XGBoost

(with `random_state=42`, `use_label_encoder=False`, `tree_method='hist'`, `eval_metric='auc'`, `scale_pos_weight=0.544`, and `device='cuda'`) evaluated each subset. The subset yielding the highest Matthews correlation coefficient (MCC) was selected; final feature choice was further guided by ROC AUC and domain expertise. This nested procedure ensured no data leakage during feature selection. The selected six predictors included breathing frequency (BF), stroke volume index (SV/FFM), high-frequency RR interval (HF RRI), heart rate signed difference (HR difference), stroke volume signed difference (SV difference), and diastolic blood pressure signed difference (dBP difference), with a binary target variable (CFS vs. HC).

Stratified 5-fold cross-validation for model training and validation was applied. Importantly, all preprocessing steps, such as feature scaling (StandardScaler) and outlier handling, were fit exclusively on the training partition within each fold and then applied to validation and test partitions. This procedure prevents data leakage. Transformer residuals (errors between predicted and true labels) were calculated using out-of-fold predictions for training folds only and incorporated as additional features for the successive XGBoost classifier. For the test set, Transformer residuals were generated without reference to true labels, adhering to strict separation between training and testing phases.

Machine learning pipeline

Transformer model architecture

The Transformer included three attention heads with a key dimension of six and residual connections [24]. Class weights addressed the imbalance by weighting the under-represented HC class more during training. Although no true temporal sequence was modeled ($t = 1$), Rotary Positional Embedding was retained as a structured feature transformation that enhances representational diversity and regularizes learning through sinusoidal mixing of input dimensions [25]. A Gated Recurrent Unit (GRU) layer with 16 units and 0.01 dropout followed [26]. The GRU layer, though not processing temporal sequences, leveraged its gating architecture to perform adaptive feature selection and nonlinear integration, effectively functioning as a trainable feature compressor that enhanced signal-to-noise ratio. Outputs were flattened and passed through two dense layers (8 and 2 units) using Leaky ReLU activations (`negative_slope = 0.1`) with `he_normal` initialization [27]. The final single-unit output layer applied sigmoid activation with `glorot_normal` initialization (see Figure S2) [28].

The model was compiled with the AdamW optimizer (learning rate=0.00002), binary cross-entropy loss, and metrics including TP, TN, FP, and FN. Early stopping and learning rate reduction callbacks prevented overfitting

and adjusted learning rates dynamically. The Transformer was trained on preprocessed data with class weights.

Although Transformers are typically used for sequential data, in this implementation each heartbeat was processed independently. The attention mechanism thus operated across input features rather than time steps, and positional encoding was included for architectural consistency but had no temporal meaning. Sequences shorter than the set window length are padded via zero filling to maintain a consistent input size. Although the architecture includes components typical of sequential models (e.g., Rotary Encoding, GRU), each heartbeat was processed independently (timesteps = 1), so the model effectively functions as a feature transformer rather than a true time-series learner. No truncation beyond fixed-window processing was applied. A sequential machine learning pipeline was implemented to differentiate CFS patients from HC using beat-to-beat autonomic measurements. The pipeline first employed a Transformer neural network trained on standardized physiological features, followed by an XGBoost classifier that incorporated both the original features and residuals (i.e., signed classification errors: $r_i = y_i - \hat{y}_i^{\text{binary}}$ derived from the Transformer's out-of-fold predictions. To prevent data leakage, subject-level stratified splitting was used: 20% of participants were held out as a final test set, while the remaining 80% underwent 5-fold cross-validation. Within each fold, the Transformer was trained on 4/5 of the subjects, and its binary predictions on the held-out validation fold were used to compute residuals, which were then appended as an additional feature exclusively for training XGBoost on that same validation fold, ensuring residuals were never derived from data used in Transformer training. All preprocessing steps, including feature scaling with StandardScaler, were fit only on training data within each fold and applied to validation and test sets. For the final test set evaluation, Transformer residuals were computed using the fully retrained Transformer model and the known true labels. —During test evaluation, Transformer residuals were generated via the final Transformer model without accessing true test labels and combined with original features for XGBoost prediction. This approach leverages the Transformer's capacity to learn complex, non-linear representations of high-dimensional physiological features, while XGBoost capitalizes on those representations, particularly by focusing on instances where the Transformer's predictions deviate from the true labels, to refine classification through powerful gradient-boosted decision rules. This sequential combination enhances overall robustness and accuracy (Figure S3) [29].

XGBoost classifier and sequential learning

XGBoost classified positive instances (CFS beat-to-beat cardiovascular and body composition data plus

Transformer errors) against negatives (HC) [30] based on 6 physiological predictors (BF, SV/FFM, HF RRI, HR difference, SV difference, DBP difference.) and Transformer errors" as the 7th input. Key parameters were: objective = binary; logistic, random_state = 42, tree_method = hist, eval_metric = auc, and scale_pos_weight adjusted for class imbalance. XGBoost was chosen for robust package support, interpretability, and computational efficiency.

Grid search optimized hyperparameters as: grow_policy = depthwise, reg_lambda = 1.0, alpha = 0.5, subsample = 0.9, colsample_bytree = 0.7, max_depth = 11, n_estimators = 100, min_child_weight = 9, max_bin = 512, gamma = 0.1, eta = 0.28 [30].

Ensemble learner diagnostics

To assess model reliability and potential data leakage, we performed a comprehensive diagnostic evaluation of our two-stage architecture (Transformer and XGBoost). We evaluated the informativeness and validity of error feature based on Transformer output passed to XGBoost as an additional feature using three complementary approaches: (1) a permutation test, where the error feature was randomly shuffled to assess its contribution to predictive performance; (2) an ablation study, comparing XGBoost performance with and without the error feature; and (3) an analysis of error distribution and correlation with true labels, including point-biserial correlation. All diagnostics were conducted on an independent validation set with strict subject-wise separation to preserve data integrity.

