World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

Systematic Integration of Artificial Intelligence and Machine Learning in the Early Detection and Management of Goitre: A Global Epidemiological and Computational Framework

Kingdom Mutala Akugri *, Prince Agbenyo, Marious Akugri and Lovelyn Keteku

Independent Researcher.

World Journal of Advanced Research and Reviews, 2026, 29(01), 1719-1725

Publication history: Received on 21 December 2025; revised on 26 January 2026; accepted on 29 January 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.29.1.0235

Abstract

Background: Goitre remains a high-signal global health indicator of thyroid dysfunction and population-level iodine status. Despite progress in salt iodization, early-stage thyroid enlargement is frequently under-detected in routine practice, especially when physical examination is confounded by body habitus and clinician subjectivity. Ultrasound is the preferred modality for early assessment, but interpretation is operator-dependent and increasingly burdened by rising thyroid nodule prevalence.

Objective: This review synthesizes evidence on Machine Learning (ML) and Artificial Intelligence (AI) methods for predicting thyroid dysfunction and diagnosing early goitre (WHO Grade 1), with a practical emphasis on multi-modal "holistic AI" systems that combine tabular laboratory markers with imaging features.

Methods: We summarize (i) supervised learning pipelines for structured clinical data (e.g., age, sex, TSH, T3, T4, T4U/FTI), (ii) deep learning architectures for ultrasound-based detection and segmentation (CNNs, U-Net variants, Vision Transformers), and (iii) deployment considerations including explainability, bias control, and reproducible benchmarking using open datasets. Following common clinical ML reporting practice, we emphasize confusion-matrixbased evaluation (precision/recall/F1/MCC) and strong ensemble baselines for tabular prediction. [30,31]

Results: For tabular prediction tasks, stacked ensembles and gradient-boosted trees repeatedly rank among the best-performing approaches, particularly when combined with careful feature engineering and imbalance mitigation. For imaging, segmentation-first pipelines that estimate thyroid volume (e.g., U-Net family) and classification models leveraging multi-channel inputs or self-attention mechanisms (e.g., ViTs) report high diagnostic performance in differentiating benign enlargement from suspicious nodular patterns. Emerging smartphone-assisted workflows and LLM-based clinical summarization show promise for low-resource settings but require rigorous validation.

Conclusion: AI can shift goitre management from late-stage detection to proactive screening by improving sensitivity for occult Grade 1 enlargement, standardizing ultrasound interpretation, and reducing unnecessary invasive procedures. Clinical adoption, however, depends on transparent explainability, external validation across diverse cohorts, and governance aligned with high-risk medical AI standards.

Keywords: Goitre; Thyroid; Artificial Intelligence; Machine Learning; Ultrasound; Vision Transformer; U-Net; Explainable AI; Global Health

* Corresponding author: Kingdom Mutala Akugri

1. Introduction

Goitre is defined as abnormal hypertrophy of the thyroid gland and functions as a clinical "surface marker" for deeper endocrine, nutritional, and autoimmune processes. In population health, thyroid enlargement is closely tied to iodine status and remains a critical surveillance signal for iodine deficiency disorders, particularly in regions with geographic iodine leaching and limited access to adequately iodized salt [2,3].

Early disease detection is clinically important because the transition from no enlargement to subtle, palpable enlargement (WHO Grade 0 to Grade 1) is the stage most likely to be missed in routine care. This under-detection is driven by (i) variability in physical examination skill, (ii) reduced sensitivity in individuals with higher body mass index, and (iii) subjective interpretation of mild enlargement. Yet Grade 1 is precisely the stage where intervention (iodine correction, monitoring, and targeted evaluation) may prevent progression to multinodular goitre, reduce long-term morbidity, and improve risk stratification for nodular disease [4,6].

Ultrasound is the preferred imaging modality for early assessment, offering non-invasive evaluation of gland size, echogenicity, vascularity, and suspicious nodular features. However, ultrasound interpretation remains operator-dependent and can vary across clinicians and sites. Meanwhile, healthcare systems face increasing diagnostic load due to rising thyroid nodule incidence and expanded screening. These pressures motivate the use of ML and AI systems that can: (1) predict thyroid dysfunction from tabular clinical data; (2) quantify thyroid volume via segmentation; (3) classify enlargement and nodules using learned image features; and (4) provide explainable outputs aligned with established clinical workflows [14,26].

