Epilepsy Research 181 (2022) 106888

Contents lists available at ScienceDirect

Epilepsy Research

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

Machine learning model to predict the efficacy of antiseizure medications in patients with familial genetic generalized epilepsy

Junhong Wu [a,1], Yan Wang [b,1], Ling Xiang [b], Yixue Gu [a], Yin Yan [a], Lulin Li [b], Xin Tian [a,*], Wei Jing [a,c,**], Xuefeng Wang [a,*]

[a] Department of Neurology, the First Affiliated Hospital of Chongqing Medical University, Chongqing Key Laboratory of Neurology, Chongqing, China
[b] School of Mathematics Science, Chongqing Normal University, Chongqing, China
[c] Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Tongji Shanxi Hospital, Third Hospital of Shanxi Medical University, Taiyuan, 030032, China

ARTICLE INFO

Keywords:
Machine learning
Antiseizure medication
Familial genetic generalized epilepsy
Random forest classifier

ABSTRACT

Objective: This study aimed to establish a machine learning model that can predict the efficacy of antiseizure medications (ASMs) in patients with familial genetic generalized epilepsy (GGE).
Methods: We prospectively followed up patients with familial GGE for at least 3 years between January 2007 and January 2017. We collected and analyzed the patients' demographic characteristics, medical history, and related auxiliary examinations. The results of the epileptic seizures were divided into two categories: seizure-free and drug-resistant epilepsy. We selected and trained thirteen classification models, i.e., random forest classifier, logistic regression, gradient boosting classifier, light gradient boosting machine, ridge classifier, linear discriminant analysis, support vector machine-linear kernel, extra tree classifier, Ada boost classifier, naive Bayes classifier, decision tree classifier, K neighbors classifier, and quadratic discriminant analysis, to get the best performing classification model.
Results: A total of 854 patients with familial GGE were included in the study after excluding 89 who were lost to follow-up. Among them, 631 patients with familial GGE became seizure-free, and 223 developed drug-resistant epilepsy with a 74.89% remission rate. Among the 13 models, the random forest classifier model was the most effective with an accuracy of 91.23% and an F1 score of 84.21%. Among the 18 patient characteristics, the most effective indicators of the final treatment results were the number of seizure types experienced, response to the first drug, prior treatment duration and number of pre-treatment seizures.
Significance: The random forest classifier model can be used to early predict the results of ASM treatment based on the clinical data of patients with familial GGE. This finding can help clinicians make timely adjustments to treatment strategies and improve patients' prognosis.

1. Introduction

Epilepsy is one of the most common diseases of the nervous system affecting approximately 70 million individuals worldwide ("Global, regional, and national burden of neurological disorders, : a systematic analysis for the Global Burden of Disease Study, ., ", 1990–, 2016); particularly, genetic generalized epilepsy (GGE) accounts for 15–20% of all the incidences of epilepsy. (Jallon and Latour, 2005) Patients with GGE usually have a positive prognosis; however, 15–40% develop drug resistance. (Mohanraj and Brodie, 2007; Seneviratne et al., 2012) This leads to impaired cognitive and social functions, which negatively affect the patients' mental health and quality of life and even cause death. (Devinsky, 1999; Laxer et al., 2014).

Studies of twins and epidemiology have shown that genetic factors

List of abbreviations: ASMs, antiseizure medications; EEG, electroencephalogram; GGE, genetic generalized epilepsy; ILAE, International Anti-Epilepsy Alliance; MRI, magnetic resonance imaging; AUC, area under the curve; ROC, receiver operating characteristic.

* Corresponding authors.
** Corresponding author at: Department of Neurology, the First Affiliated Hospital of Chongqing Medical University, Chongqing Key Laboratory of Neurology, Chongqing, China.
E-mail addresses: xintian@cqmu.edu.cn (X. Tian), jingweistar@163.com (W. Jing), xfyp@163.com (X. Wang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.eplepsyres.2022.106888
Received 16 October 2021; Received in revised form 18 January 2022; Accepted 9 February 2022
Available online 11 February 2022

are the only definitive cause of GGE (Mullen et al., 2018; Scheffer et al., 2017); however, most patients are sporadic with no family history. (Briellmann et al., 2001) However, patients with familial GGE are commonly encountered in clinical practice. In these patients, prompt and accurate prediction of the prognosis is necessary for selecting the appropriate drug therapy and treatment plan based on the patients' characteristics to improve their prognosis. To address this, considerable advancements have been made on the applicability of machine learning in the field of epilepsy. It has been widely used in different studies involving the imaging studies of patients with epilepsy, (Del Gaizo et al., 2017) electroencephalogram (EEG) studies, (V et al., 2018) prognosis of patients with epilepsy, (Munsell et al., 2015; Wissel et al., 2021, 2020, 2019) and the prediction of the patient's response to antiseizure medications (ASMs). (An et al., 2018; Devinsky et al., 2016; Yao et al., 2019) Previous studies have identified different clinical predictors of the patients' with GGE response to drug treatment; however, most studies have often used traditional methods and focused on the study of predictors of the GGE response to ASM in the whole population. (Gesche et al., 2020; Voll et al., 2015) To date, there have been no studies on the predictors of the prognosis of patients with familial GGE responding to ASMs through a machine learning model.

Therefore, this study aimed to develop a machine learning model that specifically predicts the prognosis of patients with familial GGE on drug therapy, which is of great significance for clinicians to design reasonable treatment plans for patients with familial GGE.

2. Material and methods

2.1. Patient and process

This study included patients with familial GGE diagnosed at the Epilepsy Center of the First Affiliated Hospital of Chongqing Medical University from January 1997 to January 2017.

