Epilepsy & Behavior 96 (2019) 92–97

Contents lists available at ScienceDirect

Epilepsy & Behavior

journal homepage: www.elsevier.com/locate/yebeh

Prediction of antiepileptic drug treatment outcomes of patients with newly diagnosed epilepsy by machine learning

Lijun Yao [a,*], Mengting Cai [b], Yang Chen [c], Chunhong Shen [b], Lei Shi [d], Yi Guo [b,*]

[a] Shanghai Pudong New Area Mental Health Center, Tongji University School of Medicine, Shanghai 200124, PR China
[b] Department of Neurology, Epilepsy Center, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310009, Zhejiang, PR China
[c] School of Computer Science, Fudan University, Shanghai 201203, PR China
[d] The State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing 100080, PR China

ARTICLE INFO

Article history:
Received 14 December 2018
Revised 22 March 2019
Accepted 7 April 2019
Available online 20 May 2019

Keywords:
Machine learning
Epilepsy
Epilepsy remission
Antiepileptic drug
Outcome prediction

ABSTRACT

Objective: The objective of this study was to build a supervised machine learning-based classifier, which can accurately predict the outcomes of antiepileptic drug (AED) treatment of patients with newly diagnosed epilepsy.
Methods: We collected information from 287 patients with newly diagnosed epilepsy between 2009 and 2017 at the Second Affiliated Hospital of Zhejiang University. Patients were prospectively followed up for at least 3 years. A number of features, including demographic features, medical history, and auxiliary examinations (electroencephalogram [EEG] and magnetic resonance imaging [MRI]) are selected to distinguish patients with different remission outcomes. Seizure outcomes classified as remission and never remission. In addition, remission is further divided into early remission and late remission. Five classical machine learning algorithms, i.e., Decision Tree, Random Forest, Support Vector Machine, XGBoost, and Logistic Regression, are selected and trained by our dataset to get classification models.
Results: Our study shows that 1) compared with the other four algorithms, the XGBoost algorithm based machine learning model achieves the best prediction performance of the AED treatment outcomes between remission and never remission patients with an F1 score of 0.947 and an area under the curve (AUC) value of 0.979; 2) The best discriminative factor for remission and never remission patients is higher number of seizures before treatment (>3); 3) XGBoost-based machine learning model also offers the best prediction between early remission and later remission patients, with an F1 score of 0.836 and an AUC value of 0.918; 4) multiple seizure type has the highest dependence to the categories of early and late remission patients.
Significances: Our XGBoost-based machine learning classifier accurately predicts the most probable AED treatment outcome of a patient after he/she finishes all the standard examinations for the epilepsy disease. The classifier's prediction result could help disease guide counseling and eventually improve treatment strategies.

1. Introduction

Epilepsy is one of the most devastating neurological diseases, affecting more than 50 million people worldwide [1]. Sixty to seventy percent of the patients achieve control when using a single or a combination of antiepileptic drugs (AEDs), while 12–63% of the patients still suffer from epilepsy relapsing even with medication [2]. Seizure outcomes can be classified into four different temporal patterns: early remission, late remission, relapsing–remitting course, and never remission [3]. Beside the therapeutic regimen, the clinical outcomes of patients with newly diagnosed epilepsy are related to many factors, such as seizure types, electroencephalogram (EEG) abnormality, multiple seizure types, intellectual disability, higher seizure frequency before diagnosis, symptomatic etiology, etc. [4–7].

In the majority of previous studies, researchers retrospectively investigated the prognostic factors, which may be related to the outcomes following diagnosis. It would be interesting to accurately predict the outcomes based upon patient's condition before AEDs treatment of patients with epilepsy. The booming development of artificial intelligence makes it potentially possible. Machine learning, as an important branch of artificial intelligence, has been applied to seizure detection [8,9], surgical outcome prediction [10–13], and AED selection [14,15]. Machine learning-based algorithms can combine patient's personal information to predict the outcomes of AEDs treatment and, hence, help support clinical decision making.

In the present study, we adopt supervised machine learning-based algorithms to train a classifier using the collected patient's data to predict the outcomes. We have carefully selected a number of features to distinguish between different types of treatment outcomes. We implemented and compared 5 representative classification algorithms, i.e., Decision Tree [16], Random Forest [17], Support Vector Machine

* Corresponding authors.
E-mail addresses: lyao@tongji.edu.cn (L. Yao), yiguo@zju.edu.cn (Y. Guo).

https://doi.org/10.1016/j.yebeh.2019.04.006
1525-5050/

(SVM) [18], XGBoost [19], and Logistic Regression [20]. Based on our results, one can use the trained classifier to make a prognosis of the newly diagnosed epilepsy during the patients first visit to the epileptologists, which could be valuable in helping clinicians design rational therapeutic regimen for individual patient.

