Accepted Manuscript online: 2022-09-12 Article published online: 2022-11-23

The State of Machine Learning in Outcomes Prediction of Transsphenoidal Surgery: A Systematic Review

Darrion B. Yang[1] Alexander D. Smith[1] Emily J. Smith[1]* Anant Naik[1]* Mika Janbahan[1]* Charee M. Thompson[2] Lav R. Varshney[3] Wael Hassaneen[1,4]

[1] Carle Illinois College of Medicine, University of Illinois Urbana Champaign, Champaign, Illinois, United States
[2] Department of Communication, University of Illinois Urbana Champaign, Champaign, Illinois, United States
[3] Department of Electrical and Computer Engineering, University of Illinois Urbana Champaign, Urbana, Illinois, United States
[4] Department of Neurosurgery, Carle Foundation Hospital, Urbana, Illinois, United States

Address for correspondence Wael Hassaneen, MD, PhD, 610 N Lincoln Avenue, Urbana, IL 61801, United States (e-mail: wael.mostafa@carle.com).

J Neurol Surg B Skull Base 2023;84:548–559.

Abstract

The purpose of this analysis is to assess the use of machine learning (ML) algorithms in the prediction of postoperative outcomes, including complications, recurrence, and death in transsphenoidal surgery. Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, we systematically reviewed all papers that used at least one ML algorithm to predict outcomes after transsphenoidal surgery. We searched Scopus, PubMed, and Web of Science databases for studies published prior to May 12, 2021. We identified 13 studies enrolling 5,048 patients. We extracted the general characteristics of each study; the sensitivity, specificity, area under the curve (AUC) of the ML models developed as well as the features identified as important by the ML models. We identified 12 studies with 5,048 patients that included ML algorithms for adenomas, three with 1807 patients specifically for acromegaly, and five with 2105 patients specifically for Cushing's disease. Nearly all were single-institution studies. The studies used a heterogeneous mix of ML algorithms and features to build predictive models. All papers reported an AUC greater than 0.7, which indicates clinical utility. ML algorithms have the potential to predict postoperative outcomes of transsphenoidal surgery and can improve patient care. Ensemble algorithms and neural networks were often top performers when compared with other ML algorithms. Biochemical and preoperative features were most likely to be selected as important by ML models. Inexplicability remains a challenge, but algorithms such as local interpretable model–agnostic explanation or Shapley value can increase explainability of ML algorithms. Our analysis shows that ML algorithms have the potential to greatly assist surgeons in clinical decision making.

Keywords
- transsphenoidal surgery
- pituitary adenomas
- machine learning
- artificial intelligence
- Cushing's disease
- acromegaly

* *These authors contributed equally.*

received
December 30, 2021
accepted
March 3, 2022
accepted manuscript online
September 12, 2022
article published online
November 23, 2022

DOI https://doi.org/10.1055/a-1941-3618.
ISSN 2193-6331.

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Introduction

The pituitary gland is a multifaceted center of secretion for multiple hormones located in the sella turcica of the sphenoid bone, inferior to the hypothalamus and optic chiasm. Masses that arise from the pituitary gland, called pituitary adenomas, comprise 10 to 15% of intracranial masses.[1] They are classified based on the primary cell origin, which, if the adenoma is functional, dictates the type of hormone secreted. Alternatively, pituitary adenomas that do not secrete significant amounts of hormones are considered nonfunctional and make up 28 to 37% of pituitary adenomas. Pituitary adenomas that are 10 mm or larger are classified as macroadenomas and those that are less than 10 mm are microadenomas.[1] Symptoms of pituitary adenomas are primarily caused by mass effect and/or alterations in pituitary hormone secretion. Symptoms of mass effect include headaches and bitemporal hemianopsia, given its location near the optic chiasm. For functional pituitary adenomas, additional symptoms depend on the hormones secreted. Prolactinomas are the most common, constituting 40 to 57% of pituitary adenomas. Symptoms related to these include galactorrhea, decreases in libido and infertility, gynecomastia in males, and oligo- or amenorrhea in females.[1] Growth hormone (GH)-secreting pituitary adenomas comprise 11 to 13% of pituitary adenomas and can cause symptoms of acromegaly in adults and gigantism in children.[1] One to two percent of pituitary adenomas are adrenocorticotropic hormone (ACTH)-secreting, causing symptoms of Cushing's disease.[1] Much less common are pituitary adenomas that secrete follicle-stimulating hormone, luteinizing hormone (LH), or thyroid-stimulating hormone.

Management of pituitary adenomas depends on the size and function of the tumor and includes both medical and surgical treatment. Prolactinomas are commonly managed with dopamine agonists, such as bromocriptine and cabergoline, which inhibit release of prolactin.[1] Hormonal therapy, however, is less effective for other types of functional pituitary adenomas. Regardless of type, pituitary adenomas that are causing mass effect often need to be surgically removed or debulked. Transsphenoidal surgery (TSS) is the mainstay of treatment for pituitary masses and other skull base diseases. It involves the insertion of either a microscope or an endoscope through an incision in the sphenoid bone and then into the skull base.[2] Adverse events in TSS are uncommon, with mortality rates below 1%, but the complication rates can be significant.[3,4] Complications include cranial nerve injury, vision loss, sinusitis, cerebrospinal fluid (CSF) leaks, infections, bleeding, diabetes insipidus (DI), syndrome of inappropriate secretion of antidiuretic hormone, or recurrence of tumors.[2]

Clinicians must be able to determine good surgical candidates prior to recommending surgery. This is often difficult, as the management of sellar masses may require the expertise of neurosurgeons, otolaryngologists, endocrinologists, along with the assistance of ophthalmologists, oncologists, and neurologists.[5] The ability to quantify individual risk factors has been elusive and few studies have demonstrated links between preoperative factors and postoperative complications due to the relative complexity of these diseases. A systematic review by Lobatto et al found that old age was the only risk factor for overall complications and intraventricular extension of the tumor was the only risk factor for CSF leakage, though it was noted that many of the studies had a high risk of bias and lacked clear definitions for postoperative complications.[6]

Recently, machine learning (ML) algorithms have become increasingly utilized in medical research to find patterns in health-related data and develop predictions based on those patterns.[7,8] There exist two classifications of ML algorithms: supervised and unsupervised, with some classes of algorithms capable of both. Supervised ML trains on data that has already been labeled and classified. If the algorithm has achieved an acceptable accuracy, it can be applied to new, unseen data. Examples of supervised ML algorithms include random forest (RF), boosted algorithms, K-nearest neighbor (KNN), decision tree (DT), naïve bayes (NB), and support vector machines (SVM). Unsupervised ML, on the other hand, trains on data that is unlabeled, and includes dimensionality reduction, anomaly detection, and clustering algorithms. Neural networks (NN), including deep learning NN, can be used for both supervised and unsupervised learning, depending on their type. The unsupervised algorithms may identify attributes within the data that are important in developing predictions.[9]

ML algorithms can be particularly helpful in clinical practice for issues where there are unclear predictive risk factors, such as for outcomes after TSS. Several papers have been published recently using ML to predict outcomes in TSS.[10–12] The application of ML to predictive modeling in TSS is novel, with a small but significant number of papers being published in recent years. In this present systematic review, we compile the known research on the application of ML for TSS, and we investigate the features used by these ML algorithms to predict postoperative outcomes, looking at which are deemed most predictive of poor postoperative outcomes.

