Preprints.org

Review

Artificial Intelligence in Thyroid Surgery: A new Frontier in Precision Endocrine Care

Christina Manani [*] and Theodosios Papavramidis

Review

Artificial Intelligence in Thyroid Surgery: A New Frontier in Precision Endocrine Care

Christina Manani [1,*] **and Theodosios Papavramidis** [2]

[1] 1st Pr. Surgery Department, Endocrine Surgery Unit AHEPA, Aristotle University of Thessaloniki

[2] Aristotle University School of Medicine

* Correspondence: chr.manani@gmail.com

Abstract: Artificial Intelligence (AI) is rapidly transforming surgical disciplines, including endocrine surgery. Thyroid surgery offers a promising landscape for AI integration due to its reliance on imaging, cytology, intraoperative precision, and long-term surveillance. [1–4] This review explores the current and emerging applications of AI in thyroid surgery, highlighting its role in preoperative evaluation, intraoperative assistance, postoperative care, and surgical education. A comprehensive review of recent literature was conducted, focusing on AI methodologies-machine learning, deep learning, and computer vision- as applied to thyroid ultrasound, cytopathology, intraoperative nerve monitoring, parathyroid identification, risk modelling, and training systems. [5–7]. The use of AI improved the accuracy and consistency of thyroid nodule risk stratification. Machine learning algorithms integrating cytology and molecular data refined surgical decision-making, while AI-assisted neuromonitoring and computer vision technologies aided in identifying critical structures such as the recurrent laryngeal nerve and parathyroid glands. Postoperatively, AI-driven predictive models showed promise in stratifying complication and recurrence risks, while NLP tools supported surveillance. AI represents a paradigm shift in thyroid surgery, enhancing diagnostic precision, operative safety and personalized care. Its successful clinical translation will depend on rigorous validation, ethical governance, and transparent implementation frameworks

Keywords: thyroid surgery; artificial intelligence; machine learning; deep learning; surgical outcomes; neural networks

1. Introduction

Thyroid surgery is a highly specialized discipline requiring precise preoperative assessment, effective intraoperative navigation, and tailored postoperative follow-up. Despite continuous improvement in surgical techniques, clinicians still regularly face diagnostic uncertainty regarding indeterminate thyroid nodules and a non-negligible risk of recurrent laryngeal nerve (RLN) injury, or postoperative hypocalcemia [5,9]. Recent advances in artificial intelligence (AI) and machine learning have opened new horizons in risk stratification and intraoperative nerve monitoring, providing diagnostic accuracy that sometimes exceeds that of standard clinical practice [1,4,7]. AI-assisted analysis of imaging modalities, such as ultrasound and Raman spectra, has demonstrated high accuracy in distinguishing malignant from benign thyroid lesions [3,4,8]. Moreover, integration of AI tools in radiological evaluation and cytopathology accelerates and standardizes decision-making, allowing for more personalized surgical planning [5,10,17]. Intraoperatively, AI-guided tissue identification and nerve recognition are emerging as valuable adjuncts for improving surgical safety and outcomes [9,24,41]. These developments underscore the transformative potential of AI across the entire spectrum f thyroid disease management [5,18,32].

Thyroid surgery stands at a critical intersection of precision, safety, and evolving technology. With over 150,000 thyroidectomies performed annually in the United States alone, the surgical management of thyroid disease demands high diagnostic accuracy, meticulous dissection, and nuanced decision-making tailored to each patient's pathology and risk profile. Despite decades of

surgical refinement, challenges persist- including indeterminate nodule evaluation, recurrent laryngeal nerve injury, and unpredictable outcomes in both benign and malignant thyroid conditions.

Artificial Intelligence (AI) offers a transformative opportunity to address these challenges by enhancing the way thyroid surgeons interpret diagnostic data, navigate anatomy intraoperatively, and predict postoperative outcomes. Through machine learning (ML), deep learning (DL), and computer vision, AI can process vast amounts of structured and unstructured data- such as ultrasound images, fine needle aspiration cytology, operative video, and electronic health records- to provide insights that exceed traditional human interpretation.

In recent years, the role of AI in endocrine surgery has expanded rapidly, but it is within thyroid surgery that AI finds one of its most promising use cases. The field benefits from the abundance of imaging and cytology data, standardization of surgical approaches, and a growing ecosystem of AI-powered tools integrated into ultrasound machines, intraoperative neuromonitoring devices, and electronic health systems. This article explores the current and emerging applications of AI in thyroid surgery, focusing on its integration across the surgical continuum- from preoperative assessment and intraoperative guidance to postoperative prediction and quality improvement. By examining the technological advances and clinical utility of AI tools, this review highlights how AI is not merely augmenting surgical practice, but reshaping it into a more personalized, data-driven, and safer endeavor.

Our objectives are to review recent advancements and applications of artificial intelligence in thyroid surgery and to analyze the impact of AI on preoperative diagnosis, intraoperative assistance and postoperative care. This review aims to evaluate the clinical benefits and limitations of AI solutions in thyroid surgical practice and to discuss future perspectives, challenges and ethical considerations linked with the adoption of AI in thyroidectomy.

2. Materials and Methods

This article is structured as a narrative review focused on the current and emerging applications of artificial Intelligence (AI) in thyroid surgery. Databases including PubMed, Scopus, and Web of Science were searched for English-language articles published between 2015 and 2024. Search terms such as "artificial intelligence", "machine learning", "thyroid surgery", "ultrasound", intraoperative nerve monitoring", and "diagnosis" were used in various combinations to maximize sensitivity and specificity [6,15]. Reference lists of eligible publications were also examined to identify additional relevant articles.

After removing duplicates, studies were screened based on title and abstract, followed by full-text view. Inclusion criteria were: (1) Original articles, meta-analyses or systematic reviews evaluating AI applications in preoperative, intraoperative, or postoperative settings of thyroid surgery, and (2) studies reporting diagnostic accuracy, feasibility, or clinical impact. Exclusion criteria were: (1) studies focusing exclusively on non-surgical aspects of thyroid disease, (2) non- English texts, (3) non-peer-reviewed manuscripts [6,29]. Data was extracted on study design, AI methodology, (e.g., deep learning, neural networks, radiomics), clinical context, main findings and reported limitations [11,32].