2. Methods

2.1. Patients

In this study, patients with newly diagnosed epilepsy were included between January 2009 to December 2014 at the Second Affiliated Hospital of Zhejiang University in Zhejiang Province, China. Patients were prospectively followed up until the end of December 2017, i.e., for at least three years. All of the patients' epilepsies were diagnosed by epileptologists according to the definition of International League Against Epilepsy (ILAE) [21]. Epilepsy is defined with either the following: (1) two or more unprovoked seizures occurring at least 24 h apart; (2) one unprovoked seizure with a definite epileptic focus on brain magnetic resonance imaging [MRI] or epileptiform discharges on EEG. Our definition of epilepsy also meets the standard of the 2014 ILAE definition [22]. Patients who have the following conditions were excluded: 1) refused the treatment with AEDs; 2) took more than one AED during the follow-up; 3) received epilepsy surgery during the follow-up; 4) had severe hepatic or renal diseases [23]. This study was approved by the Ethics Committee of the Second Affiliated Hospital of Zhejiang University.

In order to predict the outcomes after AED treatment of patients with epilepsy during their first epileptologist visit, we collect the following information of each patient: patient demographics (gender, living area, occupation, and education level), age at seizure onset, past medical history (brain trauma, stroke, intracranial infection, perinatal hypoxia, and brain tumor), family history of epilepsy, seizure types, number of seizures before treatment, febrile seizure, prior treatment duration, multiple seizure type, and EEG and MRI findings. Seizure type was classified into focal onset, generalized onset, and unknown onset according to the ILAE 2017 classification of seizure types basic version [24]. Multiple seizure type was defined as two or more seizure types, which includes focal aware seizure, focal impaired awareness seizure, focal to bilateral tonic–clonic seizure, generalized onset seizure, and unknown onset seizure. Monotherapy was recommended for the initial treatment, and the AED selection was determined by the caring epileptologist. One of the following AEDs was selected as monotherapy for each patient: carbamazepine, oxcarbazepine, sodium valproate, lamotrigine, levetiracetam, topiramate, and gabapentin. For patients who had seizure relapses with the first maintenance dose, dosage was further increased up to the highest tolerated dose to achieve seizure control. When the first monotherapy failed, another drug was given to patient as an alternative monotherapy (preferred) or adjunctive therapy based on the epileptologist's choice.

2.2. Definitions

In our study, we defined the timing point of epilepsy remission in accordance with Brodie's classification [3]. Remission is considered a seizure-free period if patients experience no seizures for at least one year on unchanged treatment. We classify the outcomes of AED treatment into three patterns: 1) Early remission is achieved either immediately or within six months of AED treatment; 2) Late remission means seizure freedom is delayed for more than six months after the initiation of treatment; 3) Never remission means patients never become seizure-free for any complete year until the end of follow-up. To simplify our model, we do not consider patients who have a relapsing course after remission achieved as a single type of outcome. Patients were evaluated at 4 weeks after treatment initiation and at 3-month intervals thereafter during the first year. From the second year and later, patients were evaluated at 6-months intervals. If patients' condition had been changed, a visit would be arranged accordingly. During each visit, seizure frequency, choice of AED, and response to AED were routinely recorded. The label of each patient's outcome in this study was verified by epileptologist based on medical records.

2.3. Machine learning-based classification

In this section, we evaluate the classification performance of our system. In our dataset, there are 207 remission patients and 80 patients who are never in remission; remission patients are further split into 140 early remission and 67 late remission patients (Fig. 1(2)). We

