Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Artificial Intelligence Applications in Pituitary Adenoma Management: A Scoping Review

Hector Julio Piñera-Castro

hectorpinera18100@gmail.com

Hospital Clínico Quirúrgico Hermanos Ameijeiras https://orcid.org/0000-0002-2491-489X

Marlon Manuel Ortiz-Machín

Hospital Clínico Quirúrgico Hermanos Ameijeiras https://orcid.org/0000-0001-9483-7247

Christian Borges-García

Universidad de Ciencias Médicas de Pinar del Río https://orcid.org/0000-0002-1430-8706

Omar López-Arbolay

Hospital Clínico Quirúrgico Hermanos Ameijeiras https://orcid.org/0000-0001-7948-4287

Systematic Review

Keywords: Artificial intelligence, Pituitary adenoma, Pituitary neuroendocrine tumor, Machine learning, Deep learning, Radiomics, Computer vision, Natural language processing, Large language models, Neurosurgery

Posted Date: June 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9944032/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Background

Artificial intelligence has emerged as a promising tool in pituitary adenoma management, yet its applications remain dispersed across clinical and computational disciplines. This scoping review aimed to map and analyze the current and emerging applications of artificial intelligence in the management of patients with pituitary adenomas.

Methods

Following PRISMA-ScR guidelines, Scopus and PubMed/MEDLINE and were searched for original studies published between 2015 and 2026. Studies evaluating artificial intelligence applications in pituitary adenoma management were included. Data were extracted and synthesized according to clinical application domains.

Results

Seventy-four studies met the inclusion criteria. Artificial intelligence applications were identified across seven domains: diagnosis and tumor classification, image segmentation and reconstruction, prediction of invasion and anatomical characteristics, prediction of therapeutic response and recurrence, prediction of postoperative complications and outcomes, operative workflow analysis, and natural language processing. Machine learning, deep learning, radiomics, and computer vision were the predominant methodologies. Diagnostic models demonstrated high performance in differentiating pituitary adenomas from other sellar lesions and in identifying hormonal subtypes. Radiomics- and machine learning-based approaches showed promising accuracy for predicting cavernous sinus invasion, tumor consistency, extent of resection, and response to medical or surgical treatment. Deep learning techniques enabled automated image segmentation and image quality enhancement, while emerging applications included prediction of endocrine remission, postoperative cerebrospinal fluid leakage, and the need for additional interventions. Natural language processing and large language models were explored for information extraction, patient education, and clinical decision support. Despite these advances, most studies were retrospective, single-center investigations with heterogeneous datasets, limited external validation, and inconsistent reporting standards.

Conclusions

Artificial intelligence has been applied across nearly all stages of pituitary adenoma management and demonstrates substantial potential to support diagnosis, prognostication, treatment planning, and postoperative care. However, translation into routine clinical practice remains limited by methodological

heterogeneity and insufficient external validation, highlighting the need for prospective multicenter studies and standardized evaluation frameworks.

INTRODUCTION

Pituitary adenomas are the most common neoplasms of the sellar region and remain a substantial clinical challenge in neurosurgical practice(1–3). Their burden derives not only from prevalence but also from the marked biological and functional heterogeneity that characterizes these lesions.

International epidemiological data indicate that clinically apparent pituitary tumours affect approximately 1 in every 1,100 individuals, with a prevalence of 76 to 116 cases per 100,000 inhabitants and an annual incidence of 3.9 to 7.4 cases per 100,000 persons(2). Although autopsy studies show that incidental pituitary lesions are relatively frequent, the vast majority are microadenomas that remain asymptomatic throughout life(1, 2). Among clinically apparent cases, nearly half present as macroadenomas, which are associated with greater diagnostic, surgical, and therapeutic complexity(2, 3).

The clinical expression of pituitary adenomas is highly variable. Non-functioning adenomas account for approximately 30% of clinically evident cases, whereas functioning tumours include prolactinomas, somatotropinomas, and corticotropinomas(2). Macroadenomas may produce mass effect on critical neurovascular structures, particularly the optic chiasm, the pituitary gland and stalk, the cavernous sinus and internal carotid arteries (ICAs), leading to visual field defects, headache, and hypopituitarism(1–3). Long-term morbidity is also clinically relevant: secondary hypopituitarism is associated with increased mortality, and active Cushing's disease confers higher risks of cardiovascular events, venous thromboembolism, and sepsis, even after biochemical remission(2). In this context, transsphenoidal surgery and postoperative surveillance remain demanding, and current assessment continues to rely heavily on qualitative scales with substantial interobserver variability(4, 5).

In recent years, artificial intelligence (AI) and machine learning (ML) have emerged as promising tools for the quantitative analysis of complex medical data(4, 6). Their potential value in sellar pathology lies in their ability to address limitations inherent to conventional human assessment. Computer vision (CV) and radiomics allow the extraction of high-dimensional quantitative features from imaging studies that may capture information not readily appreciable on visual inspection(7–9). Similarly, natural language processing (NLP) and large language models (LLMs) offer the possibility of synthesizing unstructured clinical information from medical records, with potential applications in risk identification and long-term follow-up(6).

Despite this promise, the literature on AI in pituitary adenomas remains fragmented across neurosurgical, endocrine, radiological, and computational disciplines(4, 7–9). The field has developed in a heterogeneous and largely isolated manner, making it difficult to define the overall state of the evidence. A comprehensive synthesis is therefore needed to clarify how AI has been applied across the

management of patients with pituitary adenomas and to identify the most relevant areas of current and emerging investigation.

This review addressed the following question: In patients with pituitary adenomas, what current and emerging applications of artificial intelligence have been reported across the continuum of clinical management, including diagnosis, treatment, prognostication, and follow-up? Accordingly, the objective of this scoping review was to systematically map and analyze the applications of artificial intelligence in the management of patients with pituitary adenomas reported in the scientific literature.

METHODS

This scoping review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA),(10) complemented with the PRISMA for Scoping Reviews (PRISMA-ScR)(11) extension.

No review protocol was prospectively registered. Nevertheless, methodological rigor was promoted through the *a priori* definition of the review objectives, eligibility criteria, information sources, search strategy, data charting framework, and thematic synthesis approach before study selection commenced.

Eligibility criteria

Studies were considered eligible if they fulfilled all of the following criteria: 1) investigated AI-based methodologies applied to pituitary adenomas; 2) included clinical, radiological, pathological, molecular, surgical, or prognostic applications; 3) employed AI-related methodologies, including but not limited to ML, deep learning (DL), convolutional neural networks (CNNs), transformer-based architectures, radiomics, CV, NLP, reinforcement learning, or LLMs; 4) were original peer-reviewed investigations; 5) were published in English or Spanish; 6) were published between 2015 and 2026.

Studies were excluded if they reported only conventional statistical analyses without an AI component, focused primarily on non-pituitary sellar lesions without providing separate data for pituitary adenomas, addressed basic science or omics-related questions without applying AI methodologies to a clinically relevant objective, or lacked sufficient methodological or outcome data for extraction.

Information sources and search

A comprehensive electronic literature search was conducted in Scopus and PubMed/MEDLINE. The final search was executed on May 1, 2026. To maximize retrieval sensitivity, both controlled vocabulary terms and free-text keywords were incorporated. Search terms encompassed concepts related to pituitary adenomas and AI. Database-specific indexing terms and Boolean operators were adapted for each platform. The complete electronic search strategies for both databases, including exact Boolean syntax, field tags, and filters, are provided in the **Supplementary Material** to ensure full reproducibility of the search process.

Selection of sources of evidence

All records retrieved from the database searches were exported into reference management software, and duplicate citations were removed before screening. Study selection was performed in two sequential phases: title and abstract screening, and full-text eligibility assessment. Screening was conducted independently by two reviewers using predefined eligibility criteria. Discrepancies between reviewers were resolved through discussion and consensus. In cases of persistent disagreement, adjudication was performed by a senior reviewer. Reasons for exclusion during full-text assessment were documented systematically to ensure methodological transparency.

Data charting process

Data extraction was performed using a standardized and pilot-tested charting form specifically developed for this review. Prior to formal extraction, the charting framework was calibrated across reviewers using a subset of representative studies to ensure consistency and reduce interpretative variability. Two reviewers independently charted data from each included study. Extracted information was cross-verified, and disagreements were resolved by consensus review. Where study characteristics or methodological details were unclear, the corresponding publication was reviewed in full detail to minimize extraction inaccuracies.

Data items

The following variables were extracted from each included study: first author and year of publication, country and institutional setting, study design, sample size and patient population, clinical domain/application, imaging modality or source of clinical data, AI architecture or computational methodology, validation strategy, comparator methods, primary and secondary clinical outcomes, performance metrics, explainability or interpretability techniques, main conclusions and reported limitations. When studies evaluated multiple algorithms or endpoints, the clinically primary model and the best-validated outcome were prioritized during synthesis.

Critical appraisal of individual sources of evidence

A formal risk-of-bias assessment or methodological quality scoring was not performed, as the purpose of this review was to map the extent, characteristics, and thematic distribution of the available evidence rather than to estimate pooled effects or establish comparative effectiveness.

Nevertheless, factors likely to influence clinical applicability and translational readiness were systematically considered during data synthesis, including study design (prospective *versus* retrospective), sample size, single-center *versus* multicenter setting, presence of external validation, class imbalance, model interpretability, reproducibility, and potential sources of overfitting. These considerations were incorporated qualitatively into the narrative synthesis to contextualize the maturity and potential clinical utility of the identified AI applications.

Synthesis of results

Due to substantial heterogeneity across studies regarding AI architectures, imaging modalities, datasets, endpoints, validation methods, and reporting standards, quantitative meta-analysis was deemed inappropriate.

Studies were grouped into seven thematic domains through an iterative and consensus-based categorization process. Domains were defined according to the primary clinical objective addressed by the AI application rather than the specific computational methodology employed. This approach facilitated the mapping of evidence across clinically relevant stages of pituitary adenoma management while minimizing overlap between categories.

Within each domain, studies were comparatively analyzed according to methodological approach, validation strategy, interpretability, and potential translational relevance to clinical practice. Particular emphasis was placed on clinically deployable models, multicenter validation, explainable AI frameworks, and emerging trends likely to influence future precision medicine paradigms in pituitary disease management.

RESULTS

The database search identified 1,125 records, including 1,057 from Scopus and 68 from PubMed/MEDLINE. After removal of 29 duplicate records, 1,096 unique citations underwent title and abstract screening. A total of 953 records were excluded at this stage. Full-text retrieval was sought through institutional library subscriptions, open-access repositories, publisher websites, and, when necessary, author contact. Despite these efforts, 38/143 articles could not be obtained. Consequently, 105 full-text articles were assessed for eligibility. Thirty-one studies were excluded after full-text review, most commonly because they reported statistical or clinical prediction models without an AI component (n = 13), focused on non-pituitary sellar lesions (n = 10), or investigated basic science or omics-related topics without implementing AI methodologies (n = 8). Ultimately, 74 studies were included in the final synthesis (Fig. 1).

The included studies encompassed a broad spectrum of clinical applications spanning diagnostic imaging, tumor classification, image segmentation, prediction of invasion and surgical complexity, treatment response modeling, postoperative outcome prediction, operative workflow analysis, and NLP applications. ML, DL, radiomics, CV, and LLMs represented the most frequently employed methodological approaches.

