Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 19(10): 10407–10423.
DOI: 10.3934/mbe.2022487
Received: 22 April 2022
Revised: 23 June 2022
Accepted: 03 July 2022
Published: 22 July 2022

Research article

Prediction of postoperative recovery in patients with acoustic neuroma using machine learning and SMOTE-ENN techniques

Jianing Wang*

School of Mathematics and Statistics, Central South University, Changsha 410083, China

* **Correspondence:** Email: wangjianinghenu@163.com.

Abstract: Acoustic neuroma is a common benign tumor that is frequently associated with postoperative complications such as facial nerve dysfunction, which greatly affects the physical and mental health of patients. In this paper, clinical data of patients with acoustic neuroma treated with microsurgery by the same operator at Xiangya Hospital of Central South University from June 2018 to March 2020 are used as the study object. Machine learning and SMOTE-ENN techniques are used to accurately predict postoperative facial nerve function recovery, thus filling a gap in auxiliary diagnosis within the field of facial nerve treatment in acoustic neuroma. First, raw clinical data are processed and dependent variables are identified based on clinical context and data characteristics. Secondly, data balancing is corrected using the SMOTE-ENN technique. Finally, XGBoost is selected to construct a prediction model for patients' postoperative recovery, and is also compared with a total of four machine learning models, LR, SVM, CART, and RF. We find that XGBoost can most accurately predict the postoperative facial nerve function recovery, with a prediction accuracy of 90.0% and an AUC value of 0.90. CART, RF, and XGBoost can further select the more important preoperative indicators and provide therapeutic assistance to physicians, thereby improving the patient's postoperative recovery. The results show that machine learning and SMOTE-ENN techniques can handle complex clinical data and achieve accurate predictions.

Keywords: XGBoost; SMOTE-ENN; machine learning; acoustic neuroma; postoperative recovery

1. Introduction

Acoustic neuroma (AN) is the most common benign tumor of the internal auditory tract and the cerebellopontine horn region, accounting for approximately 90% of cerebellopontine horn tumors and 8% of intracranial tumors in adults, for which surgery is the primary therapy [1]. With the advancement of microsurgery and improved diagnosis in recent years, postoperative mortality of acoustic neuroma has gradually decreased. The goal of treatment has shifted from saving the patient's life to preserving

facial and acoustic nerve function and improving quality of life [2]. It is still difficult to maximize the patient's facial nerve anatomy and postoperative function during surgery. Before undergoing surgery, the patient's preoperative indications must be obtained through relevant tests, and a treatment plan must be developed. Because of the large individual variability among patients and their preoperative indices, the treatment plan and postoperative recovery of facial nerve function vary greatly. Investigating the relationship between preoperative indicators and postoperative recovery, as well as predicting the recovery, can assist physicians in developing individualized treatment approaches, surgical plans, and prognostic measures. This necessitates an accurate prediction of postoperative facial nerve function recovery. Further, it can improve the patient's recovery and quality of life.

The study of tumor-related diseases has long been a priority in clinical medicine. With the advancement of computer technology, machine learning as an artificial intelligence science has become increasingly popular in biomedical fields such as clinical diagnosis, precision treatment, and tumor health monitoring [3]. The opening of the Artificial Intelligence in Medicine (AIME) conference in 1985 brought computer science, medicine, and biology even closer together, as well as the realization that computers' computational power could solve more clinical medical problems [4,5]. In comparison to traditional statistical methods, machine learning is more concerned with the model predictive ability and generalization ability, so it has precision and accuracy that other models don't have. Machine learning can provide more accurate diagnostic algorithms and postoperative predictions in the study of tumor-related diseases [6–11]. Among them, machine learning models such as logistic regression, support vector machine, decision tree and ensemble models are widely used due to their excellent performance and high accuracy [12–16].

XGBoost, a classical boosting machine learning model, has been widely used in the biomedical field since it was proposed in 2016 due to its extremely high accuracy and excellent properties [17–21]. In 2019, Fu et al. [22] used XGBoost to construct a prognostic model framework for predicting invasive disease-free survival (iDFS) in early-stage breast cancer patients. Experiments demonstrated its very competitive performance and helped physicians to develop treatment plans that may prolong patient survival. In 2020, Li et al. [23] constructed an orthopedic auxiliary classification prediction model based on XGBoost. The experiments demonstrated its ability to cope with complex and diverse medical data and better meet the requirements of timeliness and accuracy of ancillary diagnosis. In 2021, Hsiao et al. [24] used an improved XGBoost model to predict the risk of death due to ovarian cancer. Compared with other methods, the model has improved sensitivity in classifying patients for risk and helps to optimize the treatment of high-risk patients.

The studies mentioned above show that machine learning techniques are becoming more mature in tumor diagnosis and prognosis. Some traditional statistical methods, such as regression analysis and significance tests, are currently being used in studies of postoperative facial nerve function recovery in patients with AN. They are typically used to determine whether specific indicators effect the recovery, whereas machine learning techniques are less commonly used [25]. Therefore, in this study, machine learning techniques are considered to predict the recovery. After collecting the raw clinical data and quantifying them, data balancing is corrected using SMOTE-ENN technique. XGBoost is selected to construct a prediction model for the recovery, which is also compared with a total of four machine learning models: logistic regression, support vector machine, decision tree, and random forest. Among them, logistic regression is a traditional statistical method, and not exclusively a machine learning tool. It is chosen to better compare with other classical machine learning models. In terms of each

evaluation criteria, XGBoost outperforms other models. The prediction model based on XGBoost can accurately predict the recovery of patients. Further, based on the importance of the features, preoperative indicators that have a significant impact on the recovery are sought. In addition, a prognostic model framework is built to accurately predict the recovery.

The main contributions of the study are listed as follows:

1) For the first time, machine learning techniques are applied to the study of postoperative facial nerve function recovery in patients with AN. We break the limitations of traditional studies and use machine learning techniques to achieve accurate prediction of the postoperative recovery of patients.