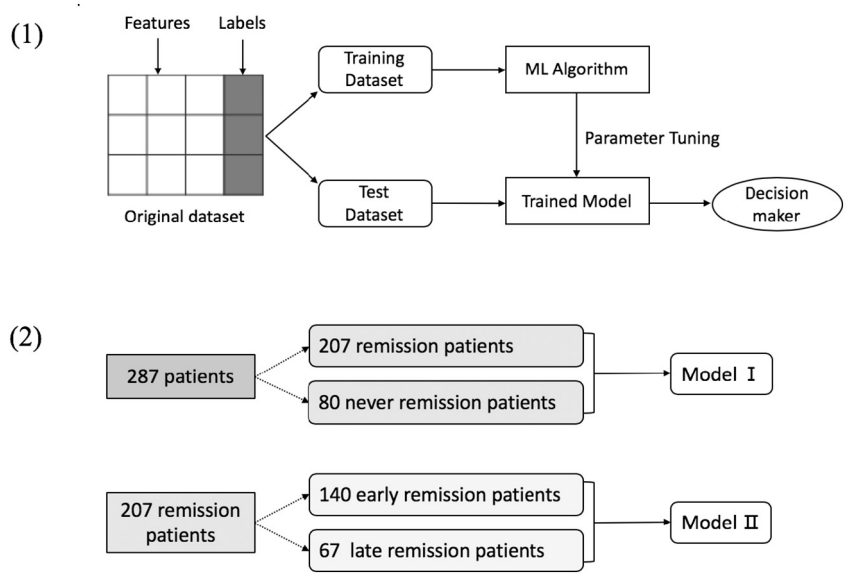

Fig. 1. The workflow of machine learning-based classifier. (1) The original dataset is split into a training dataset and a test dataset. Features are selected to describe patients' information. Five different machine learning algorithms are selected for training based on the training dataset. Trained models are achieved after parameter tuning. The final classifier is determined according to the comparison of each trained model's prediction performance. (2) Upper panel: 207 remission patients and 80 never remission patients' data are used to obtain model I; Lower panel: 140 early remission patients and 67 late remission patients' data were used to obtain model II. ML, machine learning; model I, remission/never remission prediction model; model II, early/late remission prediction model.

aimed to build a binary classifier that is able to distinguish between the following: 1) remission and never remission patients; 2) early and late remission patients, accurately. The workflow is described in Fig. 1(1). There are four key steps of our machine learning-based classification model, i.e., data preprocessing, feature selection, algorithm selection, and parameter tuning. Finally, we evaluate the prediction performance of our model. Scikit-learn, a Python-based machine learning library, was used to train the classification model [25] (refer to the website: http://scikit-learn.org/stable/).

2.3.1. Data preparation

We split the entire dataset into a training dataset and a test dataset. As in [26–28], we use 80% of the patients for training, and the other 20% of the patients for test. For both the training set and the test set, we ensure that there are equal amounts of two types of patients. We oversample the minority type by picking samples at random. We further used the 5-fold cross-validation method to prepare our training dataset, where the training dataset was randomly divided into 5 subsets with equal sizes. Of the 5 subsets, a single subset is retained as the validation data for evaluating the model, and the remaining 4 subsets are used for training. The cross-validation process is repeated 5 times, with each of the 5 subsets used once for validation.

2.3.2. Feature selection

We aimed to select a number of features that could distinguish between different outcomes after AED treatment. The following features with categorical variables are manually selected: patient's demographic information (gender (female/male), living area (urban area/rural area), occupations (student/employment/unemployment) and educational levels (illiteracy or primary school/secondary school/college or above)), medical history (age at seizure onset (≦16/>16), past medical history (negative/positive), family history (negative/positive), seizure type (focal onset/generalized onset/unknown onset), number of seizures before treatment (≦3/>3), febrile seizure (negative/positive), prior treatment duration (≦6 months/6~12 months/>12 months), multiple seizure types (no/yes)), auxiliary examination results (EEG abnormality (normal/epileptiform discharge/not available), and MRI findings (normal/abnormal/not available)). Table 1 shows the details of selected features.

2.3.3. Algorithm selection

There are a number of supervised machine learning-based classification algorithms available. In our study, we investigated several classical algorithms including the Decision Tree, Random Forest, Logistic Regression, and SVM. We also studied one emerging algorithm, i.e., XGBoost. XGBoost is a scalable end-to-end tree boosting system. It has been widely used in recent machine learning competitions on Kaggle.

2.3.4. Parameter tuning

For a selected algorithm, we need to determine an optimal set of parameters. Based on the training dataset, we apply grid search to go through the parameter space. We select a finite set of values of each parameter to form the parameter space. Grid search iterates through each parameter combination. For each combination, we evaluate the prediction performance. Finally, we record the parameters leading to the maximum F1-score based on the training set. Afterwards, the classifier can be used to judge the outcome of a patient with new diagnoses.

2.3.5. Model evaluation

To evaluate the performance of the trained model, we use the following representative metrics, i.e., precision, recall, F1-score, and the area under the curve (AUC) value [29]. Precision, similar to positive predictive value, is the fraction of model classified remission patients who are real remission patients (or classified early remission patients who are real early remission patients). Recall, similar to sensitivity, means the fraction of remission patients (or early remission patients) who have been identified by the model correctly. F1-score is the harmonic mean of precision and recall, which is defined as follows:

$$\mathrm{F1} = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

The value of F1-score is between 0 and 1. The AUC denotes the probability that this classifier will rank higher of a random positive instance than a randomly chosen negative instance. From the perspective of clinicians, high precision means that our predictions rarely over report and indicate that patients will go into remission when they will in fact continue to achieve seizure freedom. Meanwhile, high recall means that our predictions rarely under report patients that will go into remission. The F1-score is a balance between these two metrics. A higher value of F1-score indicates a better prediction performance.