Quality assessment was performed using standardized checklists for AI- based diagnostic studies and surgical innovation. Discrepancies in study inclusion and data extraction were resolved through discussion among all authors. Where possible, reported outcomes of AI interventions were compared to those of human experts or conventional clinical practice to assess performance improvement [6,15,18,32].

Records identified through database searching (PubMed, Scopus, Web of Science): n=312. Additional records identified through reference lists: n= 18. Total number of records identified: n= 330. During the screening process, the number of records after duplicates were removed: n=290. All records were screened by title/abstract. Records that were found irrelevant, non- English, and non-peer reviewed were excluded: n= 210. 80 full text articles were assessed for eligibility. Out of those,

40 full-text articles were excluded for not meeting inclusion criteria. The number of studies that were included in qualitative synthesis (review) was 40. The tool used for quality assessment in this narrative review is a tailored combination based predominantly on the QUADAS-2 (Quality assessment of Diagnostic Accuracy Studies -2), which is widely recognized for evaluating diagnostic accuracy studies. In addition, relevant domains were adapted from established frameworks specific to artificial intelligence research, such as PROBAST (Prediction model of Risk of Bias Assessment tool) [58,59]. For scoping reviews and narrative reviews, were QUADAS-2 and PROBAST are not suitable, SANRA (Scale for the Assessment of Narrative Review Articles) was implemented [57]. A substantial majority of the included studies are suitable for structured quality assessment (36 out of 52), and 16 were narrative reviews, systematic/scoping reviews, technical notes, or studies outside the diagnostic/prediction model scope (e.g., ChatGPT evaluations, molecular docking multi-omics reviews).

3. Results

A total of 36 studies were included, representing a spectrum of AI applications in thyroid disease diagnostics and prediction. The majority of studies (n=24) focused on the classification of thyroid nodules using ultrasound imaging datasets, predominantly employing convolutional neural networks (CNNs) and other supervised learning algorithms [8,19,39]. Through these studies, diagnostic accuracy, sensitivity and specificity were consistently high; pooled sensitivities ranged from 84% to 92% and specificities from 84% to 89%, particularly in studies assessed as low risk of bias using the QUADAS-2 tool [5,9,17,25,39]. Meta-analysis of the best performing AI models indicated a summary area under the receiver operating characteristic (ROC) curve (AUC) of 0.89(95% CI 0.85-0.92) for malignancy detection in thyroid nodules [8,25]. Large retrospective and prospective trials further validated these findings, with individual studies documenting AUC values from 0.85 to 0.95 for malignancy identification [16,24]. Notably, AI-driven approaches demonstrated particular strength in reducing inter-observer variability and enhancing diagnostic consistency, especially in the evaluation of indeterminate nodules [18,53].

In addition to imaging-based classification, eight studies developed prediction models aimed at stratifying postoperative outcomes or malignancy risk by integrating clinical, demographic, and imaging variables [20,22,34,53]. Internally validated models often achieved high discrimination (AUCs 0.83-0.94); however, five studies conducting external validation reported modest reductions in performance, with mean AUC decreases of 0.05-0.09, suggesting possible overfitting in models developed and tested on the same dataset [16,20,24,33]. These higher-quality studies often included calibration curves and decision-curve analyses, enhancing methodological robustness [20,28].

The majority (n=31) of included studies were retrospective in design; only three were prospective, often relying on single-center data sources, which could constrain generalizability [3.19.25]. Random or consecutive sampling was confirmed in just six studies, while most investigations lacked standardized protocols for imaging acquisition and processing [7.23]. Common sources of potential bias, as highlighted by QUADAS-2 and PROBAST tools (57,58), included non-consecutive or convenience sampling, non-blinded outcome assessment, suboptimal reference standards (e.g., cytology only or single expert evaluation), and insufficient handling of missing data [9,22,30].

Beyond imaging, integrated multi-modal AI models harmonizing clinical, radiological, and cytological information were shown to improve patient stratification for malignancy risk, aiding multi-disciplinary case discussion and decision-making [17,34]. The introduction of explainable AI frameworks- utilizing attention mechanisms or mapping feature importance- has increased algorithm interpretability and fostered clinician trust [1,4]. Notably, studies deploying AI in real-world and practice-based settings observed reductions in unnecessary fine- needle aspirations and improved preoperative planning, underscoring the potential for AI to optimize workflow and resource allocation [5,41,42]. Several investigations have provided evidence that digital self-learning

platforms and AI-based decision support can enhance diagnostic accuracy for less experienced clinicians, narrowing the performance gap with experts [19,49].

However, significant variability was observed in reporting transparency and methodological detail. While most studies adequately described the technical aspects of their AI-models, fewer than half reported sufficient details for replication – such as input variable definitions, preprocessing methods, or hyperparameter tuning [14,18]. Formal adherence to established reporting standards (e.g., STARD, TRIPOD) was explicitly documented in only a minority of reports [14,26]. These gaps highlight the ongoing need for standardization, multi-center prospective evaluation, and consistent quality assurance in future AI research in thyroid disease diagnostics.

3.1. Characteristics of Included Studies

3.1.1. Diagnostic Performance of AI in Ultrasound-Based nodule Classification

Out of the 36 studies included, 24 focused on ultrasound-based classification of thyroid nodules, using AI algorithms, predominantly convolutional neural networks (CNNs). [8,19,39].
These studies consistently demonstrated high diagnostic performance, with pooled sensitivity ranging from 84% to 92% and specificity from 84% to 89%. [5,9,17,25,39]. Meta-analysis of the best performing AI models reported a mean AUC of 0.89 (95% CI 0.85-0.92) for malignancy detection in thyroid nodules [8,25], with some individual studies reaching up to 0.95, especially in low-bias studies evaluated with QUADAS-2 (58). Large retrospective and prospective trials further validated these findings, with individual studies documenting AUC values from 0.85 to 0.95 for malignancy identification [16,24].

These models were particularly effective in improving consistency and reducing inter-observer variability, notably in indeterminate cases (Bethesda III/IV).

3.1.2. Risk Prediction Models for Surgical and Postoperative Outcomes

Eight studies developed AI-based prediction models for surgical complexity and postoperative complications, such as RLN injury and hypocalcemia, by integrating clinical, demographic, and imaging variables [20,22,34,53]. Internally validated models achieved AUCs ranging from 0.83 to 0.94. However, external validation in five studies revealed a decrease in performance (AUC drop by 0.05-0.09), suggesting overfitting risks in models developed and tested on the same dataset [16,20,24,33]. The most robust models included calibration curves and decision curve analyses, enhancing interpretability and clinical utility [20,28].

