scientific reports

Check for updates

OPEN

Predictive modeling of arginine vasopressin deficiency after transsphenoidal pituitary adenoma resection by using multiple machine learning algorithms

Yuyang Chen[1,3], Jiansheng Zhong[1,3], Haixiang Li[1,2], Kunzhe Lin[1], Liangfeng Wei[1] & Shousen Wang[1]✉

This study aimed to predict arginine vasopressin deficiency (AVP-D) following transsphenoidal pituitary adenoma surgery using machine learning algorithms. We reviewed 452 cases from December 2013 to December 2023, analyzing clinical and imaging data. Key predictors of AVP-D included sex, tumor height, preoperative and postoperative changes in sellar diaphragm height and pituitary stalk length, preoperative ACTH levels, changes in ACTH levels, and preoperative cortisol levels. Six machine learning algorithms were tested: logistic regression (LR), support vector classification (SVC), random forest (RF), decision tree (DT), k-nearest neighbors (KNN), and extreme gradient boosting (XGBoost). After cross-validation and parameter optimization, the random forest model demonstrated the highest performance, with an accuracy (ACC) of 0.882 and an AUC of 0.96. The decision tree model followed, achieving an accuracy of 0.843 and an AUC of 0.95. Other models showed lower performance: LR had an ACC of 0.522 and an AUC of 0.54; SVC had an ACC of 0.647 and an AUC of 0.67; KNN achieved an ACC of 0.64 and an AUC of 0.70; and XGBoost had an ACC of 0.794 and an AUC of 0.91. The study found that a shorter preoperative pituitary stalk length, significant intraoperative stretching, and lower preoperative ACTH and cortisol levels were associated with a higher likelihood of developing AVP-D post-surgery.

Keywords Machine learning, Predictive model, Diabetes insipidus, Pituitary adenoma, Transsphenoidal surgery

Pituitary adenomas are benign tumors that arise from adenohypophysis and frequently cause clinical symptoms by expanding and compressing surrounding tissues[1,2]. At present, transsphenoidal pituitary adenoma resection is one of the most commonly used surgical methods. It is known for its minimal invasiveness, mild postoperative symptoms, and rapid recovery[3,4]. However,this surgery frequently leads to mechanical damage to the hypothalamic–neurohypophyseal system, subsequently triggering the occurrence of central diabetes insipidus (cDI).In clinical practice, postoperative diabetes insipidus is typically classified as either transient or permanent. Given that permanent diabetes insipidus is relatively rare, this study will focus exclusively on postoperative transient diabetes insipidus.

Central diabetes insipidus is a common complication following pituitary adenoma resection, primarily due to damage to the posterior pituitary or pituitary stalk during surgery, leading to reduced central secretion of

[1]Department of Neurosurgery, Fuzhou General Clinical Medical College, Fujian Medical University (900th Hospital), Fuzhou 350025, China. [2]Department of Neurosurgery, East Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen, China, FuZhou, China. [3]These authors contributed equally: Yuyang Chen and Jiansheng Zhong. ✉email: wshsen1965@126.com

arginine vasopressin and resulting in increased urine output. It has now been officially renamed as arginine vasopressin deficiency syndrome (AVP-D)[5].It is characterized by symptoms, such as polydipsia and polyuria[6]. These symptoms can result in electrolyte imbalance and varying degrees of dehydration, and can be life-threatening in severe cases[7–9]. The incidence of AVP-D after pituitary adenoma surgery ranges from 9 to 22%[10]. It is typically associated with intraoperative damage to the hypothalamus and the infundibulum[6,11]. As a common postoperative complication[12], early AVP-D prediction is crucial for clinicians to implement early prevention and management strategies[13]. Therefore, researchers have explored various clinical factors, such as gender, age, serum pituitary hormone levels, and intraoperative cerebrospinal fluid leakage, to identify potential predictors of postoperative AVP-D[9,12–19]. However, these studies have yet to reach a consensus[10].

As an emerging method in clinical research, machine learning offers a reliable approach for predicting clinical outcomes compared with traditional methods[20]. Machine learning algorithms have been widely applied in the field of neurosurgery, demonstrating remarkable efficacy in diagnosing and prognosticating conditions, such as brain tumors and traumatic brain injuries[21]. Some researchers have recently proposed using machine learning to handle complex variables and develop predictive models for AVP-D after pituitary adenoma surgery[22]. Through a retrospective analysis, the current study aims to employ multiple machine learning algorithms to establish predictive models for AVP-D after pituitary adenoma surgery and then comprehensively evaluate the performance of these models. The objective is to assess the feasibility of machine learning-based clinical models in predicting postoperative AVP-D.

Study subjects

We continuously collected clinical and imaging data from patients who visited the Department of Neurosurgery at our hospital and underwent transsphenoidal pituitary adenoma resection under a microscope between December 2013 and December 2023. After applying strict inclusion and exclusion criteria, 452 cases were included in the analysis.This study was approved by the Ethics Committee of the 900th Hospital of Fuzhou, Fujian Province, China (Ethical Approval Number: Lunshenke No. 2024-006). The procedures used in this study comply with the principles of the Declaration of Helsinki. All individuals participating in this study provided general written informed consent during hospitalization, agreeing to the re-use of their generated data.

1. Inclusion criteria

The inclusion criteria were as follows: 1. complete imaging and clinical data, 2. underwent transsphenoidal pituitary tumor resection at our hospital, and 3. postoperative pathology and immunohistochemistry confirmed nonfunctional pituitary adenoma.

2. Exclusion criteria

The exclusion criteria were as follows: 1. preoperative or postoperative head radiotherapy; 2. previous surgery or medical treatment of the sellar region; 3. incomplete clinical or imaging data; 4. concurrent brain trauma, meningitis, brain abscess, or cerebrovascular disease; and 5. concurrent multiple intracranial or malignant tumors. 6. Patients who exhibited AVP-D preoperatively.

After strict screening, 452 cases were included in the analysis, comprising 240 males and 212 females.

Variable selection

In our study, we reviewed relevant literature and collected demographic information and disease-specific characteristics. Features and clinical data were extracted from the electronic medical records of the patients. Potential risk factors for the development of DI after pituitary tumor surgery were selected on the basis of published data, clinical expertise, and practical considerations for future clinical implementation[16,23].

Clinical data included the gender and age of patients. Imaging data included tumor height; tumor width; tumor volume; tumor classification; the presence of a posterior pituitary bright spot (PPBS), which is a high signal intensity of the posterior pituitary on magnetic resonance imaging (MRI) T1WI[24]; preoperative and postoperative pituitary stalk length and their difference (Fig. 1A,B); and preoperative and postoperative sellar diaphragm height and their difference (Fig. 1C,D).

