OEDL: an optimized ensemble deep learning method for the prediction of acute ischemic stroke prognoses using union features

Wei Ye[1†], Xicheng Chen[1†], Pengpeng Li[1†], Yongjun Tao[2†], Zhenyan Wang[1], Chengcheng Gao[1], Jian Cheng[3], Fang Li[1], Dali Yi[1,4], Zeliang Wei[1], Dong Yi[1] and Yazhou Wu[1*]

[1]Department of Health Statistics, College of Preventive Medicine, Army Medical University, Chongqing, China, [2]Department of Neurology, Taizhou Municipal Hospital, Taizhou, Zhejiang, China, [3]Department of Radiology, Taizhou Municipal Hospital, Taizhou, Zhejiang, China, [4]Department of Health Education, College of Preventive Medicine, Army Medical University, Chongqing, China

Background: Early stroke prognosis assessments are critical for decision-making regarding therapeutic intervention. We introduced the concepts of data combination, method integration, and algorithm parallelization, aiming to build an integrated deep learning model based on a combination of clinical and radiomics features and analyze its application value in prognosis prediction.

Methods: The research steps in this study include data source and feature extraction, data processing and feature fusion, model building and optimization, model training, and so on. Using data from 441 stroke patients, clinical and radiomics features were extracted, and feature selection was performed. Clinical, radiomics, and combined features were included to construct predictive models. We applied the concept of deep integration to the joint analysis of multiple deep learning methods, used a metaheuristic algorithm to improve the parameter search efficiency, and finally, developed an acute ischemic stroke (AIS) prognosis prediction method, namely, the optimized ensemble of deep learning (OEDL) method.

Results: Among the clinical features, 17 features passed the correlation check. Among the radiomics features, 19 features were selected. In the comparison of the prediction performance of each method, the OEDL method based on the concept of ensemble optimization had the best classification performance. In the comparison to the predictive performance of each feature, the inclusion of the combined features resulted in better classification performance than that of the clinical and radiomics features. In the comparison to the prediction performance of each balanced method, SMOTEENN, which is based on a hybrid sampling method, achieved the best classification performance than that of the unbalanced, oversampled, and undersampled methods. The OEDL method with combined features and mixed sampling achieved the best classification performance, with 97.89, 95.74, 94.75, 94.03, and 94.35% for Macro-AUC, ACC, Macro-R, Macro-P, and Macro-F1, respectively, and achieved advanced performance in comparison with that of methods in previous studies.

Conclusion: The OEDL approach proposed herein could effectively achieve improved stroke prognosis prediction performance, the effect of using combined data modeling was significantly better than that of single clinical

or radiomics feature models, and the proposed method had a better intervention guidance value. Our approach is beneficial for optimizing the early clinical intervention process and providing the necessary clinical decision support for personalized treatment.

Introduction

In recent years, with the increasingly serious aging phenomenon globally, the incidence of major chronic diseases represented by ischemic strokes has also increased [1]. Stroke is still the second leading cause of death in the world and the number one cause of acquired long-term disability, especially in China, which is the greatest challenge of stroke in the world, ranking third among the leading causes of death in China, second only to malignant tumors and heart disease [2]. Acute ischemic stroke (AIS) is associated with high morbidity, high mortality, and poor prognoses. They have become a major public health problem that cannot be ignored and have brought a great burden to the economy and society. In the context of limited medical resources, it is necessary to prioritize the implementation of nursing care for patients with poor prognoses, thereby reducing the incidence of disability [3]. In the era of precise diagnosis and individualized treatment, prognostic classification has become an important strategy for stroke management [4, 5]. The early prediction of prognoses is of great significance for improving the efficiency of stroke disease diagnosis and treatment and improving the levels of disease prevention and control [6].

In the past, prognosis evaluations in clinical practice mostly relied on the manual judgments of physicians, which required high-end medical technology and much physician experience, and the prediction effect of this approach was unstable, which limited its clinical promotion [7, 8]. As a new non-invasive technique, radiomics can extract high-throughput quantitative information from traditional medical images, enabling the assessment of internal tumor textures that cannot be captured by visual assessments [9, 10]. Radiomics aims to extract quantitative and high-dimensional data from digital biomedical images to facilitate the comprehensive exploration of disease information and progression, and it has been widely used in a variety of clinical fields [11, 12]. However, previous studies of this kind were mostly limited to radiomics alone and failed to comprehensively predict disease prognoses with clinical and radiomics features [13–17]. At present, there is still a lack of relevant research focusing on the predictive value of combined features for stroke prognosis, which has broad research prospects.

Compared with traditional prediction models, deep learning-based prediction models, represented by deep neural networks (DNNs), long short-term memory recurrent neural networks (LSTM-RNNs), and deep belief networks (DBNs), can automate and accurately analyze a large number of features and are suitable for various medical fields [18–21]. Ensemble learning has the advantages of fast operation and high accuracy and has been widely used in numerous fields, such as medical treatment, healthcare, and information technology [22]. Single machine learning and deep learning have the problems of limited convergence effect and difficulty in optimizing hyperparameters, which affects the improvement of prediction efficiency. Deep ensemble learning is expected to solve these problems and improve the accuracy of the model. Compared with shallow learning models and individual learning models, ensemble deep learning models can perform better on multiple learning tasks. They can also extract deeper essential features during the learning process, which can effectively improve the accuracy of the model prediction results [23].

In previous radiomics studies, the applications of deep ensemble models were relatively lacking [24–26]. If the selected network structure and parameter settings are not appropriate, this may increase the complexity of the model and reduce its overall operating efficiency [27]. Hence, the parameter optimization and layer number setting steps of deep ensemble models are still key issues that need to be solved [28]. To improve the optimization accuracy of these models and reduce the time required for the optimization process, such research usually requires the use of metaheuristic algorithms as optimization strategies [29, 30]. However, traditional algorithms often have problems such as slow convergence speeds and ease of falling into local optima [31, 32]. Research on optimization algorithms with novel optimization mechanisms, accurate solution methods, and robust computing power is still an important direction for feature and parameter selection.

This study aims to build a stroke prognosis prediction model in a deeply integrated way to provide a reference for the diagnosis and prevention of stroke. We adopt an ensemble concept involving data, methods, and algorithms and achieve excellent classification performance. Our contributions and the innovations of this study can be summarized as follows.

(1) In terms of data fusion, we innovatively extract, select, and fuse clinical features and imaging features. The combined data are beneficial to fully extract information and provide early warning for the prognosis of stroke more comprehensively and effectively.

(2) In terms of the categorical outcome, multicategorical outcome variables (normal group, mild group, and moderate-severe group) are used in this study. Compared with that of two-classification approaches, the multi-classification method is conducive to improving the pertinence of the classification, which is conducive to accurate prognosis judgment and intervention guidance.

(3) In terms of model construction, we innovatively construct the optimized ensemble of deep learning (OEDL) method. We comprehensively selected and integrated multiple deep learning methods to maximize the advantages of each method and verified the performance of the model for classification prediction. Our proposed model increases the diversity of prognosis prediction methods, enriches the methodological content of deep ensemble learning, provides new methods and ideas in its research field and clinical decision support for personalized intervention.

(4) For model optimization, we design a new Big Bang optimization algorithm (BBOA), which aims to implement the optimization process efficiently and accurately and then improve the efficiency of the feature selection and parameter search processes.

Materials and methods

The data that support the findings of this study are available from the corresponding author upon reasonable request. This study includes the following steps. (1) Data source and feature extraction: The clinical features and radiomics features are extracted in turn. (2) Data processing and feature fusion: The data filling, data noise reduction, data standardization, data screening, data splicing, data balancing, and related steps are performed. (3) Model construction: Clinical features, imaging features, and combined features from the data are included in turn. In this method, the concept of deep integration is used for modeling, and the base learner and the integration mode are selected in turn. (4) Model optimization: The proposed improved metaheuristic algorithm is used to improve the efficiency of the parameter search. Our technical route is shown in Figure 1.

Data source and feature extraction

This retrospective study was approved by the ethics committee of the Taizhou Municipal Hospital, and the requirement to obtain informed consent was waived. A total of 477 acute ischemic stroke (AIS) patients admitted to the Department of Neurology, Taizhou Municipal Hospital, Zhejiang Province, from January 2020 to April 2021 were recruited. The inclusion criteria were as follows: those who met the AIS diagnostic criteria, had complete clinical data, age more than 18 years. The patient had the first episode, and the MRI images were clear and without artifacts. Severe liver and kidney dysfunction, blood system diseases, malignant tumors, immune system diseases and other diseases, other serious nervous system diseases, and failure to cooperate with clinical treatment or follow-up as required by law were exclusion criteria. A total of 441 cases were eventually included. To ensure that the sample size met the needs of deep learning, we implemented balancing processing for the study data.

The prognosis groupings were based on the National Institute of Health Stroke Scale (NIHSS) at the time of discharge and could be divided into three groups (33, 34): a normal group (<1 point) with 106 cases; a mild group (1–4 points) with 289 cases; and a moderate-severe group (≥ 5 points) with 46 cases. In the following text, we refer to the normal, mild, and model severe groups as groups A, B, and C, respectively. The NIHSS scores could reflect the degrees of neurological deficit in patients and were used as prognostic indicators in this study.

The clinical data included NIHSS score at admission, disease type, OCSP classification, sex, age, body mass index (BMI), systolic blood pressure (SBP), left ventricular hypertrophy (LVH), homocysteinemia, history of hypertension, history of diabetes, history of coronary heart disease (CHD), history of atrial fibrillation, history of drinking, history of smoking, serum total cholesterol (TC), and low-density lipoprotein (LDL). The distribution of baseline data of each group is shown in Section Results of clinical feature selection of the results. Because of the first onset, relevant characteristics such as "stroke history" were not included in this article.
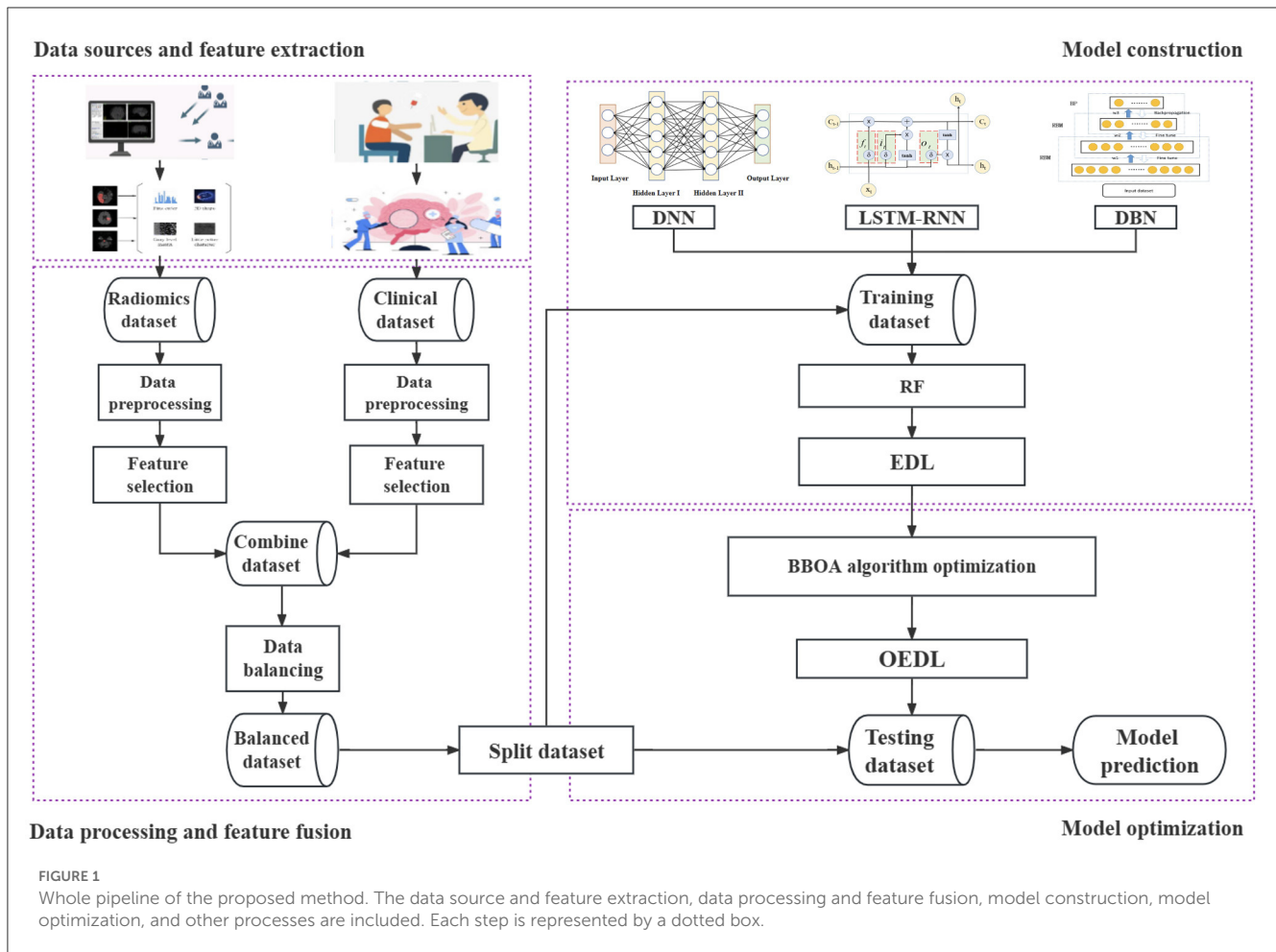
The image data were obtained from cranial MR images, and a Philips Achieve 1.5T scanner was used to obtain these data. The axial DWI sequence was acquired from all patients. To obtain DW images, the following parameters were used: the echo time was 101 ms, the repetition time was 3,211 ms, the number of excitations was 1, the slice thickness was 5 mm, the slice spacing was 1 mm, the acquisition matrix was 230 × 230, and the field of vision was 23 cm * 23 cm.