1.1. Aim and contribution of this review

This article provides a global epidemiological and computational framework for AI-enabled goitre screening, with emphasis on early detection. Specifically, we:

- Define the clinical "early goitre" detection problem around WHO Grade 1 and the limitsof physical examination,
- Summarize the strongest ML baselines for tabular thyroid prediction and the role ofimbalance-aware training,
- Review state-of-the-art ultrasound deep learning architectures for segmentation and diagnosis,
- Propose an end-to-end "holistic AI" workflow combining clinical and imaging markers,
- Discuss explainability, bias, validation, and deployment constraints required for realworld adoption.

2. Methods: Literature Search Strategy

This review followed a structured literature search and screening approach to identify studies relevant to AI/ML for early goitre detection and thyroid ultrasound decision support.

2.1. Search sources and timeframe

We searched PubMed/MEDLINE, Scopus, IEEE Xplore, and Google Scholar for studies published between 2010 and 2026. Preprints were included when accompanied by sufficient methodological detail.

2.2. Search terms

Search strings combined clinical and computational terms, including: (goitre OR "thyroid enlargement" OR "thyroid nodule" OR thyroiditis) AND (ultrasound OR sonography OR imaging) AND ("machine learning" OR "deep learning" OR CNN OR "vision transformer" OR segmentation OR XAI OR SHAP).

2.3. Inclusion and exclusion criteria

Inclusion criteria were: (i) ML/DL methods applied to thyroid/goitre-related tabular data or ultrasound imaging; (ii) studies reporting evaluation metrics; and (iii) clear description of dataset and validation. Exclusion criteria were: non-thyroid indications, non-ML methods, editorials without methods, and studies lacking reproducible evaluation.

2.4. Screening and synthesis

Titles and abstracts were screened for relevance, duplicates removed, and full texts assessed for eligibility. Findings were synthesized by modality (tabular vs imaging) and by clinical task (screening, segmentation/volume estimation, nodule risk stratification, explainability/deployment).

3. Epidemiology and Pathophysiology

3.1. Global dynamics and iodine deficiency

Historically, iodine deficiency has been the dominant driver of endemic goitre. Although iodization programs have reduced prevalence in many regions, gaps persist due to inconsistent fortification, supply constraints, and dietary patterns [2, 3]. Reported prevalence varies substantially by geography and subpopulation; representative prevalence clusters from the deep research synthesis are summarized in Table 1 [5,6,8,9].

Table 1 Representative goitre prevalence clusters and primary etiological drivers (synthesized from deep research sources).

Demographic / Geographic Cluster	Reported Goitre Prevalence	Primary Etiological Driver
Global Average (Iodine Replete)	4.7% – 5.0%	Autoimmunity / sporadic disease
Moderate Iodine Deficiency Areas	20.0% – 30.0%	Nutritional insufficiency
Severe Endemic Regions	>30.0% (up to 80.0%)	Environmental iodine leaching
Pregnant Women (Low-Income Countries)	up to 83.0%	Increased metabolic demand
Schoolchildren (e.g., Jazan, Saudi Arabia)	11.0%	Rural environmental factors
Adult Population (Thyroid Nodules)	24.83%	Age, obesity, and metabolic syndrome

3.2. Metabolic correlation and non-nutritional risk factors

Beyond iodine, thyroid enlargement increasingly co-occurs with metabolic syndrome. Obesity and insulin resistance are important non-nutritional risk factors, motivating AI models that integrate metabolic markers (BMI, lipid profiles, glucose/insulin resistance indicators where available) rather than relying solely on endocrine labs [7,8].

3.3. Clinical grading and the "occult" detection gap

The WHO grading system classifies goitre as Grade 0 (no palpable/visible enlargement), Grade 1 (palpable but not visible with neck in normal position), and Grade 2 (visible swelling). Grade 1 represents the "occult" stage most suitable for ultrasound quantification and AI assistance [1,4].

4. Data Sources and Benchmarking Considerations

AI-enabled early goitre detection relies on two primary data modalities: (i) structured clinical/laboratory data for risk prediction and triage, and (ii) ultrasound imaging for segmentation (volume estimation) and classification. Because many studies rely on single-center cohorts, external validation on at least one independent dataset is recommended for publication-quality evidence [15,21].

4.1. Representative datasets and typical tasks

Table 2 reproduces the dataset catalog from the deep research synthesis to support reproducible benchmarking [11,13,14,25].

