Supplementary attachment

Title

The Influence of Biological Rhythms on the Initial Onset of Status Epilepticus in Critically Ill Inpatients and the Study of Its Predictive Model

Authors

Nan Cheng, et al.

Supplementary Results

Supplementary Results

In this study, by integrating various rhythmic indicators and clinical factors, and employing machine learning algorithms, we adeptly developed a model to predict the onset of Status Epilepticus (SE) in critically ill patients. Compared to the currently popular Status Epilepticus Severity Score (STESS) and the Epidemiology-based Mortality score in Status Epilepticus (EMSE), our model demonstrated higher predictive accuracy and operability. Particularly, through exhaustive analysis of circadian rhythm parameters like temperature rhythm and arterial oxygen saturation rhythm, our model ingeniously identified high-risk patients, providing a valuable time window for clinical intervention.

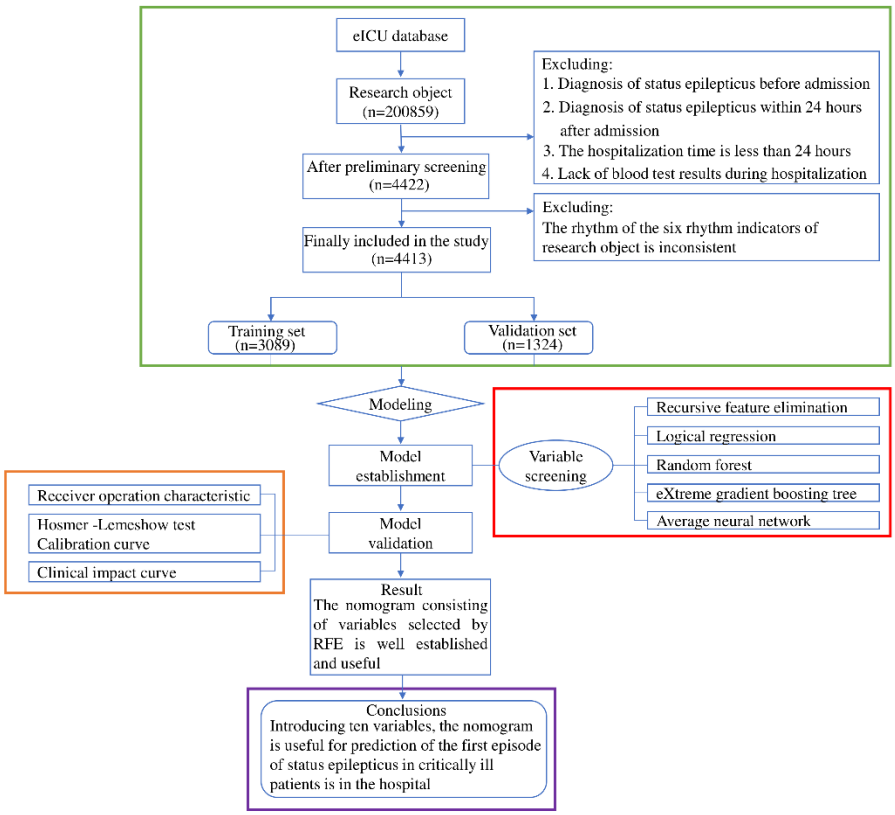
Furthermore, in comparison to traditional predictive models based on clinical variables such as STESS, EMSE, and END-IT [1][3], our model, which integrates rhythmic indicators, provides new insights into the biological mechanisms underlying SE. We observed that our model surpassed these models across a range of performance metrics, especially in terms of AUC, sensitivity, and specificity. Our model exhibited an AUC of 0.882 on the validation set, significantly higher than STESS (0.79), EMSE (0.81), and END-IT (0.83). The sensitivity of our model was 0.900, exceeding STESS (0.76), EMSE (0.78), and END-IT (0.82), while the specificity was 0.794, slightly higher than STESS (0.72), EMSE (0.74), and END-IT (0.82). These findings underscore the high accuracy and expedience of our model in identifying high-risk patients and excluding low-risk cohorts.

