

Artificial intelligence and machine learning in pediatric endocrine tumors: opportunities, pitfalls, and a roadmap for trustworthy clinical translation

Michaela Kuhlen, Fabio Hellmann, Elisabeth Pfaehler, Elisabeth André, Antje Redlich

Angaben zur Veröffentlichung / Publication details:

Kuhlen, Michaela, Fabio Hellmann, Elisabeth Pfaehler, Elisabeth André, and Antje Redlich. 2026. "Artificial intelligence and machine learning in pediatric endocrine tumors: opportunities, pitfalls, and a roadmap for trustworthy clinical translation." *Biomedicines* 14 (1): 146. <https://doi.org/10.3390/biomedicines14010146>.

Nutzungsbedingungen / Terms of use:

CC BY 4.0



Review

Artificial Intelligence and Machine Learning in Pediatric Endocrine Tumors: Opportunities, Pitfalls, and a Roadmap for Trustworthy Clinical Translation

Michaela Kuhlen ^{1,2,*}, Fabio Hellmann ³, Elisabeth Pfaehler ⁴, Elisabeth André ³ and Antje Redlich ¹

¹ Department of Pediatrics, Pediatric Hematology/Oncology, Otto-von-Guericke-University, Leipziger Str. 44, D-39120 Magdeburg, Germany; antje.redlich@med.ovgu.de

² Paediatrics and Adolescent Medicine, Faculty of Medicine, University of Augsburg, Stenglinstr. 2, D-86156 Augsburg, Germany

³ Human-Centered Artificial Intelligence, University of Augsburg, Universitaetsstrasse 6a, D-86159 Augsburg, Germany; fabio.hellmann@informatik.uni-augsburg.de (F.H.)

⁴ Institute for Neuroscience and Medicine 4 (INM-4), Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, D-52428 Jülich, Germany

* Correspondence: michaela.kuhlen@uk-augsburg.de; Tel.: +49-821-9300

Abstract

Artificial intelligence (AI) and machine learning (ML) are reshaping cancer research and care. In pediatric oncology, early evidence—most robust in imaging—suggests value for diagnosis, risk stratification, and assessment of treatment response. Pediatric endocrine tumors are rare and heterogeneous, including intra- and extra-adrenal paraganglioma (PGL), adrenocortical tumors (ACT), differentiated and medullary thyroid carcinoma (DTC/MTC), and gastroenteropancreatic neuroendocrine neoplasms (GEP-NEN). Here, we provide a pediatric-first, entity-structured synthesis of AI/ML applications in endocrine tumors, paired with a methods-for-clinicians primer and a pediatric endocrine tumor guardrails checklist mapped to contemporary reporting/evaluation standards. We also outline a realistic EU-anchored roadmap for translation that leverages existing infrastructures (EXPeRT, ERN PaedCan). We find promising—yet preliminary—signals for early non-remission/recurrence modeling in pediatric DTC and interpretable survival prediction in pediatric ACT. For PGL and GEP-NEN, evidence remains adult-led (biochemical ML screening scores; CT/PET radiomics for metastatic risk or peptide receptor radionuclide therapy response) and serves primarily as methodological scaffolding for pediatrics. Cross-cutting insights include the centrality of calibration and validation hierarchy and the current limits of explainability (radiomics texture semantics; saliency ≠ mechanism). Translation is constrained by small datasets, domain shift across age groups and sites, limited external validation, and evolving regulatory expectations. We close with pragmatic, clinically anchored steps—benchmarks, multi-site pediatric validation, genotype-aware evaluation, and equity monitoring—to accelerate safe, equitable adoption in pediatric endocrine oncology.



Academic Editor: Hermann L. Müller

Received: 29 November 2025

Revised: 1 January 2026

Accepted: 9 January 2026

Published: 11 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](#).

Keywords: pediatric oncology; endocrine tumors; machine learning; explainability; risk stratification; techquity; radiomics; ethical AI

1. Introduction

Pediatric endocrine tumors are rare and clinically diverse. Treatment choices must balance oncologic control with preservation of endocrine function, normal growth and

development, and the prevention of late effects that span a child's lifetime. This review concentrates on five entities that dominate pediatric endocrine oncology: differentiated thyroid carcinoma (DTC), medullary thyroid carcinoma (MTC), adrenocortical tumors (ACT), intra- (former termed pheochromocytoma) and extra-adrenal paraganglioma (PGL), and gastroenteropancreatic neuroendocrine neoplasms (GEP-NEN).

DTC is the most frequent endocrine malignancy in children and adolescents, with higher malignancy rates in nodules and more frequent nodal/distant spread than in adults, yet very low disease-specific mortality. Decisions aim to tailor surgery, radioactive iodine (RAI), and surveillance to reduce morbidity while safeguarding control [1–3].

Pediatric MTC is uncommon and largely multiple endocrine neoplasia type 2 (MEN2)-associated. Timing of thyroidectomy is genotype-driven (*RET* codon risk), with calcitonin and carcinoembryonic antigen used for follow-up [4,5].

ACT are ultra-rare in pediatrics and biologically heterogeneous. Surgery is central, but outcomes vary widely even within stage, motivating refined prognostication [6,7].

Pediatric PGL have a very high heritable fraction and genotype-specific metastatic risk (e.g., *SDHB*). Care depends on safe pre-operative management, resection completeness, and lifelong genotype-tailored surveillance [8–11].

GEP-NEN are exceptionally rare in children and adolescents. Contemporary care borrows adult pathways centered on somatostatin receptor (SSTR) imaging and peptide receptor radionuclide therapy (PRRT) for advanced disease. Standardized detection, non-invasive grading surrogates, and consistent response assessment remain pressing needs [12–14].

1.1. A Brief Primer on AI/ML for Clinicians

AI refers to the broader field of creating computational systems capable of performing tasks that typically require human intelligence. ML is a subset of AI focused on developing algorithms and statistical models that enable systems to learn from data.

In pediatric oncology, the most relevant applications are supervised models trained to predict predefined outcomes—recurrence, survival, treatment response, or complication risk—using structured variables (demography, staging, laboratory data), images (ultrasound, CT, MRI, PET), pathology, or genomics.

Tabular learners (regularized regression, gradient-boosted trees, and survival extensions) work well with clinical variables, whereas convolutional networks dominate image analysis. Because predictions inform care, two properties are critical: predicted probabilities should reflect observed frequencies (calibration), and clinicians should be able to see which inputs/which image part drove the estimate and why (explainability) [15–17]. Depending on the questions the user may have regarding the model's outcome, a distinct explanation must be applied. In case the user wants to investigate which input features were completely irrelevant to the outcome, Alterfactual Explanations should be used [18]. To compare the model's prediction with a healthy/sick version of the input of the same patient, Counterfactual Explanations would create appropriate complementary results [19]. Another method, primarily used for image explanations, is the use of saliency maps, where each pixel in the provided image is highlighted based on the importance leading to the model's outcome. This approach showed promising results in analyzing adult patients' pain level based on their facial expressions [20]. Reporting should therefore pair discrimination metrics with calibration curves and provide case-level explanations that are clinically coherent. In pediatrics, prospective evaluation and human oversight are prerequisites before clinical use.

Patient-facing AI—such as chatbots, digital navigators, and symptom-tracking assistants—aims to support children and families with education, logistics, and self-

management across the care pathway. Many such tools are powered by large language models (LLMs) for natural-language interaction and, increasingly, by large multimodal models (LMMs) that process both text and images to provide integrated responses. Although this review focuses on clinician-facing decision support, we briefly note patient-facing tools where they intersect with pediatric endocrine oncology.

Figure 1 summarizes the main method families, pediatric oncology domains, and evaluation and explainability concepts that structure this review, highlighting how calibration, discrimination, and case-level explanations underpin validation and implementation.

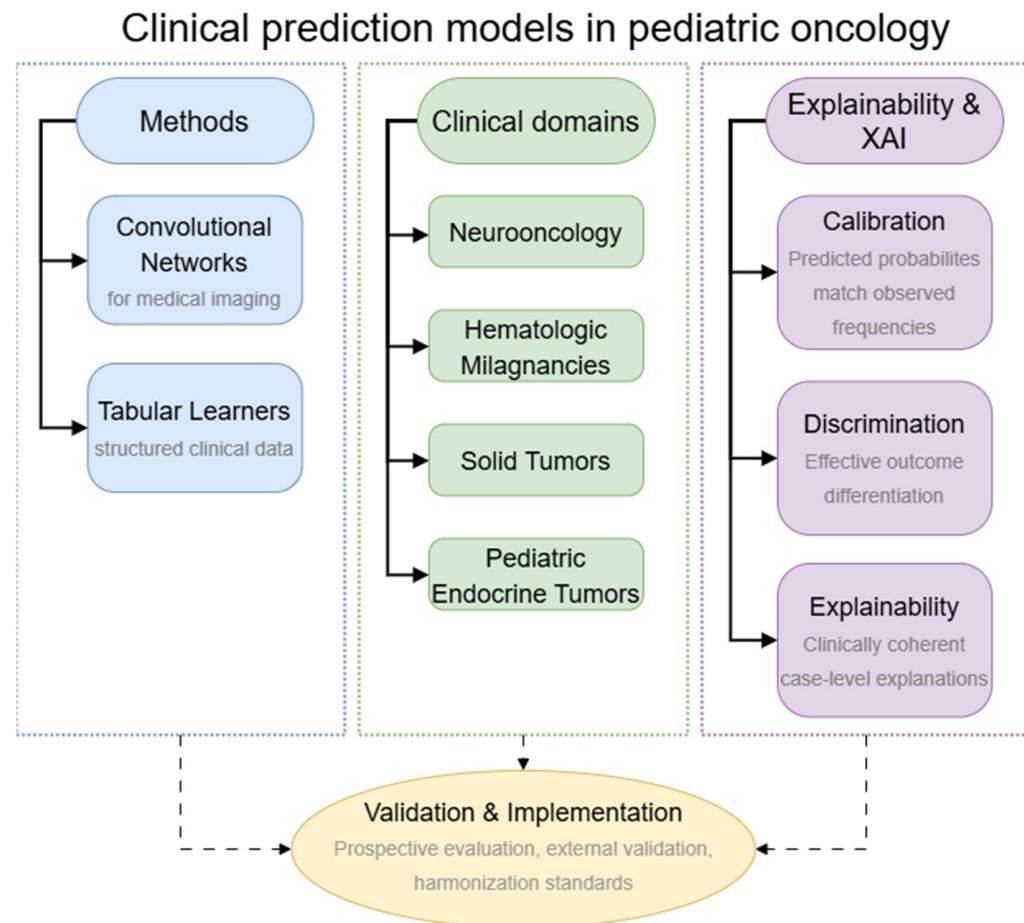


Figure 1. Schematic overview of clinical prediction models in pediatric oncology, distinguishing core method families (blue), key clinical domains (green), and evaluation and explainability pillars (purple) that together support validation and implementation through prospective and external evaluation, harmonization, and human oversight.

1.2. Current Evidence on AI/ML in Pediatric Oncology

AI/ML has matured unevenly across pediatric oncology [21–24]. Evidence is most advanced in neuro-oncology, where deep learning for detection, segmentation, grading surrogates, and treatment-response assessment appear regularly. The last two tasks are also often addressed by radiomic analysis. Nonetheless, truly independent external validation and clinical-impact studies remain uncommon, and performance can degrade across scanners and sites due to domain shift [25–28]. Hematologic malignancies have seen steady progress in risk stratification and minimal-residual-disease support tools [29–36]. In solid tumors, retrospective, single-center radiomics and deep learning pipelines are proliferating for diagnosis and survival prediction, yet many lack harmonized acquisition protocols, rigorous calibration, pre-specified action thresholds, and interpretability—all barriers to

translation [37–40]. Multi-institutional efforts to standardize imaging, define pediatric-specific outcomes, and harmonize data are emerging but are not yet routine [41,42].