1 Diagnosis and tumor classification based on imaging

The integration of AI into the management of pituitary adenomas, now increasingly referred to as pituitary neuroendocrine tumors (PitNETs), marks a transition from qualitative visual assessment to a

quantitative, data-driven paradigm (Table 1).

Table 1
Diagnosis and tumor classification based on imaging

Study	AI architecture/methodology	Domain	Performance metric
Feng *et al.* (2022)	CNN	Sellar floor invasion	96% accuracy
Qian *et al.* (2020)	CNN	General pituitary adenoma diagnosis	96.97% accuracy
Li *et al.* (2021)	Faster R-CNN	Microadenoma detection	96.5% accuracy
Liu *et al.* (2021)	ML / discrete wavelet transform	Classification prediction	~ 90% accuracy
Li *et al.* (2021)	3D CNN + Attention	Functioning pituitary adenoma vs. non-functioning pituitary adenoma	AUC-ROC: 0.847
Gargya and Jain (2024)	TL (Inception V3)	Tumor classification	96% accuracy
Demirel *et al.* (2025)	ML (XGBoost/LR)	Cushing's disease vs. ectopic ACTH syndrome (BIPSS alternative)	86% accuracy
Liu *et al.* (2025)	RF	Posterior pituitary tumor vs. other sellar lesions	AUC-ROC: 0.818
Wang *et al.* (2021)	ANN / Radiomics	Cystic pituitary adenoma vs. Rathke cleft cyst	AUC: 0.848
Peng *et al.* (2020)	SVM / Radiomics	Immunohistochemistry lineage (Pit-1, SF-1, Tpit)	89% balanced accuracy
Wu *et al.* (2025)	CNN with a Swin Transformer architecture	Lineage from hematoxylin and eosin slides	95.9% accuracy
Wang *et al.* (2026)	LASSO / SVM- Recursive Feature Elimination	PCDI for invasive PitNETs	Mean AUC: 0.89, high calibration
Baysal *et al.* (2022)	ANN / Radiomics	Hormone secretion profile	95% accuracy (prolactin)
Zhang *et al.* (2018)	SVM / Nomogram	Null cell adenoma vs. other non-functioning pituitary adenoma subtypes	High C-index
Gu *et al.* (2026)	Radiomics nomogram	p53 overexpression	AUC: 0.77
Micko *et al.* (2021)	Optical coherence tomography + CNN	Intraoperative biopsy diagnosis	100% sensitivity

Study	AI architecture/methodology	Domain	Performance metric
Yang *et al.* (2026)	RF / SVM	METTL14/Autophagy mechanism	RF AUC: 0.735, SVM AUC: 0.703

1.1 Diagnosis of pituitary adenoma

The primary challenge in initial diagnosis involves the objective identification of small lesions and the assessment of tumor extent. CNNs have emerged as the dominant architecture for these tasks due to their ability to process complex spatial hierarchies in magnetic resonance imaging (MRI) data. Early models, such as the one proposed by Qian *et al.*,(12) utilized a majority voting strategy across sagittal, coronal, and horizontal planes to achieve a diagnostic accuracy of 96.97%. While effective, such models often require substantial preprocessing, including manual segmentation and data augmentation.

Methodological evolution is evident in the work of Gargya and Jain,(13) who utilized transfer learning (TL) techniques, specifically fine-tuning pre-trained models like Inception V3 and ResNet50 (Residual network with 50 layers) on public datasets to achieve 96% accuracy. However, a critical limitation of TL-based studies using public datasets is their potential lack of clinical generalizability compared to models trained on large, multicenter hospital cohorts.

The detection of pituitary microadenomas represents a specific diagnostic hurdle. Li *et al.*(14) addressed this by developing the PM-CAD (Pituitary Microadenoma-Computer-Aided Diagnosis) system, which utilized a Faster R-CNN (Region-based CNN) architecture. This study is distinguished by its extensive multicenter validation and its ability to outperform junior neuroradiologists, highlighting the role of AI as a screening tool to reduce the diagnostic workload. Furthermore, AI has been applied to evaluate tumor behavior, such as sellar floor invasion. Feng *et al.*(15) demonstrated that a CNN could identify sellar floor destruction with high objectivity, providing a more reliable assessment than traditional clinical grading systems like the Hardy classification.

Further classification based on biological activity is addressed by Li *et al.*,(16) who developed a 3D CNN combined with an attention mechanism to distinguish between functioning and non-functioning pituitary adenomas. This model is notable for its interpretability, as attention maps allowed clinicians to visualize which tumor regions contributed most to the model's decision-making. This level of transparency is contrasted by the more abstract feature extraction methods used by Liu *et al.*,(17) who employed discrete wavelet transforms and stacking algorithms to decouple data and improve predictive stability, though with less immediate visual interpretability for the clinician.

1.2 Differential diagnosis between sellar lesions

The sellar region hosts a variety of lesions that frequently mimic pituitary adenomas on conventional MRI. AI models in this subdomain aim to resolve this "masquerade" through ML and radiomics—the extraction of high-dimensional quantitative features from medical images. A major clinical milestone

was reported by Demir *et al.*,(18) who developed an ML algorithm to differentiate Cushing's disease from ectopic adrenocorticotropic hormone (ACTH) syndrome. This is significant because it offers a non-invasive alternative to the bilateral inferior petrosal sinus sampling (BIPSS), a procedure that is technically demanding and carries procedural risks. By integrating biochemical parameters with MRI data and using SHapley Additive exPlanations (SHAP) for model interpretability, this study bridges the gap between complex AI logic and clinical trust.

Similarly, distinguishing between cystic pituitary adenoma and Rathke cleft cyst is often problematic. Wang *et al.*(19) addressed this by constructing an artificial neural network (ANN) that integrated both radiomic signatures and semantic imaging features (such as wall thickness and intracystic nodules). Their model outperformed experienced radiologists, suggesting that radiomics can detect subtle heterogeneity in cyst fluid and walls that is imperceptible to the human eye. In the context of rarer tumors, Liu *et al.*(20) utilized a random forest (RF) classifier based on contrast-enhanced T1-weighted radiomics to differentiate TTF (thyroid transcription factor)-1-positive posterior pituitary tumors from PitNETs and craniopharyngiomas, achieving a precision of 92.9%. These studies collectively demonstrate that AI can refine the differential diagnosis of sellar lesions, potentially reducing the need for invasive diagnostic surgeries.

1.3 Histological, molecular, or functional classification

Wu *et al.*(21) developed a semi-automated DL framework using a swin transformer architecture. This system predicts tumor lineage directly from hematoxylin and eosin stained slides, reducing the diagnostic workload by approximately 69% while maintaining 95.9% accuracy.

The integration of the 2017 and 2022 World Health Organization classifications has shifted the focus toward lineage-based diagnosis using transcription factors. Preoperative molecular prediction via imaging is an expanding frontier. Peng *et al.*(22) showed that a support vector machine (SVM) model could classify tumor lineages including SF-1 (steroidogenic factor 1), Tpit (T-box pituitary transcription factor), and Pit-1 (pituitary transcription factor 1) from T2-weighted MRI radiomics with high balanced accuracy. Similarly, Baysal *et al.*(23) used an ANN to predict the specific hormone secretion profile [e.g., prolactinoma *versus* growth hormone (GH)-secreting] with 95% accuracy. Zhang *et al.*(24) extended this to subtyping non-functioning pituitary adenomas, specifically identifying null cell adenomas—which are often more aggressive—using T1-weighted radiomics signatures and nomograms. To address tumor aggressiveness directly, Gu *et al.*(25) developed a model to predict p53 overexpression, a key marker of tumor proliferation, providing clinicians with a preoperative tool to identify patients who may require more aggressive initial management.

At the intersection of molecular biology and AI, Wang *et al.*(26) constructed a programmed cell death-associated index (PCDI) using ML to integrate multi-omics data, identifying 11 prognostic genes that characterize invasive PitNETs. This is complemented by Yang *et al.*,(27) who combined single-cell RNA (ribonucleic acid) sequencing with ML to reveal that METTL14 (methyltransferase like 14)-mediated autophagy drives tumor progression via the PI3K/AKT (phosphoinositide 3-kinase/ protein kinase B)

signaling pathway. Finally, Micko *et al.*(28) introduced an intraoperative application by using optical coherence tomography paired with CNNs to identify adenoma tissue *ex vivo* during surgery, reaching a sensitivity of up to 100%. These findings suggest a future where AI not only diagnoses the tumor but also provides a comprehensive molecular and mechanistic profile before or during the intervention.

2 Image segmentation, reconstruction, and automated quantification

This domain encompasses the transition from manual, subjective radiological assessment to objective, high-precision automated workflows (Table 2).

Table 2
Image segmentation, reconstruction, and automated quantification

Study	AI architecture/ methodology	Performance metric	Contribution
Wang *et al.* (2021)	GSU-Net (8-class)	Dice coefficient: 0.940	Simultaneous segmentation of tumor and seven surrounding structures (ICA, cavernous sinus, etc.).
Jiang *et al.* (2022)	Improved U-Net (Inception-dense)	Dice coefficient: 0.888	Uses cross-layer connections to handle blurred boundaries and noise.
Zhang *et al.* (2024)	MSR-Net	Dice coefficient: 0.893	Integrates dual attention (channel/spatial) to mitigate background interference.
Zhang *et al.* (2023)	PDC-Net	Dice coefficient: 0.885	Employs dilated convolutions to extract features at multiple scales without deepening the network.
Shu *et al.* (2021)	nnU-Net (3D semantic)	DSC: >0.8 (> 1,000mm³)	Demonstrates that 3D automated segmentation is highly reliable for large tumors but fails for microadenomas.
Da Mutten *et al.* (2024)	CNN Ensemble	r: 0.85 (Pre-op) / r: 0.22 (Post-op)	Automated volume assessment is objective preoperatively but struggles with postoperative blood/granulation tissue.
Park *et al.* (2023)	1-mm DLR MRI	AUC: 0.91–0.92 (CSI)	1-mm DLR provides superior delineation of the pituitary stalk compared to routine 3-mm MRI.
Kim *et al.* (2022)	1-mm DLR MRI	AUC: 0.79–0.80 (CSI)	Better identifies depth of invasion and degree of ICA contact using modified Knosp classification.
Lee *et al.* (2021)	1-mm DLR MRI (AIR Recon DL)	SNR: 1.25x increase	Detected new adenomas in 6.5% of cases that were invisible on routine 3-mm sequences.

2.1 Automated tumor segmentation

Automated segmentation of pituitary adenomas is critical for surgical planning and longitudinal follow-up, yet it remains challenging due to blurred tumor margins and the proximity of delicate structures like the optic chiasm and ICAs.(29) Recent studies have primarily utilized variations of the U-Net architecture to address these complexities.

A comparative analysis of the methodologies reveals a trend toward integrating multi-scale features and attention mechanisms to enhance performance. Wang *et al.*(30) introduced the gated-shaped U-net (GSU-Net), which uniquely performs 8-class segmentation of the sellar region (including background, pituitary adenomas, normal pituitary, ICAs, bilateral cavernous sinuses, and optic chiasm). Their model achieved a high Dice coefficient of 0.940 for tumor segmentation in key slices, demonstrating that multi-class structural context improves localized accuracy.