Methods

Searches

Our methodology for the systematic review adhered to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (►**Fig. 1**) with the study protocol included in an open access database (PROSPERO ID: CRD42021254552). We searched for studies using ML to predict outcomes after TSS on Scopus, PubMed, and Web of Science databases on May 12, 2021. The search terms are listed in the ►**Supplemental Materials** (available in the online version).

Studies were first identified using the search criteria detailed in the ►**Supplemental Materials** (available in the online version). After duplicates were removed, the title and abstract of the remaining 55 papers were reviewed by three researchers (DY, ES, MJ) using inclusion and exclusion criteria. When there was disagreement, the authors discussed their reasoning

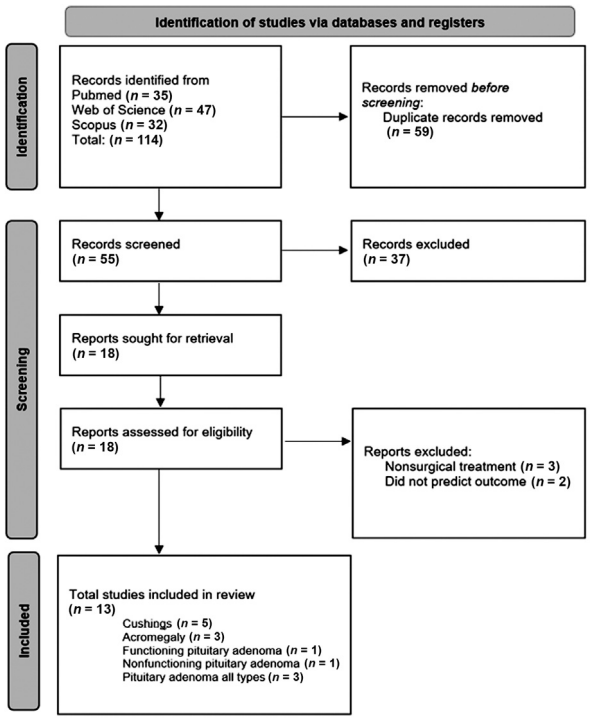

Fig. 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram for the systematic review of machine learning studies on outcomes prediction in transsphenoidal surgery.

and if an agreement could not be reached, the final decision was determined by majority rule (⅔ vote).

Inclusion/Exclusion Criteria

To be included, papers had to report on studies using specific ML algorithms to analyze preoperative, intraoperative, or postoperative patient data to predict postoperative outcomes of TSS in humans. TSS for pituitary adenoma, Cushing's syndrome, acromegaly, and craniopharyngioma were included. Postsurgical outcomes of remission, delayed remission, CSF leak, pituitary insufficiency, injury, and death were also included. Studies must have used at least one ML algorithm with analyses that include sensitivity, specificity, and area under the curve (AUC). Exclusion criteria include studies that were reviews, meta-analyses, case series, and case reports; papers written in languages other than English; and papers without full text available.

Screening

Thirteen papers met the criteria and were further divided by how the disease was classified in each paper: functional pituitary adenomas specifically causing Cushing's disease or acromegaly, nonfunctional pituitary adenomas, all classes of pituitary adenomas unspecified by type, and all classes of functional pituitary adenomas unspecified by type.

Data and Information Extraction

Due to the heterogeneous nature of the papers found, we determined that a quantitative analysis was inappropriate and opted for a qualitative synthesis. Characteristics of each study were extracted, including number of patients, the source of the patient population, the ML algorithms, and the types of pathology investigated, to obtain a general view of the kinds of studies that are being developed. All of the algorithms used by the papers as well as their corresponding AUC values, sensitivities, and specificities were extracted. Finally, the features that were analyzed by the ML algorithms and their importance as determined by the researchers, the algorithms themselves or post hoc with explanation algorithms were collected to investigate which features are most important and which have not been well investigated. The data was extracted by three investigators independently.

Results

Description of Studies

►**Table 1** includes information on number of patients, source of data, type of algorithm used, and type of disease investigated. Single-institution studies were the most common, with only one out of 13 studies being multiinstitutional. All 13 used retrospective data. Five papers focused specifically on Cushing's, three on acromegaly, two on functional pituitary adenoma, one on nonfunctional pituitary adenoma, and two on all types of pituitary adenoma. In summary, these studies included 5,048 participants. A total of 24 algorithms were used throughout the papers, with many papers sharing certain algorithms, but no papers using the same set of ML algorithms. Similarly, each paper investigated multiple features to train the ML algorithms, with no paper using the exact same combination of features.

Pituitary Adenoma, All Types

Two papers looked at all pituitary adenomas, regardless of type. They both looked at a specific outcome postsurgically, namely CSF leak and hyponatremia. Staartjes et al investigated the utility of ML models at predicting CSF leak postsurgically after TSS resection of pituitary adenomas.[11] They examined cases in 154 patients, out of which 45 had CSF leaks documented. They divided their patients into training (70%), validation (15%), and testing (15%) groups. A deep NN-based prediction model they constructed using the training set was able to identify 88% of patients in the testing set accurately, with AUC of 0.84. Age, prior surgery, and high suprasellar Hardy grade were found to be highly positively predictive and clear gross resectability was found to be negatively predictive via a polarized correlation plot constructed by the researchers. In contrast, Voglis et al investigated the prediction of hyponatremia in postoperative patients using four ML algorithms.[13] Using data from 207 patients, they divided the data into training (75%) and testing (25%) sets and were able to obtain an AUC of 0.843, 81.8% sensitivity, and 77.5% specificity with the boosted generalized linear machine (GLMBoost). The algorithm identified preoperative serum prolactin to be the most important feature, followed by preoperative insulin-like growth factor 1 (IGF-1), body mass index (BMI), and preoperative serum sodium level. In both models, the degree of classification accuracy was found to be adequate in predicting the risk of

Table 1 Characteristics of each study Included in our analysis

Author	No. of patients	Source of data	Type of data	ML algorithms used	Outcome	Disease
Dai et al, 2020[22]	306	SI	Categorical	LR, Adaboost, GBDT, XGBoost, CatBoost, RF, RFE	Delayed remission	Acromegaly secondary to pituitary adenoma
Fan et al, 2021a[19]	201	SI	Categorical	LR, GBDT, AdaBoost, XGBoost, and CatBoost	Delayed remission	Cushing's secondary to pituitary adenoma
Fan et al, 2021b[21]	354	SI	Categorical	MLP, LR, RF, GBDT, AdaBoost, XGBoost, GaussianNB, DT, PNN, DeepFM	Recurrence	Cushing's secondary to pituitary adenoma
Fan et al, 2020[12]	668	SI	Categorical	LR, RF, Adaboost, GBDT, XGBoost, LGAM	Remission	Acromegaly secondary to pituitary adenoma
Hollon et al, 2018[14]	400	SI	Categorical	NB, SVM, RF, LR-EN regularization	Early outcomes	Pituitary Adenoma, all types
Liu et al, 2019[20]	354	SI	Categorical	DT, RF, LR, NB, GBDT, AdaBoost, XGBoost	Recurrence	Cushing's secondary to pituitary adenoma
Machado et al, 2020[16]	27	SI	2D/ 3D Imaging	KNN, RF, LR, SVM, and MLP	Recurrence	Nonfunctional pituitary adenoma
Qiao et al, 2021[23]	833	MI	Categorical	Penalized LR, SVM, GBM, NN, Ensemble	Early remission	Acromegaly secondary to pituitary adenoma
Shahrestani et al, 2021[15]	348	SI	Categorical	Multilayered NN	Recurrence, progression, hormonal nonremission	Functional Pituitary Adenoma
Staartjes et al, 2020[11]	154	SI	Categorical	Deep NN Based Prediction Model	CSF leak	Pituitary Adenoma, All Types
Voglis et al, 2020[13]	207	SI	Categorical	GLMBoost, GLM, RF, NB,	Hyponatremia	Functional Pituitary Adenoma
Zhang et al, 2021[17]	1045	SI	Categorical	GBDT, RF, AdaBoost, XGBoost, Stacking, LR, NB, DT, MLP	Remission	Cushing's secondary to pituitary adenoma
Zoli et al, 2020[18]	151	SI	Categorical	SVM, RF, NN	Gross total resection, remission, long-term remission	Cushing's secondary to pituitary adenoma