1.3. AI/ML in Pediatric Endocrine Tumors

Against this backdrop, pediatric endocrine tumors present both an opportunity and a stress test: DTC offers large enough cohorts for interpretable recurrence-risk modeling, ACT and PGL demand models that remain reliable at rare-disease scale and across genotypes, and GEP-NEN require method transfer from adult datasets with careful pediatric validation. Progress has been slow because cohorts are small and genetically heterogeneous (e.g., *RET*, *SDHx*, *TP53*), outcomes accrue over long horizons, imaging and assay protocols vary across centers, and cross-border data sharing (general data protection regulation; GDPR) and assent/consent requirements add friction to aggregation.

This review responds with a pediatric-first, entity-structured synthesis of AI/ML applications in DTC, MTC, ACT, PGL, and GEP-NEN, mapped to specific clinical questions (e.g., diagnosis, prognosis, treatment response). We clearly separate pediatric from adult-only evidence (the latter used as methodological context), consolidate studies in a Table 1, and distill methodological and clinical guardrails aligned with contemporary reporting/evaluation standards in Table 2. We also outline an EU-anchored route to harmonized, multi-site validation via the European Cooperative Study Group for Pediatric Rare Tumors (EXPeRT), and the European Reference Network for Paediatric Oncology (ERN PaedCan), with priority on clinically meaningful outcomes rather than algorithmic metrics alone.

2. Methods

We reviewed peer-reviewed studies that developed, validated, or evaluated AI/ML tools for diagnosis, prognosis, treatment response, or clinical decision support in children and adolescents with the five endocrine tumor entities of interest: DTC, MTC, ACT, PGL, and GEP-NEN. “Pediatric” was defined as birth through 18 years of age. Mixed-age studies were eligible when pediatric results were reported separately, or a pediatric subgroup could be extracted with reasonable fidelity. Purely adult cohorts were excluded except where (i) the methodology was exemplary and clearly informative for pediatric translation in the same tumor family, or (ii) guidance documents (reporting standards, evaluation frameworks, or regulatory materials) were necessary to contextualize pediatric adoption. We excluded case reports, editorials, letters, conference abstracts without peer-reviewed full texts, and preprints unless later published.

Searches were performed in PubMed/MEDLINE from inception to 30 October 2025 using Boolean combinations of pediatric terms, tumor entity terms, and AI/ML terms (verbatim queries in Appendix A Table A1). We also hand-searched reference lists of included papers and recent reviews and used citation tracking to identify additional records. Contemporary healthcare-AI guidance (e.g., TRIPOD-AI, PROBAST-AI, STARD-AI, SPIRIT-AI/CONSORT-AI, DECIDE-AI, CLAIM, METRICS) and relevant European regulatory materials were consulted to frame evaluation standards [43–50]. Titles/abstracts were screened, followed by full-text review. We extracted study characteristics including population (age range, tumor entity, sample size, setting), data sources (imaging modality, laboratory/clinical variables, and—when available—omics), model class and training scheme, outcomes and time horizons, validation strategy (internal resampling, temporal split, geographic external testing), performance metrics (discrimination and calibration), explainability, and any decision-curve or utility analyses.

Given the expected small number of pediatric endocrine tumor studies and the heterogeneity of designs, inputs, and endpoints, we conducted a narrative synthesis organized by

tumor entity and clinical question. For most entity–task pairs, pediatric evidence comprised one or no studies (often $k \leq 1$), with non-comparable outcomes, variable imaging/assay protocols, and absence of calibration and action-threshold reporting. Accordingly, no meta-analysis or semi-quantitative pooling was attempted. When comparable pediatric and adult evidence existed, adult-only studies were treated as methodological scaffolding and labeled explicitly to avoid overgeneralization. Emphasis is placed on validation approach, calibration reporting, transparency/interpretability, and reporting completeness.

This review was not prospectively registered. Its scope mirrors the five entities defined in the Introduction. Any updates to the search after 30 October 2025 will be described at submission if applicable.

3. AI/ML Applications in Pediatric Endocrine Tumors

Throughout this section, for each entity, studies are discussed in relation to the decision they could inform if validated: (i) triage/detection; (ii) risk estimation for treatment planning and follow-up; (iii) peri-operative safety; and (iv) treatment response. Reported discrimination is noted alongside validation and calibration when available. Clinical use would depend on external pediatric testing and pre-specified action thresholds aligned with existing care pathways (Section 4; Table 2).

3.1. Differentiated Thyroid Carcinoma

Ultrasound malignancy triage: A transfer-learned ultrasound system (AI-Thyroid) evaluated in children and adolescents separated malignant from benign nodules with high discrimination and outperformed ACR-TIRADS/K-TIRADS in head-to-head comparisons. The retrospective design and non-standardized image acquisition temper enthusiasm, as do the absent calibration and decision-impact analyses (e.g., avoided biopsies). Explanatory attributions did not consistently map onto familiar sonographic characteristics, limiting bedside interpretability [51].

A second single-center series spanning children and young adults underscored the pediatric trade-off: sensitivity remained high, but specificity was modest compared to radiologists and TI-RADS, highlighting the need for predefined thresholds that account for pediatric tolerance of missed cancers and unnecessary fine-needle aspirations (FNA) [52].

In practice, any ultrasound model would need to present calibrated probabilities tied to FNA vs. observation policies and to perform robustly across vendors and protocols.

Early non-remission/recurrence: In a multi-center pediatric registry (GPOH-MET, $n = 250$), an interpretable gradient-boosted model predicted a 24-month composite of failure to achieve remission or structural recurrence with strong discrimination on an independent test split. Postoperative thyroglobulin, metastatic status at presentation, and very young age consistently shaped estimates in case-level explanations suitable for tumor-board review. Limitations include the retrospective design, Europe-centric cohort, and the absence—so far—of prospective maintenance calibration and decision-curve analyses aligned to pediatric management thresholds [53].

If validated prospectively, such tools could help individualize the extent of lymph-node dissection, the indication and activities of RAI, and surveillance intensity after initial therapy.

In adult DTC, AI/ML for ultrasound triage and nodal assessment is comparatively mature, and large recurrence/non-remission models increasingly report transparent variable effects and calibration checks. Generalizability across devices/vendors and consistent calibration remain persistent obstacles, so these studies serve mainly as methodological templates rather than evidence for pediatric deployment [54]. AI/ML studies in adult cohorts relevant to pediatric endocrine tumors are detailed in Appendix A Table A2.

Translation hinges on pediatric multi-site testing, assay- and scanner-level calibration (thyroglobulin and ultrasound), and predefined action thresholds linked to FNA, surgery, RAI, and follow-up. Entity-agnostic points on interpretability are summarized in Section 4 and Table 2.

3.2. Medullary Thyroid Carcinoma

No pediatric-only ML models for prognosis or surveillance were identified.

Adult cohorts suggest that ultrasound radiomics and combined ultrasound-plus-serology nomograms can stratify nodal risk pre-operatively [55,56], while *RET*-variant triage tools are being used as adjuncts in clinical genetics workflows [57]. Most adult series are retrospective and single- or dual-center with variable calibration reporting.

Direct pediatric use would require genotype-aware models that account for age and calcitonin kinetics, pediatric-tuned ultrasound features, and thresholds linked to MEN2-driven surgical timing and compartment planning—none of which have been shown to date.

3.3. Adrenocortical Tumors

Survival prediction (clinical features): A national pediatric registry (GPOH-MET) derived an interpretable survival model from four readily available variables (distant metastasis, tumor volume, pathologic T stage, and resection status). Discrimination was excellent on an internal test set, and individualized survival curves aided communication at the bedside. Explanations revealed non-linear effects, including a data-guided tumor-volume inflection slightly lower than conventional cut-points. The study is limited by single-registry derivation, retrospective data curation, incomplete germline information (e.g., *TP53*), and the absence of pediatric external validation with update rules [58].

Properly validated, such a parsimonious model fits rare-disease realities and supports equity by avoiding dependence on advanced assays.

Urinary steroid metabolomics (diagnosis/differentiation): A complementary analysis used supervised learning on targeted gas chromatography-mass spectrometry urinary steroid profiles to distinguish ACT from controls and to separate adrenocortical carcinoma (ACC) from adrenocortical adenoma (ACA).

The signal is intriguing but rests on internal validation only, without calibration or decision-impact evaluation, and with the usual concerns about batch and protocol effects in single-laboratory pipelines [59]. As an adjunct, this line of work will need multi-center assay harmonization and external testing before it can be fused with clinical and imaging data in pediatric pathways.

Adult ACC studies provide methodological templates—from clinical-only survival tools to multi-omics prognostics—but their resource demands and cohort structures limit portability to pediatrics without careful adaptation [60–67].

3.4. Pheochromocytoma and Paraganglioma

We found no pediatric-only AI/ML models.

Adult studies indicate that biochemical ML scores built from age, pre-test risk, and plasma metanephhrines/methoxytyramine can outperform clinicians' initial estimates. However, simply displaying probabilities to specialists did not meaningfully change final interpretations, underscoring that workflow integration and action thresholds are essential [68]. Clinical-parameter models have predicted intra-operative hemodynamic instability with encouraging accuracy and included calibration and decision-curve analyses; feature attributions emphasized inflammatory and coagulation markers [69]. Imaging signatures derived from venous-phase CT have demonstrated externally validated discrimination for metastatic potential and prognostic value for metastasis-free survival, but they were

trained in high $p \gg n$ settings with limited reporting on feature stability and calibration, warranting caution [70].

For pediatric translation, genotype-aware evaluation (*SDHB/SDHD/VHL*), harmonized protocols, and external pediatric testing are prerequisites. Screening or peri-operative models would need thresholds embedded in tumor-board workflows rather than stand-alone probability displays.

3.5. Gastroenteropancreatic Neuroendocrine Neoplasms

No pediatric-only AI/ML studies were identified.

In adults, multi-center radiomics across CT and MRI has repeatedly separated lower- from higher-grade disease and estimated nodal status with good discrimination, while lesion-level SSTR-PET radiomics for PRRT response show moderate performance that has not yet translated into patient-level benefit [71–74].

Any pediatric adoption would require harmonized reconstruction, external pediatric testing, and prospective evaluations that ask whether imaging-based stratification actually changes operative planning or PRRT selection. Equity monitoring is salient where access to SSTR-PET and PRRT is uneven.

3.6. Patient-Facing AI—Cross-Entities

We did not identify endocrine-specific, pediatric patient-facing tools. In broader pediatric oncology, small studies of general-purpose chatbots and digital navigators suggest that they can improve accessibility and task completion for basic education and logistics (appointments, fasting instructions, symptom diaries), but clinical accuracy varies and tools are unsuited to treatment advice or center selection without expert curation and escalation pathways [75].

Adjacent pediatric fields report similar patterns: symptom-triage assistants that escalate fever, pain, or nausea to nurses; pre-operative preparation and survivorship education delivered via reading-level-adaptive chat; and adherence reminders and care-coordination prompts for families who travel long distances. These prototypes typically demonstrate usability gains and knowledge recall rather than patient-level outcome changes, reinforcing the need for limited scopes, plain-language outputs, and explicit handoffs to clinicians [76–82].