Table 1
Features of patients with newly diagnosed epilepsy.

Category	Feature	n (%)
Demographic features	Gender	
	Female	141 (49.1%)
	Male	146 (50.9%)
	Living area	
	Urban area	144 (50.2%)
	Rural area	143 (49.8%)
	Occupation	
	Student	128 (44.3%)
	Employment	121 (34.7%)
	Unemployment	38 (13.2%)
	Educational level	
	Illiteracy or primary school	37 (12.9%)
	Secondary school	143 (49.8%)
	College or above	103 (35.9%)
Medical history	Age at seizure onset	
	≦16	82 (25.6%)
	>16	205 (71.4%)
	Past medical history	
	Negative	193 (67.2%)
	Positive (brain trauma, stroke, intracranial infection, perinatal hypoxia, and brain tumor)	94 (32.8%)
	Family history	
	Negative	276 (96.2%)
	Positive	11 (3.8%)
	Seizure type	
	Focal onset	242 (84.3%)
	Generalized onset	30 (10.5%)
	Unknown onset	15 (5.2%)
	Number of seizures before treatment	
	≦3	128 (44.6%)
	>3	159 (55.4%)
	Febrile seizure	
	Negative	20 (7.0%)
	Positive	267 (93.0%)
	Prior treatment duration	
	≦6 months	153 (53.3%)
	6~12 months	45 (15.7%)
	>12 months	89 (31.0%)
	Multiple seizure type	
	No	210 (73.2%)
	Yes	77 (26.8%)
Auxiliary examination	EEG	
	Normal	77 (26.8%)
	Epileptiform discharge (185)	185 (64.5%)
	NA	25 (8.7%)
	MRI	
	Normal	169 (58.9%)
	Abnormal	70 (24.4%)
	NA	48 (16.7%)

NA: not available.

Table 2
Comparison the performance of different machine learning algorithms for epilepsy remission prognosis.

Algorithm	Parameters	Precision/positive predictive value	Recall/sensitivity	F1-score	AUC
XGBoost	'colsample_bytree': 0.5, 'learning_rate': 0.2, 'min_child_weight': 0.1, 'n_estimators': 100, 'subsample': 1, 'max_depth': 9, 'booster': 'gbtree'	0.923	0.973	0.947	0.979
Random Forest	'n_estimators': 100, 'criterion': 'gini', 'max_depth': 8	0.951	0.886	0.918	0.974
Decision Tree	'splitter': 'best', 'max_depth': 10	0.971	0.829	0.895	0.908
SVM	'C': 100, 'kernel': 'linear'	0.738	0.838	0.785	0.897
Logistic Regression	'C': 1, 'tol': 0.001	0.711	0.865	0.780	0.908

2.4. Statistical analysis

We use the Chi-Square ($\chi 2$) Statistics [30] to evaluate the dependence of a selected feature and the categories of patients. We calculate the $\chi 2$ value based on the patient category information and feature values. A larger $\chi 2$ value indicates a better discriminative power of a feature. According to $\chi 2$ values, the top 5 ranked features, which contribute most to differentiate remission and never remission patients or early and late remission patients, are presented in Tables 3 and 5, respectively. We also calculate the odds ratio and 95% Confidence Interval (CI) of the above top five features in each prediction step, and presented them in Tables 3 and 5.

3. Results

3.1. Patient demographics

A total of 320 patients were with diagnosed epilepsy, and none of them had previously received AEDs. Thirty-three patients (10.3%) were excluded from analysis because of lacking several follow-up information. As a result, 287 patients with newly diagnosed epilepsy were included for further analysis. The mean follow-up period was 5.76 years (standard deviation [SD] = 1.24 years). The mean age at referral was 25.6 years (SD = 14.9 years). The database comprises 14 features, belonging to three different categories: 4 features from demographic features, 8 features from medical history, and 2 features from auxiliary examinations. The number of patients having each feature was shown in Table 1.