Model evaluation and interpretability

Because clinical diagnosis occurs at the patient (subject) level, not at the individual heartbeat level, all primary performance metrics are reported at the subject level. Beat-level metrics are provided in the Supplementary Material for completeness but do not reflect diagnostic utility. For each subject, beat-level predicted probabilities from the model were averaged across all heartbeats to produce a single subject-level probability. A threshold of 0.5 was applied to assign the final binary classification (CFS vs. HC). Evaluation metrics included confusion matrix statistics, AUC, accuracy, precision, recall, F1, and Matthews correlation coefficient (MCC) [31]. The subject-level classification metrics (AUC, accuracy, MCC) and their confidence intervals were calculated using a bootstrap resampling approach. Initially, beat-level predictions from the Transformer and XGBoost models were aggregated to the subject-level by averaging the predicted probabilities for each unique subject ID. Subsequently, 2000 bootstrap samples were generated by resampling the aggregated subject-level data with replacement. For each bootstrap iteration, the AUC, Accuracy, and MCC were calculated using the resampled

subject-level predictions and their corresponding true labels. The 95% confidence intervals for each metric were then determined by taking the 2.5th and 97.5th percentiles of the distribution of metrics obtained across all successful bootstrap samples, providing an estimate of the uncertainty around the point estimates derived from the original aggregated subject-level data. Global interpretability was achieved through SHapley Additive exPlanations (SHAP) values, and tree visualization was used with Dtreviz [32, 33].

Results

Population characteristics

A comparison of clinical and demographic data between CFS patients and healthy controls is shown in Table S1. HC participants were characterized by a lower percentage of adipose tissue (20.8 ± 9.9 vs. 25.7 ± 7.2 , $p < 0.001$) and visceral fat level in comparison to CFS patients (4.1 ± 3.4 units vs. 5.0 ± 2.8 , $p = 0.03$) (Table S1). Body composition data were missing for 10 CFS patients and 2 HCs. Missing body composition data were imputed using the mean value for each sex and group. Missing visceral fat data were imputed using the median value. From 112 CFS patients, there were 75 females (67%), and from 61 HC, 42 patients were females (69%). Table S2 shows a comparison between CFS patients and healthy controls in relation to participants' sex. On average, females weighed less than males, had lower free fat mass and bone mass in kilograms, and had higher adipose tissue percentage (all comparisons p (Bonferroni-corrected) < 0.001). In addition, females from the CFS group had a higher adipose tissue percentage than HC females (p (Bonferroni-corrected) < 0.001) (Table S2). Most of the cohort of patients (84%) suffered from PEM (Figure S4, panel A). 92% suffered from brain fog (Figure S4, panel A). 17% of the examined CFS patient cohort had 5, 36% had 6, 21% had 7, and 13% had 8 symptoms (Figure S4, panel B). The mean CFS symptom duration was 4.6 years (range from 0.5 to 20 years of duration). There were no severe, bed- or house-bound patients in this cohort, as we could not afford a team that would do an in-bed examination; therefore, all patients had to go to our research centre.

Table 1 Between-group comparison of values of features that serve as predictors in applied transformer and XGBoost sequential learning classifiers

Features	CFS	HC	<i>p</i> -value
	Mean \pm SD	Mean \pm SD	
HR difference [bpm]	0 \pm 4.0	0 \pm 4.0	0.44
SV/FFM [ml/kg]	1.8 \pm 0.5	1.9 \pm 0.5	< 0.001
BF [breaths/min]	15.3 \pm 5.5	15.5 \pm 5.6	0.17
HF RRI [ms^2]	550.4 \pm 954.1	681.7 \pm 1020.7	< 0.001
SV difference [ml]	0 \pm 3.0	0 \pm 3.1	0.39
dBp difference [mmHg]	0 \pm 2	0 \pm 2.2	0.65

Model performance

Using the Kendall rank correlation coefficient, a correlation heatmap between the target and features was initially assessed for their applicability as predictors of the subject group. Based on physiological and mathematical redundancy, a set of 6 features were selected: breathing frequency (BF) (measured as breaths per minute [breaths/min]), SV/FFM [ml/kg], which is stroke volume (SV) measured in milliliters [ml] divided by free-fat mass, high-frequency R-R interval (HF RRI) [ms^2], heart rate (HR) difference, SV difference, and diastolic blood pressure (dBp) difference were calculated based on the difference between subsequent heartbeats (HR difference [bpm], SV difference [ml], dBp difference [mmHg]). Figure S5 shows a correlation heatmap between the target and features included in the training and test datasets. No highly correlated features with the target (subject groups CFS vs. HC as a predicted variable) were observed. The study cohort consisted of 112 CFS patients and 61 healthy controls (HCs), leading to a slight imbalance. This imbalance was maintained in the validation and training sets (80% of the total CFS and HC sample) and the test set (20% of the total CFS and HC sample, specifically 22 CFS patients and 12 HCs). Consequently, this unequal representation could influence the prediction accuracy for CFS versus HC categories.

Feature values used in the machine learning pipeline to classify CFS vs. HC were compared using different tests. SV/FFM and HF RRI were significantly higher in the HC than in the CFS group (both p -values < 0.001) (Table 1). Figure S6 shows histograms of analyzed features (predictors) in both the CFS and control groups. Despite statistically significant between-group differences in the two features, no substantial differences in value distribution between HC and CFS patients in those predictors were noted (Figure S6). Therefore, we applied a potentially more sensitive deep learning model to the data.

This study introduces a novel approach combining a deep learning model (Transformer neural net) with a shallow learning model (tree-based eXtreme Gradient Boosting, XGBoost). The performance of this combined model was compared to that of the individual XGBoost and Transformer models. Specifically, the Transformer was initially applied, and subsequently, XGBoost was trained as a classifier to learn from the Transformer's errors (Figure S3). The process involved three stages: training, validation, and testing. First, the Transformer, XGBoost, and the combined Transformer and XGBoost models were trained using beat-to-beat data from 60% of the CFS and HC participants. Next, the validity of each model was assessed using a 5-fold stratified validation on 20% of the total CFS and HC participants. Finally, the subject-level performance of all three classifiers was evaluated on a separate, hold-out test set (20%

of the participants), with results summarized in Table 2. Notably, the combined Transformer and XGBoost model achieved a subject-level Matthews Correlation Coefficient (MCC) of 0.79, an accuracy of 0.89, and a Receiver Operating Characteristic Area Under the Curve (ROC AUC) score of 1 (Table 2). In addition, it substantially outperforms Transformer or XGBoost applied on beat-level metrics (Table S3). The mean MCC across cross-validation folds during nested feature selection was 0.58 ranged from 0.56 to 0.6. Because predictions are averaged across all heartbeats per subject, Figure S7 shows subject-level ROC curves plot reflect clinically relevant, patient-level classification rather than beat-level accuracy. Performance of Transformer, XGBoost, and an ensemble of Transformer and XGBoost models is better than the reference (Figure S7). The Decision Curve Analysis (DCA) plot evaluates the clinical utility of the same models by estimating their Net Benefit across a range of clinically plausible risk thresholds (Figure S8). An XGBoost classifier and an ensemble of Transformer and XGBoost could be considered clinically useful, because their Net Benefit exceeds both benchmarks (diagnosing none and all as CFS patients), indicating that using their predictions leads to better outcomes than default strategies.