2) Completing the collection and processing of complex clinical data. The data of AN is characterized by difficult collection, unbalanced data, and a small sample size. We finish the collection and processing of data, using SMOTE-ENN technology for data correction.

3) Filling the gap of auxiliary diagnosis within the field of AN facial nerve treatment. The XGBoost prediction model can accurately predict the postoperative recovery of patients and assist doctors in developing personalized treatment plans. This helps to form a synergy of existing medical information and further promotes the development of smart medical care.

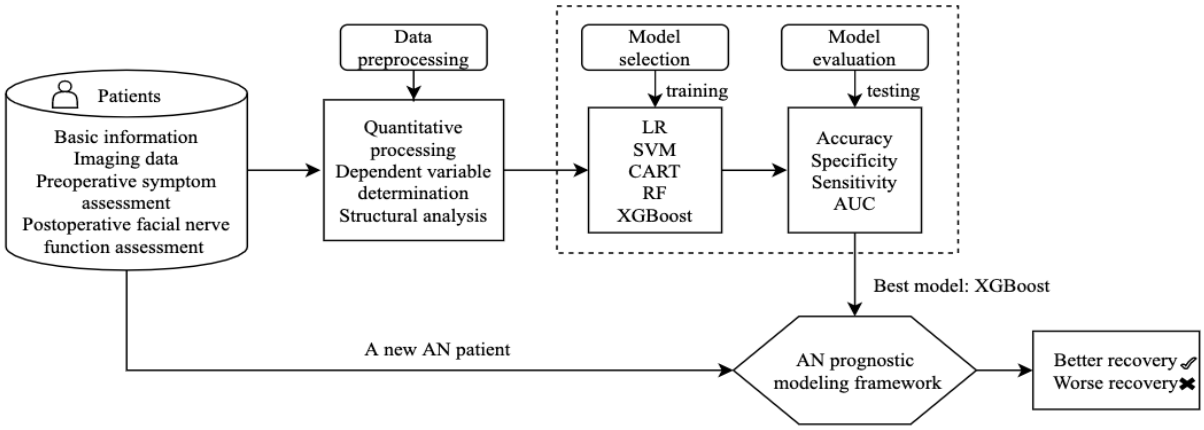

Figure 1. A prognostic model framework of postoperative recovery in AN patients.

2. Materials

2.1. Data source

The data are provided by the neurosurgery department at Central South University's Xiangya Hospital. They are collected retrospectively from patients who had microsurgery performed by the same operator between June 2018 and March 2020. Patients' data include basic patient information, tumor image data, preoperative patient symptom assessment, and postoperative facial nerve function assessment, which is obtained 6 months after surgery via follow-up visit. Due to patient privacy concerns and a follow-up walkout, 128 patients' data are included in our study. The study is approved by the Research Ethics Committee of Xiangya Hospital, and an ethical certificate for the use of human subjects is obtained.

The uniform inclusion criteria are:

1) Age is greater than 16 years and less than 80 years.

2) Pathology is proven to be AN.

3) Preoperative image suggests that the tumor originated from the auditory nerve, and surgery confirms that the tumor originated from the auditory nerve Schwann cells.

2.2. Data quantification

The data in this work include age, gender, tumor size, tumor nature, tumor with or without brainstem compression, internal auditory tract shape, Samii grading, TFIAC grading, preoperative neurological grading, presence of cerebellar symptoms, presence of posterior group neurological symptoms, and postoperative facial nerve grading, for a total of 12 variables.

The variables must then be quantified. Continuous variables such as age and tumor size don't require any special processing, whereas qualitative variables are a little more complicated. They must be quantified one by one based on their characteristics. The gender is divided into two categories, with the female being 0 and male being 1. The tumor nature is classified as solid or other, with solid being 0 and other being 1. The tumor with or without brainstem compression, presence of cerebellar symptoms, and presence of posterior group neurological symptoms are all dichotomous variables, with absence defined as 0 and presence defined as 1. The internal auditory tract shape is divided into physiological shape and other shapes, denoted by the numbers 0 and 1. Samii grading is divided into T1, T2, T3a, T3b, T4a, and T4b, and is inscribed as 1, 2, 3, 4, 5, and 6. TFIAC grading is divided into I, II, III, and IV, and is denoted as 1, 2, 3, and 4. The preoperative and postoperative facial nerve gradings are I, II, III, IV, V, and VI, and are denoted as 1, 2, 3, 4, 5, and 6.

Table 1. House-Brackman grading.

Grading	Function grading	Grading criteria
Grade I	Normal	Normal state of facial muscle movement in all regions.
Grade II	Mild abnormal	Mild facial muscle weakness, mild asymmetry of the corners of the mouth during motor status.
Grade III	Moderate abnormal	Obvious facial muscle weakness, mild asymmetry of the corners of the mouth when using maximum force in the motor state.
Grade IV	Moderate to severe abnormal	Obvious facial deformation or facial muscle weakness, and obvious asymmetry in the corners of the mouth when using maximum force in the motor state.
Grade V	Severe abnormal	Only extremely subtle facial movements are visible to the naked eye, and when in motion, the eyes cannot be completely closed after exertion, and the corners of the mouth can move slightly.
Grade VI	Completely abnormal	No movement at all.

2.3. Dependent variable determination

Our purpose is to accurately predict patients' postoperative recovery of facial nerve function. The recovery corresponded to patients' facial nerve grading at the time of follow-up, with the grading criteria referring to the House-Brackmann (H-B) scoring system (Table 1). The smaller the postoperative

facial nerve grading of the patient, the more complete the facial nerve function is preserved.

It is widely assumed that when patients' facial nerve gradings are grade I, their facial nerve function retention is more complete, and their postoperative facial nerve recovery is better. When the grading of the facial nerve are grade II, III, IV, V, and VI, the facial nerve function retention is incomplete, and the recovery is poor. As a result, the dependent variable can be identified as the patient's postoperative facial nerve recovery: better or worse, denoted by 0 and 1, respectively.

