



Machine Learning in Preoperative Prediction of Postoperative Immediate Remission of Histology-Positive Cushing's Disease

Wentai Zhang^{1†}, Mengke Sun^{2†}, Yanghua Fan^{1†}, He Wang¹, Ming Feng^{1*}, Shaohua Zhou^{2*} and Renzhi Wang^{1*}

OPEN ACCESS

Edited by:

Nicholas A. Tritos,
Massachusetts General Hospital and
Harvard Medical School, United States

Reviewed by:

Brandon Patrick Galm,
Vancouver General Hospital, Canada
Krystallenia I. Alexandraki,
National and Kapodistrian University of
Athens, Greece

*Correspondence:

Ming Feng
pumchfengming@163.com
Shaohua Zhou
zhoushaohua@ict.ac.cn
Renzhi Wang
wangrz@126.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Pituitary Endocrinology,
a section of the journal
Frontiers in Endocrinology

Received: 30 November 2020

Accepted: 25 January 2021

Published: 02 March 2021

Citation:

Zhang W, Sun M, Fan Y, Wang H,
Feng M, Zhou S and Wang R (2021)
Machine Learning in Preoperative
Prediction of Postoperative Immediate
Remission of Histology-Positive
Cushing's Disease.
Front. Endocrinol. 12:635795.
doi: 10.3389/fendo.2021.635795

¹ Department of Neurosurgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, ² Medical Imaging, Robotics, Analytic Computing Laboratory/Engineering (MIRACLE), Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, China, ³ University of Chinese Academy of Sciences, Beijing, China

Background: There are no established accurate models that use machine learning (ML) methods to preoperatively predict immediate remission after transsphenoidal surgery (TSS) in patients diagnosed with histology-positive Cushing's disease (CD).

Purpose: Our current study aims to devise and assess an ML-based model to preoperatively predict immediate remission after TSS in patients with CD.

Methods: A total of 1,045 participants with CD who received TSS at Peking Union Medical College Hospital in a 20-year period (between February 2000 and September 2019) were enrolled in the present study. In total nine ML classifiers were applied to construct models for the preoperative prediction of immediate remission with preoperative factors. The area under the receiver operating characteristic (ROC) curve (AUC) was used to evaluate the performance of the models. The performance of each ML-based model was evaluated in terms of AUC.

Results: The overall immediate remission rate was 73.3% (766/1045). First operation ($p < 0.001$), cavernous sinus invasion on preoperative MRI ($p < 0.001$), tumour size ($p < 0.001$), preoperative ACTH ($p = 0.008$), and disease duration ($p = 0.010$) were significantly related to immediate remission on logistic univariate analysis. The AUCs of the models ranged between 0.664 and 0.743. The highest AUC, i.e., the best performance, was 0.743, which was achieved by stacking ensemble method with four factors: first operation, cavernous sinus invasion on preoperative MRI, tumour size and preoperative ACTH.

Conclusion: We developed a readily available ML-based model for the preoperative prediction of immediate remission in patients with CD.

Keywords: Cushing's disease, machine learning, transsphenoidal surgery, preoperative prediction, immediate remission

INTRODUCTION

Pituitary adrenocorticotrophic hormone (ACTH) hypersecretion, also known as Cushing's disease (CD), is one of the aetiologies of Cushing syndrome (CS), causing a variety of manifestations such as fatigue, weight gain, osteoporosis, diabetes mellitus, thin skin, and ecchymoses (1). According to a consensus statement, the first-line treatment option for CD is transsphenoidal surgery (TSS) (2). In a systemic review including 2,614 CD patients who had undergone TSS, the overall remission rate was 77% (52.1%–96.6%) (3), and recurrence rates were in the range of 3%–66% (4, 5).

While several studies have revealed that postoperative hypocortisolaemia (6–8), low urinary free cortisol level (9) and low ACTH levels (10) are consistent with an increased chance of long-term remission for patients with CD, there are still no accurate models for the preoperative prediction of immediate remission.

ML (machine learning) is a discipline based on computer science and statistics, has the advantage of recognizing relationships between data by learning from datasets iteratively. ML algorithms can be used to predict the outcomes of treatment procedures based on clinical data (11).

In recent years, there have been several studies using ML methods to predict the outcomes of specific diseases. Studies using ML algorithms to predict the recurrence of CD, remission after TSS in patients with acromegaly, response after radiotherapy in patients with acromegaly, and acromegaly patients' response to somatostatin analogues indicate that the advanced processing power of ML is fit for clinical application (12–14).

Immediate remission is a potential predictive factor for long-term remission and it is important for doctor-patient communication. It is also of great significance to the choice of treatment pathways. The objective of this study was to devise a reliable ML-based model to preoperatively predict immediate remission after TSS in patients with CD.

MATERIALS AND METHODS

Study Population

The current retrospective study was authorized by the ethical review committee of Peking Union Medical College Hospital (PUMCH). All patients hospitalized in PUMCH with a diagnosis of CD between February 2000 and September 2019 were included in the study.

Abbreviations: ML, machine learning; TSS, transsphenoidal surgery; CD, Cushing's disease; ROC, receiver operating characteristic curve; AUC, area under the curve; UFC, urine free cortisol; LDDST, low-dose dexamethasone suppression test; HDDST, high-dose dexamethasone suppression test; MRI, magnetic resonance imaging; BIPSS, bilateral inferior petrosal sinus sampling; PMS-C, preoperative morning (8 a.m.) serum cortisol; P24h-UFC, preoperative 24-h urinary free cortisol; LR, logistic regression; GBDT, Gradient Boosting Decision Tree; XGBoost, Extreme Gradient Boost; AdaBoost, Adaptive Boosting; RF, random forest; NB, naïve bayes; DT, decision tree; MLP, multi-layer perception; BMI, body mass index; IOMRI, invasion of cavernous sinus on preoperative MRI.

