Systematic Review

Machine Learning and Artificial Intelligence in Thyroid Cancer Screening and Diagnosis: A Comprehensive Systematic Review

Abstract

This systematic review explores the role of artificial intelligence (AI) and machine learning (ML) technologies in the diagnosis and treatment of thyroid cancers (TC), focusing on enhancing precision, risk assessment, and tailored care. Despite advancements, challenges persist in transitioning to personalized care, including uneven prognostication and diagnostic uncertainty. The review evaluates the effectiveness of AI and ML compared to conventional methods, their ability to address diverse tumor characteristics, and their strengths and limitations in prognosis prediction. Methodologically, 10 studies were analyzed through a comprehensive literature search, focusing on AI and ML applications in TC diagnosis, prognosis, and treatment. Findings suggest AI's potential in improving precision and risk assessment, but limitations such as inconsistent approaches and biases highlight the need for larger datasets and standardized procedures. Moreover, the review emphasizes the importance of interpretability and transparency in AI models and calls for further research to validate findings in clinical settings. Despite limitations and challenges, AI's transformative potential in TC management is evident, underscoring the need for ongoing investigation and integration into clinical practice.

Keywords: Thyroid Cancer, Artificial Intelligence, Machine Learning, thyroid nodule, diagnosis, AI, ML

Introduction

Over the past two decades, there has been a notable increase in the incidence of thyroid cancers (TC); the majority of TC cases are indolent [1][2]. Addressing these trends is crucial, given the continuous rise in incidence and death rates for aggressive papillary thyroid carcinomas (PTC) and advanced thyroid malignancies [1,3,4]. Accurate and efficient risk assessment is essential in the era of customized healthcare to tailor therapy effectively. Understanding TC's biological function, characterized by diverse morphological traits and molecular elements, is the initial step. While image analysis remains the primary diagnostic method for TC, its limitations in providing a thorough evaluation are evident [5]. Primary human cell cultures from surgical biopsies and fine-needle aspiration (FNA) samples offer opportunities for customized treatments, though challenges persist in transitioning to personalized care [6], such as uneven prognostication and uncertainty surrounding cytopathological diagnosis.

Radiologists have identified computer-aided diagnostics (CAD) as valuable for identifying cancers beyond breast cancer [11]. Assessing disease phases aids in determining the extent of thyroid cancer progression. Deep learning (DL) enhances ultrasound (US) accuracy by extracting nonlinear features [12]. Artificial intelligence (AI) facilitates improved operational performance and swift access to critical information for physicians. CAD and AI simplify risk-stratification systems, enhancing thyroid nodule detection and evaluation [13]. Molecular testing combined with machine learning (ML) techniques helps forecast and detect high-risk nodules [14]. ML's intrinsic power in drawing conclusions beyond traditional statistical approaches is evident [15]. Classification models developed using ML methods show promise in improving thyroid imaging assessment CAD systems [16].

Machine learning (ML) enables completion of complex tasks, such as photo interpretation [18]. DL aids in lung cancer detection on CT images. ML applications are growing, offering a comprehensive approach to cancer diagnosis and prevention [25]. Clinical parameters influence disease prognosis, with ML generating predictions to assist in patient disease management [22,23]. Protein markers and microarray data are increasingly relied upon in cancer diagnosis [24]. ML techniques, including supervised and unsupervised methods, are expanding in healthcare domains [26]. Metabolomics technology sheds light on lung cancer characteristics [25]. ML aids in diagnosing cancer types, predicting susceptibility, and screening individuals [27]. ML models improve tumor diagnostic accuracy and optimize therapeutic approaches [27].

The most frequently used neural networks in oncology are the convolution neural network (CNN), recurrent neural network (RNN), and multilayer perceptron (MLP). Cytopathology and histology are common methods for cancer diagnosis [28]. Histology-based CNNs classify prostate, breast, and colon cancers successfully [29]. DL effectively distinguishes between benign and malignant tissues in lung cancer using whole-slide imaging. ML aids in forecasting tumor origins, even when unknown causes contribute to cancer cases [30].

Objective

The systematic review aimed to meticulously evaluate and consolidate data concerning the utilization of machine learning (ML) and artificial intelligence (AI) in thyroid tumors (TC). The primary goal of the study was to explore the potential enhancements in thyroid cancer diagnosis, prognosis, and treatment options facilitated by AI and ML technologies. The study comprehensively analyzed the merits and demerits of employing AI and ML in processing diagnostic imaging data, with a particular focus on addressing the following inquiries:

1. How effective are AI and ML technologies compared to conventional image analysis techniques in identifying thyroid cancers?
2. Can AI and ML adequately address various biological and physical characteristics of thyroid tumors to offer more precise risk assessment and tailored treatment?
3. What are the strengths and limitations of current AI and ML models in discerning the benignity or malignancy of thyroid nodules?
4. To what extent can AI and ML aid in predicting prognosis and early detection of thyroid cancer, particularly before the onset of symptoms?