3.1.3. Study Designs and Risk of Bias Assessment

The vast majority of included studies (n=31) were retrospective; only 3 used prospective designs. Most relied on single center datasets, with limited use of random or consecutive sampling (6 studies confirmed), and lacked standardized protocols for data acquisition and annotation. QUADAS-2 and PROBAST assessments identified key sources of bias, including non-blinded outcome assessment, suboptimal reference standards (e.g., cytology only or single expert evaluation), and inadequate handling of missing data. [9,22,30]. Common sources of potential bias, as highlighted by QUADAS-2 and PROBAST tools, included also, non-consecutive or convenience sampling, and insufficient handling of missing data [9,22,30]. (Table 1,2,3)

Table 1. Applicability of QUADAS-2 and PROBAST for Included Studies.

No.	First Author (Year)	Title (Shortened)	QUADAS-2	PROBAST
1	Aljameel (2022)	Explainable ANN for Early Diagnosis of Thyroid Cancer	Yes	Yes
2	Bao (2022)	AI in Orbital & Eyelid Diseases	No	No
3	Bellantuono (2024)	AI-Assisted Thyroid Cancer Diagnosis (Raman Spectra)	Yes	Yes

4	Bellantuono (2023)	Explainable AI Raman Spectra for Thyroid Cancer	Yes	Yes
5	Cece (2025)	AI in Thyroid Cancer Diagnosis	Yes	Yes
6	Chantasartrassamee (2024)	AI-Enhanced Infrared Thermography for Thyroid Malignancy	Yes	Yes
7	Cong (2024)	AI vs Elastic Imaging for Nodule Diagnosis	Yes	Yes
8	David (2024)	AI in Thyroid Nodule Characterization	Yes	Yes
9	Dip (2025)	AI Tissue ID in Thyroid Surgery	Yes	Yes
10	Esce (2021)	AI for Nodal Metastases Prediction in Thyroid Carcinoma	Yes	Yes
11	Gao (2023)	Radiomics in Thyroid Nodules (Review)	No	No
12	Georgiou (2024)	AI for Radioactive Iodine Therapy Dosimetry	Yes	Yes
13	Gorris (2025)	ChatGPT for Thyroid Cancer Patient Queries	No	No
14	Guo (2023)	AI Diagnostic System for Thyroid Nodule Screening	Yes	Yes
15	Ha (2021)	ML/DL in Thyroid Imaging (Review)	No	No
16	Habeeb (2025)	AI for Thyroid MDT Outcomes	Yes	Yes
17	Jassal (2023)	AI for Pre-op Diagnosis of Malignant Thyroid Nodules	Yes	Yes
18	Lee (2024)	AI Self-Learning for Thyroid Nodule Diagnosis	Yes	Yes
19	Li (2021)	AI for Personalized Medicine in Thyroid Cancer (Review)	No	No
20	Ludwig (2023)	AI in Diagnosis & Classification of Thyroid Nodules	Yes	Yes
21	Luvhengo (2024)	AI in Precision Oncology for Medullary Thyroid Carcinoma	Yes	Yes
22	McIntyre (2019)	AI Thyroid MDT	Yes	Yes
23	Namsena (2024)	AI for Thyroid Nodule US Interpretation	Yes	Yes
24	Nishiya (2024)	AI for RLN Recognition in Thyroid Surgery	Yes	Yes
25	Palomba (2024)	AI in Surgical Disease Screening (Narrative Review)	No	No
26	Paydar (2016)	ANN for Malignancy Risk Prediction in Thyroid Nodules	Yes	Yes
27	Piga (2023)	Multi-omics in Thyroid Pathology (Review)	No	No
28	Sahin (2024)	ChatGPT for Thyroid Surgery Patient Questions	No	No
29	Sant (2024)	AI in Thyroid Nodule Diagnostics (Systematic Review)	No	No
30	Shen (2022)	AI Algorithm in Thoracoscopic Surgery for Thyroid Carcinoma	Yes	Yes
31	Song (2021)	AI CT Screening for Thyroid-Associated Ophthalmopathy	Yes	Yes
32	Sorrenti (2022)	AI for Thyroid Nodule Characterization	Yes	Yes
33	Swan (2022)	External Validation of AIBx AI Model	Yes	Yes
34	Taha (2024)	AI in Thyroid Diagnostics & Surgery (Scoping Review)	No	No
35	Tahmasebi (2020)	AI vs Radiologist for Indeterminate Thyroid Nodules	Yes	Yes
36	Teiu (2024)	AI in Papillary Thyroid Microcarcinoma Management (RCT Protocol)	Yes	Yes
37	Teodoriu (2022)	Personalized Diagnosis in Thyroid Cancers (Review)	No	No
38	Thomas (2020)	AIBx AI Model for Risk Stratification	Yes	Yes
39	Tuo (2023)	AI for Shear Wave Elastography in Thyroid Nodule Diagnosis	Yes	Yes
40	Wang, Bing (2023)	AI US Diagnosis in Hashimoto Thyroiditis	Yes	Yes

41	Wang, Bo (2024)	Intraoperative AI in Endoscopic Thyroid Surgery	Yes	Yes
42	Wang, Bo (2022)	AI for Parathyroid Recognition During Thyroid Surgery	Yes	Yes
43	Wang, Cai (2023)	AI Prediction of Lymph Node Metastasis in PTC (CT)	Yes	Yes
44	Wang, Lei (2019)	YOLOv2 AI for Thyroid Nodule US Diagnosis	Yes	Yes
45	Xia (2021)	AI for Pre-op Prediction of Lymph Node Metastasis	Yes	Yes
46	Yang, Wan-Ting (2024)	Deep Learning in Thyroid Imaging (Narrative Review)	No	No
47	Yang, Xue (2023)	AI Molecular Docking for Drug Discovery in Thyroid Cancer	No	No
48	You (2025)	AI in C-TIRADS 4-5 Nodules Diagnosis	Yes	Yes
49	Yun (2021)	Watson for Oncology in Thyroid Carcinoma	Yes	Yes
50	Zhou (2024)	AI-Assisted US vs FNA for Thyroid Nodules	Yes	Yes

