Supplementary Methods

Diagnostic criteria

1. **Smoking history**: Smoking at least 1 cigarette every day for a consecutive year.

2. **Epilepsy**: Patients clinically present with seizures and exhibit seizure-like discharges on EEG recordings during or between seizures. Diagnosis of epilepsy is made if seizures are observed, or if there is a history of taking antiepileptic drugs in the past[1]。

3. **High blood pressure**: In the absence of antihypertensive medication, if the systolic blood pressure is ≥ 140mmHg and/or diastolic blood pressure is ≥ 90mmHg in three separate measurements on different days, the patient is diagnosed with hypertension. Even if the blood pressure is below 140/90mmHg and the patient has a history of hypertension and is currently taking antihypertensive medication, they should still be diagnosed with hypertension[2].

4. **Diabetes**: According to the diagnostic criteria proposed by the World Health Organization Expert Committee on Diabetes in 1999: fasting blood glucose ≥7.0mmol/L (confirm on another day if no typical diabetes symptoms are present); or typical diabetes symptoms with random blood glucose or venous plasma glucose ≥11.1mmol/L; or 2-hour blood glucose ≥11.1mmol/L in an oral glucose tolerance test (confirm on another day if no typical diabetes symptoms are present); or glycosylated hemoglobin ≥6.5%[3]。

5. **Coronary artery disease**: Coronary angiography shows at least one coronary artery stenosis ≥50%, or a history of coronary artery bypass grafting[4].

6. **Pulmonary embolism**: According to the "Diagnosis and Treatment Guidelines for Pulmonary Thromboembolism," a CT pulmonary angiography showing filling defects in the pulmonary artery is used to diagnose pulmonary embolism[5].

7. **Pulmonary contusion**: With a clear history of chest trauma, the patient presents with symptoms such as chest pain, shortness of breath, hemoptysis, and even respiratory distress. Physical examination reveals crackles in the lungs and decreased breath sounds. Diagnosis of pulmonary contusion is confirmed through X-ray or CT imaging[6].

8. **Pneumonia**: According to the "Guidelines for the Diagnosis and Treatment of Community-Acquired Pneumonia in Chinese Adults," chest radiography showing newly developed patchy infiltrates, lobar or segmental consolidation, ground-glass opacity, or interstitial

changes, along with any of the following features: (1) fever; (2) new onset cough, sputum production, or worsening of existing respiratory symptoms; (3) signs of lung consolidation and/or auscultation of crackles; (4) peripheral blood leukocytes $>10\times10^9$/L or $<4\times10^9$/L [7]

9. **Consciousness disorder**: The Glasgow Coma Scale was used in this study to assess the patients' level of consciousness. GCS consists of three aspects: eye opening response (1-4 points), verbal response (1-5 points), and motor response (1-6 points), with a total score of 15 points. A score of ≤8 points indicates severe consciousness disorder, 9-12 points indicates moderate consciousness disorder, 13-14 points indicates mild consciousness disorder, and 15 points indicates clear consciousness. A higher score indicates a milder consciousness disorder[8].

10. **Swallowing disorder**: This study used the Wakita water swallowing test to evaluate patients' swallowing function. The Wakita water swallowing test is graded into 5 levels: Level I (Excellent): can swallow water in one attempt; Level II (Good): can swallow water without coughing at least 2 times; Level III (Fair): can swallow water in one attempt but coughs; Level IV (Acceptable): can swallow water at least 2 times but coughs; Level V (Poor): frequently coughs and cannot swallow all the water[9].

Categorical variable label encoding

During the correlation analysis, categorical variables were converted into dummy variables for use in machine learning. For binary classification variables, 1 represented "yes" and 0 represented "no". For multiclass variables, different categories were represented by Arabic numerals from 1 to the total number of categories.

Variables	Category	Dummy variables
Meal gender	yes	1
	no	0
Wet Rales	yes	1
	no	0
Smoking	yes	1
	no	0
Consciousness Disorder	yes	1
	no	0
Dysphagia	yes	1
	no	0
Epilepsy	yes	1
	no	0

Viscous sputum	yes	1
	no	0
Upper Respiratory Tract Structural Abnormalities (URSA)	yes	1
	no	0
Hypertension	yes	1
	no	0
Coronary Heart Disease	yes	1
	no	0
Diabetes Mellitus	yes	1
	no	0
Pulmonary Infection	yes	1
	no	0
Pulmonary Embolism	yes	1
	no	0
Pulmonary Contusion	yes	1
	no	0
Mental Disorders	yes	1
	no	0
Disease Duration :	DD ≤ 30 days	1
	30 days < DD ≤ 60 days	2
	60 days < DD ≤ 90 days	3
	DD > 90 days	4
Primary Disease :	Cerebral Trauma	1
	Cerebral Hemorrhage	2
	Cerebral Infarction	3
	Subarachnoid Hemorrhage	4
	Ischemic Hypoxic Brain Disease	5
	Central Nervous System Infectious Diseases	6
	Spinal Cord Injury	7
	Other Diseases	8
Limb Paralysis :	Quadriplegia	1
	Hemiplegia	2
	Paraplegia	3
Sputum Culture Pathogens :	Pseudomonas aeruginosa	1
	Klebsiella pneumoniae pneumoniae subtype	2

	Acinetobacter baumannii	3
	Escherichia coli	4
	Serratia marcescens	5
	Other Pathogens	6
	Two or more pathogens	7
	None	8

Table S1. Categorical variables and dummy variables.