Laboratory results included perioperative serum levels of insulin-like growth factor-I, prolactin, free triiodothyronine (FT3), free thyroxine (FT4), thyroid-stimulating hormone (TSH), cortisol, adrenocorticotropic hormone (ACTH), and serum sodium levels at different postoperative stages. Surgical data included intraoperative cerebrospinal fluid leakage (CSFL). The postoperative urine output of patients was recorded for 7 days. Patients who met the following criteria were defined as having AVP-D after pituitary adenoma surgery[20,25–28]:

1. Meeting both of the following criteria:

1. Urine output increased to 3 L/day, > 5 mL/kg/hour, or > 400 mL/hour.
2. Urine specific gravity < 1.005.

2. Meeting at least one of the following criteria:

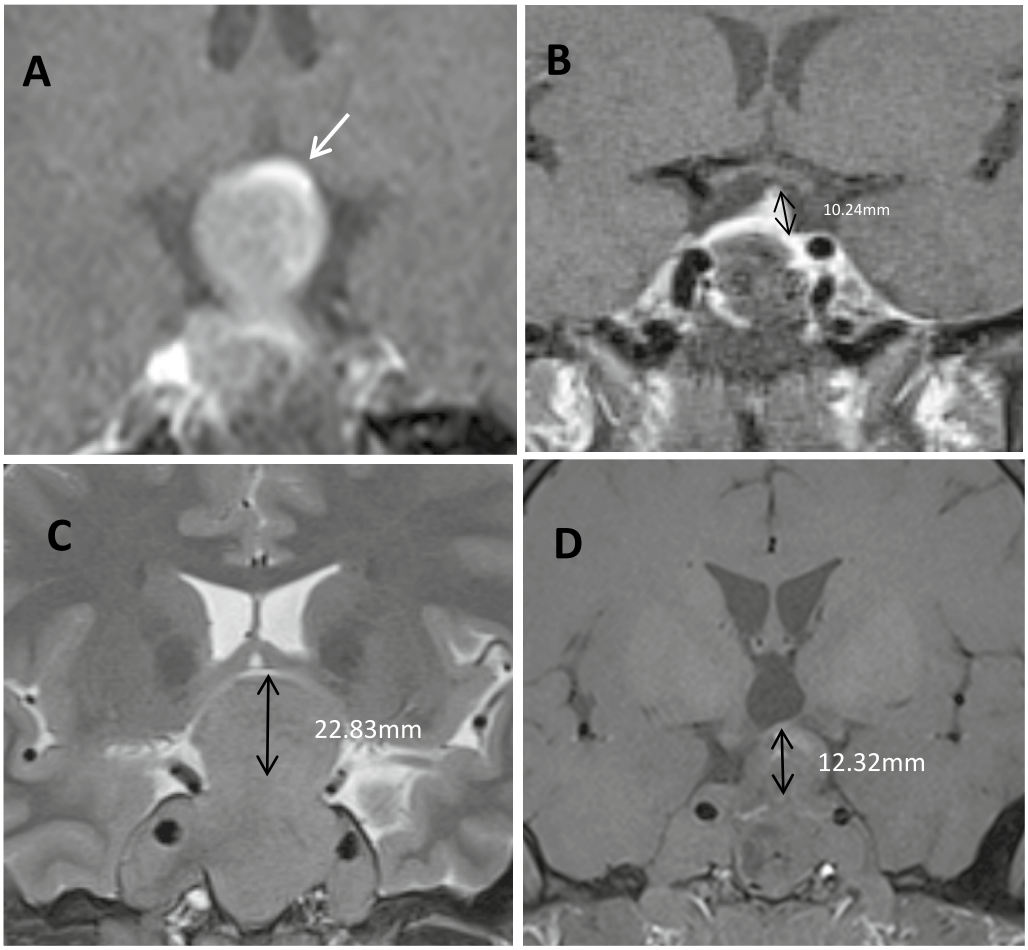


Fig. 1. (**A**,**B**) Measurement of "preoperative pituitary stalk length" on MRI T1-enhanced images. (**A**) Preoperative imaging shows a curved pituitary stalk (white arrow) compressed by a pituitary tumor, with a measured length of 0 mm. (**B**) Postoperative elongation of the pituitary stalk, with a length of 10.24 mm labeled as "postoperative pituitary stalk." (**C**,**D**) MRI-T2WI showing the measurement of the sellar diaphragm height (the distance between the plane where the sellar diaphragm height begins to ascend and the plane of its highest point). (**C**) The preoperative elevation height of the sellar diaphragm height was 22.83 mm. (**D**) The postoperative elevation sellar diaphragm height was 12.32 mm. The descent depth of the sellar diaphragm height is the difference between the two values (22.83–12.32 mm).

1. Serum osmolality > 300 mOsm/kg.
2. Serum sodium > 145 mmol/L.
3. Sensation of thirst.

Data Preprocessing and Model Construction

The data from the 452 patients were randomly divided into a training set (70%) and a test set (30%). Before model construction, chi-squared and nonparametric tests were conducted to evaluate the significance of each variable ($p < 0.05$), and thus, identify relevant feature variables. The selected variables were then used to construct models by using the following representative supervised machine learning algorithms: logistic regression (LR), support vector classification (SVC), random forest (RF), decision tree (DT), *k*-nearest neighbors (*k*-NN), and extreme gradient boosting (XGBoost).

After selecting and organizing relevant feature variables, the LR, SVC, RF, DT, *k*-NN, and XGBoost algorithms were applied to the training set data to build the corresponding models. Random fivefold cross-validation was performed on the training set to assess the internal performance of the models. In particular, the dataset was randomly divided into five groups, Each time, four groups were used for training and one group for validation. The process was repeated five times. This method allowed us to determine the optimal parameters for each model.

The established models were evaluated for predictive performance by using the test set data, comparing various metrics, such as accuracy (ACC), true positive rate (TPR), false positive rate (FPR), positive predictive value (PPV), F1 score (FSC), sensitivity (SEN), specificity (SPE), and negative predictive value (NPV). Receiver operating characteristic (ROC) curves were plotted, and area under the curve (AUC) was calculated. The goodness of fit of the models was tested using the Hosmer–Lemeshow method, and calibration curves were plotted

to verify model calibration. Clinical decision curve analysis (DCA) was conducted to evaluate the clinical utility of the models. The best machine learning model was selected on the basis of the comprehensive performance and stability of the predictions.

Statistical analysis

Sample size was based on the available data, and statistical analysis was performed using Python version 3.1.2. Categorical variables were expressed as percentages and analyzed using chi-squared tests or Fisher's exact tests. Continuous variables that followed a normal distribution were reported as means and standard deviations and analyzed using independent sample *t*-tests. Non-normally distributed continuous variables were reported as medians and interquartile ranges and analyzed using nonparametric tests. A two-sided p-value of < 0.05 was considered statistically significant.We did not correct for multiple testing in this exploratory study.

Results

Feature variable selection

In total, there were 224 cases of AVP-D occurring postoperatively.The significance analysis results (Table 1) indicated significant differences ($p < 0.05$) in several features between patients who developed postoperative AVP-D and those who did not. The selected features included gender, tumor height, preoperative sellar diaphragm height, difference in sellar diaphragm height pre- and post-operation, preoperative pituitary stalk length, difference in pituitary stalk length pre- and post-operation, preoperative ACTH levels, difference in ACTH levels pre- and post-operation, and preoperative cortisol levels. These features were incorporated into various machine learning algorithms to develop predictive models for DI after pituitary adenoma surgery.

Model Construction Results

Selection of Feature Variables

Using LR analysis, we examined the linear relationships between each feature variable and postoperative AVP-D (Fig. 2). The results indicated that tumor height, preoperative sellar diaphragm height, difference in sellar diaphragm height pre- and post-operation, and difference in pituitary stalk length pre- and post-operation were positively associated with the occurrence of postoperative AVP-D. Conversely, cortisol levels, preoperative pituitary stalk length, preoperative ACTH levels, and postoperative ACTH levels were negatively associated with postoperative AVP-D.

Evaluation Results of Different Machine Learning Models

Six different machine learning models were developed, and their ROC curves were plotted for the test and training sets. Calibration curves were also plotted to analyze the fit and generalization capabilities of the models. Shapley additive explanation (SHAP) scores were used to interpret the decision-making process of each model and identify the importance of various factors in different predictive models (Fig. 3).