Abbreviations: AdaBoost, adaptive boosting; CatBoost, categorical boosting; CSF, cerebrospinal fluid; DeepFM, factorization machine neural network; DT, decision tree; GBDT, gradient boosting decision tree; GBM, gradient boosted machine; GLM, generalized linear machine; GLMBoost, boosted generalized linear machine; LGAM, logistic generalized additive model; LR, logistic regression; MI, multiinstitutional trial; ML, machine learning; MLP, multilayer perceptron; NB, naïve bayes; RF, random forest; RFE, recursive feature elimination; SI, single Institutional trial; XGBoost, extreme gradient boost.

postoperative complications with preoperative data elements.

Functional Pituitary Adenoma

Two studies evaluated additional functional outcomes for patients with functional pituitary adenomas. Hollon et al investigated 400 consecutive pituitary surgeries for the prediction of postoperative outcomes such as functional DI, hyponatremia, length of stay in the hospital, and death using four ML algorithms.[14] They divided the data into training (75%) and testing (25%) sets. Using logistic regression with elastic net they obtained a predictive accuracy of 87% and AUC of 0.87. The SVM model they used identified six features as important: lowest perioperative sodium, age, BMI, highest perioperative sodium, Cushing's disease, and male sex. Shahrestani et al investigated 348 patients, 81 of whom had defined suboptimal outcomes, including remnant tumor, nonimprovement of preoperative visual deficit, and transient DI.[15] They divided their population into training (60%), validation (20%), and testing (20%) sets. Using a multivariate analysis, they identified features to be used in their multilayered NN, which achieved an overall accuracy of 87.1%, sensitivity of 89.5%, and an AUC of 0.917.

Nonfunctional Pituitary Adenoma

Machado et al investigated the utility of radiomics and convolutional NNs in predicting recurrence of nonfunctional pituitary adenomas after surgical resection, in 54 patients.[16] They used five ML algorithms, with RF achieving the highest AUC and accuracy of 0.962 and 92.3%, respectively, demonstrating the efficacy of three-dimensional (3D) radiomics.

Cushing's Disease Secondary to Functional Pituitary Adenoma

Our search found a total of five papers investigating the use of ML in outcomes prediction for TSS treatment of Cushing's disease caused by pituitary adenoma. The papers investigated prediction of remission, delayed remission, and recurrence.

Prediction of Remission in Cushing's Disease

Zhang et al investigated prediction of immediate remission postsurgically, defined as morning serum cortisol lower than 5 µg/dL or 24-hour urinary free cortisol (UFC) lower than 20 µg/dL postsurgically.[17] Their analysis included 1,045 participants, with a total of 766 exhibiting immediate remission. The data was split into training (80%) and testing (20%) sets. In total, nine ML models were used, and they investigated 11 features. The stacking ensemble algorithm was found to be the most effective, with an AUC of 0.743, a sensitivity of 80.4%, and a specificity of 58.9%. They found that four features were ideal to prevent overfitting, with the most accurate model being constructed from invasion of the cavernous sinus on preoperative magnetic resonance imaging (MRI), followed by tumor size, initial operation (as opposed to reoperation), and preoperative ACTH.

Zoli et al investigated rates of both short- and long-term remission.[18] They defined short-term remission as resolution of hypersecretion 1 to 6 months after surgery. Their study used 151 patient cases, with 88.1% exhibiting immediate remission after surgery. They divided their data into training (80%) and testing (20%) sets. In total they used 7 algorithms and assessed a total of 24 features. For remission, the top algorithm used was SVM with an AUC of 1.0, and a sensitivity and specificity of 100%. The model identified age and female sex, tumor visualization at preoperative MRI, size less than 10 mm, low Hardy–Wilson grade, histological confirmation of ACTH adenoma, pre-/postoperative hypopituitarism, and presurgical medical treatment as important positive prognostic values. Knosp grade, cavernous sinus invasion, and persistent ACTH hypersecretion were negative prognostic features.

Prediction of Delayed Remission in Cushing's Disease

Zoli et al also investigated prediction of long-term remission,[18] defined as long-term control of hypersecretion after surgery with additional medical treatment, such as repeated surgery, radiation, or continued medical treatment. Their analysis included 151 patients, with 13.9% exhibiting remission after surgery and additional treatment. In total they used 7 algorithms and assessed a total of 24 features. The top algorithm used was a gradient boosted machine with an AUC of 0.783, a sensitivity of 95.7%, and a specificity of 37.5%. Features were ranked on an AUC-based individual variable importance model. With this method, younger age and female sex, tumor visualization at preoperative MRI, size less than 10 mm, low Hardy–Wilson grade, histological confirmation of ACTH adenoma, pre-/postoperative hypopituitarism, and presurgical medical treatment were identified as positive prognostic features. Knosp grade, cavernous sinus invasion, and persistent ACTH hypersecretion were negative prognostic features.

Fan et al investigated spontaneous delayed remission,[19] defined as achievement of remission later than 1 week postsurgery but within 1 year. In total, 201 patients were selected due to nonremission immediately postsurgery, and 88 patients achieved delayed remission. They divided their data into training (80%) and testing (20%) sets. In total, the paper assessed 5 algorithms and 18 features. Recursive feature elimination was used to determine the best features to use. Of the five algorithms, adaptive boosting (Adaboost) was most effective, with an AUC of 0.7619, a sensitivity of 70%, and a specificity of 66.67%.

All 18 features were used in the final model, and permutation importance and local interpretable model–agnostic explanation (LIME) were used to provide weights to each of the features. In addition, LIME was also used to create a graphical interpretation of the relative prognostic values of each of the features investigated to improve interpretability by physicians. The most important feature was preoperative 24-hour UFC, with a weight of 0.136, and postoperative immediate morning serum cortisol with a weight of 0.132. Other important features include age, BMI, and disease course, which is the length of time between first disease incidence and treatment.

Prediction of Recurrence in Cushing's Disease

Liu et al investigated recurrence, defined as immediate remission postsurgery with morning cortisol levels below

5 µg or 24-hour UFC below 20 g at 7 days follow-up. After this, recurrence must occur either clinically or biochemically.[20] Of the 354 patients surveyed, 13% had recurrence. They used a fivefold cross-validation for training and testing. The paper assessed 17 features with seven ML algorithms. The most effective algorithm was RF, with an AUC of 0.779. They found that the models performed best with eight features selected. The eight features that produced the most accurate model were age, postoperative morning ACTH nadir, postoperative morning serum cortisol nadir, preoperative morning ACTH level, disease course, preoperative serum cortisol level, preoperative 24-hour UFC level, and postoperative 24-hour UFC nadir.