Ongoing reliability efforts: To reduce error and drift, current pilots increasingly (i) ground chatbot answers in curated, locally approved pediatric content (guidelines, patient leaflets) via retrieval-augmented generation; (ii) use safety classifiers/abstention rules to block dosing or treatment recommendations and trigger escalation; (iii) implement structured intent detection (education, logistics, symptom check) with role-appropriate responses; (iv) log interactions for quality review and subgroup monitoring (language, age band); and (v) provide offline/low-bandwidth modes to support equitable access.

For pediatric endocrine translation, a scoped navigator for *MEN2/PGL/DTC* could handle scheduling, test preparation (e.g., biochemical sampling requirements), and travel letters, while abstaining from advice on dose changes or surgical timing and routing such questions to the MDT.

In adult thyroid and endocrine cancers, evaluations of LLM-based chatbots provide readable answers to common questions but show variable accuracy on management topics, frequent omissions, and no evidence of calibration or patient-level benefit. Most studies are cross-sectional and platform-specific [83–85].

3.7. Multi-Omics, Gene Expression, and Network-Based AI—Cross-Entities

No pediatric endocrine tumor studies developing or externally validating gene expression, network-based, or multi-omics ML models were found.

Adult work illustrates how expression signatures, co-expression networks, radio-genomic links, and multi-omics fusion can support subtyping, risk stratification, and nodal prediction in related endocrine tumors [86–91].

These pipelines are best viewed as methodological scaffolding that would require pediatric biospecimen harmonization, explicit batch-effect control, conservative feature spaces relative to sample size, and pediatric external validation with calibration.

A summary of the identified studies on AI/ML applications in pediatric endocrine tumors is provided in Table 1.

Table 1. AI/ML studies in pediatric endocrine tumors.

Ref./Entity	Data Modality	Task/Endpoint	Algorithms	Validation	Performance	Limitations
Ha et al. 2025 [51]/Thyroid nodules, two sites, $n = 128$	Ultrasound	Benign vs. malignant nodule classification	DL model (AI-Thyroid, transfer-learned from adult data)	Two pediatric cohorts; plane-specific testing	AUROC 0.913–0.929; sensitivity 79–89%; specificity 80–92%	Pilot; retrospective; external pediatric-only training not used
Yang et al. 2023 [52]/Thyroid nodules (children and young adults), single-center, $n = 139$	Ultrasound	Compare radiologists, ACR TI-RADS, and DL algorithm	CNN-based classifier	Internal test set	Sensitivity 87.5%; specificity 36.1% (DL model)	Mixed age band; needs external validation
Redlich et al. 2025 [53]/DTC, national registry, $n = 250$	Routine clinical + biochemical, metastasis status	Predict non-remission/recurrence within 24 months	Gradient-boosted trees (XGB) with SHAP	Stratified hold-out test set with 50 bootstrap resamples	AUROC ≈ 0.86 (test); mean ≈ 0.82 across resamples	Retrospective; needs prospective and external validation
Redlich et al. 2025 [58]/ACT, national registry, $n = 97$	Routine clinical variables	Individualized survival prediction	XGB-Cox with SHAP	Stratified train/test; 500-bootstrap estimation	C-index 0.925 (test); bootstrap mean 0.891; IBS ≈ 0.09	Retrospective; single-registry
Wudy et al. 2025 [59]/ACT, national registry, $n = 46$	Urinary steroid GC-MS metabolomics	Tumor detection (ACT vs. controls) and ACC vs. ACA differentiation	Logistic regression; decision tree; PCA/clustering (exploration)	Internal only	Not provided	Multi-center external validation needed

Abbreviations: ACA, adrenocortical adenoma; ACC, adrenocortical carcinoma; AU(RO)C, area under the (receiver operating characteristic) curve; C-index, concordance index; CNN, convolutional neural network; CV, cross-validation; DL, deep learning; DTC, differentiated thyroid carcinoma; GEP-NEN, gastroenteropancreatic neuroendocrine neoplasm; LNM, lymph node metastases; MLP, multilayer perceptron; PGL, paraganglioma; PanNET, pancreatic neuroendocrine tumors; PRRT, peptide receptor radionuclide therapy; RF, random forest; SHAP, SHapley additive explanations; SVM, support vector machine; US, ultrasound; XGB, XGBoost.

4. Methodological Guardrails for Pediatrics

Checklists help readers know what to report; they do not tell pediatric teams how to build reliable tools in rare, genotype-diverse diseases. This section distills practical guardrails for pediatric ETs and indicates where common AI reporting frameworks fit—and where they do not.

Problem formulation and endpoints: Pediatric ET decisions cluster around four domains: triage/detection (e.g., thyroid FNA vs. observation), risk estimation for treatment planning and follow-up (e.g., early non-remission in DTC; survival in pACT), peri-operative

safety (e.g., PPGL hemodynamic instability), and treatment response (e.g., imaging surrogates for grade/PRRT response). Credible studies define actionable endpoints and time-at-risk windows up front, specify inclusion/exclusion with temporality (to avoid leakage), and state thresholds aligned to existing pathways. Ambiguous composites (e.g., lesion-level signals used to infer patient-level benefit) should be avoided or clearly justified. TRIPOD-AI supports clarity for prediction models [43] and STARD-AI helps diagnostic accuracy [45], but neither dictates pediatric thresholds—these must be pre-specified with clinician input.

Data provenance, labeling, and harmonization: Label quality and site effects drive most downstream failures. For imaging, acquisition and reconstruction must be documented (ultrasound presets; CT kernel/slice thickness; MRI sequence; PET reconstruction). When radiomics is used, Image Biomarker Standardization Initiative (IBSI; [92])-conformant feature definitions and reporting of resampling/quantization are essential, and feature stability (test-retest, inter-scanner, or phantom) should be shown. For clinical/biochemical variables, pediatric reference ranges and assay variability (e.g., thyroglobulin, calcitonin) should be explicit; for omics, batch correction and cross-platform normalization are required. Harmonization methods (e.g., ComBat; [93]) must be specified. CLAIM 2024 and METRICS (radiomics) cover much of this [49,50], but ultrasound specifics and pediatric ranges often need additional local detail.

Clinician's note: For radiomics, use IBSI-conformant features and show they are stable across scanners. For labs like thyroglobulin/calcitonin, name the assay and range so risks can be compared across sites.

Small-N analysis and uncertainty: Pediatric ET cohorts are small and genetically heterogeneous (RET, SDHx, TP53). Analyses should acknowledge the $p \gg n$ regime: constrain feature spaces, prefer parsimonious or regularized models when performance permits, use learning-curve plots, and avoid optimistic single splits. Internal validation should use bootstrap or nested cross-validation with leakage safeguards. Where feasible, add temporal and geographically external tests. TRIPOD-AI and PROBAST-AI help structure these choices but do not replace transparent code/configuration sharing [43,46].

Clinician's note: In small cohorts, prefer models that keep features few and stable. A slightly lower AUROC with good calibration and transparent variables usually outperforms a complex model when you move between centers.

Calibration, thresholds, and clinical utility: Discrimination alone does not determine use. Pediatric studies should report calibration (plot/metrics; calibration-in-the-large and slope), define action thresholds linked to concrete actions (FNA, extent of surgery/RAI, alpha-blockade plan, surveillance cadence), and include decision-curve analysis using those thresholds [15,94,95]. External pediatric validation should confirm both calibration and net benefit. DECIDE-AI encourages early clinical evaluation but does not prescribe thresholds; pediatric teams must do so [48].

Clinician's note: In practice, a “10% risk” should correspond to ~10 out of 100 similar patients actually experiencing the event over the defined time horizon. When that is not true, recalibration (slope/intercept adjustment) is required before you set action thresholds.

Interpretability and human oversight: In pediatrics, explanations support verification and communication, not mechanism proof. SHAP attributes which inputs most influenced a prediction but does not ensure those inputs map to meaningful clinical constructs or generalize them. Saliency/attention maps show where a model looked, not which properties were decisive. Many radiomic textures lack stable, biologically intuitive semantics and can vary with acquisition. Practical mitigations include IBSI-conformant feature spaces, stability checks, constraining features when possible, and pre-specifying the clinical concepts explanations should reflect. Evaluation should include usefulness/appropriate-

reliance endpoints (decision quality, threshold adherence, time to decision, cognitive workload) distinct from model trustworthiness (calibration/robustness/bias). CLAIM and DECIDE-AI support reporting, but clinician-co-designed interfaces are often the missing ingredient [48,49].

Clinician's note: Treat explanations as verification aids; ask whether the top features or image regions align with recognizable clinical constructs (e.g., margins, echogenicity, microcalcifications in DTC)—and if not, pause, because attribution ≠ mechanism.

Subgroups, fairness, and safety: Performance and calibration should be reported by age bands, sex, ancestry, genotype (e.g., *SDHB*, *TP53*), and site/vendor strata. Failure-mode analyses, missing-data patterns, and safeguards against data leakage are needed. For patient-facing tools, scope should remain education/navigation with explicit escalation to clinicians. Subgroup comprehension and accessibility merit monitoring. Standards mention subgroup reporting but rarely define pediatric-relevant strata—teams must choose them a priori [43,49,96].

Regulatory posture and lifecycle: If clinical deployment is envisioned, documentation should anticipate European Union (EU) AI Act (high-risk) expectations, Food and Drug Administration (FDA)/International Medical Device Regulators Forum (IMDRF) Good ML Practice, and national device rules. That includes change-control plans (pre-determined update procedures), drift surveillance, recalibration triggers, rollback procedures, and audit trails. In Europe, leveraging ERN PaedCan/EXPeRT governance can streamline oversight and cross-border collaboration. SPIRIT-AI/CONSORT-AI cover trial reporting, but pediatric ETs often require pragmatic designs (cluster/stepped-wedge) rather than classic RCTs [44].

Where the standards apply—and where they fall short:

TRIPOD-AI/PROBAST-AI: Strong for what to report and risk-of-bias appraisal in prediction models. They do not specify pediatric action thresholds or genotype-aware subgroup sets.

CLAIM 2024: Imaging reporting is comprehensive. Pediatric ultrasound variance and center-specific presets often demand extra local detail.

STARD-AI: Useful for diagnostic accuracy, but lesion-level tasks and segmentation outputs require careful mapping to patient-level decisions.

METRICS (radiomics)/IBSI: Define reporting and features but not biological meaning or stability requirements—pediatric ETs should add test-retest/inter-scanner checks.

DECIDE-AI: Orients early clinical evaluation. It does not replace pre-specification of pediatric thresholds or utility endpoints.

SPIRIT-AI/CONSORT-AI: Trial protocols/reporting are well defined. Feasibility in very rare pediatric ETs often points to stepped-wedge/cluster designs and registry-based endpoints.

A pediatric ET-focused, at-a-glance checklist that operationalizes these points appears in Table 2.

Practical examples of these methodological guardrails (e.g., calibration and decision curves) are provided in Appendix A Boxes A1–A4.

Table 2. Methodological and clinical guardrails for AI/ML in pediatric endocrine tumors, mapped to reporting/evaluation standards.