The ROC curves for the DT, RF, and XGBoost models presented excellent and stable performance on the training and test sets. Among the six models, the calibration curves for the DT, *k*-NN, and RF models performed well, while the calibration curves for the three other models were less satisfactory.

The SHAP scores revealed that the preoperative pituitary stalk length was the most important factor in the DT, RF, and XGBoost models. In the LR model, changes in the pituitary stalk length also exerted a significant effect. Preoperative ACTH, as a representative hormone of pituitary adenomas, played an important role in the *k*-NN, SVC, and DT models. Other features, such as tumor height and preoperative sellar diaphragm height, also exhibited considerable importance in various models.

Evaluation of Different Machine Learning Model Performance

Considering the predictive performance of the models (Table 2), DT excels in various metrics, particularly with an AUC of 0.95, indicating its strong ability to differentiate between positive and negative classes. Overall, high ACC and other metrics, such as TPR and FSC, are also high, demonstrating its stable performance. By contrast, *k*-NN exhibits lower performance, particularly with an AUC of 0.700 and lower overall ACC, suggesting weaker ability to differentiate between positive and negative classes. Higher FPR (0.368) indicates a higher rate of false positives. RF performs the best across all metrics, with an AUC of 0.96, demonstrating its extremely strong ability to differentiate between positive and negative classes. High ACC and low FPR demonstrate its superior performance in all aspects, making RF a highly reliable model. SVC exhibits average overall performance, with lower AUC (0.67), indicating limited discriminative ability. Although PPV is relatively high (0.768), TPR and NPV are low, indicating weaker performance in recall and predicting negative classes. XGBoost demonstrates strong overall performance, particularly with an AUC of 0.91, indicating good discriminative ability. All metrics are relatively balanced, making XGBoost a highly stable model. By contrast, LR performs the worst, with an AUC of 0.54, which is close to random guessing. All metrics are relatively low, indicating poor performance in the classification task.

Considering the goodness of fit of the models (Table 3), DT and RF exhibit more prominent performance. Both have low Brier scores (0.09), indicating highly accurate prediction probabilities, and high Hosmer–Lemeshow p-values, indicating no significant calibration issues. By contrast, XGBoost and SVC have higher Brier scores and poorer calibration than DT and RF, with XGBoost suffering from serious calibration issues. *k*-NN and LR have the highest Brier scores, indicating the poorest calibration of prediction probabilities. *k*-NN is particularly severe, with an infinite Hosmer–Lemeshow statistic and a p-value of 0, indicating that the model is completely unsuitable for the data. The low Brier score of DT (0.09) indicates a small difference between the

Variables	NOT (n = 228)	DI (n = 224)	p
Sex, n (%)			0.071
Male	111 (49)	129 (58)	
Female	117 (51)	95 (42)	
Age, median (Q1, Q3)	49 (39, 61)	50 (40.75, 57.25)	0.935
Tumor anterior posterior diameter, median (Q1, Q3)	16.79 (13, 21.56)	17.59 (13.76, 21.41)	0.376
Tumor left and right diameters, median (Q1, Q3)	21.22 (16.06, 26.19)	21.66 (17.76, 27)	0.235
Tumor height, median (Q1, Q3)	21.66 (15, 28.01)	25 (17.47, 32.23)	0.004
Tumor volume, median (Q1, Q3)	3768 (1797.21, 7064.44)	4454.94 (2117.75, 8718.04)	0.076
Preoperative sellar diaphragm height (Q1, Q3)	8.65 (4.24, 13.29)	9.39 (5.9, 14.6)	0.011
Postoperative sellar diaphragm height, median (Q1, Q3)	4.16 (0, 9.09)	4.76 (0, 9.22)	0.913
Sellar diaphragm height difference, median (Q1, Q3)	1.63 (0, 5.63)	3.5 (0, 7.01)	0.023
CSFL, n (%)			0.472
No	194 (85)	184 (82)	
Yes	34 (15)	40 (18)	
PPBS position, n (%)			0.418
Upside	104 (46)	91 (41)	
Bottom	93 (41)	91 (41)	
Both	7 (3)	12 (5)	
Without	24 (11)	30 (13)	
Stroke or not, n (%)			0.528
Yes	81 (36)	87 (39)	
No	147 (64)	137 (61)	
Preoperative pituitary stalk length, median (Q1, Q3)	4.07 (0.2, 6.8)	2.42 (0, 6.11)	< 0.001
Postoperative pituitary stalk length, median (Q1, Q3)	6.19 (4.05, 7.93)	6.3 (4.03, 8.11)	0.997
Pituitary stalk length difference, median (Q1, Q3)	0.22 (− 0.2, 3.74)	0.72 (0, 4.99)	< 0.001
Preoperative blood sodium, median (Q1, Q3)	140.7 (138.67, 142.5)	140.9 (139, 142.48)	0.762
Blood sodium on the first day after surgery, median (Q1, Q3)	139.7 (137.4, 142)	140 (138, 142.2)	0.094
Blood sodium difference 1, median (Q1, Q3)	0.75 (− 1.2, 3)	0.05 (− 2, 3)	0.162
Blood sodium on the second day after surgery, median (Q1, Q3)	140.2 (138, 143)	141 (138.17, 143)	0.372
Blood sodium difference 2, median (Q1, Q3)	0 (− 2.85, 2.5)	0 (− 3, 2.63)	0.505
FT3, median (Q1, Q3)	4.25 (3.68, 4.85)	4.18 (3.56, 4.84)	0.61
Postoperative FT3, median (Q1, Q3)	3.22 (2.81, 3.62)	3.17 (2.75, 3.68)	0.138
FT3 difference, median (Q1, Q3)	0.95 (0.54, 1.47)	1.06 (0.6, 1.55)	0.207
FT4, median (Q1, Q3)	13.52 (11.04, 15.52)	13.23 (10.3, 15.28)	0.287
Postoperative FT4, median (Q1, Q3)	13.61 (11.81, 15.89)	13.62 (11.09, 16.18)	0.796
FT4 difference, median (Q1, Q3)	− 0.47 (− 1.84, 0.85)	− 0.93 (− 2.06, 0.55)	0.109
TSH, median (Q1, Q3)	1.42 (0.81, 2.19)	1.4 (0.85, 2.31)	0.675
Postoperative TSH, median (Q1, Q3)	0.58 (0.3, 1)	0.6 (0.31, 0.95)	0.828
TSH difference, median (Q1, Q3)	0.64 (0.16, 1.38)	0.7 (0.17, 1.45)	0.497
Cortisol, median (Q1, Q3)	13.45 (8.21, 18.5)	12.56 (3.49, 17.7)	0.03
Postoperative cortisol, median (Q1, Q3)	22.45 (13.97, 36.12)	21.51 (11.43, 34.36)	0.524
Cortisol difference, median (Q1, Q3)	− 9.12 (− 19.96, 0.6)	− 10.13 (− 23.35, − 2.2)	0.231
ACTH, median (Q1, Q3)	27.13 (15.64, 42.52)	22.05 (9.75, 33.43)	0.001
Postoperative ACTH, median (Q1, Q3)	21.88 (13.91, 35.7)	18.68 (8.85, 34.76)	0.033
ACTH difference, median (Q1, Q3)	2.69 (− 7.42, 13.5)	− 0.54 (− 8.52, 10.96)	0.076

Table 1. Baseline characteristics of significant variables studied for postoperative DI after endoscopic endonasal transsphenoidal pituitary adenoma resection. *p* values were corrected for multiple testing by using the control of the false discovery rate at 5% and standard deviation. "pre", preoperative; "post, postoperative; CSFL, cerebrospinal fluid leak during surgery; PPBS, posterior pituitary bright spot.

predicted probabilities and actual results, indicating good model calibration. The high Hosmer–Lemeshow p-value (0.85) suggests a good fit between the model and the data, with no significant calibration issues.