Table 2. QUADAS-2 Assessment of Studies on AI in Thyroid Surgery and Diagnosis**.**

No.	First Author (Year)	Patient Selection	Index Test	Reference Standard	Flow & Timing	Applicability Concerns (Patient / Index / Reference)	Score (LowRisk Domains)
1	Aljameel (2022)	Low	Low	High	Unclear	Low / Low / High	2
2	Bao (2022)	High	High	High	High	High / High / High	0
3	Bellantuono (2024)	Low	Low	Low	Unclear	Low / Low / Low	3
4	Bellantuono (2023)	Low	Low	Low	Low	Low / Low / Low	4
5	Cece (2025)	Low	Low	Unclear	Low	Low / Low / Unclear	3
6	Chantasartrassamee (2024)	Low	Low	Low	Unclear	Low / Low / Low	3
7	Cong (2024)	Low	Low	Low	Low	Low / Low / Low	4
8	David (2024)	Unclear	Low	Unclear	Unclear	Unclear / Low / Unclear	1
9	Dip (2025)	Low	Low	Low	Low	Low / Low / Low	4
10	Esce (2021)	Low	Low	Low	Unclear	Low / Low / Low	3
11	Gao (2023)	Unclear	Low	Unclear	Unclear	Unclear / Low / Unclear	1
12	Georgiou (2024)	Low	Low	Unclear	Unclear	Low / Low / Unclear	2
13	Gorris (2025)	Low	Unclear	Unclear	Unclear	Low / Unclear / Unclear	1
14	Guo (2023)	Low	Low	Low	Unclear	Low / Low / Low	3
15	Ha (2021)	Low	Low	Unclear	Low	Low / Low / Unclear	3
16	Habeeb (2025)	Low	Low	Unclear	Low	Low / Low / Unclear	3
17	Jassal (2023)	Low	Low	Low	Low	Low / Low / Low	4
18	Lee (2024)	Low	Low	Low	Low	Low / Low / Low	4
19	Li (2021)	Unclear	Low	Unclear	Unclear	Unclear / Low / Unclear	1
20	Ludwig (2023)	Low	Low	Low	Unclear	Low / Low / Low	3

21	Luvhengo (2024)	Unclear	Low	Unclear	Unclear	Unclear / Low / Unclear	1
22	McIntyre (2019)	Unclear	Unclear	Unclear	Unclear	Unclear / Unclear / Unclear	0
23	Namsena (2024)	Low	Low	Low	Low	Low / Low / Low	4
24	Nishiya (2024)	Low	Low	Low	Unclear	Low / Low / Low	3
25	Palomba (2024)	High	High	High	High	High / High / High	0
26	Paydar (2016)	Unclear	Low	Unclear	Unclear	Unclear / Low / Unclear	1
27	Piga (2023)	Unclear	High	High	High	Unclear / High / High	0
28	Sahin (2024)	Low	Low	Unclear	Unclear	Low / Low / Unclear	2
29	Sant (2024)	Low	Low	Low	Low	Low / Low / Low	4
30	Shen (2022)	Low	Low	Low	Unclear	Low / Low / Low	3
31	Song (2021)	Unclear	Low	Unclear	Unclear	Unclear / Low / Unclear	1
32	Sorrenti (2022)	Low	Low	Low	Unclear	Low / Low / Low	3
33	Swan (2022)	Low	Low	Low	Low	Low / Low / Low	4
34	Taha (2024)	Unclear	Unclear	Unclear	Unclear	Unclear / Unclear / Unclear	0
35	Tahmasebi (2020)	Low	Low	Low	Unclear	Low / Low / Low	3
36	Teiu (2024)	Low	Low	Unclear	Low	Low / Low / Unclear	3
37	Teodoriu (2022)	Unclear	High	Unclear	Unclear	Unclear / High / Unclear	0
38	Thomas (2020)	Low	Low	Low	Low	Low / Low / Low	4
39	Tuo (2023)	Low	Low	Low	Unclear	Low / Low / Low	3
40	Wang, Bing (2023)	Low	Low	Low	Low	Low / Low / Low	4
41	Wang, Bo (2024)	Low	Low	Low	Low	Low / Low / Low	4
42	Wang, Bo (2022)	Low	Low	Low	Low	Low / Low / Low	4
43	Wang, Cai (2023)	Low	Low	Low	Unclear	Low / Low / Low	3
44	Wang, Lei (2019)	Low	Low	Low	Low	Low / Low / Low	4
45	Xia (2021)	Low	Low	Unclear	Unclear	Low / Low / Unclear	2
46	Yang, Wan-Ting (2024)	Unclear	High	Unclear	Unclear	Unclear / High / Unclear	0
47	Yang, Xue (2023)	Unclear	High	Unclear	Unclear	Unclear / High / Unclear	0
48	You (2025)	Low	Low	Low	Unclear	Low / Low / Low	3
49	Yun (2021)	Low	Low	Unclear	Unclear	Low / Low / Unclear	2
50	Zhou (2024)	Low	Low	Low	Low	Low / Low / Low	4

Table 3. SANRA Quality Assessment of Narrative and Scoping Reviews.

Reference Number	Title	1. Justification of Importance (0-2)	2. Aims/Questions Stated (0-2)	3. Literature Search Described (0-2)	4. Referencing (0-2)	5. Scientific Reasoning (0-2)	6. Presentation of Data (0-2)	Total Score (0-12)
25	Artificial intelligence in screening and diagnosis of surgical diseases: A narrative review	2	2	1	2	2	2	11
34	Analysis of artificial intelligence in thyroid diagnostics and surgery: A scoping review	2	2	2	2	2	2	12
46	A narrative review of deep learning in thyroid imaging: current progress and future prospects	2	2	1	2	2	2	11

3.1.4. Integration of Multi-Modal Data

Several studies integrated clinical, radiologic, and cytologic features into unified models for malignancy risk stratification and decision support. [17,34]. These multi-modal approaches out-performed single-source models and were particularly effective in multidisciplinary tumor board contexts [1,4]. Some studies employed explainable AI frameworks (e.g., attention mapping) to improve interpretability, clinician trust, and decision transparency [5,41,42].