Table 2 Public datasets commonly referenced in thyroid/goitre AI benchmarking (from deep research synthesis)

Dataset Name	Composition / Size	Diagnostic Utility
UCI Thyroid Disease Dataset	9,172 observations; 31 attributes	Tabular classification and feature selection
TN5000	5,000 B-mode ultrasound images	Largest open-access image set for detection
DDTI: Thyroid Ultrasound	349 annotated images	Interpretable malignancy prediction
TN3K	3,493 images with segmentations	Robust benchmark for U-Net segmentation

TCGA (Thyroid Cancer)	289 pathology reports	Training/validating NLP or LLM-based clinical staging extraction

5. Machine Learning Frameworks for Clinical Data

5.1. Problem formulation and features

Prediction of thyroid dysfunction and goitre-related clinical states from structured data is typically formulated as supervised classification (e.g., healthy vs. diseased, hypo/hyperthyroid

states, or risk tiers). Common features include age, sex, thyroid-stimulating hormone (TSH), triiodothyronine (T3), thyroxine (T4), and derived indices such as thyroxine utilization rate (T4U) and free thyroxine index (FTI) [11,12].

5.2. Preprocessing, imbalance, and generalization

Clinical datasets often exhibit strong class imbalance, which can bias models toward the majority ("healthy") class. Best practice includes resampling strategies (undersampling or SMOTE), stratified cross-validation, and transparent reporting of operating thresholds for screening usecases [12,27].

5.3. Algorithms and typical baselines

Ensemble approaches (bagging, boosting, stacking) are widely used for tabular medical classification due to robust performance and improved generalization. Practical baselines include logistic regression for interpretability, SVM for strong margins, random forests for non-linear interactions, and gradient boosting (e.g., XGBoost/LightGBM) for high performance on structured data [12].

Representative benchmark results summarized in the deep research are shown in Table 3 [12,20].

Table 3 Representative tabular ML benchmark performance reported in prior thyroid prediction studies (from deep research synthesis).

Algorithm Configuration	**Evaluation Metric**	**Performance Level**
Stacking (Meta-learner: XGBoost)	F1-score	0.9944
Bagging (3 decision trees)	F1-score	0.9766
Support Vector Machine (SVM)	Accuracy	99.63%
Neural Networks (MLP)	Accuracy	96.0%

6. Deep Learning Architectures for Imaging

6.1. Segmentation-first pipelines for early goitre

Accurate early goitre assessment often reduces to quantifying thyroid volume and detecting subtle echotexture changes. Segmentation models (notably U-Net variants) are commonly used to delineate thyroid boundaries and compute volume [16].

6.2. CNNs for nodule and texture characterization

CNN backbones (e.g., ResNet families) can learn discriminative patterns for nodular texture, echogenicity, and margin irregularity. Multi-channel or multi-view inputs can improve performance by capturing complementary perspectives [17,18].

6.3. Vision Transformers and long-range dependencies

Vision Transformers (ViTs) use self-attention to model long-range relationships within ultrasound images and are increasingly reported in thyroid CAD comparisons and reviews [14,19].

7. Holistic AI: A Multi-Modal Framework for Early Goitre

7.1. Screening

A clinically practical system for early goitre detection should integrate:

- Tabular risk stratification: use endocrine labs (TSH/T3/T4 and derived indices), demographics, and metabolic markers to estimate risk and prioritize imaging;
- Ultrasound segmentation and quantification: compute thyroid volume and compare to age/sex norms where available;
- Imaging classification: identify suspicious nodular patterns and gland-level abnormalities;
- Explainability and reporting: generate human-readable reasons (feature attributions and visual saliency) aligned with clinical guidelines;
- Referral decision support: recommend follow-up intervals, repeat imaging, or fineneedle aspiration (FNA) consideration based on risk tiers and explainable signals.

This “holistic AI” framing keeps the model clinically grounded: it does not attempt to replace clinical judgment, but rather standardizes early-stage detection and triage [20].

8. Emerging Technologies and Ethics

8.1. Explainable AI (XAI) for trust and adoption

Clinical adoption depends on interpretable outputs. For tabular models, SHAP-style feature attributions can demonstrate which measurements drive predictions. For imaging, saliency maps and attention visualization can highlight regions contributing to classification. Recent work in other domains has demonstrated the importance of geometric trust frameworks and explainable anomaly detection in building confidence in AI-driven diagnostic systems [26,28,29,32].

8.2. Reproducibility and external validation

Medical AI faces reproducibility challenges when models are trained only on proprietary single-center cohorts. Studies should validate on at least one external dataset and report performance consistency across subgroups and sites [15,21].

8.3. Bias, safety, and regulatory alignment

Common failure modes include data pathology (sampling bias), algorithmic bias (spurious correlations), and automation complacency. For high-stakes diagnostics, authors should document mitigation strategies and align evaluation with relevant governance expectations for high-risk medical AI [22,23].

8.4. Low-resource settings and future directions

For LMIC deployment, accessibility matters. Smartphone-assisted workflows and offline LLMs may support triage and documentation, but require careful clinical validation and clear safety boundaries [24,25].