Guardrail Topic	What It Means	Pediatric ET-Specific Application	Checklist Anchors
Problem specification, outcomes	Clearly define intended use and actionable endpoints/time-at-risk windows	e.g., DTC: 24-month non-remission/recurrence; pACT: disease-specific/overall survival horizons; PGL: intra-op instability risk; GEP-NEN: PRRT response endpoints	TRIPOD-AI, STARD-AI (diagnostic tasks), SPIRIT-AI (protocols)
Cohort construction, risk of bias	Transparent inclusion/exclusion, temporality, leakage safeguards; appraisal of bias	Exclude post-outcome variables; align imaging/biochemistry windows; report flow diagrams	TRIPOD-AI; PROBAST-AI (risk-of-bias appraisal); CLAIM
Data governance, consent	Describe consent/assent, de-identification, data use agreements, minimization of unnecessary elements	pediatric assent; family privacy; data use restrictions	CLAIM (data), SPIRIT-AI/CONSORT-AI (ethics), institutional/GDPR notes
Reference standards	Define ground truth and adjudication; report reader agreement	e.g., Thyroid nodule histology; PGL risk by Grading of Adrenal Pheochromocytoma and Paraganglioma; PRRT response definitions; centralized pathology	STARD-AI, CLAIM, TRIPOD-AI
Preprocessing, harmonization	Missing-data strategy; image normalization; batch/site correction; radiomics standards	Cross-vendor US/CT/MRI; PET recon settings; assay variability	CLAIM; METRICS (radiomics); IBSI conformance
Sample size, analysis plan	Justify size; prespecify analysis/stop rules; plan for small-N uncertainty	Rare pACT/PGL: multi-registry pooling; federated learning; learning-curve plots	TRIPOD-AI; PROBAST-AI (appraisal); DECIDE-AI (pilot evaluation)
Modeling transparency	Report algorithms, hyperparameters, versioning, and rationale	Document transfer-learning for pediatric US; share configs/code where possible	TRIPOD-AI, METRICS, CLAIM
Validation (internal, external)	Use bootstrap/nested CV; temporal split; independent multi-site tests	Train in registry A, test in registry B; temporal split around guideline changes	TRIPOD-AI; STARD-AI; DECIDE-AI (early clinical studies)
Calibration, clinical utility	Provide calibration plots/metrics; decision-curve analysis with clinical thresholds	e.g., DTC: biopsy vs. observe; pACT: adjuvant discussion; PGL: alpha-blockade intensity	TRIPOD-AI; CLAIM; DECIDE-AI; CONSORT-AI (impact)
Subgroups, fairness, safety	Prespecify subgroup analyses; report performance and calibration by subgroup; failure modes	Age bands, sex, ancestry; genotype (<i>SDHB</i> , <i>VHL</i> , <i>TP53</i>); scanner/vendor strata	TRIPOD-AI; CLAIM; SPIRIT-AI/CONSORT-AI (safety reporting)
Explainability, human-in-the-loop	Provide case-level explanations; describe clinician oversight and review points	SHAP for tabular models; heatmaps for US; pre-specified clinical concepts	TRIPOD-AI; CLAIM; DECIDE-AI (human factors)
Deployment description	Specify electronical medical record/radiology information system integration, alerting, user roles, and escalation	MDT dashboards; embargo on auto-finalization; CPMS tumor-board context	SPIRIT-AI/CONSORT-AI; DECIDE-AI
Monitoring, updates	Drift checks, recalibration, change-control plans, rollback procedures	Annual re-validation; pediatric threshold review post-guideline updates	TRIPOD-AI; CONSORT-AI; DECIDE-AI

Table 2. *Cont.*

Guardrail Topic	What It Means	Pediatric ET-Specific Application	Checklist Anchors
Data, code availability	Share de-identified/synthetic data where possible; reproducible code and model cards	Synthetic pediatric US; model cards with pediatric performance notes	TRIPOD-AI; METRICS; CLAIM
Multi-omics integration and network methods	Define fusion strategy (early, intermediate, late), batch correction, and causal/graph assumptions, document assay quality control and feature stability	e.g., DTC/MTC: integrate genotype with imaging/biochemistry; pACT: combine clinical data with urinary steroidomics; PGL: genotype-aware biochemical and imaging fusion	TRIPOD-AI; DECIDE-AI; METRICS

5. Ethics, Equity, and Patient-Facing AI

Why pediatrics is different: Children's longer life expectancy, evolving physiology, and dependence on guardians make risk–benefit trade-offs are fundamentally different from those in adult oncology. Recent pediatric ethics statements and viewpoints from the American Academy of Pediatrics emphasizes pediatric-specific governance, proportionate oversight, and the enrichment of pediatric data resources through collaboration and responsible sharing—for example, harmonized registries, age-appropriate consent/assent, and privacy-preserving analytics—so that AI systems are both safe and representative [96]. In this framing, the goal is not to collect less data *per se* but to collect the right data under robust safeguards, minimizing unnecessary elements while maximizing quality, inclusiveness, and long-term stewardship.

Patient-facing AI/LLMs—promise and pitfalls: Evaluations in pediatric oncology show that general-purpose chatbots can provide accessible information but are not adequate for treatment guidance, center selection, or nuanced counseling without expert curation and clear escalation pathways [75]. Reliability and safety can be operationalized with grounded content (answers limited to locally curated pediatric materials), hard stops for out-of-scope queries (dose changes, urgent symptoms) with automatic escalation, and appropriate-reliance metrics (comprehension checks, escalation rates, resolution times) tracked by subgroup to detect inequities. Institutions can require vendor model cards, update logs, and incident reporting and embed tools within ERN PaedCan/EXPeRT governance so consent/assent, privacy, and accessibility reviews are standardized rather than ad hoc.

Techquity—the equity lens: Digital innovations can widen disparities if poorly designed, but—as argued in recent work on pediatric and adolescents and young adult oncology—generative AI and immersive technologies can also actively reduce inequities when built and governed for techquity. Examples include multilingual, reading-level-adaptive counseling materials, culturally contextualized education co-created with families, low-bandwidth and offline delivery options, and standardized, immersive procedural preparation that narrows variation in pre-treatment information. In this framing, the actionable levers are design and measurement (co-design with under-served groups, subgroup performance and comprehension audits, continuous content localization), rather than generic calls to “ensure access”, with governance focused on documenting gaps and closing them iteratively [97].

Global and institutional governance: The World Health Organization's 2024–2025 guidance on LMMs in health frames and end-to-end, risk-managed lifecycle includes pre-deployment evaluation for clinical accuracy and harms, content provenance/labeling of AI-generated outputs, mandatory human oversight, and transparent documentation of training data sources and model limitations. It further emphasizes privacy-by-design and data minimization, bias and accessibility audits (with attention to children and guardians), and post-deployment surveillance with incident reporting and governed updates. Pro-

curement clauses (e.g., model cards, update logs, data-use restrictions, cybersecurity) are recommended to operationalize these expectations. (ISBN: 978-92-4-008475-9).

Field-applicable recommendations with illustrative cases: Translating principles into practice in pediatric ETs often comes down to scoping, governance, and measurement. Three brief scenarios illustrate how the ethics/equity guardrails can be operationalized without overpromising what AI can do.

Case A—DTC ultrasound triage: narrow scope, calibrated thresholds, and human review. In a pilot where an ultrasound model flags higher-risk thyroid nodules, outputs are limited to calibrated probabilities mapped to a pre-agreed FNA threshold set by the MDT for the local pediatric population. The model never auto-orders biopsies. Instead, a short note explains which recognizable sonographic traits (e.g., margins, echogenicity, microcalcifications) were most influential, and the radiologist/endocrinologist retains decision authority. Equity is monitored through routine reports stratified by age band, sex, and site/vendor, and a simple appeal path (MDT re-read) is available for families. This design aligns patient safety (no automation), transparency (thresholds published), and appropriate reliance (MDT sign-off).

Case B—PPGL peri-operative instability: safety-first integration. Where a clinical-parameter model estimates the risk of intra-operative hemodynamic instability, its output triggers a pre-anesthesia huddle and prompts documentation of the alpha-blockade plan. No changes occur automatically. The team agrees in advance on actions at low/indeterminate/high risk, and adverse events are tracked prospectively as part of a post-deployment safety log. To minimize bias, performance is periodically reviewed by genotype (e.g., *SDHB/SDHD/VHL*), and thresholds are re-evaluated after guideline changes. This approach embeds human oversight, genotype-aware equity, and lifecycle monitoring.

Case C—Patient-facing navigator for MEN2 families: educate, don not advise. A lightweight, multilingual navigator is scoped to education and logistics (appointments, travel letters, pre-op fasting rules) and avoids treatment advice. Content is curated from pediatric guidelines, written at adjustable reading levels, and available offline for bandwidth-constrained settings. The tool recognizes “out-of-scope” questions (e.g., whether to delay surgery) and escalates to the clinical team. Programs evaluate comprehension and trust with brief checks, track errors/omissions, and compare performance across language groups to prevent widening disparities. This keeps patient-facing AI useful while respecting limits.

Across these use cases, procurement can make expectations concrete: model cards (training data domains, pediatric performance, known limits), update logs, data use and retention terms, accessibility requirements (language, reading level, offline mode), and incident reporting pathways. Institutions can house these within existing ERN PaedCan/EXPeRT governance so that pediatric-specific assent/consent, privacy, and equity reviews are routine rather than ad hoc.

6. Roadmap for Clinical Translation in Pediatric Endocrine Tumors

Clinical translation in rare, genotype-diverse pediatric endocrine tumors will be incremental and most feasible within existing European infrastructures (EXPeRT, ERN PaedCan, Clinical Patient Management System virtual tumor boards (CPMS), PARTNER) [98–101]. Below, we outline foundations, pilot designs, and evaluation strategies using concrete, entity-specific examples. The aim is to move beyond “AI potential” toward deployable, auditable decision support.

Foundations: Rather than building pipelines from scratch, centers can map local data to minimal common elements compatible with PARTNER/EXPeRT (core clinical variables; imaging descriptors—contrast phase, slice thickness, reconstruction kernel; biochemistry with assay identifiers; genotype/variant class where available). For radiomics, report IBSI-

conformant feature definitions and stability checks (test-retest or inter-scanner). For tabular data, adopt pediatric reference ranges and units. Consent/assent text should anticipate cross-border sharing under GDPR and explicitly cover model evaluation. Within ERN PaedCan, CPMS tumor boards provide a natural venue to surface model cards (intended use, training domains, pediatric performance, limits) without automating orders or reports.

Pragmatic pilots: Pilots should target narrow questions with clear actions, use pre-specified thresholds, and log decisions and rationales.

DTC early non-remission—treatment de-escalation/escalation: Integrate a 24-month non-remission predictor into post-operative MDT review. Map probabilities to explicit actions (e.g., observation vs. compartment dissection vs. RAI activity band) agreed *ex ante*. Require a short, structured note (model estimate; top contributing factors; clinical decision) to create an audit trail.

ACT survival—risk stratification: Deploy a parsimonious four-variable survival calculator (distant metastasis, tumor volume, pT stage, resection status) as a read-only decision aid during CPMS boards. Before go-live, set a risk threshold (e.g., predicted 3-year disease-specific survival below a pre-agreed cut-point) that triggers discussion of adjuvant therapy or intensified follow-up. Monitor calibration quarterly and record whether the calculator changed the conversation (yes/no; how).

Rationale: aligns with equity (routine variables), suits small-N settings, and provides an immediately interpretable output.

PGL peri-operative instability—safety planning: Use a clinical-parameter model to stratify hemodynamic instability risk at pre-anesthesia conference. Actions are templated (alpha-blockade targets, invasive monitoring, anesthesia staffing). No automation occurs. The model simply prompts a checklist and records adherence.

GEP-NEN imaging—federated radiomics: For adolescent SSTR-PET or contrast CT/MRI tasks (grading surrogate or nodal risk), run federated/distributed training across participating sites coordinated by EXPeRT, avoiding centralization of identifiable images. Harmonize reconstruction settings up front and log site-wise performance to identify drift or bias.