The inclusion criteria were as follows: (1) patients diagnosed with GGE by the same epilepsy team according to the definition of the International Anti-Epilepsy Alliance (ILAE) in 1989 ("Proposal for revised classification of epilepsies and epileptic syndromes. Commission on Classification and Terminology of the International League Against Epilepsy., ", 1989), (2) patients with a family history of epilepsy, with their first-, second-, or third-degree relatives having epilepsy, and (3) patients with normal brain magnetic resonance imaging (MRI) findings (no cerebral infarction, hippocampal sclerosis, meningioma, and other secondary epileptic lesions).

The exclusion criteria were as follows: (1) patients who refused to receive ASM treatment, (2) patients who received treatments other than ASMs, such as epilepsy surgery and vagus nerve stimulation during the follow-up period, (3) patients who experienced severe liver or kidney disease, and (4) patients with irregular medication during treatment and poor compliance.

Upon the patient's first consultation, a specially trained physician conducted a detailed on-site interview of the witness and the patients and recorded the demographic information (sex, living area, occupation, and educational level), medical history (age at seizure onset, age at ASM treatment, seizure type at onset, number of seizures before treatment, prior treatment duration, history of febrile seizure, prior treatment duration, seizure type at onset, and family history of epilepsy, etc.), and related auxiliary examination results (EEG and cranial MRI). Finally, the same experienced epilepsy team diagnosed the patient according to the diagnostic criteria of the ILAE in 1989.

For all the diagnosed patients, the same epilepsy team selected an appropriate ASM after considering the efficacy and safety of the drug. According to the individual situation of the patient, valproic acid, levetiracetam, lamotrigine, or topiramate was mainly used. Subsequently, the patient was instructed to consult the outpatient clinic of our center for the evaluation of the treatment effect every 4 weeks for the first 6 months and then at least every 3 months for the evaluation through outpatient or telephone follow-up. In each follow-up, the epilepsy experts recorded the patient's response to ASM in detail, including the dosage form of the drug and the related adverse reactions; appropriate and timely adjustments to the treatment plan were devised according to the control of seizures (i.e., to increase, decrease, or switch medication).

This study was approved by the Ethics Committee of the First Affiliated Hospital of Chongqing Medical University. All the patients or their families were informed about the study, and signed consent forms were obtained, allowing the use of patients' medical record data for research.

2.2. Definition

Our data collection began in 1997; therefore, we classified the types of GGE according to the 1989 ILAE revised edition, which is divided into absence, myoclonic, and general tonic-clonic seizures. (Commission on Classification and Terminology of the International League Against Epilepsy, 1989).

The first drug was defined as the first drug used after the patient was diagnosed with epilepsy; if the drug was changed due to an adverse reaction, the drug without side effects was defined as the first effective treatment drug. The response to the first drug was defined as the percentage decrease in the frequency of seizures from baseline (the average number of seizures per month for the 12 months before treatment [if the time is less than 12 months, divide the total number of seizures by the time]). "Control" was defined as the absence of seizures when the first drug was administered; "excellent" was defined as a reduction in seizure frequency $> 75\%$ during the administration of the first drug; "effective" was defined as a reduction in seizure frequency by 51–75% during the administration of the first drug; "ineffective" was defined as a reduction in seizure frequency $< 50\%$ during the administration of the first drug; and "deteriorative" was defined as an increase in the frequency of seizures during the administration of the first drug.

We divided the results of ASM treatment into two modes: (1) "seizure-free" was defined as the absence of any type of seizure in the patient within 12 months at the end of the follow-up period, and (2) "drug-resistant epilepsy" was defined as the use of two or more appropriate broad-spectrum ASMs, with persistent seizures. To ensure the effectiveness of the drug concentration, we required each ASM to be used continuously for at least 3 months to evaluate the efficacy of the drug under well-tolerated conditions.

2.3. Machine learning process

2.3.1. Experimental data

In this study, the data of 854 patients with familial GGE were used, among them, 631 became seizure-free, 223 developed drug-resistant epilepsy; 80% of the original data were used as the training set, and the remaining 20% were used as the test set. Data pre-processing included two steps: (1) min-max standardization of continuous variables (age at seizure onset, age at ASM treatment, number of seizures before treatment, and prior treatment duration) and (2) one-hot encoding that was conducted for categorical variables (living area, occupation, educational level, catamenial epilepsy, seizure type at onset, number of seizure types ever experienced, respond to the first drug, psychiatric condition, history of febrile seizure, intellectual disability, developmental delay, status epilepticus, nocturnal seizures, and EEG), which were converted into a sparse vector with a value of 1 or 0. Our experiment included two outcomes, i.e., seizure-free and drug-resistant epilepsy. Table 1 shows the details of the selected variables and outcomes.

2.3.2. Related models

This study was a binary classification problem; therefore, 13 common machine learning models were used in the study for learning and prediction: random forest classifier, logistic regression, gradient boosting classifier, light gradient boosting machine, ridge classifier, linear

Table 1
The demographic and medical history characteristics of the patients.