Methodology

Search Strategy

We conducted an extensive literature search to identify relevant studies concerning the utilization of artificial intelligence (AI) and machine learning (ML) in the management of thyroid cancer. We systematically searched several electronic databases, including Embase, Web of Science, PubMed, and Scopus. To ensure a comprehensive search, various combinations of keywords related to AI, ML, thyroid cancers, and related topics were utilized. The search algorithm was developed using boolean operators and tailored to match the syntax of each database.

Eligibility Criteria

Inclusion Criteria

Studies were deemed eligible for inclusion in this systematic review if they met specific criteria designed to ensure the accuracy and relevance of the collected data. Initially, publications had to be written in English to facilitate comprehension and synthesis of results. Included publications consisted of original research articles or reviews focusing specifically on the application of AI and ML to thyroid cancers. To maintain relevance to clinical practice, only studies involving human participants were considered. Additionally, studies reporting findings related to the use of AI or ML in the diagnosis, prognosis, or treatment of thyroid cancers were included.

Exclusion Criteria

Studies failing to meet precise exclusion criteria were systematically excluded from the review to maintain its integrity and focus. Publications not presented in English were excluded to ensure coherent comprehension of the content. Conference abstracts, letters, editorials, and case reports were also excluded due to their lack of scientific rigor and depth suitable for inclusion in a systematic review. Materials not specifically addressing the application of AI or ML in the context of thyroid tumors were omitted to maintain the study's focus. To prevent duplication and preserve the originality of the data pool, duplicate articles were also excluded. Table 1 outlines the PICOS framework and eligibility criteria for this review.

Table 1: PICOS framework and eligibility criteria

Criteria	Description
Population	Human subjects diagnosed with thyroid cancers.
Intervention	Original research articles or reviews focusing on the application of AI and ML in the context of thyroid cancers.
Comparison	Not applicable (as this is not a comparative study).
Outcomes	Studies reporting outcomes related to the diagnosis, prognosis, or management of thyroid cancers using AI or ML techniques.
Study Design	Various study designs, including observational studies, clinical trials, and reviews.

Data Extraction

Thorough and meticulous data extraction procedures were applied to each included study to ensure comprehensive gathering of all relevant information for synthesizing findings. The extracted data encompassed various aspects, including authors, publication year, and research design. Demographic details of participants were carefully collected to provide valuable insights into the investigated groups. Detailed explanations of AI and ML techniques employed in each study facilitated readers' understanding of the methodological approaches. Additionally, comprehensive extractions of data regarding the diagnosis, prognosis, and treatment of thyroid malignancies using AI or ML methods were conducted. Key results and conclusions from each study were documented to enable a comprehensive review and understanding of the accumulated information.

Quality Assessment

Established instruments appropriate for the research designs were utilized to assess the quality of included studies. The Cochrane Risk of Bias tool was employed for clinical trials, while the Newcastle-Ottawa Scale was used for observational research. Each study underwent independent evaluation by two reviewers, with discrepancies resolved through discussion or consultation with a third reviewer if necessary.

Results

Study Selection

A systematic search across PubMed, Cochrane Library, and Google Scholar databases yielded a total of 218 records. After removal of duplicates, 121 records remained. Screening of titles and abstracts narrowed down the selection to 74 potentially relevant records. Following full-text screening, 10 studies met the inclusion criteria for the systematic review. Figure 1 illustrates the comprehensive flow diagram depicting the search and selection process.