3.2. Machine learning used to predict the prognosis of epilepsy remission of patients with newly diagnosed epilepsy (remission vs. never remission)

In the present study, we adopted supervised machine learning algorithms to predict the outcomes of AED treatment in patients with newly diagnosed epilepsy. In our dataset, there are 207 remission patients, and the rest of the 80 patients never entered into the remission stage during the follow-up (Fig. 1(2)). Firstly, we built a binary classifier that was able to classify remission and never remission patients accurately. Secondly, we further built a classifier to predict the outcomes of remission patients (i.e., early remission and late remission). For the first goal, we used five frequently-used machine learning algorithms — Decision Tree, Random Forest, SVM, Logistic Regression, and XGBoost — to build classification models. Our results showed that the F1-score of each of these five models (Decision Tree, Random Forest, SVM, Logistic Regression, and XGBoost) were 0.895, 0.918, 0.785, 0.780, and 0.947, respectively (Table 2). Each model's parameters, the value of precision, and recall were also described in Table 2. The AUC value of each of these five models (Decision Tree, Random Forest, SVM, Logistic Regression, and XGBoost) were 0.908, 0.974, 0.897, 0.908, and 0.979, respectively. Our data indicates that Decision Tree, Random Forest, and XGBoost were all effective in distinguishing potential remission and never remission patients. However, XGBoost-based classifier achieved the highest F1-score of 0.947 and an AUC value of 0.979, thus offering the best prediction between remission and never remission patients.

Next, we employed Chi-Square analysis to identify the discriminative power of each feature to the categories of patients. In total, 14 different features were included in this study. The top 5 ranked features that contribute most to distinguish remission and never remission patients are the number of seizures before treatment, multiple seizure type, seizure type, family history, and past medical history, each with Chi-square values of 129.63, 26.77, 7.58, 2.57, and 1.78, respectively (Table 3). The odds ratios of these five features were also described in Table 3. Higher number of seizures before treatment (>3) and having multiple seizure type increase the risk for patients who will not experience remission, while focal onset seizure will lead to a better outcome. The small odds ratio value of family history could be due to smaller sample size of patients who have positive family history (11 yes vs. 276 no family history). Our results suggested that the number of seizures before treatment was the most discriminative feature to distinguish remission and never remission patients.

3.3. Machine learning used to predict the prognosis of remission patients (early vs. late remission)

XGBoost-based classifier can accurately predict whether patients will have remission or not. For those patients who would have remission, we further predict if they would enter into early or late remission stage. In our dataset, we have 140 early remission patients and 67 late remission patients (Fig. 1(2)). As before, we used these data to train Decision Tree, Random Forest, SVM, Logistic Regression, and XGBoost algorithms to get precise binary classifiers. As a result, these models gave F1-score of 0.724, 0.774, 0.716, 0.621, and 0.836; AUC value of 0.717, 0.728, 0.666, 0.619, and 0.918, respectively (Table 4). Our results suggested that XGBoost-based classifier can better predict the outcomes of remission patients, with an F1-score of 0.836 and an AUC value of 0.918.

Finally, we analyzed each feature's dependence to the categories of early and late remission patients. Chi-square analysis showed that the top 5 ranked features were multiple seizure type, family history, age at seizure onset, number of seizures before treatment, and past medical history, with Chi-square values of 5.59, 3.60, 2.06, 1.89, and 1.11, respectively (Table 5). The odds ratios of these five features were described in Table 5. Having multiple seizure type, higher number of seizures before treatment (>3), and past medical history increase the risk for patients

Table 3
The ranking of feature importance (remission/never remission).

Rank	Features	Chi-square value	Odds ratio	95% CI
1	Number of seizures before treatment (>3)	129.63	9.984	5.003–19.924
2	Multiple seizure type (yes)	26.77	3.446	2.212–5.369
3	Seizure type (focal onset)	7.58	0.500	0.243–1.028
4	Family history (positive)	2.57	0.096	0.012–0.754
5	Past medical history (yes)	1.78	1.192	0.790–1.797

Table 4
Comparison the performance of different machine learning algorithms for epilepsy remission prognosis (early/late remission).

Algorithm	Parameters	Precision/positive predictive value	Recall/sensitivity	F1-score	AUC
XGBoost	'colsample_bytree': 1, 'learning_rate': 0.2, 'min_child_weight': 1, 'n_estimators': 100, 'subsample': 1, 'max_depth': 5, 'booster': 'gbtree'	0.852	0.821	0.836	0.918
Random forest	'n_estimators': 50, 'criterion': 'gini', 'max_depth': 10	0.706	0.857	0.774	0.728
Decision Tree	'splitter': 'best', 'max_depth': 10	0.750	0.700	0.724	0.717
SVM	'C': 100, 'kernel': 'linear'	0.667	0.774	0.716	0.666
Logistic Regression	'C': 10, 'tol': 0.001	0.667	0.581	0.621	0.619

who will enter into late remission. Younger age at seizure onset increases the risk of experiencing late remission. Our data suggested that the multiple seizure type feature has the largest discriminative power for early or late remission patients.