Each patient's first MR image was collected after admission. Two attending physicians independently segmented the regions of interest (ROIs) from the lesions, and ITK-SNAP 3.6.0 software was used for segmentation to obtain the 3D structural data of the lesions. The radiomics features of each annotated lesion were then obtained by using a radiomics analysis tool (the Pyradiomics package). The 2D mask labeling process for each patient is shown in Figure 2. The radiomics features included shape features (14 features), first-order statistics (162 features), gray-level dependence matrix features (GLDM features, 126 features), gray-level cooccurrence matrix features (GLCM features, 216 features), gray-level run length matrix features (GLRLM features, 144 features), gray-level size zone matrix features (GLSZM features, 144 features), and neighboring gray-tone difference matrix features (NGTDM features, 45 features). Finally, 17 clinical features and 851 radiomics features were initially included in this study.

Data processing and feature fusion

After feature extraction, data preprocessing was performed, including data filling, data noise reduction, data standardization, data screening, data splicing, data balancing, and other steps, as shown in Figure 3.

First, in data imputation, we used multiple imputation methods (35). Multiple imputation is a commonly used method to deal with missing values in data. Its basic principle is to generate multiple complete data sets through simulation, and each data set uses different methods to impute missing values. In this study, we use a multiple imputation method based on the multiple Monte Carlo Method to imputation the samples to reduce the impact of missing data on model construction. The multiple Monte Carlo method is a statistical method that uses multiple independent random samples to estimate the expected value. Suppose the expected value $E[f(x)]$

FIGURE 1
Whole pipeline of the proposed method. The data source and feature extraction, data processing and feature fusion, model construction, model optimization, and other processes are included. Each step is represented by a dotted box.

of some function $f(x)$, where $x \in R^d$ is a vector of dimension d. The formula for the multiple Monte Carlo method is as follows:

$$\bar{f}_n = \frac{1}{N}\sum_{i=1}^{N} f(x_n, i) \tag{1}$$

$$\bar{f}_m = \frac{1}{M}\sum_{i=1}^{M} f_{n(j)} \tag{2}$$

where $\bar{f}_n$ is the sample mean value of $f(x)$ obtained from the n-th sampling, x_n, i is the i-th sample point obtained from the n-th sampling, and N is the number of samples for each sampling; $\bar{f}_m$ is the multiple Monte Carlo estimator obtained by averaging $\bar{f}_n$ obtained from n samples for m times, where $n(j)$ is the sample set used for the j-th sample.

Second, for high-dimensional imaging features, a large number of useless noise features will affect the screening of meaningful features and increase the difficulty of model construction (36). In data denoising, we chose the variance selection filtering method to perform variance-based feature screening (37) and then filtered out features with small differences. The variance of each feature was

calculated, and features with variances greater than the threshold were selected. If the variance is small, it means that there is a small difference between these samples with respect to the feature, and this feature is not conducive to sample discrimination. We filtered features with zero or less variance to preferentially exclude features with lower contributions.

Third, the data were standardized by distinguishing clinical features from radiologic features. (A) Clinical characteristics were assigned to a range between 0 and 1 by one-hot encoding because one-hot encoding can extend the values of discrete features to a Euclidean geometry space and thus fuse standardized imaging features (38, 39). The mathematical formula for one-hot encoding is as follows. Let the value of a discrete feature x with n different values be $\{x_1, x_2, ..., x_n\}$; then, the one-hot encoding of this feature is an n-dimensional vector v, where if and only if the value of x is x_i, $v[i] = 1$; if x is not x_i, $v[i] = 0$. (B) For radiomics features, we selected a normalization method for feature processing, aiming to eliminate dimensional differences between different features, avoid possible deviation in the model training, improve the convenience of data processing, and speed up the model convergence (40, 41). Normalization helps to ensure dimensional unity between different features, thus improving the robustness and generalizability of the model (42, 43). Normalization refers to scaling the data so that it falls within a specific interval. Standardization (Sz) is the
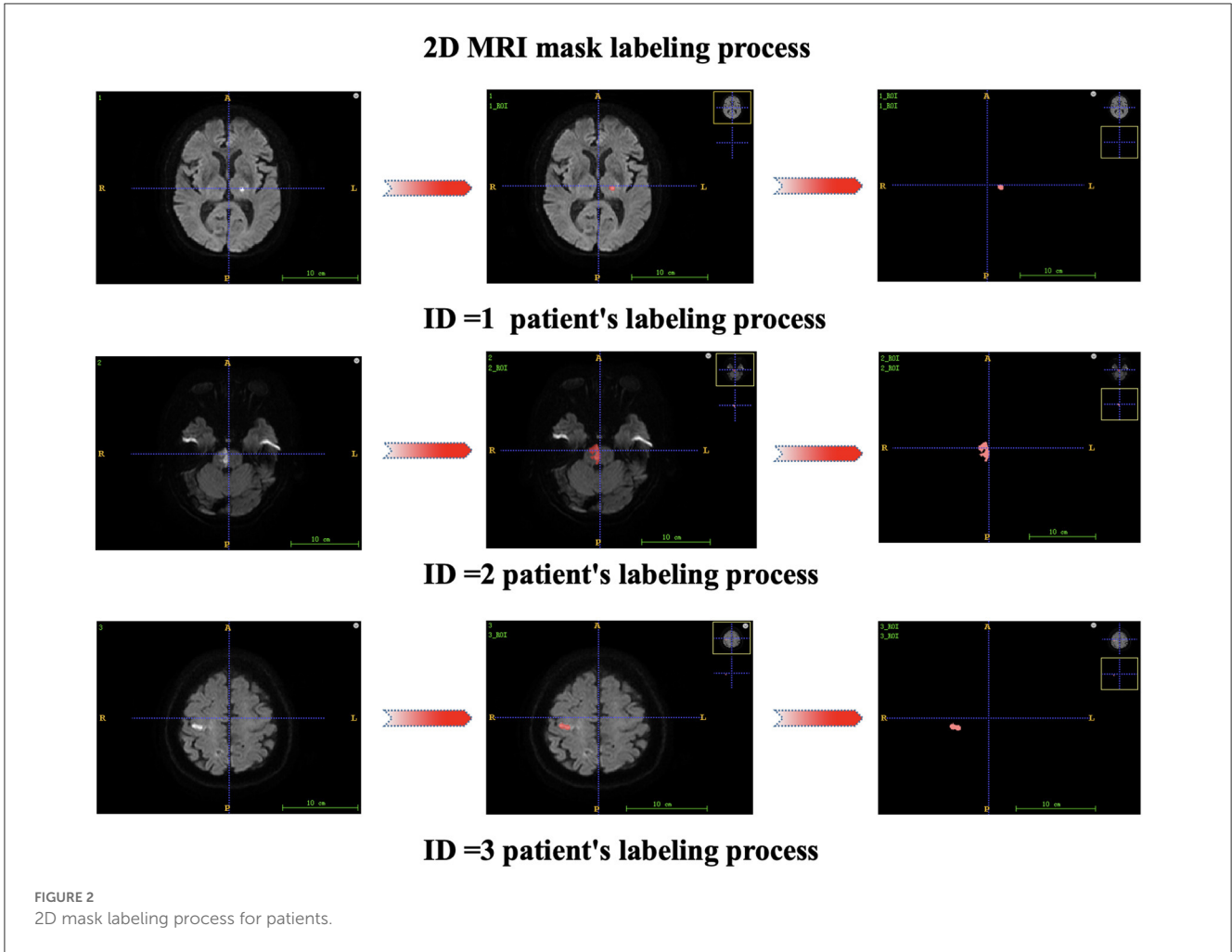
FIGURE 2
2D mask labeling process for patients.

transformation of data into a normal distribution with a mean of 0 and a standard deviation of 1 (44). Suppose that there are N samples, each sample has n features, and the value of the i-th feature of all N samples is $x_{i1}, x_{i2}, ..., x_{in}$. Then, the standardized mathematical formula of the feature is as follows:

$$Sz = \frac{x_{ij} - \mu_i}{\sigma_i} \quad (3)$$

where μ_i denotes the mean of the i-th feature over all N samples and σ_i denotes the standard deviation of the i-th feature over all N samples. For the i-th feature in each sample, a new value can be obtained from this formula, representing the relative size and distribution of the feature across the entire data set.

Fourth, we use the embedded method to filter and reduce the dimensions of the data. The LightGBM and XGBoost algorithms are selected to perform feature importance scoring and selection, the top 50 most important features in terms of weight are screened out, and the features appearing in both methods and the top 10 features in terms of weight in each method are sorted out. (A) LightGBM is a gradient-boosting framework based on a decision tree (DT). It uses a node segmentation strategy based on leaves, seeks the leaf with the largest gain among all the current leaves, and

finally generates a boosted tree (45, 46). The LightGBM algorithm is based on the selection of partition points based on the histogram algorithm and reduces the number of samples and features required in the training and learning processes through two methods, namely, gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB), to maintain high learning performance and reduce the resource occupation in terms of time and space in the training process (47, 48). Let X^s be the input space, s be the feature dimension, and Y be the output space. The given training dataset is $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$, where $\vec{x}_i = (\vec{x}_i^{(1)}, \vec{x}_i^{(2)}, ..., \vec{x}_i^{(s)}), i = 1, 2, ..., n$ represents the input instance and $\{g_1, g_2, ..., g_n\}$ represents the negative gradient direction of the loss function relative to the model output at each enhancement iteration. Let n represent the number of samples, and let O be the training set of the DT on a node. Then, the information gain $V_{j|o}(d)$ (49) of feature j at node d can be defined as

$$V_{j|o}(d) = \frac{1}{n_0} \left(\frac{(\sum_{x_i \in o : x_{ij} \leq d} g_i)^2}{n_{l|o}^j(d)} + \frac{(\sum_{x_i \in o : x_{ij} > d} g_i)^2}{n_{r|o}^j(d)} \right) \quad (4)$$

where $n_0 = \sum I(x_i \in o)$, $n_{l|o}^j = \sum I[x_i \in o : x_{ij} \leq d]$, and $n_{r|o}^j = \sum I[x_i \in o : x_{ij} > d]$. (B) The XGBoost algorithm can generate a second-order Taylor expansion of the utilized

FIGURE 3
Feature fusion process.