Hyperparameter optimization and parameter tuning

In this study, a total of eight models were constructed, including K-Nearest Neighbors (K-NN), Support Vector Machine (SVM) [10], Gaussian Naive Bayes (GNB), Decision Tree (DT), Random Forest (RF) [11], Extra Trees Classifier (ET), Gradient Boosting (GB), and Logistic Regression (LR), to compare the predictive performance of each model. Modeling parameters were selected using cross-validation and hyperparameter optimization. The screened features collected in this study were used as feature variables, with extubation outcomes as the target variable. In the dataset of 501 patients, 80% of the data was used as the training set, while 20% was used as the test set. Accuracy was calculated using the relevant formulas [10]. The raw data processing was performed using the Numpy package in Python 2020, while dataset splitting and model construction were carried out using the Sklearn package[12]. The details of hyperparameter optimization were provided in Supplementary Material, Table S2.

Feature	Range	Optimal value
KNN		
k_values	1-40	24
SVM		
Kernel	linear', 'poly', 'rbf', 'sigmoid'	linear'
max_iter	10-1000	500
Regularization parameter (C)	0.01-100	0.1
probability	Ture-False	Ture
Gaussian Naive Bayes		
var_smoothing	10^{-1}-10^{-9}	0.01
Decision Tree		
max_depth	3-20	3
min_samples_leaf	1-50	3
criterion	gini', 'entropy'	entropy'
min_samples_split	2-50	5
splitter	best', 'random'	best'
Random Forest		

max_depth	3-20	5
max_features	sqrt', 'log2',	sqrt'
min_samples_leaf	1-20	2
n_estimators	100-5000	700
random_state	1-20	3
max_leaf_nodes	10-100	30
min_samples_split	2-20	10
Extra Tree Classifier		
max_depth	3-20	10
max_features	sqrt', 'log2',	sqrt'
min_samples_leaf	1-20	2
n_estimators	100-5000	1000
random_state	1-20	3
max_leaf_nodes	10-100	30
min_samples_split	2-20	10
Gradient Boosting		
learning_rate	0.001-0.1	0.01
loss	deviance', 'exponential'	exponential'
max_depth	3-20	9
max_leaf_nodes	5-50	5
min_samples_leaf	1-50	25
min_samples_split	2-20	5
n_estimators	100-5000	500
subsample	0.1-1	1
AdaBoosting		
n_estimators	100-5000	200
learning_rate	0.001-0.1	0.01
Logistic Regression		
max_iter	0.001-1000	100
nverse of regularization strength (C)	0.001-200	100
solver	liblinear', 'newton-cg', 'lbfgs', 'sag', 'saga'	liblinear'
penalty	L1', 'L2'	L2'

Table S2. **Range of hyperparameter optimization values.**

Model building methods and code

The raw data processing was performed via the *Numpy* package, while dataset splitting and model construction were carried out via the *Sklearn* package in *Python 3.12.4*[12]. The study constructed a total of 8 independent models, and the specific methods and construction process of these models are as follows:

1. Logistic Regression

Logistic regression is a regression method widely used for modeling binary outcomes. It applies the logit function to the linear combination of independent variables to map the predicted results to the [0, 1] interval, estimating the probability of event occurrence[13,14]. Compared to traditional linear regression, logistic regression has significant advantages in dealing with non-continuous dependent variables (such as presence or absence of disease, survival or not). The method's statistical inference is based on maximum likelihood estimation, with a key focus on ensuring independence, avoiding multicollinearity, and maintaining a linear relationship for continuous variables on a log-odds scale[15]. In numerous clinical and epidemiological studies, logistic regression is widely used to investigate disease risk factors, predict clinical outcomes, and assess intervention effects.

Code:

```
lr = LogisticRegression(max_iter=100,C = 100,solver='liblinear',penalty='l2');
```

2. K-Nearest Neighbors

The K-nearest neighbors (K-NN) algorithm has been widely recognized as a non-parametric classification and regression method based on instance learning since the 1990s[16]. The core idea is to calculate the distance between the sample to be classified and all samples in the training set, selecting the K nearest neighbors, and making predictions based on the neighbors' categories. The standard K-NN algorithm was first applied in academia, its main advantage being its simplicity in implementation, allowing for direct prediction using distance calculations without the need for a training phase[16].

