International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Operating Publisher
SciFormat Publishing Inc.
ISNI: 0000 0005 1449 8214

2734 17 Avenue SW,
Calgary, Alberta, T3E0A7,
Canada
+15878858911
editorial-office@sciformat.ca

ARTICLE TITLE	INTELLIGENT DIAGNOSTICS IN ENDOCRINOLOGY: THE POTENTIAL OF ARTIFICIAL INTELLIGENCE IN THE DETECTION AND EVALUATION OF THYROID NODULES
DOI	https://doi.org/10.31435/ijitss.1(49).2026.3998
RECEIVED	19 December 2025
ACCEPTED	03 March 2026
PUBLISHED	16 March 2026
LICENSE	The article is licensed under a **Creative Commons Attribution 4.0 International License.**

© The author(s) 2026.
This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

INTELLIGENT DIAGNOSTICS IN ENDOCRINOLOGY: THE POTENTIAL OF ARTIFICIAL INTELLIGENCE IN THE DETECTION AND EVALUATION OF THYROID NODULES

Agnieszka Kasprzak *(Corresponding Author, Email: akasprzak2810@gmail.com)*
Medical University of Gdańsk, Gdańsk, Poland
ORCID ID: 0009-0004-2252-5163

Angelika Malec
Medical University of Bialystok, Białystok, Poland
ORCID ID: 0009-0000-8152-476X

Katarzyna Oświeczyńska
Silesian Academy, Zbigniew Religa Faculty of Medical Sciences, Poland
ORCID ID: 0009-0000-8383-9175

Agnieszka Zaleszczyk
Rzeszów University, Collegium Medicum, Rzeszów, Poland
ORCID ID: 0009-0005-7639-2156

Patrycja Jędrzejewska-Rzezak
M.D., The John Paul II Catholic University of Lublin, Lublin, Poland
ORCID ID: 0000-0003-2144-5810

ABSTRACT

Advancements in artificial intelligence (AI) havesignificantly influenced the development of diagnostic tools in medicine, including endocrinology.Thyroid nodules are a prevalent clinical occurrence, and their accurate evaluation about malignancy risk is essential for treatment decisions. Conventional diagnostic techniques, like ultrasonography and fine-needle aspiration biopsy (FNAB),although effective, are burdened with subjectivity and limited availability of specialists.This article discusses the contemporary applications of artificial intelligence, including machine learning and deep learning algorithms, in the detection, classification, and evaluation of thyroid nodules utilizing ultrasound pictures and clinical data. The text compares the performance of AI models with expert evaluations and examines the possible advantages of their application in clinical practice, including enhanced diagnostic accuracy, standardized assessments, and decreased time to diagnosis.Despite the promising research findings, additional efforts are required to validate models in extensive populations and to integrate AI systems with current clinical protocols. Artificial intelligence possesses the potential to support as a substantial aid in the diagnosis of thyroid disorders; nevertheless, its comprehensive implementation necessitates the evaluation of ethical, legal, and organizational factors.

Aim of the study: The aim of this article is to assess the role of artificial intelligence in the diagnosis of thyroid nodules, with particular emphasis on the use of machine learning (ML) and deep learning (DL) algorithms in the analysis of ultrasound images, cancer risk classification, and clinical decision support. The article aims to compare the effectiveness of AI-based systems with radiological assessment performed by specialists and to discuss the potential benefits and limitations of implementing these technologies in endocrinology practice.

Materials and methods: A review of the literature available in the PubMed and Google Scholar databases was performed, using the key words: “Artificial Intelligence”, “AI”, “Thyroid Nodules”, “Fine Needle Aspiration Biopsy”, “FNAB”, “TIRADS”.

KEYWORDS

Artificial Intelligence, AI, Thyroid Nodules, Fine Needle Aspiration Biopsy, FNAB, TIRADS

CITATION

Agnieszka Kasprzak, Angelika Malec, Katarzyna Oświeczyńska, Agnieszka Zaleszczyk, Patrycja Jędrzejewska-Rzezak. (2026) Intelligent Diagnostics in Endocrinology: The Potential of Artificial Intelligence in the Detection and Evaluation of Thyroid Nodules. *International Journal of Innovative Technologies in Social Science*. 1(49). doi: 10.31435/ijitss.1(49).2026.3998

COPYRIGHT
© The author(s) 2026. This article is published as open access under the **Creative Commons Attribution 4.0 International License (CC BY 4.0)**, allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

1. Introduction

Thyroid nodules are a common clinical problem. They are detected by palpation in about 5% of people, while they can be incidentally detected by ultrasound in 60-70% of the population, especially in the elderly.Focal thyroid lesions incidentally revealed by PET scan may have up to a 35% risk of being malignant lesions. Evaluation of thyroid nodules is important because of the need to determine the hormonal function of the gland, the presence of local symptoms and the potential risk of malignancy[1].

Evaluation of a thyroid nodule begins with determination of thyrotropin (TSH) levels and ultrasound of the thyroid gland with evaluation of regional neck lymph nodes. Finding a normal or elevated TSH level suggests that the nodule has no hormonal activity. On the other hand, a decreased or suppressed TSH level may indicate primary hyperthyroidism, in which case thyroid scintigraphy with technetium-99 or iodine-123 is indicated. Increased radionuclide uptake within the nodule (“hot” nodule) argues for hyperthyroidism, which is usually associated with a low risk of malignancy and does not require fine-needle aspiration biopsy (FNAB). In contrast, inactive (“cold”) nodules that meet appropriate clinical or ultrasound criteria should undergo further diagnostics, including FNAB[2].