FIGURE 4
SMOTEENN balancing algorithm.
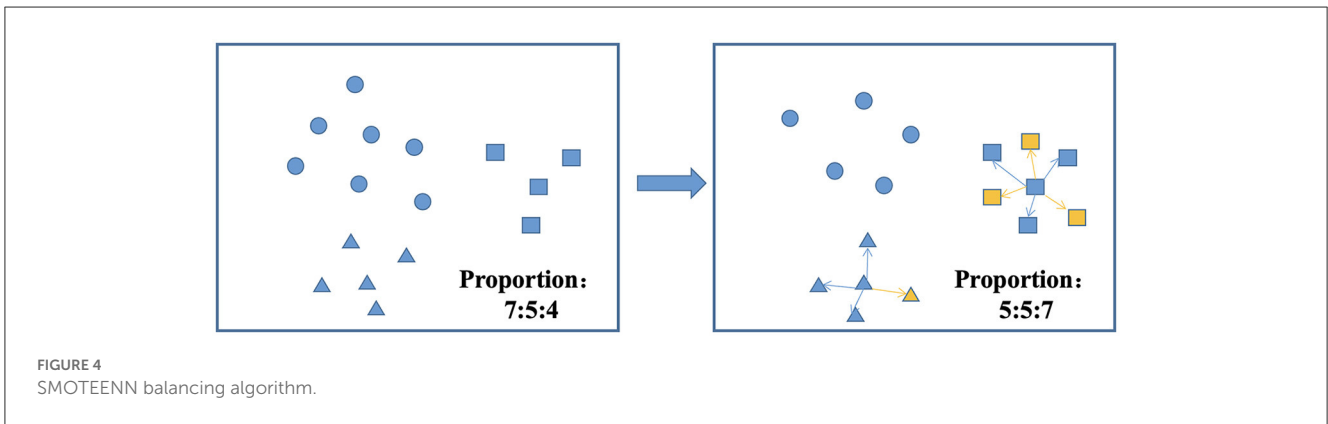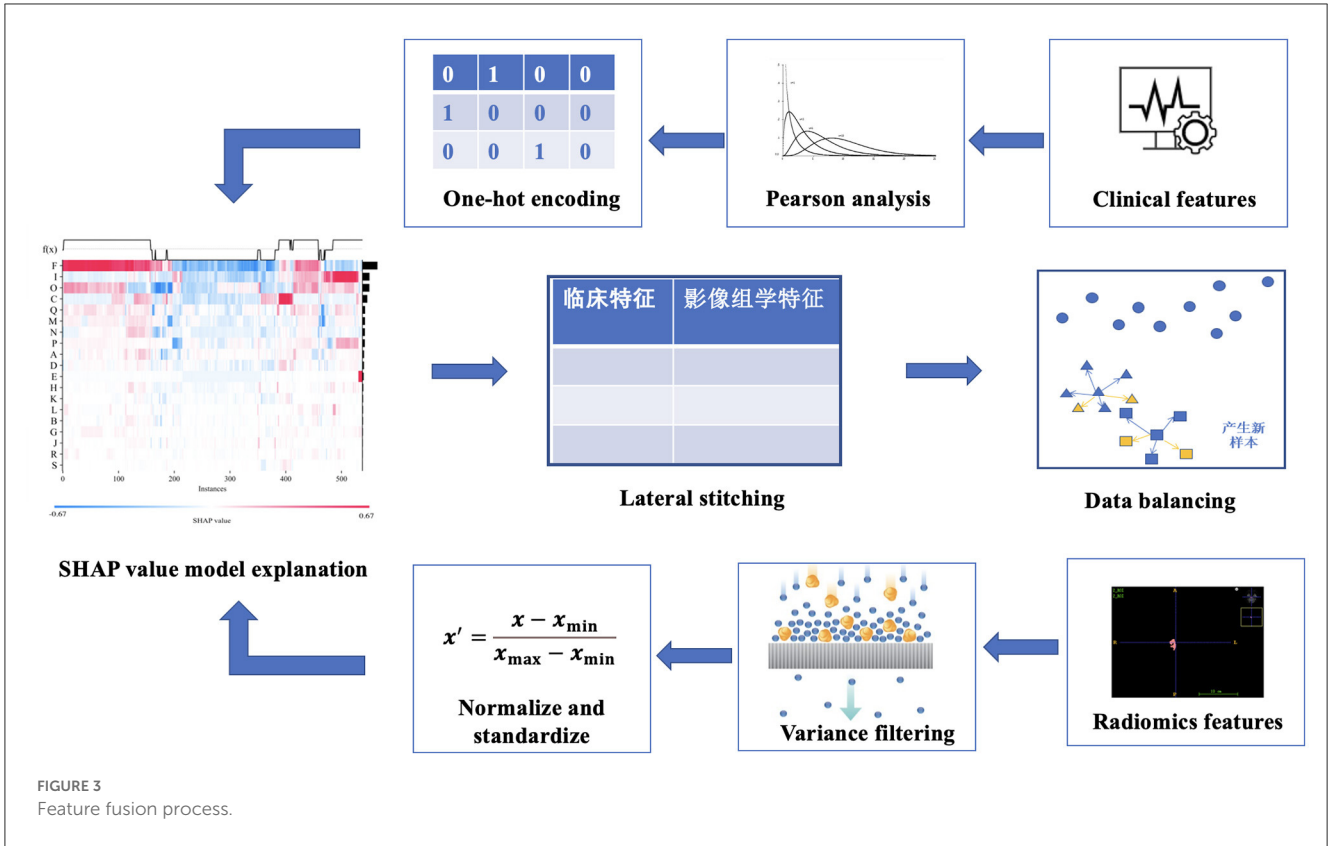
loss function and obtain the optimal solution for the regular term outside the loss function (50, 51). The larger the weight of a feature and the more times it is selected by the boosted tree, the more important the feature is considered to be (52, 53). Suppose that the model has t DTs, n represents the total number of samples, f_t represents the t-th regression tree, F represents the collective space of all DTs, and $\hat{y}_i^t$ represents the total predicted value for the i-th sample after adding the outputs of the t DTs. Then, the predicted value of XGBoost (54) can be expressed as

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i), f_k \in F, i \in n \qquad (5)$$

Its loss function is

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^{t} \Omega(f_k) \qquad (6)$$

$$\Omega(f_k) = \lambda T + \frac{1}{2}\lambda \|w\|^2 \qquad (7)$$

where l represents the error between the predicted value and the actual value, T and w represent the number and weight of the leaf nodes, respectively, and γ and λ represent regularization coefficients. The k-th tree is represented by k, and the complexity of k trees is represented by $\sum_{k=1}^{t} \Omega(f_k)$. (C) The Pearson

FIGURE 5
Construction process of the deep integration learning method.

FIGURE 6
Schematic diagram of each base learner. **(A)** DNN, **(B)** LSTM-RNN, and **(C)** DBN.

correlation coefficient can measure the strength and direction of the linear relationship between two variables, and different correlation coefficients can be selected according to different data characteristics (55). If two features have a high correlation, this indicates that the information contained in the two features is highly similar, and too much similar information can reduce the performance of the chosen algorithm (56). Hence, only one feature must be reserved for features whose correlations are higher than

a certain threshold. To avoid the negative impact of collinearity features on outcome variables, we randomly retained only one of many features with Pearson correlation coefficients greater than the threshold (0.9 in our study). (D) In this study, the SHAP model interpreter tool is used to explain the operation mechanism of the model. SHAP can construct a weighted explanatory model to calculate the contribution of each feature to the results (57, 58). In the interpretation of radiomics and clinical features using LightGBM and XGBoost, respectively, each sample can generate a predictive value, and the SHAP value is expressed as $f(x)$, which can represent the numerical value assigned to each feature in a sample. Red represents features that act positively, and blue represents features that act negatively (5). After the screening of clinical and radiomics features, the combined features were constructed by stitching.

Fifth, there are three common approaches to dealing with class imbalance: undersampling, oversampling, and hybrid sampling techniques. Undersampling techniques include the random undersampling technique, and oversampling techniques include the random oversampling, SMOTE, adaptive synthetic (ADASYN), and borderline-SMOTE techniques (59). SMOTEENN is a method that combines oversampling and undersampling to handle both sample imbalance and noisy data. The SMOTE method increases the number of minority class samples by random oversampling, while the ENN method reduces the number of majority class samples by removing majority class samples. The combination of these two methods can better balance the class distribution in the dataset, thus improving the performance of the classifier (60, 61). The balancing algorithm can balance the number of samples for each classification, thus effectively improving the prediction performance of the model with unbalanced datasets (62, 63). Figure 4 shows the process of the SMOTEENN balancing algorithm, which not only synthesizes new samples for minority classes but also prunes duplicate samples to improve the difference between groups.

Model construction and optimization

The content in this section can be divided into the selection of the base learner, model construction, model optimization, and other steps. The construction process combines the ideas of ensemble learning and deep learning to construct an ensemble of deep learning (EDL) model with a multilayer cascade structure. The optimized ensemble of deep learning (OEDL) model is established by adding an optimization algorithm. The model is built as shown in Figure 5.

First, the selection of base learners is needed. DNN, LSTM-RNN and DBN are used as base learners, and the schematic diagram of each base learner is shown in Figure 6. (A) A deep neural network (DNN) refers to a neural network with more than one hidden layer [64]. The input layer and hidden layer, hidden layer and hidden layer, and hidden layer and output layer all have linear relationships, which can be expressed as

$$y_i = \sigma\left(\sum w_i^n x_i^n + b_i\right) \tag{8}$$

where y_i is the next neuron, x_i is a feature or neuron connected to y_i, σ is an activation function in a layer, n is the number of neurons or features connected to the neuron, w_i is a weight coefficient between a feature and a neuron or between neurons, and b is a constant. (B) Long short-term memory (LSTM) is proposed to solve the problem of vanishing or exploding gradients in recurrent neural networks (64, 65). The unit structure records the patient characteristic information of the current state by introducing a new internal state and carries out internal information transmission. First, an input gate i_t, a forget gate f_t, and an output gate o_t are calculated by using the patient characteristic information x_t of the current state and the hidden state h_{t-1} of the last time. Then, the input gate i_t and the forget gate f_t are used to control the retained historical characteristic information and the current state characteristic information of the patient, respectively, to obtain a new C_t. Finally, the input gate o_t is used to transfer the patient characteristic information of the internal state to the hidden state h_t. To achieve the classification effect, an RNN fully connected layer is added behind the LSTM unit to construct an LSTM-RNN to obtain a multi-classification result. (C) Deep belief networks (DBNs) are probabilistic generative models consisting of multiple layers of restricted Boltzmann machines. The main structure combines several layers of RBM and one layer of a BP network and outputs the results by the BP network. The specific steps are as follows. First, the features are trained in each layer of the RBM network separately in an unsupervised manner to ensure that the feature information is reused and retained. Then, the trained features enter the BP network to train the classifier through supervision. Finally, a backpropagation network fine-tunes the training error information direction of each RBM layer so that optimization can be achieved throughout the whole network. The parameters of the DBN are given by w (connection weights), b (visible unit bias), and c (hidden unit bias). The probability of input vector v and output vector h is given by

$$p(v, h) = \frac{e^{-E(v,h)}}{Z} \tag{9}$$

where $-E(v, h)$ is the energy function

$$E(v, h) = -b^T v - c^T h - h^T W v \tag{10}$$

Z is the normalizing factor obtained by summing the numerator over all possible statuses of h and v:

$$Z = \sum_{v,h} -E(v, h) \tag{11}$$

Second, after considering bagging, boosting, stacking, and other methods (see the section Results), we chose the stacking algorithm as the ensemble method in this study, and the model constructed by it is named as the ensemble of deep learning (EDL). Stacking is a method that combines the outputs of multiple base learners according to a certain combination strategy (66, 67). We chose classical and representative deep learning models such as DNN, LSTM-RNN, and DBN as the base learner (68–70) and

```
1: ----------------Step 1: Input
 Dataset D ={(x₁₍₁₎,x₂₍₁₎,...,y₍₁₎),(x₁₍₂₎,x₂₍₂₎,...,y₍₂₎),(x₁₍ₙ₎,x₂₍ₙ₎,...,y₍ₙ₎)};
Primary learning algorithm:
PLA={DNN,LSTM-RNN,DBN};
Secondary learning algorithm:RandomForest(RF)
3: ----------------Step 2: Process
1: split D:Train_data,Testing_data
2: for t = 1,2,...,T do
3:         h₍ₜ₎ = Stratified Fold (Train_data);
4: end for
5:New_Train_data =∅;
6: for i in PLA do
7:         for t = 1,2,...,T do
8:             Zᵢₜ =h₍ₜ₎(PLA₍ᵢ₎);
9:         end for
10:     New_Train_ data = ∅ ∪ ((Zᵢ₁,Zᵢ₁,...,Zᵢ₁),yᵢ);
11: end for
12: New_Test_data = ∅;
13: for i in PLA do
14:         for t = 1,2,...,T do
15:             Zᵢₜ= h₍ₜ₎(PLA₍ᵢ₎);
16:         end for
17:     New_Test_data = ∅∪ ((Zᵢ₁,Zᵢ₁,...,Zᵢ₁),yᵢ);
18: end for
19: Training_RF = RF(New_Train_data)
3: --------------Step 3: Output
Testing_RF = Training_RF
        (h₍₁₎(New_Test_ data₍₁₎),h₍₂₎(New_Test_ data₍₁₎),...,
        h₍ₜ₎(New_Test_ data₍₁₎))
```

Algorithm 1. Stacking pseudocode.

random forest (RF) as the meta-learner. After the training of each base learner was completed, we used the stacking algorithm for analysis; that is, the outputs of multiple base learners were taken as a new dataset that was incorporated into the meta-learner (random forest was selected in this study) for learning and prediction. We integrated the results of the three neural networks and formed probabilities for the three classifications to obtain the final prediction for each sample. The deep learning system was iterated 100 times, and finally, the optimal model was selected by using a greedy strategy. The pseudocode for the stacking algorithm is shown in Algorithm 1.