9. Conclusion

AI and ML can materially improve early goitre detection by increasing sensitivity for occult WHO Grade 1 enlargement, standardizing ultrasound quantification, and integrating clinical markers to prioritize high-risk patients for imaging and specialist review. The strongest practical direction is multi-modal “holistic AI,” where tabular lab-based risk prediction complements imaging-based segmentation and classification rather than competing with it.

For future research and publication-quality evidence, we recommend:

- Multi-modal evaluation: report gains from combining tabular + imaging features over single-modality baselines;
- External validation: test models on at least one public or external dataset and include subgroup analysis;
- Explainability by design: integrate XAI outputs in the primary workflow (not as an
- afterthought);

- Deployment realism: include threshold selection for screening and monitoring plans for post-deployment drift.

Compliance with ethical standards

Disclosure of conflict of interest

The author declares no conflict of interest.

References

[1] Can AS, Rehman A. Goiter. [Updated 2023 Aug 14]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK562161/

[2] Iodine Global Network. IGN Annual Report 2024 [Internet]. Available from: https://www.scribd.com/document/882236151/11959-IGN-AnnualReport-2024-1-6-Digital

[3] World Health Organization. Goitre as a determinant of the prevalence and severity of iodine deficiency disorders in populations [Internet]. Geneva: WHO; Available from: https://iris.who.int/bitstream/handle/10665/133706/WHO_NMH NHD EPG 14.5 eng.pdf

[4] International Council for Control of Iodine Deficiency Disorders., UNICEF., and World Health Organization. Assessment of Iodine Deficiency Disorders and Monitoring Their Elimination : A Guide for Programme Managers. 2nd ed. Geneva: World Health Organization, 2001. Print.

[5] Lin J, Tan HL, Ge H. Global, Regional, and National Burden of Iodine Deficiency in Reproductive Women From 1990 to 2019, and Projections to 2035: A Systematic Analysis for the Global Burden of Disease Study in 2019. Int J Womens Health. 2025 Jun 21;17:1863-1875. doi: 10.2147/IJWH.S513856. PMID: 40574983; PMCID: PMC12198354.

[6] Mu C, Ming X, Tian Y, Liu Y, Yao M, Ni Y, Liu Y, Li Z. Mapping global epidemiology of thyroid nodules among general population: A systematic review and meta-analysis. Front Oncol. 2022 Nov 10;12:1029926. doi: 10.3389/fonc.2022.1029926. PMID: 36439513; PMCID: PMC9685339.

[7] Kalra, Sanjay & Sahay, Rakesh & A G, Unnikrishnan. (2013). The global burden of thyroid disease. Thyroid Research and Practice. 10. 89. 10.4103/0973-0354.116129.

[8] The epidemiology of thyroid disease. ResearchGate; Available from: https://www.researchgate.net/publication/51619844 The epidemiology _of thyroid disease

[9] Alsanosy, Rashad & Gaffar, Abdelrahim & Khalafalla, Husam & Mahfouz, Mohamed & Zaid, Abdelnaser & Bani, Ibrahim. (2012). Current iodine nutrition status and progress toward elimination of iodine deficiency disorders in Jazan, Saudi Arabia. BMC public health. 12. 1006. 10.1186/1471-2458-12-1006.

[10] Bahn RS, Castro MR. Approach to the patient with nontoxic multinodular goiter. J Clin Endocrinol Metab. 2011 May;96(5):1202-12. doi: 10.1210/jc.2010-2583. PMID: 21543434.

[11] Arel-Bundock V. Thyroid disease dataset (Rdatasets/MLDataR) [Internet]. Available from: https://vincentarelbundock.github.io/Rdatasets/doc/MLDataR/thyroid disease.html

[12] Enhancing thyroid disease prediction using machine learning (PDF). ResearchGate; Available from: https://www.researchgate. net/publication/388941244 Enhancing Thyroid Disease Prediction Using Machine_ Learning

[13] Zhang H, Liu Q, Han X, Niu L, Sun W. TN5000: An Ultrasound Image Dataset for Thyroid Nodule Detection and Classification. Sci Data. 2025 Aug 16;12(1):1437. doi: 10.1038/s41597-025-05757-4. PMID: 40819171; PMCID: PMC12357846.

[14] Yang WT, Ma BY, Chen Y. A narrative review of deep learning in thyroid imaging: current progress and future prospects. Quant Imaging Med Surg. 2024 Feb 1;14(2):2069-2088. doi: 10.21037/qims-23-908. Epub 2024 Jan 19. PMID: 38415152; PMCID: PMC10895129.