Figure 1; PRISMA Flow Chart

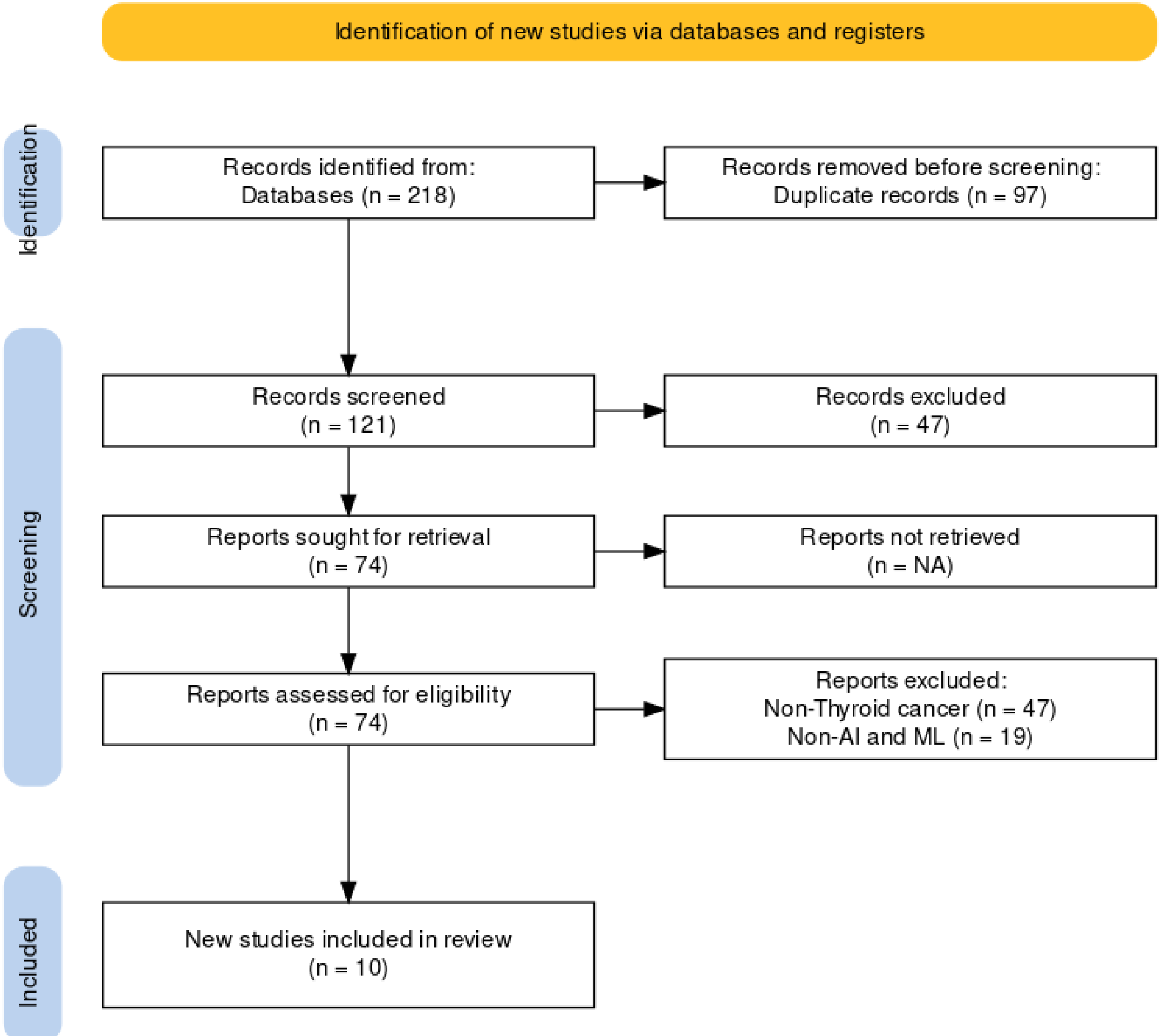
Identification of new studies via databases and registers
Identification
Records identified from:
Databases (n = 218)
Records removed before screening:
Duplicate records (n = 97)
Screening
Records screened
(n = 121)
Records excluded
(n = 47)
Reports sought for retrieval
(n = 74)
Reports not retrieved
(n = NA)
Reports assessed for eligibility
(n = 74)
Reports excluded:
Non-Thyroid cancer (n = 47)
Non-AI and ML (n = 19)
Included
New studies included in review
(n = 10)

Table 2 outlines the studies incorporated in this article.

Table 2: Characteristics of included studies

Author & Year	Study Design	Interventions	Population	Outcome Measures	Findings

Yoon et al. (2020)	Retrospective Observational Study	Computer-aided diagnosis (CAD) program using deep learning Convolutional Neural Network (CNN)	469 patients with thyroid cancer (380 positive, 89 negative for BRAFV600E mutation)	Association of CAD value with BRAFV600E mutation, Area Under the Receiver Operating Characteristic (AUC) of Receiver Operating Characteristic (ROC) curves for CAD value and multivariable model	Older age, smaller size, and higher CAD value significantly associated with BRAFV600E mutation. CAD value yielded an AUC of 0.646 for predicting BRAFV600E mutation. Multivariable model (age, size, and CAD value) had an AUC of 0.706, significantly better than CAD value alone. Deep learning-based CAD program shows promise in

					predicting BRAFV600E mutation in thyroid cancer. Multicenter studies with larger sample sizes are recommended for further validation.

Bellantuono et al. (2023)	eXplainable Artificial Intelligence analysis	Machine Learning procedure for discrimination of healthy/benign vs. malignant nodules using Raman spectra, Boruta feature selection, Synthetic Minority Over-sampling Technique (SMOTE) algorithm for imbalanced dataset, Random Forest, eXtreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and	Patients with thyroid nodular pathology, 54 subjects (34 females, 20 males), aged 46.3 years on average, who underwent surgery (total thyroidectomy) after a cytological diagnosis of indeterminate, suspicious, or malignant nodules	Classification performance of Machine Learning algorithms (Random Forest, XGBoost, SVM, Gaussian Naïve Bayes) quantified by AUC, feature importance using Boruta, and synthetic data generation using SMOTE	Random Forest is identified as the best classifier (median AUC 0.9441, interquartile range 0.0049) for healthy/benign vs. cancer tissue classification. XGBoost, SVM, and Gaussian Naïve Bayes also explored. eXplainable Artificial Intelligence (XAI) analysis (SHapley Additive exPlanations - SHAP values) for interpretabilit

		Gaussian Naïve Bayes classifiers			y. Performance evaluated on 72 samples (59 unambiguous and 13 ambiguous). Identified limitations in classifying ambiguous spectra with reduced AUC (median 0.7949, IQR 0.0135). Impactful features include carotenoid and oxidized cytochrome bands.

