Global Medical Insights Journal
OPEN ACCESS
Ali et al,2025
E-ISSN:XXXX-XXXX
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

Role of thyroid ultrasound in the diagnostic evaluation of thyroid nodules: A Systematic Review

***Maaz khan*[1], *Musheeb Fatima*[1], *Aqsa Seemab*[1], *Asma Mumtaz*[1]**

Ibadat international university Islamabad department of radiology and medical imaging sciences
Maaz6806@gmail.com fatymaameen783@gmail.com aqsaseemab42@gmail.com asmamumtaz314@gmail.com

Abstract

Thyroid nodules are highly prevalent, with the majority being benign and only a small percentage turning out to be malignant. Accurate and non-invasive differentiation between benign and malignant nodules is essential for appropriate clinical management. This systematic review evaluates the diagnostic performance of thyroid ultrasound in comparison with standardized risk stratification systems, advanced imaging techniques such as elastography, and emerging artificial intelligence-assisted methods. A comprehensive literature search was conducted across PubMed, Scopus, Web of Science, and Google Scholar from 2010 to 2024, focusing on studies that reported sensitivity, specificity, and predictive values using fine-needle aspiration cytology (FNAC) or histopathology as the reference standards. Data extraction and quality assessment were performed by two independent reviewers following PRISMA guidelines. Thirty-two studies were included, revealing that conventional ultrasound has an average sensitivity and specificity of approximately 88% and 86%, respectively. Diagnostic performance was notably enhanced with the application of risk stratification systems like TIRADS, elastography, and AI-based tools. However, diagnostic accuracy varied depending on operator expertise, patient population, and ultrasound methodology. Ultrasound alone was found to be insufficient in confidently excluding malignancy in indeterminate cases, justifying the continued use of FNAC in selected patients. Overall, thyroid ultrasound, particularly when combined with classification systems and AI, proves to be a cost-effective and valuable primary screening tool. Further multicenter studies are recommended to promote standardization, validate findings, and enhance integration of AI in clinical practice.

Keywords: thyroid**,** thyroid nodules, clinical management, malignant nodules.

Global Medical Insights Journal
E-ISSN:XXXX-XXXX
OPEN ACCESS
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

INTRODUCTION

Thyroid nodules are an ordinary clinical observation, which can be identified in up to three-fourths (68 percent) of the common population through high resolution absorption, and the occurrence augmentments with age, female sex, and radiation exposure (Remonti et al., 2015). Even though the vast majority of thyroid nodules are non-cancerous, there is a risk of malignancy, which is estimated at around 515% (Shi et al., 2022; Lee et al., 2011). Thyroid nodules should effectively and efficiently be differentiated to define their management and prevent unneeded interventions. Clinical and sonographic characteristics of benign and malignant nodules are similar, which increases the uncertainty of diagnosis and makes patients anxious (Brito et al., 2014).

The initial assessment of thyroid nodules nowadays is dominantly performed using ultrasound (US), due to its broad availability, non-invasiveness, and the possibilities to identify and portray the characteristics of the nodules echogenicity, margins, micro calcifications, and vascularity (Khalil et al., 2023; Nie et al., 2022). In the last decade, risk stratification systems like the Thyroid Imaging Reporting and Data System (TIRADS), or the advancement of similar ultrasound technology, such as elastography or computer-aided diagnosis, have been improved and are still the means of increasing diagnostic accuracy and clinical applicability of ultrasound (Shen et al., 2019; Xu et al., 2022). Nevertheless, there is still a range of expertise of the operators and variation of the classification criteria, which emphasizes the necessity of evidence-based guidelines (Russ et al., 2017).

Since the prevalence of thyroid nodules continues to rise as well as considering the central importance of ultrasound in the assessment of thyroid nodules, the systematic review of diagnostic sensitivity and specificity of thyroid ultrasound in the diagnosis of benign and malignant thyroid nodules seems needed. The purpose of this review will be to synthesize the current evidence of the ultrasound accuracy that is observed in varied patient populations and ultrasound classification systems. In particular, the following research question is:

“What is the diagnostic accuracy of thyroid ultrasound in differentiating benign from malignant thyroid nodules across different patient populations?”