Code:

```
knn = KNeighborsClassifier(n_neighbors=24);
```

3. Support Vector Machine

Support Vector Machine (SVM) is based on the principle of "structural risk minimization" in statistical learning theory, aiming to find a separating hyperplane with the maximum margin to effectively enhance the model's generalization ability[17]. In the basic linear SVM model, the algorithm transforms the training data into a high-dimensional space and determines the optimal hyperplane by solving a convex quadratic programming problem. Kernel function technique allows SVM to create non-linear decision boundaries for non-linearly separable problems[18].Commonly used kernel functions include radial basis function (RBF), polynomial kernel, and sigmoid kernel, among others. The choice of different kernel functions significantly impacts the performance of the model. In addition, SVM introduces slack variables during model training to handle noise and outliers in the data, allowing it to maintain high accuracy while preventing overfitting[19].

Code:

```
svc = SVC(kernel='linear', max_iter=500, C=0.1, probability=True);
```

4. Gaussian Naive Bayes

The Gaussian Naive Bayes (GNB) classifier is based on Bayes' theorem, assuming all features are conditionally independent and continuous variables follow a Gaussian distribution. The essence of this approach is to compute the posterior probability and make classification decisions by maximizing the probability. In the field of disease prediction and classification, GNB is widely used for the early diagnosis of chronic diseases, cancer, and internal medicine conditions. In real medical data, it is often difficult to fully satisfy, which may affect the accuracy of the model's predictions. Therefore, scholars improve feature selection and data preprocessing techniques to minimize the impact of correlations, or use ensemble methods to integrate results from multiple classifiers[20].

Code:

```
bays = GaussianNB(var_smoothing=0.01);
```

5. Decision Tree

Decision tree is a model with a tree-like structure, consisting of nodes that split data based on certain criteria until reaching a predetermined termination condition. The construction of a decision tree mainly involves two steps: attribute selection and tree pruning. During the attribute selection process, commonly used quantification metrics such as information gain, entropy, and Gini index are used to determine the optimal splitting attribute, thereby maximizing the inter-class differences at each branch[21]. To avoid overfitting during the construction process, pruning techniques are usually employed to simplify the complexity of the tree[21].

Code:

```
tree = DecisionTreeClassifier( max_depth=3, min_samples_leaf=3, criterion='entropy',
min_samples_split=5, splitter='best');
```

6. Random Forest

Random forest (RF), as a non-parametric, ensemble machine learning algorithm, has shown many significant advantages in medical data mining and risk prediction. Firstly, RF's random sampling and majority voting mechanism make it robust in handling noisy and high-dimensional data. Additionally, its built-in variable importance assessment provides clinicians with an intuitive feature selection tool, effectively reducing the feature space and revealing underlying biological mechanisms[22]. Recent research has successfully applied random forests to various fields such as neuroimaging, cardiovascular diseases, diabetes prediction, and monitoring of drug adverse reactions, demonstrating high practical value of RF in all these areas[23,24].

Code:

```
rf = RandomForestClassifier(max_depth=6, max_features='sqrt', min_samples_leaf=2,
n_estimators=700, random_state=7, max_leaf_nodes=30, min_samples_split=10);
```

7. Extreme Random Tree

Extreme random tree classifier is a type of decision tree ensemble method, similar to RF, where the core idea is to vote on the predictions of multiple decision trees. However, ET achieves higher levels of randomness and generalization ability through the following key features:

- Random split points: During node splitting, ET does not select the data-driven optimal split point, but randomly selects from possible split values of features to further increase model diversity[25].

- Ensemble decision: ET uses the predictions of multiple trees for majority voting or averaging to enhance model robustness and stability[26].

Given these characteristics, ET has significant advantages in computational efficiency and model robustness compared to traditional decision trees and random forests. It also performs well in handling high-dimensional features and heterogeneous data, which is particularly important for the complexity of medical data [27].

Code:

```
extraTreeclassfier = ExtraTreesClassifier(max_depth=10, max_features='sqrt',
min_samples_leaf=2, n_estimators=1000, random_state=3, max_leaf_nodes=30,
min_samples_split=10);
```

8. Gradient Boosting

The gradient boosting algorithm works by creating a sequence of weak learners (often decision trees) and updating each model by using the residuals from the previous model as the new target, moving closer to the true value with each iteration. This approach offers powerful nonlinear fitting capabilities and robustness, making it well-suited for high-dimensional data and complex pattern recognition tasks[28]. GB methods typically involve the following steps: initializing the model, usually with a simple mean prediction. In each iteration, calculate the current model's residual and fit it with a new weak learner. Update the overall model based on the predictions of weak learners, and introduce a learning rate to prevent overfitting. Repeat this process until reaching a predetermined number of iterations or the error decreases to a certain level. To improve model performance, researchers often use techniques such as cross-validation and Bayesian optimization for hyperparameter tuning, while incorporating regularization techniques to mitigate the risk of overfitting complex models[29]. In addition, there are efficient implementations like XGBoost, LightGBM, which have significantly optimized computing speed and memory usage, making them advantageous in large-scale medical data processing[30].