The subject-level confusion matrix (Table S6) visualizes the diagnostic performance of the Transformer and XGBoost ensemble model at the clinically relevant

Table 2 Subject-level metrics with 95% confidence interval (CI)

Metric	Point estimate Transformer (95% CI)	Point estimate XGBoost (95% CI)	Point estimate Transformer and XGBoost (95% CI)
AUC	0.77 (0.60–0.92)	0.99 (0.97–1)	1 (0.98–1)
Accuracy	0.46 (0.29–0.63)	0.91 (0.83–1)	0.89 (0.77–0.97)
MCC	0.26 (0.12–0.40)	0.83 (0.66–1)	0.79 (0.61–0.94)

ROC AUC-Receiver Operating Characteristic Area Under the Curve, MCC-Matthews Correlation Coefficient

patient level, showing that all 12 healthy controls were correctly classified (100% specificity; 0 false positives), while 14 of 23 CFS patients were correctly identified as cases (61% sensitivity), with 9 false negatives. A confusion matrix for the Transformer and XGBoost classifier on beat-level data is shown in Fig. 1. A confusion matrix for the Transformer and XGBoost classifier on beat-level data is shown in Figure S9. Subject-level confusion matrix for the Transformer model is presented in Table S4, for XGBoost in Table S5. Leakage diagnostics confirmed that the Transformer error feature was highly informative but not indicative of data leakage. Permuting the error feature reduced XGBoost accuracy from 99.55% to 61.05% ($\Delta = 38.5\%$), demonstrating its strong predictive contribution. Ablation analysis showed that including the error feature improved accuracy by 4.21% (from 95.34% to 99.55%). The error feature exhibited a moderate but

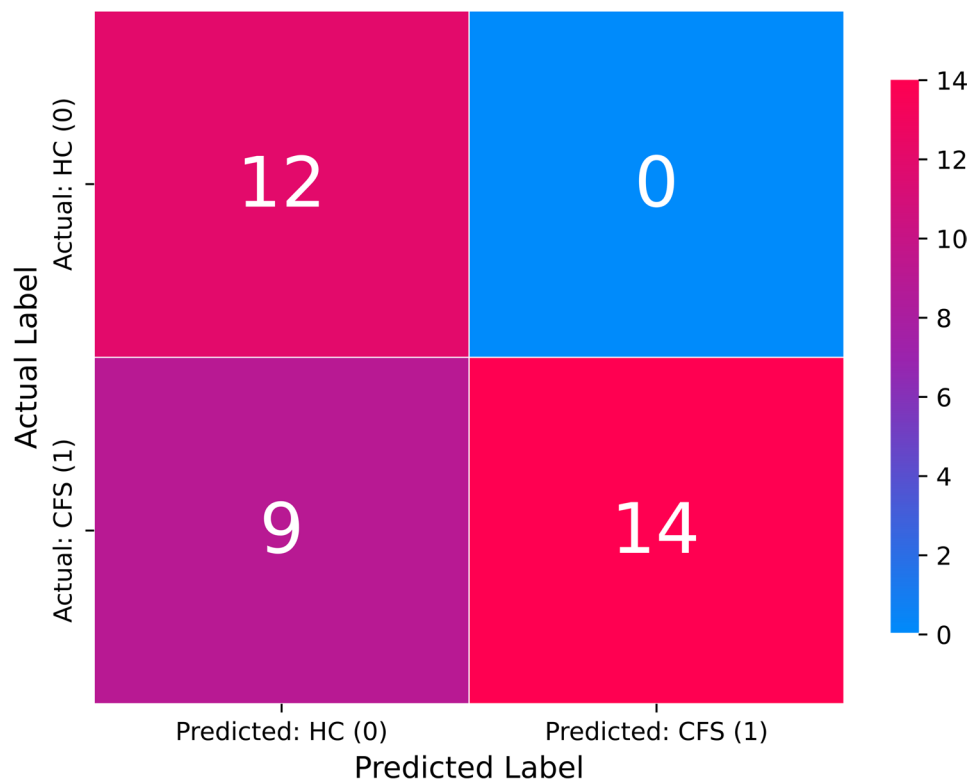


Fig. 1 Transformer and XGBoost sequential learning classifier confusion matrix based on beat-level data