2.4. Data Structure

We have processed all data materials up to this point, and there are 11 independent variables and 1 dependent variable, with 2 continuous and 9 qualitative variables included in the independent variables. The dependent variable is qualitative that reflects the patients' postoperative facial nerve recovery (Table 2).

Table 2. Comparison of variables.

Variable Name	Description	Range of value
Facial nerve function recovery	Qualitative Variable (2 levels)	Better / Worse
Patient gender	Qualitative Variable (2 levels)	Male / Female
Patient age	Continuous Variable (years)	17–73
Tumor size	Continuous Variable (cm^3)	0.585–93.897
tumor nature	Qualitative Variable (2 levels)	Solid / Other
Tumor with brainstem compression	Qualitative Variable (2 levels)	Yes / No
Internal auditory tract shape	Qualitative Variable (2 levels)	Physiological shape / Others
TFIAC grading	Qualitative Variable (4 levels)	Grade I - Grade IV
Samii grading	Qualitative Variable (6 levels)	Grade T1–Grade T4b
Cerebellar symptoms resence	Qualitative Variable (2 levels)	Yes / No
Posterior group neurological symptoms	Qualitative Variable (2 levels)	Yes / No
Preoperative neurological grading	Qualitative Variable (6 levels)	Grade I–Grade VI

3. Methods

3.1. SMOTE-ENN

The problem of data imbalance frequently arises in practical medical data, which means that the number of samples in different categories in the dataset varies greatly [26]. Methods for dealing with data imbalance are broadly classified as algorithmic and dataic [27]. Algorithm-level methods frequently encounter new problems in their application. Data-level methods primarily employ sampling techniques to reconstruct data in order to change the data sample distribution. Oversampling and undersampling are two common sampling techniques.

If there is a large difference in the number of samples from different categories in a dataset S, it is considered unbalanced. Typically, sampling techniques can be used to solve this problem. Chawla proposed SMOTE in 2002 as a classic comprehensive minority class oversampling algorithm [28]. The primary goal of this algorithm is to increase the number of minority class samples through linear interpolation, thereby balancing the dataset. The following are the specific steps.

Step 1: For the imbalanced dataset, divide it into majority class S_{maj} and minority class S_{min}.

Step 2: For each minority class sample, calculate its K-nearest neighbors.

Step 3: Based on the imbalance ratio of the dataset, determine the number N of new samples to be synthesized for each minority class sample. For each minority class sample, choose N nearest neighbors at random from its K-nearest neighbors. If the nearest neighbor is chosen in the process of synthesizing a new sample x_n from sample x, then the new sample is built according to equation $x_{new} = x + rand(0, 1) * (x_n - x)$.

Since the SMOTE algorithm proved to be prone to problems such as sample overlap and noisy samples, the SMTOE-ENN algorithm was proposed by Batista et al. [29]. It is based on the SMOTE algorithm and uses the ENN algorithm to achieve further cleaning of the data. The SMTOE-ENN algorithm has been shown to outperform other classical sampling methods in many fields [30, 31].

3.2. Logistic regression

Logistic regression is a classical classification model that uses regression ideas to solve classification problems. A Sigmoid function is introduced to transform a regression problem into a classification problem for a general dichotomous classification problem based on linear regression $y = wTx + b$. The general formulation of the logistic regression problem is as follows:

$$y(x) = \frac{1}{1 + e^{w^T x+b}}, \tag{3.1}$$

where $y(x)$ is the label value returned by the logistic regression, w is the regression coefficient, and b is the constant term. Because the values of $y(x)$ are between [0,1], the sum of $y(x)$ and $1 - y(x)$ must be 1. From this, the binary logistic regression expression can be derived:

$$p(y = 0|x) = \frac{1}{1 + e^{w^T x+b}}, \tag{3.2}$$

$$p(y = 1|x) = \frac{1}{1 + e^{w^T x+b}}, \tag{3.3}$$

the loss function of the logistic regression algorithm is obtained using the great likelihood method to estimate the parameters w and b. The solution of its parameters is equivalent to the minimization of the loss function:

$$min\ l(w, b) = min \sum_{i=1}^{N} [-y_i(w^T x + b) + ln(1 + e^{w^T x+b})] \tag{3.4}$$

3.3. Support vector machine

The main idea of the support vector machine is to establish an optimal decision hyperplane that minimizes the distance between the closest two classes of samples. To solve the classification problem more effectively, it is also known as interval maximization. The linear equation is frequently used to describe the hyperplane in the sample space:

$$w^T x + b = 0, \tag{3.5}$$

where w is the normal vector that determines the hyperplane's direction. The displacement term, b, determines the distance between the hyperplane and the origin. The hyperplane is determined by w and

b. Support vectors are the few samples closest to the hyperplane, and the sum of the distances from two dissimilar support vectors to the hyperplane is:

$$\gamma = \frac{2}{\|w\|}, \tag{3.6}$$

this sum of distances is also made the interval, and $\|w\|$ is the L_2-parametrization of w. To find the maximum interval dividing the hyperplane, that is, to find the parameters w and b that maximize γ, the objective function for solving the optimal hyperplane can then be written as:

$$\begin{cases} min_{w,b} \quad \frac{1}{2}\|w\|^2 \\ s.t. \quad y_i(w^T x + b) \geq 1, i = 1, 2, ..., m \end{cases} \tag{3.7}$$

3.4. Decision tree

A decision tree is a classical tree model in machine learning. It can extract decision rules from a dataset with features and labels and present them in a tree structure, which is commonly used to solve classification and regression problems. A decision tree is typically built in three stages: feature selection, decision tree generation, and decision tree pruning. The key to classification decision tree algorithms is determining the best partitioning attributes. As the partitioning process progresses, it is natural to anticipate that the samples contained in each branch node will belong to the same category as much as possible. In other words, the "purity" of the nodes is increasing over time. The most commonly used indicators for measuring purity are information gain, information gain rate, and Gini index.