Code:

```
gb = GradientBoostingClassifier(learning_rate = 0.01, loss ='exponential', max_depth = 9,
max_leaf_nodes = 5, min_samples_leaf = 25, min_samples_split = 5, n_estimators = 500,
subsample= 1);
```

9. WPV

The basic principle of WPV integration method lies in the fact that different models may exhibit complementary performance advantages when facing complex tasks. By using the predicted probabilities of each model and assigning different weights based on their accuracy on the validation set, we can mitigate the negative impact of individual model errors on overall predictions. In addition, when designing the WPV model, it is necessary to consider not only the synergistic effects between different models, but also to effectively address the potential issues of data redundancy and high dimensionality that may exist between models. The voting ensemble learning method can produce results using either hard or soft voting. Hard voting is ideal for predicting distinct class labels, whereas soft voting is more suitable for predicting continuous values.This study focused on tweets categorized into different groups, suggesting that hard voting is preferable over soft voting[31].

Code:

```
WPV_model = VotingClassifier(estimators=[('rf', rf),    ('bays', bays), ('knn', knn)],
voting='soft', weights=[1, 1, 1], verbose=1)
```

Supplementary Results

Rank	Model	Accuracy (%)	AUC	Sensitivity (%)	Specificity (%)	Youden index (%)	PPV (%)	NPV (%)	Threshold
1	**WPV**	**85.15**	**0.87**	**81.08**	**87.50**	**68.58**	**78.95**	**88.89**	**NA**
2	Extra Tree Classifier	83.17	0.87	70.27	90.63	60.90	81.25	84.06	0.51
3	Random Forest	84.15	0.85	72.97	90.62	63.60	81.82	85.29	0.50
4	K-Nearest Neighbour	82.18	0.85	70.27	89.06	59.33	78.79	83.82	0.58
5	Gradient Boosting	81.19	0.85	72.97	85.93	58.91	75.00	84.61	0.53
6	Support Vector Machine	79.21	0.83	75.67	81.25	56.93	70.00	85.24	0.43
7	Decision Tree	79.21	0.82	75.68	81.25	56.92	70.00	85.24	0.27
8	Logistic Regression	77.23	0.84	75.67	78.13	53.80	66.67	84.75	0.60
9	Gaussian Naive Bayes	73.28	0.88	86.49	65.63	52.11	59.26	89.36	0.79

Table S3. Summary of the performance of all models.