Other researchers have focused on architectural refinements to capture richer contextual information. Jiang *et al.*(29) utilized cross-layer connections and inception-dense blocks to enlarge the effective receptive field, achieving a Dice coefficient of 0.888 despite a small patient sample. Similarly, Zhang *et al.*(31) developed multi-scale residual network (MSR-Net), which incorporates multi-scale residuals and dual (channel and spatial) attention mechanisms to mitigate interference from complex backgrounds, reaching a Dice coefficient of 0.893. Another study by Zhang *et al.*(32) proposed parallel dilated convolutional network (PDC-Net), employing parallel dilated convolutions to extract multi-level features without increasing computational load, resulting in a Dice coefficient of 0.885.

From a clinical utility perspective, Shu *et al.*(33) evaluated the nnU-Net (no-new-Net) framework for 3D semantic segmentation. Their findings highlight a significant performance gap based on tumor size: the models performed excellently (Dice coefficient > 0.8] for tumors larger than 1,000 mm^3 but were unsatisfactory (Dice coefficient < 0.5) for smaller lesions. This suggests that while AI is highly effective for macroadenomas, manual or semi-automated intervention remains necessary for microadenomas in current clinical practice.

2.2 Automated volumetry and quantification

While segmentation defines the boundaries, automated volumetry translates these shapes into clinically actionable data. Da Mutten *et al.*(34) developed an automated pipeline using an ensemble of CNNs to assess tumor volume and the extent of resection (EOR).

Their results underscore a major discrepancy between preoperative and postoperative performance. While the preoperative automated volumes correlated strongly with manual measurements ($r = 0.85$), the correlation for postoperative residual tumor volume was significantly lower ($r = 0.22$). The model struggled to detect gross-total resection (GTR), achieving only 50% accuracy. This limitation is attributed to the presence of postoperative granulation tissue, blood products, and the small size of residual fragments, which "ground truth" labels often fail to distinguish consistently.

2.3 Image reconstruction and enhancement

Deep learning-based reconstruction (DLR) has emerged as a solution to the traditional trade-off between signal-to-noise ratio (SNR) and spatial resolution in thin-slice MRI.

Studies by Park *et al.*(35) and Kim *et al.*(36) demonstrated that 1-mm DLR MRI significantly outperforms routine 3-mm MRI in predicting cavernous sinus invasion (CSI). Park *et al.*(35) reported that DLR improved the area under the curve (AUC) for CSI diagnosis from 0.87 to 0.91–0.92. Furthermore, Kim *et al.*(36) found that 1-mm DLR MRI allowed for a more accurate estimation of the depth of invasion and the degree of ICA contact, which is vital for assessing surgical risk.

Beyond invasiveness, Lee *et al.*(37) found that 1-mm DLR MRI could detect new adenomas in 6.5% of cases that were entirely invisible on routine 3-mm MRI. The DLR software (e.g., AIR Recon DL) utilizes CNNs to remove noise and truncation artifacts, resulting in a 1.21 to 1.25-fold increase in signal-to-noise ratio. Notably, this technological enhancement was most beneficial for inexperienced readers, who showed a stronger preference for DLR images when delineating the normal pituitary axis compared to experts.

3 Prediction of invasion, tumor extension, and anatomical characteristics

The preoperative assessment of anatomical features, specifically CSI and tumor consistency, is a critical determinant of surgical strategy, EOR, and long-term endocrinological remission(38). Traditional radiological classifications, such as the Knosp and Hardy scales, often fail to accurately predict invasion in intermediate grades and cannot objectively quantify physical tumor properties(39). AI, through radiomics and DL, has emerged to bridge this gap by extracting high-dimensional data that exceed human visual perception(40).

3.1 Cavernous sinus invasion

CSI is a major predictor of suboptimal surgical outcomes, yet its preoperative identification remains challenging due to the limited resolution of the medial wall of the cavernous sinus on standard MRI (Fig. 2).

A recurring theme in the literature is the critique of the Knosp grading system, particularly for grades 2 and 3, where surgical confirmation of invasion varies significantly. To address this, Fang *et al.*(38) implemented a CNN based on the ResNet50 architecture for an end-to-end diagnosis using T1 contrast-enhanced (CE-T1) images. Their model significantly outperformed the Knosp system, achieving an AUC of up to 0.98. The advantage of this DL approach is the reduction of interobserver bias inherent in manual segmentation.

Conversely, handcrafted radiomics remains prevalent for its interpretability. Niu *et al.*(41) and Chen *et al.*(40) utilized SVM and LightGBM (light gradient-boosting machine) algorithms, respectively, to predict

invasion. Niu *et al.*(41) demonstrated that combining a radiomics signature with clinico-radiological factors (e.g., periarterial enhancement) into a nomogram provides superior predictive value for Knosp 2–3 adenomas compared to using radiological markers alone. Chen *et al.*(40) further innovated by identifying the "pedunculated satellite tumor" as a novel imaging marker, which, when integrated with radiomics and clinical data, reached an AUC of 0.86.

The importance of the peritumoral environment was highlighted by Zhang *et al.*,(42) who utilized a "mask expansion" technique to analyze infiltration levels beyond the visible tumor margin. By incorporating coronal, sagittal, and axial views, they improved 3D feature representation, achieving an accuracy of 85% in their independent validation set.

3.2 Tumor consistency and resectability

Tumor consistency (soft *versus* fibrous) dictates the technical difficulty of transsphenoidal surgery. While soft tumors are easily suctioned, fibrous tumors often require extracapsular dissection or result in incomplete resection.

Cuocolo *et al.*(43) utilized handcrafted radiomics on T2-weighted MRI without contrast, demonstrating that texture analysis combined with an extra trees (ET) classifier can predict consistency with 93% accuracy. Their study advocates for the use of simple, non-invasive sequences to streamline the preoperative pipeline. In contrast, Yang *et al.*(39) emphasized the value of multiparametric MRI, integrating T1, T2, and CE-T1 sequences. Their comparative analysis of 2D *versus* 3D regions of interest (ROIs) revealed that 3D ROI models generally provide richer quantitative data, leading to better predictive performance (AUC 0.813 in validation). This multicenter study provides high robustness across different image acquisition protocols.

Beyond consistency, AI has been applied to predict the EOR. Patel *et al.*(44) developed a regression-based CNN specifically for giant pituitary adenomas. Unlike previous binary models (GTR *versus* non-GTR), this pilot study predicted EOR as a continuous percentage directly from raw DICOM® (Digital Imaging and Communications in Medicine) images. Their findings suggest that the coronal plane is the most informative axis for giant pituitary adenomas due to the vertical and horizontal growth trajectories of these lesions. This model serves as an objective tool for preoperative counseling and multidisciplinary care planning.

4 Prediction of therapeutic response, remission, recurrence, and tumor control

The shift toward precision medicine in the management of PitNETs is increasingly driven by the application of AI to predict complex clinical trajectories. This domain focuses on extracting predictive patterns from high-dimensional datasets that traditional statistical models fail to capture, ranging from biochemical and clinical parameters to sub-visual radiomic features and unstructured clinical narratives (Table 3).

Table 3
Prediction of therapeutic response, remission, recurrence, and tumor control

Study	AI architecture/methodology	Domain	Performance metric
Shinya *et al.* (2025)	LR / DT	Acromegaly intervention-free rate	AUC: 0.84
Zanier *et al.* (2022)	Generalized linear model / Gradient boosting machine	Acromegaly GTR	AUC: 0.75
Huber *et al.* (2022)	ML / Super Learner	Prolactinoma dopamine agonist dependency	AUC: 0.90
Park *et al.* (2021)	Ensemble ML	Prolactinoma dopamine agonist response	AUC: 0.81
Qiao *et al.* (2021)	ML	Acromegaly early remission	AUC: 0.81
Kocak *et al.* (2019)	k-NN	Acromegaly somatostatin analog response	AUC: 0.847
Sharifi *et al.* (2025)	Nomogram	Cushing's disease persistence/recurrence	AUC: 0.70
Zhang *et al.* (2024)	XGBoost	Cushing's disease delayed remission	AUC: 0.819
Zhang *et al.* (2021)	LR + NLP	Cushing's disease immediate remission	AUC: 0.793
Shinya *et al.* (2025)	DT	Cushing's disease survival (intervention-free survival)	91% accuracy
Fan *et al.* (2021)	AdaBoost	Cushing's disease delayed remission	AUC: 0.761
Fan *et al.* (2021)	DeepFM	Cushing's disease recurrence	AUC: 0.869
Liu *et al.* (2019)	RF	Cushing's disease recurrence	AUC: 0.781
Speckter *et al.* (2025)	SVM / LASSO	Gamma Knife radiosurgery volumetric response	AUC: 0.924
Shen *et al.* (2023)	LR	Non-functioning -PitNET regrowth	AUC: 0.929
Fan *et al.* (2019)	SVM	Invasive functional pituitary adenoma treatment response	AUC: 0.811

4.1 Acromegaly and prolactinomas

The management of GH-secreting and prolactin-secreting adenomas requires precise triage between medical and surgical interventions. In acromegaly, predicting long-term outcomes after endoscopic transsphenoidal surgery is a primary focus.

Shinya *et al.*(45) developed decision tree (DT) and RF models to predict the intervention-free rate, identifying that GTR and patient age are critical determinants of long-term control. This finding is partially mirrored by Zanier *et al.*,(46) who provided a crucial external validation using cohorts from Bologna and Zurich, achieving an AUC of 0.75 for GTR and 0.77 for intraoperative cerebrospinal fluid (CSF) leaks. However, Zanier *et al.*(46) noted that biochemical remission remains more difficult to predict (AUC 0.63), suggesting that surgical success and hormonal normalization are governed by different biological and technical factors.

Qiao *et al.*(47) further contributed to this area through multicenter modeling of early remission. A significant technical advancement was demonstrated by Kocak *et al.*(48), who utilized quantitative texture analysis on T2-weighted MRI to predict response to somatostatin analogues. Their k-nearest neighbors (k-NN) model achieved an AUC of 0.847, significantly outperforming qualitative visual evaluation (AUC 0.599), thereby proving that radiomic heterogeneity serves as a superior biomarker for chemosensitivity compared to conventional signal intensity.

Regarding prolactinomas, research addresses the clinical conflict between primary surgery and dopamine agonist therapy. Huber *et al.*(49) employed a stacked "super learner"—an ensemble of diverse algorithms—to predict long-term dopamine agonist dependency, finding that remission status at 30 days postoperatively was the most robust predictor (AUC 0.9, Matthews correlation coefficient 0.85). Conversely, Park *et al.*(50) addressed the challenge of medical resistance preoperatively using an ensemble ML model (including LightGBM and ET) on radiomic features. Their model achieved an AUC of 0.81, indicating that the radiomic phenotype can identify drug-resistant patients before initiating therapy, whereas Huber's model serves to stabilize follow-up strategies after surgical intervention.

4.2 Cushing's disease

Predicting outcomes in Cushing's disease is uniquely complicated by high recurrence rates and the phenomenon of delayed remission, where patients initially fail immediate remission but remit spontaneously without adjuvant therapy. Fan *et al.*(51) and Zhang *et al.*(52) targeted delayed remission specifically; while Fan *et al.*(51) utilized the AdaBoost algorithm (AUC 0.76) to identify preoperative 24-hour urine free cortisol and postoperative morning cortisol as key features, Zhang *et al.*(52) enhanced this approach by incorporating multicenter radiomics and automated DL segmentation, achieving a superior AUC of 0.819. A distinctive technical contribution is provided by Zhang *et al.*,(53) who NLP via word embedding to process unstructured electronic medical records (EMR). Their model (AUC 0.793) demonstrated that doctor-written clinical summaries contain high-order information—reflecting subjective disease severity—that biochemical data alone cannot capture.