The innovations of the OEDL method proposed in this study can be reflected in the following aspects. (A) When splitting the training set and the test set, the random stratification method is improved to the label percentage stratification approach to achieve the effect of label balancing. (B) During data selection, we selected clinical features and radiomics features in turn, analyzed clinical features and high-dimensional, abstract radiomics information as combined features, and finally built a combined feature model. (C) We chose the method of deep integration and comprehensively utilized the advantages of each deep learning model to improve its effectiveness and generalization. (D) We innovatively used an improved metaheuristic algorithm (see Section Model training for details) for optimization purposes to ensure the excellence of the classification results.

Third, we performed model optimization based on a newly proposed optimization algorithm. The above EDL proposal combines deep learning and ensemble learning ideas, but there is still the problem of the slow hyperparameter search. To solve this problem, we considered introducing a metaheuristic algorithm. Based on the stacking idea and the framework of particle swarm optimization, we proposed the big bang optimization algorithm (BBOA), which aims to solve the parameter optimization problem in deep networks and applied it to the OEDL method. During the analysis, the algorithm draws on the particle swarm optimization algorithm and the black hole theory of the Big Bang (71), as shown in Figure 7. In the process of constructing the algorithm, we used a sinusoidal chaotic map, an adaptive inertia weight, a greedy strategy, and other optimization methods. The symbol descriptions of the algorithm are shown in Table 1.

(A) The galaxy's initial position can be expressed as

$$x_j^H = U_x + m_{ij}^H * (D_x - U_x) \tag{12}$$

$$x_j^L = U_x + m_{ij}^L * (D_x - U_x) \tag{13}$$

$$m_{ij}^H = a(m_{i-1j}^H)^2 \sin(\pi m_{i-1j}^H) \tag{14}$$

$$m_{ij}^L = a(m_{i-1j}^L)^2 \sin(\pi m_{i-1j}^L) \tag{15}$$

In the study, Formula (14) and Formula (15) add the sinusoidal chaotic map, which acts as an initial randomization to make the distribution range of the star group more dispersed (72). A sinusoidal chaotic map is a non-linear map that can produce chaotic phenomena (73), where a is any constant, and the initial values m_{ij}^H and m_{ij}^L can be any number, but for the depth rule suitable for the deep learning neural network, the initial value is a random integer between 0 and 100, and the chaotic map is rounded to indicate the number of neurons in the deep learning.

(B) The rates of expansion of the galaxies can be expressed as

$$E_i^H = w_w E_{1i}^{t-1} + ef_1 \cdot r(y_{oi}^t - H_i^t) + ef_2 \cdot r(y_{oi}^t - H_i^t) \tag{16}$$

$$E_i^L = w_w E_{2i}^{t-1} + ef_1 \cdot r(y_{oi}^t - L_i^t) + ef_2 \cdot r(y_{oi}^t - L_i^t) \tag{17}$$

$$w_w = w_s - (w_s - w_e)\frac{t}{T} \tag{18}$$

In the study, Formula (18) is the added adaptive inertia weight, and its function is to regulate the initial expansion speed. Adaptive inertia weights are a variant of inertia weights. Each galaxy should constantly consider its historical and global best position when updating its expansion speed. The adaptive inertia weight can dynamically adjust the value of the inertia weight according to the historical state of the galaxy so that the algorithm converges to the
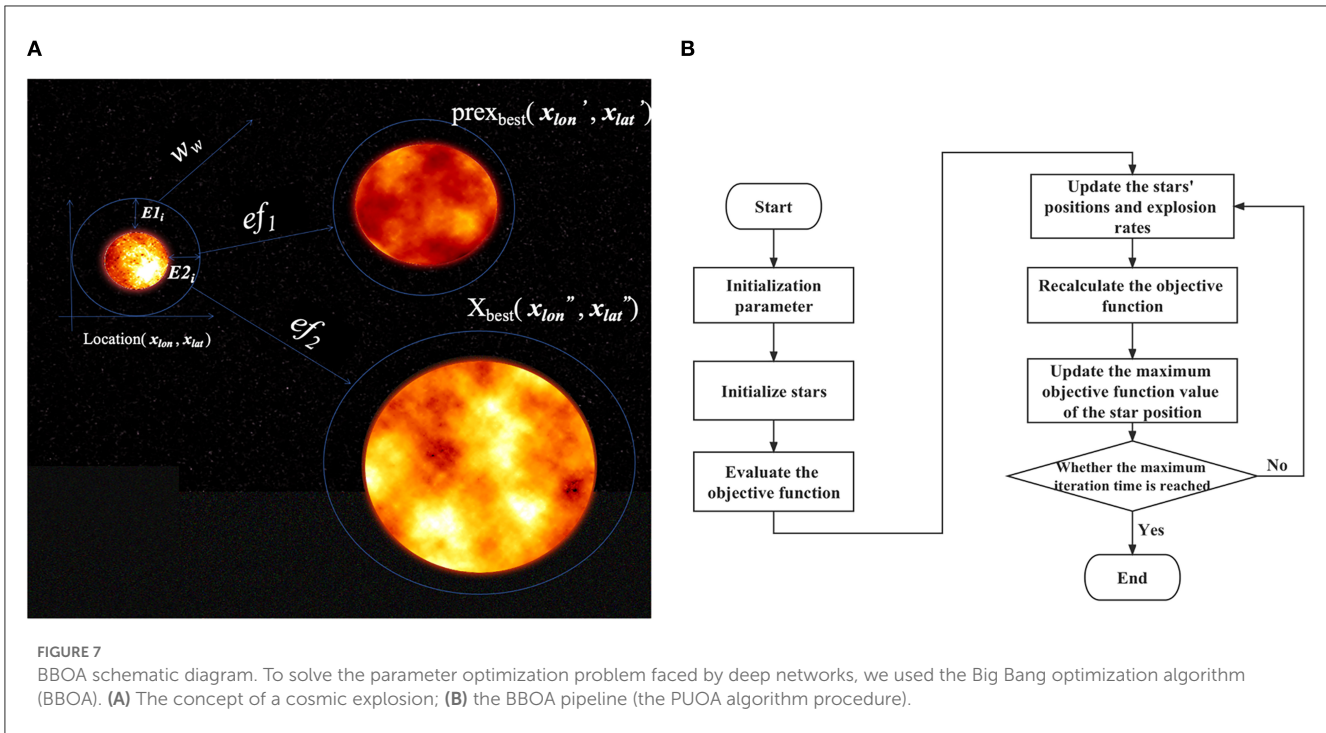
FIGURE 7
BBOA schematic diagram. To solve the parameter optimization problem faced by deep networks, we used the Big Bang optimization algorithm (BBOA). **(A)** The concept of a cosmic explosion; **(B)** the BBOA pipeline (the PUOA algorithm procedure).

optimal solution faster. The function of the adaptive inertia weight is to regulate the initial expansion speed.

(C) The transformation of the production expansion center of galaxies affected by higher expansion velocities is expressed as

$$x(H)_i = x(H)_i + E_i^H \qquad (19)$$

$$x(L)_i = x(L)_i + E_i^L \qquad (20)$$

(D) The optimal solution is the velocity of the largest star in the universe, which can be expressed as

$$Y_o = \begin{cases} y_o, & if\, y_o > Y_o; \\ Y_o, & else. \end{cases} \qquad (21)$$

In this study, to update the optimal solution, the greedy strategy is used, which can be expressed as

$$y_o = \max(y_o^{\,t}) \qquad (22)$$

BBOA was used in OEDL to optimize the number of hidden layers and the number of neurons in each layer of the DNN, LSTM-RNN, and DBN. The stacking integration algorithm was used to integrate the three models after each model was optimized.

Model training

This research was carried out on a Linux workstation equipped with a GPU. The software platform was based on Python

3.7. The proposed algorithms were implemented based on the TensorFlow 2.8 framework. The GPU was used to accelerate the training process. Among the study population, 70% of the data were randomly selected for the training set, and the remaining 30% were used as the test set. Statistical analysis was performed using Python 3.7.0, SPSS 26.0 (SPSS Inc., Chicago, IL, USA).

For various algorithm models, the algorithm was implemented using data split based on a ratio of 7:3 for training and testing. The model was fitted with the training set, the hyperparameters except for the number of hidden layers and the number of neurons were determined by the grid search method, and the best parameter model was selected after 50% cross-validation. Finally, the model was tested with the test set to evaluate the generalizability of each model. The hyperparameters determined by the grid search method in this study were based on a learning rate of 10-4, a batch size of 20, a momentum term of 0.9, and 1,000 epochs. In addition, the imbalanced distribution of the sample size in each category will lead to the prediction bias of the model. To eliminate this effect, we used the SMOTEENN algorithm to enhance the fused features.

In this study, the implementation of the deep learning network was based on the Keras package in TensorFlow 2.8. The Adam optimizer was used to optimize the gradient of the deep learning model, and the cross-entropy loss function was combined with the softmax activation function to obtain better classification results. The neural network of the three base learners was initially set as a double layer, and the number of neurons in each layer was 10. The BBOA optimizes the numbers of layers and neurons of each base learner to fix the model.

The evaluation indicators in the classification model were obtained based on a confusion matrix, and they include the

TABLE 1 Algorithm process symbol description.

Notation	Meaning
x_j^H	Initial position of the j-th galaxy (horizontal axis)
x_j^L	Initial position of the j-th galaxy (vertical axis)
m_{ij}^H	Initial position of the i-th star in the j-th galaxy (horizontal axis)
m_{ij}^L	Initial position (vertical axis) of the i-th star in the j-th galaxy
U_x	Upper boundary of the galaxy (horizontal axis)
D_x	Lower boundary of the galaxy (vertical axis)
ef_1	The speed of the fastest star in the galaxy
ef_2	The speed of the fastest star in the universe
r	Random number between 0 and 1
E_j^H	Galactic expansion speed (horizontal axis)
E_j^L	Expansion speed of the galaxies (vertical axis)
H_i^t	Time position of the largest star t in the galaxy (horizontal axis)
L_i^t	Time position of the largest star t in the galaxy (vertical axis)
y_{oi}^t	The speed of the fastest star in the galaxy
Y_{oi}^t	The speed of the fastest star in the universe
w_s	The initial inertia of stars in the galaxies
w_e	The final inertia of stars in the galaxies
t	Evolution time (number of iterations)
T	Total evolution time (iterations)
w_w	Adaptive inertia weight
Y_o	Velocity of the largest star in the universe (optimal solution)
y_o	The velocity of the largest star in the galaxy

following four basic indicators: "true positive" (TP) means that the prediction is true and the actual value is also true; "true negative" (TN) means that the prediction is false and the actual value is also false; "false positive" (FP) means that the prediction is true but the actual value is false; "false negative" (FN) means that the prediction is false but the actual value is true. Among multiple classes, each class i has values TP_i, TN_i, FP_i, and FN_i. T_iP_i represents that the true class i is correctly predicted as class i, and F_jP_i represents that the true class j is incorrectly predicted to be class i.

In this study, the evaluation indices include Macro-AUC, accuracy (ACC), macrosensitivity (Macro-R), macrospecificity (Macro-P), and Macro-F1 score (Macro-F1) (74). The ROC curve for each classification was plotted with the true positive rate of each classification as the vertical axis and the false positive rate of each classification as the horizontal axis. The area under the ROC curve of each category is the AUC value of each category, and Macro-AUC is the sum of all types of areas and the average. The value range is [0 ∼ 1]. The greater the value is, the more accurate the classification. The indicators can be expressed as follows:

$$ACC = \frac{\sum_{i=1}^{n=3} TP_i}{\sum_{i=1}^{n=3} (TP_i + FP_i)} \quad (23)$$

$$macro\_PRE = \frac{1}{n} \sum_{i=1}^{n=3} \left(\frac{TP_i}{TP_i + FP_i} \right) \quad (24)$$

$$macro\_SEN = \frac{1}{n} \sum_{i=1}^{n=3} \left(\frac{TP_i}{TP_i + FN_i} \right) \quad (25)$$

$$macro\_F1score = \frac{2 \cdot macro\_PRE \cdot macro\_SEN}{macro\_PRE + macro\_SEN} \quad (26)$$

Results

The results include the results of clinical feature selection, imaging feature selection, comparison of the prediction performance of each method, comparison of the prediction performance of each feature, comparison of the prediction performance of each balanced method, and comparison with previous studies.