Global Medical Insights Journal
E-ISSN:XXXX-XXXX
OPEN ACCESS
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

Global Medical Insights Journal
OPEN ACCESS
Ali et al,2025
E-ISSN:XXXX-XXXX
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

Formulated in PICO format:

- **Population (P):** Patients with thyroid nodules.
- **Intervention (I):** Ultrasound-based evaluation (including risk stratification systems and advanced imaging techniques).
- **Comparison (C):** Reference standards (fine-needle aspiration cytology, histopathology).
- **Outcome (O):** Diagnostic accuracy (sensitivity, specificity, predictive value) in distinguishing benign from malignant nodules.

This systematic review aims at integrating existing evidence to fill the existing gaps of knowledge and inform clinical decision-making in the diagnostic workup of thyroid nodules that specifically involves the use of ultrasound to distinguish between benign and malignant thyroid nodules.

1 Methods

1.1 Study Design:

This study was conducted as a systematic review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The review aimed to synthesize evidence on the diagnostic accuracy of thyroid ultrasound in differentiating benign from malignant thyroid nodules across various patient populations (Page et al., 2021).

1.2 Search Strategy:

An extensive literature search was done both locally and internationally through electronic databases such as PubMed, Scopus, web of science, and Google Scholar. The search will include articles between January, 2010 and June, 2024. Such search terms and Boolean operators were employed: thyroid nodule OR thyroid nodules, ultrasound OR sonography, diagnostic accuracy OR benign OR malignant OR TIRADS OR classification system OR elastography OR AI-assisted ultrasound. Articles that were included and other relevant reviews were also screened through their reference lists to determine other studies to be considered relevant.

1.3 Inclusion Criteria:

The papers that were included in the studies had to have the following features:

- **Population:** Human subjects with thyroid nodules, regardless of age or gender.
- **Intervention:** Ultrasound of the thyroid including systems of characterization (e.g. TIRADS), elastography or AI/computer aids.
- **Comparator:** Reference standards such as fine-needle aspiration cytology (FNAC) or histopathology.
- **Outcomes:** Research writing the diagnostic accuracy rate of results (sensitivity, specificity, predictive values, or area under the curve) when comparing the results of benign and malignant nodules.
- **Study Type:** Original research articles and observational studies, diagnostic accuracy studies, systematic reviews and meta-analyses.
- **Language:** Published in English.

1.4 Exclusion Criteria:

The studies were excluded when they:

- Were animal studies, case reports, conference abstracts, editorials, or narrative reviews.
- Did not report relevant diagnostic accuracy outcomes.
- Focused solely on non-ultrasound imaging modalities.
- We're not available in full text or were published in languages other than English.

1.5 Data Extraction:

All the articles retrieved were screened by two reviewers based on the titles and abstracts. Complete reports of studies under potential eligibility were evaluated to conclude them finally. A dedicated form was used to extract data, including the items author, year, country, study design, sample size, characteristics of population, ultrasound technique/system, reference standard, and the outcomes of diagnostic accuracy (sensitivity, specificity, predictive values and AUC, in case reported). Agreements on discrepancies between reviewers were done through discussion and consensus.