Variable	Total (n = 854)	Seizure-free (n = 631)	Drug-resistant epilepsy (n = 223)	*P*
Living area				
Urban area	354	269	85	0.240
Rural area	500	362	138	
Occupation				
Student	484	354	130	0.851
Employment	211	158	53	
Unemployment	159	119	40	
Educational level				
Illiteracy or primary school	300	281	82	0.654
Secondary school	300	226	74	
Senior school	181	130	51	
College or above	73	57	16	
Catamenial epilepsy				
Male	455	339	116	< 0.001
Female with catamenial epilepsy	16	3	13	
Female with no catamenial epilepsy	383	289	94	
Age at seizure onset (year)	13 (8–19)	13.00 (8–19)	12(6–19)	0.055
Age at ASM treatment (year)	15 (10–22)	14 (10–22)	15 (10–22)	0.378
Seizure type at onset				
Absence seizures	284	195	89	0.014
Myoclonic seizures	295	234	61	
General tonic-clonic seizures	275	202	73	
Number of seizures before treatment	3 (3–14)	3 (3–10)	14 (9–18)	< 0.001
Prior treatment duration (year)	0.83 (0.33–3)	0.5 (0.25–2)	2.5 (1.5–4)	< 0.001
Number of seizure types ever experienced				
1	458	443	15	< 0.001
2	180	140	40	
3	216	48	168	
Respond to the first drug				
Control	181	170	11	< 0.001
Excellent	287	236	51	
Effective	102	92	10	
Ineffective	192	112	80	
Deteriorative	92	21	71	
Psychiatric condition				
Positive	88	63	25	0.605
Negative	766	568	198	
History of febrile seizure				
Negative	760	568	192	0.108
Positive	94	63	31	
Intellectual disability				
Positive	16	9	7	0.105
Negative	838	622	216	
Developmental delay				
Negative	841	623	218	0.307
Positive	13	8	5	
Status epilepticus				
Positive	223	114	109	< 0.001
Negative	631	517	114	
Nocturnal seizures				
Positive	45	28	17	0.067
Negative	809	603	206	
EEG				
Epileptiform discharge	431	300	131	0.013
Normal	298	236	62	
NA	125	95	30	

Note: NA: not available, ASM: antiseizure medication, EEG: electroencephalogram.

discriminant analysis, support vector machine-linear kernel, extra tree classifier, Ada boost classifier, naive Bayes classifier, decision tree classifier, K neighbors classifier, and quadratic discriminant analysis.

2.3.3. Model performance evaluation metrics

In this study, five evaluation indicators were used to measure the model performance: accuracy, area under the receiver operating characteristic (ROC) curve, sensitivity, specificity, and F1-score.

2.3.4. Model comparison

In this study, we used the PyCaret library to establish an environment, i.e., data type inference, data preprocessing, and dataset division. Pycaret will automatically infer whether the feature is a numerical type or a categorical type, which need to be checked if its feature inference is true by the experimenter manually. Then, it will perform the corresponding preprocessing, and divide the dataset into a training set and a test set at a ratio of 8:2. The 13 machine learning models introduced earlier (parameter defaults) were used on the divided training set to perform ten-fold cross-validation to obtain the accuracy, area under the ROC curve (AUC), sensitivity, specificity and F1-score of each model. The data were arranged in descending order according to the average value of the accuracy (Table 2). Each model's parameters were also described in Table 2.

As shown in Table 2, the averages of the accuracy, AUC, F1-score, and specificity of the random forest classifier were higher compared to those of the other models, which were 90.2%, 94.9%, 80.6%, and 87.5% respectively. Moreover, its sensitivity was also relatively high at 90.8%. Therefore, the random forest classifier model was selected as the prognostic prediction model for patients with familial GGE.

2.3.5. Parameter selection of prediction model

In developing the random forest classifier model, the number of decision trees and variables tried at each split was tuned according to the model errors. The association between the model error and the number of decision trees is shown in Fig. 1. It can be observed that when the number of decision trees reaches 80, the random forest classifier model tends to stabilize. Simultaneously, the three errors had smaller values. Therefore, the numbers of trees and variables tried at each split were set to 80 and 4, respectively, allowing the final determination of the prediction model. Regarding the training set of 683 cases, the out-of-bag error of the model was 10.83%.

2.4. Statistical analyses

Statistical analyses were performed using the R software (version 4.0.5; Comprehensive R Archive Network: https://cran.r-project.org/mirrors.html). Quantitative data following non-normal distributions were reported as medians and interquartile ranges (M [P25, P75]) and were compared using the nonparametric test. Categorical variables were analyzed using the chi-squared test.

3. Results

3.1. Patients characteristics

Overall, 89 patients who were lost to follow-up and seven patients who died from January 1997 to January 2017 were excluded. Finally, 854 patients with familial GGE were included in this study (Fig. 2). Among them, 631 patients with familial GGE became seizure-free, and 223 developed drug-resistant epilepsy with a 74.89% remission rate. Catamenial epilepsy, seizure type at onset, number of seizures before treatment, prior treatment duration, number of seizure types ever experienced, response to the first drug, status epilepticus, and

Table 2
Performance comparison of the different models in the training set.