Prior to the analysis of circadian rhythm consistency, we conducted a detailed comparison of vital signs data between two groups of patients – those with Status Epilepticus (SE) and those without SE (non-SE). By observing the 24-hour variations in indicators such as body temperature, heart rate, respiratory rate, oxygen saturation (SaO2), systolic and diastolic blood pressure, we found notable differences in the amplitude and peak timing of certain indicators between the SE and non-SE groups. To further understand these differences in the context of circadian rhythms, we employed the cosine similarity method to quantitatively assess the diurnal rhythm of vital signs between the two groups. The cosine similarity between the two groups was close to 1, indicating that despite differences in individual indicators, their overall diurnal rhythm patterns were very similar, as detailed in Supplementary Material 2. This finding is significant for understanding the impact of Status Epilepticus on patients' daily biological circadian rhythms.

A logistic regression model was developed to assess the relationship between diurnal rhythm parameters and the incidence of SE, as shown in Figure S6. After adjusting for confounders like age and sex, it was found that the temperature mesor and peak time were negatively correlated with the incidence of SE ($P<0.05$), while the SaO2 mesor and heart rate were positively correlated ($P<0.05$). These findings suggest that the stability and balance of these vital signs' diurnal rhythms are protective factors against SE. For instance, the Odds ratio for heart rate mesor was 1.015, with a p-value of 0.028. An Odds ratio slightly greater than 1 indicates a positive correlation between heart rate mesor and SE incidence, albeit a small effect. For every unit increase in heart rate mesor,

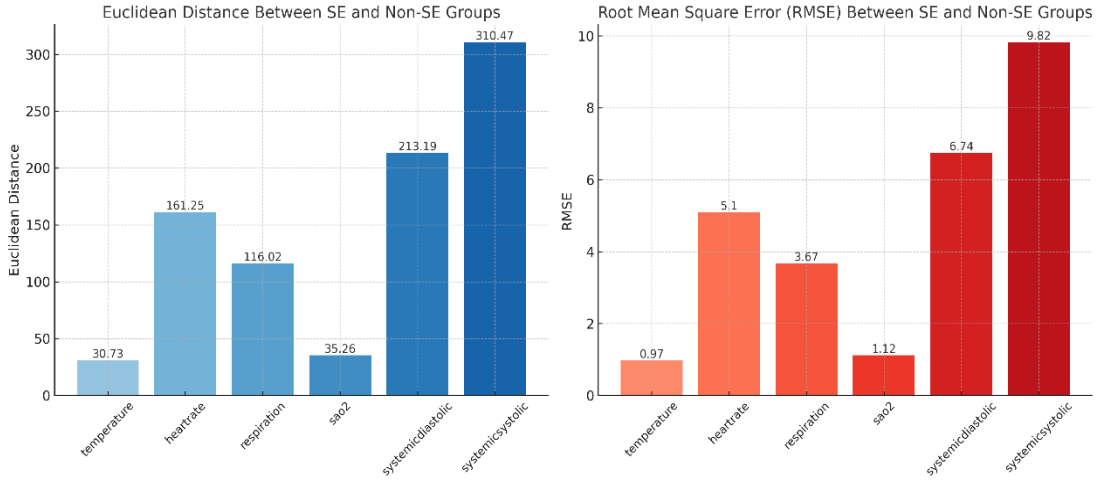
the incidence of SE increased by approximately 1.5% (95% CI: 1.002 ~ 1.029).