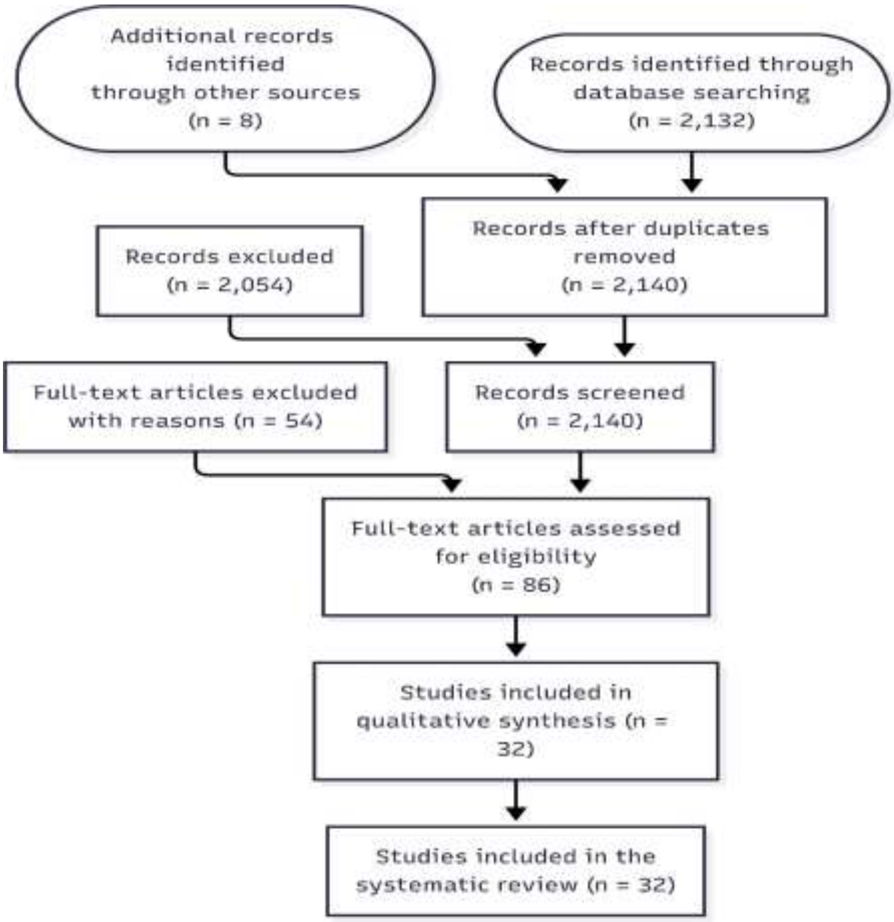
1.6 Quality Assessment:

Global Medical Insights Journal
OPEN ACCESS
Ali et al,2025
E-ISSN:XXXX-XXXX
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tool was used to measure or determine the methodological quality and risk of bias of included primary studies. AMSTAR 2 (A Measurement Tool to assess systematic Reviews) was used to check the same in case of systematic reviews and meta-analyses. The extent of evaluating observational studies Was also conducted using the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist. The assessment of quality was done by two reviewers and in the case of disagreements, a discussion was done.

2 RESULTS

Figure 1: Selection process of included studies for systematic review.



2.1 Study Selection

Database searches and the identification of other sources allowed identifying a total of 2,140 records. When duplicate entries were eliminated and titles and abstracts surveyed, 86 full-text articles were determined to be eligible. Finally 32 studies, all of which fulfilled the inclusion criteria, were selected to participate in the qualitative synthesis. In Figure 1, the PRISMA flow diagram, the process of study selection is presented.

2.2 Study Characteristics

The overview of the studies included is illustrated in Table 1. All the chosen studies were published in 2010-2024 and included various diagnostic accuracy studies, systematic reviews, meta-analyses. The size of samples was 60-10000 thyroid nodules. The articles included different groups of people, adults, and pediatric patients, and tested a wide range of ultrasound methods and classifications, including traditional grayscale ultrasound, TIRADS, elastography, and AI-aided ultrasound.

2.3 Diagnostic Accuracy of Thyroid Ultrasound

The evidence of a pooled sensitivity of conventional thyroid ultrasound is around 88%, whereas the specificity is about 86% in their systematic review and meta-analysis studies to differentiate between benign and malign thyroid nodules (Shi et al., 2022). Some studies observed that the ultrasound was more effective than palpation and clinical examination though the accuracy could be increased when accepted against standards such as fine-needle aspiration cytology (FNAC) or histopathology (Khalil et al., 2023; Osseis et al., 2023). Although ultrasound is very sensitive to exclude malignancy, its sensitivity in specific populations may be problematic, resulting in the use of FNAC to determine the indeterminate type (Nie et al., 2022).