The inclusion criteria were as follows: 1) clinical manifestations of hypercortisolaemia (1); 2) pituitary tumour on magnetic resonance imaging (MRI); 3) exclusion of the possibility of ectopic ACTH syndrome; 4) meeting the criteria for endocrine diagnosis; and 5) surgical and pathological confirmation of tumour. The exclusion criteria were as follows: 1) abortion of surgery; 2) >1 missing value; 3) lack of pathological proof of tumour; 4) lack of record of postoperative prognosis within 7 days (morning (8 am) serum cortisol concentration lower than 5 µg/dl or 24h-UFC lower than 20 µg/d); and 5) incomplete medical records.

In total, 1,045 participants were enrolled in the study and 341 participants were excluded with the exclusion criteria (including 150 pathology-negative patients). All participants were randomly divided into two datasets, a training dataset and a test dataset in a ratio of 8:2.

Diagnosis of CD

The participants regularly underwent T1-weighted, T2-weighted, T1-weighted gadolinium-enhanced, or dynamic gadolinium-enhanced T1-weighted MRI. In most cases, T1-weighted gadolinium-enhanced MRI was sufficient to demonstrate a suspicion of CD. A pituitary adenoma was suspected if there was a relatively isolated hypointense region within the pituitary gland. In some cases where the profile of the suspected existing tumour was inconspicuous, a dynamic gadolinium-enhanced T1-weighted MRI was needed to outline the tumour and the existence of a tumour would be suspected if there was a relatively hypointense region trailing off over time.

All participants underwent combined low-dose and high-dose dexamethasone suppression tests (LDDST and HDDST) to detect hypercortisolism and identify the location of the tumour, respectively. In the LDDST, a dose of 0.5 mg dexamethasone was given to participants every 6 h for 2 consecutive days, and 24-h urinary free cortisol (24h-UFC) was considered suppressed if it fell below 12.3 µg/24 h on the second day or the plasma cortisol level was lower than 1.8 µg/dl in the morning of the third day. In the HDDST, 2 mg dexamethasone was given to participants every 6 h for 2 consecutive days; 24h-UFC was measured on the second day of drug administration, and plasma cortisol was measured on the morning of the third day. Cortisol was considered to be suppressed if it was reduced by more than 50% compared to the original level. The combination of failure of suppression on the LDDST and successful suppression on HDDST is an indicator of Cushing disease.

The position of the ACTH-secreting adenoma was also confirmed by bilateral petrosal sinus sampling (BIPSS) using desmopressin (instead of CRH) stimulation test in cases where there was uncertainty about the tumour location on preoperative MRI. A 10-mg dose of desmopressin was administered to participants to stimulate ACTH. If the ratio of ACTH concentration in the inferior petrosal sinus to that in peripheral vein (the elbow vein) was greater than 2 in the basal state or greater than 3 after desmopressin stimulation, a diagnosis of CD could be made. The reason that we did not use CRH test in our practice is that CRH is not accessible in China.

The diagnosis of CD was made on the basis of synthesized evidence, including but not limited to MRI scans, combined LDDST and HDDST, BIPSS using desmopressin stimulation test, as well as clinical manifestations, signs, and chief complaints.

All participants underwent TSS by the same two experienced pituitary neurosurgeons (RW and MF). The details of TSS are discussed in our previous study (5). No glucocorticoids or adjuvant medical therapy were given to participants before or during the operation.

Postoperative Management and Immediate Remission

Plasma cortisol was routinely monitored in the first 3 days after TSS and if the concentration was lower than 5 µg/dl, we initiated glucocorticoid replacement therapy consisting of intravenous hydrocortisone 100 mg two times per day for 3 consecutive days, followed by 30 mg once per day. After being discharged from the hospital, participants decreased their hydrocortisone dosage by 2.5 mg per week, till a dosage of 2.5–5 mg per day. The cessation of glucocorticoid intake was suggested by an endocrinologist or a neurosurgeon after the evaluation of pituitary function.

Immediate remission was defined as morning (8 a.m.) serum cortisol concentration lower than 5 µg/dl or 24-h UFC lower than 20 µg/d (15). Serum cortisol, ACTH and 24-h UFC were all measured with immunoassay method.

Study Design and ML Algorithms

In ML, data determine the final result and algorithms can only approximate it. Therefore, data preprocessing before model training is particularly important. In the experiment, we performed feature selection and standardized preprocessing on the data.

The selection of features is a process of selecting relevant characteristics and ranking their importance. In our study, we used a univariate feature selection method based on F-test. The F-test, also known as ANOVA, assumes homogeneity of variance and serves as a filtering method to capture the linear relationship between each feature and the label. The highest 1-11 ranked characteristics were sequentially introduced into each ML algorithm. The AUCs of models with different numbers of variables were calculated.

Data standardization can eliminate errors caused by different dimensions or large differences in values. To maintain high prediction performance, we applied min–max normalization to all the data (16). Z-score and non-normalization were also applied, but it turns out that min-max is suitable for normalization to get the model with the highest AUC.

Algorithms

Five ensemble learning algorithms were applied, namely, the gradient boosting decision tree (GBDT) (17), random forest (RF) (18), adaptive boosting (AdaBoost) (19), extreme gradient boost (XGBoost), and stacking algorithms (20). In addition, 4 non-ensemble learning algorithms, i.e., the logistic regression (LR), naïve bayes (NB) (21), decision tree (DT) (22), and multi-layer perceptron (MLP) algorithms (23) were also used in our

experiment. Among these algorithms, LR, NB, and DT are interpretable, but GBDT, RF, AdaBoost, XGBoost, MLP, and stacking are not interpretable, meaning that the functions connecting characteristics and labels are invisible to users. Stacking, which showed the best performance, usually incorporates multiple heterogeneous weak learners, in parallel, combining them by training a “meta-model”, whose inputs are the prediction results of these different weak models. On that basis, the model then outputs a final result.