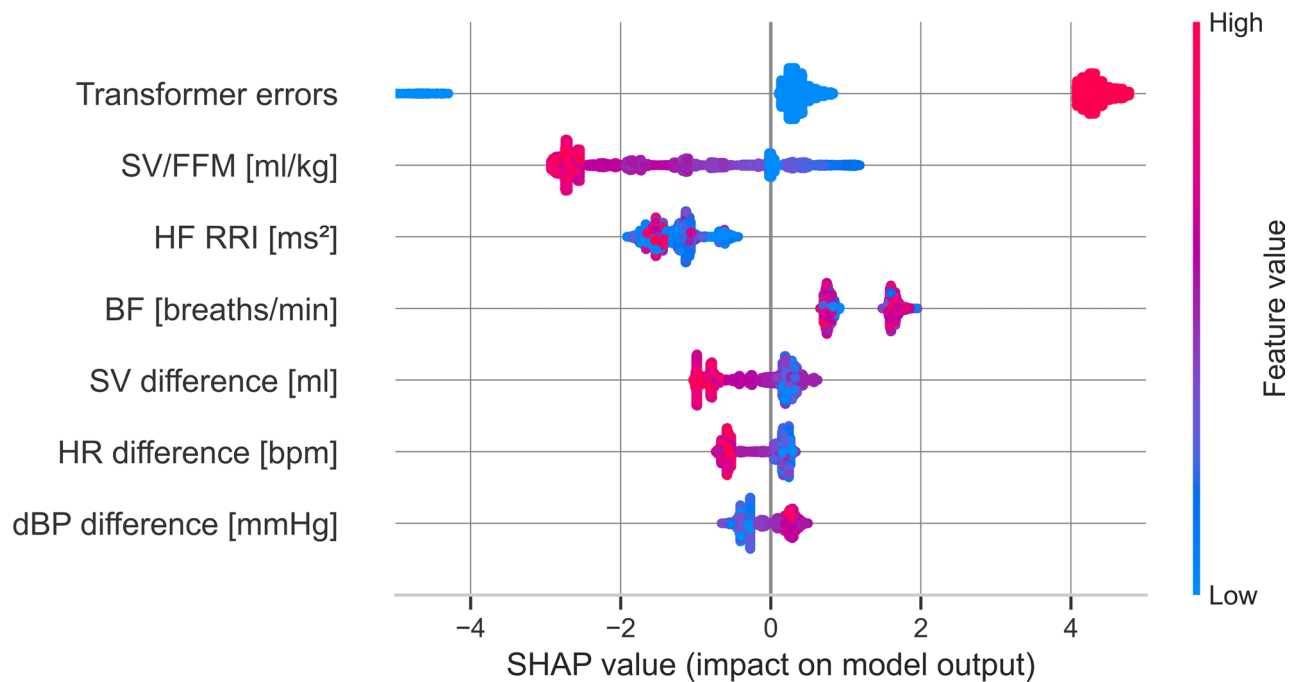


Fig. 2 SHapley Additive exPlanations (SHAP) values for the applied classifier predicting being classified as CFS patients. Transformer errors are values calculated based on errors related to the prediction of a Transformer model, based on which XGBoost classifier was learning, *breathing frequency* (BF) (measured as breaths per minute [breaths/min]), *stroke volume index* (SV/FFM) which is stroke volume (SV) measured in milliliters [ml] divided by free-fat mass (FFM), *high-frequency R-R interval* (HF RRI) [ms²]. Heart rate (HR) difference, SV difference, and diastolic blood pressure (dB) difference were calculated based on the difference between subsequent heartbeats (HR difference [bpm], SV difference [ml], dBP difference [mmHg])

significant point-biserial correlation with the true label ($r = 0.605$, $p < 0.001$), and error distributions differed systematically between classes, indicating that the Transformer made class-dependent mistakes rather than random errors. Collectively, it was concluded that these results confirm that the error feature captures meaningful, non-leaky signal related to model uncertainty.

Key predictive features and model interpretability

To enhance model transparency, we employed explainable artificial intelligence (XAI) techniques. Specifically, SHapley Additive exPlanations (SHAP) values were utilized to explain why the model made particular predictions for the predictors in both the CFS and HC groups. These values quantify each predictor's contribution (each ANS parameter), providing detailed insights into individual predictions rather than offering a single global measure of feature importance. This means that SHAP values reveal the specific reasons behind a model's prediction for each data point (that is, the ANS parameter value from each heartbeat). Figure 2 shows the SHAP values for the classifier that predicts whether a patient has CFS based on single values from beat-to-beat measurements. In the figure, the features are ranked in descending order based on their importance, with the most influential predictors at the top. According to the ranking, the most important predictors are, in order: transformer model errors, SV/

FFM, HF RRI, BF, the difference in SV, the difference in HR, and the difference in dBP. Specifically, higher transformer errors (displayed as red dots) are linked to an increased probability of being classified as CFS. In addition, lower differences in HR (a larger decrease in HR, i.e., more beat-to-beat heart decelerations) and raw values of HF RRI, which are shown with blue dots and indicate reduced cardiac vagal tone and higher differences in dBP (shown with red dots, i.e., more consecutive beats, in which dBP raised from beat to beat), are associated with a higher likelihood of CFS classification. Finally, low to reduced beat-to-beat changes (i.e., variability in SV, denoted as blue dots in SV difference), and low SV/FFM values (indicating less effective cardiac hemodynamics) also contribute to the CFS classification, while the association between BF and classification remains less defined.

Discussion

In this study, we applied an automated machine learning (ML)/artificial intelligence (AI) pipeline incorporating automated feature engineering, which included feature extraction, feature selection, and automatic tuning. Using a classifier composed of Transformer and XGBoost in a Sequential Learning framework, the model achieved strong subject-based results: a Matthews correlation coefficient (MCC) of 0.79, accuracy of 0.89, and

ROC AUC of 1. An AUC ROC of 1 indicates perfect discrimination between beat-to-beat autonomic nervous system (ANS) data from chronic fatigue syndrome (CFS) patients versus healthy controls [34]. The accuracy indicates 89% of predictions were correct, but caution is warranted due to data imbalance: twice as many CFS patients as healthy controls, which may affect performance. The high-resolution time-series nature meant each subject generated hundreds of rows.

The model was trained, validated, and evaluated on this data, predicting features derived from single heartbeats and assessed on beat-level and subject-level data. The subject-level confusion matrix for the final Transformer and XGBoost sequential learning ensemble reveals a highly specific but moderately sensitive diagnostic performance in distinguishing CFS from HCs. Out of 12 healthy controls, all were correctly classified as non-CFS (12 true negatives, 0 false positives), yielding 100% specificity. It is a critical strength for a diagnostic tool, as it ensures no healthy individual is mislabeled as having CFS, thereby avoiding unnecessary psychological burden, further invasive testing, or inappropriate interventions. Among the 23 CFS patients in the test set, 14 were correctly identified (true positives), while 9 were misclassified as healthy (false negatives), resulting in a sensitivity of approximately 61%. This indicates that while the model is exceptionally reliable at confirming the absence of CFS, it misses a non-trivial subset of actual CFS cases, potentially those with milder autonomic dysfunction or atypical physiological profiles. The MCC of 0.79 highlights robust performance across all confusion matrix metrics, signifying effective classification of CFS versus healthy states [35]. This approach compares favorably to previous CFS diagnostic methods [36]. We propose further research integrating big data sequential learning, combining tree-based models and neural networks across multi-systemic and cognitive biomarkers to improve sensitivity and specificity.

In the current study, to enhance model transparency, we employed explainable artificial intelligence (XAI) techniques. Specifically, we used SHapley Additive exPlanations (SHAP) to quantify the contribution of each input feature to the model's predictions, allowing us to identify which cardiovascular parameters most strongly influenced the classification of CFS versus healthy controls. This global interpretability supports clinical insight and trust in the model's decisions.