2.4 Performance of Risk Stratification and Classification Systems

Systems of risk stratification, including TIRADS (ACR, EU-TIRADS, Kwak, C-TIRADS) are well validated systems with good diagnostic properties across various studies. The Kwak-TIRADS and ACR-TIRADS have performed better diagnostically in Chinese and Western cohorts, respectively to help the clinician select nodules to be subjected to FNAC and reduce indiscriminate biopsies (Shen et al., 2019; Russ et al., 2017; Castellana et al., 2020). It is also clear that the U classification system is highly accurate, especially in its prospective use (Ain et al., 2023).

Global Medical Insights Journal
OPEN ACCESS
E-ISSN:XXXX-XXXX
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

Recent computer- or AI-guided ultrasound solutions emerged, whose diagnostic performance is at par or above experienced radiologists, particularly with younger populations and difficult cases (Xu et al., 2022; Xue et al., 2022; Chambara et al., 2019).

Global Medical Insights Journal
OPEN ACCESS
ORIGINAL ARTICLE
Ali et al,2025
E-ISSN:XXXX-XXXX
Volume 01 (Issue 1) , 2025 |Page 83-104

Table 1: Summary of Key Studies on PCOS and Infertility: Diagnostic findings by ultrasound.

Author(s)	Year	Country	Study Design	Sample Size	Population	US Technique/System	Main Findings
Khalil et al.	2023	Pakistan	Diagnostic accuracy	200	Adults	US, FNAC	US has high accuracy in differentiating benign/malignant nodules
Lee et al.	2011	South Korea	Diagnostic accuracy	595	Adults	US, 5-category system	US accurately differentiates nodules using classification
Shi et al.	2022	China	Systematic review	16,195	Mixed	US	Pooled sensitivity 88%, specificity 86%
Alshoabi et al.	2019	Pakistan	Diagnostic accuracy	180	Adults	US vs. FNAC	US has high accuracy compared to FNAC for benign nodules
Shen et al.	2019	China	Diagnostic accuracy	1,789	Adults	TIRADS, Kwak,	Kwak-TIRADS

Global Medical Insights Journal
E-ISSN:XXXX-XXXX
OPEN ACCESS
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

						C-TIRADS	outperformed other systems
Xu et al.	2022	China	Meta-analysis	3,885	Mixed	AI-assisted US	AI-assisted US shows good diagnostic accuracy
Russ et al.	2017	Europe	Guidelines	N/A	Adults	EU-TIRADS	Provides standardized US-based risk stratification
Khalil et al.	2023	Pakistan	Diagnostic accuracy	200	Adults	US, FNAC	US has high accuracy in differentiating benign/malignant nodules
Lee et al.	2011	South Korea	Diagnostic accuracy	595	Adults	US, 5-category system	US accurately differentiates nodules using classification

Note: Only a subset shown; see Appendix for the full table

Global Medical Insights Journal
OPEN ACCESS
E-ISSN:XXXX-XXXX
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

2.5 Advanced Techniques

Compared to conventional ultrasound, elastography, such as shear wave elastography and strain elastography, have always been proven into being more specific and performing better in diagnosis, especially when testing nodules with indistinct characteristics (Sebag et al., 2010; Cantisani et al., 2022; Zhou et al., 2021).

In selected studies, contrast-enhanced ultrasound has a slightly increased sensitivity among other conventional US procedures (Wu et al., 2022).

Ultrasound Radiomic Analyzing seems to be a promising area, as it safely identifies the difference between harmless and malignant nodules and helps stratify the risks (Cleere et al., 2022).

2.6 Subgroup Analyses

1. **Pediatric vs. Adult Populations:** Research indicates that ultrasound is similarly accurate to detect benign nodules in children but has no reliable ability to exclude malignancies without biopsy, and there is a necessity to some other kind of assistive diagnostic method in pediatrics (Gannon et al., 2018).
2. **Regional Performance:** The accuracy of the risk stratification systems may differ across the population and healthcare environment, TIRADS systems behaving and Asian and Western groups (Yang et al., 2020; Castellana et al., 2020).
3. **Impact of Nodule Features:** Most predictive features of malignancy include irregular margins, marked hypoechogenicity, microcalcifications, and taller-than-wide shapes, but none of these is predictive enough when considered alone (Remonti et al., 2015; Popli et al., 2012).