Information entropy is one of the most commonly used metrics to measure the purity of D sample set, assuming that the proportion of class k samples in the sample set D is p^k. Then information entropy can be defined as:

$$E(D) = \sum_{k=1}^{K} p_k log_2 p_k, \tag{3.8}$$

the smaller the value, the higher the purity. From this, the concept of information gain can be further introduced. Assuming that the discrete attribute a has V possible values $\left\{a^1, a^2, ..., a^V\right\}$. If feature a is used to partition the sample set D, this results in V branch nodes, where the v-th node contains the total number of all samples in the sample set D that take value a^v on feature a, denoted as D^v. Then the information gain of attribute m can be defined as:

$$G(D, a) = E(D) - \sum_{v}^{V=1} \frac{|D^v|}{|D|} E(D^v), \tag{3.9}$$

the higher its value, the higher the purity. Further, the information gain rate of attribute a can be defined as:

$$R_{Gain}(D, a) = \frac{G(D, a)}{I(a)}, \tag{3.10}$$

$$I(a) = -\sum_{v}^{V=1} \frac{|D^v|}{|D|} log_2 \frac{|D^v|}{|D|}, \tag{3.11}$$

in addition to this, the Gini index can also be used to measure the purity of the sample set, which is defined as:

$$G(D) = 1 - \sum_{k=1}^{K} {p_k}^2, \tag{3.12}$$

the smaller its value, the higher the purity. Further, the Gini index of attribute a can be defined as:

$$I_{Gini}(D, a) = -\sum_{v}^{V=1} \frac{|D^v|}{|D|} G(D), \tag{3.13}$$

the smaller its value, the higher the purity. The feature division, that is, the purity metric chosen by different indicators, creates different decision trees. The decision trees constructed from the indicators information gain, information gain rate, and Gini index are $ID3$, $C4.5$, and $CART$, respectively.

3.5. Random forest

Random forest is a bagging ensemble learning algorithm first proposed by Breiman in 2001 [32]. It is not a standalone supervised learning algorithm, but rather integrates the modeling results of all models by building multiple models. The basic unit in Random Forest is the $CART$ decision tree. The random forest algorithm's specific steps are divided into four major steps, which are as follows:

Step 1: Using the bootstrap method, generate m training sets and train m decision trees.

Step 2: When selecting features for splitting at the node of each decision tree, a subset of the features is chosen at random.

Step 3: Each decision tree is grown to its full potential without being pruned.

Step 4: The generated multiple decision trees are used for decision making. For the classification problem, the classification result is decided by multiple decision trees voting. For the regression problem, the regression result is decided by the mean of the predicted values of multiple decision trees.

3.6. XGBoost

XGBoost is a boosting ensemble learning algorithm proposed by Tianqi Chen et al. in 2016 [33]. It is a massively parallel boosting tree tool that employs a regression decision tree as its base unit. Unlike the bagging ensemble learning algorithm, the boosting ensemble learning algorithm doesn't train multiple base models at the same time. It employs a boosting approach in which each base model learns from the previous base model. The final prediction is produced by combining the predictions of all base models.

The main process of XGBoost's modeling is to grow a tree based on feature splitting and keep adding trees, each time adding a tree, actually losing the residuals of the last prediction to obtain a new function, iterating to improve the model performance. It is an additive operator consisting of k base models:

$$y_i(x) = \sum_{t=1}^{k} f_t(x_i), \tag{3.14}$$

where f_t is the t-th base model and $y_i(x)$ is the predicted value of the i-th sample. The loss function can be expressed as:

$$L = \sum_{i=1}^{n} I(y_i, y_i(x)), \tag{3.15}$$

where n is the number of samples. The objective function of the XGBoost model consists of a loss function and a regular term Ω that suppresses the complexity of the model:

$$Obj = \sum_{i=1}^{n} I(y_i, y_i(x)) + \sum_{t=1}^{k} \Omega(f_t), \tag{3.16}$$

how to minimizie the objective function, XGBoost's idea is to approximate it using the Taylor second-order expansion of f at $t = 0$. The final objective function relies only on the first-order derivatives and second-order derivatives of each data point on the error function.

The regularization term is added to the objective function to prevent overfitting, and this construction and solution of the objective function gives XGBoost good performance. To improve the accuracy of the results, not only the first-order derivatives but also the second-order derivatives are used in the solution of the objective function. Furthermore, XGBoost can solve classification and regression problems.

3.7. Evaluation criteria

The study's core is a dichotomous problem in which five classical and representative machine learning models are chosen for prediction. Model evaluation is used to judge and compare the learning ability of these models. In machine learning, accuracy, sensitivity, specificity, the ROC curve, and the AUC value are frequently used to assess model performance for dichotomous classification problems [34].

For any binary classification problem. Instances can be classified into positive and negative classes. And predictions can be done in four ways. The instance is a positive class and prediction is a positive class (TP). The instance is a positive class and prediction is a negative class (FN). The instance is a negative class and prediction is a negative class (TN). The instance is a negative class and prediction is a positive class (FP). The confusion matrix for the dichotomous classification problem can be constructed from this (Table 3).

Table 3. Confusion matrix of binary classification problem.