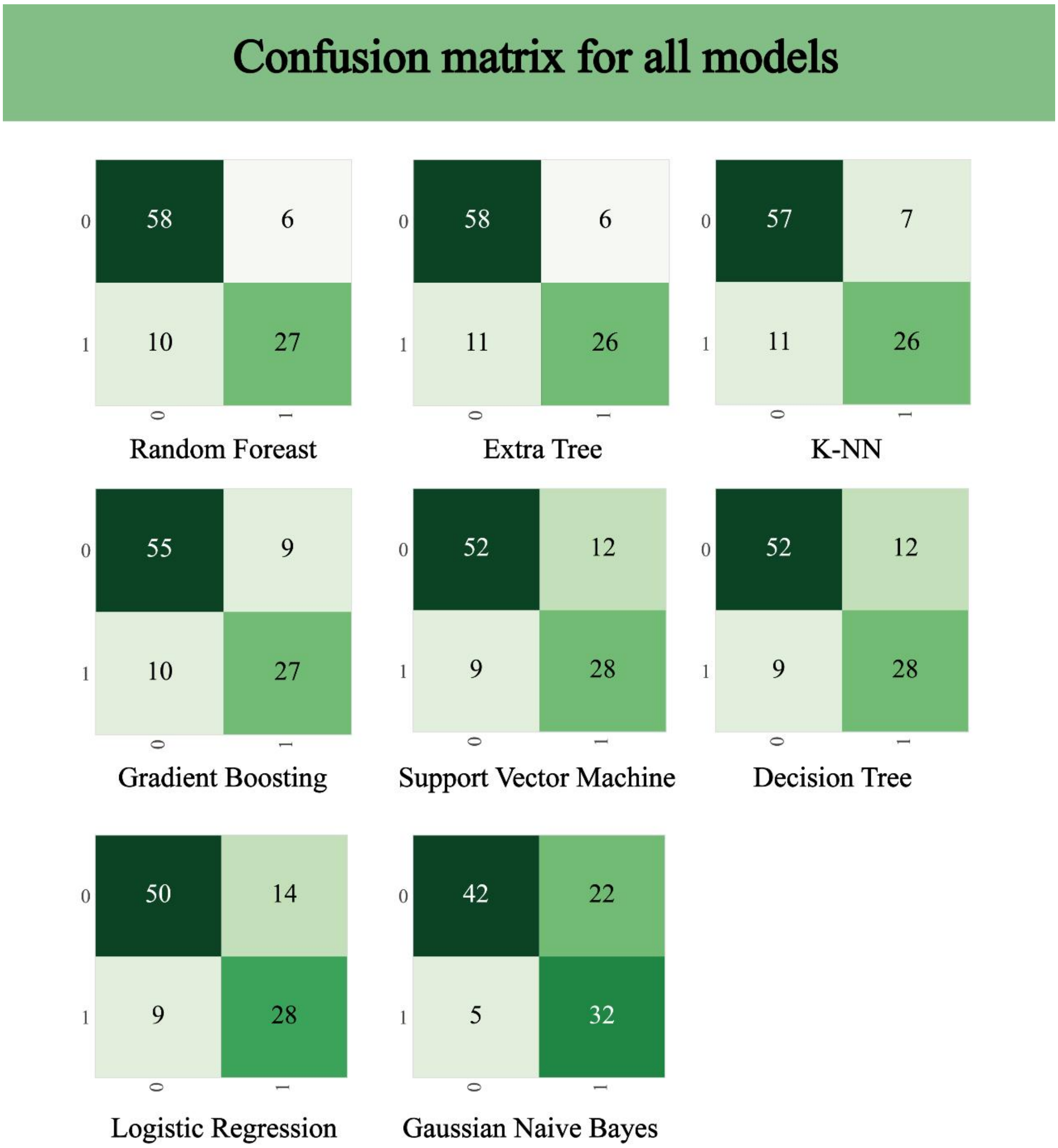

Figure S1. Confusion matrices of all models. K-Nearest Neighbors (K-NN).

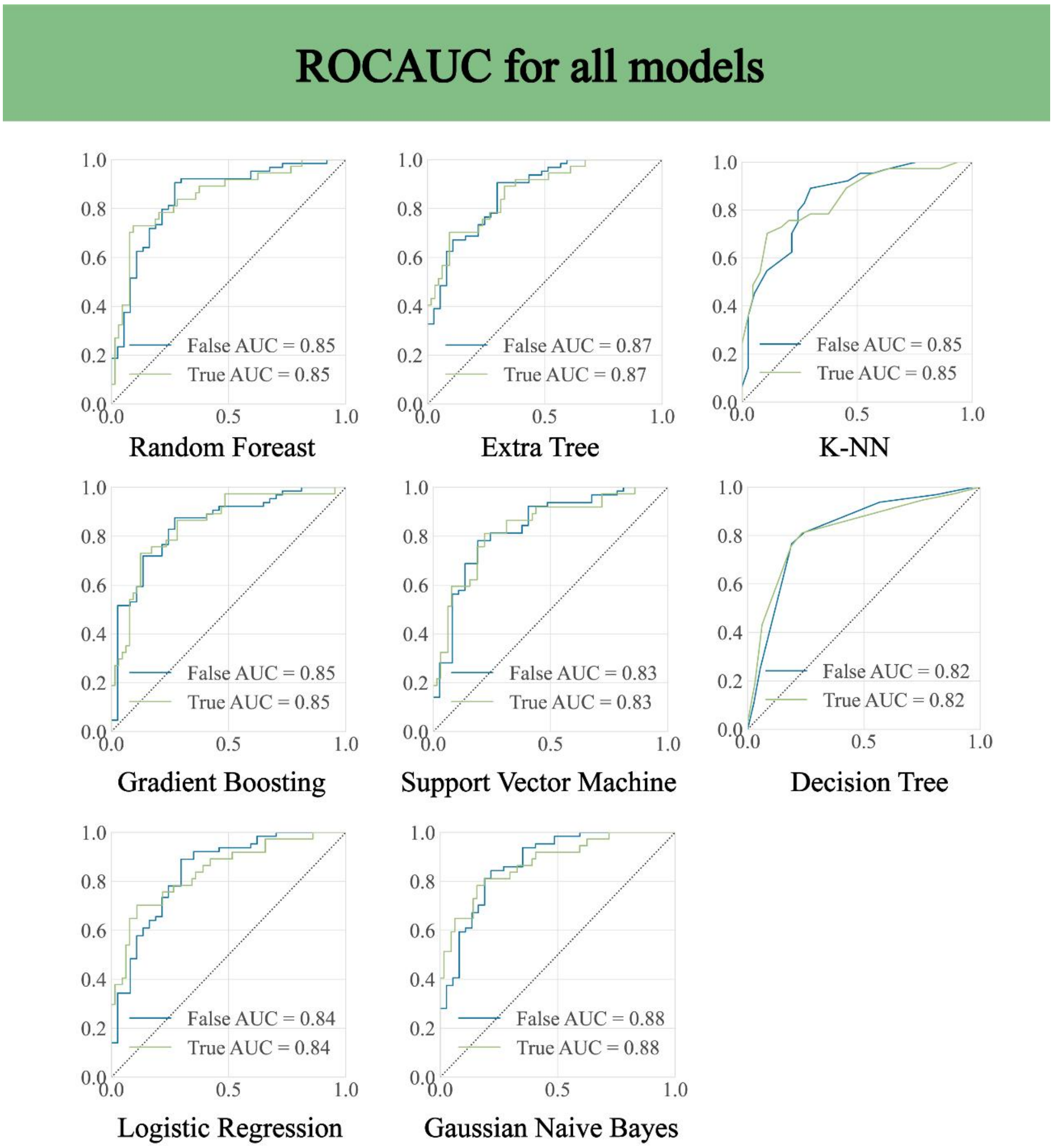

Figure S2. AUC curves of all models. Area under receiver operating characteristic curve (AUC), K-Nearest Neighbors (K-NN).