Figure S1. Flowchart Illustrating the Development Process for the Prediction of First Onset of Status Epilepticus in Critically Ill Patients



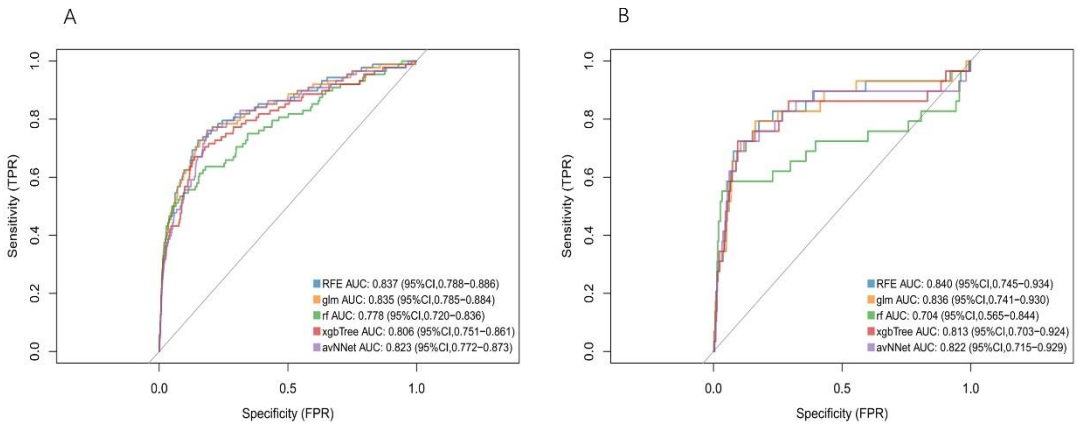
This flow chart delineates the evaluative steps and findings of a study assessing a method to predict status epilepticus (SE) risk in epilepsy patients. It segments into three categories: methodology, results, and conclusion. Different colors signify distinct information types: blue for methodology, green for results, and red for conclusion. Varied shapes represent outcome types: rectangles for intermediate outcomes, diamonds for final outcomes, and ovals for comments.

Figure S2. Comparative Analysis of Diurnal Rhythmicity in Vital Signs Between Status Epilepticus and Non-Status Epilepticus Patients



The graph illustrates the Euclidean Distance and Root Mean Square Error (RMSE) between SE (Status Epilepticus) and non-SE patient groups across six vital sign parameters. The Euclidean Distance reflects the overall disparity in the diurnal patterns, while the RMSE provides a measure of the average magnitude of variation between the groups. These metrics are derived from the cosine transformation of the vital signs' diurnal rhythms, indicating potential differences in physiological stability and rhythmicity. The left panel shows the Euclidean Distance, and the right panel shows the RMSE for temperature, heart rate, respiration, blood oxygen saturation (SaO2), diastolic and systolic blood pressure. Higher values suggest greater dissimilarity in the rhythmic patterns between the two patient groups.

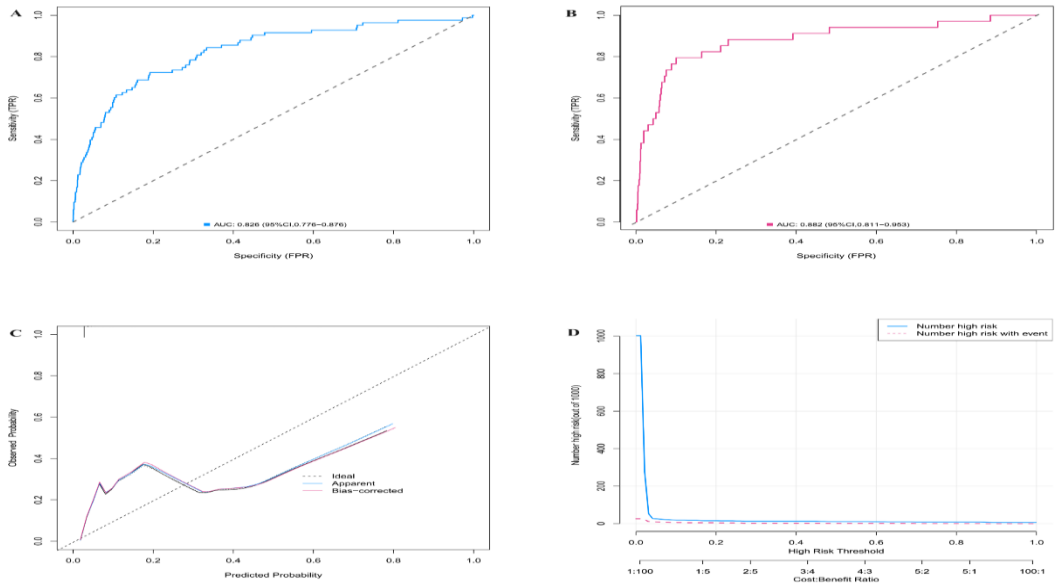
Figure S3. ROC curve compares the ability of different methods to predict status epilepticus