We applied a total of nine algorithms to predict immediate remission using all 11 variables. First, we selected features according to their rankings from 1 to 11 based on the F-test in the training dataset. Then, we chose the same features in the test dataset. For each iteration, 10-fold cross-validation was applied for training and validation in the training dataset. Later, the grid-search approach (24) was used to select the optimal hyperparameters in each model. Finally, the model that performed best on the validation set was evaluated on the test dataset with the values of area under the receiver operating characteristic curve (AUC), sensitivity (true positive rate), specificity (true negative rate) and Youden's index. All models were applied and evaluated with the programming language Python version 3.6 (Python Software Foundation of Delaware, USA) and scikit-learn (an ML tool based on the Python language).

Statistical Analysis

We performed all statistical analyses in Python, version 3.6 (Python Software Foundation, Delaware, United States), RStudio software (1.2.5042) and IBM SPSS Statistics 23 (IBM Corporation). Matplotlib 3.1.3 (comprehensive library for creating static, animated, and interactive visualizations in Python) was used to draw statistical figures. The proportions of missing values for disease duration, body mass index (BMI), tumour size, tumour invasion of the cavernous sinus and combined LDDST and HDDST were 3.9%, 2.8%, 1.7%, 2.6%, and 2.9%, respectively. There were no missing values for any other characteristics. The missing values were imputed with the K-nearest neighbour (KNN) algorithm. The normality of continuous variables was evaluated with the Shapiro-Wilk test. Continuous variables that were normally distributed are displayed as mean ± standard deviation. Continuous variables that were non-normally distributed are displayed as the interquartile range. The Wilcoxon test was used to compare non-normal continuous variables. Categorical variables were displayed as frequencies or percentages and were analysed by the chi-squared test or Fisher's exact test.

RESULTS

Participant Characteristics

In total 1,045 participants treated for CD from February 2000 to September 2019 were enrolled in this study.

The values of the 11 predictors in all participants (836 from the training dataset and 209 from the test dataset) are shown in **Table 1**. A total of 766 participants (73.3%) exhibited immediate

TABLE 1 | Participants' characteristics in training and test datasets.

Characteristic	Total	Training dataset	Test dataset	P-value
N	1045	836	209	
Gender				
Male	201(19.23%)	168(20.10%)	33(15.79%)	0.158
Female	844(80.77%)	668(79.90)	176(84.21%)	
Age(year)	35(26–45)	35(26–45)	34(26–44)	0.858
First operation				
Yes	925(88.52%)	740(88.52%)	185(88.52%)	1.000
No	120(11.48%)	96(11.48)	24(11.48%)	
Disease duration(month)	46(24–72)	46(21–73)	46(24–72)	0.883
BMI(kg/m²)	25.96(23.76–28.60)	25.96(23.75–28.65)	26.10(23.82–28.39)	0.886
P24h-UFC(μg/d)	467.52(279–791.7)	467.61(278.95–790.85)	466.9(283.6–812.8)	0.845
Combined LDDST and HDDST				
Positive	856(81.91%)	679(81.22%)	177(84.69%)	0.244
Negative	189(18.09%)	157(18.78%)	32(15.31%)	
preoperative ACTH(pg/ml)	71.26(48.2–110)	71.10(47.88–111.00)	72.60(50.90–109.00)	0.748
PMS-C(μg/dl)	26.38(21.46–32.12)	26.31(21.34–31.96)	26.61(21.86–33.40)	0.589
Tumour size				
microadenoma	845(80.86%)	679(81.22%)	166(79.43%)	0.555
macroadenoma	200(19.14%)	157(18.78%)	43(20.57%)	
IOMRI				
Positive	66(6.32%)	50(5.98%)	16(7.66%)	0.373
Negative	979(93.68%)	786(94.02%)	193(92.34%)	
Prognosis				0.828
remission	766(73.30%)	613(73.33%)	153(73.21%)	
non-remission	279(26.70%)	223(26.67%)	56(26.79%)	

BMI, body mass index; P24h-UFC, preoperative 24-h urinary free cortisol; PMS-C, preoperative morning (8 a.m.) serum cortisol; IOMRI, invasion of cavernous sinus on preoperative MRI.

remission postoperatively. No obvious interclass differences were found in gender, age, first operation, disease duration, BMI, 24-h preoperative UFC (P24h-UFC), combined LDDST and HDDST (combined LH), preoperative morning serum cortisol (PMS-C), preoperative ACTH, tumour size (microadenoma or macroadenoma) or cavernous sinus invasion evaluated on MRI (IOMRI). Microadenoma was defined as the largest diameter of the tumour <10 mm, and macroadenoma was defined as the largest diameter ≥10 mm. IOMRI was defined as Knosp 3 or Knosp 4 adenoma on MRI. The characteristics of patients in remission and non-remission groups are shown in **Table 2**.

Predictive Performance of the Nine Models

To clarify the relationship between predictive performance and the number of characteristics selected in the model, we introduced all features in order of importance to 9 algorithms. We found that, as the number of selected characteristics increased, the prediction performance of the model initially improved, but once the optimal number of features was reached, the performance declined if the number of selected characteristics was further increased. All models except DT performed best when the number of features is 4, achieving a much greater AUC than it when all features were used (**Figure 1**). **Figure 2** shows the performance of each model at the optimal number of features. The best predictive performance was achieved by the stacking algorithm (AUC = 0.743, 95% confidence interval (0.677, 0.806), ACC=0.746) which outperformed GBDT (AUC = 0.734), RF (AUC = 0.726), LR (AUC = 0.701), XGB (AUC = 0.712), MLP (AUC = 0.700), and AdaBoost (AUC = 0.699) and significantly outperformed NB

TABLE 2 | Patients' characteristics in remission and non-remission groups.