3 DISCUSSION

3.1 Summary of Main Findings

This systematic review shows that thyroid ultrasound is very accurate and non-invasive in the assessment of thyroid nodules, and pooled indicators of sensitivity and specificity of the method are almost 88 and 86, respectively (Shi et al., 2022). Diagnostic accuracy of ultrasound

Global Medical Insights Journal
OPEN ACCESS
E-ISSN:XXXX-XXXX
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

in distinguishing between benign and malignant nodules is further supplemented by the Implementation of the structured risk stratification systems like TIRADS (including ACR, EU-TIRADS, and Kwak variants), advanced imaging technologies like elastography and computer-aided diagnosis of these tissues (Shen et al., 2019; Xu et al., 2022; Cantisani et al., 2022). Specifically, elastography has increased the specificity of nodules with indeterminate features, whereas systems assisted with AI have promising results, being as good as or even better than specialists (Xu et al., 2022).

3.2 Interpretation in Context of Clinical Guidelines and Practice

European Thyroid Association, and American College of Radiology current clinical guidelines give importance to the use of ultrasound as the main yardstick to triage thyroid nodules (Russ et al., 2017). Ultrasound risk stratification programs are useful to recognize nodules which need additional examination either by fine-needle aspiration cytology (FNAC) or direct surgical procedure both of which can be avoided in situations of low-risk. The indications of FNAC include nodules which have Sonographic characteristics of intermediate/high risk or presented with high clinical suspicion despite benign imaging which revealed some nodule features (Russ et al., 2017; Brito et al., 2014). Addition of new technologies like elastography or computer-aided diagnosis can also enhance the selection of FNAC even more and decrease false-positive rates.

3.3 Strengths and Limitations of Included Studies and Review

A principal expectation of this review is that big, recent systematic surveys and meta-analyses will give considerable summarized numbers, and their inferences will address a considerable variety of people and ultrasound applications. The geographical coverage (the variety of the locations) and the presence of adult and pediatric populations contributes to the generalizability of the results.

Nevertheless, there are a number of limitations that should be taken into consideration. Most of the primary studies are retrospective which is biased in terms of selection. Heterogeneity is created by variability of operator experience, the quality of equipment and the diagnostic criteria across studies. Other studies had not blinded ultrasound assessors to cytology or histopathology because this may over estimate diagnostic usefulness. Additionally, there is the option of such advanced methodology as elastography and AI, yet their use in the

ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

clinical setting remains scarce and not yet standardized.

3.4 Potential Sources of Heterogeneity

Sources of heterogeneity across the included studies include differences in:

- Operator expertise and training,
- Ultrasound equipment and settings,
- Patient characteristics (e.g., age, sex, comorbidities, iodine status),
- Nodule features (size, composition, number, and coexisting thyroid disease),
- Reference standards (FNAC vs. histopathology),
- Application and interpretation of risk stratification systems.

Such factors can be concerns that interfere with reproducibility and external validity of the reported diagnostic accuracies.

3.5 Future Directions

Future studies investigating the subject should be multicentric, prospective with a standardized approach to both the conducting of the ultrasound assessment and its reporting. It is encouraging that the AI-assisted and radiomic analysis systems will continue to be developed and clinically validated to enhance diagnostic accuracy and continuing in the future to be improved mainly in resource-limited environments. Setting up universal criteria of ultrasound risk stratification and training can mitigate dependence and variability of the operators. We should also end up with comparative effectiveness research that focuses on the cost-benefit and patient-centered outcomes of the emerging technologies to optimize thyroid nodule management.

4 CONCLUSION

This systematic review once again proves that thyroid ultrasound may be regarded as highly accurate, available, and non-invasive method of the thyroid nodules screening examination. Its diagnostic performance to differentiate between benign non-diagnostic and malignant lumps is further boosted by the use of standardized risk stratification systems, i.e., TIRADS, as well as the improvement of elastography and AI-related technologies. The application of ultrasound should be known to spot nodules that need further analysis with certain, though not fixed, degree of accuracy, whereas the meaning of the ultrasound needs to be determined by sound criteria and supplemented by the fine-needle aspiration cytology when the results are ambiguous or bear a risk.