Displayed is a Receiver Operating Characteristic (ROC) curve for evaluating various methods in predicting status epilepticus. A) represents the results from the independent test set, while B) reflects those of the validation set. The horizontal axis denotes specificity (FPR), and the vertical axis

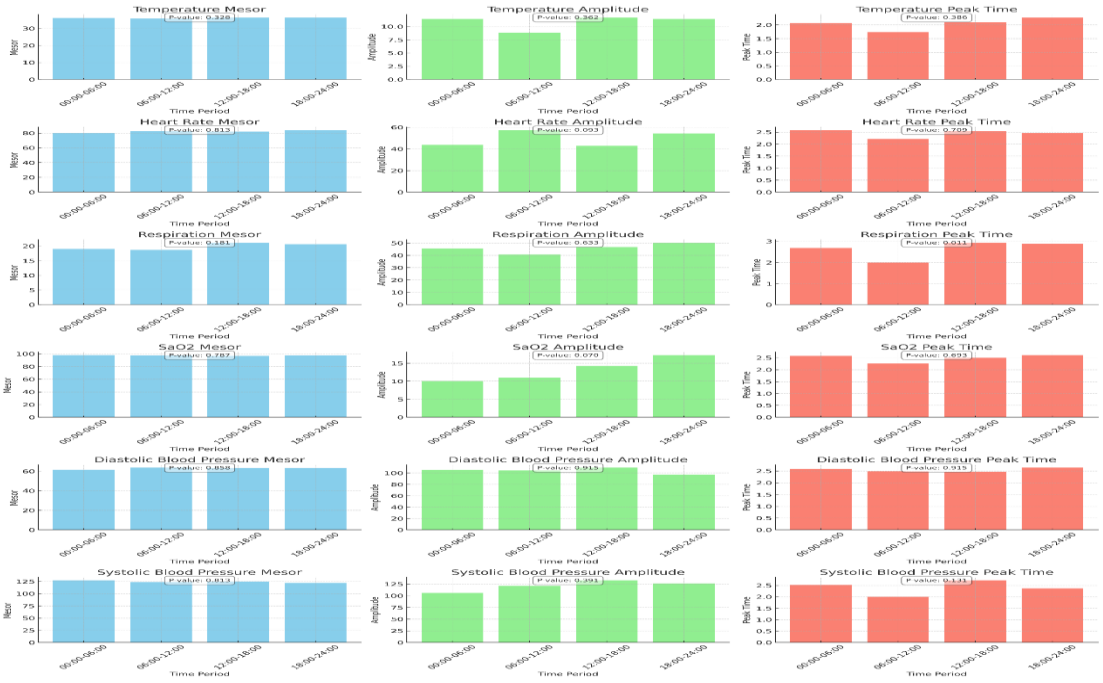
signifies sensitivity (TPR). Different colors and styles depict distinct prediction methods for ease of differentiation. The Area Under the Curve (AUC) value is annotated in the legend, with a value closer to 1 indicating a more precise prediction method.