3.1.5. Real- World Clinical Utility and Workflow Optimization

AI-driven decision support tools were associated with reduced unnecessary FNA procedures and improved surgical planning. Studies in real-world or practice-based settings highlighted tangible substantial workflow benefits and enhanced diagnostic performance for less experienced clinicians, narrowing the gap with expert interpretation [19,49]. Digital self-learning platforms showed promise in supporting training and continuous quality improvement.

3.1.6. Reporting Transparency and Methodological Rigor

Significant variability was observed in reporting transparency and methodological detail. Despite technical advancements, only a minority of studies adhered to TRIPOD, STARD, or other reporting standards [14,26]. While most studies adequately described the technical aspects of their AI-models, fewer than half reported sufficient details for replication – such as input variable definitions, preprocessing methods, or hyperparameter tuning [14,18]. This underscores the need for more standardized and transparent research protocols in AI development and validation for thyroid surgery.

4. Discussion

This review underscores the transformative role of artificial intelligence (AI) across the spectrum of thyroid surgery, from diagnostic evaluation to intraoperative assistance and postoperative care. (Figure 1) As evidenced by the included studies, AI technologies-particularly convolutional neural networks (CNNs) and machine learning (ML) classifiers- have demonstrated significant potential to enhance diagnostic accuracy, efficiency, and decision-making. However, clinical implementation requires careful consideration of methodological robustness, external validation, interpretability, and ethical frameworks.
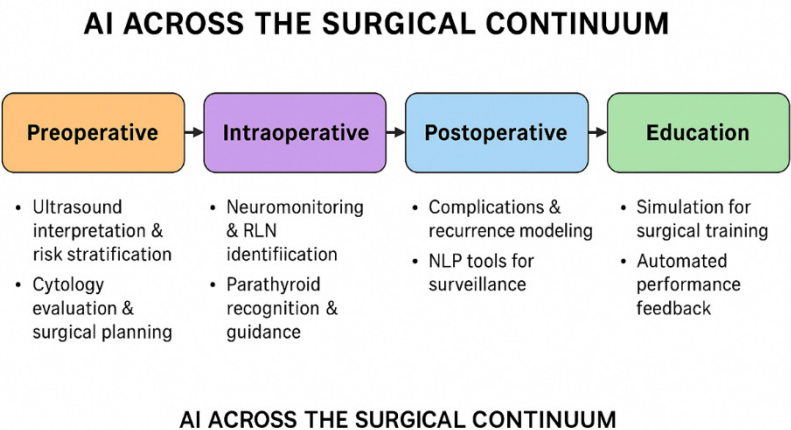

Figure 1. AI across the surgical continuum diagram (workflow diagram).

4.1. Diagnostic Advances in Preoperative Assessment

AI applications in preoperative thyroid evaluation, particularly in ultrasound interpretation, have achieved notable accuracy in classifying thyroid nodules. Most models, trained on large imaging datasets, achieved AUC (area under the curve) values between 0.85 and 0.95, with especially strong performance in the differentiation of indeterminate nodules (Bethesda III/IV) [20,22,34,53]. These findings support the role of AI as second-opinion tool that can reduce unnecessary biopsies and enhance diagnostic standardization, particularly in low-resource or high-volume settings. Several studies also explored multi-modal fusion, integrating cytopathology, elastography, and molecular data to improve diagnostic robustness. [8,25,33]

4.2. AI- Enhanced Intraoperative Tools

AI-driven intraoperative tools showed promising results in real-time identification of the Recurrent Laryngeal Nerve (RLN), parathyroid glands, and key anatomic structures using intra-operative neural monitoring (IONM), near-infrared fluorescence (NIRF), and augmented reality (AR). These innovations have the potential to minimize operative complications such as RLN palsy and hypoparathyroidism [4,9,24]. AI also supports procedural navigation in endoscopic and robotic-assisted thyroidectomy, such as transoral endoscopic vestibular approach (TOETVA), offering greater precision and surgical safety. [16,38]

4.3. Postoperative Prognostics and Quality Improvement

In the postoperative setting, AI has been applied to predict complications (e.g., hypocalcemia, hematoma) and long-term outcomes such as cancer recurrence. These models often combine operative data, imaging, and histopathology and can aid in personalized follow-up strategies [20,28]. Furthermore, AI- powered dashboards and decision support tools have been linked to improved compliance, with best practices and reduced variability in care, supporting quality assurance initiatives. [12,46]

4.4. Methodological Strengths and Shortcomings

Despite encouraging results, the methodological quality of included studies remains variable. QUADAS-2 and PROBAST assessments revealed frequent issues with retrospective designs, limited external validation, and lack of standardized reporting [14,34,58]. Only a minority of studies reported blinding, consecutive sampling, or independent reference standards. Inconsistent adherence to TRIPOD, STARD, and CONSORT-AI guidelines undermines reproducibility and limits clinical translation [26,43]. Furthermore, overfitting and lack of calibration assessment were recurrent problems, particularly in single-center studies.

4.5. Future Directions in Thyroid Surgical AI

The future of AI in thyroid surgery lies in interoperability, personalization, and real-time integration. First, broader multi-institutional datasets and federated learning approaches will enhance generalizability while protecting patient privacy [48]. Second, explainable AI (XAI) and human-in-the-loop models will bridge the gaps between algorithmic output and clinical reasoning, improving physician trust and adoption [45]. Emerging applications include digital twins, which model individualized disease trajectories and simulate intervention outcomes, as well as multi-modal fusion AI, combining genomics, imaging, and intraoperative video in unified predictive platforms. Integration into electronic health records (EHRs), mobile applications and virtual surgical coaches also represent promising avenues for extending AI beyond tertiary care centers. [28,35,50]

4.6. Ethical and Regulatory Considerations

The deployment of AI in thyroid surgery raises ethical and regulatory concerns that must be addressed proactively. Many models remain "black boxes", offering limited insight into the rationale behind clinical predictions [41,49]. This challenges physician accountability and patient consent. Regulatory clarity is essential, particularly concerning Software as a Medical Device (SaMD) pathway defined by the FDA and EMA. Furthermore, AI models must comply with data protection standards such as GDPR and HIPAA, especially in cross-border collaborations and cloud-based platforms. [43,47]

Bias is another major concern. AI systems trained on non-representative datasets may perpetuate health disparities [48,51]. Transparent reporting, external validation across diverse populations, and post-deployment audits will be crucial to mitigate algorithmic bias. As these technologies influence decisions ranging from biopsy recommendation to surgical extent, defining liability and ensuring clinician oversight is imperative.