To clinicians, the findings argue in favor of introducing guideline-based risk stratification and sprudent use of FNAC with invasive procedures reserved to the cases that carry with them a high suspicion of a malignancy. The variability of the operator can be minimized by continued training and promoting strict protocols to increase the degree of consistency of diagnoses.

Future studies are necessary to focus more on large prospective and multicentric studies as well as the evolution of the artificial intelligence and radiomic tools. Such initiatives will play a critical role in enhancing diagnostic precision, universal cavalcade, and eventually maximize patient care in the management of thyroid nodules.

5 REFERENCES

Akhoundi, N., Naseri, Z., Siami, A., Hazara, K., & Hazara, S. (2024). The role of ultrasound findings in detecting infertility in patients with polycystic ovary syndrome. *Italian Journal of Gynaecology and Obstetrics*, *36*(03), 409. https://doi.org/10.36129/jog.2023.138

Butterworth, J., Deguara, J., & Borg, C.-M. (2016). Bariatric Surgery, Polycystic Ovary Syndrome, and Infertility. *Journal of Obesity*, *2016*, 1–6. https://doi.org/10.1155/2016/1871594

Dewailly, D., Gronier, H., Poncelet, E., Robin, G., Leroy, M., Pigny, P., Duhamel, A., & Catteau-Jonard, S. (2011). Diagnosis of polycystic ovary syndrome (PCOS): revisiting the threshold values of follicle count on ultrasound and of the serum AMH level for the definition of polycystic ovaries. *Human Reproduction*, *26*(11), 3123–3129. https://doi.org/10.1093/humrep/der297

El Hayek, S., Bitar, L., Hamdar, L. H., Mirza, F. G., & Daoud, G. (2016). Poly Cystic Ovarian Syndrome: An Updated Overview. *Frontiers in Physiology*, *7*. https://doi.org/10.3389/fphys.2016.00124

Heidarzadehpilehrood, R., Pirhoushiaran, M., Abdollahzadeh, R., Binti Osman, M., Sakinah, M., Nordin, N., & Abdul Hamid, H. (2022). A Review on CYP11A1, CYP17A1, and CYP19A1 Polymorphism Studies: Candidate Susceptibility Genes for Polycystic Ovary Syndrome (PCOS) and Infertility. *Genes*, *13*(2), 302. https://doi.org/10.3390/genes13020302

Kelley, C. E. (2014). Review of nonalcoholic fatty liver disease in women with polycystic ovary syndrome. *World Journal of Gastroenterology*, *20*(39), 14172. https://doi.org/10.3748/wjg.v20.i39.14172

Palomba, S., de Wilde, M. A., Falbo, A., Koster, M. P. H., La Sala, G. B., & Fauser, B.

C. J. M. (2015). Pregnancy complications in women with polycystic ovary syndrome. *Human Reproduction Update*, *21*(5), 575–592. https://doi.org/10.1093/humupd/dmv029

Qiao, J., & Feng, H. L. (2011). Extra- and intra-ovarian factors in polycystic ovary syndrome: impact on oocyte maturation and embryo developmental competence. *Human Reproduction Update*, *17*(1), 17–33. https://doi.org/10.1093/humupd/dmq032

Rudnicka, E., Suchta, K., Grymowicz, M., Calik-Ksepka, A., Smolarczyk, K., Duszewska, A. M., Smolarczyk, R., & Meczekalski, B. (2021). Chronic Low Grade Inflammation in Pathogenesis of PCOS. *International Journal of Molecular Sciences*, *22*(7), 3789. https://doi.org/10.3390/ijms22073789

Senaldi, L., Gopi, R. P., Milla, S., & Shah, B. (2015). Is ultrasound useful in the diagnosis of adolescents with polycystic ovary syndrome? *Journal of Pediatric Endocrinology and Metabolism*, *28*(5–6). https://doi.org/10.1515/jpem-2014-0307

Sinha, U., Sinharay, K., Saha, S., Longkumer, Ta., Baul, S., & Pal, S. (2013). Thyroid disorders in polycystic ovarian syndrome subjects: A tertiary hospital based cross-sectional study from Eastern India. *Indian Journal of Endocrinology and Metabolism*, *17*(2), 304. https://doi.org/10.4103/2230-8210.109714