Figure S4. Performance and internal verification analysis of the model



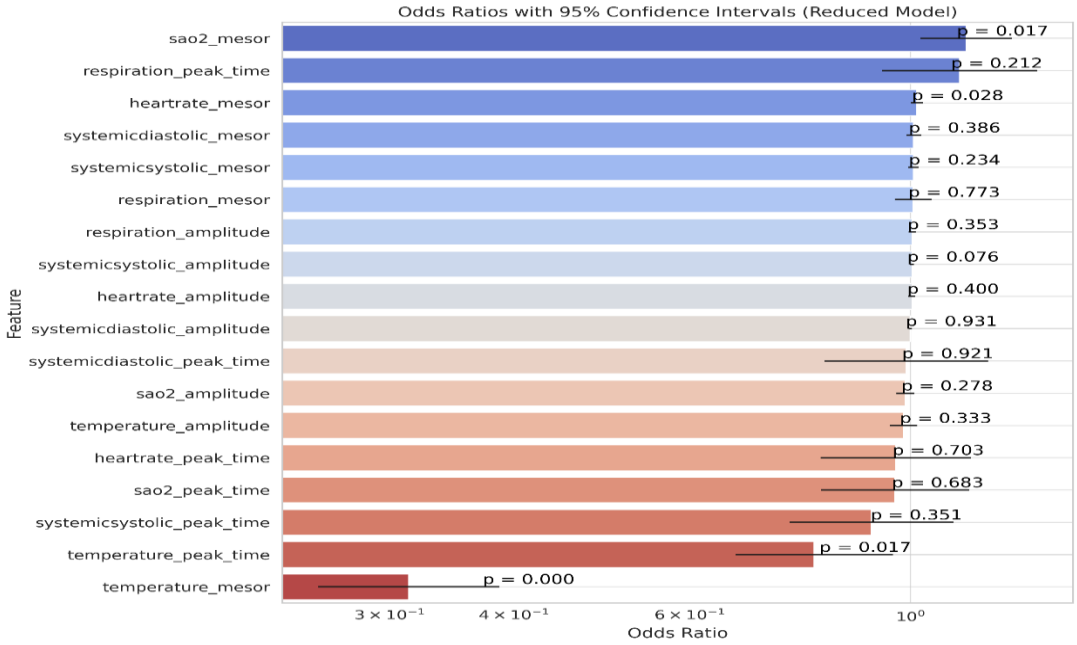
The model's performance is demonstrated on the test set. C) depicts the calibration curve of the model across 1000 bootstrap resamples. D) presents the CIC diagram of the model under various thresholds. The horizontal axis denotes specificity (FPR), and the vertical axis signifies sensitivity (TPR). The curve, delineated in blue and pink, facilitates easy distinction. The Area Under the Curve (AUC) value is annotated in the legend, with proximity to 1 indicating enhanced prediction accuracy. The calibration curve illustrates the congruence between predicted probabilities and observed probabilities, ideally aligning closely with a 45-degree diagonal. The CIC diagram underscores the model's utility in decision support, where the region above the diagram signifies superior performance over random guessing, and the region below denotes inferior performance.

Figure S5. Differences of biorhythm parameters in patients with SE at different onset times



Circadian Variation in Physiological Parameters Among Status Epilepticus Patients. This figure displays an in-depth comparison of six biorhythm parameters (Temperature, Heart Rate, Respiration, SaO2, Diastolic Blood Pressure, Systolic Blood Pressure) in patients with Status Epilepticus (SE) across four key onset time periods (00:00-06:00, 06:00-12:00, 13:00-18:00, 18:00-24:00). For each parameter, the mean (mesor), amplitude, and peak time are presented. The P-values, derived for each parameter across different time periods, provide a measure of statistical significance for the observed variations, with a threshold of 0.05 indicating significant differences. The bar graphs in each subplot facilitate a comparative analysis of the circadian patterns in these clinical parameters among patients with SE. The visualization underscores discernible circadian rhythms and highlights statistically significant differences in certain parameters as indicated by the P-values. This comprehensive analysis is crucial in understanding the temporal dynamics of physiological changes in SE patients, offering valuable insights for tailored therapeutic approaches and optimizing clinical management based on the onset time of SE.