Real situation	Predicted result	
	Positive	Negative
Positive	TP (True positive)	FN (False negative)
Negative	FP (False positive)	TN (True negative)

Further, the formulae for the calculation of precision, sensitivity and specificity are derived:

$$Acurracy = \frac{TP + TN}{TP + FP + FN + TN}, \tag{3.17}$$

$$Sensitity = \frac{TP}{TP + FN}, \tag{3.18}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{3.19}$$

the full name of the ROC curve is Receiver Operating Characteristic (ROC) [35]. The ROC curve is created by ranking the samples based on the model prediction results, then calculating the values of two

significant quantities and plotting them as the horizontal and vertical axes, respectively. Its horizontal and vertical axes are the false positive rate (FPR) and the true case rate (TPR), which are calculated as follows:

$$FPR = \frac{FP}{FP + TN}, \tag{3.20}$$

$$TPR = \frac{TP}{TP + FN}, \tag{3.21}$$

the area under the ROC curve is the AUC value. It is widely assumed that the closer the AUC value is to 1, the better the model's learning ability.

4. Results

4.1. Data correction

There are 96 patients with better postoperative facial nerve function recovery and 32 patients with poor postoperative facial nerve function recovery among 128 patients with AN. There is a significant difference in sample size between the two categories. As a result, the AN dataset should be balanced before selecting models and making predictions. However, the commonly used SMOTE algorithm tends to cause problems such as overfitting, so the SMOTE-ENN algorithm is considered. Based on the balanced data, it further performs data cleaning to better avoid overfitting. After completing the data imbalance correction, the data in this paper changed from 128 to 94.

Table 4. Data distribution table (before and after correcting).

Classification	Sample size (before)	Sample size (after)	Proportion (before)	Proportion (after)
Better	96	39	0.75	0.41
Worse	32	55	0.25	0.58

4.2. Model comparison

Five machine learning models are used in this study: logistic regression (LR), support vector machine (SVM), decision tree (CART), random forest (RF), and XGBoost. To more reasonably compare the effects of the models, we use five-fold cross-validation. The corrected dataset is brought into the five models for training, and the average accuracy, sensitivity, specificity, and AUC values of each model are obtained. The performance of the five models is compared to determine the best model for postoperative facial nerve function recovery in AN patients.

Table 5. Performance of five machine learning models.

Model	Accuracy	Sensitivity	Specificity	AUC
LR	0.79 ± 0.08	0.82 ± 0.17	0.74 ± 0.09	0.78 ± 0.07
SVM	0.82 ± 0.09	0.84 ± 0.18	0.80 ± 0.12	0.81 ± 0.08
CART	0.87 ± 0.10	0.91 ± 0.09	0.82 ± 0.12	0.86 ± 0.10
RF	0.86 ± 0.16	0.87 ± 0.15	0.85 ± 0.16	0.86 ± 0.15
XGboost	0.90 ± 0.14	0.93 ± 0.14	0.85 ± 0.21	0.90 ± 0.15

The means and standard deviations of the evaluation criteria for the five models are given in Table 5. Combining precision, specificity, sensitivity, and AUC values, XGBoost outperforms the other four models. LR, on the other hand, performs the worst and is lower than the other four models in terms of each criterion. In terms of stability of precision, specificity, and AUC values, LR is more stable and XGBoost is less stable. In the stability of sensitivity, CART is more stable and SVM is less stable.

All five models perform better in predicting postoperative facial nerve function in patients with AN, but XGBoost is the best in each criterion. Not only does it have high prediction accuracy, sensitivity, and specificity, but it also has an AUC value of 0.90. XGBoost has a disadvantage that it may be slightly less stable. However, considering all other aspects, it is still considered that XGBoost is the best model to predict the recovery of facial nerve function after surgery in patients with AN. As a result, based on known basic patient information, tumor imaging data, preoperative symptoms assessment, postoperative facial nerve function assessment, and other clinical information, it is concluded that XGBoost is the most accurate model among the five models.

4.3. Model validation

To better validate the models, we select a portion of the original data as the validation set (n = 30). It includes 15 positive samples and 15 negative samples. The validation set is used to further compare the prediction effectiveness of the five models.