Evaluation and sustainability: When signal and feasibility are shown, scale within networks using quasi-experimental designs (stepped-wedge or cluster roll-out across centers). Outcomes should be clinically anchored and entity-specific: time-to-diagnosis, reduction in avoidable FNAs (DTC) or peri-operative complications (PGL), timeliness of genetics referral, or alignment with MEN2 surgical timing rather than AUROC alone. Each deployment should include the following: (i) calibration maintenance (plots, calibration-in-the-large, slope) at planned intervals; (ii) equity dashboards (performance and calibration by age, sex, ancestry, genotype, site/vendor); (iii) drift surveillance with triggers for recalibration or rollback; and (iv) governed change-control (versioned releases, update logs, incident reporting) aligned to EU AI Act “high-risk” expectations and national device rules. Within ERN PaedCan/EXPeRT, steering groups can serve as oversight bodies for approvals and equity monitoring.

To accelerate replication, each exemplar can ship with a short pack: (1) one-page model card; (2) minimum data dictionary (fields, units, ranges); (3) threshold rationale; (4) calibration-check template; (5) CPMS note template; and (6) monitoring checklist (equity slices; drift flags). These lightweight artifacts make pilots reproducible across centers with different resources.

7. Conclusions and Future Directions

AI/ML holds credible promise across pediatric endocrine tumors, provided models are developed on harmonized data, are calibrated and interpretable, and are evaluated

prospectively within pediatric care pathways. Progress should prioritize multi-center collaboration, federated analyses, and pragmatic prospective studies with predefined actions and clinically meaningful endpoints, coupled to equity audits and lifecycle monitoring. With this disciplined approach, collaborative networks can translate retrospective signals into safe, reproducible, and equitable clinical benefit for children and adolescents.

Author Contributions: Conceptualization, M.K.; methodology, M.K.; formal analysis, M.K. and E.P.; resources, M.K. and A.R.; writing—original draft preparation, M.K.; writing—review and editing, F.H., E.P., A.R., and E.A.; visualization, F.H.; funding acquisition, M.K. and A.R. All authors have read and agreed to the published version of the manuscript.

Funding: The German MET studies were funded by the Deutsche Kinderkrebsstiftung (grants DKS 2021.11, DKS 2024.16, DKS 2025.07, and DKS 2025.16) and the Magdeburger Förderkreis krebskranker Kinder e.V. The research on pediatric adrenocortical tumors was funded by Mitteldeutsche Kinderkrebsforschung and Medical Faculty, University of Augsburg (intramurale Forschungsförderung). Mitteldeutsche Kinderkrebsforschung funded the research on pediatric endocrine tumors.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: During the preparation of this manuscript, the author(s) used OpenEvidence and ChatGPT 5 for the purposes of supplementary literature search (OpenEvidence) and language improvement (ChatGPT). The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ACA	Adrenocortical adenoma
ACC	Adrenocortical carcinoma
ACT	Adrenocortical tumor
AI	Artificial intelligence
AUROC	Area under the (receiver operating characteristic) curve
DTC	Differentiated thyroid carcinoma
ERN PaedCan	European Reference Network for Paediatric Oncology
EXPeRT	European Cooperative Study Group for Pediatric Rare Tumors
FNA	Fine-needle aspiration
GEP-NEN	Gastroenteropancreatic neuroendocrine neoplasm
IBSI	Image Biomarker Standardization Initiative
LLM	Large language models
LMM	Large multimodal models
MDT	Multidisciplinary tumor board
ML	Machine learning
MTC	Medullary thyroid carcinoma
PGL	Intra- and extra-adrenal paraganglioma
PRRT	Peptide receptor radionuclide therapy
SHAP	Shapely Additive exPlanations
XAI	Explainable artificial intelligence

Appendix A

Table A1. PubMed/MEDLINE search strategy.

Search Block	Verbatim Boolean Query (PubMed Syntax)
All endocrine entities (master query)	(“pediatric” OR child * OR adolescen *) AND ((“differentiated thyroid carcinoma” OR “papillary thyroid carcinoma” OR “follicular thyroid carcinoma” OR DTC OR PTC OR FTC OR “medullary thyroid carcinoma” OR MTC OR adrenocortical OR “adrenal cortical” OR “adrenocortical carcinoma” OR ACC OR “adrenocortical tumor **” OR pheochromocytoma OR paraganglioma OR PPGL OR (neuroendocrine AND (gastroenteropancreatic OR pancreatic OR PanNET OR “pancreatic NET” OR “pancreatic neuroendocrine” OR “small intestinal” OR “small bowel” OR midgutOR GEP))) AND (“artificial intelligence” OR “machine learning” OR “deep learning” OR radiomics OR radiogenomics OR “multi-omics” OR omics OR “neural network **” OR “support vector” OR “random forest” OR “gradient boosting” OR XGBoost OR “risk model **” OR “prediction model **” OR “survival model **”) (“pediatric” OR child * OR adolescen *) AND (thyroid AND (“differentiated thyroid carcinoma” OR “papillary thyroid carcinoma” OR “follicular thyroid carcinoma” OR DTC OR PTC OR FTC)) AND (“artificial intelligence” OR “machine learning” OR “deep learning” OR radiomics OR “risk model **” OR “prediction model **”))
DTC-focused	(“pediatric” OR child* OR adolescen *) AND (“medullary thyroid carcinoma” OR MTC) AND (“artificial intelligence” OR “machine learning” OR radiomics)
MTC-focused	(“pediatric” OR child * OR adolescen *) AND (adrenocortical OR “adrenal cortical” OR ACC OR “adrenocortical carcinoma” OR “adrenocortical tumor **”) AND (“artificial intelligence” OR “machine learning” OR radiomics OR “risk model **” OR “survival model **”))
ACT-focused	(“pediatric” OR child * OR adolescen *) AND (pheochromocytoma OR paraganglioma OR PPGL) AND (“artificial intelligence” OR “machine learning” OR radiomics)
PGL-focused	(“pediatric” OR child * OR adolescen *) AND (neuroendocrine AND (gastroenteropancreatic OR pancreatic OR PanNET OR “pancreatic NET” OR “pancreatic neuroendocrine” OR “small intestinal” OR “small bowel” OR midgut OR GEP)) AND (“artificial intelligence” OR “machine learning” OR radiomics OR radiogenomics OR “multi-omics”))
GEP-NEN-focused	

Table A2. AI/ML studies in adult cohorts with pediatric relevance.

Ref./Entity	Data Modality	Task/Endpoint	Algorithms	Validation	Performance	Limitations
Pamporaki et al. 2025 [68]/PGL, multi sites, n = 2046	Biochemical screening + age + pre-test risk	Screening/ diagnostic support: disease-probability score for PGL	ML classifiers (logistic regression/tree-based)	External, multi-site validation; comparison of specialists' pre- vs. post-score interpretations	ML scores outperformed specialists' pre-test estimates; negligible change in specialists' final interpretations	Thresholds not pre-specified; assay standardization required; minimal demonstrated clinical impact; pediatric validation needed
Zhao et al. 2025 [69]/PGL, single center, n = 197	Clinical variables, imaging features	Predict intra-op hemodynamic instability	RF, SVM, LightGBM, MLP ensembles	Internal train/test; calibration and decision-curve analyses	Best AUROC ≈ 0.86; good calibration	Etiology and physiology differ in children
Zhou et al. 2025 [70]/PGL, three sites, n = 249	CT venous-phase radiomics, DL (ResNet), clinical	Pre-op metastatic potential/high-risk (GAPP ≥ 3)	Six ML models + ResNet features; combined model	External validation across two test cohorts	AUCs > 0.87 across datasets; prognostic for MFS	Requires pediatric imaging harmonization

Table A2. *Cont.*

Ref./Entity	Data Modality	Task/Endpoint	Algorithms	Validation	Performance	Limitations
Gu et al. 2023 [71]/ PanNET, two sites, $n = 320$	Contrast-enhanced CT or MRI	Predict grade (G1–G3), LNM, or aggressiveness	DL signatures + radiomics; nomograms	External validation (multi-center)	Typical AUCs 0.85–0.93 depending on task	Prospective pediatric validation lacking
Laudicella et al. 2022 [74]/GEP-NEN, single center, $n = 38$	[⁶⁸ Ga]DOTATOC PET/CT radiomics \pm clinical	Predict response to PRRT at lesion level	Feature selection + logistic/discriminant analysis	k-fold CV; per-site analyses	AUC ~0.74–0.75 for histogram skewness; SUVmax non-predictive	Heterogeneous scanners; lesion-level outcomes

Box A1. Calibration in practice (pediatric DTC example).

Calibration links predicted risk to what actually happens. Suppose a model estimates a 24-month non-remission risk of 30% for a child after DTC surgery. In a calibrated model, among 100 similar children, about 30 will not remit by 24 months. If, on testing, the observed rate is 18%, the model is over-predicting and should not guide decisions until recalibrated (e.g., adjust intercept/slope). Report a calibration plot and simple statistics (calibration-in-the-large, slope). Decisions (e.g., extent of lymph-node dissection, RAI activity) should then be tied to pre-agreed thresholds (e.g., intervene if risk $\geq X\%$) set by the MDT.

Box A2. Decision curves and thresholds (pediatric PGL example).

Decision-curve analysis (DCA) checks whether using a model is better than “treat/test all” or “treat/test none” at clinically relevant thresholds. In PGL biochemical screening, a threshold of, say, 10% might reflect the point at which you proceed to additional imaging. Plotting net benefit across thresholds shows whether the model adds value where you actually plan to act. Report the thresholds before you run DCA and interpret the curve where the team will use it—not across the entire 0–100% range.

Box A3. Validation hierarchy.

- Internal validation (bootstrap/nested CV): guards against overfitting in the same data distribution.
- Temporal split: tests resilience to changes over time (protocols, practice).
- Geographically external testing: challenges the model with different scanners, assays, and case-mix—often where pediatric tools fail.
- Prospective evaluation/early deployment: checks usability, calibration maintenance, and safety signals.
- Impact evaluation (cluster/stepped-wedge): asks whether decisions and outcomes actually improve (e.g., fewer avoidable FNAs; fewer peri-operative complications).

Box A4. Reading explanations responsibly (SHAP, saliency, radiomics).

Use SHAP to verify that influential tabular features match clinical expectations. Be cautious when high-order radiomic textures dominate without stability evidence. For images, a saliency map that lights up a suspicious margin is reassuring; one that highlights unrelated regions warrants review. Document when explanations changed a decision (or prevented an over-reliant one).