Ha, E. J., & Baek, J. H. (2021)	Review and developmental overview of AI-based CAD systems	Application of CAD systems by loading ultrasound images from Picture Archiving and Communication System (PACS). Real-time application during Ultrasound (US) examinations	Patients with thyroid nodules undergoing ultrasound imaging	Analysis of sonographic characteristics (echogenic foci, echogenicity, texture, margin, anechoic areas, height/width ratio, nodule shape, and size) and risk of malignancy based on Thyroid Imaging Reporting and Data System (TI-RADS) classifications	AmCAD-UT: Similar sensitivity (87.0%) but lower specificity (68.8%) compared to clinical experts using TI-RADS. Food and Drug Administration (FDA) 510(k) cleared. S-Detect 1: Comparable sensitivities (80.0%-92.0%) but lower specificity (74.6%-88.1%) compared to experienced radiologists. FDA approval

					in progress. S-Detect 2: Comparable sensitivities (81.4%) but lower specificity (68.2%-81.9%) compared to experienced radiologists.

Agarwal et al. (2021)	Evaluation and comparison of AI algorithms	Implementation of AI algorithms and machine learning models to analyze diagnostic imaging data	Individuals undergoing diagnostic tests for cancer, including imaging tests, endoscopic procedures, biopsy, and cytology	Assessment of the diagnostic accuracy of AI algorithms and machine learning models in differentiating benign and malignant tumors	AI improves diagnostic accuracy by analyzing large imaging datasets, leveraging technical advances and hardware enhancements for neural network training. It excels in early diagnosis, particularly in breast and lung cancer, surpassing human specialists in breast cancer prognosis and providing early lung cancer predictions. In gastric cancer, Convolutional

					Neural Networks aid in invasion depth diagnosis through gastric endoscopy. AI techniques, coupled with imagery, enable early identification of oral cancer. Overall, AI significantly enhances cancer diagnosis precision and extends forecasting capabilities.

Xi et al. (2022)	Prospective study using machine learning	Six machine learning models trained on a clinical dataset from 724 patients undergoing thyroidectomy. Models included Gradient Boosting, Logistic Regression, Linear Discriminant Analysis, SVM, and Random Forest	724 patients at Shengjing Hospital, China, with demographic info, ultrasound features, and blood test results	Models demonstrated superior accuracy, with Random Forest leading. Gradient Boosting excelled in sensitivity, Logistic Regression in specificity. Variable importance analysis highlighted key predictors. Models outperformed expert assessment in accuracy and F1 score	Machine learning, especially Random Forest and Gradient Boosting, improved thyroid nodule malignancy prediction compared to expert assessment. Models offered valuable insights into nodule characteristics, enhancing preoperative thyroid cancer diagnosis.

Peng et al. (2021)	Multicentre Diagnostic Study	Development and application of the deep-learning AI model (ThyNet) for differentiating thyroid nodules	Patients aged 18 or older with thyroid nodules at least 3 mm in diameter identified via ultrasound	Primary Endpoint: Area Under the Receiver Operating Characteristic Curve (AUROC) for thyroid nodule diagnosis. Secondary Endpoints: Accuracy, Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV)	ThyNet AUROC: 0.922 (95% CI 0.910–0.934) was significantly higher than radiologists ($p < 0.0001$). ThyNet-assisted strategy improved radiologists' AUROC from 0.837 to 0.875 ($p < 0.0001$). In a simulated scenario, ThyNet-assisted strategy reduced unnecessary fine needle aspirations by 26.7%. Missed malignancy decreased

						from 18.9% to 17.0% with ThyNet-assisted strategy.

Olatunji et al. (2021)	Retrospective Case Study	Machine learning-based tools development for early detection of thyroid cancer (TC)	Techniques used: Random Forest (RF), Artificial Neural Network (ANN), Support Vector Machine (SVM), Naïve Bayes (NB)	Patients from the Kingdom of Saudi Arabia	RF technique demonstrated the highest accuracy at 90.91%. SVM, ANN, and NB achieved accuracy rates of 84.09%, 88.64%, and 81.82%, respectively. Emphasis on early detection at pre-symptomatic stages. RF recommended as the preferred technique for this specific problem.

Zhou et al. (2022)	Experimental Study (Comparison of new ultrasound technologies)	Ultrasonic intelligent diagnosis of papillary thyroid cancer based on machine learning, involving Contrast-Enhanced Ultrasound (CEUS) and Ultrasound Elastography (UE) technologies	Patients with papillary thyroid carcinoma (PTC), 70 cases (10 male, 60 female), tumor diameter ≤10 mm, 107 lymph nodes	Characteristics of ultrasound images (CEUS and UE). Diagnostic effectiveness of new ultrasound technologies (CEUS and UE) in distinguishing between benign and malignant nodules	CEUS and UE showed significant differences in enhancement mode, intensity, early regression, and edge enhancement between micro benign and malignant tumors. UE demonstrated higher sensitivity and diagnostic efficiency compared to CEUS in the differential diagnosis of thyroid micro benign and malignant nodules. Combined use of CEUS and

						UE resulted in 78.43% sensitivity and 78.67% specificity for diagnosing thyroid micro benign and malignant nodules.