For long-term surveillance, Liu *et al.*(54) established an RF model for recurrence prediction (AUC 0.781), whereas Fan *et al.*(55) reported a factorization-machine based neural network (DeepFM) model with an AUC of 0.869 for recurrence prediction in Cushing disease. The DeepFM model (AUC 0.869) automatically models both low- and high-order feature interactions, providing significantly higher discriminative power than traditional RF or logistic regression (LR). Shinya *et al.*(56) refined these survival assessments by using DT to predict intervention-free survival with 91% accuracy, identifying body mass index as a critical prognostic factor. Finally, Sharifi *et al.*(57) developed the CuPeR model, a dynamic online nomogram that bridges the gap between research and practice by including patients with a history of prior surgery or radiotherapy, thereby enhancing its real-world utility.

4.3 Tumor regrowth and control

Predicting the behavior of residual or invasive tumors is essential for long-term management strategies. Shen *et al.*(58) developed a clinical-radiomic model to predict the regrowth of residual non-functioning PitNETs. This study was the first to combine both pre- and postoperative radiomic features, achieving a high AUC of 0.929. For patients undergoing radiotherapy, Speckter *et al.*(59) pioneered the use of radiomics to predict the volumetric response to Gamma Knife radiosurgery. Their T1-weighted models (AUC 0.924) significantly outperformed standard clinicopathological parameters (AUC 0.846), allowing for individualized radiation planning based on inferred tumor biology. Lastly, Fan *et al.*(60) addressed invasive functional pituitary adenomas; by integrating a 7-feature radiomic signature with the Knosp grade in a SVM, they achieved an AUC of 0.811, providing a non-invasive method to identify patients who will likely require aggressive multimodal therapy.

5 Prediction of postoperative complications and outcomes

The integration of AI in the postoperative management of pituitary adenomas represents a paradigm shift from traditional clinical observation to data-driven proactive risk stratification. This domain utilizes diverse ML architectures to navigate the clinical heterogeneity inherent in sellar pathology, addressing complications ranging from electrolyte imbalances to permanent sensory deficits (Table 4).

Table 4
Prediction of postoperative complications and outcomes

Study	Complication	AI architecture/methodology	Performance metric	Key predictor(s)
Yin *et al.* (2026)	Early hyponatremia	LR	AUC: 0.85	Tumor size > 25mm, Knosp grade
Lin *et al.* (2025)	Late hyponatremia	XGBoost	AUC: 0.831	Stalk angle, day 2 sodium
Liu *et al.* (2025)	Hypernatremia	LASSO-LR	AUC: 0.894	Pituitary stalk involvement
Chen *et al.* (2024)	Arginine vasopressin deficiency (diabetes insipidus)	Random Forest	AUC: 0.96	Preoperative stalk length
Hussein *et al.* (2023)	Long-term endocrine	SVM	94% accuracy	1-year postoperative status
Chang *et al.* (2023)	CSF leak (end-to-end)	Deep neural network (U-Net)	84.6% accuracy	CE-T1 raw image features
Villalonga *et al.* (2022)	CSF leak (anatomical)	One-class SVM	90% specificity	Sellar barrier thickness
Qiao *et al.* (2023)	Venous thromboembolism	LDA	AUC: 0.869	Age > 65, chordoma pathology
Zhang *et al.* (2023)	Visual recovery	Delta-radiomics	AUC: 0.840	Microscopic stability of chiasm

5.1 Hydroelectrolytic and endocrine complications

Sodium regulation and the restoration of the hypothalamic-pituitary-adrenal axis are the primary focus of postoperative neuroendocrinology. Research in this subdomain is notably partitioned by the temporal onset of complications. Regarding hyponatremia, Yin *et al.*(61) focused on early postoperative hyponatremia, defined as occurring within the first three days post-surgery, identifying that a tumor diameter exceeding 25 mm and the ACTH-secreting subtype are potent independent risk factors. Their model achieved an area under the receiver operating characteristic curve (AUC-ROC) of 0.85, providing a simplified six-point scoring system for bedside use.

In contrast, Lin *et al.*(62) and Tan *et al.*(63) investigated delayed postoperative hyponatremia, which typically triggers unplanned readmissions. Lin *et al.*(62) demonstrated that the extreme gradient boosting (XGBoost) algorithm was superior to six other models (AUC 0.831), emphasizing that changes

in the pituitary stalk deviation angle and the length of the measurable stalk are critical morphological predictors. This reliance on "black-box" models was balanced by the use of SHAP to improve interpretability. Tan *et al.*(63) reached an even higher validation AUC of 0.959 using a nomogram that incorporated preoperative prolactin levels and suprasellar cistern height, suggesting that endocrine biomarkers may be as predictive as radiological ones for late-stage electrolyte failure.

While hyponatremia is more frequent, Liu *et al.*(64) addressed the lack of predictive tools for hypernatremia, which also significantly extends hospital stays. By utilizing a least absolute shrinkage and selection operator (LASSO) penalized LR, they identified pituitary stalk involvement and surgical duration as the primary drivers of postoperative high sodium levels (AUC 0.894). Closely related to these water-balance disturbances is arginine vasopressin deficiency, formerly known as diabetes insipidus. Chen *et al.*(65) compared multiple algorithms to predict transient arginine vasopressin deficiency, finding that RF models (AUC 0.96) outperformed support vector classification and LR. Their analysis confirmed that shorter preoperative pituitary stalks are more susceptible to intraoperative traction, leading to higher deficiency rates.

Predicting long-term endocrine outcomes remains a complex challenge due to the slow evolution of hormonal recovery. Fang *et al.*(66) conducted a multicenter study using only preoperative hormone levels as input, proving that AdaBoost (AUC 0.82) can effectively predict uniaxial hypofunction in male patients with non-functioning tumors. They notably excluded female patients to avoid the confounding variables of menstrual cycles, thereby increasing the internal validity of their data at the cost of broader generalizability. Finally, Hussein *et al.*(67) provided an unprecedented 15-year longitudinal analysis, demonstrating that SVM can predict panhypopituitarism with 94% accuracy. A critical insight from their comparison was that the endocrine status at one year post-surgery is the most robust predictor of the status at 10 years, suggesting a threshold for permanent deficiency.

5.2 Other surgical complications

The prediction of intraoperative CSF leaks is essential for determining the necessity of complex sellar reconstructions. Two distinct AI philosophies have emerged here. Villalonga *et al.*(68) utilized novelty detection algorithms, such as one-class SVM, to validate the anatomical concept of the "sellar barrier," finding a 90% specificity in cases where MRI features indicated a "weak" barrier. Conversely, Chang *et al.*(69) bypassed anatomical concepts by training a deep neural network exclusively on raw contrast-enhanced T1 images. Their model, based on a modified U-Net architecture, achieved an 84.6% accuracy in clinical tests, outperforming expert neurosurgeons by over 30%, which suggests that DL can capture sub-visual patterns of dural fragility that escape human experts.

Broader surgical outcomes were evaluated by Hollon *et al.*,(70) who built a model to predict a composite of poor early outcomes including major medical events, readmissions, or death. Their LR with elastic net (LR-EN) model achieved 87% accuracy, identifying lowest perioperative sodium and age as the most influential variables. This holistic approach is useful for general risk stratification but lacks the granularity of subtype-specific models. For instance, Qiao *et al.*(71) focused specifically on venous

thromboembolism, which is often overlooked in pituitary surgery. They integrated NLP to extract variables from clinical records, and their linear discriminant analysis (LDA) model identified chordoma pathology and high-volume CSF leakage as significant risks (AUC 0.869). Finally, Chen *et al.*(72) established a risk model for olfactory disorders, which affect up to 73% of transnasal patients. Using RF with LASSO selection, they found that surgery duration and the formation of blood crusts were key determinants of postoperative anosmia (AUC 0.846).

5.3 Visual outcomes

Visual recovery is the primary goal for many patients with macroadenomas, and AI-based radiomics provides a more granular assessment of axonal damage than routine MRI. Zhang *et al.*(73) first established that ML-based radiomics of the optic chiasm, specifically using SVM (AUC 0.824), could predict visual field recovery better than traditional parameters like tumor diameter or age. This was further refined by multiparametric MRI integration. Zhang *et al.*(74) combined T1, T2, and contrast-enhanced sequences, demonstrating that a multiparametric approach (AUC 0.851) significantly outperforms any single-sequence model, as different sequences capture unique pathophysiological aspects of nerve compression.

The most recent innovation in this field is delta-radiomics, which quantifies the relative change between preoperative and postoperative imaging. Zhang *et al.*(75) observed that while macroscopic morphological recovery of the chiasm does not strictly correlate with functional recovery, the stability of microscopic imaging patterns is a robust prognostic indicator. Their combined model, which integrated delta-radiomics with patient age, achieved an AUC of 0.840, suggesting that functional reversibility is a dynamic process best captured by longitudinal image analysis.

6 Artificial intelligence applied to operative workflow

The integration of AI into the surgical workflow of endoscopic transsphenoidal surgery for pituitary adenomas represents a paradigm shift from purely diagnostic applications to real-time intraoperative optimization and postoperative educational refinement. This domain focuses on the systematic deconstruction of surgical maneuvers to enhance safety, efficiency, and training.

6.1 Automated operative workflow analysis

Current research in automated workflow analysis has progressed from foundational phase recognition to the sophisticated assessment of surgical fluency and the development of AI-driven coaching platforms (Table 5).

Table 5
Applications in automated operative workflow analysis

Study	AI architecture/methodology	Domain	Performance metric
Khan *et al.* (2022)	CNN-RNN (touch surgery)	Phase/step recognition	91% phase accuracy, 76% step accuracy
Wong *et al.* (2024)	OpenCV (color-based)	Surgical interruptions and fluency	98.5% accuracy (collaborative), 74% time saving
Mao *et al.* (2025)	ConsisTNet (HRNet + ConvLSTM)	Spatio-temporal consistency	43.8% reduction in landmark distance error
Khan *et al.* (2024)	AI-assisted coaching	Performance and clinical outcomes	Improved mOSATS, reduced endocrine deficits

Khan *et al.*(76) pioneered the first neurosurgical application of automated workflow analysis using a combined CNN and recurrent neural network model. In an IDEAL (Idea, Development, Exploration, Assessment, Long-term study) stage 0 study, the model achieved high accuracy in recognizing surgical phases (91%) but demonstrated a performance gap when identifying more granular surgical steps (76%). This study highlighted significant intraoperative variability, with only 50% of procedures following a sequential 7-step numerical order, emphasizing the need for models capable of handling non-linear surgical trajectories.

Addressing the limitations of frame-by-frame analysis, Mao *et al.*(77) introduced ConsisTNet, a spatio-temporal model designed to reduce prediction "volatility". While earlier models often produced inconsistent results between consecutive frames, ConsisTNet utilizes ConvLSTM (convolutional long short-term memory) and a semi-supervised approach with CoTracker2 to ensure temporal continuity. Their results showed a significant improvement in anatomical localization consistency—improving sella segmentation by 4.56% and reducing landmark detection error by 43.86%. This highlights a methodological shift from simple classification to stable, real-time guidance.