Figure S6. Analysis of Influential Factors on Diurnal Rhythms of Vital Signs in Patients with and without Conversion to Status Epilepticus



This figure illustrates the associations between various physiological rhythm indicators and status epilepticus using logistic regression analysis. Each forest plot represents the Odds Ratio (OR) and its 95% Confidence Interval (CI) for corresponding physiological rhythm parameters (body temperature, respiration, blood oxygen saturation (SaO2), systemic systolic, and systemic diastolic). A dashed line indicates the absence of effect or difference (i.e., OR=1). An OR greater than 1, with a CI not crossing 1, suggests a positive correlation with status epilepticus. Conversely, an OR less than 1, with a CI not crossing 1, implies a negative correlation. If the CI includes 1, it indicates that the variable has no statistically significant association with status epilepticus. Specifically, for each parameter, we observe that the OR CIs of most parameters include 1, suggesting no significant statistical association with the occurrence of status epilepticus. However, parameters like the mesor (24-hour average) of body temperature and SaO2 show statistical significance (P-value <0.05), indicating that rhythmic variations in these parameters may be associated with the occurrence of status epilepticus.

Supplementary Table 1 Percentage of missing data in the variables of the included cases

Variables	Research data set (n = 4413)
Baseline characteristics	
Age	2.10%
Sex	0.00%
Laboratory	
RBC	0.28%

WBC	0.83%
Monos	9.90%
MCH	0.26%
Hgb	0.01%
Hct	0.01%
Lymphs	8.54%
Platelets	0.15%
PT	15.50%
Glucose	1.12%
Sodium	2.00%
Potassium	1.01%
Bicarbonate	2.08%
Comorbidities	
Pulmonary infections	0.00%
Hypertension	0.00%
Hyperlipidemia	0.00%
Hyperthyroidism	0.00%
Hypothyroidism	0.00%
Fracture of skull	0.00%
Intracranial injury	0.00%
Encephalitis	0.00%
Meningitis	0.00%
Diabetes mellitus	0.00%
Stroke	0.00%
Coronary artery disease	0.00%
Head infections	0.00%
Pulmonary embolism	0.00%
Abscess	0.00%
Anxiety	0.00%
Atrial fibrillation	0.00%
Bipolar disorder	0.00%
Brain edema	0.00%
Cirrhosis	0.00%
CNS infection	0.00%
Coma	0.00%
Contusion	0.00%
Depression	0.00%
Fever	0.00%
Fracture	0.00%
Heart failure	0.00%
Hypocalcemia	0.00%
Hypoglycemia	0.00%
Hyponatremia	0.00%
Infections	0.00%

Ischemia	0.00%
Myocardial infarction	0.00%
Trauma	0.00%
Vasculitis	0.00%
COPD	0.00%
Scoring system	
GCS	3.16%
Rhythm index	
Temperature	0.06%
Heartrate	0.06%
Respiration	0.05%
SaO2	0.01%
Systemic diastolic	0.05%
Systemic systolic	0.05%

In this data analysis, the approach taken involved a customized strategy for managing missing values, distinguishing between numerical and categorical variables. For numerical variables, multiple imputation was applied, a technique that harnessed the inherent data structure and correlations to impute missing values. To ensure robustness, five iterations of multiple imputation were carried out, resulting in five complete datasets. This meticulous process accounted for the uncertainty associated with imputation and bolstered the accuracy of the results. Conversely, for categorical variables, mode imputation was employed, with missing values being replaced by the mode of the respective category. This approach upheld data integrity while addressing categorical attributes. By systematically addressing missing data using these methods, the completeness and reliability of the dataset were ensured, a crucial step in minimizing potential bias and enhancing the quality of the analyses.