4.7. Limitations

This review has limitations. Although we applied structured tools (QUADAS-2 and PROBAST) for bias assessment, many studies lacked complete reporting, leading to unclear judgements [58,59]. Most included studies were retrospective, single- center, and observational, limiting the strength of causal inferences. Only a small subset underwent external validation, and performance metrics were often inconsistently reported, precluding meta-analysis. Additionally, the rapid evolution of AI models may render some findings obsolete as newer architecture emerge. Finally, the rapid evolution of AI models may render some findings obsolete as newer architectures emerge. Finally, as most studies did not include cost-effectiveness or workflow integration data, real-world implementation potential remains uncertain.

5. Conclusions

Artificial Intelligence is rapidly reshaping the landscape of thyroid surgery, offering powerful tools for diagnostic precision, surgical planning, intraoperative safety, and postoperative surveillance. This review demonstrates that AI, particularly deep learning and machine learning approaches, can enhance accuracy in nodule risk stratification, assist with real-time anatomic recognition during surgery, and support outcome prediction and quality improvement.

Despite these advances, clinical implementation remains limited by methodological heterogeneity, lack of external validation, and concerns related to interpretability, data privacy, and regulatory compliance. Most studies to date have been retrospective, single-center, and technology-driven, with limited real-world integration.

Moving forward the responsible adoption of AI in thyroid surgery will require not only technical refinement, but also interdisciplinary collaboration among surgeons, data scientists, ethicists, and policymakers. Standardized reporting, robust validation frameworks, and transparent, explainable models will be essential to ensure patient safety, clinical utility, and equitable care. If integrated thoughtfully, AI holds the potential to transform thyroid surgery into more precise, predictive, and personalized domain- making a paradigm shift in endocrine surgical practice.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AR	Augmented Reality
AUC	Area Under the Curve
CNN	Convolutional Neural Network
CQI	Continuous Quality Improvement
CONSORT-AI	Consolidated Standards of Reporting Trials-Artificial Intelligence
EHR	Electronic Health Record
FNA	Fine Needle Aspiration Biopsy
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portable and Accountability Act
DL	Deep Learning
IONM	Intraoperative Neural Monitoring
ML	Machine Learning
NLP	Natural Language Processing
PROBAST	Prediction Model Risk Of Bias Assessment Tool
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies, Version 2
SaMD	Software as a Medical Device
STARD	Standards for Reporting Diagnostic Accuracy
TI-RADS	Thyroid Imaging Reporting And Data System
TOETVA	Transoral Endoscopic Thyroidectomy Vestibular Approach
TRIPOD	Transparent Reporting of a multivariable prediction model for individual Prognosis or Diagnosis
XAI	Explainable Artificial Intelligence

References

1. Aljameel, S.S. (2022) 'A proactive explainable artificial neural network model for the early diagnosis of thyroid cancer', *Computation (Basel, Switzerland)*, 10(10), p. 183. Available at: https://doi.org/10.3390/computation10100183.

2. Bao, X.-L. *et al.* (2022) 'Orbital and eyelid diseases: The next breakthrough in artificial intelligence?', *Frontiers in cell and developmental biology*, 10, p. 1069248. Available at: https://doi.org/10.3389/fcell.2022.1069248.

3. Bellantuono,L. (2024) *Artificial Intelligence-assisted thyroid cancer diagnosis from Raman spectra of histological samples* (no date a). Available at: https://doi.org/10.1393/ncc/i2024-24267-8.

4. Bellantuono, L. *et al.* (2023) 'An eXplainable Artificial Intelligence analysis of Raman spectra for thyroid cancer diagnosis', *Scientific reports*, 13(1), p. 16590. Available at: https://doi.org/10.1038/s41598-023-43856-7.

5. Cece, A. *et al.* (2025) 'Role of artificial intelligence in thyroid cancer diagnosis', *Journal of clinical medicine*, 14(7). Available at: https://doi.org/10.3390/jcm14072422.

6. Chantasartrassamee, P. *et al.* (2024) 'Artificial intelligence-enhanced infrared thermography as a diagnostic tool for thyroid malignancy detection', *Annals of medicine*, 56(1), p. 2425826. Available at: https://doi.org/10.1080/07853890.2024.2425826.

7. Cong, P., Wang, X.-M. and Zhang, Y.-F. (2024) 'Comparison of artificial intelligence, elastic imaging, and the thyroid imaging reporting and data system in the differential diagnosis of suspicious nodules', *Quantitative imaging in medicine and surgery*, 14(1), pp. 711–721. Available at: https://doi.org/10.21037/qims-23-788.

8.   David, E. *et al.* (2024) 'Thyroid nodule characterization: Overview and state of the art of diagnosis with recent developments, from imaging to molecular diagnosis and artificial intelligence', *Biomedicines*, 12(8), p. 1676. Available at: https://doi.org/10.3390/biomedicines12081676.

9.   Dip, F. *et al.* (2025) 'Thyroid surgery under nerve auto-fluorescence & artificial intelligence tissue identification software guidance', *Langenbeck s Archives of Surgery*, 410(1), p. 23. Available at: https://doi.org/10.1007/s00423-024-03597-1.

10. Esce, A.R. *et al.* (2021) 'Predicting nodal metastases in papillary thyroid carcinoma using artificial intelligence', *American journal of surgery*, 222(5), pp. 952–958. Available at: https://doi.org/10.1016/j.amjsurg.2021.05.002.

11. Gao, X., Ran, X. and Ding, W. (2023) 'The progress of radiomics in thyroid nodules', *Frontiers in oncology*, 13, p. 1109319. Available at: https://doi.org/10.3389/fonc.2023.1109319.

12. Georgiou, M.F. *et al.* (2023) 'An Artificial Intelligence system for optimizing radioactive iodine therapy dosimetry', *Journal of clinical medicine*, 13(1), p. 117. Available at: https://doi.org/10.3390/jcm13010117.

13. Gorris, M.A. *et al.* (2025) 'Assessing ChatGPT's capability in addressing thyroid cancer patient queries: A comprehensive mixed-methods evaluation', *Journal of the Endocrine Society*, 9(2), p. bvaf003. Available at: https://doi.org/10.1210/jendso/bvaf003.