Model	Parameters	Accuracy (%)	AUC (%)	F1-score (%)	Sensitivity (%)	Specificity (%)
Random forest classifier	'criterion': gini, 'max features': auto, 'min samples leaf': 1, 'min samples split': 2, 'n estimators': 100	90.2	94.9	80.6	90.8	87.5
Ridge classifier	'alpha': 1.0, 'normalize': False,'solver': auto,'tol': 0.001	89.2	0.00	78.8	94.3	83.3
Light gradient boosting machine	'boosting type': gbdt, 'learning rate': 0.1, 'max depth': − 1, 'n estimators': 100, 'num leaves': 31, 'subsample for bin': 200000	88.9	94.2	78.5	91.3	80.0
Logistic regression	'C': 1.0, 'intercept scaling': 1, 'max iter': 1000, 'penalty': l2, 'tol': 0.0001	88.7	94.0	77.9	95.0	82.0
Linear discriminant analysis	'solver': svd, 'tol': 0.0001	88.7	94.3	78.4	94.2	80.0
Extra tree classifier	'criterion': gini, 'max features': auto, 'min samples leaf': 1, 'min samples split': 2, 'n estimators': 100	88.3	93.9	77.3	93.5	83.0
Gradient boosting classifier	'learning rate': 0.1, 'max depth': 3, 'min samples leaf': 1, 'min samples split': 2, 'n estimators': 100, 'subsample': 1.0	88.3	94.0	76.9	91.5	85.7
Ada boost classifier	'algorithm': SAMME.R, 'learning rate': 1.0, 'n estimators': 50	86.8	93.9	73.7	93.5	81.3
K neighbors classifier	'leaf size': 30, 'n neighbors': 5, 'p': 2, 'weights': uniform	86.5	89.6	70.4	91.3	80.0
Support vector machine-linear kernel	'alpha': 0.0001, 'epsilon': 0.1, 'l1 ratio': 0.15, 'penalty': l2	86.5	0.00	71.9	89.2	80.5
Naive Bayes classifier	'var moothing': 1e-09	85.5	92.2	71.5	91.4	67.3
Decision tree classifier	'criterion': gini, 'min samples leaf': 1, 'min samples split': 2, 'splittert': best	83.6	78.4	67.5	85.8	75.7
Quadratic discriminant analysis	'store covariance': False, 'tol': 0.0001	65.1	60.1	39.1	73.6	32.3

Note: AUC: area under the receiver operating curve.

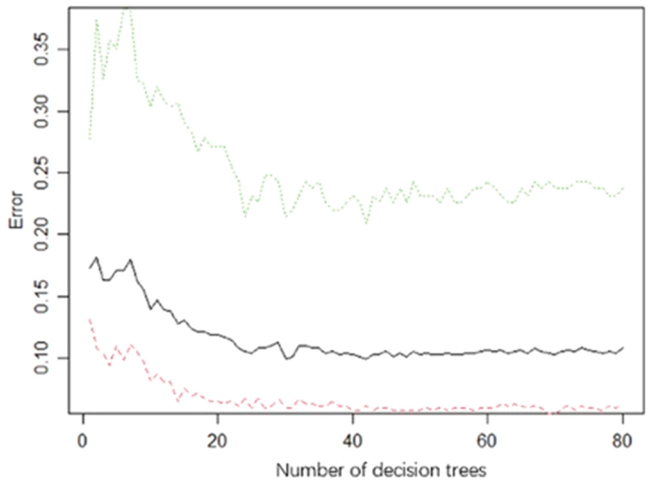

Fig. 1. Graph of the association between the model error and the number of decision trees. The red curve represents the prediction error of the model for class 0 data under different numbers of decision trees, and the green curve represents the model's prediction error for class 1 data under different numbers of decision trees. The black curve represents the model's out-of-bag (OOB) estimate of the error rate, that is, OOB error.

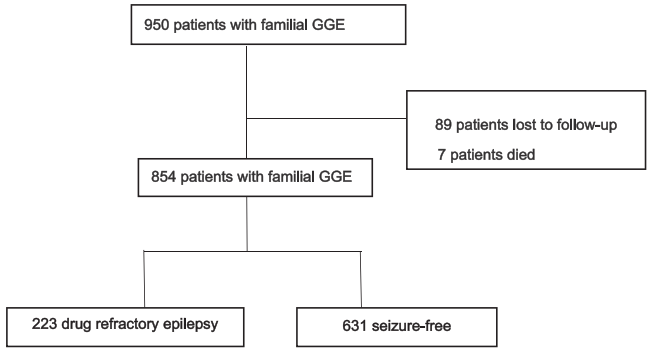

Fig. 2. Study flowchart.

epileptiform discharge differed between the seizure free group and drug-resistance epilepsy group. Table 1 shows the demographic and medical histories of the patients.

3.2. *Using machine learning for early prediction of the efficacy of ASMs in patients with familial GGE*

3.2.1. *Evaluation of the Random Forest Classifier model*

After the model was successfully constructed, predictions were made for the test set of 171 cases. A confusion matrix was used (Fig. 3); additionally, the AUC (Fig. 4), sensitivity, specificity, accuracy, and F1 score (Table 3) were used to measure the prediction effect of the model.

The confusion matrix is an important tool for evaluating the model performance. It can be observed from Fig. 4 that the FP (false positive) and FN (false negative) cases predicted by the model are 2 and 13, respectively, which indicates the good performance of the constructed prediction model.

The receiver operating characteristic (ROC) curve of the model is shown in Fig. 4, and the AUC was 92.6%, indicating the good

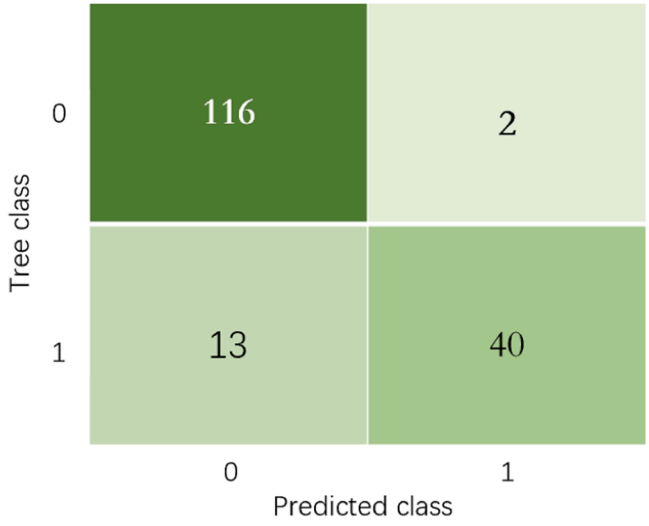

Fig. 3. Confusion matrix and receiver operating characteristic curves of the random forest classifier model.