Zhao et al. (2019)	Meta-analysis	Evaluation of computer-aided diagnosis system (CAD) for thyroid nodules on ultrasound	536 patients with 723 thyroid nodules	Sensitivity, Specificity, Positive Likelihood Ratio (LR), Negative LR, Diagnostic Odds Ratio (DOR)	Findings (CAD System): Sensitivity: 0.87 (95% CI, 0.73–0.94), Specificity: 0.79 (95% CI, 0.63–0.89), Positive LR: 4.1 (95% CI, 2.5–6.9), Negative LR: 0.17 (95% CI, 0.09–0.32), DOR: 25 (95% CI, 15–42), Summary Receiver Operating Characteristic (SROC) AUC: 0.90 (95% CI, 0.87–0.92). Findings (Experienced Radiologists): Sensitivity: 0.82 (95% CI,

					0.69–0.91), Specificity: 0.83 (95% CI, 0.76–0.89), Positive LR: 4.9 (95% CI, 3.4–7.0), Negative LR: 0.22 (95% CI, 0.12–0.38), DOR: 23 (95% CI, 11–46), SROC AUC: 0.96 (95% CI, 0.94–0.97).

Kuang et al. (2022)	Metabolomic analysis, machine-learning model development	Analysis of existing data of thyroid cancer (TC) metabolites, development of a machine-learning model using metabolite biomarkers	The study involved datasets related to papillary thyroid cancer (PTC) patients	Classification accuracy of machine-learning models (LogitBoost, AdaBoostM1, RandomForest, etc.) through 10-fold cross-validation. Identification of metabolic pathways related to TC	Highest classification accuracy: LogitBoost - 87.30%, Various classifiers achieved accuracies above 80%. Independent testing showed 100% accuracy in identifying TC-related metabolites .

Discussion

A deep learning convolutional neural network (CNN)-based computer-aided diagnostic (CAD) tool was created for the diagnosis of thyroid cancer in a retrospective observational research carried out by Yoon et al. in 2020 [30]. Of the 469 thyroid cancer patients included in the research, 380 had a positive BRAFV600E mutation test and 89 did not. The relationship between the CAD value and the BRAFV600E mutation was evaluated by calculating the area under the receiver operating characteristic (ROC) curve (AUC) for the CAD value and a multivariable model. The BRAFV600E mutant was shown to be significantly correlated with higher CAD values, smaller sizes, and older ages. An AUC of 0.646 was discovered for the BRAFV600E mutant prediction using the CAD value. Age, size, and CAD value were all incorporated in the multivariable model, which had an AUC of 0.706, considerably higher than CAD value by itself. Based on the results, the deep learning-based CAD program may be able to forecast the BRAFV600E mutation in thyroid cancer. Nonetheless, the authors recommend additional validation via multicenter research with bigger sample sizes.

Thyroid cancer was identified by Bellantuono et al. in a paper published in 2023 using an eXplainable Artificial Intelligence (XAI) analysis of Raman spectra [31]. Using a variety of classifiers, such as Random Forest, XGBoost, Support Vector Machine, and Gaussian Naïve Bayes, along with Boruta feature selection, the synthetic minority oversampling technique (SMOTE) algorithm for imbalanced datasets, and Raman spectra, the machine learning process involved identifying benign and malignant nodules. 54 patients with thyroid nodular pathology (mean age 46.3 years) who had total thyroidectomy following a cytological diagnosis of ambiguous, suspicious, or malignant nodules were included in the research. The Random Forest classifier was found to be the best, with a median AUC of 0.9441 for classifying tissue as either cancerous or healthy. Important characteristics like oxidized and carotenoid cytochrome bands were found using XAI analysis and SHAP values for interpretability. The work demonstrated the difficulties of identifying ambiguous spectra, but it also demonstrated the potential application of machine learning to diagnose thyroid cancer with the assistance of XAI.

Ha and Baek provided a thorough analysis and developing overview of AI-based computer-aided diagnostic (CAD) techniques for thyroid nodules in 2021 [32]. During ultrasonography testing, CAD systems were used in real time by importing ultrasound pictures from Picture Archiving and Communication Systems (PACS). Among those having ultrasounds were individuals who had thyroid nodules. Sonographic features were the main focus of the inquiry, and the Thyroid Imaging Reporting and Data System (TI-RADS) categories were used to determine the risk of malignancy. In the study, a variety of CAD systems, such as AmCAD-UT, S-Detect 1, and S-Detect 2, were evaluated against radiologists and clinical specialists. The results showed that CAD systems may be used in real-time ultrasound exams, even though their sensitivity was either similar to or somewhat lower than that of clinical specialists.

Agarwal et al. examined and compared machine learning models and AI algorithms for cancer diagnosis in 2021 [33]. Using AI algorithms and machine learning models, the study examined diagnostic imaging data from patients undergoing various cancer diagnostic procedures. The assessment focused on how well AI systems can differentiate between benign and malignant cancers. The findings showed how artificial intelligence (AI) techniques might increase the precision of cancer diagnosis by using hardware updates and technology breakthroughs to train neural networks. Significantly, it has been shown that AI algorithms may be used for early detection, prognostication, and enhanced accuracy of a number of cancer types, including stomach, oral, lung, and breast cancer. The focus of the study was on how AI-generated predictions that are lengthier and more precise may enhance cancer detection overall.