Considering the clinical utility of the models (Fig. 4), we calculated the net benefit at different thresholds by using the true labels and model predicted probabilities on the test set and the plotted decision curves for the six models. These decision curves show the net benefit of each model at different thresholds. From the graph, the DT, RF, and XGBoost models evidently demonstrate higher net benefits across a wide range of thresholds.

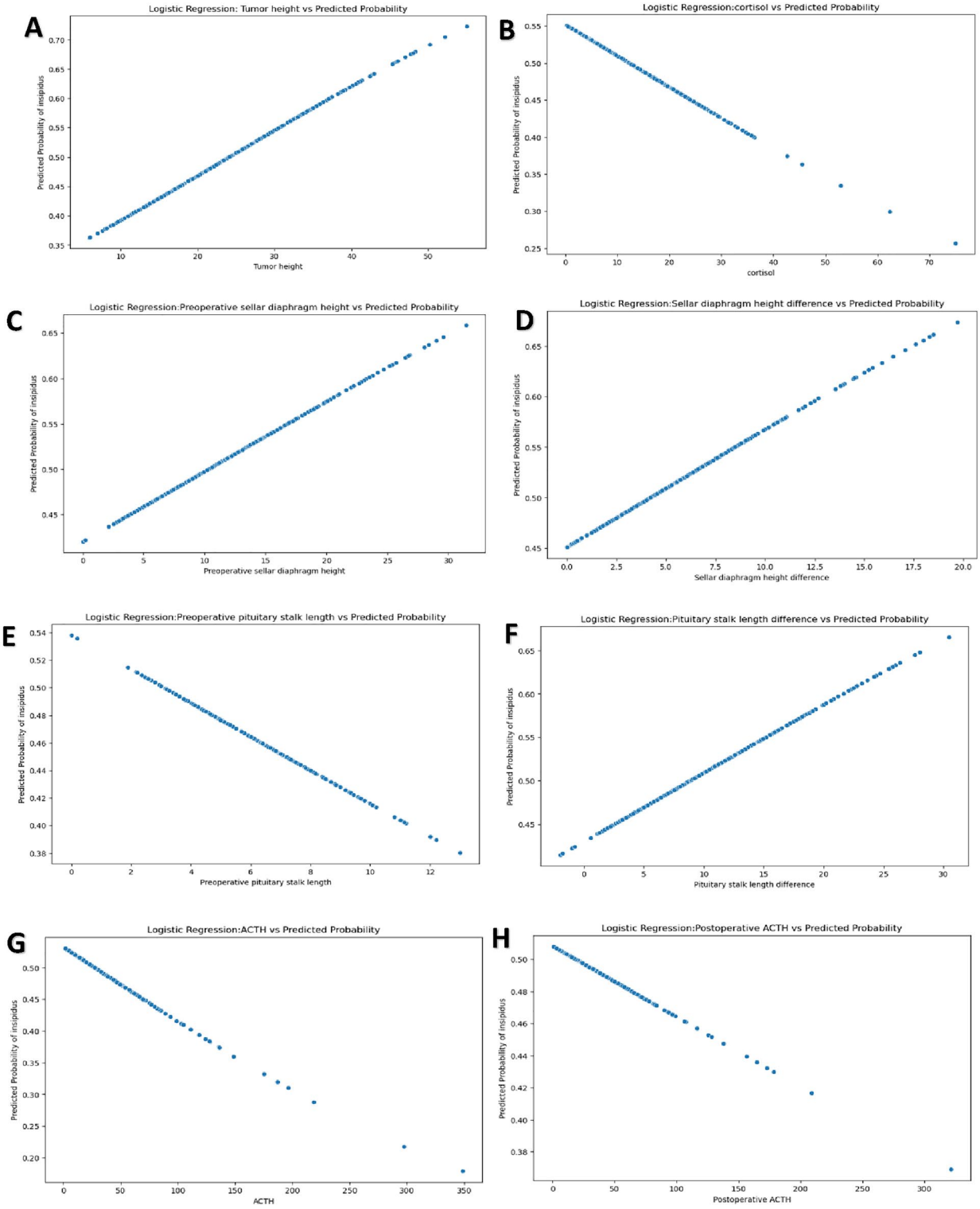


Fig. 2. LR analysis of the linear relationship between each feature variable and postoperative DI. (**A**). Positive correlation between tumor height and postoperative AVP-D. As tumor height increases, the likelihood of postoperative AVP-D occurrence increases. (**B**). Negative correlation between cortisol hormone level and postoperative AVP-D. As cortisol level increases, the likelihood of postoperative AVP-D occurrence decreases. (**C**) Positive correlation between preoperative sellar diaphragm height and postoperative AVP-D. As preoperative sellar diaphragm height increases, the likelihood of postoperative AVP-D occurrence increases. (**D**) Positive correlation between difference in sellar diaphragm height pre- and post-surgery and postoperative AVP-D. As difference in sellar diaphragm height between pre- and post-surgery increases, the likelihood of postoperative AVP-D occurrence increases. (**E**) Negative correlation between preoperative pituitary stalk length and postoperative AVP-D. As preoperative measurable pituitary stalk length increases, the likelihood of postoperative AVP-D occurrence decreases. (**F**) Positive correlation between difference in pituitary stalk length pre- and post-surgery and postoperative AVP-D. The the greater the difference in pituitary stalk length, the higher the likelihood of postoperative AVP-D occurrence. (**G**) Negative correlation between ACTH level and postoperative AVP-D. As ACTH level increases, the likelihood of postoperative AVP-D occurrence decreases. (**H**) Negative correlation between postoperative ACTH level and postoperative AVP-D. As postoperative ACTH level increases, the likelihood of postoperative AVP-D occurrence decreases.

Among them, the net benefit of the RF model is higher than those of the two other models, indicating that the RF model can provide greater clinical utility at these thresholds.

In summary, RF is the optimal model, closely followed by DT. The two models not only exhibit excellent performance in classification but also demonstrate excellent probability calibration.

Discussion

This study systematically applied various machine learning techniques to predict the occurrence of AVP-D after pituitary adenoma surgery. To our knowledge,it is the first to include multiple preoperative radiological markers of the pituitary in machine learning research, establishing machine learning-based clinical prediction models. The results of this study demonstrate that the RF model exhibits high accuracy and precision in predicting AVP-D after pituitary adenoma surgery. Changes in pituitary stalk length and pituitary hormone concentrations play extremely important roles in predicting AVP-D. We validated the effectiveness of MR features in the field of machine learning for predicting AVP-D after pituitary adenoma surgery and further refined the diagnostic benefits of MR imaging and hormone levels for AVP-D[28].This study established an accurate and reliable prediction model for AVP-D after pituitary adenoma surgery and verified the important predictive functions of intraoperative pituitary stalk stretching and perioperative hormone levels on the occurrence of AVP-D.