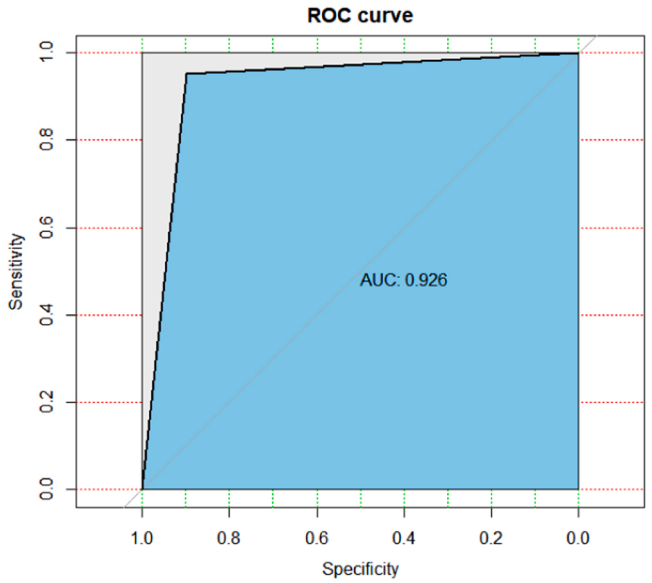

Fig. 4. Receiver operating characteristic curves of the random forest classifier model.

Table 3
Performance of the developing random forest classifier model on the test set.

Model	Accuracy (%)	AUC (%)	F1-score (%)	Sensitivity (%)	Specificity (%)
Random forest classifier	91.23	92.6	84.21	89.92	95.24

performance and clinical diagnostic value of the model.

The accuracy of the model was defined as the proportion of the number of samples with accurate predictions to the total number of samples, and the accuracy of the constructed model was 91.2%, which indicates that the model has a good classification effect. The sensitivity (89.9%) and specificity (95.2%) of the model represent the proportions of the samples with correct predictions in all the positive and negative cases, respectively. It can be observed that the random forest classifier model has a good ability to recognize positive and negative examples. The F1 value is 84.2, which simultaneously considers both the accuracy and recall rate to maximize these parameters and achieve a balance.

3.2.2. Prognosis prediction of familial GGE patients

The importance of the variables was determined according to the mean decrease in the Gini impurity. As shown in Fig. 5, the number of seizure types ever experienced was the most important variable in predicting the prognosis patients with familial GGE, followed by prior treatment duration (year), number of pre-treatment seizures, and response to the first drug.

4. Discussion

In this study, we evaluated 13 machine learning models; among which, the random forest classifier model had the best performance in terms of the averages of accuracy, AUC, and F1-score through parameter optimization. Furthermore, we analyzed the 18 characteristics of patients with familial GGE and found that the best distinguishing factors for final treatment outcomes were the number of seizure types experienced, response to the first drug, prior treatment duration and number of pre-treatment seizures. These factors can help clinicians better predict the prognosis of patients with familial GGE and make appropriate and

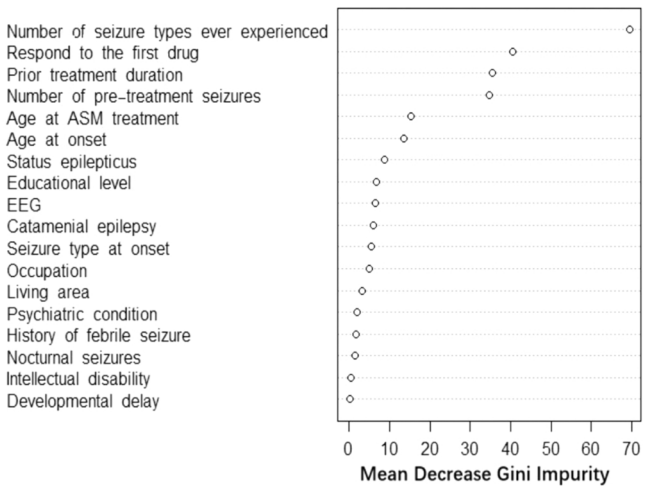

Fig. 5. Rank of importance of variables from the random forest classifier model.

timely adjustments to the treatment plan to improve the patient's prognosis.

Among the 854 patients with familial GGE, 74.89% demonstrated a satisfactory response to the drug administered, which is approximately consistent with the findings of previous studies regarding the remission rate of the whole population of patients with GGE. (Choi et al., 2020; Seneviratne et al., 2012) Many factors may influence the outcomes of patients with GGE; among which, research background, study design, study cohort, follow-up time, and the definition of duration of remission of seizures (Seneviratne et al., 2012) are particularly important. Simultaneously, among the 18 characteristics, we found that the number of seizure types experienced is an important predictive factor of the prognosis of patients with familial GGE who are taking ASMs, which is consistent with previous studies. (Trinka et al., 2004) Moreover, Gelisse et al. (Gelisse et al., 2001) conducted a 1-year follow-up study of patients with juvenile myoclonic epilepsy. Among the 155 patients, they found that 62.5% of patients with drug-resistant epilepsy had absent, myoclonic, and generalized tonic-clonic seizures. In contrast, 23.3% of non-resistant patients had three types of seizures. All drug-resistant patients had no myoclonus as the only type of seizure, and there were no seizures or myoclonic seizures.