Characteristic	remission	nonremission	p-value
Gender	766	279	
Male	143(18.67%)	58(20.79%)	0.442
Female	623(81.33%)	221(79.21%)	
Age(year)	35(27–45)	33(25–43)	0.170
First operation			
Yes	700(91.38%)	225(80.65%)	<0.001
No	66(8.62%)	54(19.35%)	
Disease duration (month)	39(22–72)	48(24–84)	0.010
BMI(kg/m²)	25.90(23.65–28.45)	26.43(24.02–29.09)	0.146
P24h-UFC(μg/d)	459.1(287.6–793.3)	482.4(258.0–796.5)	0.830
Combined LDDST and HDDST			
Positive	634(82.77%)	222(79.57%)	0.235
Negative	132(17.23%)	57(20.43%)	
PMS-Cortisol(μg/dl)	26.44(21.34–32.19)	26.07(21.86–31.97)	0.840
preoperative ACTH(pg/ml)	68.3(47.1–108.0)	80.0(51.0–119.0)	0.008
Tumour size			
Micro	646(84.33%)	199(71.33%)	<0.001
Macro	120(15.67%)	80(28.67%)	
IOMRI			
Yes	30(3.92%)	36(12.90%)	<0.001
No	736(96.08%)	243(87.10%)	

BMI, body mass index; P24h-UFC, preoperative 24-h urinary free cortisol; combined low-dose and high-dose dexamethasone suppression test; PMS-C, preoperative morning (8 a.m.) serum cortisol; IOMRI, invasion of cavernous sinus on preoperative MRI.

(AUC = 0.681) and DT (AUC = 0.664). The outcomes of different models are shown in **Table 3**.

The highest Youden's index for stacking was 0.393, and the sensitivity and specificity were 80.4% and 58.9%, respectively.

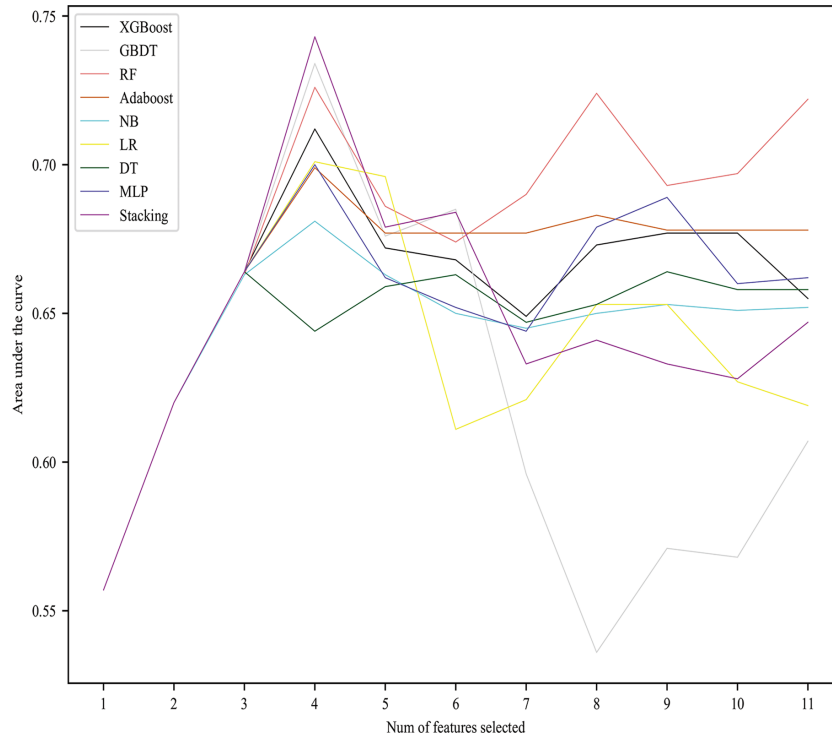


FIGURE 1 | AUC values of nine models with different numbers of features selected.

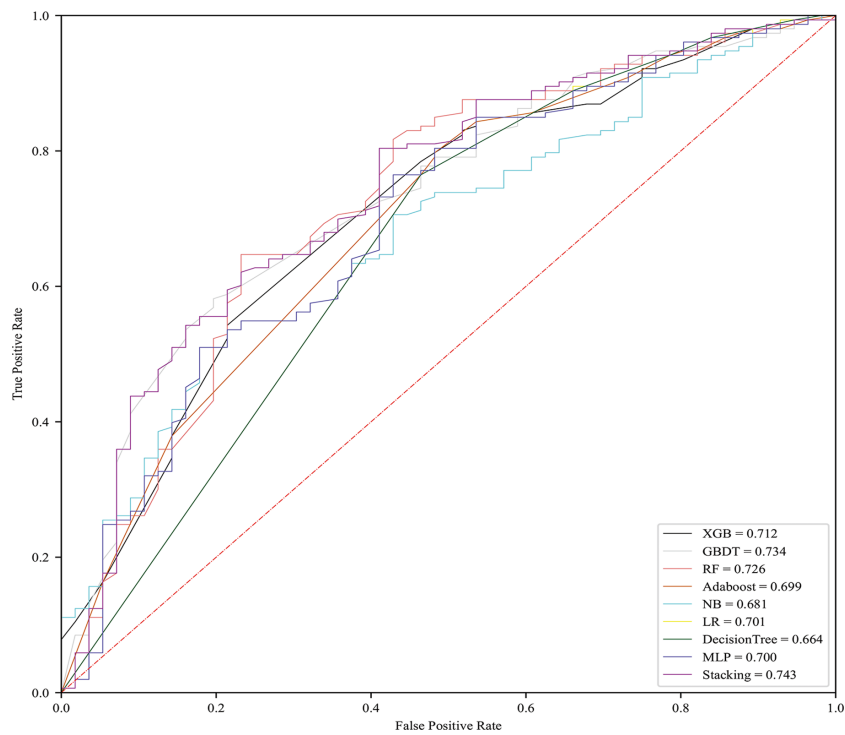


FIGURE 2 | Performances of models with optimal number of features.