In contrast to DL approaches, Wong *et al.*(78) utilized CV algorithm based on OpenCV to analyze surgical fluency by detecting interruptions (e.g., endoscope removal for cleaning). Their collaborative human-CV strategy achieved 98.5% accuracy and reduced video review time by 74%. Critically, this study quantified the learning curve, demonstrating that increased surgeon experience correlates with a significant reduction in both the number and duration of interruptions during the sellar phase.

Moving beyond analytics into clinical impact, Khan *et al.*(79) implemented a stage 1 proof-of-concept AI-assisted surgical coaching program. By using AI to index operative videos for targeted review, the program significantly improved surgical performance scores [modified Objective Structured Assessment of Technical Skills (mOSATS)] across all phases. Most importantly, this intervention was associated with a reduction in new postoperative anterior pituitary hormone deficits (from 48.8% to 41.2%), providing empirical evidence that AI-mediated feedback can directly improve patient outcomes.

6.2 Navigation and intraoperative support

AI applications in this subdomain aim to augment the surgeon's spatial orientation by providing advanced visualization techniques that traditional 2D endoscopy lacks.

Zanier *et al.*(80) explored the feasibility of monocular depth estimation using Meta AI's DINOv2 foundation model. This approach generates 3D-like depth maps from standard 2D video feeds without requiring specialized hardware. While the qualitative assessment by senior surgeons was promising, the study reported a mean Dice coefficient of 0.48, indicating that the model, while feasible, requires further fine-tuning on specific neurosurgical datasets to reach clinical reliability. Furthermore, real-time application faced latency issues (3–10 frames per second), suggesting that hardware optimization is necessary for seamless intraoperative integration.

7 Natural language processing, clinical records, and large language models

The integration of NLP and LLMs in the management of pituitary adenomas represents a significant shift from traditional structured data analysis toward the extraction and synthesis of knowledge from unstructured clinical narratives. This domain explores how AI can transform free-text EMR into computable data and how generative models perform as clinical decision-support tools.

7.1 Natural language processing on clinical records and data extraction

Recent research has prioritized the transition from manual data entry to automated clinical named entity recognition (NER) to handle the complexity and heterogeneity of pituitary adenomas clinical notes. Two primary studies illustrate this evolution through the construction of specialized corpora and the application of DL architectures.

Fang *et al.*(81) conducted the first study focused on clinical information extraction from Chinese neurosurgical EMRs. They compared several methodologies, including dictionary-based matching, conditional random fields (CRF), and DL models [BiLSTM (bidirectional long short-term memory)-CRF and BERT (bidirectional encoder representations from transformers)-BiLSTM-CRF]. Their results demonstrate that dictionary-based methods were the least effective (F1 strict < 20% for most entities) due to their inability to utilize context. In contrast, the BERT-BiLSTM-CRF model achieved the highest strict F1 value (91.27%), highlighting that transformer-based architectures significantly outperform traditional ML by capturing nuanced medical semantics. However, the authors noted that even advanced models struggle with entities separated by punctuation or complex qualifiers, suggesting that post-processing remains necessary.

Complementing this, Hu *et al.*(82) introduced the Text Corpus of Pituitary Adenomas (TCPA), a high-quality annotated dataset designed to characterize pituitary adenomas through a semi-automatic framework. Their study critiqued the limitations of general-purpose LLMs in specialized domains; they found that without fine-tuning, models like ChatGLM3-6B achieved an F1 score of 0 in identifying pituitary adenomas-specific entities. By employing low-rank adaptation (LoRA) fine-tuning with the TCPA, performance improved to an F1 of 0.8179. This underscores that while LLMs possess vast general knowledge, their clinical utility in niche fields like neuroendocrinology is entirely dependent on domain-specific fine-tuning and high-quality human-annotated data.

7.2 Evaluation of large language models and clinical chatbots

The evaluation of LLMs as supplementary tools for both clinicians and patients has revealed significant variability in reliability based on the complexity of the query.

Aliyeva *et al.*(83) provided a prospective comparative analysis of three state-of-the-art models: ChatGPT-5.0, Claude Opus 4.1, and Gemini 2.5 Flash. Using validated instruments [QAMAI (Quality Analysis of Medical Artificial Intelligence), AIPI (Artificial Intelligence Performance Instrument), and AISPE-Q (Artificial Intelligence Satisfaction and Performance Evaluation Questionnaire)], they found that Claude Opus 4.1 demonstrated superior informational quality (4.39 ± 0.66) and clinical reasoning depth compared to its peers. A significant finding was the strong correlation ($r = 0.71$) between objective information quality and clinical reasoning performance, suggesting that models with better data synthesis are more trusted by expert evaluators.

However, the "reliability gap" between general and professional information remains a major concern. Chen *et al.*(84) evaluated ChatGPT-4's performance on 256 pituitary adenomas-related questions and found that while it was highly reliable for general patient-oriented queries (95.0%), its reliability dropped significantly for professional clinical questions (78.6%). Surgical management questions, in particular, yielded the highest rate of completely incorrect responses (19.4%), highlighting the danger of using LLMs for autonomous surgical decision-making.

Similarly, Şenoymak *et al.*(85) focused on prolactinomas and identified a critical failure point regarding pregnancy-related queries, where ChatGPT-3.5 provided inaccurate or misleading information (e.g., suggesting routine prolactin monitoring during pregnancy, which contradicts established guidelines). These findings collectively argue that while LLMs can enhance health literacy, they lack the "human-like" contextual understanding required for sensitive or high-risk clinical scenarios.

DISCUSSION

Across the literature, AI has progressively shifted pituitary adenoma research from descriptive imaging toward quantitative, prediction-oriented, and workflow-integrated decision support. The most mature applications were concentrated in diagnostic imaging and outcome prediction, where CNNs, radiomics,

and ensemble ML models repeatedly achieved strong discrimination for tumor detection, invasion, consistency, treatment response, and postoperative complications. A recurring pattern across domains was the ability of AI to extract sub-visual information from MRI, histology, operative video, or clinical text that is not readily captured by conventional assessment. At the same time, the evidence base remains dominated by retrospective, single-center studies with limited external validation, small sample sizes, and heterogeneous reporting, which constrains reproducibility and clinical generalizability.

Each thematic domain showed a distinct profile of strengths and limitations. In diagnosis and classification, AI was especially valuable for differentiating sellar lesions, predicting lineage, and supporting non-invasive inference of biological behavior, thereby reducing reliance on technically demanding or postoperative methods. In segmentation, reconstruction, and volumetry, AI performed well for larger tumors and preoperative planning, but performance deteriorated in microadenomas and postoperative scans, where blood products, granulation tissue, and partial volume effects complicate ground truth definition. In prediction of invasion, consistency, and resectability, radiomics and DL models provided clinically relevant risk stratification, although many still required manual preprocessing and remained dependent on scanner- and protocol-specific features.

The prognostic literature followed a similar trajectory: models for remission, recurrence, and therapeutic response often showed high apparent accuracy, particularly when multimodal inputs were combined, but their clinical maturity was uneven. Ensemble methods, gradient boosting, and DL frequently outperformed traditional regression, yet the strongest studies were those that incorporated interpretability tools, external cohorts, or clinically meaningful endpoints. Postoperative complication models were particularly relevant because they translated directly into bedside risk stratification; however, their utility depends on calibration, transportability, and consistent perioperative definitions. Likewise, operative workflow applications demonstrated that AI can move beyond static prediction into real-time support, training, and coaching, but these studies are still early-stage and mostly proof-of-concept.

NLP and LLMs represented the most rapidly evolving but also the most fragile domain. Their main value lay in structured extraction from unstructured records and patient-facing knowledge support, yet performance was highly dependent on model version, language, and domain-specific fine-tuning. The evidence supports a human-in-the-loop paradigm rather than autonomous use, especially for high-stakes questions such as surgery, pregnancy, and endocrine management. Overall, the literature suggests that AI is already useful as an adjunct for pattern recognition, triage, and risk prediction, but not yet mature enough for unsupervised deployment in most pituitary workflows.

The main limitations of the field are methodological rather than conceptual. The repeated reliance on retrospective datasets, limited multicenter validation, manual segmentation, and opaque model logic indicates that many studies remain at the development stage rather than implementation stage. Future work should prioritize prospective multicenter cohorts, standardized reporting, calibration and external validation, and transparent interpretability frameworks. In practical terms, the next step is not simply

higher AUCs, but clinically reliable models that are robust across institutions, usable at the point of care, and clearly tied to patient-centered outcomes.

Limitations of the review

This scoping review has some limitations. First, the search was restricted to Scopus and PubMed/MEDLINE, which may have resulted in the omission of relevant studies indexed exclusively in other databases or sources of grey literature. Second, only studies published in English or Spanish were considered, introducing the possibility of language bias. Third, consistent with the objectives of a scoping review, no formal risk-of-bias assessment or methodological quality appraisal was performed; therefore, the findings should be interpreted as a mapping of the available evidence rather than an evaluation of the certainty or strength of that evidence. Fourth, the included studies were highly heterogeneous with respect to patient populations, imaging protocols, AI architectures, outcome definitions, and validation strategies, precluding quantitative synthesis and limiting direct comparisons across studies. Finally, because AI is a rapidly evolving field, the evidence summarized herein represents a temporal snapshot of a continuously expanding literature, and some technologies may have advanced substantially since the completion of the search.

Conclusions

This scoping review demonstrates that AI has been applied across virtually all stages of pituitary adenoma management, including diagnosis, tumor classification, image segmentation, prediction of invasion and surgical complexity, treatment response assessment, postoperative outcome prediction, operative workflow optimization, and analysis of unstructured clinical data. Collectively, the evidence suggests that AI has evolved from experimental proof-of-concept applications toward clinically relevant decision-support tools capable of extracting diagnostically and prognostically meaningful information beyond conventional assessment methods.

However, despite consistently promising performance across multiple domains, most available evidence originates from retrospective, single-center studies with limited external validation, restricting the current level of clinical implementation. Future research should prioritize prospective multicenter validation, standardized reporting frameworks, model interpretability, and integration into real-world clinical workflows.