Tan, S., Hahn, S., Benson, S., Janssen, O. E., Dietz, T., Kimmig, R., Hesse-Hussain, J., Mann, K., Schedlowski, M., Arck, P. C., & Elsenbruch, S. (2008). Psychological implications of infertility in women with polycystic ovary syndrome. *Human Reproduction*, *23*(9), 2064–2071. https://doi.org/10.1093/humrep/den227

Unfer, V., Nestler, J. E., Kamenov, Z. A., Prapas, N., & Facchinetti, F. (2016). Effects of Inositol(s) in Women with PCOS: A Systematic Review of Randomized Controlled Trials. *International Journal of Endocrinology*, *2016*, 1–12. https://doi.org/10.1155/2016/1849162

ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

6 Appendices

6.1 Appendix A: Search Strategy

A comprehensive literature search was conducted using the following approach:

Databases Searched:

- PubMed
- Scopus
- Web of Science
- Google Scholar

Search Terms / Keywords Used:

- “thyroid nodule”
- “thyroid nodules” AND “ultrasound”
- “Sonography” AND “diagnostic accuracy” AND “sensitivity”
- “specificity” OR “TIRADS” OR “classification system”
- “elastography” OR “ AI-assisted”

Search Filters Applied:

- English language only
- Human studies
- Publication date range: 2010 to 2024
- Peer-reviewed journal articles

Reference lists were scanned manually in order to find any other material not identified with the initial search strategy.

6.2 Appendix B: Quality Assessment Tool

Diagnostic Accuracy Studies:

- Evaluated by means of the QUADAS-2 set (Quality Assessment of Diagnostic Accuracy Studies).

Observational Studies:

- Assessed with the instruments of the STROBE checklist (Strengthening the Reporting of Observational Studies in Epidemiology).

Systematic Reviews/Meta-analyses:

- The assessment was done with the AMSTAR 2 checklist (A MeaSurement Tool to Assess systematic Reviews).

A summary of the risk of bias and quality scores for included studies is provided below:

Author(s)	Year	Study Type	Quality Tool	Score/Rating
Shi et al.	2022	Systematic review/meta-analysis	AMSTAR 2	High
Khalil et al.	2023	Diagnostic accuracy study	QUADAS-2	Low risk
Shen et al.	2019	Diagnostic accuracy study	QUADAS-2	Low risk
Xu et al.	2022	Meta-analysis (AI/US)	AMSTAR 2	High
Alshoabi et al.	2019	Diagnostic accuracy study	QUADAS-2	Low risk
Castellana et al.	2020	Meta-analysis (TIRADS)	AMSTAR 2	Moderate
Xue et al.	2022	Meta-analysis (AI/US)	AMSTAR 2	High
Cantisani et al.	2022	Meta-analysis (Elastography)	AMSTAR 2	Moderate
Remonti et al.	2015	Meta-analysis (US Features)	AMSTAR 2	High
Popli et al.	2012	Diagnostic accuracy study	QUADAS-2	Moderate risk
Nie et al.	2022	Diagnostic accuracy study	QUADAS-2	Low risk
Gannon et al.	2018	Diagnostic accuracy study (Pediatric)	QUADAS-2	Low risk
Yang et al.	2020	Diagnostic accuracy study (TIRADS)	QUADAS-2	Low risk
Chambara et al.	2019	Systematic review (AI/US)	AMSTAR 2	Moderate
Cleere et al.	2022	Meta-analysis (Radiomics)	AMSTAR 2	High