Fan et al also investigated recurrence.[21] Of the 354 patients surveyed, 13% had recurrence. They divided their data into training (80%) and testing (20%) sets. The paper assessed 17 features with 10 ML algorithms. A Factorization Machine Neural Network (DeepFM) was found to be the most effective, with an AUC of 0.884 and the lowest log loss value, at 0.256. The top five features were ACTH level, age, postoperative morning serum cortisol nadir, disease course, and postoperative 24-hour UFC nadir level. They also used a LIME to provide weights and relative negative and positive prognostic values of each of the features as a visual display to enhance physician interpretability.

Acromegaly Secondary to Functional Pituitary Adenoma

Our search found a total of three papers investigating the use of ML in predicting the outcomes of TSS treatment for acromegaly secondary to gonadotropin-releasing hormone and GH-secreting pituitary adenoma. The papers investigated prediction of delayed remission and recurrence.

Dai et al investigated delayed remission in acromegaly after surgical treatment, defined as remission occurring after 6 months follow-up.[22] They had 306 patients, with 55 (17.97%) exhibiting delayed remission following surgery. The patients were randomly assigned to the training set and test set and were included in a study of six ML algorithms using 18 clinical features. They found that extreme gradient boost (XGBoost) was the most effective in predicting delayed remission with an AUC of 0.8349 and sensitivity of 0.8889. Using a LIME, the authors identified 6-month postoperative IGF-1 and nadir growth hormone (nGH) were the most predictive of delayed remission.

Fan et al investigated remission, defined as "3 months after TSS, either nadir GH $<$ 0.4 ng/ mL after OGTT [oral glucose tolerance test] or GH $<$1.0 ng/mL in a random sample that is associated with a normal IGF-1 level (age and gender matched)"[12]. Their analysis used 668 patient cases with 349 (52.2%) exhibiting remission with good response to TSS. Patients were randomly divided into a training set including 534 patients (80%) and a test set including 134 (20%). Twelve preoperative features were studied, and six supervised ML algorithms were trained and gradient-boosting decision tree (GBDT) was shown to be most effective in predicting postsurgical outcome, with AUC of 0.8555, sensitivity of 85.25%, and specificity of 84.83%. The authors noted that the GBDT algorithm was *inexplicable*, meaning that the features calculate the outputs without clinical reason, and therefore, despite increased predictive value, may be less clinically valuable, especially when explaining results to patients. Using a classifier-specific feature evaluator, individual features were ranked on importance, with GH value and Knosp grade being the most important features.

Qiao et al investigated recurrence of acromegaly, defined as "off-medication GH levels (nadir GH $<$ 0.4 µg/L during an oral glucose tolerance test, and/or random GH $<$ 1.0 µg/L) or normalized IGF-1 ($<$ 1) at 6-month follow-up after surgery."[23] The authors used 833 patient cases with 434 (52.1%) expressing endocrine remission at 6 months after surgery as the training set. The algorithms were prospectively validated using 151 additional patients. Partial and full models were constructed with 15 preoperative and 20 perioperative features, respectively. The partial model that had the best results was an ensemble of penalized logistic regression, SVM, gradient boost machine, and NN, with AUC of 0.803, sensitivity of 90.3%, and specificity of 53.1%. The full model found to have the best results was a gradient boost machine, with AUC of 0.888, sensitivity of 90.5%, and specificity of 69.6%. Using Shapley additive explanations, the authors identified postoperative day 1 GH level, total resection, and Knosp grade as the most important features.

Discussion

The use of ML for outcome prediction in TSS is a recent innovation in the field of neurosurgery. Outcomes for TSS currently rely on population statistics and focus less on the outcomes expected for individuals despite heterogeneity within the population at risk for pituitary adenomas.[14] Since 2018, several studies on TSS have assessed the predictability of patient outcomes based on preoperative, intraoperative, and postoperative outcomes (►**Table 1**). Some show promise in outcomes prediction on an individual basis with algorithms approaching AUC of 1 as seen in Zoli et al, indicating perfect accuracy in predicting the outcomes for a given patient population.[18]

According to a 2015 study, the outcomes of TSS between 2008 and 2011 found postoperative complication rates that included 12.5% of patients experiencing central DI, 11.4% of patients experiencing electrolyte abnormalities, 8.1% of patients experiencing some level of neurological deficit, 4.2% of patients experiencing cranial nerve II or III deficits, and 0.4% of patients experiencing mortality.[24] Prior to surgery, neurosurgeons currently do not have methods to predict outcomes of surgery, especially specific outcomes such as delayed remission, recurrence, or CSF leaks.

Outcomes prediction in TSS would greatly benefit the field for a variety of reasons. First, predictions of unfavorable outcomes prior to surgery would help the surgeon and patient determine the best next steps in care of a pituitary adenoma. Having the means of predicting patient outcomes using features such as patient demographics, tumor characteristics, and surgery timing would be ideal for decision making, particularly for patients with high-risk adenomas.

Additionally, the outcomes predicted by these algorithms could assist in the long-term planning, care, and workup for a patient, providing additional information such as expected costs, insurance coverage, new surgical protocols, and additional assurance of favorable outcomes.

Because several variables associated with pituitary adenomas may be characterized by a combination of linear and nonlinear functions, the problem of preoperative outcomes prediction appears to be an excellent application for ML. For diagnostic purposes, ML algorithms will have to be highly effective in predicting outcomes, especially surgical outcomes such as central DI, electrolyte changes, neurological deficits, visual changes, mortality, and more, to provide proper planning for the future for patients. AUC is generally used as a way to compare ML algorithm accuracy by describing what percentage of the outcomes are typically correctly predicted by the algorithm. An AUC of 0.7 to 0.8 is usually considered acceptable for diagnosis, and 0.9 and above is considered excellent.[25] All of the papers in our study were able to surpass the 0.7 AUC benchmark needed, and thus would be useful in clinical practice (►**Fig. 2**).

Using ML to predict an individual's expected outcomes after TSS would be preferred to population statistics because the population of patients with pituitary adenomas requiring surgical consult is extremely diverse, and predicting outcomes based on features carefully selected for the algorithms and indications would be better applied to individual cases. In determining the ideal model for a given pituitary indication, many algorithms have been tested on retrospective patient data, and more research is currently underway.

Algorithms and Algorithm Selection

Algorithm selection is vital to successfully creating ML models. We found that most papers used multiple ML algorithms and compared the outputs of each to find the most effective. Though there was much heterogeneity among the papers, over half used LR and RF, with Adaboost, XGBoost, and GBDT used by eight, six, and six studies, respectively. No papers used unsupervised learning methods to determine outcomes. This is unsurprising, as the outcomes of all the patients were already known. In general, unsupervised algorithms are useful for categorizing uncategorized data, analyzing images or visualizing data, and work best with large datasets. They have shown promise in both the fields of dermatology and pathology.[26,27] Deep learning methods in particular work well for large, raw datasets.[28] The studies investigated in our analysis used relatively small, labeled datasets, which makes supervised learning the ideal option.