The occurrence of AVP-D is often accompanied by hyponatremia. For most patients with AVP-D, early control of urine output is a crucial factor affecting the quality of life post-pituitary adenoma surgery[28]. Prolonged polyuria can lead to symptoms such as nausea, dehydration, and electrolyte disturbances. More severely, as urine output further increases, the late use of AVP analogs often fails to alleviate irreversible damage caused by dehydration and necrosis of brain cells. Therefore, early prediction of AVP-D can assist clinicians in controlling urine output sooner and promptly using AVP analogs to manage diabetes insipidus symptoms, thereby preventing further damage to the body.

The occurrence of AVP-D is often due to abnormalities in antidiuretic hormone secretion caused by damage to the hypothalamus.In many patients with pituitary adenomas, the pituitary stalk is frequently compressed due to the growth of pituitary adenoma.Studies have shown that more than 50% of pituitary adenoma patients experience compression of the pituitary stalk, with this occurrence nearly doubling in those with macroadenomas[29]. Preoperative measurement of tumor and sellar diaphragm height can reflect the extent of tumor invasion into the sella, thereby indirectly assessing the degree of compression on the pituitary stalk.Temporary decompression of the tumor cavity during surgery may cause the pituitary tumor to retract into the sphenoid sinus cavity, resulting in the temporary stretching of the compressed pituitary stalk. This significant stretching may indirectly lead to damage to the hypothalamus, resulting in abnormalities in antidiuretic hormone secretion, and further leading to the occurrence of AVP-D after pituitary adenoma surgery. Therefore, appropriate intrasellar filling methods can avoid the temporary stretching of the pituitary stalk, reducing the possibility of postoperative AVP-D[30,31].

The effect of age on the formation of AVP-D after pituitary adenoma surgery has not yet been definitively concluded[12,18, 20, 32, 33]. However, most researchers believe that factors, such as immunosuppression in elderly patients and their slower postoperative recovery, may be important reasons for the development of postoperative AVP-D. In the current study, age did not exhibit significant differences, and thus, it was not included in the research. Gender is frequently used in the establishment of clinical prediction models. Although the results did not show a statistically significant difference, we included it in this study because of its importance as a basic clinical element. Changes in serum pituitary hormones often indicate the occurrence of AVP-D.In existing studies, patients with prolactin-secreting adenomas and ACTH-secreting adenomas often experience transient AVP-D after surgery[10,34]. Burke et al. found that patients with TSH-secreting adenomas and follicle-stimulating hormone/luteinizing hormone-secreting adenomas exhibited a higher probability of developing AVP-D after surgery, providing new research ideas for analyzing AVP-D after pituitary adenoma surgery[12]. In the current study, ACTH similarly demonstrated extremely high predictive efficacy for AVP-D in various models,it may be due to arginine vasopressin being an important secretagogue for ACTH. Damage to the hypothalamus or pituitary gland can lead to a decrease in arginine vasopressin, resulting in reduced ACTH secretion. The negative correlation between these factors is consistent with the results of the previous linear analysis.Although cortisol hormone performed well in significant research, it did not exhibit good predictive performance through SHAP analysis after model establishment, which might be related to the misdiagnosis of AVP-D caused by a decrease in urine concentration. In summary, attention to electrolyte abnormalities during the perioperative period can reduce the

Fig. 3. Confusion Matrices, ROC Curves, Calibration Curves, and SHAP Values for the Six Machine Learning Models. (**A**) DT: AUC of 0.95 for training and test sets, indicating excellent model performance. The calibration curve demonstrates good model fit and generalization. SHAP values signify that preoperative pituitary stalk length and ACTH are the most important features. (**B**) *k*-NN: AUC of 0.76 for the training set and 0.70 for the test set, indicating moderate model performance. The calibration curve demonstrates good model fit and generalization. SHAP values signify that postoperative ACTH and tumor height are the most important features. (**C**) RF: AUC of 1.0 for the training set and 0.96 for the test set, indicating outstanding model performance. The calibration curve demonstrates good model fit and generalization. SHAP values signify that preoperative pituitary stalk length and preoperative sellar diaphragm height are the most important features. (**D**) SVC: AUC of 0.68 for the training set and 0.67 for the test set, indicating poor model performance. The calibration curve demonstrates poor model fit and generalization. SHAP values signify that ACTH and preoperative sellar diaphragm height are the most important features. (**E**) XGBoost: AUC of 1.0 for the training set and 0.91 for the test set, indicating excellent model performance. The calibration curve demonstrates poor model fit and generalization. SHAP values signify that preoperative pituitary stalk length and the difference in preoperative sellar diaphragm height are the most important features. (**F**) LR: AUC of 0.62 for the training set and 0.54 for the test set, indicating very poor model performance. The calibration curve demonstrates extremely poor model fit and generalization. SHAP values signify that the difference in preoperative pituitary stalk length and tumor height are the most important features.

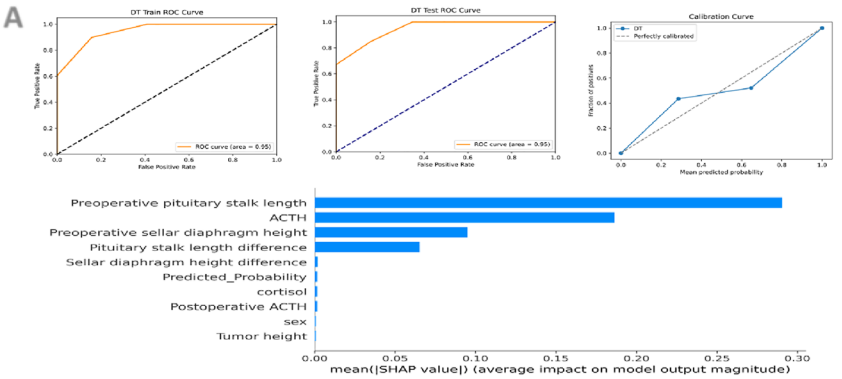
incidence of AVP-D after pituitary adenoma surgery.Based on the above analysis, we believe that the compression and traction of the pituitary stalk before and during surgery indirectly cause damage to the hypothalamus, leading to abnormal release of AVP, which in turn results in abnormal changes in serum pituitary hormones.

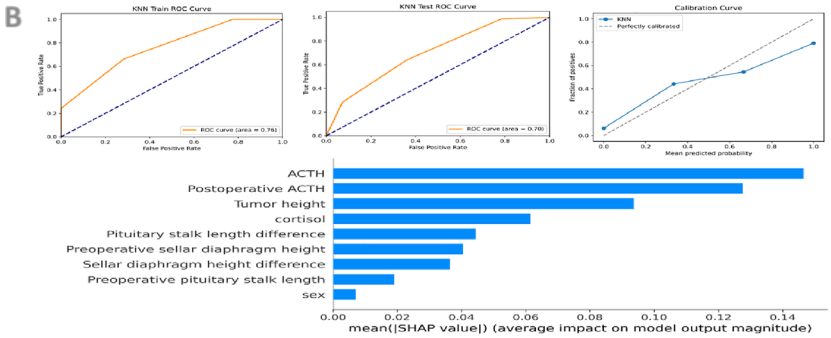
Traditional clinical statistical models have provided excellent clinical performance results when dealing with limited variables, high homogeneity, and small-scale data. In recent years, however, factors used to predict complications after pituitary adenoma surgery have gradually increased, including patients' basic information, clinical radiological features, and clinical endocrine test result indicators, which are all popular research topics. The mechanism of complications after pituitary adenoma surgery is highly complex, with these variables interacting and influencing one another and demonstrating inseparable relationships. Establishing effective and accurate clinical models with many indicators has become increasingly difficult with traditional statistical models. Machine learning models incorporate a broader range of potential risk factors and predictive variables in predictive analysis. The input and output of machine learning models are known, but the specific decision-making process is unclear. Compared with traditional clinical models, the interpretability of machine learning models is decreased, but they can achieve higher predictive accuracy. The use of machine learning models to predict AVP-D after pituitary adenoma surgery is an effective and feasible clinical research method.