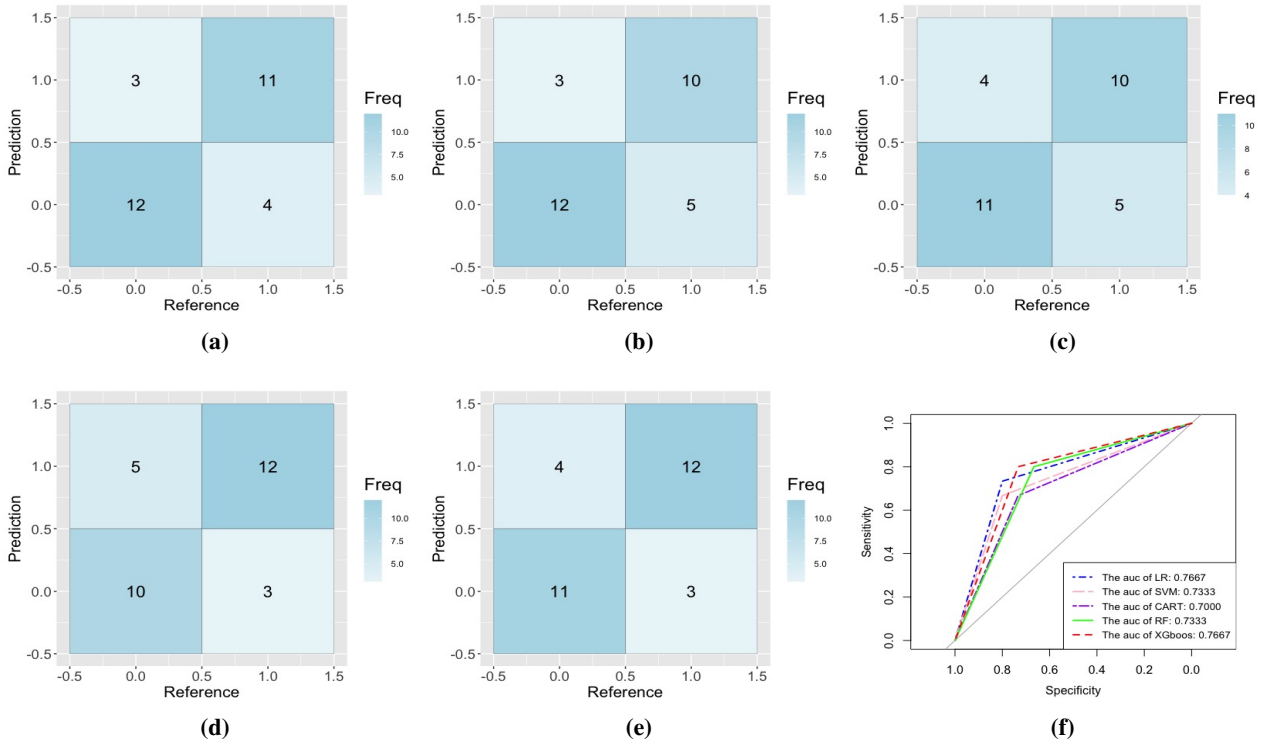

Figure 2. Confusion matrix and ROC curves of five machine learning models. (a) Confusion matrix of LR, (b) Confusion matrix of SVM, (c) Confusion matrix of CART, (d) Confusion matrix of RF, (e) Confusion matrix of XGBoost, (f) ROC curves of five models.

Figure 2 shows the confusion matrix and ROC curves of the models on the validation set. From the

confusion matrix, the SVM and RF have the same prediction results with 22 cases of samples correctly predicted (Figure 2b,2d). Among them, XGBoost and LR have the best prediction results with 23 cases of samples correctly predicted (Figure 2a,2e). And CART has the worst prediction result, with 21 cases of samples correctly predicted (Figure 2c). In terms of ROC curves and AUC values, XGBoost and LR have the same AUC value of 0.77. SVM and RF also have the same AUC value of 0.73. While CART has the smallest AUC value of 0.70 (Figure 2f). XGBoost and LR have the best prediction results based solely on the confusion matrix and ROC curves, and further calculate the five models for accuracy, sensitivity, specificity, and 95% CI of AUC (Table 6).

Table 6. Performance of five machine learning models.

Model	Accuracy	Sensitivity	Specificity	AUC	AUC (95% CI)
LR	0.77	0.73	0.80	0.7667	0.6105–0.9228
SVM	0.73	0.67	0.80	0.7333	0.5714–0.8953
CART	0.70	0.67	0.73	0.7000	0.5307–0.8693
RF	0.73	0.80	0.67	0.7333	0.5714–0.8953
XGboost	0.77	0.80	0.73	0.7667	0.6105–0.9228

In terms of prediction accuracy, XGBoost and LR have a high prediction accuracy of 0.77, while CART has the lowest prediction accuracy of 0.70. Given sensitivity, XGBoost and RF have the highest sensitivity of 0.80, while SVM and CART have the lowest sensitivity of 0.67. LR and SVM have the highest specificity of 0.80, while RF has the lowest specificity was 0.67. Overall, XGBoost and LR performed better on the validation set with the same 95% CI of AUC: 0.6105–0.9228. Combined with the model prediction results of the five-fold cross-validation on the corrected dataset, we conclude that XGBoost remains the best model for predicting patients' postoperative facial nerve function recovery.

4.4. Feature importance

This study includes 11 features that corresponded to patients' preoperative metrics. The features that contribute more to the construction of the predictive models are the preoperative indicators that should be given more importance by patients and physicians. Among the five machine learning models constructed, CART, RF, and XGBoost provides a direct importance score for each feature, which measures the value of the feature in the construction of the model. The greater the contribution of a feature in the construction of the model, the higher its relative importance. CART, RF, and XGBoost use the Gini index, the average Gini index and the average information gained to measure the contribution of each feature, respectively.

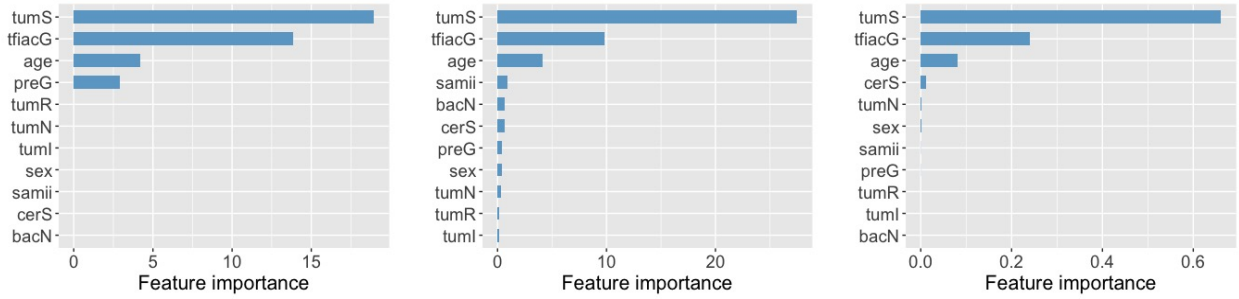

Figure 3. Feature importance. (a) Feature importance of CART, (b) Feature importance of RF, (c) Feature importance of XGBoost.

Figure 3 depicts the relative importance of features in CART, RF, and XGBoost models. In the construction of CART, the more important features are tumS (tumor size), tfiacG (TFIAC grading), age, and preG (preoperative facial nerve gradings) (Figure 3a). In the RF construction, the more important features are tumS (tumor size), tfiacG (TFIAC grading), and age (Figure 3b). In contrast, in the XGBoost construction, the more important features were tumS (tumor size), tfiacG (TFIAC grading), and age (Figure 3c). It can be found that in the construction of these three models, the most important one is tumS (tumor size), which is much more important than other features. The next more important features are tfiacG (TFIAC grading), age, and preG (preoperative facial nerve gradings).