Abbreviations

ACTH
Adrenocorticotropic hormone
AI
Artificial intelligence
AIPI
Artificial Intelligence Performance Instrument

AISPE-Q
Artificial Intelligence Satisfaction and Performance Evaluation Questionnaire
ANN
Artificial neural network
AUC
Area under the curve
AUC-ROC
Area under the receiver operating characteristic curve
BERT
Bidirectional encoder representations from transformers
BiLSTM
Bidirectional long short-term memory
BIPSS
Bilateral inferior petrosal sinus sampling
CE-T1
Contrast-enhanced T1
CNNs
Convolutional neural networks
ConvLSTM
Convolutional long short-term memory
CRF
Conditional random fields
CSF
Cerebrospinal fluid
CSI
Cavernous sinus invasion
CV
Computer vision
DICOM®
Digital Imaging and Communications in Medicine
DeepFM
Factorization-machine based neural network
DL
Deep learning
DLR
Deep learning-based reconstruction
DT
Decision tree
EMR
Electronic medical records

EOR
Extent of resection
ET
Extra trees
GH
Growth hormone
GSU-Net
Gated-shaped U-net
GTR
Gross-total resection
IDEAL
Idea, Development, Exploration, Assessment, Long-term study
ICAs
Internal carotid arteries
k-NN
K-nearest neighbors
LASSO
Least absolute shrinkage and selection operator
LightGBM
Light gradient-boosting machine
LLMs
Large language models
LoRA
Low-rank adaptation
LR
Logistic regression
LR-EN
LR with elastic net
mOSATS
Modified Objective Structured Assessment of Technical Skills
ML
Machine learning
MRI
Magnetic resonance imaging
MSR-Net
Multi-scale residual network
NER
Named entity recognition
NLP
Natural language processing

nnU-Net
No-new-Net
PCDI
Programmed cell death-associated index
PDC-Net
Parallel dilated convolutional network
PI3K/AKT
Phosphoinositide 3-kinase/protein kinase B
Pit-1
Pituitary transcription factor 1
PitNETs
Pituitary neuroendocrine tumours
PM-CAD
Pituitary Microadenoma-Computer-Aided Diagnosis
PRISMA
Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR
PRISMA for Scoping Reviews
QAMAI
Quality Analysis of Medical Artificial Intelligence
R-CNN
Region-based CNN
ResNet50
Residual network with 50 layers
RF
Random forest
RNA
Ribonucleic acid
ROI
Regions of interest
SF-1
Steroidogenic factor 1
SHAP
SHapley Additive Explanations
SNR
Signal-to-noise ratio
SVM
Support vector machine
TCPA
Text Corpus of Pituitary Adenomas

TL
Transfer learning
Tpit
T-box pituitary transcription factor
TTF-1
Thyroid transcription factor 1
XGBoost
Extreme gradient boosting

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Use of artificial intelligence tools

Grammarly and DeepL were used for language editing and translation. The authors reviewed all outputs and approved the final manuscript.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS

H.J.P.C. conceptualized the study, developed the methodology, conducted the investigation, curated the data, performed formal analysis and validation, created the visualizations, and wrote the original draft of the manuscript. C.B.G. contributed to validation, data curation, and manuscript review and editing. M.M.O.M. contributed to formal analysis, manuscript review and editing, and supervision. O.L.A. supervised the research and administered the project. All authors reviewed and approved the final manuscript.

ACKNOWLEDGEMENTS

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

No datasets were generated or analyzed during the current study. Supplementary material is available online and includes the complete search strategies

References

1. Melmed S, Kaiser UB, Lopes MB, Bertherat J, Syro LV, Raverot G et al (2022) Clinical Biology of the Pituitary Adenoma. Endocr Rev 43(6):1003–1037. 10.1210/endrev/bnac010
2. Tritos NA, Miller KK (2023) Diagnosis and Management of Pituitary Adenomas: A Review. JAMA 329(16):1386–1398. 10.1001/jama.2023.5444
3. Daly AF, Beckers A (2020) The Epidemiology of Pituitary Adenomas. Endocrinol Metab Clin North Am 49(3):347–355. 10.1016/j.ecl.2020.04.002
4. Dai C, Sun B, Wang R, Kang J (2021) The Application of Artificial Intelligence and Machine Learning in Pituitary Adenomas. Front Oncol 11:784819. 10.3389/fonc.2021.784819
5. Khan DZ, Hanrahan JG, Baldeweg SE, Dorward NL, Stoyanov D, Marcus HJ (2023) Current and Future Advances in Surgical Therapy for Pituitary Adenoma. Endocr Rev 44(5):947–959. 10.1210/endrev/bnad014
6. Qiao N (2025) Artificial Intelligence in Pituitary Disease. Endocrinol Metab Clin North Am 54(4):715–728. 10.1016/j.ecl.2025.06.004
7. Zilka T, Benesova W (2024) Radiomics of pituitary adenoma using computer vision: a review. Med Biol Eng Comput 62(12):3581–3597. 10.1007/s11517-024-03163-3
8. Koong K, Preda V, Jian A, Liquet-Weiland B, Di Ieva A (2022) Application of artificial intelligence and radiomics in pituitary neuroendocrine and sellar tumors: a quantitative and qualitative synthesis. Neuroradiology 64(4):647–668. 10.1007/s00234-021-02845-1
9. Saha A, Tso S, Rabski J, Sadeghian A, Cusimano MD (2020) Machine learning applications in imaging analysis for patients with pituitary tumors: a review of the current literature and future directions. Pituitary 23(3):273–293. 10.1007/s11102-019-01026-x
10. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD et al (2021) PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 372:n160. 10.1136/bmj.n160
11. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D et al (2018) PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 169(7):467–473. 10.7326/M18-0850

12. Qian Y, Qiu Y, Li CC, Wang ZY, Cao BW, Huang HX et al (2020) A novel diagnostic method for pituitary adenoma based on magnetic resonance imaging using a convolutional neural network. Pituitary 23(3):246–252. 10.1007/s11102-020-01032-4
13. Gargya S, Jain S (2024) CAD System Design for Pituitary Tumor Classification based on Transfer Learning Technique. Curr Med Imaging 20(1):1–14. 10.2174/0115734056246146231018110415
14. Li Q, Zhu Y, Chen M, Guo R, Hu Q, Lu Y et al (2021) Development and Validation of a Deep Learning Algorithm to Automatic Detection of Pituitary Microadenoma From MRI. Front Med 8. 10.3389/fmed.2021.758690
15. Feng T, Fang Y, Pei Z, Li Z, Chen H, Hou P et al (2022) A Convolutional Neural Network Model for Detecting Sellar Floor Destruction of Pituitary Adenoma on Magnetic Resonance Imaging Scans. Front Neurosci 16. 10.3389/fnins.2022.900519
16. Li H, Zhao Q, Zhang Y, Sai K, Xu L, Mou Y et al (2021) Image-driven classification of functioning and nonfunctioning pituitary adenoma by deep convolutional neural networks. Comput Struct Biotechnol J 19:3077–3086. 10.1016/j.csbj.2021.05.023
17. Liu A, Xiao Y, Wu M, Tan Y, He Y, Deng Y et al (2022) Diagnosis and classification prediction model of pituitary tumor based on machine learning. Neural Comput Appl 34(12):9257–9272. 10.1007/s00521-021-06277-z
18. Demir AN, Ayata D, Oz A, Sulu C, Kara Z, Sahin S et al (2025) Machine Learning May Be an Alternative to BIPSS in the Differential Diagnosis of ACTH-dependent Cushing Syndrome. J Clin Endocrinol Metab 110(2):e412–e422. 10.1210/clinem/dgae180
19. Wang Y, Chen S, Shi F, Cheng X, Xu Q, Li J et al (2021 Aug) MR-Based Radiomics for Differential Diagnosis between Cystic Pituitary Adenoma and Rathke Cleft Cyst. Comput Math Methods Med 10:20216438861. 10.1155/2021/6438861
20. Liu Y, Zhou Y, Zhou C, Wang Z, Fan Z, Tang K et al (2025) Machine learning method based on radiomics help differentiate posterior pituitary tumors from pituitary neuroendocrine tumors and craniopharyngioma. Sci Rep 15(1):19967. 10.1038/s41598-025-05143-5
21. Wu G, Ning Z, Yan X, Li J, Ma C, Cheng H et al (2025) Deep learning based semi-automated model can predict lineage in patients with pituitary neuroendocrine tumors. Acta Neuropathol Commun 13(1):200. 10.1186/s40478-025-02104-x
22. Peng A, Dai H, Duan H, Chen Y, Huang J, Zhou L et al (2020) A machine learning model to precisely immunohistochemically classify pituitary adenoma subtypes with radiomics based on preoperative magnetic resonance imaging. Eur J Radiol 125. 10.1016/j.ejrad.2020.108892
23. Baysal B, Eser MB, Dogan MB, Kursun MA, Baysal B, Eser MB et al (2022) Multivariable Diagnostic Prediction Model to Detect Hormone Secretion Profile From T2W MRI Radiomics with Artificial Neural Networks in Pituitary Adenomas. Medeni Med J 37(1):36–43. 10.4274/MMJ.galenos.2022.58538
24. Zhang S, Song G, Zang Y, Jia J, Wang C, Li C et al (2018) Non-invasive radiomics approach potentially predicts non-functioning pituitary adenomas subtypes before surgery. Eur Radiol

28(9):3692–3701. 10.1007/s00330-017-5180-6

25. Gu L, Zhou F, Wu B, Yang J, Li B, Fan Y et al (2026) Preoperative prediction of p53 overexpression in pituitary neuroendocrine tumors using MRI radiomics. Front Neurol 16. 10.3389/fneur.2025.1693959
26. Wang G, Yan S, Zhang L, Lin L, Liu R, Han Y et al (2026) Machine Learning-driven PCDI Classifier for Invasive PitNETs. Curr Gene Ther 26(2):e15665232399193. 10.2174/0115665232399193250529074831
27. Yang S, Pu C, Deng C, Bai X, Tian C, Zhang W et al (2026) Single-cell and machine learning approaches reveal METTL14-mediated autophagy via PI3K/AKT signaling in invasive PitNET. Cancer Cell Int 26(1):72. 10.1186/s12935-026-04192-8
28. Micko A, Placzek F, Fonollà R, Winklehner M, Sentosa R, Krause A et al (2021) Diagnosis of Pituitary Adenoma Biopsies by Ultrahigh Resolution Optical Coherence Tomography Using Neuronal Networks. Front Endocrinol 12. 10.3389/fendo.2021.730100
29. Jiang X, Xiao J, Zhang Q, Wang L, Jiang J, Lan K et al (2023) Improved U-Net based on cross-layer connection for pituitary adenoma MRI image segmentation. Math Biosci Eng 20(1):34–51. 10.3934/mbe.2023003
30. Wang H, Zhang W, Li S, Fan Y, Feng M, Wang R (2021) Development and Evaluation of Deep Learning-based Automated Segmentation of Pituitary Adenoma in Clinical Task. J Clin Endocrinol Metab 106(9):2535–2546. 10.1210/clinem/dgab371
31. Zhang Q, Jiang X, Huang X, Zhou C (2024) MSR-Net: Multi-Scale Residual Network Based on Attention Mechanism for Pituitary Adenoma MRI Image Segmentation. IEEE Access 12:119371–119382. 10.1109/ACCESS.2024.3449925
32. Zhang Q, Cheng J, Zhou C, Jiang X, Zhang Y, Zeng J et al (2023) PDC-Net: parallel dilated convolutional network with channel attention mechanism for pituitary adenoma segmentation. Front Physiol 14. 10.3389/fphys.2023.1259877
33. Shu X, Zhou Y, Li F, Zhou T, Meng X, Wang F et al (2021) Three-Dimensional Semantic Segmentation of Pituitary Adenomas Based on the Deep Learning Framework-nnU-Net: A Clinical Perspective. Micromachines 12(12):1473. 10.3390/mi12121473
34. Da Mutten R, Zanier O, Ciobanu-Caraus O, Voglis S, Hugelshofer M, Pangalu A et al (2024) Automated volumetric assessment of pituitary adenoma. Endocrine 83(1):171–177. 10.1007/s12020-023-03529-x
35. Park H, Nam YK, Kim HS, Park JE, Lee DH, Lee J et al (2023) Deep learning-based image reconstruction improves radiologic evaluation of pituitary axis and cavernous sinus invasion in pituitary adenoma. Eur J Radiol 158. 10.1016/j.ejrad.2022.110647
36. Kim M, Kim HS, Park JE, Park SY, Kim YH, Kim SJ et al (2022) Thin-Slice Pituitary MRI with Deep Learning–Based Reconstruction for Preoperative Prediction of Cavernous Sinus Invasion by Pituitary Adenoma: A Prospective Study. Am J Neuroradiol 43(2):280–285. 10.3174/ajnr.A7387