Higher-magnitude heart rate (HR) decelerations, observed more frequently in CFS, indicate pronounced parasympathetic (vagal) bursts modulating cardiac rhythm. In contrast, healthy controls exhibit more accelerations. These large decelerations in CFS reflect hypersensitive or dysregulated vagal control, consistent with autonomic imbalance studies in CFS showing both

sympathetic overactivity and parasympathetic irregularities [7, 37]. Reduced cardiac vagal tone, evidenced by decreased absolute values and diminished beat-to-beat differences in high-frequency RR intervals (HF RRI), further confirms diminished parasympathetic regulation. Lower HF RRI and heart rate variability (HRV) have been documented as characteristic markers of CFS, with meta-analyses showing reductions in high-frequency HRV components at rest and during stress testing [7]. These changes indicate weakened autonomic flexibility and reduced cardiac adaptability [37]. Additionally, the increased frequency of beat-to-beat increases in diastolic blood pressure (DBP) suggests heightened sympathetic vasoconstrictor activity or baroreflex dysfunction, both of which have been reported in multiple studies on CFS cardiovascular pathology. This is consistent with the sympatho-excitatory dominance seen in CFS patients and likely relates to, or is a maladaptive consequence of, altered blood pressure regulation and hemodynamic instability [38]. Impaired venous compliance and hypovolemia (low blood volume) are key factors reducing preload, the amount of blood entering the heart during diastole. Since stroke volume (SV) variability depends on the heart's dynamic capacity to adjust SV in response to changes in venous return and autonomic inputs, limited preload reduces SV variability. Reduced blood volume decreases the stretch on cardiac muscle fibers (sarcomere length), weakening the Frank-Starling mechanism, which normally boosts SV in response to increased filling. Consequently, in CFS patients, the heart operates on a lower, flatter part of the Frank-Starling curve, limiting beat-to-beat SV fluctuations [39]. Autonomic dysfunction, including disrupted sympathetic and parasympathetic balance, may aggravate this by impairing vascular tone regulation and baroreflex sensitivity, further compromising venous return and cardiac filling. Additionally, cardiac bioenergetic impairments could reduce myocardial contractility, limiting SV responsiveness [40]. This diminished SV variability reflects hemodynamic inefficiency. Further studies should examine whether this contributes to symptoms like orthostatic intolerance, exercise intolerance, and fatigue by restricting the cardiovascular system's ability to dynamically respond to physiological demands and maintain adequate oxygen delivery during activity. A lower stroke volume index (SV/FFM), defined as stroke volume divided by fat-free mass (FFM), was associated with a higher likelihood of CFS classification. This aligns with previous findings of a 10.2% reduction in stroke volume in CFS patients [41]. Additionally, diminished end-diastolic volume, cardiac output, stroke index, and cardiac index have been reported in this population [42–44]. Indexing stroke volume to FFM rather than total body mass better reflects oxygen delivery to metabolically active tissues, providing clinical advantages by

aligning with the body's metabolic demands [45]– [46]. The reduced SV/FFM in CFS may reflect impaired preload due to hypovolemia, autonomic dysfunction limiting ventricular filling, or intrinsic myocardial inefficiency. These factors might collectively compromise the heart's ability to meet the metabolic demands of active tissues. This pathophysiological pattern aligns with the exercise intolerance and post-exertional malaise characteristic of CFS, suggesting that SV/FFM serves not only as a hemodynamic marker but also as an indicator of the bioenergetic deficit in this condition.

An increased magnitude of HR decelerations, reduced cardiac vagal tone (HF RRI), increased sympathetic DBP variability, and lowered SV variability and SV/FFM portray a complex interplay of overactive, irregular parasympathetic bursts and sympathetic dominance combined with cardiac mechanical impairment. This constellation underpins the autonomic-cardiac dysregulation defining CFS physiology, supporting prior findings from comprehensive meta-analyses and mechanistic studies addressing the autonomic nervous system and cardiovascular involvement in CFS [7]. What is novel is that this constellation of parameter values might indicate disruption of the baroreceptor reflex in CFS patients. Typically, a change in blood pressure is accompanied by a consecutive change in heart rate and/or stroke volume. In this study, despite higher beat-to-beat fluctuations in diastolic blood pressure, diminished heart rate and stroke volume responses were noted in CFS patients.

Although respiratory sinus arrhythmia is reduced in CFS, respiratory mechanics or thoracic pressure changes may influence stroke volume dynamics; however, no clear link between breathing frequency and CFS classification was found. Therefore, consistent with other authors, the lower cardiac volume in CFS is unlikely to indicate classical cardiovascular disease [47]. Reduced SV/FFM may relate to lower cardiac contractility or preload, but we hypothesize hypovolemia as the primary correlate rather than impaired contractility. Further studies are needed to confirm this, with the ongoing SIMPLE study investigating saline infusion treatment [48].

In summary, CFS patients likely exhibit decreased vagal cardiac input and baroreceptor reflex function, alongside increased sympathetic vascular outflow. These findings may improve diagnostic criteria by combining exclusionary factors with objective inclusion criteria, such as autonomic nervous system dysfunction, standardized cognitive testing, and cardiopulmonary exercise testing to confirm post-exertional malaise [49]. AI classifiers analyzing beat-to-beat autonomic nervous system data could become valuable diagnostic tools, facilitating patient access to necessary support, including financial assistance.

Study strengths and limitations

This is the first extensive ML/AI pipeline study applying automated feature engineering, including extraction, selection, and tuning to improve CFS diagnostics. We incorporated multiple classification metrics, including MCC, which better accounts for class imbalance compared to accuracy alone [50]. The use of SHAP enabled the identification of key predictors and their relationship to outcomes, enhancing interpretability.

Limitations include moderate class imbalance (112 CFS patients vs. 61 controls), preserved in training and testing. While imbalanced data remains challenging in ML, this ratio is modest relative to common highly skewed datasets (1:100 to 1:10,000) [51]. To address the class imbalance (112 CFS vs. 61 controls), we implemented algorithmically informed class weighting strategies. Specifically, the `scale_pos_weight` parameter in the XGBoost classifier was tuned to emphasize minority class examples. Additionally, corresponding class weights were applied during Transformer training. These approaches mitigate bias against the underrepresented class. Future studies could explore synthetic oversampling methods, ensemble learning, or focal loss functions as complementary techniques to further enhance model robustness. Although the sample size ($n = 112$) is relatively small for ML, it is substantial for CFS research. Tabular data remains challenging for deep learning, with gradient-boosted trees still the industry standard [52]. Larger datasets generally improve performance, underscoring the need for clinical data sharing nationwide [53]. In addition, there are multiple methods and effectors that should be measured in rest and in response to stressor(s) to give an overall picture of ANS dysfunction of ME/CFS [54]. Therefore, further studies should extend methods beyond focusing on cardiovascular function as a sole representation of ANS function. In addition, adding non-standard indicators of the cardiovascular system, as SV normalized by FFM in the feature engineering process, might limit the generalizability of the obtained results [54]. Recent computational physiology studies have adopted more robust features to reflect dynamic fluctuations. For instance, Chen et al. applied corrected amplitude envelope correlation across multiple features representing time series [55]. The methods used for preprocessing the RR interval data, specifically the replacement of outliers exceeding 3 standard deviations with the preceding value and the manual removal of ectopic beats, carry a potential risk of introducing artifacts into the signal. As highlighted by Liu et al. [56], signal preprocessing steps, including data replacement and interpolation, can significantly alter the features of interest if not carefully controlled, potentially impacting the integrity of the derived physiological parameters. Future work should consider employing more robust outlier detection

and replacement strategies, potentially validated against physiological models or alternative preprocessing pipelines, to minimize this risk.