According to the feature importance ranking, tumor size, TFIAC grading, age, preoperative facial nerve gradings should be prioritized among the 11 preoperative indicators for patients with AN. Simultaneously, these four indicators are critical for predicting facial nerve function recovery after surgery. This can assist physicians in developing individualized treatment plans, saving medical resources, and improving patients' recovery of facial nerve function and quality of life after surgery.

5. Conclusions

This paper focuses on the uses of machine learning and SMOTE-ENN techniques in the postoperative recovery of patients with AN, with a particular focus on the accurate prediction of facial nerve function. Thanks to the datum collected from the Xiangya Hospital of Central South University, our main working procedures based on the investigation of 128 surgical cases can be summarized as follows.

1) We assign all indices proper values before loading the statistical method. Due to the complexity of the clinical data, it cannot be directly used for modeling. It may need to be quantified according to the clinical characteristics of the disease and the purpose of the study. The dependent variable studied in this paper is not directly given, so it is inscribed as whether the facial nerve function is improved.

2) The samples are corrected using the SMOTE-ENN algorithm. The accuracy of traditional machine learning models using unbalanced datasets is very poor, which eventually leads to unreliable predictions of the models. So it is necessary to balance the datasets before modeling. However, using the SMOTE algorithm to expand a few classes of samples generates noise and tends to cause model overfitting. Therefore, the SMOTE-ENN algorithm is considered to correct the data set.

3) XGBoost is selected to construct a prediction model for the recovery, which is also compared

with a total of four machine learning models. After using five-fold cross-validation, XGBoost is found to have the highest prediction accuracy and AUC values. This indicates that the XGBoost model can accurately predict the postoperative recovery of patients and can build prognostic models to provide medical assistance to physicians.

The recent development of artificial intelligence has led to the gradual realization of precision medicine. Based on machine learning and SMOTE-ENN techniques, we present a new methodology to further study the accurate prediction of postoperative facial nerve function recovery for patients with AN, instead of studying them solely by using traditional statistical techniques reported in previous literature. In our experience, the prediction accuracy of the XGBoost model reaches 90.0% and the AUC value is 0.90. Due to the difficulty of data collection for AN disease, the sample size used in this paper is not very large. In our future studies, a better and more reliable prediction model can be established if more data can be collected from clinical practice.

Acknowledgments

The author is grateful to Mr. Fengqi Zhang from the Xiangya Hospital of Central South University for his kind support in collecting the datum for this project. This work is partly supported by the Fundamental Research Funds for the Central Universities of Central South University (No. 1053320211909).

Conflict of interest

The author declares that there is no conflict of interest to report regarding the present study.

References

1. J. Halliday, S. A. Rutherford, M. G. McCabe, D. G. Evans, An update on the diagnosis and treatment of vestibular schwannoma, *Expert Rev. Neurother.*, **18** (2018), 29–39. https://doi.org/10.1080/14737175.2018.1399795
2. D. Starnoni, R. T. Daniel, C. Tuleasca, M. George, M. Levivier, M. Messerer, Systematic review and meta-analysis of the technique of subtotal resection and stereotactic radiosurgery for large vestibular schwannomas: a "nerve-centered" approach, *Neurosurg. Focus*, **44** (2018), E4. https://doi.org/10.3171/2017.12.FOCUS17669
3. B. Acs, M. Rantalainen, J. Hartman, Artificial intelligence as the next step towards precision pathology, *J. Intern. Med.*, **288** (2020), 62–81. https://doi.org/10.1111/joim.13030
4. J. Goecks, V. Jalili, L. Heiser, J. Gray, How machine learning will transform biomedicine, *Cell*, **181** (2020), 92–101. https://doi.org/10.1016/j.cell.2020.03.022
5. G. S. Handelman, H. K. Kok, R. V. Chandra, A. H. Razavi, M. J. Lee, H. Asadi, eDoctor: machine learning and the future of medicine, *J. Intern. Med.*, **284** (2018), 603–609. https://doi.org/10.1111/joim.12822
6. M. M. Hasan, M. A. Alam, W. Shoombuatong, H. W. Deng, B. Manavalan, H. Kurata, NeuroPred-FRL: an interpretable prediction model for identifying neuropeptide using feature representation learning, *Briefings Bioinf.*, **22** (2021). https://doi.org/10.1093/bib/bbab167