Results of clinical feature selection

In the clinical feature selection stage, 17 features were included in the model, and all of these features passed the correlation test (as shown in Figure 8). The clinical characteristics according to the discharge NIHSS classification are presented in Table 2.
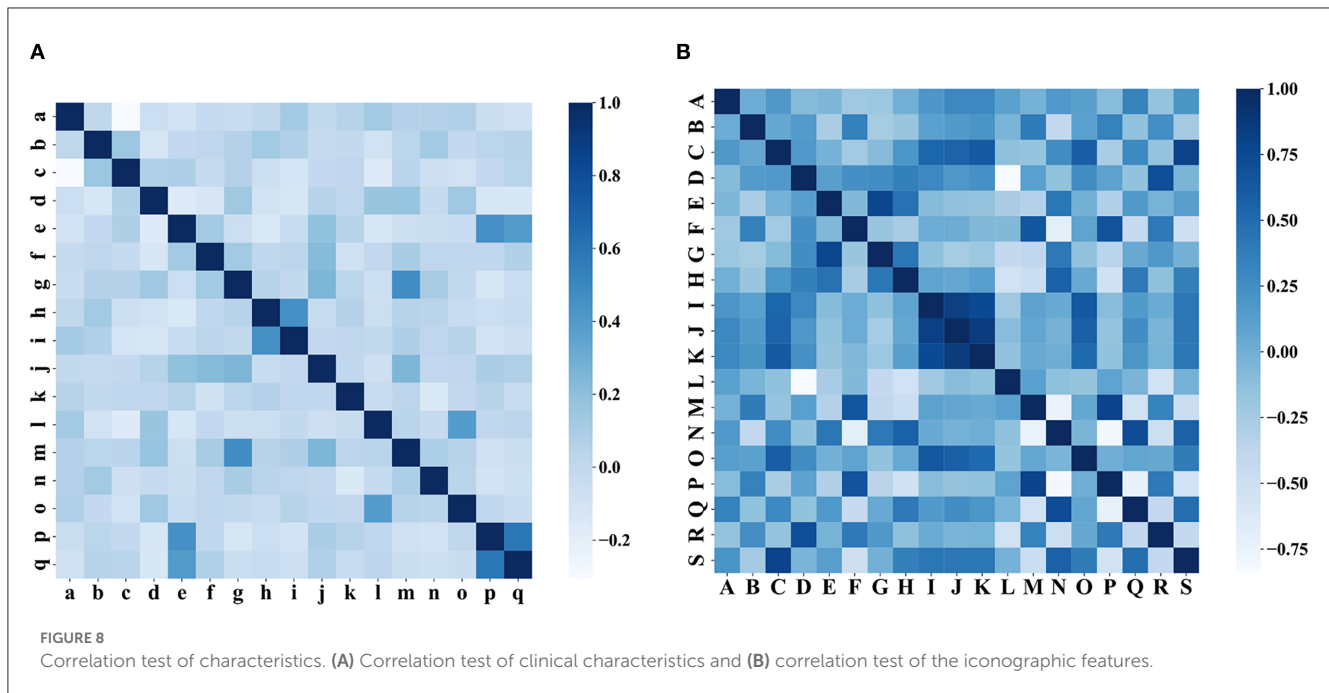
Results of radiomics feature selection

In radiomics feature selection, we first selected 328 features from 851 features using variance selection (threshold 0.3), used the LightGBM and XGBoost algorithms to screen out 81 more important features with the top 50 weights, and sorted out 19 features that appeared in both methods and the top 10 features in their respective methods. After the correlation test, 19 image features were selected. A rose plot was drawn based on the 19 features and their importance weights to the model, as shown in Figure 9.

To visualize the importance of the selected features, we used SHAP to illustrate the degrees to which these features influenced the prediction results, as shown in Figure 10. The SHAP value represents the contribution of each feature to the final prediction and can effectively explain the model prediction for each sample. The feature ranking (y-axis) represents the importance of the prediction model, and the corresponding SHAP value (x-axis) represents the degree of feature influence.

Comparison of the prediction performance of each method

Using the joint dataset as an example, the classification results of these methods are compared. The results show that the OEDL constructed based on the concept of ensemble optimization

FIGURE 8
Correlation test of characteristics. **(A)** Correlation test of clinical characteristics and **(B)** correlation test of the iconographic features.

obtained the best classification performance, and its Macro-AUC, ACC, Macro-R, Macro-P, and Macro-F1 reached 97.89, 95.74, 94.75, 94.03, and 94.35%, respectively, as shown in Table 3.

Comparison of the prediction performance of each feature

The classification results of EDL and OEDL were compared. The results show that compared with that using the clinical and radiomics features, the method using the combined data had better classification performance, and the EDL method achieved a Macro-AUC of 96.68% and an ACC of 92.55%. The OEDL method achieved a Macro-AUC of 97.89% and an ACC of 95.74%, as shown in Table 4. We also visualized the classification results of the OEDL method with the three features in the form of ROC curves, as shown in Figure 11.

Comparison of the prediction performance of each balanced method

The classification results of the combined features and the OEDL method were compared. Compared with the unbalanced, oversampled, and undersampled techniques, SMOTEENN based on a mixed sampling method achieved the best classification performance, and its Macro-AUC, ACC, Macro-R, Macro-P, and Macro-F1 reached 97.89, 95.74, 94.75, 94.03, and 94.35%, respectively, as shown in Table 5.

Comparison with previous studies

To demonstrate the advanced performance of the method proposed in this study, we reviewed relevant research in the field

of AIS classification and prediction and compared the AUC and ACC of each study. Although the datasets, classification numbers, and other aspects of these studies differed, the differences in the results have some implications for the excellence of the methods. The comparison results show that the proposed method has significantly better classification performance in terms of AUC, ACC, and other aspects than that of previous studies, and it has better classification advantages. For more information, see Table 6.

Discussion

AIS is one of the many diseases that endangers the health of Chinese residents. It is difficult and expensive to check, and it is difficult to evaluate the early prognosis (75). We used joint features to train the OEDL model to predict the prognosis of AIS, which is of great significance to improve the diagnosis and prevention system of AIS and promote the optimal allocation of medical resources.

In terms of data collection and processing, we used clinical features and radiomics features creatively to build joint features, and we built a complete and feasible data processing operation process. Compared with the clinical and imaging feature models, the method using the combined data had better classification performance, and the EDL method achieved a Macro-AUC of 96.68% and an ACC of 92.55%. The OEDL method obtained a Macro-AUC of 97.89% and an ACC of 95.74%. Joint-feature modeling produces better results than single-feature modeling. The reason for the analysis is that simple clinical feature model information is easy to collect, but prediction efficiency is limited due to clinical feature instability; radiomics features can be used to achieve high prediction efficiency, but the inclusion of purely influencing omics features is limited; joint feature modeling can incorporate more comprehensive and objective information. Feature selection is the process of finding the feature subset that

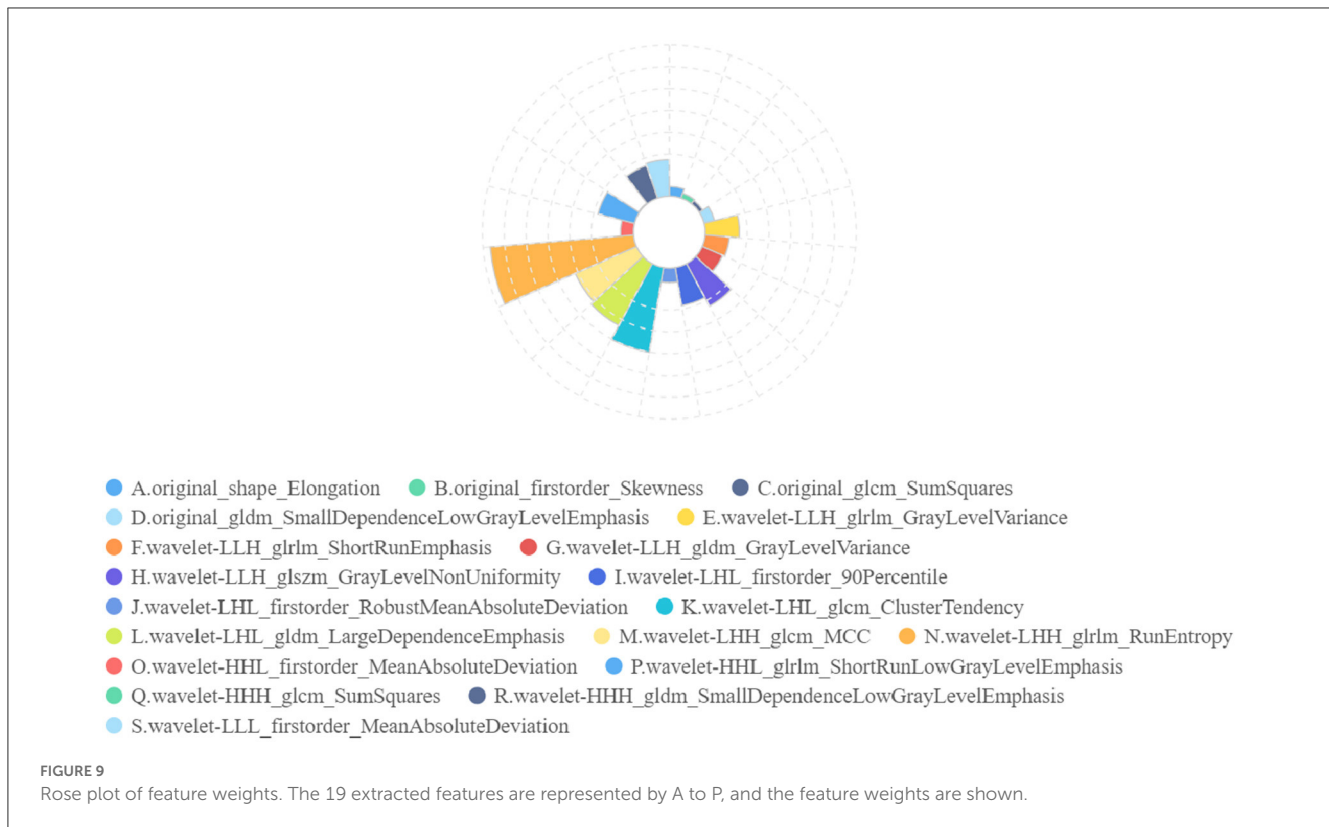TABLE 2 Clinical characteristics according to the discharge NIHSS category.

Number	Clinical factors		Group A (n = 106)	Group B (n = 289)	Group C (n = 46)
a	NIHSS on admission		0 (0, 1)	2 (1, 3)	7 (5, 9)
b	Position	Telencephalon	53 (50.00%)	122 (42.21%)	22 (47.83%)
		Diencephalon	15 (14.15%)	91 (31.49%)	15 (32.61%)
		Cerebellum	23 (21.70%)	5 (1.73%)	0 (0.00%)
		Brain stem	15 (14.15%)	71 (24.57%)	9 (19.57%)
c	OCSP typing	TACI	0 (0.00%)	2 (0.69%)	9 (19.57%)
		PACI	22 (20.75%)	116 (40.14%)	26 (56.52%)
		LACI	44 (41.51%)	79 (27.34%)	10 (21.74%)
		POCI	40 (37.74%)	92 (31.83%)	1 (2.17%)
d	Age		67.17 ± 11.63	66.92 ± 12.09	65.63 ± 14.82
e	Sex	Male	66 (62.26%)	167 (57.79%)	24 (52.17%)
		Female	40 (37.74%)	122 (42.21%)	22 (47.83%)
f	BMI		24.29 ± 3.10	24.13 ± 3.94	25.29 ± 4.15
g	SBP		160.79 ± 14.96	162.90 ± 13.97	161.85 ± 18.70
h	TC		4.48 ± 1.05	4.60 ± 1.11	4.38 ± 1.14
i	LDL		2.61 ± 0.92	2.70±1.01	2.64 ± 1.14
j	LVH	Yes	61 (57.55%)	161 (55.71%)	31 (67.39%)
		No	45 (42.45%)	128 (44.29%)	15 (32.61%)
k	HCY	Yes	10 (9.43%)	30 (10.38%)	6 (13.04%)
		No	96 (90.57%)	259 (89.62%)	40 (86.96%)
l	AF	Yes	9 (8.49%)	21 (7.27%)	10 (21.74%)
		No	97 (91.51%)	268 (92.73%)	36 (78.26%)
m	Hypertension	Yes	93 (87.74%)	265 (91.70%)	43 (93.48%)
		No	13 (12.26%)	24 (8.30%)	3 (6.52%)
n	Diabetes	Yes	35 (33.02%)	124 (42.91%)	23 (50.00%)
		No	71 (66.98%)	165 (57.09%)	23 (50.00%)
o	CHD	Yes	6 (5.66%)	17 (5.88%)	6 (13.04%
		No	100 (94.34%)	272 (94.12%)	40 (86.96%)
p	Smoking	Yes	32 (30.19%)	62 (21.45%)	13 (28.26%)
		No	74 (69.81%)	227 (78.55%)	33 (71.74%)
q	Drinking	Yes	25 (23.58%)	52 (17.99%)	8 (17.39%)
		No	81 (76.42%)	237 (82.01%)	38 (82.61%)

Groups A, B, and C refer to the normal group, mild group, and moderate-severe group, respectively. SBP, systolic blood pressure; TC, total cholesterol; LDL, low-density lipoprotein; LVH, left ventricular hypertrophy; HCY, homocysteinemia; AF, atrial fibrillation; CHD, coronary heart disease; TACI, OCSP classification, total anterior circulation infarcts; PACI, partial anterior circulation infarcts; LACI, lacunar circulation infarcts; and POCI, posterior circulation infarcts.

yields the best model performance, which is conducive to removing redundant features and avoiding the risk of overfitting. Our data collection and processing techniques can be actively promoted in future radiomics research.