While signed mean differences of heart rate (HR difference), stroke volume (SV difference), and diastolic blood pressure (dBP difference) were used as features in the present model, future studies should explore employing absolute differences, root mean square of successive differences (RMSSD), or standard deviation-based metrics, which capture beat-to-beat variability more robustly. This advancement aims to enhance the physiological interpretability and clinical relevance of dynamic autonomic fluctuations. We did not apply parameters that are point parameters based on larger time window frames. All parameters used were based on raw beat-to-beat data to not decrease the amount of potential information that would serve as input to classifiers.

The single-center retrospective design limits generalizability; future multi-center validation, including COVID-19-related CFS cases, is needed. This study represents a secondary analysis of prospectively collected baseline data from two intervention studies (IAP and WBC+SS) conducted at a single center. The single-center nature and exclusion of post-COVID-19 CFS patients may limit the generalizability of our findings. Future research should aim for multi-center studies recruiting diverse populations, including post-viral fatigue syndromes to validate and broaden the applicability of the proposed AI pipeline. Recruitment of patients and controls followed rigorous inclusion/exclusion criteria, including screening for comorbidities, medication use, and lifestyle factors, to minimize confounding and selection bias.

The input to the Transformer was a single time step per patient, therefore ROPE did not encode real temporal order. ROPE might act as a structured feature transformation that regularized learning, while the GRU, though not processing a true sequence, provided richer nonlinear feature interactions through its gating mechanism. Those elements might enhance the model's representational capacity, leading to improved performance, not through temporal modeling, but rather via better feature representation and implicit regularization on tabular data. The Transformer-XGBoost ensemble leveraged complementary strengths: Transformers captured complex non-linear patterns, while XGBoost mitigated overfitting and aided interpretability. Sequential learning, where XGBoost learns from Transformer errors, enhanced classification but added complexity, requiring careful tuning. Ablation studies confirmed that this design improved generalization without introducing data leakage, as all error features were derived strictly from in-sample (cross-validated) or post-hoc test predictions. Implementing the proposed AI pipeline in clinical practice presents several challenges that must be carefully

considered. The 'Transformer error' feature was defined as the difference between true label and binary prediction. During cross-validation, this was computed using out-of-fold predictions; at test time, it was computed using the held-out test labels. While this feature is not available in real-world deployment, it serves as a proxy for instance difficulty and enables controlled ablation to assess the value of residual modeling. A deployable version would replace this with an unsupervised uncertainty estimate. The Task Force Monitor, required for continuous beat-to-beat cardiovascular and autonomic measurements, while highly effective and non-invasive, is sophisticated and represents a significant capital investment, limiting its accessibility primarily to specialized centers. Integration of this technology into routine diagnostic protocols would require infrastructural upgrades, trained personnel, and workflow adjustments, which may pose barriers in resource-limited settings. Nevertheless, increasing automation and decreasing costs of monitoring technologies may improve accessibility over time. Nevertheless, the reliance on Transformer residuals as a feature limits clinical applicability, as residuals require knowledge of the true diagnosis. Future work will replace residuals with uncertainty estimates or ensemble disagreement metrics to enable real-world application. Regarding confounding factors, the study cohort was rigorously screened using clinical and neurological evaluations to exclude participants with known medical conditions or treatments that could affect autonomic function, thereby minimizing confounding due to medication use or subclinical comorbidities. However, residual confounding cannot be entirely excluded given the complexity of the condition and heterogeneity of the population. Future research should incorporate detailed medication and comorbidity profiling to further elucidate their impact on autonomic biomarkers. Compared to existing diagnostic methods for chronic fatigue syndrome, which predominantly rely on subjective clinical criteria with variable sensitivity and specificity, this AI-based approach offers objective, quantifiable measures of autonomic nervous system dysfunction with promising accuracy and robustness both on beat-level and subject-level data. While blood-based biomarkers and cognitive testing provide valuable complementary data, their accessibility and standardization remain limited. Thus, the AI pipeline utilizing beat-to-beat cardiovascular measurements may provide a scalable, reproducible, and sensitive adjunct for improving diagnosis, patient stratification, and monitoring in clinical practice, once practical and economic barriers are addressed.

Abbreviations

AI	Artificial Intelligence
ANS	Autonomic Nervous System
BF	Breathing Frequency

CFS	Chronic Fatigue Syndrome
dBp	Diastolic Blood Pressure
FFMI	Fat-Free Mass Index
GRU	Gated Recurrent Unit
HR	Heart Rate
HF RRI	High-Frequency R-to-R Interval
MCC	Matthews Correlation Coefficient
ML	Machine Learning
PEM	Post-Exertional Malaise
ROC AUC	Receiver Operating Characteristic Area Under the Curve
SHAP	SHapley Additive exPlanations
SI	Stroke Index
SV/FFM	Stroke Volume Indexed to Fat-Free Mass Index
SV	Stroke Volume
TFM	Task Force Monitor
XAI	Explainable Artificial Intelligence

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-025-07433-y>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7
Supplementary Material 8
Supplementary Material 9
Supplementary Material 10

Acknowledgements

We would like to express gratitude for important cues on how to improve the clinical context of an early draft version from Prof Carmen Scheibenbogen, MD.

Authors' contributions

JS, and PZ conceived and designed the study, collected and interpreted data, drafted and revised the manuscript, supervised the study. SK analyzed data, drafted and revised the manuscript, supervised the study. All other authors helped in pre-processing data and revised the manuscript. SK, JS, and PZ had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Funding

None.

Data availability

Written requests for access to the data reported in this paper will be considered by SK and PZ and a decision made about the appropriateness of the use of data. If the use is appropriate, a data sharing agreement will be put in place before a fully de-identified version of the dataset used for analysis with individual participant data is made available.