Additionally, we found that that first drug response affected the patient's overall drug response, which is in line with the result of a previous study. Mohanraj et al. (Mohanraj and Brodie, 2006) followed up 780 patients newly diagnosed with epilepsy in Glasgow and found that the overall response rates of the patients to the first, second, and third drugs were 50.4%, 10.7%, and 2.7%, respectively. Kwan et al. (Kwan and Brodie, 2000) conducted a prospective follow-up of 525 patients with epilepsy. They found that among the 470 patients who received no prior treatment, 222 had no seizures during treatment with the first ASM. Furthermore, among the patients with no first drug response, 11% who failed treatment due to lack of efficacy were seizure-free, which was less than the treatment failure due to unbearable side effects (41%) or special reactions (55%). Other studies have drawn similar conclusions. (Dlugos et al., 2001; Elwes et al., 1984; Mohanraj and Brodie, 2013a; Sillanpää, 1993).

Our study also found prior treatment duration and number of seizures before treatment had a significant effect on patients' prognosis. Previous studies have shown that recurrent seizures may cause shrinking of the hippocampus, resulting in hippocampal neuron loss and mossy fiber sprouting, which further enhance its excitatory recurring cycle. (Berg et al., 2010; Kälviäinen and Salmenperä, 2002; Kwan and Brodie, 2000; Theodore et al., 1999) Additionally, failure to control seizures with drugs can progressively lead to drug-resistant epilepsy. Mohanraj

et al. (Mohanraj and Brodie, 2013b) believe that early treatment of patients with epilepsy minimizes the number of subsequent seizures and improves the prognosis of patients. Similarly, Reynolds et al. (Reynolds et al., 1983) hold that early treatment of seizures is the key to preventing drug-resistant seizures. In contrast, Feksi et al. (Feksi et al., 1991) found different results. They prospectively evaluated the anti-epileptic treatment plan of patients with general tonic-clonic seizures in rural and semi-urban areas of Kenya. Among the 249 patients, 202 (81%) had not received any treatment, 52% had seizures for more than 5 years, and 38% experienced at least 100 episodes of seizures. Finally, they found that prior ASM treatment duration and number of seizures before treatment had no influence on the patient's treatment effect.

The performance of the random forest model in our study achieved better accuracy and F1-score on the test dataset than on the training set, which is likely because the data distribution of the test set is more in line with the rules model learned from the training set, and the small number of our dataset. More data can be collected to get better model in the future. Moreover, this can also be explained by the fact that the final results of accuracy and F1-score in the training set is the average of the random forest model's ten-fold cross-validation on the training set (ten random validation sets). Therefore, there will be different degree of accuracy and F1-score in the results of ten times of verification, so it is also possible that the result after averaging is slightly lower than the result of only one run.

To the best of our knowledge, this is the first study to apply machine learning methods to predict the efficacy of ASMs in patients with familial GGE. The advantages of our study include having a large number of patients with familial GGE, with the same team performed the whole diagnosis, and subsequent follow-up adjustments to drug treatment. In addition, this is the first study to use the newly open source PyCaret to compare the performance of 13 classic machine learning prediction models on the training set and obtain a prediction model with the best performance, namely, the random forest classifier model. Furthermore, we used parameter optimization, a function of random forest classifier model, to obtain the results for predicting the prognosis of patients with familial GGE. The prediction results and performance evaluation indicators of the model on the test set data were realized using the R software. The method adopted in this study makes appropriate use of the advantages of the novel PyCaret and traditional R software, making the study more efficient and effective. Despite these strengths, our study has some limitations. First, there is a certain randomness in the data selection when we conducted model training, and the distribution of the amount of data for the final outcome in the experimental data was unbalanced. In the future, some advanced machine learning techniques especially for unbalanced data should be considered, such as the combination of machine learning methods with techniques addressing the problem of unbalanced classification (Eitrich et al., 2007; Fernández et al., 2018; Scholar et al., 2021), ensemble learning algorithm (Tang et al., 2021), ect. Second, this was a single-center study, and there may have been selection bias in the inclusion of cases. Large-scale multicenter prospective randomized controlled trial studies can be conducted in the future, and more relevant factors can be included to improve the accuracy of prediction for the whole population.

In conclusion, our study shows that the number of seizure types experienced, response to the first drug, prior treatment duration and number of pre-treatment seizures were the most effective indicators for the final treatment results of patients with familial GGE. Furthermore, the random forest classifier model can be used to make early predictions regarding the efficacy of ASMs in patients with familial GGE, which can help clinicians select the best treatment strategies earlier and improve the patient's prognosis.

Funding

This work was supported by the National Natural Science Foundation of China (No. 82001378 and No. 11901071).

Declaration of Competing Interest

The authors report no declarations of interest.

References

An, S., Malhotra, K., Dilley, C., Han-Burgess, E., Valdez, J.N., Robertson, J., Clark, C., Westover, M.B., Sun, J., 2018. Predicting drug-resistant epilepsy - A machine learning approach based on administrative claims data. Epilepsy Behav. 89, 118–125. https://doi.org/10.1016/j.yebeh.2018.10.013.

Berg, A.T., Berkovic, S.F., Brodie, M.J., Buchhalter, J., Cross, J.H., Van Emde Boas, W., Engel, J., French, J., Glauser, T.A., Mathern, G.W., Moshé, S.L., Nordli, D., Plouin, P., Scheffer, I.E., 2010. Revised terminology and concepts for organization of seizures and epilepsies: Report of the ILAE Commission on Classification and Terminology, 2005-2009. Epilepsia 51, 676–685. https://doi.org/10.1111/j.1528-1167.2010.02522.x.