7. A. Hoshino, H. S. Kim, L. Bojmar, K. E. Gyan, M. Cioffi, J. Hernandez, et al., Extracellular vesicle and particle biomarkers define multiple human cancers, *Cell*, **18** (2020), 1044–1061. https://doi.org/10.1016/j.cell.2020.07.009
8. B. Koo, J. K. Rhee, Prediction of tumor purity from gene expression data using machine learning, *Briefings Bioinf.*, **22** (2021). https://doi.org/10.1093/bib/bbab163
9. H. Luo, Q. Zhao, W. Wei, L. Zheng, S. Yi, G. Li, et al., Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer, *Sci. Transl. Med.*, **12** (2020). https://doi.org/10.1126/scitranslmed.aax7533
10. L. Huang, L. Wang, X. Hu, S. Chen, Y. Tao, H. Su, et al., Machine learning of serum metabolic patterns encodes early-stage lung adenocarcinoma, *Nat. Commun.*, **11** (2020), 3556. https://doi.org/10.1038/s41467-020-17347-6
11. M. K. Abd Ghani, M. A. Mohammed, N. Arunkumar, S. A. Mostafa, D. A. Ibrahim, M. K. Abdullah, et al., Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques, *Neural Comput. Appl.*, **32** (2020), 625–638. https://doi.org/10.1007/s00521-018-3882-6
12. P. Achilli, C. Magistro, M. A. A. E. Aziz, G. Calini, C. L. Bertoglio, G. Ferrari, et al., Modest agreement between magnetic resonance and pathological tumor regression after neoadjuvant therapy for rectal cancer in the real world, *Int. J. Cancer*, (2022), 1–8. https://doi.org/10.1002/ijc.33975
13. Z. M. Zhuang, Z. B. Yang, S. X. Zhuang, A. N. J. Raj, Y. Yuan, R. Nersisson, Multi-features-based automated breast tumor diagnosis using ultrasound image and support vector machine, *Comput. Intell. Neurosci.*, **2021** (2021). https://doi.org/10.1155/2021/9980326
14. M. M. Ghiasi, S. Zendehboudi, Application of decision tree-based ensemble learning in the classification of breast cancer, *Comput. Biol. Med.*, **128** (2021). https://doi.org/10.1016/j.compbiomed.2020.104089
15. A. Moncada-Torres, M. C. van Maaren, M. P. Hendriks, S. Siesling, G. Geleijnse, Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival, *Sci. Rep.*, **11** (2021). https://doi.org/10.1038/s41598-021-86327-7
16. C. Yang, X. W. Huang, Y. Li, J. F. Chen, Y. Y. Lv, S. X. Dai, et al., Prognosis and personalized treatment prediction in TP53-mutant hepatocellular carcinoma: an in silico strategy towards precision oncology, *Briefings Bioinf.*, **22** (2021). https://doi.org/10.1093/bib/bbaa164
17. Y. Q. Wu, N. Jiao, R. X. Zhu, Y. D. Zhang, D. F. Wu, A. J. Wang, et al., Identification of microbial markers across populations in early detection of colorectal cancer, *Nat. Commun.*, **12** (2021). https://doi.org/10.1038/s41467-021-23265-y
18. L. Zhang, H. X. Ai, W. Chen, Z. M. Yin, H. Hu, J. F. Zhu, et al., CarcinoPred-EL: Novel models for predicting the carcinogenicity of chemicals using molecular fingerprints and ensemble learning methods, *Sci. Rep.*, **7** (2017). https://doi.org/10.1038/s41598-017-02365-0
19. A. Tahmassebi, G. J. Wengert, T. H. Helbich, Z. Bago-Horvath, S. Alaei, R. Bartsch, et al., Impact of machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy and survival outcomes in breast cancer patients, *Invest. Radiol.*, **54** (2019), 110–117. https://doi.org/10.1097/RLI.0000000000000518

20. J. Li, Z. Shi, F. Liu, X. Fang, K. Cao, Y. H. Meng, et al., XGBoost classifier based on computed tomography radiomics for prediction of tumor-infiltrating CD8+T-Cells in patients with pancreatic ductal adenocarcinoma, *Front. Oncol.*, **11** (2021). https://doi.org/10.3389/fonc.2021.671333

21. K. Thedinga, R. Herwig, A gradient tree boosting and network propagation derived pan-cancer survival network of the tumor microenvironment, *iScience*, **25** (2021), 103617. https://doi.org/10.1016/j.isci.2021.103617

22. W. Tang, H. Zhou, T. H. Quan, X. Y. Chen, H. N. Zhang, Y. Lin, et al., XGboost prediction model based on 3.0T diffusion kurtosis imaging improves the diagnostic accuracy of MRI BiRADS 4 masses, *Front. Oncol.*, **12** (2022), 833680. https://doi.org/10.3389/fonc.2022.833680

23. B. Fu, P. Liu, J. Lin, L. Deng, K. Hu, H. Zheng, Predicting invasive disease-free survival for early stage breast cancer patients using follow-up clinical data, *IEEE Trans. Biomed. Eng.*, **66** (2019), 2053–2064. https://doi.org/10.1109/TBME.2018.2882867

24. S. L. Li, X. J. Zhang, Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm, *Neural Comput. Appl.*, **32** (2020), 1971–1979. https://doi.org/10.1007/s00521-019-04378-4

25. Y. M. Hsiao, C. L. Tao, E. Y. Chuang, T. P. Lu, A risk prediction model of gene signatures in ovarian cancer through bagging of GA-XGBoost models, *J. Adv. Res.*, **30** (2021), 113–122. https://doi.org/10.1016/j.jare.2020.11.006

26. B. Krawczyk, Learning from imbalanced data: open challenges and future directions, *Prog. Artif. Intell.*, **5** (2016), 221–232. https://doi.org/10.1007/s13748-016-0094-0

27. S. Fotouhi, S. Asadi, M. W. Kattan, A comprehensive data level analysis for cancer diagnosis on imbalanced data, *J. Biomed. Inf.*, **90** (2019). https://doi.org/10.1016/j.jbi.2018.12.003

28. N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, *J. Artif. Intell. Res.*, **6** (2001), 321–357. https://doi.org/10.1613/jair.953

29. G. Batista, R. C. Prati, M. C. Monard, A study of the behavior of several methods for balancing machine learning training data, *ACM SIGKDD Explor. Newsl.*, **6** (2004), 20–29. https://doi.org/10.1145/1007730.1007735

30. S. Fotouhi, S. Asadi, M. W. Kattan, A comprehensive data level analysis for cancer diagnosis on imbalanced data, *J. Biomed. Inf.*, **90** (2019). https://doi.org/10.1016/j.jbi.2018.12.003

31. X. Huang, T. Y. Cao, L. Z. Q. Chen, J. P. Li, Z. H. Tan, B. J. M. Xu, et al., Novel insights on establishing machine learning-based stroke prediction models among hypertensive adults, *Front. Cardiovasc. Med.*, **9** (2022). https://doi.org/10.3389/fcvm.2022.901240

32. L. Breiman, Random forest, *Mach. Learn.*, **45** (2001), 5–32. https://doi.org/10.1023/A:1010933404324

33. T. Q. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2016), 785–794. https://doi.org/10.1145/2939672.2939785

34. M. Sokolova, G. Lapalme, TA systematic analysis of performance measures for classification tasks, *Inf. Process. Manage.*, **45** (2009), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

35. A. P. Bradley, The use of the area under the ROC curve in evaluation of machine learning algorithms, *Pattern Recognit.*, **30** (1997), 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2

AIMS Press