Declarations

Ethics approval and consent to participate

The research was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee, Ludwik Rydygier Memorial Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University, Toruń (KB 332/2013 and KB 660/2017). All study participants gave written, informed consent for their data to be used for research purposes.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 16 September 2025 / Accepted: 1 November 2025

Published online: 23 December 2025

References

- Aoun MS, Hainselin M, Gounden Y, Sirbu CA, Sekulic S, Lorusso L, et al. Systematic review and meta-analysis of cognitive impairment in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Sci Rep*. 2022;12:2157. <https://doi.org/10.1038/s41598-022-06119-1>.
- Van Cauwenbergh D, Nijs J, Kos D, Van Weijnen L, Struyf F, Meeus M. Malfunctioning of the autonomic nervous system in patients with chronic fatigue syndrome: a systematic literature review. *Clin Physiol Funct Imaging*. 2014;34:335–52. <https://doi.org/10.1111/eci.12256>.
- Estévez-López F, Mudie K, Wang-Steering X, Bakken IJ, Ivanovs A, Castro-Marrero J, et al. Systematic review of the epidemiological burden of myalgic encephalomyelitis/chronic fatigue syndrome across Europe: current evidence and EUROMENE research recommendations for epidemiology. *J Clin Med*. 2020;9:1557. <https://doi.org/10.3390/jcm9051557>.
- Brown A, Jason LA. Meta-analysis investigating post-exertional malaise between patients and controls. *J Health Psychol*. 2020;25:2053–71. <https://doi.org/10.1177/1359105320914448>.
- Nacul L, Authier FJ, Scheibenbogen C, Lorusso L, Helland IB, Martin JA, et al. European network on myalgic encephalomyelitis/chronic fatigue syndrome (EUROMENE): expert consensus on the diagnosis, service provision, and care of people with ME/CFS in Europe. *Med (Kaunas)*. 2021;57:510. <https://doi.org/10.3390/medicina57050510>.
- Bateman L, Basted AC, Bonilla HF, Chheda BV, Chu L, Curtin JM, et al. Myalgic encephalomyelitis/chronic fatigue syndrome: essentials of diagnosis and management. *Mayo Clin Proc*. 2021;96:2861–78. <https://doi.org/10.1016/j.jm.2021.07.004>.
- Nelson MJ, Bahl JS, Buckley JD, Thomson RL, Davison K. Evidence of altered cardiac autonomic regulation in myalgic encephalomyelitis/chronic fatigue syndrome: A systematic review and meta-analysis. *Med (Baltim)*. 2019;98:e17600. <https://doi.org/10.1097/MD.00000000000017600>.
- Słomko J, Estévez-López F, Kujawski S, Zawadka-Kunikowska M, Tafil-Klawe M, Klawe JJ, et al. Autonomic phenotypes in chronic fatigue syndrome (CFS) are associated with illness severity: A cluster analysis. *J Clin Med*. 2020;9:2531. <https://doi.org/10.3390/jcm9082531>.
- De Mauro A, Greco M, Grimaldi M. What is big data? A consensual definition and a review of key research topics. *AIP Conf Proc*. 2015;1644:97–104. <https://doi.org/10.1063/1.4907823>.
- Xu J, Lodge T, Kingdon C, Strong JW, MacLennan J, Lacerda E, et al. Developing a blood cell-based diagnostic test for myalgic encephalomyelitis/chronic fatigue syndrome using peripheral blood mononuclear cells. *Adv Sci (Weinh)*. 2023;10:2302146. <https://doi.org/10.1002/adv.202302146>.
- Bentéjac C, Csörgo A, Martínez-Muñoz G. A comparative analysis of Xgboost. *ArXiv*. 2019;1911:392. [abs/1911.392](https://arxiv.org/abs/1911.392).
- Shwartz-Ziv R, Armon A. Tabular data: deep learning is not all you need. *Inf Fusion*. 2022;81:84–90. <https://doi.org/10.1016/j.inffus.2021.11.003>.
- Rudin C. Why black box machine learning should be avoided for high-stakes decisions, in brief. *Nat Rev Methods Primers*. 2022;2:81. <https://doi.org/10.1038/s43586-022-00167-9>.
- Kujawski S, Cossington J, Słomko J, Zawadka-Kunikowska M, Tafil-Klawe M, Klawe JJ, et al. Relationship between cardiopulmonary, mitochondrial and autonomic nervous system function improvement after an individualised activity programme upon chronic fatigue syndrome patients. *J Clin Med*. 2021;10:1542. <https://doi.org/10.3390/jcm10071542>.
- Kujawski S, Słomko J, Godlewska BR, Cudnoch-Jędrzejewska A, Murovska M, Newton JL, et al. Combination of whole body cryotherapy with static stretching exercises reduces fatigue and improves functioning of the autonomic nervous system in chronic fatigue syndrome. *J Transl Med*. 2022;20:273. <https://doi.org/10.1186/s12967-022-03483-6>.
- Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A, International Chronic Fatigue Syndrome Study Group. The chronic fatigue syndrome: a comprehensive approach to its definition and study. *Ann Intern*