Xi et al. [2022] used ten-fold cross-validation and machine learning with bootstrap analysis in a prospective research to predict the malignancy of thyroid nodules. Using a clinical dataset of 724 patients undergoing thyroidectomy, six machine learning models—Gradient Boosting, Logistic Regression, Linear Discriminant Analysis, Support Vector Machine, and Random Forest—were developed. The models included blood test results, ultrasound characteristics, and demographic data. The study claims that machine learning models are more accurate than ever, with Random

Forest performing best overall. While Logistic Regression showed excellent specificity, Gradient Boosting had better sensitivity. Important predictors were identified via variable significance analysis, and the models performed better in terms of accuracy and F1 score than expert judgment. In predicting the malignancy of thyroid nodules, the results demonstrated that machine learning—more especially, Random Forest and Gradient Boosting—performed better than expert evaluation. Additionally, the nodule traits that machine learning revealed might improve preoperative thyroid cancer diagnosis [34].

Peng et al. [2021] created and used the ThyNet deep-learning AI model to distinguish thyroid nodules in a multicenter diagnostic investigation. Patients with thyroid nodules detected by ultrasonography that were at least 3 mm in diameter were included in the research. The area under the Receiver Operating Characteristic Curve (AUROC) was the main goal for thyroid nodule identification; other outcomes were accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). ThyNet outperformed radiologists with an AUROC of 0.922 ($p < 0.0001$). When radiologists used the approach, their AUROC increased from 0.837 to 0.875 ($p < 0.0001$) with ThyNet's assistance. Within a simulated setting, the ThyNet-assisted approach lowered the percentage of missed malignancies from 18.9% to 17.0% and the frequency of needless microneedle aspirations by 26.7%. The results show that ThyNet is a useful tool for thyroid nodule identification since radiologists may use it to identify thyroid nodules more accurately and with fewer needless procedures [35].

Olatunji et al. conducted a retrospective case study in 2021 on the creation of machine learning-based tools for the early identification of thyroid cancer (TC) in Saudi Arabian nationals. Methods such as Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN) were employed in the study. Among these, RF had the highest accuracy (90.91%), followed by SVM, ANN, and NB, with 84.09%, 88.64%, and 81.82%, in that order. The study emphasized the need for early diagnosis in the pre-symptomatic stages, and radiation therapy was recommended as the most effective approach for this specific problem. The results suggest that machine learning techniques can lead to high accuracy rates for the early identification of thyroid cancer [36].

Zhou et al. [2022] assessed the efficacy of machine learning between ultrasound elastography (UE) and contrast-enhanced ultrasound (CEUS) for the ultrasonic intelligent diagnosis of papillary thyroid cancer (PTC). PTC patients participated in the study, and the ultrasound scans were used to examine their features. Comparing UE with CEUS, UE showed superior diagnostic sensitivity and efficacy in differentiating between benign and malignant nodules. In contrast to 90.21% and 78.45% for UE, the sensitivity and specificity of CEUS were 85.34% and 67.54%, respectively. When CEUS and UE were combined, the diagnostic accuracy for thyroid micro benign and malignant nodules was 78.43% with a specificity of 78.67%. According to the results, UE and CEUS together may improve thyroid nodule identification accuracy, especially when it comes to identifying benign from malignant instances [37].

Zhao et al. evaluated a computer-aided diagnostic system (CAD) for thyroid nodules on ultrasonography in a 2019 meta-analysis that included 536 people and 723 thyroid lesions. The following outcomes were obtained with the CAD system: a diagnostic odds ratio of 25, a positive likelihood ratio of 4.1, a negative likelihood ratio of 0.17, a sensitivity of 0.87, and a specificity of 0.79. It was 0.90 for the AUC of the summary receiver operating characteristic (SROC). In comparison, experienced radiologists had an SROC AUC of 0.96, sensitivity of 0.82, specificity of 0.83, positive likelihood ratio of 4.9, negative likelihood ratio of 0.22, and diagnostic odds ratio of 23. The findings demonstrate that the CAD system outperforms skilled radiologists in thyroid nodule identification, offering competitive sensitivity and specificity [38].

Kuang et al. [2022] used the available information on the metabolites of thyroid cancer (TC) to conduct a metabolomic analysis and create a machine-learning model. Although patient demographics were not specifically specified, the study included patient data for papillary thyroid carcinoma (PTC). The objective of the study was to categorize metabolite biomarkers linked to TC. During 10-fold cross-validation, a number of classifiers were able to reach accuracy levels above 80%; the LogitBoost model had the greatest classification accuracy, at 87.30%. Tests conducted independently revealed 100% accuracy in detecting metabolites associated with TC. The findings imply that machine learning methods, particularly LogitBoost, may be able to classify metabolites linked to thyroid cancer in an efficient manner, potentially improving the diagnosis of thyroid cancer [39].