References

- Lebbink, C.A.; Links, T.P.; Czarniecka, A.; Dias, R.P.; Elisei, R.; Izatt, L.; Krude, H.; Lorenz, K.; Luster, M.; Newbold, K.; et al. 2022 European thyroid association guidelines for the management of pediatric thyroid nodules and differentiated thyroid carcinoma. *Eur. Thyroid J.* **2022**, *11*, e220146. [[CrossRef](#)] [[PubMed](#)]
- Francis, G.L.; Waguestock, S.G.; Bauer, A.J.; Angelos, P.; Benvenga, S.; Cerutti, J.M.; Dinauer, C.A.; Hamilton, J.; Hay, I.D.; Luster, M.; et al. Management guidelines for children with thyroid nodules and differentiated thyroid cancer. *Thyroid* **2015**, *25*, 716–759. [[CrossRef](#)]

3. Kuhlen, M.; Kunstreich, M.; Redlich, A. Towards harmonised paediatric thyroid cancer care: Adult comparisons and gaps. *Endocr. Relat. Cancer* **2025**, *32*, e250191. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Waguespack, S.G.; Rich, T.A.; Perrier, N.D.; Jimenez, C.; Cote, G.J. Management of medullary thyroid carcinoma and men2 syndromes in childhood. *Nat. Rev. Endocrinol.* **2011**, *7*, 596–607. [\[CrossRef\]](#)
5. Kuhlen, M.; Fruhwald, M.C.; Dunstheimer, D.P.A.; Vorwerk, P.; Redlich, A. Revisiting the genotype-phenotype correlation in children with medullary thyroid carcinoma: A report from the gpoh-met registry. *Pediatr. Blood Cancer* **2020**, *67*, e28171. [\[CrossRef\]](#)
6. Ilanchezian, M.; Varghese, D.G.; Glod, J.W.; Reilly, K.M.; Widemann, B.C.; Pommier, Y.; Kaplan, R.N.; Del Rivero, J. Pediatric adrenocortical carcinoma. *Front. Endocrinol.* **2022**, *13*, 961650. [\[CrossRef\]](#)
7. Michalkiewicz, E.; Sandrini, R.; Figueiredo, B.; Miranda, E.C.; Caran, E.; Oliveira-Filho, A.G.; Marques, R.; Pianovski, M.A.; Lacerda, L.; Cristofani, L.M.; et al. Clinical and outcome characteristics of children with adrenocortical tumors: A report from the international pediatric adrenocortical tumor registry. *J. Clin. Oncol.* **2004**, *22*, 838–845. [\[CrossRef\]](#)
8. Redlich, A.; Pamporaki, C.; Lessel, L.; Fruhwald, M.C.; Vorwerk, P.; Kuhlen, M. Pseudohypoxic pheochromocytomas and paragangliomas dominate in children. *Pediatr. Blood Cancer* **2021**, *68*, e28981. [\[CrossRef\]](#)
9. Casey, R.T.; Hendriks, E.; Deal, C.; Waguespack, S.G.; Wiegering, V.; Redlich, A.; Akker, S.; Prasad, R.; Fassnacht, M.; Clifton-Bligh, R.; et al. International consensus statement on the diagnosis and management of phaeochromocytoma and paraganglioma in children and adolescents. *Nat. Rev. Endocrinol.* **2024**, *20*, 729–748. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Pamporaki, C.; Hamplova, B.; Peitzsch, M.; Prejbisz, A.; Beuschlein, F.; Timmers, H.; Fassnacht, M.; Klink, B.; Lodish, M.; Stratakis, C.A.; et al. Characteristics of pediatric vs adult pheochromocytomas and paragangliomas. *J. Clin. Endocrinol. Metab.* **2017**, *102*, 1122–1132. [\[CrossRef\]](#)
11. Kuo, M.J.M.; Nazari, M.A.; Jha, A.; Pacak, K. Pediatric metastatic pheochromocytoma and paraganglioma: Clinical presentation and diagnosis, genetics, and therapeutic approaches. *Front. Endocrinol.* **2022**, *13*, 936178. [\[CrossRef\]](#)
12. Virgone, C.; Roganovic, J.; Rindi, G.; Kuhlen, M.; Jamsek, J.; Panagopoulou, P.; Bajcova, V.; Ben-Ami, T.; Raphael, M.F.; Seitz, G.; et al. Appendiceal neuroendocrine tumors in children and adolescents: The european cooperative study group for pediatric rare tumors (expert) diagnostic and therapeutic recommendations. *Surgery* **2025**, *184*, 109451. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Tasto, O.; Raitio, A.; Losty, P.D. Management and outcomes of pediatric neuroendocrine tumors—A systematic review of published studies. *Eur. J. Surg. Oncol.* **2025**, *51*, 110388. [\[CrossRef\]](#)
14. Brisset, C.; Roumy, M.; Lacour, B.; Hescot, S.; Bras, M.L.; Dijoud, F.; Brisse, H.; Delehaye, F.; Desandes, E.; Philippe-Chomette, P.; et al. Bronchial carcinoid tumors in children and adolescents. *Pediatr. Blood Cancer* **2025**, *72*, e31822. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Huang, Y.; Li, W.; Macheret, F.; Gabriel, R.A.; Ohno-Machado, L. A tutorial on calibration measurements and calibration models for clinical prediction models. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 621–633. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Alba, A.C.; Agoritsas, T.; Walsh, M.; Hanna, S.; Iorio, A.; Devereaux, P.J.; McGinn, T.; Guyatt, G. Discrimination and calibration of clinical prediction models: Users' guides to the medical literature. *JAMA* **2017**, *318*, 1377–1384. [\[CrossRef\]](#)
17. Salih, A.M.; Menegaz, G.; Pillay, T.; Boyle, E.M. Explainable artificial intelligence in paediatric: Challenges for the future. *Health Sci. Rep.* **2024**, *7*, e70271. [\[CrossRef\]](#)
18. Mertes, S.; Karle, C.; Huber, T.; Weitz, K.; Schlagowski, R.; André, E. Alterfactual explanations—The relevance of irrelevance for explaining ai systems. *arXiv* **2022**, arXiv:2207.09374.
19. Mertes, S.; Huber, T.; Weitz, K.; Heimerl, A.; Andre, E. Ganterfactual-counterfactual explanations for medical non-experts using generative adversarial learning. *Front. Artif. Intell.* **2022**, *5*, 825565. [\[CrossRef\]](#)
20. Prajod, P.; Huber, T.; André, E. Using explainable ai to identify differences between clinical and experimental pain detection models based on facial expressions. In *MultiMedia Modeling*; Jónsson, B.B., Gurrin, C., Tran, M.-T., Dang-Nguyen, D.-T., Hu, A.M.-C., Thanh, B.H.T., Huet, B., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 311–322.
21. Tozzi, A.E.; Fabozzi, F.; Eckley, M.; Croci, I.; Dell, V.A.; Colantonio, E.; Mastronuzzi, A. Gaps and opportunities of artificial intelligence applications for pediatric oncology in european research: A systematic review of reviews and a bibliometric analysis. *Front. Oncol.* **2022**, *12*, 905770. [\[CrossRef\]](#)
22. Ramesh, S.; Chokkara, S.; Shen, T.; Major, A.; Volchenboum, S.L.; Mayampurath, A.; Applebaum, M.A. Applications of artificial intelligence in pediatric oncology: A systematic review. *JCO Clin. Cancer Inform.* **2021**, *5*, 1208–1219. [\[CrossRef\]](#)
23. Hassan, M.; Shahzadi, S.; Kloczkowski, A. Harnessing artificial intelligence in pediatric oncology diagnosis and treatment: A review. *Cancers* **2025**, *17*, 1828. [\[CrossRef\]](#)
24. Hashem, H.; Sultan, I. Revolutionizing precision oncology: The role of artificial intelligence in personalized pediatric cancer care. *Front. Med.* **2025**, *12*, 1555893. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Bian, C.; Wang, H.; Zhang, L.; Chen, H.; Wang, F. Integration of artificial intelligence in the clinical management of medulloblastoma: From precision diagnosis to dynamic prognosis. *Expert Rev. Neurother.* **2025**, *25*, 1411–1423. [\[CrossRef\]](#)
26. Dalboni da Rocha, J.L.; Lai, J.; Pandey, P.; Myat, P.S.M.; Loschinskey, Z.; Bag, A.K.; Sitaram, R. Artificial intelligence for neuroimaging in pediatric cancer. *Cancers* **2025**, *17*, 622. [\[CrossRef\]](#)

27. Kann, B.H.; Vossough, A.; Bruningk, S.C.; Familiar, A.M.; Aboian, M.; Linguraru, M.G.; Yeom, K.W.; Chang, S.M.; Hargrave, D.; Mirsky, D.; et al. Artificial intelligence for response assessment in pediatric neuro-oncology (ai-rapno), part 1: Review of the current state of the art. *Lancet Oncol.* **2025**, *26*, e597–e606. [\[CrossRef\]](#)

28. Kazerooni, A.F.; Familiar, A.M.; Aboian, M.; Bruningk, S.C.; Vossough, A.; Linguraru, M.G.; Huang, R.Y.; Hargrave, D.; Peet, A.C.; Resnick, A.C.; et al. Artificial intelligence for response assessment in pediatric neuro-oncology (ai-rapno), part 2: Challenges, opportunities, and recommendations for clinical translation. *Lancet Oncol.* **2025**, *26*, e607–e618. [\[CrossRef\]](#) [\[PubMed\]](#)

29. Abunadi, I.; Senan, E.M. Multi-method diagnosis of blood microscopic sample for early detection of acute lymphoblastic leukemia based on deep learning and hybrid techniques. *Sensors* **2022**, *22*, 1629. [\[CrossRef\]](#) [\[PubMed\]](#)

30. Makinen, V.P.; Rehn, J.; Breen, J.; Yeung, D.; White, D.L. Multi-cohort transcriptomic subtyping of b-cell acute lymphoblastic leukemia. *Int. J. Mol. Sci.* **2022**, *23*, 4574. [\[CrossRef\]](#)

31. Monaghan, S.A.; Li, J.L.; Liu, Y.C.; Ko, M.Y.; Boyiadzis, M.; Chang, T.Y.; Wang, Y.F.; Lee, C.C.; Swerdlow, S.H.; Ko, B.S. A machine learning approach to the classification of acute leukemias and distinction from nonneoplastic cytopenias using flow cytometry data. *Am. J. Clin. Pathol.* **2022**, *157*, 546–553. [\[CrossRef\]](#)

32. Karar, M.E.; Alotaibi, B.; Alotaibi, M. Intelligent medical iot-enabled automated microscopic image diagnosis of acute blood cancers. *Sensors* **2022**, *22*, 2348. [\[CrossRef\]](#) [\[PubMed\]](#)

33. Jawahar, M.; H, S.; L, J.A.; Gandomi, A.H. Alnett: A cluster layer deep convolutional neural network for acute lymphoblastic leukemia classification. *Comput. Biol. Med.* **2022**, *148*, 105894. [\[CrossRef\]](#) [\[PubMed\]](#)

34. Sanchez, R.; Mackenzie, S.A. Integrative network analysis of differentially methylated and expressed genes for biomarker identification in leukemia. *Sci. Rep.* **2020**, *10*, 2123. [\[CrossRef\]](#)

35. Jiang, H.; Ou, Z.; He, Y.; Yu, M.; Wu, S.; Li, G.; Zhu, J.; Zhang, R.; Wang, J.; Zheng, L.; et al. Dna methylation markers in the diagnosis and prognosis of common leukemias. *Signal Transduct. Target. Ther.* **2020**, *5*, 3. [\[CrossRef\]](#)

36. Huang, F.; Guang, P.; Li, F.; Liu, X.; Zhang, W.; Huang, W. Aml, all, and cml classification and diagnosis based on bone marrow cell morphology combined with convolutional neural network: A stard compliant diagnosis research. *Medicine* **2020**, *99*, e23154. [\[CrossRef\]](#) [\[PubMed\]](#)

37. Koelsche, C.; Schrimpf, D.; Stichel, D.; Sill, M.; Sahm, F.; Reuss, D.E.; Blattner, M.; Worst, B.; Heilig, C.E.; Beck, K.; et al. Sarcoma classification by dna methylation profiling. *Nat. Commun.* **2021**, *12*, 498. [\[CrossRef\]](#)

38. Eweje, F.R.; Bao, B.; Wu, J.; Dalal, D.; Liao, W.H.; He, Y.; Luo, Y.; Lu, S.; Zhang, P.; Peng, X.; et al. Deep learning for classification of bone lesions on routine mri. *eBioMedicine* **2021**, *68*, 103402. [\[CrossRef\]](#)

39. Pan, D.; Liu, R.; Zheng, B.; Yuan, J.; Zeng, H.; He, Z.; Luo, Z.; Qin, G.; Chen, W. Using machine learning to unravel the value of radiographic features for the classification of bone tumors. *Biomed. Res. Int.* **2021**, *2021*, 8811056. [\[CrossRef\]](#)