Traditional methods for predicting AIS prognosis are shallow and deep machine learning methods. Their ability to represent complex problems is limited, as is their learning ability. To design a new OEDL and apply it to the prediction of AIS prognosis, we creatively combined the ideas of deep learning, integrated learning, and metaheuristic optimization. A comparison of the prediction performance of the various methods shows that the best classification performance was obtained by OEDL based on ensemble optimization, with Macro-AUC, ACC, Macro-R, Macro-P, and Macro-F1 reaching 97.89, 95.74, 94.75, 94.03, and 94.35%, respectively. The main reasons can be analyzed as follows. (1) In complex problems, the deep learning model can outperform the traditional shallow learning model in terms of feature learning ability. Deep learning multilayer networks can effectively represent

FIGURE 9
Rose plot of feature weights. The 19 extracted features are represented by A to P, and the feature weights are shown.
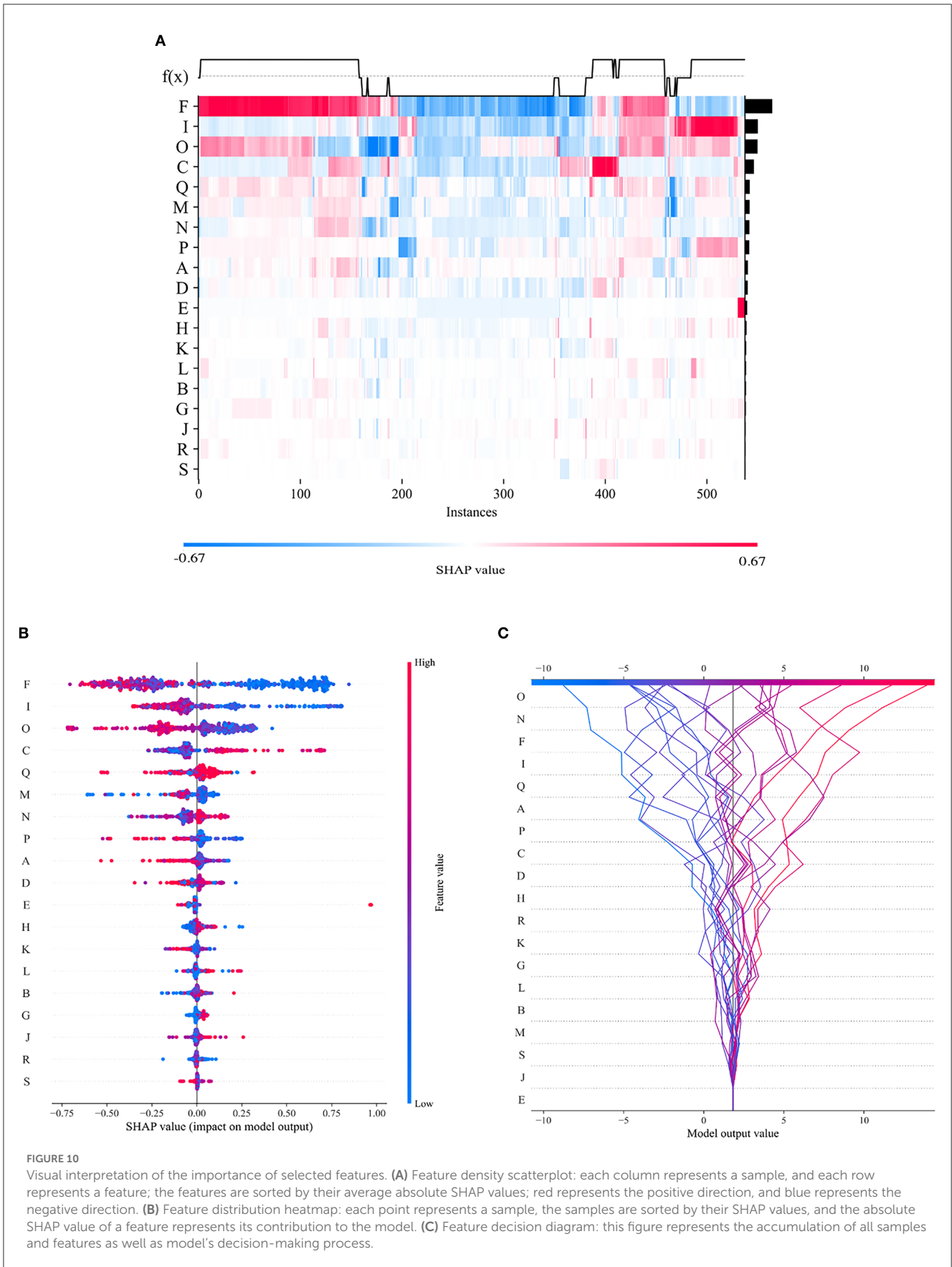
the complexity of prediction results and are adept at discovering complex relationships between a large number of input features, resulting in high prediction performance (76). (2) When compared to a single learner, the advantage of integrated learning is that it ensures classifier diversity and richness, as well as better prediction effect and stability through stacking combination (77). (3) We developed a new parameter optimization strategy based on the traditional metaheuristic algorithm to address the problem of superparameter optimization in machine learning algorithms. Our optimization algorithm can effectively avoid the problem that traditional optimization methods have of falling into a local optimal solution, and it can also effectively improve the model efficiency (78).
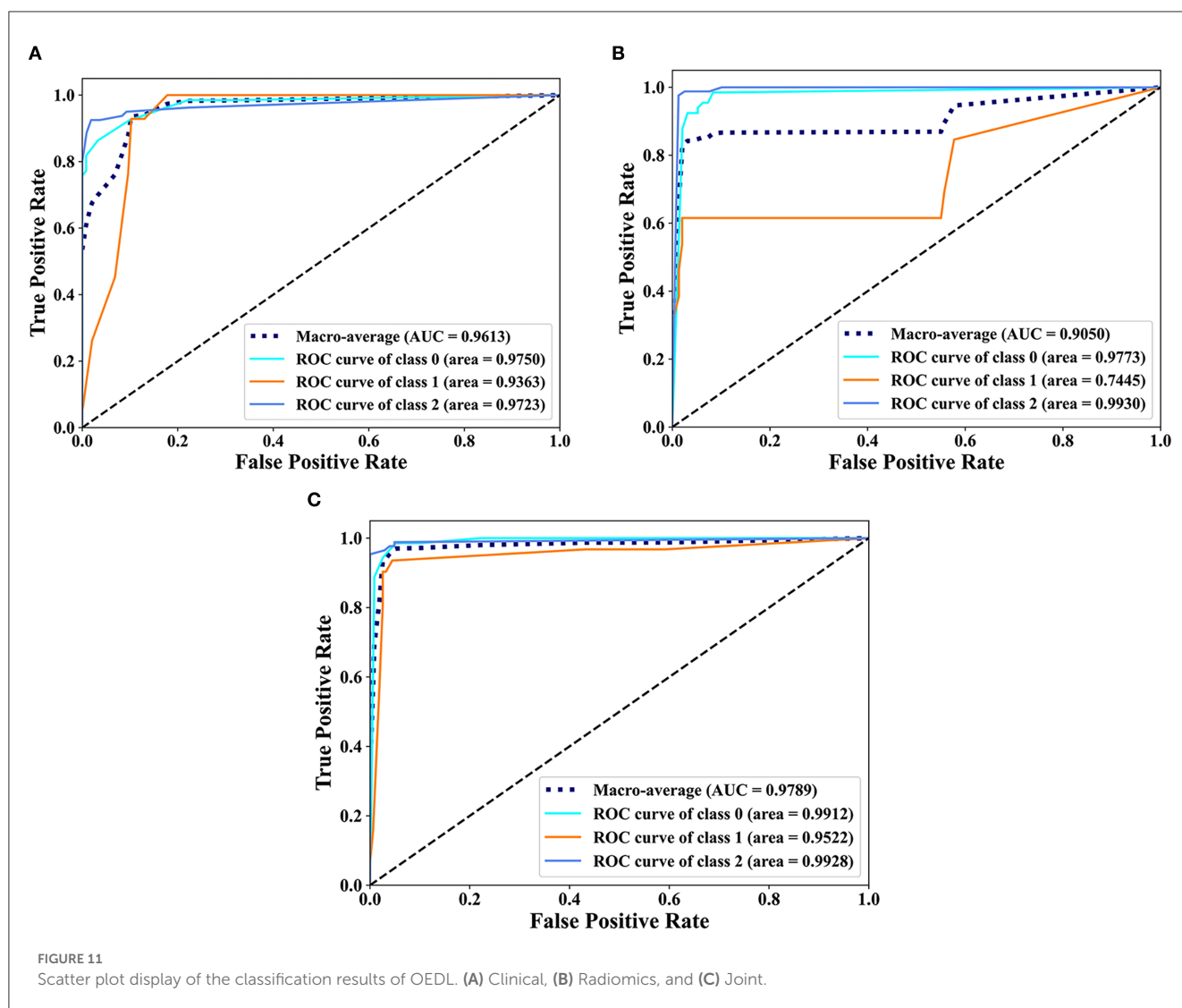
In deep learning, the quality and quantity of data have a crucial impact on the training effect of the model. If the training data are imbalanced, i.e., the number of samples in some classes is too small, then the model will be biased toward those classes with a high proportion during the training process and will perform poorly for those classes with a low proportion. This results in poor model performance with test data may lead to overfitting. Therefore, we introduced a data balancing method to ensure the balance of the training data. The classification results of various balance methods were compared. Compared with the unbalanced, oversampled, and undersampled techniques, SMOTEENN based on mixed sampling can achieve the best classification performance. The results suggest that SMOTEENN, which combines undersampling and oversampling, is the most suitable balancing technique for this study.

A comparison with actual scenarios can be described as follows. Qiu (16) used a linear SVM method to analyze the optimal imaging

group thrombus characteristics of IV protease recanalization with AIS patients on noncontrast CT (NCCT) and CT angiography and obtained (0.85 ± 0.03) ACC in the comparison of actual data. Multiple regression and machine learning models were used by Alaka (17) to predict the related dysfunction of AIS patients after intravascular therapy. Using an internal dataset, the model had an AUC of 0.65–0.72, and using an external dataset, the model had an AUC of 0.66–0.71. Hofmeister (26) investigated the predictive value of radiomics features extracted from clots on the first thrombosis recanalization using SVM, with an ACC of 0.88. Wang (24) obtained an ACC of 0.73 by using the modified Rankin scale (mRS) to predict the prognosis of AIS. Traditional methods for complex problems have limited expression and learning ability, so it is necessary to design a deep integration model with a multilevel cascade structure to improve the model's learning ability in complex problems (35). When compared to that of other single and integrated methods, OEDL can achieve the best classification performance when compared to the control method. The proposed method outperformed previous methods in terms of classification performance (AUC, ACC, etc.) and classification superiority. Furthermore, to address the problem of poor interpretability that frequently exists in deep learning (79), we used interpretable machine learning technology to understand the model's applicability to clinical prediction, with the goal of revealing the reasons behind the prediction results.

In a clinical sense, the combined feature model we developed can serve as a reliable clinical diagnostic tool for predicting stroke prognoses. Our modeling method is more suitable for the clinical model application scenario; it is convenient for radiologists to understand the differences between clinical features,

FIGURE 10
Visual interpretation of the importance of selected features. **(A)** Feature density scatterplot: each column represents a sample, and each row represents a feature; the features are sorted by their average absolute SHAP values; red represents the positive direction, and blue represents the negative direction. **(B)** Feature distribution heatmap: each point represents a sample, the samples are sorted by their SHAP values, and the absolute SHAP value of a feature represents its contribution to the model. **(C)** Feature decision diagram: this figure represents the accumulation of all samples and features as well as model's decision-making process.