Briellmann, R.S., Torn-Broers, Y., Berkovic, S.F., 2001. Idiopathic generalized epilepsies: Do sporadic and familial cases differ? Epilepsia 42, 1399–1402. https://doi.org/10.1046/j.1528-1157.2001.03201.x.

Choi, H., Detyniecki, K., Bazil, C., Thornton, S., Crosta, P., Tolba, H., Muneeb, M., Hirsch, L.J., Heinzen, E.L., Sen, A., Depondt, C., Perucca, P., Heiman, G.A., 2020. Development and validation of a predictive model of drug-resistant genetic generalized epilepsy. Neurology. https://doi.org/10.1212/WNL.0000000000010597.

Commission on Classification and Terminology of the International League Against Epilepsy, 1989. 1989. Proposal for Revised Classification of Epilepsies and Epileptic Syndromes - - 1989 - Epilepsia - Wiley Online Library. Epilepsia 30, 389–399.

Devinsky, O., 1999. Patients with refractory seizures. N. Engl. J. Med 340, 1565–1570. https://doi.org/10.1056/NEJM199905203402008.

Devinsky, O., Dilley, C., Ozery-Flato, M., Aharonov, R., Goldschmidt, Y., Rosen-Zvi, M., Clark, C., Fritz, P., 2016. Changing the approach to treatment choice in epilepsy using big data. Epilepsy Behav. 56, 32–37. https://doi.org/10.1016/j.yebeh.2015.12.039.

Dlugos, D.J., Sammel, M.D., Strom, B.L., Farrar, J.T., 2001. Response to first drug trial predicts outcome in childhood temporal lobe epilepsy. Neurology 57, 2259–2264. https://doi.org/10.1212/WNL.57.12.2259.

Eitrich, T., Kless, A., Druska, C., Meyer, W., Grotendorst, J., 2007. Classification of highly unbalanced CYP450 data of drugs using cost sensitive machine learning techniques. ChemInform 38, 92–103. https://doi.org/10.1002/chin.200715216.

Elwes, R.D., Johnson, A.L., Shorvon, S.D., Reynolds, E.H., 1984. The prognosis for seizure control in newly diagnosed epilepsy. N. Engl. J. Med 311, 944–947. https://doi.org/10.1056/NEJM198410113111503.

Feksi, A.T., Kaamugisha, J., Sander, J.W.A.S., Shorvon, S.D., Gatiti, S., 1991. Comprehensive primary health care antiepileptic drug treatment programme in rural and semi-urban Kenya. Lancet 337, 406–409. https://doi.org/10.1016/0140-6736(91)91176-U.

Fernández, A., García, S., Herrera, F., Chawla, N.V., 2018. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. J. Artif. Intell. Res 61, 863–905. https://doi.org/10.1613/jair.1.11192.

Del Gaizo, J., Mofrad, N., Jensen, J.H., Clark, D., Glenn, R., Helpern, J., Bonilha, L., 2017. Using machine learning to classify temporal lobe epilepsy based on diffusion MRI. Brain Behav. 7, 1–7. https://doi.org/10.1002/brb3.801.

Gelisse, P., Genton, P., Thomas, P., Rey, M., Samuelian, J.C., Dravet, C., 2001. Clinical factors of drug resistance in juvenile myoclonic epilepsy. J. Neurol. Neurosurg. Psychiatry 70, 240–243. https://doi.org/10.1136/jnnp.70.2.240.

Gesche, J., Hjalgrim, H., Rubboli, G., Beier, C.P., 2020. The clinical spectrum of familial and sporadic idiopathic generalized epilepsy. Epilepsy Res 165, 106374. https://doi.org/10.1016/j.eplepsyres.2020.106374.

Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016, 2019. Lancet Neurol. 18, 459–480. https://doi.org/10.1016/S1474-4422(18)30499-X.

Jallon, P., Latour, P., 2005. Epidemiology of idiopathic generalized epilepsies. Epilepsia 46, 10–14. https://doi.org/10.1111/j.1528-1167.2005.00309.x.

Kälviäinen, R., Salmenperä, T., 2002. Do recurrent seizures cause neuronal damage? A series of studies with MRI volumetry in adults with partial epilepsy. Prog. Brain Res 135, 279–295. https://doi.org/10.1016/S0079-6123(02)35026-X.

Kwan, P., Brodie, M.J., 2000. Early identification of refractory epilepsy. N. Engl. J. Med. 342, 314–319. https://doi.org/10.1056/NEJM200002033420503.

Laxer, K.D., Trinka, E., Hirsch, L.J., Cendes, F., Langfitt, J., Delanty, N., Resnick, T., Benbadis, S.R., 2014. The consequences of refractory epilepsy and its treatment. Epilepsy Behav. 37, 59–70. https://doi.org/10.1016/j.yebeh.2014.05.031.

Mohanraj, R., Brodie, M.J., 2013a. Early predictors of outcome in newly diagnosed epilepsy. Seizure 22, 333–344. https://doi.org/10.1016/j.seizure.2013.02.002.

Mohanraj, R., Brodie, M.J., 2013b. Early predictors of outcome in newly diagnosed epilepsy. Seizure 22, 333–344. https://doi.org/10.1016/j.seizure.2013.02.002.

Mohanraj, R., Brodie, M.J., 2007. Outcomes of newly diagnosed idiopathic generalized epilepsy syndromes in a non-pediatric setting. Acta Neurol. Scand. 115, 204–208. https://doi.org/10.1111/j.1600-0404.2006.00791.x.

Mohanraj, R., Brodie, M.J., 2006. Diagnosing refractory epilepsy: Response to sequential treatment schedules. Eur. J. Neurol. 13, 277–282. https://doi.org/10.1111/j.1468-1331.2006.01215.x.