40. Zhang, X.; Wang, S.; Rudzinski, E.R.; Agarwal, S.; Rong, R.; Barkauskas, D.A.; Daescu, O.; Cline, L.F.; Venkatramani, R.; Xie, Y.; et al. Deep learning of rhabdomyosarcoma pathology images for classification and survival outcome prediction. *Am. J. Pathol.* **2022**, *192*, 917–925. [\[CrossRef\]](#)

41. Pfaehler, E.; van Sluis, J.; Merema, B.B.J.; van Ooijen, P.; Berendsen, R.C.M.; van Velden, F.H.P.; Boellaard, R. Experimental multicenter and multivendor evaluation of the performance of pet radiomic features using 3-dimensionally printed phantom inserts. *J. Nucl. Med.* **2020**, *61*, 469–476. [\[CrossRef\]](#)

42. Boellaard, R.; Delgado-Bolton, R.; Oyen, W.J.; Giammarile, F.; Tatsch, K.; Eschner, W.; Verzijlbergen, F.J.; Barrington, S.F.; Pike, L.C.; Weber, W.A.; et al. Fdg pet/ct: Eanm procedure guidelines for tumour imaging: Version 2.0. *Eur. J. Nucl. Med. Mol. Imaging* **2015**, *42*, 328–354. [\[CrossRef\]](#) [\[PubMed\]](#)

43. Collins, G.S.; Moons, K.G.M.; Dhiman, P.; Riley, R.D.; Beam, A.L.; Van Calster, B.; Ghassemi, M.; Liu, X.; Reitsma, J.B.; van Smeden, M.; et al. Tripod+ai statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **2024**, *385*, e078378. [\[CrossRef\]](#) [\[PubMed\]](#)

44. Ibrahim, H.; Liu, X.; Rivera, S.C.; Moher, D.; Chan, A.W.; Sydes, M.R.; Calvert, M.J.; Denniston, A.K. Reporting guidelines for clinical trials of artificial intelligence interventions: The spirit-ai and consort-ai guidelines. *Trials* **2021**, *22*, 11. [\[CrossRef\]](#) [\[PubMed\]](#)

45. Sounderajah, V.; Guni, A.; Liu, X.; Collins, G.S.; Karthikesalingam, A.; Markar, S.R.; Golub, R.M.; Denniston, A.K.; Shetty, S.; Moher, D.; et al. The stard-ai reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nat. Med.* **2025**, *31*, 3283–3289. [\[CrossRef\]](#)

46. Moons, K.G.M.; Damen, J.A.A.; Kaul, T.; Hooft, L.; Navarro, C.A.; Dhiman, P.; Beam, A.L.; Van Calster, B.; Celi, L.A.; Denaxas, S.; et al. Probast+ai: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ* **2025**, *388*, e082505. [\[CrossRef\]](#)

47. Rivera, S.C.; Liu, X.; Chan, A.W.; Denniston, A.K.; Calvert, M.J.; The SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: The spirit-ai extension. *BMJ* **2020**, *370*, m3210. [\[CrossRef\]](#)

48. Vasey, B.; Nagendran, M.; Campbell, B.; Clifton, D.A.; Collins, G.S.; Denaxas, S.; Denniston, A.K.; Faes, L.; Geerts, B.; Ibrahim, M.; et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: Decide-ai. *Nat. Med.* **2022**, *28*, 924–933. [\[CrossRef\]](#)

49. Tejani, A.S.; Klontzas, M.E.; Gatti, A.A.; Mongan, J.T.; Moy, L.; Park, S.H.; Kahn, C.E., Jr.; Panel, C.U. Checklist for artificial intelligence in medical imaging (claim): 2024 update. *Radiol. Artif. Intell.* **2024**, *6*, e240300. [\[CrossRef\]](#)

50. Kocak, B.; Akinci, D.T.; Mercaldo, N.; Alberich-Bayarri, A.; Baessler, B.; Ambrosini, I.; Andreychenko, A.E.; Bakas, S.; Beets-Tan, R.G.H.; Bressem, K.; et al. Methodological radiomics score (metrics): A quality scoring tool for radiomics research endorsed by eusomii. *Insights Imaging* **2024**, *15*, 8. [\[CrossRef\]](#)

51. Ha, E.J.; Lee, J.H.; Mak, N.; Duh, A.K.; Tong, E.; Yeom, K.W.; Meister, K.D. A deep learning-based artificial intelligence model assisting thyroid nodule diagnosis and management: Pilot results for evaluating thyroid malignancy in pediatric cohorts. *Thyroid* **2025**, *35*, 652–661. [\[CrossRef\]](#)

52. Yang, J.; Page, L.C.; Wagner, L.; Wildman-Tobriner, B.; Bisset, L.; Frush, D.; Mazurowski, M.A. Thyroid nodules on ultrasound in children and young adults: Comparison of diagnostic performance of radiologists' impressions, acr ti-rads, and a deep learning algorithm. *AJR Am. J. Roentgenol.* **2023**, *220*, 408–417. [\[CrossRef\]](#)

53. Redlich, A.; Pfaehler, E.; Kunstreich, M.; Schmutz, M.; Lapa, C.; Kuhlen, M. MI prediction of recurrence in pediatric thyroid cancer: Met cohort analysis using xgboost and shap. *J. Clin. Endocrinol. Metab.* **2025**, dgaf487. [\[CrossRef\]](#) [\[PubMed\]](#)

54. Pozdeyev, N.; White, S.L.; Bell, C.C.; Haugen, B.R.; Thomas, J. Artificial intelligence applications in thyroid cancer care. *J. Clin. Endocrinol. Metab.* **2025**, dgaf530. [\[CrossRef\]](#)

55. Lu, Q.; Zhu, X.; Li, M.; Zhan, W.; Feng, F. Ultrasound radiomics for preoperative prediction of cervical lymph node metastasis in medullary thyroid carcinoma. *Br. J. Hosp. Med.* **2025**, *86*, 1–21. [\[CrossRef\]](#)

56. Jin, Z.; Xu, L.; Chen, C.; Li, C.; Zhu, X.; Yan, Y.; Sui, L.; Xu, B.; Zheng, Y.; Chen, X.; et al. The clinical utility of ultrasound and serological features derived nomogram for the prediction of lateral lymph node metastases in medullary thyroid cancer. *Ultrasound Med. Biol.* **2025**, *51*, 1797–1804. [\[CrossRef\]](#) [\[PubMed\]](#)

57. Neocleous, V.; Fanis, P.; Skordis, N.; Phylactou, L.A. Ret proto-oncogene variants in patients with medullary thyroid carcinoma from the mediterranean basin: A brief report. *Life* **2023**, *13*, 1332. [\[CrossRef\]](#) [\[PubMed\]](#)

58. Redlich, A.; Pfaehler, E.; Kunstreich, M.; Schmutz, M.; Slavetinsky, C.; Jüttner, E.; Holterhus, P.-M.; Warncke, G.; Vokuhl, C.; Fuchs, J.; et al. Interpretable machine learning model for survival prediction in pediatric adrenocortical tumors. *J. Endocr. Soc.* **2025**, bvaf177. [\[CrossRef\]](#)

59. Wudy, S.A.; Pons-Kuhnemann, J.; Kunstreich, M.; Redlich, A.; Hartmann, M.F.; Kuhlen, M. Delineating pediatric adrenocortical tumors by gc-ms urinary steroid metabolome analysis: Observations from the met study. *J. Clin. Endocrinol. Metab.* **2025**, dgaf613. [\[CrossRef\]](#)

60. Saygili, E.S.; Elhassan, Y.S.; Prete, A.; Lippert, J.; Altieri, B.; Ronchi, C.L. Machine learning-based survival prediction tool for adrenocortical carcinoma. *J. Clin. Endocrinol. Metab.* **2025**, *110*, e3185–e3192. [\[CrossRef\]](#)

61. Martin-Hernandez, R.; Espeso-Gil, S.; Domingo, C.; Latorre, P.; Hervas, S.; Mora, J.R.H.; Kotelnikova, E. Machine learning combining multi-omics data and network algorithms identifies adrenocortical carcinoma prognostic biomarkers. *Front. Mol. Biosci.* **2023**, *10*, 1258902. [\[CrossRef\]](#)

62. Tang, J.; Fang, Y.; Xu, Z. Establishment of prognostic models of adrenocortical carcinoma using machine learning and big data. *Front. Surg.* **2022**, *9*, 966307. [\[CrossRef\]](#) [\[PubMed\]](#)

63. Yang, Y.; Wang, X.; Wu, L.; Zhao, S.; Chen, R.; Yu, G. Identification and validation of susceptibility modules and hub genes of adrenocortical carcinoma through wgcna and machine learning. *Discov. Oncol.* **2025**, *16*, 663. [\[CrossRef\]](#)

64. Vogg, N.; Muller, T.; Floren, A.; Dandekar, T.; Riester, A.; Dischinger, U.; Kurlbaum, M.; Kroiss, M.; Fassnacht, M. Simplified urinary steroid profiling by lc-ms as diagnostic tool for malignancy in adrenocortical tumors. *Clin. Chim. Acta* **2023**, *543*, 117301. [\[CrossRef\]](#) [\[PubMed\]](#)

65. Yan, X.; Guo, Z.X.; Yu, D.H.; Chen, C.; Liu, X.P.; Yang, Z.W.; Liu, T.Z.; Li, S. Identification and validation of a novel prognosis prediction model in adrenocortical carcinoma by integrative bioinformatics analysis, statistics, and machine learning. *Front. Cell Dev. Biol.* **2021**, *9*, 671359. [\[CrossRef\]](#)

66. Marquardt, A.; Landwehr, L.S.; Ronchi, C.L.; di Dalmazi, G.; Riester, A.; Kollmannsberger, P.; Altieri, B.; Fassnacht, M.; Sbiera, S. Identifying new potential biomarkers in adrenocortical tumors based on mrna expression data using machine learning. *Cancers* **2021**, *13*, 4671. [\[CrossRef\]](#)

67. Chortis, V.; Bancos, I.; Nijman, T.; Gilligan, L.C.; Taylor, A.E.; Ronchi, C.L.; O'Reilly, M.W.; Schreiner, J.; Asia, M.; Riester, A.; et al. Urine steroid metabolomics as a novel tool for detection of recurrent adrenocortical carcinoma. *J. Clin. Endocrinol. Metab.* **2020**, *105*, e307–e318. [\[CrossRef\]](#)

68. Pamporaki, C.; Pommer, G.; Apostolopoulos, I.D.; Filippatos, A.; Peitzsch, M.; Remde, H.; Constantinescu, G.; Berends, A.M.A.; Nazari, M.A.; Beuschlein, F.; et al. Utility of disease probability scores to guide decision-making during screening for phaeochromocytoma and paraganglioma: A machine learning modelling cross sectional study. *eClinicalMedicine* **2025**, *82*, 103181. [\[CrossRef\]](#)

69. Zhao, H.; Tang, L.; Li, Z.; Li, X.; Jia, T.; Luo, J.; Dong, Y.; Li, S.; Ma, X.; Zhang, P. Clinical parameters-based machine learning models for predicting intraoperative hemodynamic instability in hypertensive pheochromocytomas and paragangliomas patients. *World J. Urol.* **2025**, *43*, 555. [\[CrossRef\]](#)