Some consistent trends emerged from our data showing that, by far, ensemble methods were most often top performers in the studies investigated, with nine studies identifying them as most predictive. Ensemble methods allow for

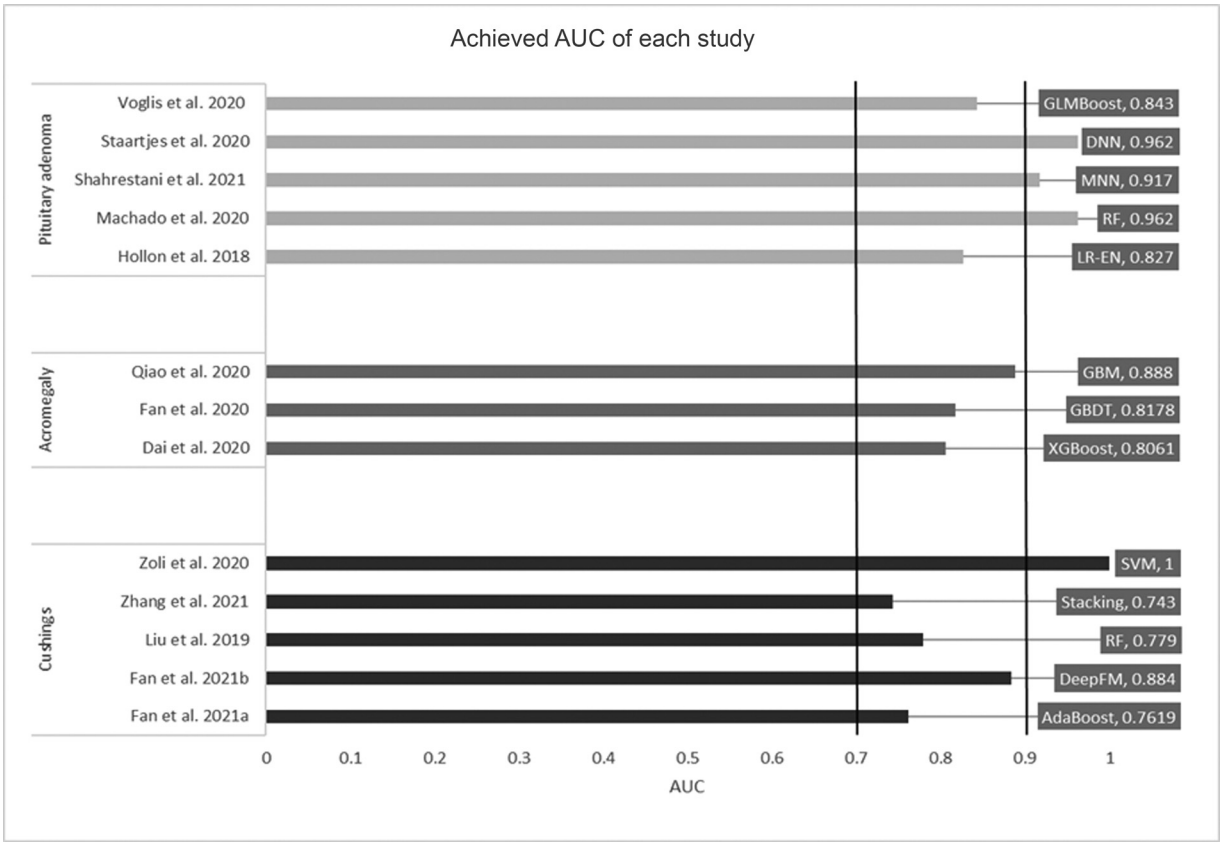

Fig. 2 Achieved area under the curve (AUC) of top performing algorithm in each study, with algorithm used, compared with the 0.7 and 0.9 AUC benchmarks of being considered clinically useful or highly predictive, respectively. AdaBoost, adaptive boosting; DNN, deep neural networks; GBDT, gradient boosting decision tree; GBM, gradient boosted machine; boosted machine; GLMBoost, boosted generalized linear machine; LR-EN, logistic regression model with elastic net; logistic regression; MNN, multilayered neural network; RF, random forest; XGBoost, extreme gradient boost.

multiple different ML algorithms or multiple instances of the same ML algorithm to be combined into a single algorithm allowing for better predictive performance than could be obtained from one source alone.[8] They have been used in the prediction of diabetes and cardiovascular disease,[29] prediction of disease course in multiple sclerosis,[30] prediction of fetal macrosomia,[31] mortality prediction,[32] and the diagnosis of breast cancer.[33] Although ensemble models are powerful, their use of multiple individual algorithms and then further compilation of their results can cause these algorithms to be resource intensive.[34] Clinicians and researchers looking to use these algorithms will have to take this into account if they wish to use them in practice or research. Multiple ensemble methods exist, such as bagging, boosting, or stacking methods. Bagging (or bootstrap aggregating) refers to a specific way of creating small random samples derived from a larger training set and feeding those smaller randomized samples to multiple learning algorithms. One such example of this is RF, which is a bagging algorithm that uses the averaged outputs of multiple DTs to produce an aggregated result that is more accurate than the sum of its individual parts.[34] RF was used by 77% of the papers we investigated. It proved useful in a wide variety of applications, including prediction using 3D radiomics[16] and prediction of recurrence in Cushing's disease.[20]

Boosted algorithms were also highly effective in our study. Boosting refers to a method where multiple weaker ML models are trained on the same training set one after another, with each subsequent model focusing on the mistakes the previous model made. A multitude of boosted algorithms were used throughout the studies, with AdaBoost, XGBoost, and GBDT being the most common. Adaboost uses single split DTs as weak learners and evaluates the efficacy of the individual DTs using an exponential loss function. Gradient boosting is similar but is not limited to using exponential loss as its method of grading individual DTs.[35] XGBoost is an open-source gradient boosting software that focuses on execution speed and accuracy.[36] Overall, boosting was found to be particularly effective in the prediction of remission in acromegaly,[12,23] with two of the papers identifying boosted algorithms as their top algorithm. Stacking is when the outputs of multiple ML algorithms are used in parallel and compiled into a single result via some other algorithm. Usually, linear regression is used in stacking to compile the results of multiple algorithms. Stacking proved to be useful in the prediction of early remission of both Cushing's[17] and acromegaly.[23]

NNs are ML algorithms that are an older, well-known technology in the medical field, having been used in research for cancer detection and diagnosis for nearly 20 years, with huge leaps being made in the complexity and predictive ability of NNs during that time.[37] Loosely inspired by the neurons in a human brain, NNs utilize inputs that stimulate neurons, which then go on to stimulate other neurons. In a NN, the neurons are arranged in layers, with each neuron in a layer communicating with certain neurons in the next layer. Once an input is plugged into the algorithm, each neuron will analyze its input and will decide whether to send a signal into the neurons in the layer below. Each layer then undergoes a similar process with its individual neurons until it reaches the final layer, where the output could be something like a yes or no, or a specific category.[38] NN's have been proven to be useful in the setting of high dimensional data, such as imaging from radiomics and histology.[28] However, the papers included in this study show that they can also perform well in categorical settings. They were found to be highly effective in multiple fields, notably recurrence of Cushing's, where a factorization based NN was used,[21] as well as the prediction of CSF leaks, recurrence, progression and hormonal nonremission in pituitary adenoma.[15]