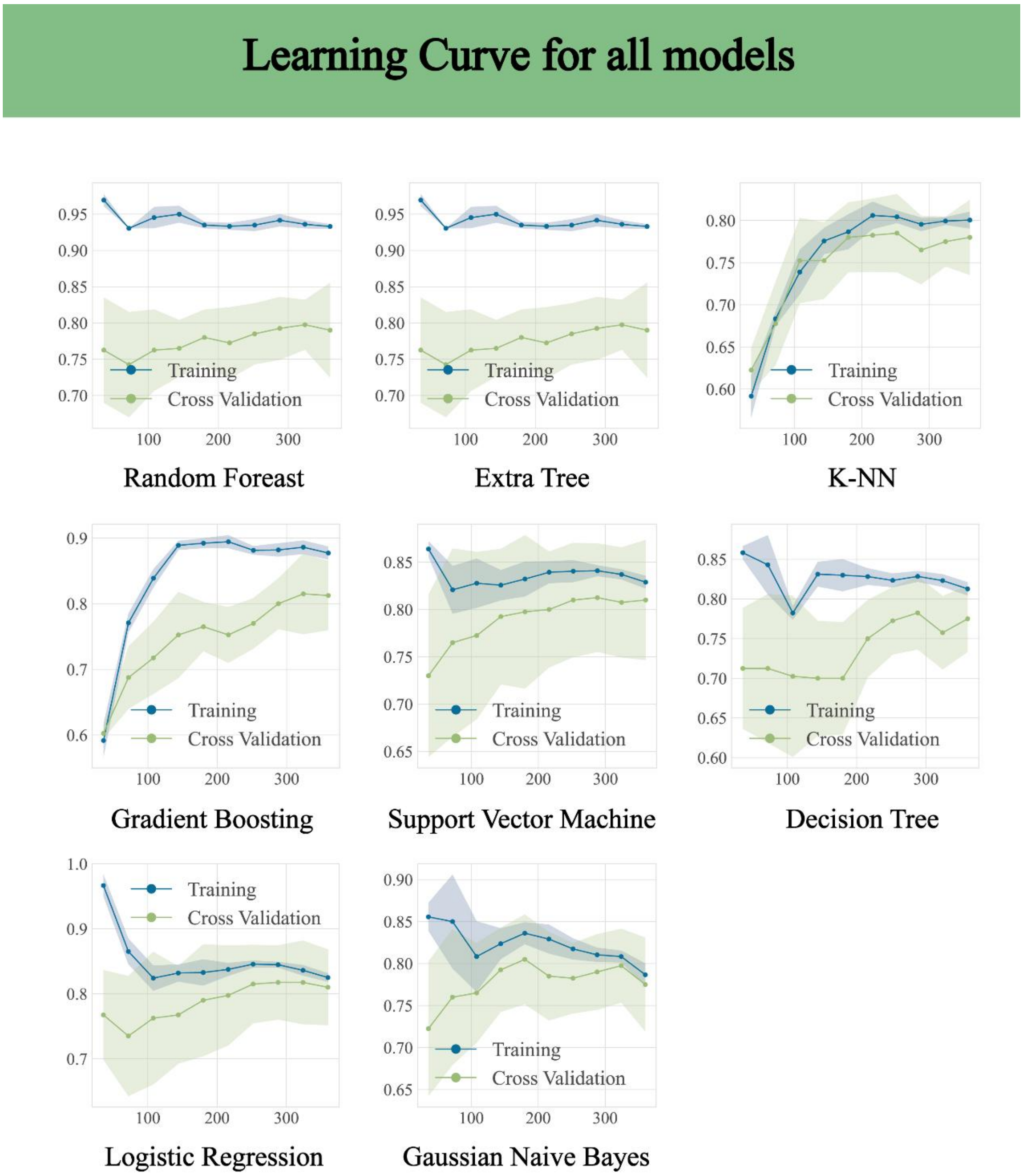

Figure S3. Learning curves of all models. K-Nearest Neighbors (K-NN).