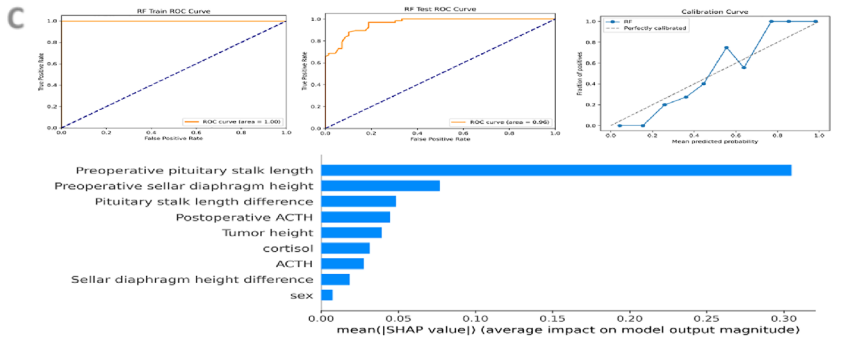
This study has certain limitations. (1) This study is retrospective, and certain information bias should be considered when interpreting a model and related results. (2) Compared with large-sample machine learning, the number of samples in this study is small, and the samples are from a single center. This condition may affect the accuracy of a model, leading to overfitting and poor generalization ability.

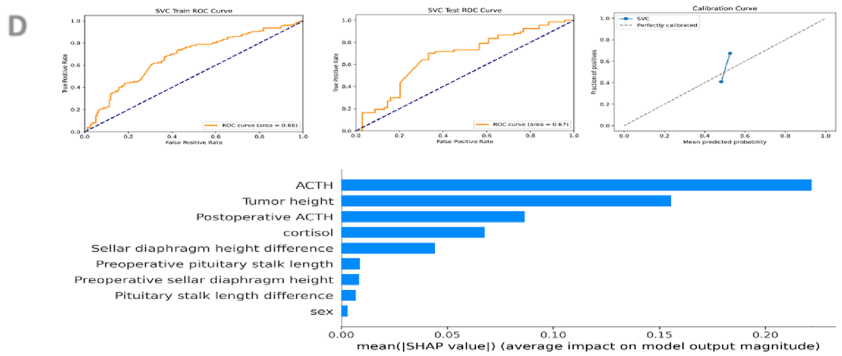
Conclusion

This study established multiple machine learning models to predict the occurrence of DI after pituitary adenoma surgery. Among which, the RF model demonstrated the best predictive performance. The length of the preoperative pituitary stalk and the difference in pituitary stalk length between preoperative and postoperative periods are important predictors of DI after pituitary adenoma surgery. A shorter preoperative pituitary stalk length is associated with a higher degree of pituitary stalk elongation during surgery, increasing the likelihood of postoperative DI. The preoperative and postoperative concentrations of ACTH and the preoperative concentrations of cortisol also exhibit predictive efficacy for DI after pituitary adenoma surgery. Lower levels of ACTH and preoperative cortisol are associated with a higher risk of postoperative DI.


A
DT Train ROC Curve
DT Test ROC Curve
Calibration Curve
Preoperative pituitary stalk length
ACTH
Preoperative sellar diaphragm height
Pituitary stalk length difference
Sellar diaphragm height difference
Predicted_Probability
cortisol
Postoperative ACTH
sex
Tumor height
mean(|SHAP value|) (average impact on model output magnitude)


B
KNN Train ROC Curve
KNN Test ROC Curve
Calibration Curve
ACTH
Postoperative ACTH
Tumor height
cortisol
Pituitary stalk length difference
Preoperative sellar diaphragm height
Sellar diaphragm height difference
Preoperative pituitary stalk length
sex
mean(|SHAP value|) (average impact on model output magnitude)


C
RF Train ROC Curve
RF Test ROC Curve
Calibration Curve
Preoperative pituitary stalk length
Preoperative sellar diaphragm height
Pituitary stalk length difference
Postoperative ACTH
Tumor height
cortisol
ACTH
Sellar diaphragm height difference
sex
mean(|SHAP value|) (average impact on model output magnitude)


D
SVC Train ROC Curve
SVC Test ROC Curve
Calibration Curve
ACTH
Tumor height
Postoperative ACTH
cortisol
Sellar diaphragm height difference
Preoperative pituitary stalk length
Preoperative sellar diaphragm height
Pituitary stalk length difference
sex
mean(|SHAP value|) (average impact on model output magnitude)

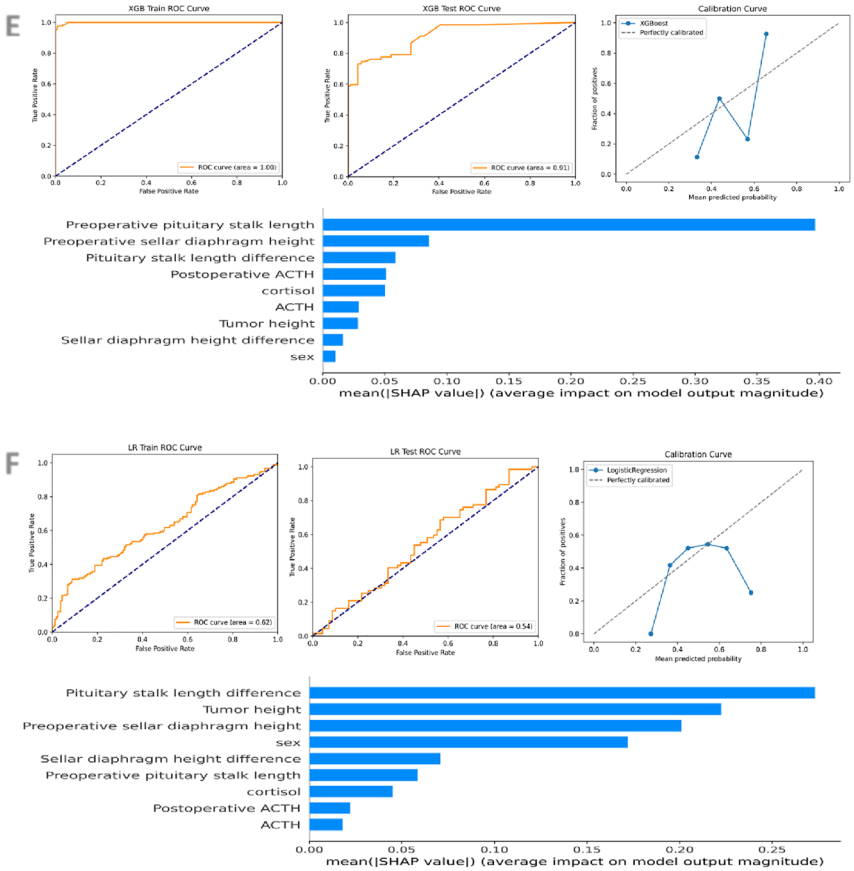


Fig. 3. (continued)

Model	AUC	ACC	TPR	FPR	PPV	SPE	SEN	NPV	FSC
DT	0.95	0.843	0.848	0.162	0.836	0.838	0.848	0.851	0.842
k-NN	0.700	0.640	0.647	0.368	0.638	0.632	0.647	0.642	0.642
RF	0.96	0.882	0.873	0.108	0.899	0.892	0.873	0.866	0.886
SVC	0.67	0.647	0.624	0.314	0.768	0.686	0.624	0.522	0.689
XGBoost	0.91	0.794	0.797	0.209	0.797	0.791	0.797	0.689	0.797
LR	0.54	0.522	0.527	0.484	0.565	0.516	0.527	0.478	0.545

Table 2. Performance comparison of the six different machine learning models. AUC, area under the curve; ACC, accuracy; TPR, true positive rate; FPR, false positive rate; PPV, positive predictive value; SPE, specificity; SEN, sensitivity; NPV, negative predictive value; FSC, F1 Score.