TABLE 3 | Highest AUCs of different models.

Algorithm	AUC	95% confidence interval	Accuracy
Extreme Gradient Boost	0.712	0.645–0.776	0.732
Gradient Boosting Decision Tree	0.734	0.669–0.796	0.732
Random Forest	0.726	0.656–0.792	0.746
Adaptive Boost	0.699	0.632–0.764	0.737
Naïve Bayes	0.681	0.614–0.744	0.684
Logistic regression	0.701	0.630–0.767	0.732
Decision Tree	0.664	0.598–0.726	0.742
Multi-layer Perceptron	0.700	0.629–0.766	0.732
Stacking	0.743	0.677–0.806	0.746

AUC, area under the curve; Accuracy, the percentage of participants whose postoperative immediate prognoses were correctly predicted with the ML-based model among the cohort.

Variable Importance

We selected the best features through a statistical test based on F-test univariate analysis, and the features were ranked for importance as follows, IOMRI, tumour size, first operation, preoperative ACTH, disease duration, combined LH, BMI, age, P24hUFC, gender, PMS-C. Six models also output their own variable importance rankings, and the rankings are shown in **Figure 3**. The top 3 most important predictors were IOMRI, first operation and tumour size. For the top 3 most important predictors, we also performed chi-squared analysis to evaluate the remission rates in different groups (**Figure 4**).

DISCUSSION

TSS is the first-line treatment for CD by neurosurgeons. According to a previous study by Ioachimescu AG et al., the immediate remission rate ranges between 59% and 96.6% (25). In the present cohort, the overall immediate remission rate was

73.3%, which approximates the 76% remission rate reported by a meta-analysis (26). According to a recent study, immediate remission may be a compelling predictor of long-term remission (6). Therefore, it is important to develop a preoperative prediction model for immediate remission.

To our knowledge, this is the largest cohort and first ML-based model for the preoperative prediction of immediate remission after TSS. Conventional statistical approaches have usually been the first choice in previous studies. In a previous study by our team (27), conventional biostatistical methods were used to detect the preoperative predictors for immediate remission in patients with Cushing's disease after TSS. The AUC of receiver operating characteristic curve for single preoperative ACTH predicting immediate remission is 0.553 (cutoff value=67.35 ng/L, sensitivity 60.9%, specificity 49.5%). Disease duration ≤64.5 months predicted immediate remission with an AUC of 0.552 (40.5% sensitivity and 71.0% specificity). The AUC value(0.743) of the present ML-based model is conspicuously higher than the ones(0.553 and 0.552) in our previous study which indicates that the ML-based model can more accurately predict immediate remission.

The ML-based model is completely different from conventional methods in that it can learn the patterns of health trajectories from electronic health datasets themselves. ML models may help clinicians to gain knowledge from a vast quantity of data far beyond any individual clinician's experience (28). The result from ML methods has its own scalability and flexibility advantages over traditional statistics. In our analysis, therefore, an ML-based method including nine algorithms was applied, leading us to the conclusion that a model constructed with IOMRI, tumour size, first operation and preoperative ACTH had the best performance (AUC=0.743).

We developed an online calculator to facilitate the prediction which can output the probability of immediate remission with