Although the majority of high-performing ML algorithms in our study were ensemble or NN algorithms, the studies we looked at used a multitude of other algorithms as well, with some even outperforming the ensemble or NNs to which they were compared with. SVMs were found to be highly accurate. They function by plotting all outcomes in some multidimensional space and finding the best divider to separate the outcomes. In a 2D plane, that divider can be a line. In a 3D space, that divider would usually be a plane. Subsequent data can then be plotted on the multidimensional space and classified by where they fall relative to that line.[39] This algorithm was found to be effective in both predicting early outcomes of pituitary surgery and 3D radiomics. SVMs were also identified as most effective by Zoli et al for the prediction of remission of Cushing's disease.[18] Logistic regression was the most tested algorithm after RF, despite it not being considered a proper ML algorithm. Rather it is considered a statistical model. Christodoulou et al found that there was no demonstrable difference in accuracy between LR and ML models.[40] In addition, it is simple and not resource intensive to use, making it highly accessible.[41] In our study, LR performed well, often outperforming some of the ML algorithms when tested against them. Hollon et al found that LR, when combined with an elastic net, could outperform all other algorithms including SVM and RF in the prediction of postsurgical outcomes after TSS.[14] Another notable advantage it has is that it is explicable, meaning that its reasoning behind its output is easily understandable by humans. Its relative accuracy, ease of use, and interpretability justify its inclusion in the majority of ML studies.

A major problem with most ML algorithms is that they are inexplicable, which reduces the trust that clinicians have in the algorithm's judgement. Inexplicability can be decomposed into two concepts: inscrutability and nonintuitiveness. Inscrutability is related to the ability of ML algorithms to uncover relationships in data that are subtle and can result in models that depend on the contributions of multiple factors in ways that are exceedingly complex and impossible for people to follow. Nonintuitiveness is defined as the inability for statistical relationships between features and outcomes to be understood. The relationships identified by ML algorithms may not appear to have any intuitive explanation, despite having a sound statistical basis.[42] Some researchers have focused on making the inexplicable explainable, while others have insisted that high stakes decisions must be made using explicable functions. However,

nearly every paper identified inexplicable algorithms as most effective. Inexplicable algorithms include all ensemble algorithms, SVM and NNs. Interpretable algorithms include LR, GLM, LGAM, NB, and DTs. Multiple papers tackled this problem, using various techniques to determine the importance of certain features to a certain algorithm. One method that is used to explain such ML algorithms is LIME, which was used in Fan et al 2021a,[19] Fan et al 2021b,[21] and Dai et al 2020.[22] LIME is able to display the weight of each feature used by the ML algorithm in its final prediction. Qiao et al used a similar method called Shapley additive explanations, which assigns an importance value to each feature used in a prediction.[23] Staartjes et al used a polarity correlation plot to explain the outputs of its NN-based ML model.[11] These papers demonstrate that the field is beginning to move past just algorithmic accuracy; translating these technologies into clinical practice requires that clinicians trust and understand the decisions made by these algorithms.

Feature Selection

In our study, we also found that certain trends appeared in the features that were selected identified as important. Many of the features that were found to be important by the ML algorithms were also proven to have associations with postoperative outcome in the literature via traditional statistical methods. By far, biochemical measures such as serum concentrations of hormones were the most often deemed as important (►**Fig. 3**). Biochemical measures were also the most commonly tested features, likely because surveillance of biochemical markers is a mainstay of treatment, making the data readily available. Similarly, preoperative features were also commonly investigated and commonly identified as important (►**Fig. 4**). This is likely because biochemical testing, imaging, and patient demographic data are usually collected preoperatively. Intraoperative features are limited to intraoperative histology, categorical determinations by the surgeon, and other intraoperative events. Postoperative imaging and biochemical tests are also done postsurgically, but as the pathology is already diagnosed and treated, usually less data are collected after the surgery. Other important features included patient characteristics such as sex and age and preoperative imaging results, such as evidence of cavernous sinus invasion or Knosp grade. In addition, the papers noted that pituitary tumors are unique in that they are often considered functioning, releasing hormones that can result in serious systemic illnesses, further modifying preoperative risk. In other tumors, location and histology play an important role in prognosis, but for pituitary masses, tumor morphology, location, patient features, and other comorbidities can make patient selection and outcome prediction extremely complex. This illustrates the need for powerful models to assist in clinical decision making.

Biochemical features have long been used to assess risk of negative outcomes in patients after pituitary surgery. In particular, DI, an important feature reported by Shahrestani et al, is a known postoperative indicator of damage to the pituitary gland, with increased damage resulting in more severe DI. Poorly marginated or aggressive tumors tend to result in more damage when they are removed and thus are associated with poorer outcomes.[43] Moreover, risk of recurrence or lack of remission of functional tumors postsurgically has also been consistently linked to increased levels of hormone after surgery.[44–48] Though less research has been done into the association of preoperative hormone levels with postoperative outcomes, higher preoperative levels of IGF-1 have been shown to be predictive of nonremission postsurgically in acromegaly,[49] and lower remission rates of Cushing's disease postoperatively were associated with higher preoperative levels of cortisol.[50] Nearly all the studies we investigated rank pre- and postoperative hormone levels as important features, with postoperative hormone levels being more commonly reported as significant. Interestingly, multiple papers in our study also noted hypopituitarism as a positive predictive feature for remission,[15,18] despite DI also being a sign of poor postoperative outcome.

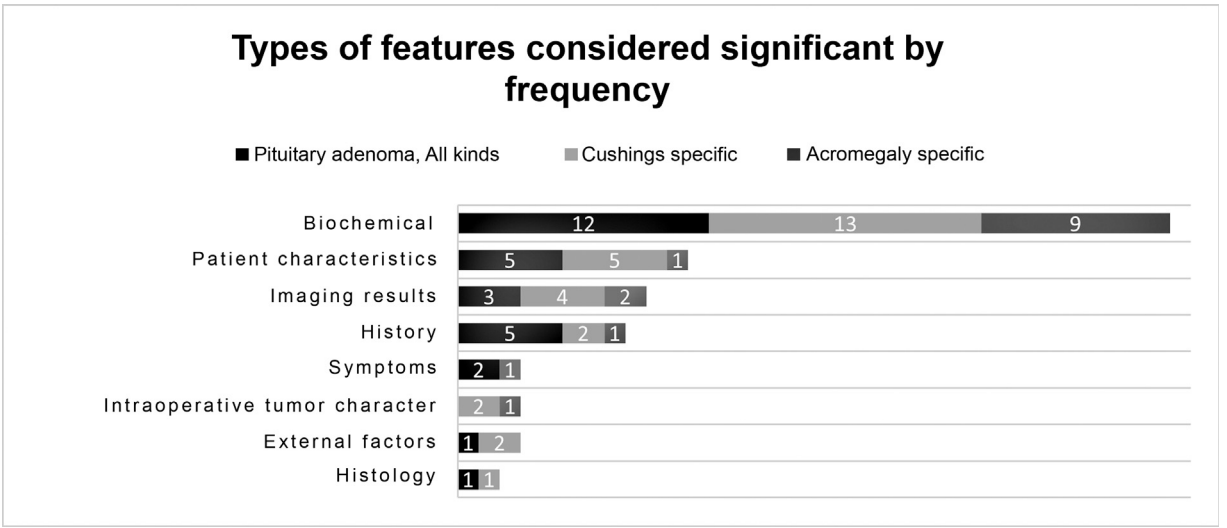

Fig. 3 Number of times a feature was mentioned by a study as important to their machine learning model, by type. See ►**Supplementary Material** for list of individual features.