4. Discussion

In the present study, five machine learning algorithms have been carefully evaluated to predict the outcomes of AED treatment for patients with newly diagnosed epilepsy. Based on our results, we conclude the following: 1) Our XGBoost-based machine learning model offers precise predictions of the outcomes of AED treatment between remission and never remission patients with an F1-score of 0.947 and AUC value of 0.979. 2) The feature "number of seizures before treatment (>3)" has the largest dependence to the patient categories (remission/never remission ;). 3) The XGBoost-based machine learning model also offers the best prediction of early and later remission patients;. 4) The feature "multiple seizure type" has the highest dependence to the patient category (early remission/late remission). In short, we can accurately predict the AED treatment outcomes, which reflect the most probable prognosis when a patient receives treatment in tertiary Grade-A hospitals at current medical level.

Among the 14 features we have selected, it is important to identify which features are more relevant to the classification of patients. We have used Chi-Square analysis to demonstrate the top 5 ranked features which contribute most to distinguish remission and never remission patients, as well as early remission and late remission patients. Consider the classification between remission and never mission patients, the feature "number of seizures before treatment (>3)" had the highest Chi-square value at 129.6 and odds ratio at 9.984, which indicates patients who have a greater number of seizures before treatment are more susceptible to severe disease conditions, and associated with poor outcomes. Our result is consistent with a study that systematically analyzed cohort studies from Medline and Embase [4], where they found a greater number of seizures before diagnosis is a consistent predictor of less likelihood of remission, as well as consistent with a study from two cohort studies performed by Geelhoed et al. [31]. Besides, we have shown that the most relevant feature to distinguish early vs. late remission is "multiple seizure types". Shen C et al. identified more than 3 seizure onsets prior to treatment or multiple seizure types were two prognostic predictors associated with late remission using a multivariable logistic regression model [23], which is also consistent with our results. Based on our results, "multiple seizure type" ranks first, and "number of seizures before treatment" ranks fourth to distinguish early vs. late remission. The possible reason of the ranking difference could be due to the fact that we used a different methodology to rank features, as well as the different study population. In addition, the Chi-Square values of the top 5 ranked features are fairly close, other features also contribute to the final classification, but may be with less significance.

Table 5
The ranking of feature importance (early/late remission).

Rank	Features	Chi-square value	Odds ratio	95% CI
1	Multiple seizure type (yes)	5.59	3.032	1.682–5.466
2	Family history (positive)	3.60	0.361	0.094–1.391
3	Age at seizure onset (>16)	2.06	0.449	0.268–0.752
4	Number of seizures before treatment (>3)	1.89	1.598	0.992–2.575
5	Past medical history (yes)	1.11	1.523	0.923–2.514

The ultimate desired outcome of the patients with epilepsy is complete seizure freedom without any further medications. Predicting the outcomes based upon early features is of great interest. Machine learning-based methods have been applied in the field of epilepsy study to some degree [10,15,32–34]. In the present study, we focus on the outcomes of AED treatment. By evaluating a number of algorithms, we demonstrate the usefulness of using supervised machine learning algorithms for the prediction. In particular, the three classification algorithms (Decision Tree, Random Forest, and XGBoost) used in the present study were all effective as tools to predict between remission and never remission patients, with each showing an F1-score around or above 0.9. The XGBoost-based model performs best not only in predicting remission vs. never remission (F1-score of 0.947), but also early vs. late remission outcomes (F1-score of 0.836). Actually, some machine learning-based algorithms have been used to predict AED treatment outcomes, but mainly for predicting remission and never remission patients. For example, Geelhoed M et al. used classification tree model and stepwise logistic regression model to predict the outcome from two cohort studies of childhood epilepsy with 70% accuracy rate [31]. Devinsky O et al. used Random Forest to identify AED regimens from a large cohort of population, with 72% area under receiver operating characteristic (ROC) curve [14]. Berg AT et al. used logistic regression model to study a cohort of 613 children with newly diagnosed epilepsy; the overall accuracy of their model ranges from 72% to 85% depends on different feature selection [35]. Obviously, our classifier performs much better compared with previous studies. The main reason could be due to detailed scheme of feature selection, algorithms selection, and parameter tuning. First, we select a set of relevant features to distinguish between different types of patients. These features cover patient's demographic information, medical history, and auxiliary examinations. Second, we carefully chose and compared 5 representative supervised machine learning algorithms in this study. In particular, the XGBoost algorithm is an implementation of gradient boosted decision trees, which has recently been dominantly applied in Kaggle competitions due to its great performance. Last but not least, we conducted grid search to go through a number of parameter combinations, and obtain a set of "best" parameters of the classification model. Hughes DM et al. used baseline information plus follow-up data to predict patients who will not achieve seizure remission within 5 years on AEDs. Their model correctly classified 95% of no remission patients [36]. Despite different feature selection and ranking, our work only used baseline information but also achieved a good prediction performance. As a result, our XGBoost-based model can be used as an accurate classifier to predict the potential outcome of a patient during her/his first visit to an epileptologist.