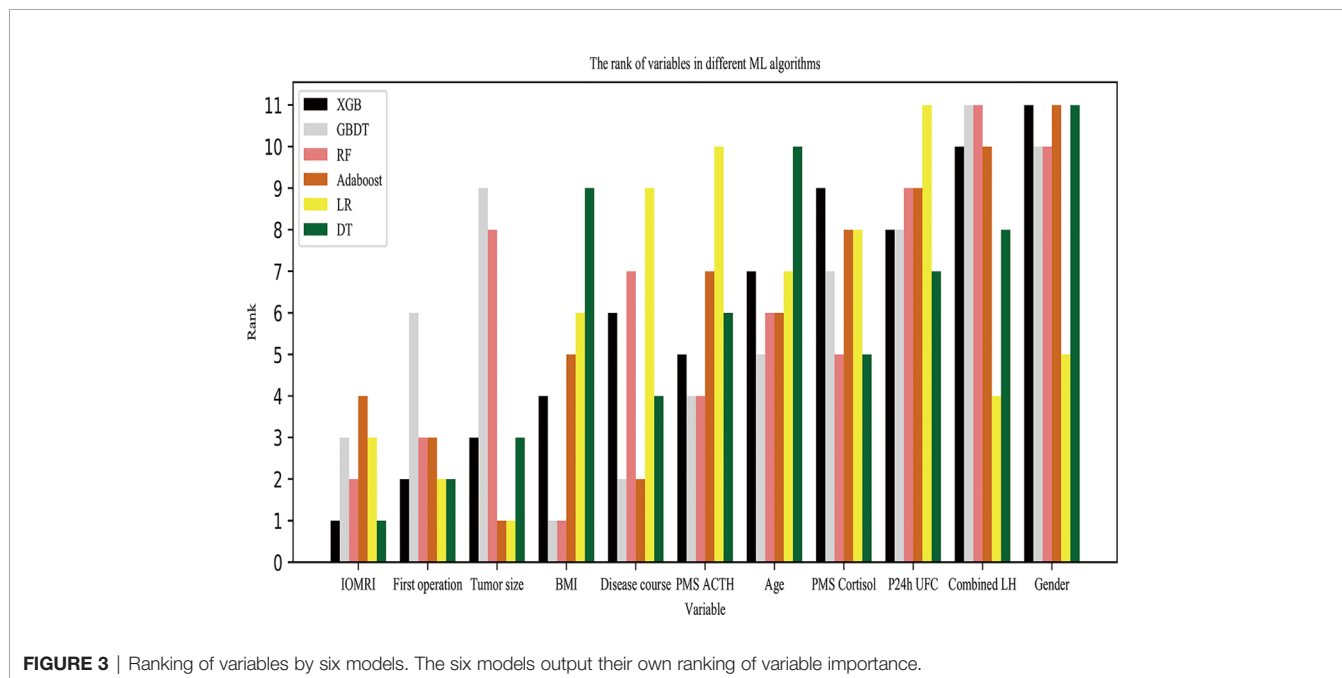


FIGURE 3 | Ranking of variables by six models. The six models output their own ranking of variable importance.

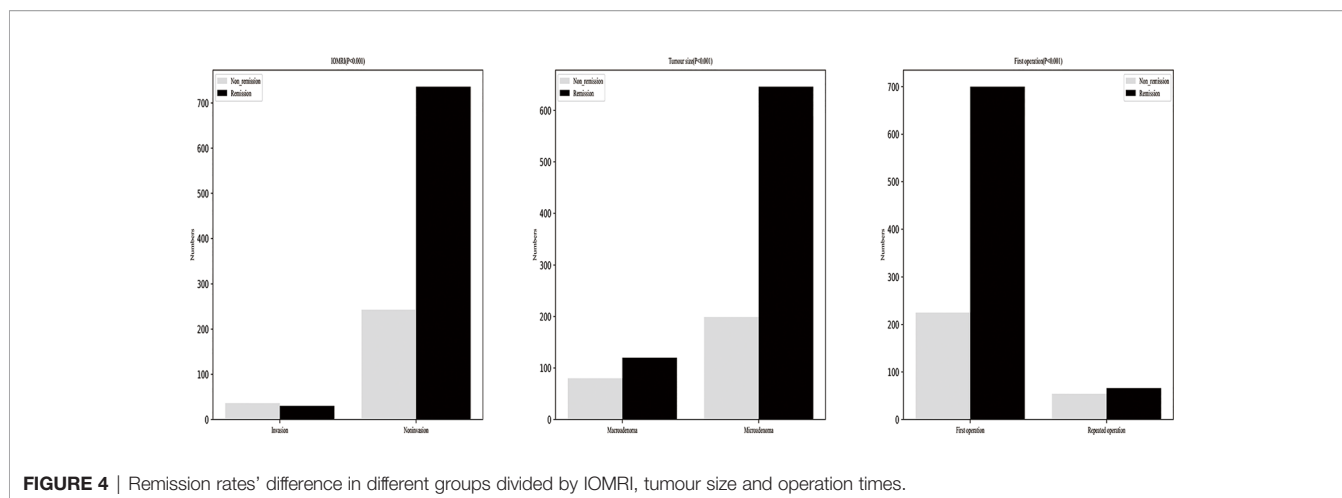


FIGURE 4 | Remission rates' difference in different groups divided by IOMRI, tumour size and operation times.

input of preoperative factors (<http://smk921101.pythonanywhere.com/index>). With the output from the calculator, surgeons can talk to patients about the possible prognosis. Thus, the ML-based model is important in providing information about prognosis and doctor–patient communication.

There are advantages as well as disadvantages of ML algorithms. Simple statistical method such as single-factor logistic regression is a kind of simple algorithm of ML as well as a traditional statistical method. It is a linear classifier, so it cannot handle the correlation between features. At the same time, the performance may be poor (AUC) and the accuracy may be low. In contrast, we used a variety of ensemble learning methods in our experiment. The idea is to combine multiple weak learners into one strong learner. The advantage of ensemble learning is to ensure the diversity of weak classifiers, so that the results obtained are often better than those of a single learner. There are also some disadvantages for ML-based models. On the one hand, ML requires enough data to train the model, otherwise it will easily lead to overfitting. Our study presented the largest cohort which is beneficial for the analysis of ML method. On the other hand, some ML methods have poor interpretability, thus it needs an independent test dataset. We divided the cohort into a training dataset and an independent test dataset, but it would be better to include data from another centre as a test dataset in the future.

In the present study, the model with the best performance showed that IOMRI was ranked as the top characteristic related to immediate remission which is consistent with previous studies and our experience (29). Tumour size is ranked as the second most important characteristic. Patients with microadenomas tended to have a higher immediate remission rate than patients with macroadenomas, which is consistent with a previous study (26). Similarly, other studies also found that patients with macroadenomas tended to have lower immediate remission rates than patients with microadenomas after TSS (30, 31). Patients who have a history of TSS is ranked as the third most important characteristic. Patients with a history of TSS tended to have an increased probability of non-remission in our study, which is in

accordance with previous studies (32, 33). This may be because a second TSS often indicates a more invasive and aggressive behaviour of the tumour in which there may be a residual tumour after TSS. Preoperative ACTH is ranked as the fourth most important predictor. The four predictors mentioned above were included in the stacking algorithm with highest AUC.