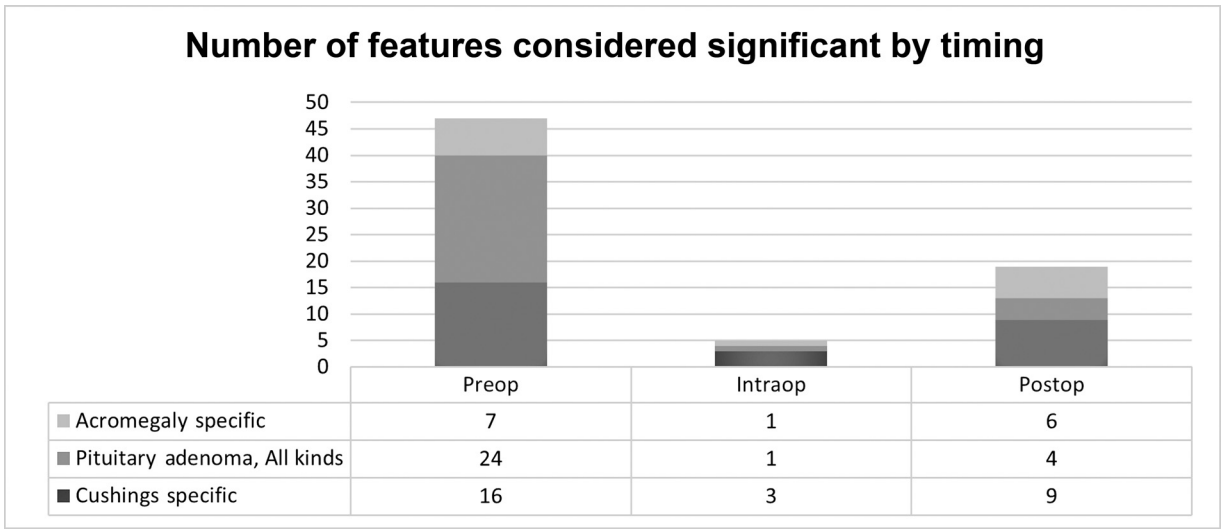

Fig. 4 Timing of features relative to surgery, by type of pathology.

Patient characteristics such as age and sex were also commonly cited as important features in outcomes prediction, with age being identified as the most significant. Younger age has been correlated with decreased rates of remission and increased rates of recurrence in both Cushing's[50] and acromegaly.[51] BMI was also commonly investigated in the papers we included in our study. It is a preoperative feature that is known to have a positive correlation with postsurgical CSF leaks[52] and a negative correlation with postsurgical hyponatremia.[53] Despite being tested in nearly every study in this review, it was only identified as an important feature in two studies looking at the specific outcomes of CSF leak[11] and postsurgical hyponatremia.[13]

Other features that were sometimes considered important include preoperative imaging and history of previous surgery or diagnosis. Preoperative imaging is always performed prior to surgery, as an invasion of the cavernous sinus makes pituitary tumors not fully resectable, which can lead to nonremission, recurrence, and other poor outcomes postsurgically.[51,54] Preoperative images such as MRI can be analyzed to estimate Knosp grade or Hardy–Wilson grade which are both helpful in determining probability of cavernous sinus invasion. Invasion of cavernous sinus on preoperative MRI was found to be the single most important feature in the prediction of immediate remission of Cushing's disease,[17] and Knosp grade was found to be highly important in the prediction of remission of acromegaly.[12,23]

Limitations

The main limitation with this study was the limited quantitative analysis possible given the heterogeneity of the studies analyzed. In each study, different patient features were collected from the medical record or prospective data collection process to predict outcomes. Because each study utilized different variables to predict outcomes, a comparison across the different algorithms was not possible, as it became difficult to parse the difference between the strength of the algorithm compared with the strength of the variables utilized. For this reason, statistical analysis comparing different algorithms or features was not performed.

Another associated limitation of this study stems from the weaknesses of the individual studies included. The studies considered were regional cohort studies analyzing data from primarily single centers. Given this, heterogeneity of the collection approaches and the patient populations are important to consider. To overcome this heterogeneity, large multiinstitutional studies ought to be considered for training algorithms. This would allow a fair assessment of the performance of algorithms that may not be robust on small datasets but may be more accurately trained on less heterogeneous datasets constructed.

Conclusion

ML has the potential to be used to predict postsurgical outcomes in multiple applications of TSS. The studies were highly heterogeneous in their definitions of outcomes, the features used to train the ML algorithms, and the ML algorithms used. Ensemble algorithms and NNs were found to be highly effective in the development of ML models for outcomes prediction of TSS. Features that have support in literature are biochemical in nature, or include patient characteristics tend to be useful features for ML programs to use. Once this technology is transferred into the hands of clinicians, it is certain to decrease complications rates and help guide clinicians in clinical and patient-oriented decision making.

Conflict of Interest
None declared.

Authors' Contributions
D.Y. conceptualized and designed the work. M.J., E.S., C.T., and D.Y. did systematic review. A.N., A.S., and D.Y. did data extraction, did data analysis and interpretation. D.Y., E.S., M.J., A.N., and A.S. were involved in drafting

the article. W.M., A.N., A.S., E.S., C.T., and L.V. did critical revision of the article. W.M., D.Y. gave final approval of the version to be published.