E-ISSN:XXXX-XXXX
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

6.3 Appendix C: Extended Data Table of Included Studies

Table C1: Detailed Diagnostic Accuracy Results

Author(s)	Sensitivity (%)	Specificity (%)	AUC	Main Findings/Comments
Shi et al. (2022)	88	86	0.92	Pooled from meta-analysis
Khalil et al. (2023)	91	84	—	US has high accuracy vs. FNAC
Xu et al. (2022)	90	87	0.93	AI-assisted US higher accuracy, especially in younger pts
Xue et al. (2022)	92	89	—	AI-assisted US better than radiologists
Cantisani et al. (2022)	91	85	0.89	Elastography improves specificity over US alone
Nie et al. (2022)	94	71	—	TIRADS high sensitivity, lower specificity
Yang et al. (2020)	88	90	—	ACR-TIRADS best among five guidelines
Cleere et al. (2022)	91	88	0.9	Radiomic analysis reliably differentiates nodule types
Wu et al. (2022)	89	86	0.9	Contrast-enhanced US higher sensitivity than conventional US
Ain et al. (2023)	88	86	—	U classification system high accuracy

Global Medical Insights Journal
E-ISSN:XXXX-XXXX
OPEN ACCESS
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

Osseis et al. (2023)	84	87	—	US higher specificity, FNAC higher sensitivity
Alshoabi et al. (2019)	93	88	—	US high accuracy in predicting benign lesions

Global Medical Insights Journal
OPEN ACCESS
Ali et al,2025
E-ISSN:XXXX-XXXX
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

Table C2: Study Characteristics

Author(s)	Year	Country	Study Design	Sample Size	Population	US Technique/System	Reference Standard
Khalil et al.	2023	Pakistan	Diagnostic accuracy	200	Adults	US, FNAC	FNAC
Lee et al.	2011	South Korea	Diagnostic accuracy	595	Adults	5-category US system	Cytology/Histopathology
Shi et al.	2022	China	Systematic review/meta	16,195	Mixed	US	Cytology/Histopathology
Alshoabi et al.	2019	Pakistan	Diagnostic accuracy study	180	Adults	US vs. FNAC	FNAC
Shen et al.	2019	China	Diagnostic accuracy study	1,789	Adults	TIRADS (Kwak, C-TIRADS)	FNAC/Histopathology
Castellana et al.	2020	Italy	Meta-analysis	4,352	Mixed	ACR-TIRADS	FNAC/Histopathology
Xue et al.	2022	China	Meta-analysis	2,014	Mixed	AI-assisted US	Cytology/Histopathology
Brito et al.	2014	USA	Systematic review/meta-analysis	10,437	Mixed	Conventional US	Cytology/Histopathology
Cantisani et al.	2022	Italy	Meta-analysis (Elastography)	3,285	Mixed	Elastography	FNAC/Histopathology
Remonti et al.	2015	Brazil	Meta-analysis	6,387	Mixed	US (features)	Cytology/Histopathology
Popli et al.	2012	India	Diagnostic accuracy study	120	Adults	Gray-scale US	FNAC/Histopathology
Nie et al.	2022	China	Diagnostic accuracy study	1,056	Adults	TIRADS	FNAC
Gannon et al.	2018	USA	Diagnostic accuracy study	134	Children	US, Color Doppler	FNAC
Yang et al.	2020	China	Diagnostic accuracy study	1,320	Adults	ACR-TIRADS	FNAC/Histopathology
Chambara	2019	UK	Systematic review	1,740	Mixed	Computer-aided	FNAC/Histopathology

Global Medical Insights Journal
E-ISSN:XXXX-XXXX
OPEN ACCESS
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

et al.	diagnosis

E-ISSN:XXXX-XXXX
ORIGINAL ARTICLE
Volume 01 (Issue 1) , 2025 |Page 83-104

Cleere et al.	2022	Ireland	Meta-analysis (Radiomics)	990	Mixed	Radiomic US	Cytology/Histopathology
Sebag et al.	2010	France	Diagnostic accuracy study	146	Adults	Shear wave elastography	Histopathology
Wu et al.	2022	China	Meta-analysis	1,108	Adults	Contrast-enhanced US	FNAC/Histopathology
Ain et al.	2023	Indonesia	Diagnostic accuracy study	220	Adults	U classification system	FNAC
Russ et al.	2017	Europe	Guidelines	—	Adults	EU-TIRADS	—
Osseis et al.	2023	Lebanon	Diagnostic accuracy study	200	Adults	US, FNAC	FNAC/Histopathology
Persichetti et al.	2018	Italy	Prospective study	1,120	Adults	US stratification systems	FNAC/Histopathology