For the next step, we will collect more patient samples to train our classifier and build a prediction software that connects to the hospital's patient database. The software will automatically read patient's

information and calculate the prediction outcome, which could be helpful for personalized therapeutic regimen design. For instance, if the predicted outcome by the classifier is never remission, epileptologist could try strong dose at early stage of the treatment, which would improve the prognosis of the treatment and bypass some unnecessary remedy attempts. Besides, knowing the potential treatment outcome may also affect patient and family members' psychology and reduce anxiety level. As a result, machine learning-based methods described in this study may become a powerful tool to be included in standard pretreatment evaluation for patients with epilepsy.

The predictive accuracy of a model depends on the large scale of dataset, the number and quality of features, and the design of the algorithm. Our study still has some limitations. First, we included 14 different features, and the weight of each feature in the final model differs. In the future, when we introduce more relevant features into our model, the value of F1-score will be further increased. Second, we collected the information of 287 patients in our study, yet a larger sample size of patients needs to be integrated into our dataset to increase the predictive accuracy. Third, we do not consider relapsing–remitting course as a single type of outcome in this study. Epilepsy relapse may happen at different time windows. At present, our classifier cannot predict if a patient will have relapse or not. Fourth, our dataset was collected at a local tertiary hospital, and the predictive accuracy may not be representative for all the regions in China and other countries. In the future, we will include more patient samples with longer follow-up years, introduce more relevant features, and incorporate epilepsy relapsing into our model. It is expected that our prediction classifier will become more accurate, and easier-to-use for clinic purposes.

5. Conclusion

In this study, we proposed a supervised machine learning-based approach that can accurately predict the outcomes of AED treatment in patients with newly diagnosed epilepsy. The information provided by our model is an important reference for neurologists and could be useful for treatment regimen decision making. Our solution only relies on general-purpose computers and open-source software, and can be adopted by hospitals conveniently.

Acknowledgments

This work was supported by the Stake Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences (No. SYSKF1801); National Natural Science Foundation of China (No. 81871010); Health and Family Planning commission of Shanghai Pudong New Area (No. PWYgy2018-10)

Conflicts of interest

None of the authors has any conflict of interest to disclose. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