37. Lee DH, Park JE, Nam YK, Lee J, Kim S, Kim YH et al (2021) Deep learning-based thin-section MRI reconstruction improves tumour detection and delineation in pre- and post-treatment pituitary adenoma. Sci Rep 11(1):21302. 10.1038/s41598-021-00558-2
38. Fang Y, Wang H, Feng M, Chen H, Zhang W, Wei L et al (2022) Application of Convolutional Neural Network in the Diagnosis of Cavernous Sinus Invasion in Pituitary Adenoma. Front Oncol 12. 10.3389/fonc.2022.835047
39. Yang Q, Wang Y, Wu J, Hu H, He Y, Wang Y et al (2025) Preoperative prediction of pituitary neuroendocrine tumor consistency based on multiparametric MRI radiomics: a multicenter study. BMC Cancer 25(1):1501. 10.1186/s12885-025-14799-1
40. Chen Y, Zhong J, Hou P, Wang X, Li J, Li Z et al (2025) Radiomics-based multiple machine learning approaches for investigating medial wall invasion of the cavernous sinus in pituitary adenomas. Front Oncol 15. 10.3389/fonc.2025.1706895
41. Niu J, Zhang S, Ma S, Diao J, Zhou W, Tian J et al (2019) Preoperative prediction of cavernous sinus invasion by pituitary adenomas using a radiomics method based on magnetic resonance images. Eur Radiol 29(3):1625–1634. 10.1007/s00330-018-5725-3
42. Zhang C, Heng X, Neng W, Chen H, Sun A, Li J et al (2022) Prediction of high infiltration levels in pituitary adenoma using MRI-based radiomics and machine learning. Chin Neurosurg J 8(1):21. 10.1186/s41016-022-00290-4
43. Cuocolo R, Ugga L, Solari D, Corvino S, D'Amico A, Russo D et al (2020) Prediction of pituitary adenoma surgical consistency: radiomic data mining and machine learning on T2-weighted MRI. Neuroradiology 62(12):1649–1656. 10.1007/s00234-020-02502-z
44. Patel BK, Tariciotti L, DiRocco L, Mandile A, Lohana S, Rodas A et al (2025) Regression modeling with convolutional neural network for predicting extent of resection from preoperative MRI in giant pituitary adenomas: a pilot study. J Neurosurg 143(1):174–183. 10.3171/2024.10.JNS241527
45. Shinya Y, Ghaith AK, Hong S, Herndon JS, Palit SR, Erickson D et al (2025) A supervised machine learning approach for predicting the need for postsurgical intervention in acromegaly. Neurosurg Focus 59(1):E10. 10.3171/2025.4.FOCUS2597
46. Zanier O, Zoli M, Staartjes VE, Guaraldi F, Asioli S, Rustici A et al (2022) Machine learning-based clinical outcome prediction in surgery for acromegaly. Endocrine 75(2):508–515. 10.1007/s12020-021-02890-z
47. Qiao N, Shen M, He W, He M, Zhang Z, Ye H et al (2021) Machine learning in predicting early remission in patients after surgical treatment of acromegaly: a multicenter study. Pituitary 24(1):53–61. 10.1007/s11102-020-01086-4
48. Kocak B, Durmaz ES, Kadioglu P, Polat Korkmaz O, Comunoglu N, Tanriover N et al (2019) Predicting response to somatostatin analogues in acromegaly: machine learning-based high-dimensional quantitative texture analysis on T2-weighted MRI. Eur Radiol 29(6):2731–2739. 10.1007/s00330-018-5876-2

49. Huber M, Luedi MM, Schubert GA, Musahl C, Tortora A, Frey J et al (2022) Machine Learning for Outcome Prediction in First-Line Surgery of Prolactinomas. Front Endocrinol 13. 10.3389/fendo.2022.810219
50. Park YW, Eom J, Kim S, Kim H, Ahn SS, Ku CR et al (2021) Radiomics With Ensemble Machine Learning Predicts Dopamine Agonist Response in Patients With Prolactinoma. J Clin Endocrinol Metab 106(8):e3069–e3077. 10.1210/clinem/dgab159
51. Fan Y, Li Y, Bao X, Zhu H, Lu L, Yao Y et al (2021) Development of Machine Learning Models for Predicting Postoperative Delayed Remission in Patients With Cushing's Disease. J Clin Endocrinol Metab 106(1):e217–e231. 10.1210/clinem/dgaa698
52. Zhang W, Zhang D, Liu S, Wang H, Liu X, Dai C et al (2024) Predicting delayed remission in Cushing's disease using radiomics models: a multi-center study. Front Oncol 13. 10.3389/fonc.2023.1218897
53. Zhang W, Li D, Feng M, Hu B, Fan Y, Chen Q et al (2021) Electronic Medical Records as Input to Predict Postoperative Immediate Remission of Cushing's Disease: Application of Word Embedding. Front Oncol 11. 10.3389/fonc.2021.754882
54. Liu Y, Liu X, Hong X, Liu P, Bao X, Yao Y et al (2019) Prediction of Recurrence after Transsphenoidal Surgery for Cushing's Disease: The Use of Machine Learning Algorithms. Neuroendocrinology 108(3):201–210. 10.1159/000496753
55. Fan Y, Li D, Liu Y, Feng M, Chen Q, Wang R (2021) Toward better prediction of recurrence for Cushing's disease: a factorization-machine based neural approach. Int J Mach Learn Cybern 12(3):625–633. 10.1007/s13042-020-01192-6
56. Shinya Y, Ghaith AK, Hong S, Erickson D, Bancos I, Herndon JS et al (2025) Machine learning–based model to predict long-term tumor control and additional interventions following pituitary surgery for Cushing's disease. J Neurosurg 143(1):184–193. 10.3171/2024.10.JNS241338
57. Sharifı G, Paraandavaji E, Akbari Dilmaghani N, Emami Meybodi T, Mohammadzadeh I, Sadeghi N et al (2025) The CuPeR model: A dynamic online tool for predicting Cushing's disease persistence and recurrence after pituitary surgery. J Clin Transl Endocrinol 41:100417. 10.1016/j.jcte.2025.100417
58. Shen C, Liu X, Jin J, Han C, Wu L, Wu Z et al (2023) A Novel Magnetic Resonance Imaging-Based Radiomics and Clinical Predictive Model for the Regrowth of Postoperative Residual Tumor in Non-Functioning Pituitary Neuroendocrine Tumor. Med (Kaunas) 59(9):1525. 10.3390/medicina59091525
59. Speckter H, Radulovic M, Lazo E, Hernandez G, Bido J, Rivera D et al (2025) Prediction of Pituitary Adenoma's Volumetric Response to Gamma Knife Radiosurgery Using Machine Learning-Supported MRI Radiomics. J Clin Med 14(9):2896. 10.3390/jcm14092896
60. Fan Y, Liu Z, Hou B, Li L, Liu X, Liu Z et al (2019) Development and validation of an MRI-based radiomic signature for the preoperative prediction of treatment response in patients with invasive functional pituitary adenoma. Eur J Radiol 121. 10.1016/j.ejrad.2019.108647
61. Yin C, Xu F, Jiang X, Wang Y, Xie T, Wang X et al (2026) Early postoperative hyponatremia after pituitary adenoma surgery: risk factors, predictive model, and clinical implications. BMC Endocr

Disord 26(1):99. 10.1186/s12902-026-02212-2

62. Lin K, Zhang J, Zhao L, Wei L, Wang S (2025) Machine learning algorithms for predicting delayed hyponatremia after transsphenoidal surgery for patients with pituitary adenoma. Sci Rep 15(1):1463. 10.1038/s41598-024-83319-1
63. Tan H, Miao X, Pei Y, Miao F (2025) Predictive model of delayed hyponatremia after endoscopic endonasal transsphenoidal resection of pituitary adenoma. Front Hum Neurosci 19. 10.3389/fnhum.2025.1674519
64. Liu X, Wang Z, Zhang Y, Luan X (2025) Construction and validation of a LASSO penalized logistic regression model predicting hypernatremia after pituitary adenoma surgery. Sci Rep 15(1):38552. 10.1038/s41598-025-22512-2
65. Chen Y, Zhong J, Li H, Lin K, Wei L, Wang S (2024) Predictive modeling of arginine vasopressin deficiency after transsphenoidal pituitary adenoma resection by using multiple machine learning algorithms. Sci Rep 14(1):22210. 10.1038/s41598-024-72486-w
66. Fang Y, Wang H, Feng M, Zhang W, Cao L, Ding C et al (2021) Machine-Learning Prediction of Postoperative Pituitary Hormonal Outcomes in Nonfunctioning Pituitary Adenomas: A Multicenter Study. Front Endocrinol 12:748725. 10.3389/fendo.2021.748725
67. Hussein Z, Slack RW, Marcus HJ, Mazomenos EB, Baldeweg SE (2023) Post-Operative Medium- and Long-Term Endocrine Outcomes in Patients with Non-Functioning Pituitary Adenomas—Machine Learning Analysis. Cancers 15(10):2771. 10.3390/cancers15102771
68. Villalonga JF, Solari D, Cuocolo R, De Lucia V, Ugga L, Gragnaniello C et al (2022) Clinical application of the sellar barrier's concept for predicting intraoperative CSF leak in endoscopic endonasal surgery for pituitary adenomas with a machine learning analysis. Front Surg 9. 10.3389/fsurg.2022.934721
69. Chang H, Zhao K, Qiu J, Ji X-J, Chen W-G, Li B-Y et al (2023) Prediction of intraoperative cerebrospinal fluid leaks in endoscopic endonasal transsphenoidal pituitary surgery based on a deep neural network model trained with MRI images: a pilot study. Front Neurosci 17:1203698. 10.3389/fnins.2023.1203698
70. Hollon TC, Parikh A, Pandian B, Tarpeh J, Orringer DA, Barkan AL et al (2018) A machine learning approach to predict early outcomes after pituitary adenoma surgery. Neurosurg Focus 45(5):E8. 10.3171/2018.8.FOCUS18268
71. Qiao N, Zhang Q, Chen L, He W, Ma Z, Ye Z et al (2023) Machine learning prediction of venous thromboembolism after surgeries of major sellar region tumors. Thromb Res 226:1–8. 10.1016/j.thromres.2023.04.007
72. Chen M, Li Y, Zhou S, Zou L, Yu L, Deng T et al (2024) Establishment of a risk prediction model for olfactory disorders in patients with transnasal pituitary tumors by machine learning. Sci Rep 14(1):12514. 10.1038/s41598-024-62963-7
73. Zhang Y, Chen C, Huang W, Cheng Y, Teng Y, Zhang L et al (2021) Machine Learning-Based Radiomics of the Optic Chiasm Predict Visual Outcome Following Pituitary Adenoma Surgery. J Pers