FIGURE 11
Scatter plot display of the classification results of OEDL. **(A)** Clinical, **(B)** Radiomics, and **(C)** Joint.

morphological features, and high-dimensional omics features, as well as diagnostic performance differences. In addition, when building the training set, we also built a data validation set and performed in-model validation at a single center. This study confirmed the validity and scientific nature of the combined data, provided an important reference for similar subsequent studies, and facilitated further verification through the use of more external multi-center data. Compared with traditional radiomics analysis, our combined feature model could extract more statistical features, thereby providing a comprehensive stroke description. In addition, computerized tools overcome the instability of human empirical judgments, allowing clinicians to quickly and accurately predict long-term outcomes.

This study still has room for improvement. First, this study is a single-center, retrospective study with a limited sample size, and it is expected that a multi-center study with larger samples will be implemented in future to further verify the generalizability of the model. Second, lesion labeling comes from manual delineation and may be affected by the subjective judgment of investigators. Subsequent semiautomatic or fully automatic labeling algorithms need to be further explored to

improve the stability and consistency of feature extraction. Third, the radiomics features constructed in this study are based on noncontrast-enhanced MR only, requiring further advanced MR acquisitions such as contrast-enhanced DWI to obtain a high level of evidence for clinical application. Fourth, more efficient image preprocessing tools (80) need to be incorporated to improve the robustness and versatility of the method.

Conclusion

In conclusion, using a combination of clinical features and radiomics, we developed and validated a set of methods for the early prediction of stroke prognoses. We combined DNN ideas with ensemble learning to use OEDL as an effective tool for the early and non-invasive prediction of prognosis levels, thereby optimizing the clinical decision-making process and improving the efficiency of clinical intervention. The ideas in this study can provide new research directions for the effective establishment of stroke prevention and control mechanisms.

TABLE 3 Comparison of the classification effects of different methods (%).

Type	Name	Macro-AUC	ACC	Macro-R	Macro-P	Macro-F1
Machine learning	DT	90.53	75.53	59.62	50.37	54.41
	SVM	97.69	82.45	72.56	82.71	74.11
	RF	95.90	87.77	79.67	85.28	81.14
Deep learning	DNN	93.13	82.96	83.26	83.02	83.12
	LSTM-RNN	94.56	84.81	84.81	84.66	84.63
	DBN	94.39	83.70	83.70	83.61	83.61
Deep learning + Ensemble learning	Hard-voting	95.36	87.23	86.86	87.01	86.90
	Soft-voting	93.45	87.23	86.86	87.10	86.92
	EDL	96.68	92.55	92.10	91.42	91.72
OEDL	OEDL	**97.89**	**95.74**	**94.75**	**94.03**	**94.35**

EDL represents a deep ensemble learning model based on DNN, LSTM-RNN, DBN, and stacking ensemble; OEDL is an optimization algorithm based on EDL and BBOA.
These bold characters represent the predictive performance of the optimal method.

TABLE 4 Comparison of classification performance of various feature combinations (%).

Feature	Model	Macro-AUC	ACC	Macro-R	Macro-P	Macro-F1
Clinical	EDL	97.15	88.30	87.60	86.78	86.82
Radiomics	EDL	90.79	90.74	74.10	80.28	75.82
Joint	EDL	96.68	92.55	92.10	91.42	91.72
Clinical	OEDL	96.13	90.43	90.57	89.29	89.35
Radiomics	OEDL	90.50	93.21	82.19	86.27	83.87
Joint	OEDL	**97.89**	**95.74**	**94.75**	**94.03**	**94.35**

EDL represents a deep ensemble learning model based on DNN, LSTM-RNN, DBN, and stacking ensemble; OEDL is an optimization algorithm based on EDL and BBOA.
These bold characters represent the predictive performance of the optimal method.

TABLE 5 Comparison of the classification performance of various balancing methods (%).

Method	Type	Macro-AUC	ACC	Macro-R	Macro-P	Macro-F1
Original	None	92.60	81.16	85.01	74.62	78.75
Random Oversample	Oversampling	91.84	80.37	80.37	81.14	80.49
Random Undersample	Oversampling	82.78	73.81	71.82	74.52	72.47
SMOTE	Undersampling	90.98	85.93	86.10	85.85	85.85
ADASYN	Undersampling	95.74	84.84	85.05	85.35	84.28
Borderline-SMOTE	Undersampling	96.10	87.04	87.12	87.68	86.75
SMOTEENN	Mixed-sampling	97.89	95.74	94.75	94.03	94.35

TABLE 6 Comparison of classification performance with previous studies.

Authors	Data	Methods	Number of categories	AUC (%)	ACC (%)
Qiu (16)	Radiomics	SVM	2	85	-
Alaka (17)	Clinical	LR + Machine Learning	2	66–71	-
Hofmeister (26)	Clinical + Radiomics	SVM	2	88.1	-
Wang (24)	Clinical + Radiomics	LR	2	73	-
Ours	Clinical + Radiomics	OEDL	3	**97.89**	**95.74**

These bold characters represent the predictive performance of the optimal method.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the patients/participants or patients/participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

WY: conceptualization, data curation, resources, methodology, software, formal analysis, validation, investigation, and writing—original draft. XC: methodology, software, formal analysis, validation, investigation, writing—original draft, and editing and polishing. PL: data curation, resources, software, and formal analysis. YT: software, validation, investigation, and data curation. ZWa: validation, investigation, data curation, and editing and polishing. CG: investigation, resources, and writing—polishing. JC: validation and resources. FL: formal analysis and validation. DaY and ZWe: writing—polishing. DoY: conceptualization, investigation, validation,

and resources. YW: funding acquisition, conceptualization, investigation, resources, methodology, writing—review and editing, project administration, and supervision. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Wu S, Wu B, Liu M, Chen Z, Wang W, Anderson CS, et al. Stroke in China: advances and challenges in epidemiology, prevention, and management. *Lancet Neurol.* (2019) 18:394–405. doi: 10.1016/S1474-4422(18)30500-3

2. Katan M, Luft A. Global burden of stroke. *Semin Neurol.* (2018) 38:208–11. doi: 10.1055/s-0038-1649503

3. Purroy F, Montalà N. Epidemiología del ictus en la última década: revisión sistemática. *Rev Neurol.* (2021) 73:321. doi: 10.33588/rn.7309.2021138

4. Pohl M, Hesszenberger D, Kapus K, Meszaros J, Feher A, Varadi I, et al. Ischemic stroke mimics: a comprehensive review. *J Clin Neurosci.* (2021) 93:174–82. doi: 10.1016/j.jocn.2021.09.025

5. Karatzetzou S, Tsiptsios D, Terzoudi A, Aggeloussis N, Vadikolias K. Transcranial magnetic stimulation implementation on stroke prognosis. *Neurol Sci.* (2022) 43:873–88. doi: 10.1007/s10072-021-05791-1

6. Lau L-H, Lew J, Borschmann K, Thijs V, Ekinci EI. Prevalence of diabetes and its effects on stroke outcomes: a meta-analysis and literature review. *J Diabetes Investig.* (2019) 10:780–92. doi: 10.1111/jdi.12932

7. Vasudeva K, Munshi A. miRNA dysregulation in ischaemic stroke: focus on diagnosis, prognosis, therapeutic and protective biomarkers. *Eur J Neurosci.* (2020) 52:3610–27. doi: 10.1111/ejn.14695

8. Tang G, Cao Z, Luo Y, Wu S, Sun X. Prognosis associated with asymptomatic intracranial hemorrhage after acute ischemic stroke: a systematic review and meta-analysis. *J Neurol.* (2022) 269:3470–81. doi: 10.1007/s00415-022-11046-6

9. Harding-Theobald E, Louissaint J, Maraj B, Cuaresma E, Townsend W, Mendiratta-Lala M, et al. Systematic review: radiomics for the diagnosis and prognosis of hepatocellular carcinoma. *Aliment Pharmacol Ther.* (2021) 54:890–901. doi: 10.1111/apt.16563

10. Sotoudeh H, Sarrami AH, Roberson GH, Shafaat O, Sadaatpour Z, Rezaei A, et al. Emerging applications of radiomics in neurological disorders: a review. *Cureus.* (2021) 13:e20080. doi: 10.7759/cureus.20080

11. Wang X, Xie T, Luo J, Zhou Z, Yu X, Guo X. Radiomics predicts the prognosis of patients with locally advanced breast cancer by reflecting the heterogeneity of tumor cells and the tumor microenvironment. *Breast Cancer Res.* (2022) 24:20. doi: 10.1186/s13058-022-01516-0

12. Chen Q, Xia T, Zhang M, Xia N, Liu J, Yang Y. Radiomics in stroke neuroimaging: techniques, applications, and challenges. *Aging Dis.* (2021) 12:143. doi: 10.14336/AD.2020.0421

13. Hatami N, Cho T-H, Mechtouff L, Eker OF, Rousseau D, Frindel C. CNN-LSTM based multimodal MRI and clinical data fusion for predicting functional outcome in stroke patients. *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC).* Glasgow, Scotland, United Kingdom: IEEE (2022). p. 3430–4. doi: 10.1109/EMBC48229.2022.9871735

14. Sarioglu O, Sarioglu FC, Capar AE, Sokmez DF, Mete BD, Belet U. Clot-based radiomics features predict first pass effect in acute ischemic stroke. *Interv Neuroradiol.* (2022) 28:160–8. doi: 10.1177/15910199211019176

15. Wen X, Li Y, He X, Xu Y, Shu Z, Hu X, et al. Prediction of malignant acute middle cerebral artery infarction via computed tomography radiomics. *Front Neurosci.* (2020) 14:708. doi: 10.3389/fnins.2020.00708

16. Qiu W, Kuang H, Nair J, Assis Z, Najm M, McDougall C, et al. Radiomics-based intracranial thrombus features on CT and CTA predict recanalization with intravenous alteplase in patients with acute ischemic stroke. *AJNR Am J Neuroradiol.* (2019) 40:39–44. doi: 10.3174/ajnr.A5918

17. Alaka SA, Menon BK, Brobbey A, Williamson T, Goyal M, Demchuk AM, et al. Functional outcome prediction in ischemic stroke: a comparison of machine learning algorithms and regression models. *Front Neurol.* (2020) 11:889. doi: 10.3389/fneur.2020.00889

18. Dashtbani Moghari M, Zhou L, Yu B, Young N, Moore K, Evans A, et al. Efficient radiation dose reduction in whole-brain CT perfusion imaging using a 3D GAN: performance and clinical feasibility. *Phys Med Biol.* (2021) 66:075008. doi: 10.1088/1361-6560/abe917

19. Kriegeskorte N, Golan T. Neural network models and deep learning. *Curr Biol.* (2019) 29:R231–6. doi: 10.1016/j.cub.2019.02.034

20. Liu X, Liu C, Huang R, Zhu H, Liu Q, Mitra S, et al. Long short-term memory recurrent neural network for pharmacokinetic-pharmacodynamic modeling. *CP.* (2021) 59:138–46. doi: 10.5414/CP203800

21. Chen X, Li T-H, Zhao Y, Wang C-C, Zhu C-C. Deep-belief network for predicting potential miRNA-disease associations. *Brief Bioinformatics.* (2021) 22:bbaa186. doi: 10.1093/bib/bbaa186

22. Wu O, Winzeck S, Giese A-K, Hancock BL, Etherton MR, Bouts MJRJ, et al. Big data approaches to phenotyping acute ischemic stroke using automated lesion segmentation of multi-center magnetic resonance imaging data. *Stroke.* (2019) 50:1734–41. doi: 10.1161/STROKEAHA.119.025373

23. Poirion OB, Jing Z, Chaudhary K, Huang S, Garmire LX. DeepProg: an ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data. *Genome Med.* (2021) 13:112. doi: 10.1186/s13073-021-00930-x

24. Wang H, Sun Y, Ge Y, Wu P-Y, Lin J, Zhao J, et al. Clinical-radiomics nomogram for functional outcome predictions in ischemic stroke. *Neurol Ther.* (2021) 10:819–32. doi: 10.1007/s40120-021-00263-2

25. Tang N, Zhang R, Wei Z, Chen X, Li G, Song Q, et al. Improving the performance of lung nodule classification by fusing structured and unstructured data. *Information Fusion.* (2022) 88:161–74. doi: 10.1016/j.inffus.2022.07.019

26. Hofmeister J, Bernava G, Rosi A, Vargas MI, Carrera E, Montet X, et al. Clot-based radiomics predict a mechanical thrombectomy strategy for successful recanalization in acute ischemic stroke. *Stroke.* (2020) 51:2488–94. doi: 10.1161/STROKEAHA.120.030334

27. Bahaddad AA, Ragab M, Ashary EB, Khalil EM. Metaheuristics with deep learning-enabled Parkinson's disease diagnosis and classification model. *J Healthc Eng.* (2022) 2022:9276579. doi: 10.1155/2022/9276579