Disease duration was not included in the final ML-based model, but it was ranked as the fifth most important predictor and had obvious significance in univariate analysis. The results of a study by Elena Y et al. showed that the disease duration is not correlated with the recurrence of CD (34). Thus, the actual relationship between disease duration and immediate remission remains to be investigated in a larger or multicentre cohort. In a previous study, Jessica K et al. reported that older age was a strong predictor of mortality following TSS (35), but in other studies, age was not correlated with recurrence (6, 36). In our study, age was not included in the final model. And the model also excluded the combined LDDST and HDDST results and BMI which were not significantly correlated with immediate remission according to univariate analysis. There have been several noteworthy contradictions among different studies investigating predictors of remission and these differences may be attributable to the small cohorts and difficulty in unveiling the complex relationships of the vast number of variables with conventional biostatistical methods. This is exactly where the advantages of ML lie. We believe that an ML-based model may be more effective than previous methods as a way to understand the disease in the future.

LIMITATIONS AND STRENGTHS

The present study has some limitations. First, the clinical data still had some missing values. Second, a larger cohort would be more suitable for ML analysis, although, to our knowledge this study already has the largest CD cohort in the literature. Third, this is a single centre study. However, our cohort is larger than most cohorts in previous studies and immediate remission was

evaluated accurately. One strength of this study is that it is the first to carry out preoperative prediction of immediate remission using ML methods, whose performance is better than that of traditional methods, such as LR and NB.

CONCLUSION

TSS is the preferred choice of neurosurgeons to treat CD, and ML-based models can be used to preoperatively predict immediate remission. In this study, stacking algorithm outperformed the traditional ones, such as LR and NB. Invasion of cavernous sinus on MRI, the largest diameter of tumour, first operation, and preoperative ACTH were included in the final model.

DATA AVAILABILITY STATEMENT

Some or all datasets generated and/or analysed during the current study are not publicly available but are available from the corresponding author on reasonable request. If the datasets are needed, the corresponding author RW would be in contact.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by ethical review committee of Peking Union Medical

College Hospital (PUMCH). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

WZ, MS, and YF contributed equally to the present study. Each author contributes to the article in data collecting and analysis. RW, SZ, and MF take final responsibility for this article. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Graduate Innovation Fund of Peking Union Medical College (2018-1002-01-10) and by the, Natural Science Foundation of Beijing Municipality (grant number 7182137).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fendo.2021.635795/full#supplementary-material>

REFERENCES

1. Arnaldi G, Angeli A, Atkinson AB, Bertagna X, Cavagnini F, Chrousos GP, et al. Diagnosis and complications of Cushing's syndrome: a consensus statement. *J Clin Endocrinol Metab* (2003) 88(12):5593–602. doi: 10.1210/jc.2003-030871
2. Biller BM, Grossman AB, Stewart PM, Melmed S, Bertagna X, Bertherat J, et al. Treatment of adrenocorticotropic-dependent Cushing's syndrome: a consensus statement. *J Clin Endocrinol Metab* (2008) 93(7):2454–62. doi: 10.1210/jc.2007-2734
3. Petersenn S, Beckers A, Ferone D, van der Lely A, Bollerslev J, Boscaro M, et al. Therapy of endocrine disease: outcomes in patients with Cushing's disease undergoing transsphenoidal surgery: systematic review assessing criteria used to define remission and recurrence. *Eur J Endocrinol* (2015) 172(6):R227–39. doi: 10.1530/EJE-14-0883
4. Aranda G, Ensenat J, Mora M, Puig-Domingo M, Martinez de Osaba MJ, Casals G, et al. Long-term remission and recurrence rate in a cohort of Cushing's disease: the need for long-term follow-up. *Pituitary* (2015) 18(1):142–9. doi: 10.1007/s11102-014-0567-8
5. Feng M, Liu Z, Liu X, Bao X, Yao Y, Deng K, et al. Diagnosis and Outcomes of 341 Patients with Cushing's Disease Following Transsphenoid Surgery: A Single-Center Experience. *World Neurosurg* (2018) 109:e75–80. doi: 10.1016/j.wneu.2017.09.105
6. Ironside N, Chatain G, Asuzu D, Benzo S, Lodish M, Sharma S, et al. Earlier post-operative hypocortisolemia may predict durable remission from Cushing's disease. *Eur J Endocrinol* (2018) 178(3):255–63. doi: 10.1530/EJE-17-0873
7. Mayberg M, Reintjes S, Patel A, Moloney K, Mercado J, Carlson A, et al. Dynamics of postoperative serum cortisol after transsphenoidal surgery for Cushing's disease: implications for immediate reoperation and remission. *J Neurosurg* (2018) 129(5):1268–77. doi: 10.3171/2017.6.JNS17635
8. Costenaro F, Rodrigues TC, Rollin GA, Ferreira NP, Czepielewski MA. Evaluation of Cushing's disease remission after transsphenoidal surgery based on early serum cortisol dynamics. *Clin Endocrinol (Oxf)* (2014) 80(3):411–8. doi: 10.1111/cen.12300
9. Pendharkar AV, Sussman ES, Ho AL, Hayden Gephart MG, Katznelson L. Cushing's disease: predicting long-term remission after surgical treatment. *Neurosurg Focus* (2015) 38(2):E13. doi: 10.3171/2014.10.FOCUS14682
10. Abdelmannan D, Chaiban J, Selman WR, Arafah BM. Recurrences of ACTH-secreting adenomas after pituitary adenectomy can be accurately predicted by perioperative measurements of plasma ACTH levels. *J Clin Endocrinol Metab* (2013) 98(4):1458–65. doi: 10.1210/jc.2012-3910
11. Deo RC. Machine Learning in Medicine. *Circulation* (2015) 132(20):1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593
12. Liu Y, Liu X, Hong X, Liu P, Bao X, Yao Y, et al. Prediction of Recurrence after Transsphenoidal Surgery for Cushing's Disease: The Use of Machine Learning Algorithms. *Neuroendocrinology* (2019) 108(3):201–10. doi: 10.1159/000496753
13. Fan Y, Li Y, Li Y, Feng S, Bao X, Feng M, et al. Development and assessment of machine learning algorithms for predicting remission after transsphenoidal surgery among patients with acromegaly. *Endocrine* (2020) 67(2):412–22. doi: 10.1007/s12020-019-02121-6
14. Kocak B, Durmaz ES, Kadioglu P, Polat Korkmaz O, Comunoglu N, Tanriover N, et al. Predicting response to somatostatin analogues in acromegaly: machine learning-based high-dimensional quantitative texture analysis on T2-weighted MRI. *Eur Radiol* (2019) 29(6):2731–9. doi: 10.1007/s00330-018-5876-2
15. Nieman LK, Biller BM, Findling JW, Murad MH, Newell-Price J, Savage MO, et al. Treatment of Cushing's Syndrome: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* (2015) 100(8):2807–31. doi: 10.1210/jc.2015-1818
16. Markov I, Arampatzis A, Crestani F. *Unsupervised linear score normalization revisited*. (2012) 1161. doi: 10.1145/2348283.2348519