References

[1] Reynolds EH. Introduction: epilepsy in the world. Epilepsia 2002;43(Suppl. 6):1–3.
[2] Britton JW. Antiepileptic drug withdrawal: literature review. Mayo Clin Proc 2002;77(12):1378–88.
[3] Brodie MJ, Barry SJ, Bamagous GA, Norrie JD, Kwan P. Patterns of treatment response in newly diagnosed epilepsy. Neurology 2012;78(20):1548–54.
[4] Abimbola S, Martiniuk AL, Hackett ML, Glozier N, Mohamed A, Anderson CS. Early predictors of remission in newly diagnosed epilepsy: a systematic approach to reviewing prognostic factor studies. Neurol Res 2014;36(1):1–12.
[5] Park KM, Hur Y, Kim HY, Ji KH, Hwang TG, Shin KJ, et al. Initial response to antiepileptic drugs in patients with newly diagnosed epilepsy. J Clin Neurosci 2014;21(6):923–6.
[6] Schiller Y, Najjar Y. Quantifying the response to antiepileptic drugs: effect of past treatment history. Neurology 2008;70(1):54–65.
[7] Lagunju IA, Asinobi A. Predictors of early seizure remission in Nigerian children with newly diagnosed epilepsy. Afr J Med Med Sci 2011;40(3):239–45.
[8] Hosseni M-P, Soltanian-Zadeh H, Elisevich KV, Pompili D. Cloud-based deep learning of big EEG data for epileptic seizure prediction. 2016 IEEE global conference on signal and information processing (GlobalSIP); 2016. p. 1151–5.
[9] Temko A, Thomas E, Marnane W, Lightbody G, Boylan GB. Performance assessment for EEG-based neonatal seizure detectors. Clin Neurophysiol 2011;122(3):474–82.
[10] Armananzas R, Alonso-Nanclares L, Defelipe-Oroquieta J, Kastanauskaite A, de Sola R, DeFelipe J, et al. Machine learning approach for the outcome prediction of temporal lobe epilepsy surgery. PLoS One 2013;8(4):e62819.
[11] Senders JT, Staples PC, Karhade AV, Zaki MM, Gormley WB, Broekman MLD, et al. Machine learning and neurosurgical outcome prediction: a systematic review. World Neurosurg 2018;109:476–486.e1.
[12] Memarian N, Kim S, Dewar S, Engel Jr J, Staba RJ. Multimodal data and machine learning for surgery outcome prediction in complicated cases of mesial temporal lobe epilepsy. Comput Biol Med 2015;64:67–78.
[13] Taylor PN, Sinha N, Wang Y, Vos SB, de Tisi J, Miserocchi A, et al. The impact of epilepsy surgery on the structural connectome and its relation to outcome. Neuroimage Clin 2018;18:202–14.
[14] Devinsky O, Dilley C, Ozery-Flato M, Aharonov R, Goldschmidt Y, Rosen-Zvi M, et al. Changing the approach to treatment choice in epilepsy using big data. Epilepsy Behav 2016;56:32–7.
[15] Colic S, Wither RG, Lang M, Zhang L, Eubanks JH, Bardakjian BL. Prediction of antiepileptic drug treatment outcomes using machine learning. J Neural Eng 2017;14(1):016002.
[16] Quinlan JR. C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc; 1993 (302 pp.).
[17] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
[18] Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Syst Appl 1998;13(4):18–28.
[19] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; San Francisco, California, USA 2939785. ACM; 2016. p. 785–94.
[20] Cox DR. The regression analysis of binary sequences. J R Stat Soc B Methodol 1958;20(2):215–42.
[21] Fisher RS, Boas WvE, Blume W, Elger C, Genton P, Lee P, et al. Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). Epilepsia 2005;46(4):470–2.
[22] Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, et al. ILAE official report: a practical clinical definition of epilepsy. Epilepsia 2014;55(4):475–82.
[23] Shen C, Du Y, Lu R, Zhang Y, Jin B, Ding Y, et al. Factors predictive of late remission in a cohort of Chinese patients with newly diagnosed epilepsy. Seizure 2016;37:20–4.
[24] Fisher RS, Cross JH, D'Souza C, French JA, Haut SR, Higurashi N, et al. Instruction manual for the ILAE 2017 operational classification of seizure types. Epilepsia 2017;58(4):531–42.
[25] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825–30.
[26] Rus ID, Marc P, Dinsoreanu M, Potolea R, Muresan RC. Classification of EEG signals in an object recognition task. Proc IEEE Intell Comput Commun Process; 2017. p. 391–5.
[27] Zhang JH, Han X, Zhao HW, Zhao D, Wang N, Zhao T, et al. Personalized prediction model for seizure-free epilepsy with levetiracetam therapy: a retrospective data analysis using support vector machine. Br J Clin Pharmacol 2018;84(11):2615–24.
[28] Eliades C, Papadopoulos H. Detecting seizures in EEG recordings using conformal prediction. Proceedings of machine learning Research; 2018. p. 171–86.
[29] Fawcett T. An introduction to ROC analysis. Pattern Recogn Lett 2006;27(8):861–74.
[30] Yang Y, Pedersen JO. A comparative study on feature selection in text categorization. Proceedings of the fourteenth international conference on machine learning. 657137. Morgan Kaufmann Publishers Inc; 1997. p. 412–20.
[31] Geelhoed M, Boerrigter AO, Camfield P, Geerts AT, Arts W, Smith B, et al. The accuracy of outcome prediction models for childhood-onset epilepsy. Epilepsia 2005;46(9):1526–32.
[32] V KR, Rajagopalan SS, Bhardwaj S, Panda R, Reddam VR, Ganne C, et al. Machine learning detects EEG microstate alterations in patients living with temporal lobe epilepsy. Seizure 2018;61:8–13.
[33] Del Gaizo J, Mofrad N, Jensen JH, Clark D, Glenn R, Helpern J, et al. Using machine learning to classify temporal lobe epilepsy based on diffusion MRI. Brain Behav 2017;7(10):e00801.
[34] Munsell BC, Wee CY, Keller SS, Weber B, Elger C, da Silva LA, et al. Evaluation of machine learning algorithms for treatment outcome prediction in patients with epilepsy based on structural connectome data. Neuroimage 2015;118:219–30.
[35] Berg AT, Rychlik K, Levy SR, Testa FM. Complete remission of childhood-onset epilepsy: stability and prediction over two decades. Brain 2014;137:3213–22 Pt 12.
[36] Hughes DM, Bonnett LJ, Czanner G, Komárek A, Marson AG, García-Fiñana M. Identification of patients who will not achieve seizure remission within 5 years on AEDs. Neurology 2018;91(22):e2035–44.