28. Gao C, Zhang R, Chen X, Yao T, Song Q, Ye W, et al. Integrating Internet multisource big data to predict the occurrence and development of COVID-19 cryptic transmission. *NPJ Digit Med* (2022) 5:161. doi: 10.1038/s41746-022-00704-8

29. Dehghani M, Trojovská E, Zuščák T. A new human-inspired metaheuristic algorithm for solving optimization problems based on mimicking sewing training. *Sci Rep.* (2022) 12:17387. doi: 10.1038/s41598-022-22458-9

30. Kumar Pandey R, Aggarwal S, Nath G, Kumar A, Vaferi B. Metaheuristic algorithm integrated neural networks for well-test analyses of petroleum reservoirs. *Sci Rep.* (2022) 12:16551. doi: 10.1038/s41598-022-21075-w

31. Alshareef AM, Alsini R, Alsieni M, Alrowais F, Marzouk R, Abunadi I, et al. Optimal deep learning enabled prostate cancer detection using microarray gene expression. *J Healthc Eng.* (2022) 2022:7364704. doi: 10.1155/2022/7364704

32. Nematzadeh S, Kiani F, Torkamanian-Afshar M, Aydin N. Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: a bioinformatics study on biomedical and biological cases. *Comput Biol Chem.* (2022) 97:107619. doi: 10.1016/j.compbiolchem.2021.107619

33. Larsen K, Jæger HS, Hov MR, Thorsen K, Solyga V, Lund CG, et al. Streamlining acute stroke care by introducing national institutes of health stroke scale in the emergency medical services: a prospective cohort study. *Stroke.* (2022) 53:2050–7. doi: 10.1161/STROKEAHA.121.036084

34. Elsaid N, Bigliardi G, Dell'Acqua ML, Vandelli L, Ciolli L, Picchetto L, et al. Evaluation of stroke prognostication using age and NIH Stroke Scale index (SPAN-100 index) in delayed intravenous thrombolysis patients (beyond 45 hours). *J Stroke Cerebrovasc Dis.* (2022) 31:106384. doi: 10.1016/j.jstrokecerebrovasdis.2022.106384

35. Ye W, Zhang L, Zhang W, Wu X, Yi D, Wu Y, et al. Comparison of single imputation and multiple imputation methods for missing data in different oncogene expression profiles. *Biostat Epidemiol.* (2022) 6:113–27. doi: 10.1080/24709360.2021.2023805

36. Staartjes VE, Kernbach JM, Stumpo V, van Niftrik CHB, Serra C, Regli L. Foundations of feature selection in clinical prediction modeling. *Acta Neurochir Suppl.* (2022) 134:51–7. doi: 10.1007/978-3-030-85292-4_7

37. Bommert A, Welchowski T, Schmid M, Rahnenführer J. Benchmark of filter methods for feature selection in high-dimensional gene expression survival data. *Brief Bioinform.* (2022) 23:bbab354. doi: 10.1093/bib/bbab354

38. Riaz S, Latif S, Usman SM, Ullah SS, Algarni AD, Yasin A, et al. Malware detection in internet of things (IoT) devices using deep learning. *Sensors.* (2022) 22:9305. doi: 10.3390/s22239305

39. Diao X, Huo Y, Zhao S, Yuan J, Cui M, Wang Y, et al. Automated ICD coding for primary diagnosis via clinically interpretable machine learning. *Int J Med Inform.* (2021) 153:104543. doi: 10.1016/j.ijmedinf.2021.104543

40. Salvi M, Michielli N, Molinari F. Stain color adaptive normalization (SCAN) algorithm: separation and standardization of histological stains in digital pathology. *Comput Methods Programs Biomed.* (2020) 193:105506. doi: 10.1016/j.cmpb.2020.105506

41. Dorgalaleh A, Favaloro EJ, Bahraini M, Rad F. Standardization of prothrombin time/international normalized ratio (PT/INR). *Int J Lab Hematol.* (2021) 43:21–8. doi: 10.1111/ijlh.13349

42. Miyoshi T, Tanaka H. Standardization of normal values for cardiac chamber size in echocardiography. *J Med Ultrason.* (2001) 49:21–33. doi: 10.1007/s10396-021-01147-6

43. Risso D. Normalization of single-cell RNA-seq data. *Methods Mol Biol.* (2021) 2284:303–29. doi: 10.1007/978-1-0716-1307-8_17

44. Wang S-F, Ku F-L, Zhang H-X. Improved PCA facial recognition with bootstrap and data standardization in small sample case. *Proceedings of 2011 International Conference on Computer Science and Network Technology.* Harbin, China: IEEE (2011). p. 2618–22. doi: 10.1109/ICCSNT.2011.6182504

45. Zeng X. Length of stay prediction model of indoor patients based on light gradient boosting machine. *Comput Intell Neurosci.* (2022) 2022:9517029. doi: 10.1155/2022/9517029

46. Zhang C, Lei X, Liu L. Predicting metabolite-disease associations based on lightGBM model. *Front Genet.* (2021) 12:660275. doi: 10.3389/fgene.2021.660275

47. Du Z, Zhong X, Wang F, Uversky VN. Inference of gene regulatory networks based on the Light Gradient Boosting Machine. *Comput Biol Chem.* (2022) 101:107769. doi: 10.1016/j.compbiolchem.2022.107769

48. Zeng H, Yang C, Zhang H, Wu Z, Zhang J, Dai G, et al. LightGBM-based EEG analysis method for driver mental states classification. *Comput Intell Neurosci.* (2019) 2019:3761203. doi: 10.1155/2019/3761203

49. Hu X, Yin S, Zhang X, Menon C, Fang C, Chen Z, et al. Blood pressure stratification using photoplethysmography and light gradient boosting machine. *Front Physiol.* (2023) 14:1072273. doi: 10.3389/fphys.2023.1072273

50. Gan L. XGBoost-based e-commerce customer loss prediction. *Comput Intell Neurosci.* (2022) 2022:1858300. doi: 10.1155/2022/1858300

51. Shin H. XGBoost regression of the most significant photoplethysmogram features for assessing vascular aging. *IEEE J Biomed Health Inform.* (2022) 26:3354–61. doi: 10.1109/JBHI.2022.3151091

52. Song X, Zhu J, Tan X, Yu W, Wang Q, Shen D, et al. XGBoost-based feature learning method for mining COVID-19 novel diagnostic markers. *Front Public Health.* (2022) 10:926069. doi: 10.3389/fpubh.2022.926069

53. Zhang J, Ma X, Zhang J, Sun D, Zhou X, Mi C, et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. *J Environ Manage.* (2023) 332:117357. doi: 10.1016/j.jenvman.2023.117357

54. Khan MS, Salsabil N, Alam MGR, Dewan MAA, Uddin MZ. CNN-XGBoost fusion-based affective state recognition using EEG spectrogram image analysis. *Sci Rep.* (2022) 12:14122. doi: 10.1038/s41598-022-18257-x

55. Eslami T, Saeed F. Fast-GPU-PCC: a GPU-based technique to compute pairwise pearson's correlation coefficients for time series data-fMRI study. *High Throughput.* (2018) 7:11. doi: 10.3390/ht7020011

56. Lu L-L, Hu X-J, Yang Y, Xu S, Yang S-Y, Zhang C-Y, et al. Correlation of myopia onset and progression with corneal biomechanical parameters in children. *World J Clin Cases.* (2022) 10:1548–56. doi: 10.12998/wjcc.v10.i5.1548

57. Dickinson Q, Meyer JG. Positional SHAP (PoSHAP) for interpretation of machine learning models trained from biological sequences. *PLoS Comput Biol.* (2022) 18:e1009736. doi: 10.1371/journal.pcbi.1009736

58. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed.* (2022) 214:106584. doi: 10.1016/j.cmpb.2021.106584

59. Wu Y, Fang Y. Stroke prediction with machine learning methods among older Chinese. *Int J Environ Res Public Health.* (2020) 17:1828. doi: 10.3390/ijerph17061828

60. Wang B, Zhang C, Wong YD, Hou L, Zhang M, Xiang Y. Comparing resampling algorithms and classifiers for modeling traffic risk prediction. *Int J Environ Res Public Health.* (2022) 19:13693. doi: 10.3390/ijerph192013693

61. Xing M, Zhang Y, Yu H, Yang Z, Li X, Li Q, et al. Predict DLBCL patients' recurrence within two years with Gaussian mixture model cluster oversampling and multi-kernel learning. *Comput Methods Programs Biomed.* (2022) 226:107103. doi: 10.1016/j.cmpb.2022.107103

62. Sajjadian M, Lam RW, Milev R, Rotzinger S, Frey BN, Soares CN, et al. Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis. *Psychol Med.* (2021) 51:2742–51. doi: 10.1017/S0033291721003871

63. Xu Y, Ye W, Song Q, Shen L, Liu Y, Guo Y, et al. Using machine learning models to predict the duration of the recovery of COVID-19 patients hospitalized in Fangcang shelter hospital during the Omicron BA. 22 pandemic. *Front Med (Lausanne).* (2022) 9:1001801. doi: 10.3389/fmed.2022.1001801

64. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* (1997) 9:1735–80. doi: 10.1162/neco.1997.9.8.1735

65. Gers FA, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM. *Neural Comput.* (2000) 12:2451–71. doi: 10.1162/089976600300015015

66. Attallah O. A deep learning-based diagnostic tool for identifying various diseases via facial images. *Digital Health.* (2022) 8:205520762211244. doi: 10.1177/20552076221124432

67. Yan J, Cai X, Chen S, Guo R, Yan H, Wang Y. Ensemble learning-based pulse signal recognition: classification model development study. *JMIR Med Inform.* (2021) 9:e28039. doi: 10.2196/28039

68. Kim J, Kang U, Lee Y. Statistics and deep belief network-based cardiovascular risk prediction. *Healthc Inform Res.* (2017) 23:169. doi: 10.4258/hir.2017.23.3.169

69. Mahbobi M, Kimiagari S, Vasudevan M. Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks. *Ann Oper Res.* (2021). doi: 10.1007/s10479-021-04114-z

70. Abdel-Nasser M, Mahmoud K. Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Comput & Applic.* (2019) 31:2727–40. doi: 10.1007/s00521-017-3225-z

71. Zhang J, Sheng J, Lu J, Shen L, UCPSO. A uniform initialized particle swarm optimization algorithm with cosine inertia weight. *Comput Intell Neurosci.* (2021) 2021:1–18. doi: 10.1155/2021/8819333

72. Mirjalili S, Gandomi AH. Chaotic gravitational constants for the gravitational search algorithm. *Appl Soft Comput.* (2017) 53:407–19. doi: 10.1016/j.asoc.2017.01.008

73. Zhu H, Ge J, Qi W, Zhang X, Lu X. Dynamic analysis and image encryption application of a sinusoidal-polynomial composite chaotic system. *Math Comput Simul.* (2022) 198:188–210. doi: 10.1016/j.matcom.2022.02.029

74. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol.* (2018) 63:07TR01. doi: 10.1088/1361-6560/aab4b1

75. Ma Q, Li R, Wang L, Yin P, Wang Y, Yan C, et al. Temporal trend and attributable risk factors of stroke burden in China, 1990–2019: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health.* (2021) 6:e897–906. doi: 10.1016/S2468-2667(21)00228-0

76. Muhammad Usman S, Khalid S, Bashir S. A deep learning based ensemble learning method for epileptic seizure prediction. *Comput Biol Med.* (2021) 136:104710. doi: 10.1016/j.compbiomed.2021.104710

77. Cheng L-H, Hsu T-C, Lin C. Integrating ensemble systems biology feature selection and bimodal deep neural network for breast cancer prognosis prediction. *Sci Rep.* (2021) 11:14914. doi: 10.1038/s41598-021-92864-y

78. Zhang R, Gao C, Chen X, Li F, Yi D, Wu Y. Genetic algorithm optimised Hadamard product method for inconsistency judgement matrix adjustment in AHP and automatic analysis system development. *Expert Syst Appl.* (2023) 211:118689. doi: 10.1016/j.eswa.2022.118689

79. Ogami C, Tsuji Y, Seki H, Kawano H, To H, Matsumoto Y, et al. An artificial neural network–pharmacokinetic model and its interpretation using Shapley additive explanations. *CPT Pharmacometrics Syst Pharmacol.* (2021) 10:760–8. doi: 10.1002/psp4.12643

80. Versaci M, Angiulli G, di Barba P, Morabito FC. Joint use of eddy current imaging and fuzzy similarities to assess the integrity of steel plates. *Open Physics.* (2020) 18:230–40. doi: 10.1515/phys-2020-0159