[15] Savelonas, Michalis. (2025). An Overview of AI-Guided Thyroid Ultrasound Image Segmentation and Classification for Nodule Assessment. Big Data and Cognitive Computing. 9. 255. 10.3390/bdcc9100255.

[16] Prochazka, Antonin & Zeman, Jan. (2025). Thyroid nodule segmentation in ultrasound images using U-Net with ResNet encoder: achieving state-of-the-art performance on all public datasets. AIMS Medical Science. 12. 124-144. 10.3934/medsci.2025009.

[17] Zhu YC, Jin PF, Bao J, Jiang Q, Wang X. Thyroid ultrasound image classification using a convolutional neural network. Ann Transl Med. 2021 Oct;9(20):1526. doi: 10.21037/atm-21-4328. PMID: 34790732; PMCID: PMC8576712.

[18] Zhang, Amy & Lee, Vincent & Rong, Jia & Liu, Feng & Kong, Haoyu. (2022). Multi-channel convolutional neural network architectures for thyroid cancer detection. PLOS ONE. 17. e0262128. 10.1371/journal.pone.0262128.

[19] A deep learning framework for thyroid nodule segmentation and malignancy classification from ultrasound images. arXiv; Available from: https://www.arxiv.org/pdf/2511.11937

[20] Ateeq Almutairi S. Advancing thyroid diagnosis: integrating AI-driven CAD framework with numerical data and ultrasound images. PeerJ Comput Sci. 2025 Jul 29;11:e3063. doi: 10.7717/peerj-cs.3063. PMID: 40989437; PMCID: PMC12453860.

[21] Li Y, Yi X, Fu J, Yang Y, Duan C, Wang J. Reducing misdiagnosis in AI-driven medical diagnostics: a multidimensional framework for technical, ethical, and policy solutions. Front Med (Lausanne). 2025 Oct 31;12:1594450. doi: 10.3389/fmed.2025.1594450. PMID: 41244777; PMCID: PMC12615213.

[22] Reducing misdiagnosis in AI-driven medical diagnostics: a multidimensional framework for technical, ethical, and policy solutions. PubMed Central; Available from: https://pmc.ncbi.nlm.nih.gov/

[23] Mennella C, Maniscalco U, De Pietro G, Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. Heliyon. 2024 Feb 15;10(4):e26297. doi: 10.1016/j.heliyon.2024.e26297. PMID: 38384518; PMCID: PMC10879008.

[24] Assaad S, Dov D, Davis R, Kovalsky S, Lee WT, Kahmke R, Rocke D, Cohen J, Henao R, Carin L, Range DE. Thyroid Cytopathology Cancer Diagnosis from Smartphone Images Using Machine Learning. Mod Pathol. 2023 Jun;36(6):100129. doi: 10.1016/j.modpat.2023.100129. Epub 2023 Feb 13. PMID: 36931041; PMCID: PMC10293075.

[25] Vlachakis D, Dragoumani K, Papakonstantinou E, Chrousos GP. The Intersection of Artificial Intelligence and Precision Endocrinology. EMBnet J. 2024;30:e1052. doi: 10.14806/ej.30.0.1052. Epub 2024 Sep 9. PMID: 39323646; PMCID: PMC11423795.

[26] Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, Cronan JJ, Beland MD, Desser TS, Frates MC, Hammers LW, Hamper UM, Langer JE, Reading CC, Scoutt LM, Stavros AT. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. J Am Coll Radiol. 2017 May;14(5):587-595. doi: 10.1016/j.jacr.2017.01.046. Epub 2017 Apr 2. PMID: 28372962.

[27] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority oversampling technique. J Artif Intell Res. 2002;16:321–357.

[28] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS). 2017.

[29] Lu Q, Wu Y, Chang J, Zhang L, Lv Q, Sun H. Application progress of artificial intelligence in managing thyroid disease. Front Endocrinol (Lausanne). 2025 Jun 17;16:1578455. doi: 10.3389/fendo.2025.1578455. PMID: 40600013; PMCID: PMC12208828.

[30] Azigi JA, Adrah F. Toward Smart Biosensing: A Machine Learning Approach for Early Diabetes Detection. Int J Comput Appl. 2025;187(19):8–11.

[31] Agboklu M, Lartey B, Adrah F, LaJeunesse D. Towards Proactive Heart Health: A Machine Learning-Powered Approach for Chronic Heart Failure Detection. Int J Comput Appl. 2024;186(28):30–35.

[32] Kingdom Mutala Akugri, Prince Agbenyo, Marious Akugri. Generative Identity Forensics and Trust System (GIFTS): Geodesic anomaly detection and manifold diffusion for cloud identity telemetry