17. Friedman JH. Stochastic gradient boosting. *Comput Stats Data Analysis* (2002) 38(4):367–78. doi: 10.1016/S0167-9473(01)00065-2
18. Surhone LM, Tennoe MT, Henssonow SF, Breiman L. Random Forest. *Mach Learn* (2010) 45(1):5–32. doi: 10.1023/A:1010933404324
19. Schapire RE, Singer Y. BoosTexter: A Boosting-based System for Text Categorization. *Mach Learn* (2000) 39(2):135–68. doi: 10.1023/A:1007649029923
20. Wolpert DH. Stacked generalization. *Neural Netw* (2017) 5(2):241–59. doi: 10.1007/978-1-4899-7687-1_969
21. Zhang H. *The Optimality of Naive Bayes*. (2004).
22. Quinlan JR. Induction on decision tree. *Mach Learn* (1986) 1:81–106. doi: 10.1007/BF00116251
23. Gardner MW, Dorling SR. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment* (1998) 32. doi: 10.1016/S1352-2310(97)00447-0
24. Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, et al. API design for machine learning software: experiences from the scikit-learn project. *Eprint Arxiv* (2013).
25. Ioachimescu AG. Prognostic Factors of Long-Term Remission After Surgical Treatment of Cushing's Disease. *Endocrinol Metab Clin North Am* (2018) 47(2):335–47. doi: 10.1016/j.ecl.2018.02.002
26. Abu Dabrh AM, Singh Ospina NM, Al Nofal A, Farah WH, Barrionuevo P, Sarigianni M, et al. Predictors of Biochemical Remission and Recurrence after Surgical and Radiation Treatments of Cushing Disease: A Systematic Review and Meta-Analysis. *Endocr Pract* (2016) 22(4):466–75. doi: 10.4158/EP15922.RA
27. Dai C, Fan Y, Liu X, Bao X, Yao Y, Wang R, et al. Predictors of Immediate Remission after Surgery in Cushing's Disease Patients: a large retrospective study from a single center. *Neuroendocrinology* (2020). doi: 10.1159/000509221
28. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med* (2019) 380(14):1347–58. doi: 10.1056/NEJMra1814259
29. Pivonello R, De Leo M, Cozzolino A, Colao A. The Treatment of Cushing's Disease. *Endocr Rev* (2015) 36(4):385–486. doi: 10.1210/er.2013-1048
30. Blevins LSJr., Christy JH, Khajavi M, Tindall GT. Outcomes of therapy for Cushing's disease due to adrenocorticotropin-secreting pituitary macroadenomas. *J Clin Endocrinol Metab* (1998) 83(1):63–7. doi: 10.1210/jcem.83.1.4525
31. Chandler WF, Barkan AL, Hollon T, Sakharova A, Sack J, Brahma B, et al. Outcome of Transsphenoidal Surgery for Cushing Disease: A Single-Center Experience Over 32 Years. *Neurosurgery* (2016) 78(2):216–23. doi: 10.1227/NEU.0000000000001011
32. Starke RM, Reames DL, Chen CJ, Laws ER, Jane JAJr. Endoscopic transsphenoidal surgery for cushing disease: techniques, outcomes, and predictors of remission. *Neurosurgery* (2013) 72(2):240–7; discussion 247. doi: 10.1227/NEU.0b013e31827b966a
33. Valderrabano P, Aller J, Garcia-Valdecasas L, Garcia-Uria J, Martin L, Palacios N, et al. Results of repeated transsphenoidal surgery in Cushing's disease. Long-term follow-up. *Endocrinol Nutr* (2014) 61(4):176–83. doi: 10.1016/j.endonu.2013.10.008
34. Nadezhdina EY, Rebrova OY, Grigoriev AY, Ivaschenko OV, Azizyan VN, Melnichenko GA, et al. Prediction of recurrence and remission within 3 years in patients with Cushing disease after successful transnasal adenomectomy. *Pituitary* (2019) 22(6):574–80. doi: 10.1007/s11102-019-00985-5
35. Lambert JK, Goldberg L, Fayngold S, Kostadinov J, Post KD, Geer EB. Predictors of mortality and long-term outcomes in treated Cushing's disease: a study of 346 patients. *J Clin Endocrinol Metab* (2013) 98(3):1022–30. doi: 10.1210/jc.2012-2893
36. Alexandraki KI, Kaltsas GA, Isidori AM, Storr HL, Afshar F, Sabin I, et al. Long-term remission and recurrence rates in Cushing's disease: predictive factors in a single-centre study. *Eur J Endocrinol* (2013) 168(4):639–48. doi: 10.1530/EJE-12-0921

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

Copyright © 2021 Zhang, Sun, Fan, Wang, Feng, Zhou and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.