Model	Brier score	Hosmer–Lemeshow statistic	Hosmer–Lemeshow *p*-value
DT	0.09	4.11	0.85
k-NN	0.23	inf	0.00
RF	0.09	13.53	0.09
XGBoost	0.17	35.08	0.00
SVC	0.24	12.08	0.15
LR	0.25	8.87	0.35

Table 3. Results of Hosmer–Lemeshow test.

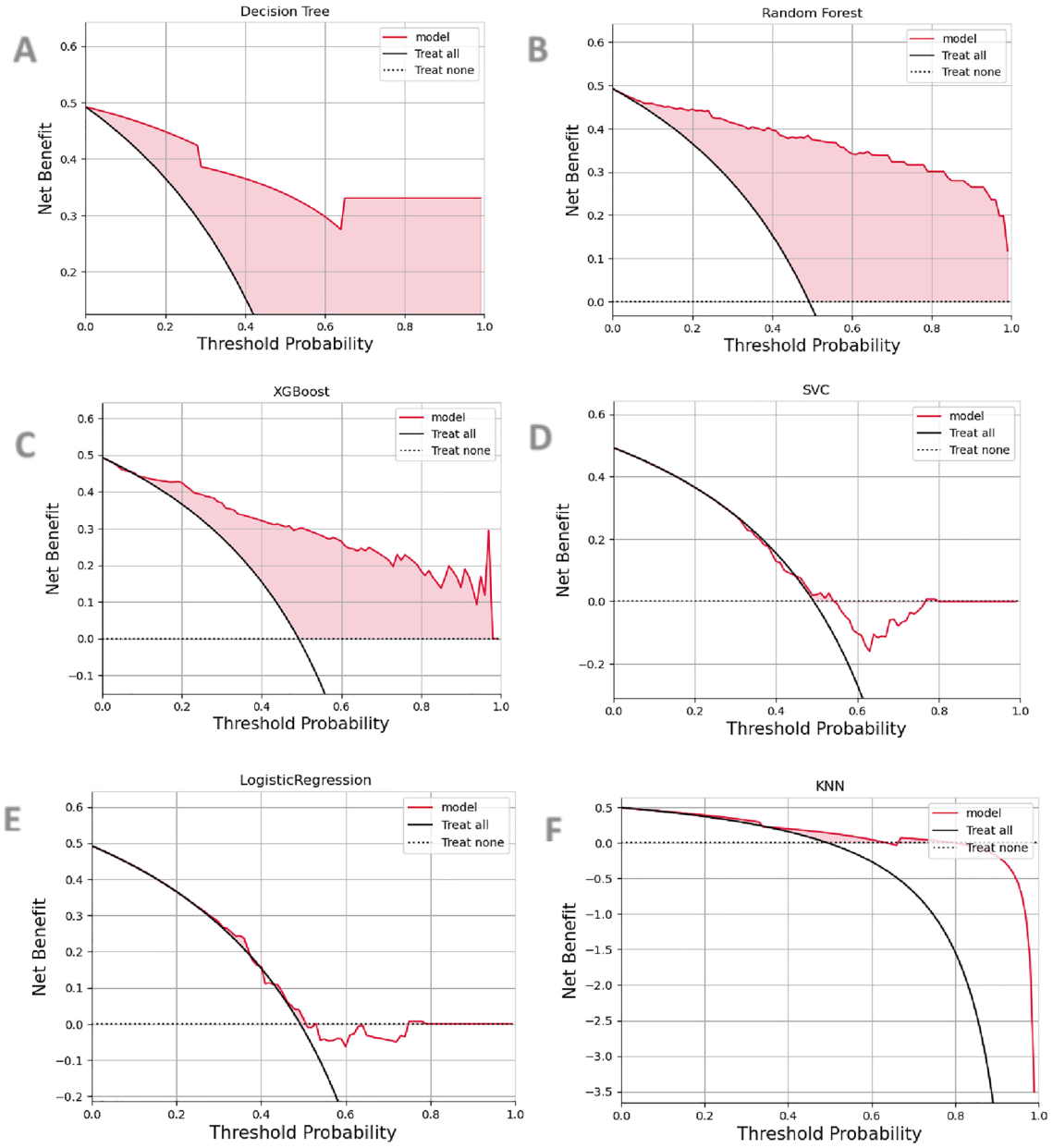


Fig. 4. Comparison of the decision curves of the six machine learning models. (**A**) The net benefit curve is consistently higher than "Treat all" and "Treat none" at all thresholds and stabilizes gradually as threshold increases. (**B**) The net benefit curve is higher than "Treat all" and "Treat none" at most thresholds, showing a slow decline as threshold increases. (**C**) The net benefit curve is higher than "Treat all" and "Treat none" at all thresholds, but decreases gradually and becomes unstable as threshold increases. (**D**) The net benefit is close to "Treat none" at most thresholds, resulting in low overall net benefit. (**E**) The net benefit curve nearly overlaps with "Treat all" when threshold is below 0.4 and tends toward zero when threshold is above 0.6, resulting in low overall net benefit. (**F**) The net benefit is consistently low at all thresholds, even lower than "Treat none," resulting in low overall net benefit.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 12 June 2024; Accepted: 9 September 2024
Published online: 27 September 2024