Limitations

The reviewed literature, despite its promising implications, exhibits significant limitations. The wide variations in research methodologies, patient cohorts, and AI models make it challenging to draw definitive conclusions applicable across different contexts. Given that many studies rely on observational or retrospective designs, larger-scale randomized controlled trials may be essential to ascertain the therapeutic effectiveness of AI and ML in diagnosing and treating thyroid cancer. Moreover, the diversity of data sources, imaging techniques, and diagnostic criteria among studies introduces potential biases and hampers the generalizability of findings. For instance, Bellantuono et al.'s study illustrates the challenges in distinguishing ambiguous spectra, underscoring practical limitations.

Yoon et al.'s suggestion for larger, multicenter investigations underscores the need for more comprehensive validation and highlights the constraints imposed by current sample sizes. Addressing issues related to interpretability and transparency of AI models, as demonstrated by Bellantuono et al.'s XAI project, is imperative before widespread adoption of AI technology in healthcare settings. Furthermore, the focus on specific AI applications and exclusion of non-English research may introduce publication bias and limit our understanding of the field.

Conclusion

In conclusion, while significant advancements have been made in AI and ML for diagnosing and prognosticating thyroid cancer, caution is warranted due to discrepancies and shortcomings noted in various studies. Future research in this area should prioritize standardized methodologies, larger and more diverse sample sizes, and inclusion of various patient demographics to enhance therapeutic utility and generalizability. Additionally, considerations of interpretability, transparency, and ethical implications will be pivotal in the seamless integration of AI and ML into routine clinical practice. Overall, this systematic review underscores the need for further investigation and validation in real-world clinical settings and emphasizes the transformative potential of AI in the management of thyroid cancer.

References

1. Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in thyroid cancer incidence and mortality in the United States, 1974-2013. Jama. 2017 Apr 4;317(13):1338-48.

2. La Vecchia C, Malvezzi M, Bosetti C, Garavello W, Bertuccio P, Levi F, Negri E. Thyroid cancer mortality and incidence: a global overview. International journal of cancer. 2015 May 1;136(9):2187-95.

3. Ho AS, Luu M, Barrios L, Chen I, Melany M, Ali N, Patio C, Chen Y, Bose S, Fan X, Clair JM. Incidence and mortality risk spectrum across aggressive variants of papillary thyroid carcinoma. JAMA oncology. 2020 May 1;6(5):706-13.

4. Jegerlehner S, Bulliard JL, Aujesky D, Rodondi N, Germann S, Konzelmann I, Chiolero A, NICER Working Group. Overdiagnosis and overtreatment of thyroid cancer: A population-based temporal trend study. PloS one. 2017 Jun 14;12(6):e0179387.

5. Koelzer VH, Sirinukunwattana K, Rittscher J, Mertz KD. Precision immunoprofiling by image analysis and artificial intelligence. Virchows Archiv. 2019 Apr 1;474:511-22.

6. Fallahi P, Ferrari SM, Elia G, Ragusa F, Patrizio A, Paparo SR, Marone G, Galdiero MR, Guglielmi G, Foddis R, Cristaudo A. Primary cell cultures for the personalized therapy in aggressive thyroid cancer of follicular origin. InSeminars in cancer biology 2022 Feb 1 (Vol. 79, pp. 203-216). Academic Press.

7. Persichetti A, Di Stasio E, Coccaro C, Graziano F, Bianchini A, Di Donna V, Corsello S, Valle D, Bizzarri G, Frasoldati A, Pontecorvi A. Inter-and intraobserver agreement in the assessment of thyroid nodule ultrasound features and classification systems: a blinded multicenter study. Thyroid. 2020 Feb 1;30(2):237-42.

8. Cibas ES, Ali SZ. The 2017 Bethesda system for reporting thyroid cytopathology. Thyroid. 2017 Nov 1;27(11):1341-6.

9. Baloch ZW, LiVolsi VA. Follicular-patterned lesions of the thyroid: the bane of the pathologist. American journal of clinical pathology. 2002 Jan 1;117(1):143-50.

10. LiVolsi VA, Baloch ZW. Follicular neoplasms of the thyroid: view, biases, and experiences. Advances in Anatomic Pathology. 2004 Nov 1;11(6):279-87.

11. Hadjiiski L, Sahiner B, Chan HP. Advances in CAD for diagnosis of breast cancer. Current opinion in obstetrics & gynecology. 2006 Feb;18(1):64.

12. Brattain LJ, Telfer BA, Dhyani M, Grajo JR, Samir AE. Machine learning for medical ultrasound: status, methods, and future opportunities. Abdominal radiology. 2018 Apr;43:786-99.

13. Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX. Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing. 2019 Jan 15;115:213-37.

14. Daniels K, Gummadi S, Zhu Z, Wang S, Patel J, Swendseid B, Lyshchik A, Curry J, Cottrill E, Eisenbrey J. Machine learning by ultrasonography for genetic risk stratification of thyroid nodules. JAMA Otolaryngology–Head & Neck Surgery. 2020 Jan 1;146(1):36-41.

15. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer informatics. 2006 Jan;2:117693510600200030.

16. Li X, Zhang S, Zhang Q, Wei X, Pan Y, Zhao J, Xin X, Qin C, Wang X, Li J, Yang F. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. The Lancet Oncology. 2019 Feb 1;20(2):193-201.