- Med. 1994;121:953–9. <https://doi.org/10.7326/0003-4819-121-12-19941215-0-00009>.
17. Kouri EM, Pope HG, Jr, Katz DL, Oliva P. Fat-free mass index in users and nonusers of anabolic-androgenic steroids. *Clin J Sport Med.* 1995;5:223–8. <https://doi.org/10.1097/00042752-199510000-00005>.
 18. Fortin J, Klinger T, Wagner C, Sterner H, Madritsch C, Grullenberger R. (1998). The task force monitor – a non-invasive beat-to-beat monitor for hemodynamic and autonomic function of the human body. *Proc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.* 20:—.
 19. Zalewski P, Słomiński K, Klawe J, Tafil-Kławe M. Ocena czynnościowa autonomicznego układu Nerwowego z użyciem systemu task force monitor. *Acta Bio-Opt Inf Med Inżynieria Biomedyczna.* 2008;14:228–34.
 20. Wickham H, François R, Henry L, Müller K, Vaughan D. (2023). dplyr: a grammar of data manipulation. Available from: <https://dplyr.tidyverse.org>
 21. McKinney W. (2010). Data structures for statistical computing in Python. *Proc. 9th Python Sci. Conf.* 61:—.
 22. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with numpy. *Nature.* 2020;585:357–62. <https://doi.org/10.1038/s41586-020-2649-2>.
 23. Van Rossum G, Drake FL. Python 3 reference manual. Scotts Valley, CA: CreateSpace; 2009.
 24. Vaswani A, et al. Attention is all you need. *Adv Neural Inf Process Syst.* 2017;30.
 25. Su J, Ahmed M, Lu Y, Pan S, Bo W, Liu Y. Roformer: enhanced transformer with rotary position embedding. *Neurocomputing.* 2024;568:127063. <https://doi.org/10.1016/j.neucom.2023.127063>.
 26. Cho K. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *ArXiv:* 1406.1078.
 27. He K, Zhang X, Ren S, Sun J. (2015). Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. *Proc. IEEE Int. Conf. Comput. Vis.* 1026–1034.
 28. Glorot X, Bengio Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proc. 13th Int. Conf. Artif. Intell. Stat.* 249–256.
 29. Chen T, Guestrin C. (2016). XGBoost: a scalable tree boosting system. *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 785–794.
 30. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* 2011;12:2825–30.
 31. Boughorbel S, Jarray F, El-Anbari M. Optimal classifier for imbalanced data using Matthews correlation coefficient metric. *PLoS ONE.* 2017;12:e0177678. <https://doi.org/10.1371/journal.pone.0177678>.
 32. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;30.
 33. Parr T, Grover P. (2023). dtreeviz: a python library for decision tree visualization and model interpretation. Available from: <https://github.com/parr/dtreeviz>.
 34. Çorbacioğlu ŞK, Aksel G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turk J Emerg Med.* 2023;23:195. https://doi.org/10.4103/tjem.tjem_75_23.
 35. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics.* 2020;21:1–3. <https://doi.org/10.1186/s12864-019-6413-7>.
 36. Huang K, Lidbury BA, Thomas N, Gooley PR, Armstrong CW. Machine learning and multi-omics in precision medicine for ME/CFS. *J Transl Med.* 2025;23:68. <https://doi.org/10.1186/s12967-024-02797-3>.
 37. Natelson BH, Brunjes DL, Mancini D. Chronic fatigue syndrome and cardiovascular disease: JACC state-of-the-art review. *J Am Coll Cardiol.* 2021;78(10):1056–67. <https://doi.org/10.1016/j.jacc.2021.06.045>.
 38. Newton JL, Okonkwo O, Sutcliffe K, Seth A, Shin J, Jones DEJ. Symptoms of autonomic dysfunction in chronic fatigue syndrome. *QJM.* 2007;100(8):519–26. <https://doi.org/10.1093/qjmed/hcm057>.
 39. Newton JL, Finkelmeyer A, Petrides G, Frith J, Hodgson T, MacLachlan L et al. (2016). Reduced cardiac volumes in chronic fatigue syndrome associate with plasma volume but not length of disease: a cohort study. *Open heart,* 3(1). DOI: 0.1136/openhrt-2015-000381.
 40. Wirth K, Scheibenbogen C. A unifying hypothesis of the pathophysiology of myalgic Encephalomyelitis/Chronic fatigue syndrome (ME/CFS): recognitions from the finding of autoantibodies against β 2-adrenergic receptors. *Autoimmun Rev.* 2020;19(6):102527. <https://doi.org/10.1016/j.autrev.2020.102527>.
 41. Hurwitz BE, Coryell VT, Parker M, Martin P, Laperriere A, Klimas NG, et al. Chronic fatigue syndrome: illness severity, sedentary lifestyle, blood volume and evidence of diminished cardiac function. *Clin Sci (Lond).* 2010;118:125–35. <https://doi.org/10.1042/CS20090055>.
 42. Hollingsworth KG, Hodgson T, Macgowan GA, Blamire AM, Newton JL. Impaired cardiac function in chronic fatigue syndrome measured using magnetic resonance cardiac tagging. *J Intern Med.* 2012;271:264–70. <https://doi.org/10.1111/j.1365-2796.2011.02429.x>.
 43. Miwa K. Cardiac dysfunction and orthostatic intolerance in patients with myalgic encephalomyelitis and a small left ventricle. *Heart Vessels.* 2015;30:484–9. <https://doi.org/10.1007/s00380-014-0510-y>.
 44. van Campen CLMC, Visser FC. (2018). The abnormal cardiac index and stroke volume index changes during a normal tilt table test in ME/CFS patients compared to healthy volunteers are not related to deconditioning. *J Thromb Circ* 2018;10. <https://doi.org/10.29011/JTC-107.000007>.
 45. Nookaew I, Svensson PA, Jacobson P, Jernäs M, Taube M, Larsson I, et al. Adipose tissue resting energy expenditure and expression of genes involved in mitochondrial function are higher in women than in men. *J Clin Endocrinol Metab.* 2013;98:E370–8. <https://doi.org/10.1210/jc.2012-2985>.
 46. Hwaung P, Bony-Westphal A, Muller MJ, Geisler C, Heo M, Thomas DM, et al. Obesity tissue: composition, energy expenditure, and energy content in adult humans. *Obes (Silver Spring).* 2019;27:1472–81. <https://doi.org/10.1002/oby.22554>.
 47. Stewart JM. Chronic fatigue syndrome: comments on deconditioning, blood volume and resulting cardiac function. *Clin Sci (Lond).* 2010;118:121–3. <https://doi.org/10.1042/CS20090342>.
 48. Simple study: impact of saline infusion on ME/CFS POTS and long covid. Available from: <https://www.omfcana.org/simple-study-impact-of-saline-infusion-on-me-cfs-pots-and-long-covid/> (Accessed March 11, 2025).
 49. Jason LA, Ravichandran S, Katz BZ, Natelson BH, Bonilla HF. Establishing a consensus on ME/CFS exclusionary illnesses. *Fatigue Biomed Health Behav.* 2023;11:1–3. <https://doi.org/10.1080/21641846.2022.2163719>.
 50. Jurman G, et al. A comparison of MCC and CEN error measures in assessing the performance of binary classifiers. *PLoS ONE.* 2012;7:e44138. <https://doi.org/10.1371/journal.pone.0044138>.
 51. Zhang P, Jia Y, Shang Y. Research and application of XGBoost in imbalanced data. *Int J Distrib Sens Netw.* 2022;18:15501329221106935. <https://doi.org/10.1177/15501329221106935>.
 52. Gorishniy Y, Rubachev I, Khrlukov V, Babenko A. Revisiting deep learning models for tabular data. *Adv Neural Inf Process Syst.* 2021;34:18932–43. <https://doi.org/10.48550/arXiv.2106.11959>.
 53. Kokol P, Kokol M, Zagoranski S. Machine learning on small size samples: A synthetic knowledge synthesis. *Sci Prog.* 2022;105:00368504211029777. <https://doi.org/10.1177/00368504211029777>.
 54. Zhang H, et al. Autonomic nerve and its modulation approaches for heart failure. *Brain Heart.* 2023;1(2):0913. <https://doi.org/10.36922/bh.0913>.
 55. Chen J, Wang Y, Cui Y, Wang H, Polat K, Alenezi F. EEG-based multi-band functional connectivity using corrected amplitude envelope correlation for identifying unfavorable driving States. *Comput Methods Biomech Biomed Engin.* 2025;1–13. <https://doi.org/10.1080/10255842.2025.2488502>.
 56. Liu X, Tan J, Long S. Multi-axis fatigue load spectrum editing for automotive components using generalized S-transform. *Int J Fatigue.* 2024;188:108503.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.