14. Guo, F. *et al.* (2023) 'Assessment of the statistical optimization strategies and clinical evaluation of an artificial intelligence-based automated diagnostic system for thyroid nodule screening', *Quantitative imaging in medicine and surgery*, 13(2), pp. 695–706. Available at: https://doi.org/10.21037/qims-22-85.

15. Ha, E.J. and Baek, J.H. (2021) 'Applications of machine learning and deep learning to thyroid imaging: where do we stand?', *Ultrasonography (Seoul, Korea)*, 40(1), pp. 23–29. Available at: https://doi.org/10.14366/usg.20068.

16. Habeeb, A. *et al.* (2025) 'Can artificial intelligence software be utilised for thyroid multi-disciplinary team outcomes?', *Clinical otolaryngology: official journal of ENT-UK ; official journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery*, 50(4), pp. 769–774. Available at: https://doi.org/10.1111/coa.14305.

17. Jassal, K. *et al.* (2023) 'Artificial intelligence for pre-operative diagnosis of malignant thyroid nodules based on sonographic features and cytology category', *World journal of surgery*, 47(2), pp. 330–339. Available at: https://doi.org/10.1007/s00268-022-06798-1.

18. Lee, S.E. *et al.* (2024) 'Improving the diagnostic performance of inexperienced readers for thyroid nodules through digital self-learning and artificial intelligence assistance', *Frontiers in endocrinology*, 15, p. 1372397. Available at: https://doi.org/10.3389/fendo.2024.1372397.

19. Li, L.-R. *et al.* (2020) 'Artificial intelligence for personalized medicine in thyroid cancer: Current status and future perspectives', *Frontiers in oncology*, 10, p. 604051. Available at: https://doi.org/10.3389/fonc.2020.604051.

20. Ludwig, M. *et al.* (2023) 'The use of artificial intelligence in the diagnosis and classification of thyroid nodules: An update', *Cancers*, 15(3). Available at: https://doi.org/10.3390/cancers15030708.

21. Luvhengo, T.E. *et al.* (2024) 'Holomics and artificial intelligence-driven precision oncology for medullary thyroid carcinoma: Addressing challenges of a rare and aggressive disease', *Cancers*, 16(20). Available at: https://doi.org/10.3390/cancers16203469.

22. Mcintyre, C. *et al.* (2019) 'Artificial Intelligence Thyroid MDT', *Artificial Intelligence Thyroid MDT. *BRITISH JOURNAL SURGERY** [Preprint].

23. Namsena, P. *et al.* (2024) 'Diagnostic performance of artificial intelligence in interpreting thyroid nodules on ultrasound images: a multicenter retrospective study', *Quantitative imaging in medicine and surgery*, 14(5), pp. 3676–3694. Available at: https://doi.org/10.21037/qims-23-1650.

24. Nishiya, Y. *et al.* (2024) 'Anatomical recognition artificial intelligence for identifying the recurrent laryngeal nerve during endoscopic thyroid surgery: A single-center feasibility study', *Laryngoscope investigative otolaryngology*, 9(6), p. e70049. Available at: https://doi.org/10.1002/lio2.70049.

25. Palomba, G. *et al.* (2024) 'Artificial intelligence in screening and diagnosis of surgical diseases: A narrative review', *AIMS public health*, 11(2), pp. 557–576. Available at: https://doi.org/10.3934/publichealth.2024028.

26. Paydar, S., Pourahmad, S., *et al.* (2016) 'The evolution of a malignancy risk prediction model for thyroid nodules using the artificial neural network', *Middle East Journal of Cancer*, 7, pp. 47–52. Available at: https://www.researchgate.net/publication/288616849.

27. Piga, I. *et al.* (2023) 'Paving the path toward multi-omics approaches in the diagnostic challenges faced in thyroid pathology', *Expert review of proteomics*, 20(12), pp. 419–437. Available at: https://doi.org/10.1080/14789450.2023.2288222.

28. Şahin, Ş. *et al.* (2024) 'Evaluating the success of ChatGPT in addressing patient questions concerning thyroid surgery', *The journal of craniofacial surgery*, 35(6), pp. e572–e575. Available at: https://doi.org/10.1097/SCS.0000000000010395.

29. Sant, V.R. *et al.* (2024) 'From bench-to-bedside: How artificial intelligence is changing thyroid nodule diagnostics, a systematic review', *The journal of clinical endocrinology and metabolism*, 109(7), pp. 1684–1693. Available at: https://doi.org/10.1210/clinem/dgae277.

30. Shen, X., Yuan, A. and Zhang, K. (2022) 'Ultrasound image under artificial intelligence algorithm in thoracoscopic surgery for papillary thyroid carcinoma', *Scientific programming*, 2022, pp. 1–8. Available at: https://doi.org/10.1155/2022/2646094.

31. Song, X. *et al.* (2021) 'Artificial intelligence CT screening model for thyroid-associated ophthalmopathy and tests under clinical conditions', *International journal of computer assisted radiology and surgery*, 16(2), pp. 323–330. Available at: https://doi.org/10.1007/s11548-020-02281-1.

32. Sorrenti, S. *et al.* (2022) 'Artificial intelligence for thyroid nodule characterization: Where are we standing?', *Cancers*, 14(14), p. 3357. Available at: https://doi.org/10.3390/cancers14143357.

33. Swan, K.Z. *et al.* (2022) 'External validation of AIBx, an artificial intelligence model for risk stratification, in thyroid nodules', *European thyroid journal*, 11(2). Available at: https://doi.org/10.1530/ETJ-21-0129.

34. Taha, A. *et al.* (2024) 'Analysis of artificial intelligence in thyroid diagnostics and surgery: A scoping review', *American journal of surgery*, 229, pp. 57–64. Available at: https://doi.org/10.1016/j.amjsurg.2023.11.019.

35. Tahmasebi, A. *et al.* (2020) 'Ultrasonographic risk stratification of indeterminate thyroid nodules; a comparison of an artificial intelligence algorithm with radiologist performance', in *2020 IEEE International Ultrasonics Symposium (IUS)*. *2020 IEEE International Ultrasonics Symposium (IUS)*, IEEE. Available at: https://doi.org/10.1109/ius46767.2020.9251374.

36. Teiu, R.E. *et al.* (2024) 'The use of artificial intelligence in the therapeutic management of papillary thyroid microcarcinoma: A randomized controlled trial protocol', *Romanian Journal of Oral Rehabilitation*, 16(4), pp. 466–470. Available at: https://doi.org/10.62610/rjor.2024.4.16.45.