Med 11(10):991. 10.3390/jpm11100991

74. Zhang Y, Huang Z, Zhao Y, Xu J, Chen C, Xu J (2025) Radiomics using multiparametric magnetic resonance imaging to predict postoperative visual outcomes of patients with pituitary adenoma. Asian J Surg 48(1):166–172. 10.1016/j.asjsur.2024.07.132
75. Zhang Y, Zheng J, Huang Z, Teng Y, Chen C, Xu J (2023) Predicting visual recovery in pituitary adenoma patients post-endoscopic endonasal transsphenoidal surgery: Harnessing delta-radiomics of the optic chiasm from MRI. Eur Radiol 33(11):7482–7493. 10.1007/s00330-023-09963-9
76. Khan DZ, Luengo I, Barbarisi S, Addis C, Culshaw L, Dorward NL et al (2021) Automated operative workflow analysis of endoscopic pituitary surgery using machine learning: development and preclinical evaluation (IDEAL stage 0). J Neurosurg. 10.3171/2021.6.JNS21923. 2021.6.JNS21923
77. Mao Z, Das A, Khan DZ, Williams SC, Hanrahan JG, Stoyanov D et al (2025) ConsisTNet: a spatio-temporal approach for consistent anatomical localization in endoscopic pituitary surgery. Int J Comput Assist Radiol Surg 20(6):1239–1248. 10.1007/s11548-025-03369-2
78. Wong CE, Chen PW, Hsu HJ, Cheng SY, Fan CC, Chen YC et al (2024) Collaborative Human–Computer Vision Operative Video Analysis Algorithm for Analyzing Surgical Fluency and Surgical Interruptions in Endonasal Endoscopic Pituitary Surgery: Cohort Study. J Med Internet Res 26(1):e56127. 10.2196/56127
79. Khan DZ, Newall N, Koh CH, Das A, Aapan S, Layard Horsfall H et al (2024) Video-Based Performance Analysis in Pituitary Surgery - Part 2: Artificial Intelligence Assisted Surgical Coaching. World Neurosurg 190:e797–808. 10.1016/j.wneu.2024.07.219
80. Zanier O, Alakmeh A, Mutten RD, Carretta A, Zoli M, Mazzatenta D et al (2026) Real-time intraoperative depth estimation in transsphenoidal surgery using deep learning: A feasibility study. J Clin Neurosci 147:111910. 10.1016/j.jocn.2026.111910
81. Fang A, Hu J, Zhao W, Feng M, Fu J, Feng S et al (2022) Extracting clinical named entity for pituitary adenomas from Chinese electronic medical records. BMC Med Inf Decis Mak 22(1):72. 10.1186/s12911-022-01810-z
82. Hu J, Fu J, Zhao W, Lou P, Feng M, Ren H et al (2024) Characterizing pituitary adenomas in clinical notes: Corpus construction and its application in LLMs. Health Inf J 30(4):14604582241291442. 10.1177/14604582241291442
83. Aliyeva A, Nevzati E, Grassia F, Riaz M, Mann S, Nasirov R (2026) AI at the Sella Turcica: Multi-Model Large Language Model Evaluation in Pituitary Adenomas. Brain Spine 6:105997. 10.1016/j.bas.2026.105997
84. Chen Y, Chen Y, Chen L, Feng T, Wang S (2025) Evaluation of ChatGPT as a supplementary tool for pituitary adenomas: An observational study based on simulated consultations. Med (Baltim) 104(46):e45928. 10.1097/MD.0000000000045928
85. Şenoymak MC, Erbatur NH, Şenoymak İ, Fırat SN (2024) The Role of Artificial Intelligence in Endocrine Management: Assessing ChatGPT's Responses to Prolactinoma Queries. J Pers Med 14(4):330. 10.3390/jpm14040330

Figures

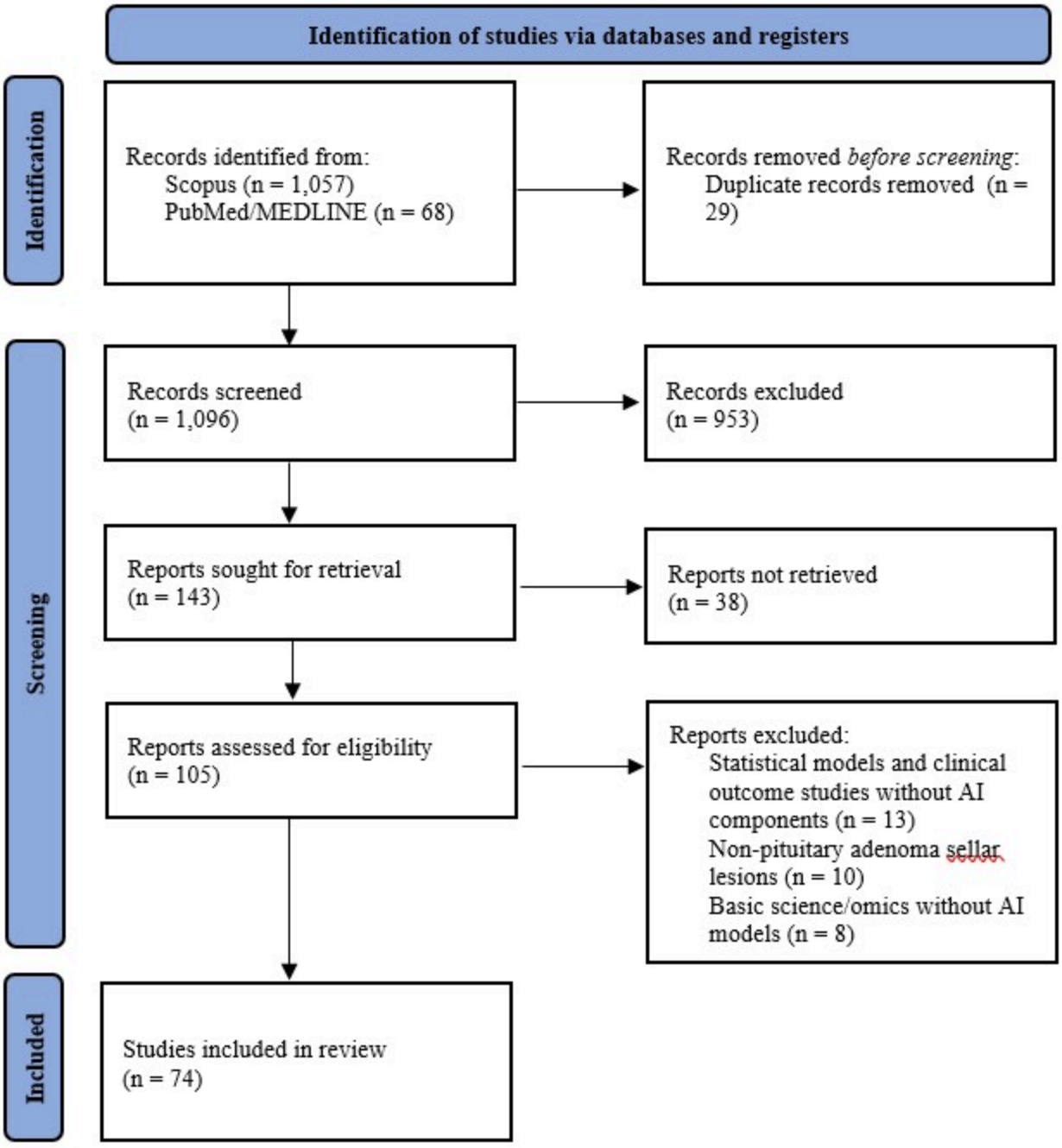


Figure 1

Flowchart of the search and study selection processes

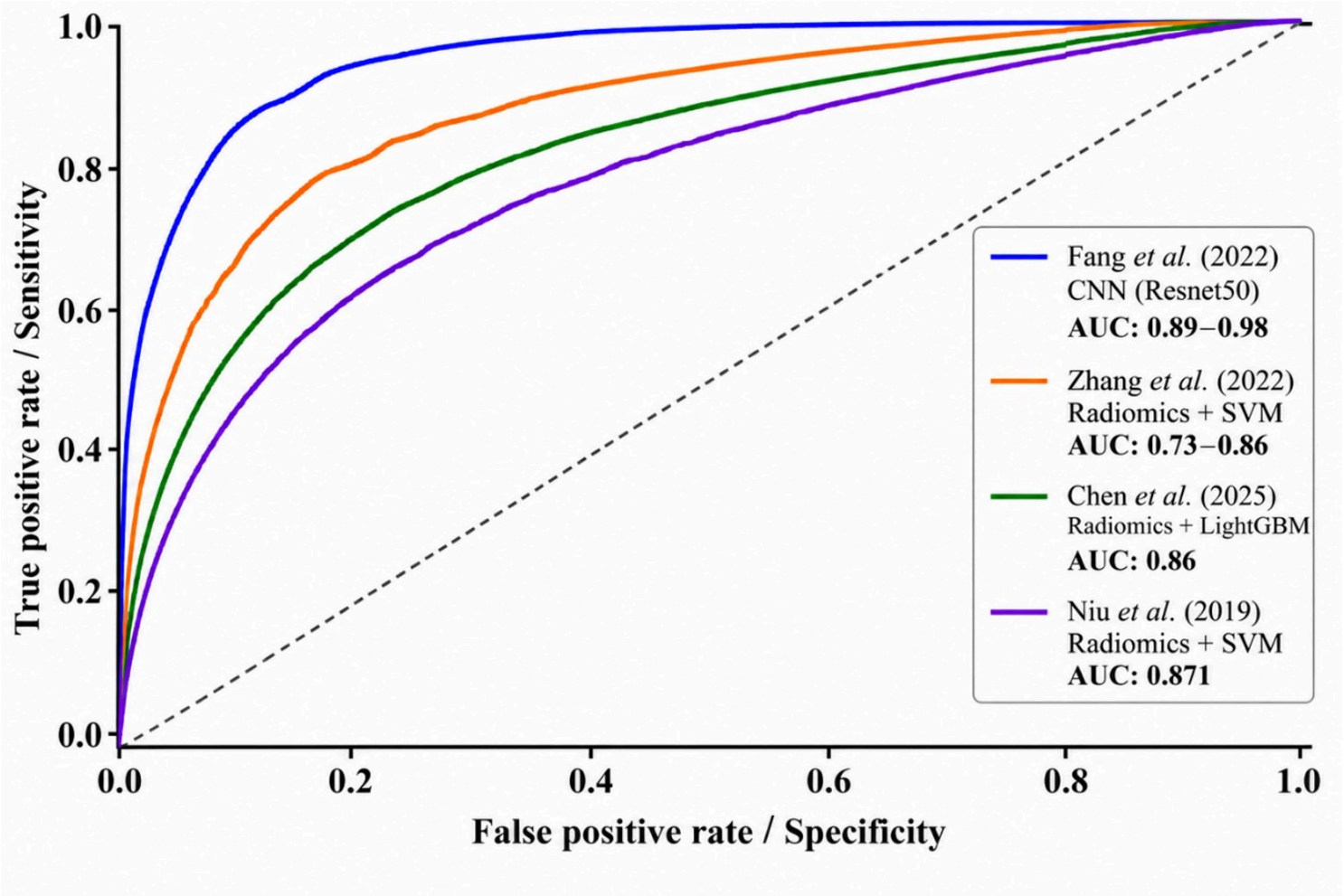


Figure 2

Prediction of cavernous sinus invasion. *Note:* This is a representative illustration based on reported area under the curve values

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SUPPLEMENTARYMATERIAL.docx