17. Mongan J, Moy L, Kahn Jr CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiology: Artificial Intelligence. 2020 Mar 25;2(2):e200029.

18. Simes RJ. Treatment selection for cancer patients: application of statistical decision theory to the treatment of advanced ovarian cancer. Journal of chronic diseases. 1985 Jan 1;38(2):171-86.

19. Astion ML, Wilding P. Application of neural networks to the interpretation of laboratory data in cancer diagnosis. Clinical Chemistry. 1992 Jan 1;38(1):34-8.

20. Fielding LP, Fenoglio-Preiser CM, Freedman LS. The future of prognostic factors in outcome prediction for patients with cancer. Cancer. 1992 Nov 1;70(9):2367-77.

21. Cochran AJ. Prediction of outcome for patients with cutaneous melanoma. Pigment cell research. 1997 Jun;10(3):162-7.

22. Burke HB, Bostwick DG, Meiers I, Montironi R. Prostate cancer outcome: epidemiology and biostatistics. Analytical and quantitative cytology and histology. 2005 Aug 1;27(4):211-7.

23. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer informatics. 2006 Jan;2:117693510600200030.

24. Sun R, Fei F, Wang M, Jiang J, Yang G, Yang N, Jin D, Xu Z, Cao B, Li J. Integration of metabolomics and machine learning revealed tryptophan metabolites are sensitive biomarkers of pemetrexed efficacy in non-small cell lung cancer. Cancer Medicine. 2023 Sep;12(18):19245-59.

25. Zhang B, Shi H, Wang H. Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach. Journal of Multidisciplinary Healthcare. 2023 Dec 31:1779-91.

26. Iqbal MJ, Javed Z, Sadia H, Qureshi IA, Irshad A, Ahmed R, Malik K, Raza S, Abbas A, Pezzani R, Sharifi-Rad J. Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future. Cancer cell international. 2021 Dec;21(1):1-1.

27. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Medicine. 2021 Dec;13(1):1-7.

28. Ryu HS, Jin MS, Park JH, Lee S, Cho J, Oh S, Kwak TY, Woo JI, Mun Y, Kim SW, Hwang S. Automated Gleason scoring and tumor quantification in prostate core needle biopsy images using deep neural networks and its comparison with pathologist-based assessment. Cancers. 2019 Nov 25;11(12):1860.

29. Pavlidis N, Pentheroudakis G. Cancer of unknown primary site. The Lancet. 2012 Apr 14;379(9824):1428-35.

30. Yoon J, Lee E, Koo JS, Yoon JH, Nam KH, Lee J, Jo YS, Moon HJ, Park VY, Kwak JY. Artificial intelligence to predict the BRAFV600E mutation in patients with thyroid

cancer. PloS one. 2020 Nov 25;15(11):e0242806.

31. Bellantuono L, Tommasi R, Pantaleo E, Verri M, Amoroso N, Crucitti P, Di Gioacchino M, Longo F, Monaco A, Naciu AM, Palermo A. An eXplainable Artificial Intelligence analysis of Raman spectra for thyroid cancer diagnosis. Scientific Reports. 2023 Oct 3;13(1):16590.

32. Ha EJ, Baek JH. Applications of machine learning and deep learning to thyroid imaging: where do we stand?. Ultrasonography. 2021 Jan;40(1):23.

33. Agarwal S, Yadav AS, Dinesh V, Vatsav KS, Prakash KS, Jaiswal S. By artificial intelligence algorithms and machine learning models to diagnosis cancer. Materials Today: Proceedings. 2023 Jan 1;80:2969-75.

34. Xi NM, Wang L, Yang C. Improving the diagnosis of thyroid cancer by machine learning and clinical data. Scientific Reports. 2022 Jul 1;12(1):11143.

35. Peng S, Liu Y, Lv W, Liu L, Zhou Q, Yang H, Ren J, Liu G, Wang X, Zhang X, Du Q. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. The Lancet Digital Health. 2021 Apr 1;3(4):e250-9.

36. Olatunji SO, Alotaibi S, Almutairi E, Alrabae Z, Almajid Y, Altabee R, Altassan M, Ahmed MI, Farooqui M, Alhiyafi J. Early diagnosis of thyroid cancer diseases using computational intelligence techniques: A case study of a Saudi Arabian dataset. Computers in Biology and Medicine. 2021 Apr 1;131:104267.
37. Zhou H, Liu B, Liu Y, Huang Q, Yan W. Ultrasonic intelligent diagnosis of papillary thyroid carcinoma based on machine learning. Journal of Healthcare Engineering. 2022 Jan 10;2022.

38. Zhao WJ, Fu LR, Huang ZM, Zhu JQ, Ma BY. Effectiveness evaluation of computer-aided diagnosis system for the diagnosis of thyroid nodules on ultrasound: A systematic review and meta-analysis. Medicine. 2019 Aug;98(32).

39. Kuang A, Kouznetsova VL, Kesari S, Tsigelny IF. Diagnostics of Thyroid Cancer Using Machine Learning and Metabolomics. Metabolites. 2023 Dec 22;14(1):11.