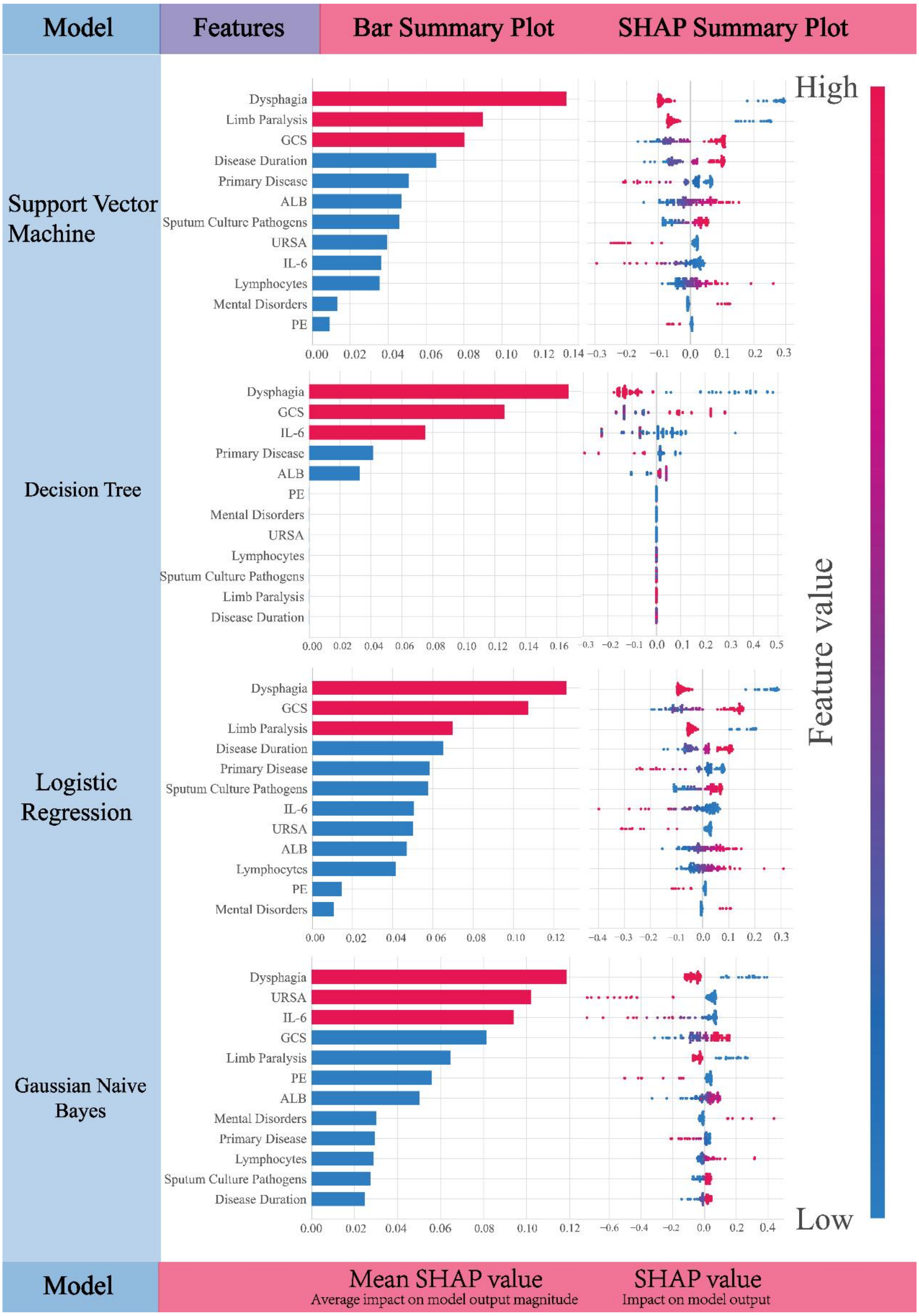

Figure S4. Bar summary plots and SHAP summary plots for the remaining four models. Shapley additive explanations (SHAP), Glasgow Coma Scale (GCS), Upper Respiratory Tract Structural Abnormalities (URSA), Interleukin 6 (IL-6), Albumin (ALB), Pulmonary Embolism (PE), K-Nearest Neighbors (K-NN).

1 Bandopadhyay, R. *et al.* Recent Developments in Diagnosis of Epilepsy: Scope of MicroRNA and Technological Advancements. *Biology* **10**, doi:10.3390/biology10111097 (2021).

2 Malta, D. C. *et al.* Prevalence of arterial hypertension according to different diagnostic criteria, National Health Survey. *Revista brasileira de epidemiologia = Brazilian journal of epidemiology* **21**, e180021, doi:10.1590/1980-549720180021.supl.1 (2018).

3 Alberti, K. G. & Zimmet, P. Z. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabetic medicine : a journal of the British Diabetic Association* **15**, 539-553, doi:10.1002/(sici)1096-9136(199807)15:7<539::Aid-dia668>3.0.Co;2-s (1998).

4 Knuuti, J. *et al.* 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. *European heart journal* **41**, 407-477, doi:10.1093/eurheartj/ehz425 (2020).

5 李积凤 & 杨媛华. 《肺血栓栓塞症诊治与预防指南》——关于静脉血栓栓塞症预防的推荐意见 %J 中国医刊. **53**, 1078-1082 (2018).

6 高二击, 李扬, 郭翔 & 杨异. 肺挫伤的诊断与治疗进展 %J 中华胸部外科电子杂志. **6**, 28-32 (2019).

7 于翠香 & 王西艳. 《中国成人医院获得性肺炎与呼吸机相关性肺炎诊断和治疗指南（2018 年版）》解读 %J 中国医刊. **56**, 951-953 (2021).

8 Waterhouse, C. Practical aspects of performing Glasgow Coma Scale observations. *Nursing standard (Royal College of Nursing (Great Britain) : 1987)* **31**, 40-46, doi:10.7748/ns.2017.e10189 (2017).