70. Zhou, Y.; Zhan, Y.; Zhao, J.; Zhong, L.; Zou, F.; Zhu, X.; Zeng, Q.; Nan, J.; Gong, L.; Tan, Y.; et al. Ct-based radiomics deep learning signatures for non-invasive prediction of metastatic potential in pheochromocytoma and paraganglioma: A multicohort study. *Insights Imaging* **2025**, *16*, 81. [\[CrossRef\]](#) [\[PubMed\]](#)

71. Gu, W.; Chen, Y.; Zhu, H.; Chen, H.; Yang, Z.; Mo, S.; Zhao, H.; Chen, L.; Nakajima, T.; Yu, X.; et al. Development and validation of ct-based radiomics deep learning signatures to predict lymph node metastasis in non-functional pancreatic neuroendocrine tumors: A multicohort study. *eClinicalMedicine* **2023**, *65*, 102269. [\[CrossRef\]](#)

72. Mileva, M.; Marin, G.; Levillain, H.; Artigas, C.; Van Bogaert, C.; Marin, C.; Danieli, R.; Deleporte, A.; Picchia, S.; Stathopoulos, K.; et al. Prediction of (177)lu-dotatate prrt outcome using multimodality imaging in patients with gastroenteropancreatic neuroendocrine tumors: Results from a prospective phase ii lumen study. *J. Nucl. Med.* **2024**, *65*, 236–244. [\[CrossRef\]](#)

73. Behmanesh, B.; Abdi-Saray, A.; Deevband, M.R.; Amoui, M.; Haghigatkhah, H.R.; Shalbaf, A. Predicting the response of patients treated with (177)lu-dotatate using single-photon emission computed tomography-computed tomography image-based radiomics and clinical features. *J. Med. Signals Sens.* **2024**, *14*, 28. [\[CrossRef\]](#)

74. Laudicella, R.; Comelli, A.; Liberini, V.; Vento, A.; Stefano, A.; Spataro, A.; Croce, L.; Baldari, S.; Bambaci, M.; Deandreas, D.; et al. [(68)Ga]dotatoc pet/ct radiomics to predict the response in gep-nets undergoing [(177)lu]dotatoc prrt: The “Theragnomics” Concept. *Cancers* **2022**, *14*, 984. [\[CrossRef\]](#) [\[PubMed\]](#)

75. Clerici, C.A.; Bernasconi, A.; Lasalvia, P.; Bisogno, G.; Milano, G.M.; Trama, A.; Chiaravalli, S.; Bergamaschi, L.; Casanova, M.; Massimino, M.; et al. Being diagnosed with a rhabdomyosarcoma in the era of artificial intelligence: Whom can we trust? *Pediatr. Blood Cancer* **2024**, *71*, e31256. [\[CrossRef\]](#) [\[PubMed\]](#)

76. Min, C.; Lim, R.X.C.; Tan, S.W.; Ganapathy, S. Experience developing a pediatric medical chatbot in singapore: A digital innovation for improved emergency care. *Front. Digit. Health* **2025**, *7*, 1557804. [\[CrossRef\]](#)

77. Ganapathy, S.; Chang, S.Y.S.; Tan, J.M.C.; Lim, C.; Ng, K.C. Acute paediatrics tele-support for caregivers in singapore: An initial experience with a prototype chatbot: Upal. *Singapore Med. J.* **2023**, *64*, 335–342. [\[CrossRef\]](#)

78. Silva, E.; Pinto, P.; Reis, L.P. Communicating with children with cancer: Development of a chatbot-based educational support tool. *J. Cancer Educ.* **2025**. [\[CrossRef\]](#) [\[PubMed\]](#)

79. Bulduk, M.; Can, V.; Aktas, E.; Ipekci, B.; Bulduk, B.; Nas, I. Artificial intelligence-assisted virtual reality for reducing anxiety in pediatric endoscopy. *J. Clin. Med.* **2025**, *14*, 1344. [\[CrossRef\]](#)

80. Swallow, V.; Horsman, J.; Mazlan, E.; Campbell, F.; Zaidi, R.; Julian, M.; Branchflower, J.; Martin-Kerry, J.; Monks, H.; Soni, A.; et al. Digibete, a novel chatbot to support transition to adult care of young people/young adults with type 1 diabetes mellitus: Outcomes from a prospective, multimethod, nonrandomized feasibility and acceptability study. *JMIR Diabetes* **2025**, *10*, e74032. [\[CrossRef\]](#)

81. Boggiss, A.L.; Babbott, K.; Milford, A.; Ellett, S.; Consedine, N.; Reid, S.; Cao, N.; Cavadino, A.; Hopkins, S.; Jefferies, C.; et al. The usability and feasibility of a self-compassion chatbot (compass) for youth living with type 1 diabetes. *Diabet. Med.* **2025**, *42*, e70115. [\[CrossRef\]](#)

82. Sezgin, E.; Jackson, D.I.; Kocaballi, A.B.; Bibart, M.; Zupanec, S.; Landier, W.; Audino, A.; Ranalli, M.; Skeens, M. Can large language models aid caregivers of pediatric cancer patients in information seeking? A cross-sectional investigation. *Cancer Med.* **2025**, *14*, e70554. [\[CrossRef\]](#)

83. Gorris, M.A.; Randle, R.W.; Obermiller, C.S.; Thomas, J.; Toro-Tobon, D.; Dream, S.Y.; Fackelmayer, O.J.; Pandian, T.K.; Mayson, S.E. Assessing chatgpt’s capability in addressing thyroid cancer patient queries: A comprehensive mixed-methods evaluation. *J. Endocr. Soc.* **2025**, *9*, bvaf003. [\[CrossRef\]](#)

84. Campbell, D.J.; Estephan, L.E.; Sina, E.M.; Mastrolonardo, E.V.; Alapati, R.; Amin, D.R.; Cottrill, E.E. Evaluating chatgpt responses on thyroid nodules for patient education. *Thyroid* **2024**, *34*, 371–377. [\[CrossRef\]](#) [\[PubMed\]](#)

85. Cavnar Helvaci, B.; Hepsen, S.; Candemir, B.; Boz, O.; Durantas, H.; Houssein, M.; Cakal, E. Assessing the accuracy and reliability of chatgpt’s medical responses about thyroid cancer. *Int. J. Med. Inform.* **2024**, *191*, 105593. [\[CrossRef\]](#) [\[PubMed\]](#)

86. Drozdov, I.; Kidd, M.; Nadler, B.; Camp, R.L.; Mane, S.M.; Hauso, O.; Gustafsson, B.I.; Modlin, I.M. Predicting neuroendocrine tumor (carcinoid) neoplasia using gene expression profiling and supervised machine learning. *Cancer* **2009**, *115*, 1638–1650. [\[CrossRef\]](#)

87. Assie, G.; Giordano, T.J.; Bertherat, J. Gene expression profiling in adrenocortical neoplasia. *Mol. Cell Endocrinol.* **2012**, *351*, 111–117. [\[CrossRef\]](#)

88. Meng, K.; Hu, X.; Zheng, G.; Qian, C.; Xin, Y.; Guo, H.; He, R.; Ge, M.; Xu, J. Identification of prognostic biomarkers for papillary thyroid carcinoma by a weighted gene co-expression network analysis. *Cancer Med.* **2022**, *11*, 2006–2019. [\[CrossRef\]](#)

89. Tong, Y.; Sun, P.; Yong, J.; Zhang, H.; Huang, Y.; Guo, Y.; Yu, J.; Zhou, S.; Wang, Y.; Wang, Y.; et al. Radiogenomic analysis of papillary thyroid carcinoma for prediction of cervical lymph node metastasis: A preliminary study. *Front. Oncol.* **2021**, *11*, 682998. [[CrossRef](#)]

90. Inoue, K. Causal inference and machine learning in endocrine epidemiology. *Endocr. J.* **2024**, *71*, 945–953. [[CrossRef](#)] [[PubMed](#)]

91. Yu, Q.; Hao, W.; He, Y.; Ruan, X.; Liu, L.; Yun, X.; Li, D.; Zhao, J.; Cao, W.; Yin, Y.; et al. Multi-omics analysis unveils dysregulation of the tumor immune microenvironment and development of a machine learning-based multi-gene classifier for predicting lateral lymph node metastasis in papillary thyroid carcinoma. *Endocrine* **2025**, *90*, 172–187. [[CrossRef](#)]

92. Zwanenburg, A.; Vallieres, M.; Abdallah, M.A.; Aerts, H.; Andrearzyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beuvinga, R.J.; Boellaard, R.; et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **2020**, *295*, 328–338. [[CrossRef](#)]

93. Orlhac, F.; Eertink, J.J.; Cottreau, A.S.; Zijlstra, J.M.; Thieblemont, C.; Meignan, M.; Boellaard, R.; Buvat, I. A guide to combat harmonization of imaging biomarkers in multicenter studies. *J. Nucl. Med.* **2022**, *63*, 172–179. [[CrossRef](#)]

94. Vickers, A.J.; van Calster, B.; Steyerberg, E.W. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn. Progn. Res.* **2019**, *3*, 18. [[CrossRef](#)]

95. Van Calster, B.; McLernon, D.J.; van Smeden, M.; Wynants, L.; Steyerberg, E.W.; Topic Group ‘Evaluating Diagnostic Tests and Prediction Models’ of the STRATOS Initiative. Calibration: The achilles heel of predictive analytics. *BMC Med.* **2019**, *17*, 230. [[CrossRef](#)]

96. Johnson, K.B.; Simonian, M.; Adams, L.L.; Schneider, J.H. Toward trustworthy pediatric ai: A call to action from the national academy of medicine. *Pediatrics* **2025**, *156*, e2025073304. [[CrossRef](#)]

97. Puthenpura, V.; Hunter, M.; Marks, A.M. Techquity In pediatric, adolescent, and young adult oncology: Addressing inequities through artificial intelligence and immersive technologies. *Pediatr. Blood Cancer* **2025**, *72*, e31909. [[CrossRef](#)] [[PubMed](#)]

98. Schneider, D.T.; Ferrari, A.; Orbach, D.; Virgone, C.; Reguerre, Y.; Godzinski, J.; Bien, E.; Roganovic, J.; Farinha, N.R.; Ben-Ami, T.; et al. A virtual consultation system for very rare tumors in children and adolescents—An initiative of the european cooperative study group in rare tumors in children (expert). *EJC Paediatr. Oncol.* **2024**, *3*, 100137. [[CrossRef](#)]

99. Orbach, D.; Ferrari, A.; Schneider, D.T.; Reguerre, Y.; Godzinski, J.; Bien, E.; Stachowicz-Stencel, T.; Surun, A.; Almaraz, R.L.; Dragomir, M.; et al. The european paediatric rare tumours network—European registry (partner) project for very rare tumors in children. *Pediatr. Blood Cancer* **2021**, *68*, e29072. [[CrossRef](#)]

100. Ferrari, A.; Schneider, D.T.; Bisogno, G.; Reguerre, Y.; Godzinski, J.; Bien, E.; Stachowicz-Stencel, T.; Cecchetto, G.; Brennan, B.; Roganovic, J.; et al. Facing the challenges of very rare tumors of pediatric age: The european cooperative study group for pediatric rare tumors (expert) background, goals, and achievements. *Pediatr. Blood Cancer* **2021**, *68*, e28993. [[CrossRef](#)]

101. Roganovic, J.; Bien, E.; Ferrari, A.; Vassal, G.; Trama, A.; Casali, P.G.; Kienesberger, A.; Bisogno, G.; Virgone, C.; Ami, T.B.; et al. Solutions for optimal care and research for children and adolescents with extremely rare cancers developed within the joint action for rare cancers (jarc). *EJC Paediatr. Oncol.* **2023**, *2*, 100130. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.