Mullen, S.A., Berkovic, S.F., Lowenstein, D.H., Kato, M., Cross, H., Satishchandra, P., De Jonghe, P., Goldman, A., Petrou, S., Tan, N.C.K., Helbig, I., Mefford, H.C., Jiang, Y.,

2018. Genetic generalized epilepsies. Epilepsia 59, 1148–1153. https://doi.org/10.1111/epi.14042.

Munsell, B.C., Wee, C.-Y., Keller, S.S., Weber, B., Elger, C., da Silva, L.A.T., Nesland, T., Styner, M., Shen, D., Bonilha, L., 2015. Evaluation of machine learning algorithms for treatment outcome prediction in patients with epilepsy based on structural connectome data. Neuroimage 118, 219–230. https://doi.org/10.1016/j.neuroimage.2015.06.008.

Proposal for revised classification of epilepsies and epileptic syndromes. Commission on Classification and Terminology of the International League Against Epilepsy, 1989. Epilepsia 30, 389–399. https://doi.org/10.1111/j.1528-1157.1989.tb05316.x.

Reynolds, E.H., Elwes, R.D., Shorvon, S.D., 1983. Why does epilepsy become intractable? Prevention of chronic epilepsy. Lancet (Lond., Engl.) 2, 952–954. https://doi.org/10.1016/s0140-6736(83)90462-2.

Scheffer, I.E., Berkovic, S., Capovilla, G., Connolly, M.B., French, J., Guilhoto, L., Hirsch, E., Jain, S., Mathern, G.W., Moshé, S.L., Nordli, D.R., Perucca, E., Tomson, T., Wiebe, S., Zhang, Y.H., Zuberi, S.M., 2017. ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. Epilepsia 58, 512–521. https://doi.org/10.1111/epi.13709.

Scholar, M.T., Singh, K., Prof, A., Avinash, P., 2021. Data Balanc. Apply Mach. Learn. Approach Unbalanced Cancer Dataset 7, 2737–2742.

Seneviratne, U., Cook, M., D'Souza, W., 2012. The prognosis of idiopathic generalized epilepsy. Epilepsia 53, 2079–2090. https://doi.org/10.1111/j.1528-1167.2012.03723.x.

Sillanpää, M., 1993. Remission of seizures and predictors of intractability in long-term follow-up. Epilepsia 34, 930–936. https://doi.org/10.1111/j.1528-1157.1993.tb02114.x.

Tang, J., Henderson, A., Gardner, P., 2021. Exploring AdaBoost and Random Forests machine learning approaches for infrared pathology on unbalanced data sets. Analyst 146, 5880–5891. https://doi.org/10.1039/d0an02155e.

Theodore, W.H., Bhatia, S., Hatta, J., Fazilat, S., DeCarli, C., Bookheimer, S.Y., Gaillard, W.D., 1999. Hippocampal atrophy, epilepsy duration, and febrile seizures in patients with partial seizures. Neurology 52, 132–136. https://doi.org/10.1212/wnl.52.1.132.

Trinka, E., Baumgartner, S., Unterberger, I., Unterrainer, J., Luef, G., Haberlandt, E., Bauer, G., 2004. Long-term prognosis for childhood and juvenile absence epilepsy. J. Neurol. 251, 1235–1241. https://doi.org/10.1007/s00415-004-0521-1.

Voll, A., Hernández-Ronquillo, L., Buckley, S., Téllez-Zenteno, J.F., 2015. Predicting drug resistance in adult patients with generalized epilepsy: A case-control study. Epilepsy Behav. 53, 126–130. https://doi.org/10.1016/j.yebeh.2015.09.027.

V, K.R., Rajagopalan, S.S., Bhardwaj, S., Panda, R., Reddam, V.R., Ganne, C., Kenchaiah, R., Mundlamuri, R.C., Kandavel, T., Majumdar, K.K., Parthasarathy, S., Sinha, S., Bharath, R.D., 2018. Machine learning detects EEG microstate alterations in patients living with temporal lobe epilepsy. Seizure 61, 8–13. https://doi.org/10.1016/j.seizure.2018.07.007.

Wissel, B.D., Greiner, H.M., Glauser, T.A., Holland-Bouley, K.D., Mangano, F.T., Santel, D., Faist, R., Zhang, N., Pestian, J.P., Szczesniak, R.D., Dexheimer, J.W., 2020. Prospective validation of a machine learning model that uses provider notes to identify candidates for resective epilepsy surgery. Epilepsia 61, 39–48. https://doi.org/10.1111/epi.16398.

Wissel, B.D., Greiner, H.M., Glauser, T.A., Mangano, F.T., Santel, D., Pestian, J.P., Szczesniak, R.D., Dexheimer, J.W., 2019. Investigation of bias in an epilepsy machine learning algorithm trained on physician notes. Epilepsia 60, e93–e98. https://doi.org/10.1111/epi.16320.

Wissel, B.D., Greiner, H.M., Glauser, T.A., Pestian, J.P., Kemme, A.J., Santel, D., Ficker, D.M., Mangano, F.T., Szczesniak, R.D., Dexheimer, J.W., 2021. Early identification of epilepsy surgery candidates: A multicenter, machine learning study. Acta Neurol. Scand. 1–10. https://doi.org/10.1111/ane.13418.

Yao, L., Cai, M., Chen, Y., Shen, C., Shi, L., Guo, Y., 2019. Prediction of antiepileptic drug treatment outcomes of patients with newly diagnosed epilepsy by machine learning. Epilepsy Behav. 96, 92–97. https://doi.org/10.1016/j.yebeh.2019.04.006.