9 Dziewas, R. *et al.* European Stroke Organisation and European Society for Swallowing Disorders guideline for the diagnosis and treatment of post-stroke dysphagia. *European stroke journal* **6**, Lxxxix-cxv, doi:10.1177/23969873211039721 (2021).

10 Hao, P. Y., Chiang, J. H. & Chen, Y. D. Possibilistic classification by support vector networks. *Neural networks : the official journal of the International Neural Network Society* **149**, 40-56, doi:10.1016/j.neunet.2022.02.007 (2022).

11 Becker, T., Rousseau, A. J., Geubbelmans, M., Burzykowski, T. & Valkenborg, D. Decision trees and random forests. *American journal of orthodontics and dentofacial orthopedics : official publication of the American Association of Orthodontists, its constituent societies, and the American Board of Orthodontics* **164**, 894-897, doi:10.1016/j.ajodo.2023.09.011 (2023).

12 Abraham, A. *et al.* Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics* **8**, 14, doi:10.3389/fninf.2014.00014 (2014).

13 Stoltzfus, J. C. Logistic regression: a brief primer. *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine* **18**, 1099-1104, doi:10.1111/j.1553-2712.2011.01185.x (2011).

14 Besag, J., Green, P., Higdon, D. & Mengersen, K. Bayesian Computation and Stochastic Systems. **10 %J Statistical Science**, 3-41, 39 (1995).

15 The Risk of Determining Risk with Multivariable Models. **118**, 201-210, doi:10.7326/0003-4819-118-3-199302010-00009 %m 8417638 (1993).

16 Xu, L., Krzyzak, A. & Suen, C. Y. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics* **22**, 418-435, doi:10.1109/21.155943 (1992).

17 El-Naqa, I., Yongyi, Y., Wernick, M. N., Galatsanos, N. P. & Nishikawa, R. M. A support vector machine approach for detection of microcalcifications. *IEEE Transactions on Medical Imaging* **21**, 1552-1563, doi:10.1109/TMI.2002.806569 (2002).

18 Jin, Z., Hei, T. G. & Jin, K. *Application of support vector machines for categorizing biological and medical data*. Vol. 13105 CGAIDP (SPIE, 2024).

19 Jui, S. J. J. *et al.* Application of Entropy for Automated Detection of Neurological Disorders With Electroencephalogram Signals: A Review of the Last Decade (2012–2022). *IEEE Access* **11**, 71905-71924, doi:10.1109/ACCESS.2023.3294473 (2023).

20 Amiri, Z., Heidari, A., Navimipour, N. J., Esmaeilpour, M. & Yazdani, Y. The deep learning applications in IoT-based bio- and medical informatics: a systematic literature review. *Neural Computing and Applications* **36**, 5757-5797, doi:10.1007/s00521-023-09366-3 (2024).

21 Jin, H. & Ling, C. X. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering* **17**, 299-310, doi:10.1109/TKDE.2005.50 (2005).

22 Khalilia, M., Chakraborty, S. & Popescu, M. Predicting disease risks from highly imbalanced data using random forest. *BMC Medical Informatics and Decision Making* **11**, 51, doi:10.1186/1472-6947-11-51 (2011).

23 Ebrahim, O. A. & Derbew, G. Application of supervised machine learning algorithms for classification and prediction of type-2 diabetes disease status in Afar regional state, Northeastern Ethiopia 2021. *Scientific reports* **13**, 7779, doi:10.1038/s41598-023-34906-1 (2023).

24 Bhatt, C. M., Patel, P., Ghetia, T. & Mazzeo, P. L. Effective Heart Disease Prediction Using Machine Learning Techniques. **16**, 88 (2023).

25 Jiang, H. *et al.* Mobile Application and Machine Learning-Driven Scheme for Intelligent Diabetes Progression Analysis and Management Using Multiple Risk Factors. **11**, 1053 (2024).

26 Izonin, I., Tkachenko, R., Shakhovska, N., Ilchyshyn, B. & Singh, K. K. A Two-Step Data Normalization Approach for Improving Classification Accuracy in the Medical Diagnosis Domain. **10**, 1942 (2022).

27 Reshi, A. A. *et al.* Diagnosis of vertebral column pathologies using concatenated resampling with machine learning algorithms. *PeerJ. Computer science* **7**, e547, doi:10.7717/peerj-cs.547 (2021).

28 Natekin, A. & Knoll, A. Gradient boosting machines, a tutorial. **7**, doi:10.3389/fnbot.2013.00021 (2013).

29 Bentéjac, C., Csörgő, A. & Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review* **54**, 1937-1967, doi:10.1007/s10462-020-09896-5 (2021).

30 Rufo, D. D., Debelee, T. G., Ibenthal, A. & Negera, W. G. Diagnosis of Diabetes Mellitus Using Gradient Boosting Machine (LightGBM). **11**, 1714 (2021).

31 Mutanga, R. T., Naicker, N., Olugbara, O. O. J. I. J. o. A. C. S. & Applications. Detecting

Hate Speech on Twitter Network using Ensemble Machine Learning. (2022).