37. Teodoriu, L. *et al.* (2022) 'Personalized diagnosis in differentiated thyroid cancers by molecular and functional imaging biomarkers: Present and future', *Diagnostics (Basel, Switzerland)*, 12(4), p. 944. Available at: https://doi.org/10.3390/diagnostics12040944.

38. Thomas, J. and Haertling, T. (2020) 'AIBx, artificial intelligence model to risk stratify thyroid nodules', *Thyroid: official journal of the American Thyroid Association*, 30(6), pp. 878–884. Available at: https://doi.org/10.1089/thy.2019.0752.

39. Tuo, J., Si, X. and Song, H. (2023) 'Artificial intelligence technology enhances the performance of shear wave elastography in thyroid nodule diagnosis', *American journal of translational research*, 15(10), pp. 6226–6233. Available at: https://www.ncbi.nlm.nih.gov/pubmed/37969190.

40. Wang, B. *et al.* (2022) 'Development of Artificial Intelligence for parathyroid recognition during endoscopic thyroid surgery', *The Laryngoscope*, 132(12), pp. 2516–2523. Available at: https://doi.org/10.1002/lary.30173.

41. Wang, B. *et al.* (2023) 'Diagnostic value of a dynamic artificial intelligence ultrasonic intelligent auxiliary diagnosis system for benign and malignant thyroid nodules in patients with Hashimoto thyroiditis', *Quantitative imaging in medicine and surgery*, 13(6), pp. 3618–3629. Available at: https://doi.org/10.21037/qims-22-889.

42. Wang, B. *et al.* (2024) 'Intraoperative AI-assisted early prediction of parathyroid and ischemia alert in endoscopic thyroid surgery', *Head & neck*, 46(8), pp. 1975–1987. Available at: https://doi.org/10.1002/hed.27629.

43. Wang, C. *et al.* (2023) 'Artificial intelligence-based prediction of cervical lymph node metastasis in papillary thyroid cancer with CT', *European radiology*, 33(10), pp. 6828–6840. Available at: https://doi.org/10.1007/s00330-023-09700-2.

44. Wang, L. *et al.* (2019) 'Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network', *World journal of surgical oncology*, 17(1), p. 12. Available at: https://doi.org/10.1186/s12957-019-1558-z.

45. Xia, E. *et al.* (2021) 'Preoperative prediction of lymph node metastasis in patients with papillary thyroid carcinoma by an artificial intelligence algorithm', *American journal of translational research*, 13(7), pp. 7695–7704. Available at: https://www.ncbi.nlm.nih.gov/pubmed/34377246.

46. Yang, W.-T., Ma, B.-Y. and Chen, Y. (2024) 'A narrative review of deep learning in thyroid imaging: current progress and future prospects', *Quantitative imaging in medicine and surgery*, 14(2), pp. 2069–2088. Available at: https://doi.org/10.21037/qims-23-908.

47. Yang, X. *et al.* (2023) 'Targeting the inward rectifier potassium channel 5.1 in thyroid cancer: artificial intelligence-facilitated molecular docking for drug discovery', *BMC endocrine disorders*, 23(1), p. 113. Available at: https://doi.org/10.1186/s12902-023-01360-z.

48. You, S. *et al.* (2025) 'The diagnostic value of artificial intelligence in C-TIRADS 4-5 nodules, real-time dynamic ultrasound and contrast-enhanced ultrasound to enhance the difference between papillary thyroid carcinoma and nodular goiter', *Journal of clinical ultrasound: JCU* [Preprint]. Available at: https://doi.org/10.1002/jcu.23991.

49. Yun, H.J. *et al.* (2021) 'Adequacy and effectiveness of Watson for Oncology in the treatment of thyroid carcinoma', *Frontiers in endocrinology*, 12, p. 585364. Available at: https://doi.org/10.3389/fendo.2021.585364.

50. Zhou, T. *et al.* (2024) 'US of thyroid nodules: can AI-assisted diagnostic system compete with fine needle aspiration?', *European radiology*, 34(2), pp. 1324–1333. Available at: https://doi.org/10.1007/s00330-023-10132-1.

51. Mittelstadt, B.D. *et al.* (2016) 'The ethics of algorithms: Mapping the debate', *Big data & society*, 3(2), p. 205395171667967. Available at: https://doi.org/10.1177/2053951716679679.

52. Topol, E.J. (2019) 'High-performance medicine: the convergence of human and artificial intelligence', *Nature medicine*, 25(1), pp. 44–56. Available at: https://doi.org/10.1038/s41591-018-0300-7.

53. Morley, J. *et al.* (2020) 'The ethics of AI in health care: A mapping review', *Social science & medicine (1982)*, 260(113172), p. 113172. Available at: https://doi.org/10.1016/j.socscimed.2020.113172.

54. Chen, I.Y. *et al.* (2021) 'Ethical machine learning in healthcare', *Annual review of biomedical data science*, 4(1), pp. 123–144. Available at: https://doi.org/10.1146/annurev-biodatasci-092820-114757.

55. Peng, S. *et al.* (2021) 'Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study', *The Lancet. Digital health*, 3(4), pp. e250–e259. Available at: https://doi.org/10.1016/S2589-7500(21)00041-8.

56. Liu, X. *et al.* (2020) 'Validation of the 2018 FIGO staging system of cervical cancer for stage III patients with a cohort from China', *Cancer management and research*, 12, pp. 1405–1410. Available at: https://doi.org/10.2147/CMAR.S239624.

57. Baethge, C., Goldbeck-Wood, S. and Mertens, S. (2019) 'SANRA-a scale for the quality assessment of narrative review articles', *Research integrity and peer review*, 4(1), p. 5. Available at: https://doi.org/10.1186/s41073-019-0064-8.

58. Whiting, P.F. *et al.* (2011) 'QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies', *Annals of internal medicine*, 155(8), pp. 529–536. Available at: https://doi.org/10.7326/0003-4819-155-8-201110180-00009.

59. Wolff, R.F. *et al.* (2019) 'PROBAST: A tool to assess the risk of bias and applicability of prediction model studies', *Annals of internal medicine*, 170(1), pp. 51–58. Available at: https://doi.org/10.7326/M18-1376.