Supplementary Table 2 Comparative Analysis of Variable Importance Across Machine Learning Models

RFE		LR		RF		Xg Boost		ANN	
variables		variables		variables		variables		variables	
temperature mesor	29.31	temperature mesor	22.51	temperature mesor	25.32	temperature mesor	5.18	temperature mesor	25.27
Stroke	22.88	GCS	19.93	SaO2 mesor	17.84	age	2.65	GCS	13.14
Age	22.50	age	16.92	PT	16.15	stroke	2.39	coma	9.18
hypoglycemia	20.87	hypoglycemia	15.83	heartrate mesor	16.14	GCS	1.84	stroke	8.15
Potassium	19.93	potassium	15.25	bicarbonate	15.87	potassium	1.62	SaO2 mesor	7.11
GCS	17.98	stroke	15.22	age	15.55	SaO2 mesor	1.23	hypoglycemia	6.93
myocardial infarction	17.93	heartrate mesor	14.86	glucose	13.70	bicarbonate	1.15	age	6.26
hyponatremia	17.80	hyponatremia	13.45	MCH	13.20	hypoglycemia	1.00	meningitis	5.36
temperature peak time	16.92	depression	13.11	potassium	12.80	coma	0.90	depression	4.67
SaO2 mesor	15.83	heartrate mesor	12.74	sodium	11.33	WBC	0.89	hyponatremia	4.13

The table enumerates the top ten variables from each machine learning algorithm—RFE, LR, RF, Xg Boot, and ANN—based on their ranked importance for the prediction of status epilepticus. These

variables, include clinical signs, patient demographics, and laboratory values. These variables were selected for their significant roles as indicated by preliminary machine learning analysis and were incorporated into the development of advanced models aimed at predicting the onset of status epilepticus.

Supplementary Table 3 Zero-Amplitude Test Results for Key Physiological Indicators

Parameter	Mean Amplitude	Standard Deviation	t-Statistic	p-Value
Body Temperature Amplitude	11.57	6.15	124.95	0.00
Heart Rate Amplitude	42.38	22.99	122.42	0.00
Respiration Rate Amplitude	41.42	23.08	119.19	0.00
SaO2 Amplitude	14.73	12.26	79.83	0.00
Diastolic Blood Pressure Amplitude	95.64	70.38	90.26	0.00
Systolic Blood Pressure Amplitude	110.36	61.34	119.49	0.00

This table details the results of zero-amplitude tests conducted on the circadian amplitudes of six physiological indicators. Statistical measures provided include mean amplitude, standard deviation, t-statistic, and p-value. The zero-amplitude test assesses whether the amplitude significantly deviates from zero, using a one-sample t-test. A p-value close to zero confirms significant rhythmicity, indicating that the observed fluctuations in the data are not due to random variation but represent true biological rhythms.

Supplementary Table 4 Hyperparameter Settings for Evaluated Classifier Models

Classifier models	**Hyperparameters**
Logistic Regression	C=0.4, multi class=multinomial, random state =8, solver=saga
Random Forests	Random state=8, max depth=60, max features=sqrt, min samples split=5, min samples leaf=4, n estimators =400
Xg Boost	Random state=8, max depth=10, max features=sqrt, min samples split=50, n estimators=800, learning rate=0.5, subsample=0.5
Artificial Neural Network	solver=lbfgs, learning rate=adaptive, activation=identity, alpha=0.0001, batch size=auto, hidden layer sizes=7, learning rate init=0.001, max iter=500
Recursive Feature Elimination	N/A

This table summarizes the optimized hyperparameters for the classifier models utilized in the study. For each model, we tuned parameters to maximize the validation accuracy. The "N/A" for RFE indicates that Recursive Feature Elimination is used as a feature selection technique prior to the application of classifier models and does not have hyperparameters in the conventional sense.

Supplementary References

[1] Leitinger, M., Höller, Y., Kalss, G. et al. The Status Epilepticus Severity Score (STESS): A Tool to Orient Early Treatment Strategy. J Neurol 257, 1561–1566 (2010).

[2] Rossetti, A. O., Logroscino, G., Bromfield, E. B. A Clinical Score for Prognosis of Status Epilepticus in Adults. Neurology 66, 1736–1738 (2006).

[3] Sutter, R., Kaplan, P. W., Marsch, S. et al. Anesthetic Drugs in Status Epilepticus: Risk or Rescue? A 6-Year Cohort Study. Neurology 82, 656–664 (2014).