References

1 Lake MG, Krook LS, Cruz SV. Pituitary adenomas: an overview. Am Fam Physician 2013;88(05):319–327
2 Zubair A, Das JM. Transsphenoidal hypophysectomy. In: *StatPearls*. StatPearls Publishing; 2021. Accessed September 6, 2022 at: http://www.ncbi.nlm.nih.gov/books/NBK556142/
3 Halvorsen H, Ramm-Pettersen J, Josefsen R, et al. Surgical complications after transsphenoidal microscopic and endoscopic surgery for pituitary adenoma: a consecutive series of 506 procedures. Acta Neurochir (Wien) 2014;156(03): 441–449
4 Charalampaki P, Ayyad A, Kockro RA, Perneczky A. Surgical complications after endoscopic transsphenoidal pituitary surgery. J Clin Neurosci 2009;16(06):786–789
5 Araujo-Castro M, Pascual-Corrales E, Martínez San Millan J, et al. Multidisciplinary protocol of preoperative and surgical management of patients with pituitary tumors candidates to pituitary surgery. Ann Endocrinol (Paris) 2021;82(01):20–29
6 Lobatto DJ, de Vries F, Zamanipoor Najafabadi AH, et al. Preoperative risk factors for postoperative complications in endoscopic pituitary surgery: a systematic review. Pituitary 2018;21(01): 84–97
7 Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019;380(14):1347–1358
8 Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol 2019;19 (01):64
9 Berry MW, Mohamed A, Yap BW, Eds. Supervised and Unsupervised Learning for Data Science. Springer International Publishing; 2020
10 Soldozy S, Farzad F, Young S, et al. Pituitary tumors in the computational era: exploring novel approaches to diagnosis, and outcome prediction with machine learning. World Neurosurg 2021;146:315–321.e1
11 Staartjes VE, Zattra CM, Akeret K, et al. Neural network-based identification of patients at high risk for intraoperative cerebrospinal fluid leaks in endoscopic pituitary surgery. J Neurosurg 2019;133(02):1–7
12 Fan Y, Li Y, Li Y, et al. Development and assessment of machine learning algorithms for predicting remission after transsphenoidal surgery among patients with acromegaly. Endocrine 2020;67 (02):412–422
13 Voglis S, van Niftrik CHB, Staartjes VE, et al. Feasibility of machine learning based predictive modelling of postoperative hyponatremia after pituitary surgery. Pituitary 2020;23(05):543–551
14 Hollon TC, Parikh A, Pandian B, et al. A machine learning approach to predict early outcomes after pituitary adenoma surgery. Neurosurg Focus 2018;45(05):E8
15 Shahrestani S, Cardinal T, Micko A, et al. Neural network modeling for prediction of recurrence, progression, and hormonal nonremission in patients following resection of functional pituitary adenomas. Pituitary 2021;24(04):523–529
16 Machado LF, Elias PCL, Moreira AC, Dos Santos AC, Murta Junior LO. MRI radiomics for the prediction of recurrence in patients with clinically non-functioning pituitary macroadenomas. Comput Biol Med 2020;124:103966
17 Zhang W, Sun M, Fan Y, et al. Machine learning in preoperative prediction of postoperative immediate remission of histology-positive Cushing's disease. Front Endocrinol (Lausanne) 2021; 12:635795
18 Zoli M, Staartjes VE, Guaraldi F, et al. Machine learning-based prediction of outcomes of the endoscopic endonasal approach in Cushing disease: is the future coming? Neurosurg Focus 2020;48 (06):E5
19 Fan Y, Li Y, Bao X, et al. Development of machine learning models for predicting postoperative delayed remission in patients with Cushing's disease. J Clin Endocrinol Metab 2021a106(01): e217–e231
20 Liu Y, Liu X, Hong X, et al. Prediction of recurrence after transsphenoidal surgery for Cushing's disease: The use of machine learning algorithms. Neuroendocrinology 2019;108(03): 201–210
21 Fan Y, Li D, Liu Y, Feng M, Chen Q, Wang R. Toward better prediction of recurrence for Cushing's disease: a factorization-machine based neural approach. Int J Mach Learn Cybern 2021b12 (03):625–633
22 Dai C, Fan Y, Li Y, et al. Development and interpretation of multiple machine learning models for predicting postoperative delayed remission of acromegaly patients during long-term follow-up. Front Endocrinol (Lausanne) 2020;11:643
23 Qiao N, Shen M, He W, et al. Machine learning in predicting early remission in patients after surgical treatment of acromegaly: a multicenter study. Pituitary 2021;24(01):53–61
24 Villwock JA, Villwock MR, Goyal P, Deshaies EM. Current trends in surgical approach and outcomes following pituitary tumor resection. Laryngoscope 2015;125(06):1307–1312
25 Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010;5(09):1315–1316
26 Acharya P, Mathur M. Artificial intelligence in dermatology: the 'unsupervised' learning. Br J Dermatol 2020;182(06):1507–1508
27 Roohi A, Faust K, Djuric U, Diamandis P. Unsupervised machine learning in pathology: the next frontier. Surg Pathol Clin 2020;13 (02):349–358
28 Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med 2019;25(01):24–29
29 Dinh A, Miertschin S, Young A, Mohanty SD. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Med Inform Decis Mak 2019;19 (01):211
30 Zhao Y, Wang T, Bove R, et al; SUMMIT Investigators. Ensemble learning predicts multiple sclerosis disease course in the SUMMIT study. NPJ Digit Med 2020;3:135
31 Ye S, Zhang H, Shi F, Guo J, Wang S, Zhang B. Ensemble learning to improve the prediction of fetal macrosomia and large-for-gestational age. J Clin Med 2020;9(02):E380
32 Bergquist T, Schaffter T, Yan Y, et al. Evaluation of crowdsourced mortality prediction models as a framework for assessing AI in medicine. 2021
33 Hosni M, Abnane I, Idri A, Carrillo de Gea JM, Fernández Alemán JL. Reviewing ensemble classification methods in breast cancer. Comput Methods Programs Biomed 2019;177:89–112
34 Doupe P, Faghmous J, Basu S. Machine learning for health services researchers. Value Health 2019;22(07):808–815
35 Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013;7:21
36 XGBoost Documentation—xgboost 1.6.0-dev documentation. Accessed September 6, 2022, at: https://xgboost.readthedocs.io/en/latest/
37 Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Inform 2007;2:59–77
38 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(01):44–56
39 Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2(04):230–243
40 Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12–22
41 Boulesteix A-L, Schmid M. Machine learning versus statistical modeling. Biom J 2014;56(04):588–593

42 Selbst AD, Barocas S. The intuitive appeal of explainable machines. SSRN Journal 2018
43 Woods C, Thompson CJ. Risk of diabetes insipidus after pituitary surgery. Expert Rev Endocrinol Metab 2008;3(01):23–27
44 Ironside N, Chatain G, Asuzu D, et al. Earlier post-operative hypocortisolemia may predict durable remission from Cushing's disease. Eur J Endocrinol 2018;178(03):255–263
45 Bansal P, Lila A, Goroshi M, et al. Duration of post-operative hypocortisolism predicts sustained remission after pituitary surgery for Cushing's disease. Endocr Connect 2017;6(08):625–636
46 Wang YY, Waqar M, Abou-Zeid A, et al. Value of early post-operative growth hormone testing in predicting long-term remission and residual disease after transsphenoidal surgery for acromegaly. Neuroendocrinology 2022;112(04):345–357
47 Butenschoen VM, von Werder A, Bette S, et al. Transsphenoidal pituitary adenoma resection: do early post-operative cortisol levels predict permanent long-term hypocortisolism? Neurosurg Rev 2022;45(02):1353–1362
48 Lu C, Lin XT, Yang DT, Liu YC, He WT, Zhong XL. Pre- and post-operative hypothalamic-pituitary-thyroidal axis function in patients with prolactinoma, growth hormone tumour and ACTH tumour. Chin Med J (Engl) 1989;102(04):306–312
49 Agrawal N, Ioachimescu AG. Prognostic factors of biochemical remission after transsphenoidal surgery for acromegaly: a structured review. Pituitary 2020;23(05):582–594
50 Shirvani M, Motiei-Langroudi R, Sadeghian H. Outcome of microscopic transsphenoidal surgery in Cushing disease: a case series of 96 patients. World Neurosurg 2016;87:170–175
51 Bourdelot A, Coste J, Hazebroucq V, et al. Clinical, hormonal and magnetic resonance imaging (MRI) predictors of transsphenoidal surgery outcome in acromegaly. Eur J Endocrinol 2004;150(06):763–771
52 Patel PN, Stafford AM, Patrinely JR, et al. Risk factors for intra-operative and postoperative cerebrospinal fluid leaks in endoscopic transsphenoidal sellar surgery. Otolaryngol Head Neck Surg 2018;158(05):952–960
53 Hussain NS, Piper M, Ludlam WG, Ludlam WH, Fuller CJ, Mayberg MR. Delayed postoperative hyponatremia after transsphenoidal surgery: prevalence and associated factors. J Neurosurg 2013;119(06):1453–1460
54 Braileanu M, Hu R, Hoch MJ, et al. Pre-operative MRI predictors of hormonal remission status post pituitary adenoma resection. Clin Imaging 2019;55:29–34