References

1. Songtao, Q. Anatomical study of membranous structures related to pituitary adenomas and its clinical significance. *Chin. J. Neurosurg.* **33**(2), 109–112. https://doi.org/10.3760/cma.j.issn.1001-2346.2017.02.001 (2017).
2. Zhang, X., Yang, F. & Han, N. Recurrence rate and exploration of clinical factors after pituitary adenoma surgery: A systematic review and meta-analysis based on computer artificial intelligence system. *Comput. Intell. Neurosci.* **2022**, 6002672. https://doi.org/10.1155/2022/6002672 (2022).
3. Mortini, P. *et al.* Pituitary Surgery. *Presse Med.* **50**(4), 104079. https://doi.org/10.1016/j.lpm.2021.104079 (2021).
4. Artico, M. *et al.* The contribution of Davide Giordano (1864–1954) to pituitary surgery: The transglabellar-nasal approach. *Neurosurgery* **42**(4), 909–912. https://doi.org/10.1097/00006123-199804000-00121 (1998).
5. Arima, H. *et al.* Changing the name of diabetes insipidus: A position statement of The Working Group for Renaming Diabetes Insipidus. *Endocr. Connect.* **11**(11), P1. https://doi.org/10.1530/eje-22-0751 (2022).
6. Schreckinger, M., Szerlip, N. & Mittal, S. Diabetes insipidus following resection of pituitary tumors. *Clin. Neurol. Neurosurg.* **115**(2), 121–126. https://doi.org/10.1016/j.clineuro.2012.08.009 (2013).
7. Lin, K. *et al.* Diaphragma sellae sinking can predict the onset of hyponatremia after transsphenoidal surgery for pituitary adenomas. *J. Endocrinol. Invest.* **44**(11), 2511–2520. https://doi.org/10.1007/s40618-021-01611-7 (2021).
8. Senior, B. A. *et al.* Minimally invasive pituitary surgery. *Laryngoscope* **118**(10), 1842–1855. https://doi.org/10.1097/MLG.0b013e31817e2c43 (2008).
9. Xue, L. *et al.* Change in the pituitary stalk deviation angle after transsphenoidal surgery can predict the development of diabetes insipidus for pituitary adenomas. *Endocr. Connect.* **11**(11), e220187. https://doi.org/10.1530/EC-22-0187 (2022).
10. Nemergut, E. C. *et al.* Predictors of diabetes insipidus after transsphenoidal surgery: A review of 881 patients. *J. Neurosurg.* **103**(3), 448–454. https://doi.org/10.3171/jns.2005.103.3.0448 (2005).
11. Seckl, J. & Dunger, D. Postoperative diabetes insipidus. *BMJ* **298**(6665), 2–3. https://doi.org/10.1136/bmj.298.6665.2 (1989).
12. Burke, W. T. *et al.* Diabetes insipidus after endoscopic transsphenoidal surgery. *Neurosurgery* **87**(5), 949–955. https://doi.org/10.1093/neuros/nyaa148 (2020).
13. Almalki, M. H. *et al.* Management of diabetes insipidus following surgery for pituitary and suprasellar tumours. *Sultan Qaboos Univ. Med. J.* **21**(3), 354–364. https://doi.org/10.18295/squmj.4.2021.010 (2021).
14. Joshi, R. S. *et al.* Identifying risk factors for postoperative diabetes insipidus in more than 2500 patients undergoing transsphenoidal surgery: A single-institution experience. *J. Neurosurg.* **55**, 1–11. https://doi.org/10.3171/2021.11.JNS211260 (2022).
15. Rajan, R. *et al.* Predictors of delayed hyponatraemia after surgery for pituitary tumour. *Horm. Metab. Res.* **55**(6), 395–401. https://doi.org/10.1055/a-2074-9329 (2023).
16. Kinoshita, Y. *et al.* Predictive factors of postoperative diabetes insipidus in 333 patients undergoing transsphenoidal surgery for non-functioning pituitary adenoma. *Pituitary* **25**(1), 100–107. https://doi.org/10.1007/s11102-021-01175-y (2022).
17. Angelousi, A. *et al.* Diabetes insipidus secondary to sellar/parasellar lesions. *J. Neuroendocrinol.* **33**(3), e12954. https://doi.org/10.1111/jne.12954 (2021).
18. Ma, J. *et al.* Correlation of pituitary descent and diabetes insipidus after transsphenoidal pituitary macroadenoma resection. *Neurosurgery* **92**(6), 1269–1275. https://doi.org/10.1227/neu.0000000000002360 (2023).
19. Yasuda, M. E. *et al.* Risk factors related to transient diabetes insipidus development following transsphenoidal pituitary adenoma resection: A multicentric study. *World Neurosurg.* **175**, e636–e643. https://doi.org/10.1016/j.wneu.2023.03.150 (2023).
20. Hollon, T. C. *et al.* A machine learning approach to predict early outcomes after pituitary adenoma surgery. *Neurosurg. Focus* **45**(5), E8. https://doi.org/10.3171/2018.8.FOCUS18268 (2018).
21. Deo, R. C. Machine learning in medicine. *Circulation* **132**(20), 1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593 (2015).
22. Staartjes, V. E. *et al.* Neural network-based identification of patients at high risk for intraoperative cerebrospinal fluid leaks in endoscopic pituitary surgery. *J. Neurosurg.* **133**, 1–7. https://doi.org/10.3171/2019.4.JNS19477 (2019).
23. Kadir, M. L. *et al.* Incidence of diabetes insipidus in postoperative period among the patients undergoing pituitary tumour surgery. *Mymensingh Med. J. MMJ* **26**(3), 642–649 (2017).
24. Wang, S. *et al.* MR imaging analysis of posterior pituitary in patients with pituitary adenoma. *Int. J. Clin. Exp. Med.* **8**(5), 7634–7640 (2015).
25. de Vries, F. *et al.* Postoperative diabetes insipidus: How to define and grade this complication?. *Pituitary* **24**(2), 284–291. https://doi.org/10.1007/s11102-020-01083-7 (2021).
26. Nayak, P. *et al.* Predictors of postoperative diabetes insipidus following endoscopic resection of pituitary adenomas. *J. Endocr. Soc.* **2**(9), 1010–1019. https://doi.org/10.1210/js.2018-00121 (2018).
27. Lin, K. *et al.* Novel nomograms to predict delayed hyponatremia after transsphenoidal surgery for pituitary adenoma. *Front. Endocrinol.* **13**, 900121. https://doi.org/10.3389/fendo.2022.900121 (2022).
28. Hou, S. *et al.* A machine learning-based prediction of diabetes insipidus in patients undergoing endoscopic transsphenoidal surgery for pituitary adenoma. *World Neurosurg.* **175**, e55–e63. https://doi.org/10.1016/j.wneu.2023.03.027 (2023).
29. Vivancos Sánchez, C. *et al.* Role of pituitary stalk and gland radiological status on endocrine function and outcome after endoscopic transsphenoidal surgery for non-functioning pituitary adenomas. *Endocrine* **73**, 416–423. https://doi.org/10.1007/s12020-021-02726-w (2021).
30. Zhan, R. *et al.* Pure endoscopic endonasal transsphenoidal approach for nonfunctioning pituitary adenomas in the elderly: Surgical outcomes and complications in 158 patients. *World Neurosurg.* **84**(6), 1572–1578. https://doi.org/10.1016/j.wneu.2015.08.035 (2015).
31. Wang, S. *et al.* Magnetic resonance imaging characteristics of residual pituitary tissues following transsphenoidal resection of pituitary macroadenomas. *Neurol. India* **69**(4), 867–873. https://doi.org/10.4103/0028-3886.325377 (2021).
32. Hoang, A. N. *et al.* Pituitary stalk stretch predicts postoperative diabetes insipidus after pituitary macroadenoma transsphenoidal resection. *Oper. Neurosurg.* **24**(3), 248–255 (2023).
33. Lin, K. *et al.* The difference between preoperative and postoperative pituitary stalk deviation angles can predict delayed hyponatremia after transsphenoidal surgery. *World Neurosurg.* **155**, e637–e645 (2021).
34. Salata, R. A. *et al.* Vasopressin stimulation of adrenocorticotropin hormone (ACTH) in humans. In vivo bioassay of corticotropin-releasing factor (CRF) which provides evidence for CRF mediation of the diurnal rhythm of ACTH. *J. Clin. Investig.* **81**(3), 766–774. https://doi.org/10.1172/JCI113382 (1988).

Author contributions

-Yuyang Chen:Data Curation, Conceptualization, Methodology, Formal Analysis, Writing ; -Jiansheng Zhong: Data Curation -Haixiang Li:Methodology -Kunzhe Lin : Writing—Review & Editing -Liangfeng Wei:Statistical calculations -Shousen Wang:Writing—Review & Editing.

Funding

This study was funded by the Joint Logistics Medical Key Specialty Project (LQZD-SW) and the Fujian Provincial Science and Technology Program Science and Technology Innovation Platform Project (2022Y2017).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.