Fine-needle aspiration biopsy (FNAB) is a widely used, minimally invasive diagnostic procedure that involves collecting cells from an organ—most commonly the thyroid—for microscopic evaluation. The procedure is performed using a thin needle inserted directly into the focal lesion, allowing for the extraction of cellular material. This sample is then analyzed to detect abnormalities such as malignant features or signs of inflammation.

Fine-needle aspiration biopsy is a key diagnostic tool in the evaluation of thyroid nodules, providing the most precise information about their nature. It is a simple, safe procedure that is widely regarded as reliable. In cases where the nodule is difficult to locate by palpation or is cystic in nature, it is recommended to perform FNAB under ultrasound guidance, which increases the accuracy and safety of the procedure[3].

Algorithms based on deep learning utilize images fromimaging examinations (such as computed tomography, magnetic resonance imaging, USG or mammography) to identify subtle changes that may escape the human eye. The possibilities of using artificial intelligence methods in oncological diagnosis and therapy are widely researched in many cancers. As a result, AI increases the chances of early detection of cancer, while also minimizing the number of false positive results, which helps reduce patient stress and costs associated with unnecessary tests[4]. Artificial intelligence also enables more personalized diagnostics. By processing vast amounts of data, AI can tailor the analysis to an individual patient's characteristics, such as age, gender, medical history, or genetic factors, leading to more accurate diagnoses and more effective treatment plans.

2. Fine-needle biopsy as the gold standard for diagnosing thyroid nodules

Fine-needle aspiration biopsy (FNAB) is considered the gold standard in diagnosing thyroid nodules. The test is safe for the patient and precise with very high sensitivity and specificity near to 100% in the diagnosis of thyroid cancer[5]. Ultrasound-guided FNA is frequently conducted in response to suspicious sonographic characteristics observed in an initial thyroid ultrasound or alterations in imaging features during subsequent evaluations. The primary objective of FNA is to categorize the aspirated lesion as benign, suspicious, or malignant, hence enabling a more accurate assessment of the reasons for surgical intervention.

To mitigate diagnostic errors and enhance clinical decision-making in thyroid fine-needle aspiration cytology, a standardized reporting terminology for thyroid fine-needle aspiration cytology and all thyroid aspirates should be implemented. The initial recommendation for a cytological classification of thyroid lesions was the Papanicolaou Society classification established in 2005. The subsequent approach for reporting thyroid cytology, released a few years later, was The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC). The third edition from 2023 is presently in effect[6]. TBSRTC is a worldwide accepted reporting system for thyroid FNAs[7]. It comprises six categories classified based on the malignancy risk. The Bethesda categories III (atypia of undetermined significance [AUS] or follicular lesion of undetermined significance

[FLUS]) and IV (follicular neoplasm [FN] or suspected follicular neoplasm [SFN]) represent the most challenging classifications, referred to as "indeterminate" nodules, which exhibit a significantly elevated risk of malignancy, ranging from 15.7% to 54.6% for category III and from 16.8% to 72.4% for category IV[8]. Consequently, these ambiguous lesions present numerous diagnostic issues due to their diverse histological findings, ranging from benign to malignant, with considerable variability in malignancy incidence[9].

Alternative reporting terminologies are also accessible. The British *Thy* nomenclature, the Italian TIR terminology, the Australian terminology, and the Japanese system exist[10]. All of these systems exhibit significant similarities, albeit with slight distinctions. All efforts to categorize and systematize thyroid FNAC based on the likely histological diagnosis aim to provide an indicator of malignancy risk and to recommend suitable therapeutic strategies.

Regrettably, despite the implementation of uniformity in cytological reporting, considerable discrepancies persist in the interpretation of results and the associated malignancy risk[11]. Analyses demonstrate considerable diversity in the cytological evaluation of thyroid nodules, both among different pathologists and within the same observer over multiple occasions[12]. Findings demonstrate the necessity for ongoing evaluation and consultation in instances of ambiguous results.

The incidence of fine-needle biopsies of the thyroid has markedly escalated in recent decades. This results from advancements in imaging techniques, as progressively superior ultrasound apparatus enables the identification of diminutive, clinically irrelevant nodules (microcancers) and the apprehensions of physicians and patients regarding the oversight of a malignant tumor. The advancement of artificial intelligence techniques is anticipated to enhance the patient qualification procedure for biopsy, hence increasing the precision of diagnostic determinations.

3. Consequences of thyroid biopsy: potential complications, costs and patient stress

Potential Complications:

FNAB is a simple and safe procedure that has reduced the percentege of patients undergoing unnecessary thyroidectomy. It is a much less invasive method than core-needle or surgical biopsy, which involves a skin incision and may require general anesthesia. Recovery time after FNAB is short, and most patients are able to return to their daily activities on the same day. This is not the case with surgical methods, where recovery often exceeds 24 hours and involves certain limitations in daily functioning [13]

The entire FNAB procedure is performed under ultrasound guidance, which minimizes the risk of complications. Local pain and hematomas are the most common problems, while serious adverse events are rare.Other serious complications are extremely rare and include acute transient thyroid swelling, infection, tracheal puncture, needle seeding (i.e. tumor implantation), and recurrent laryngeal nerve injury with vocal cord paralysis[13]

Costs:

Fine-needle aspiration biopsy is associated with significant costs for both the healthcare system and patients. The prevalence of thyroid nodules in the population is quite high and is estimated at 60-70%, as already stated, with higher occurrence in older individuals and in women compared to men. Most detected nodules are benign, and the malignancy rate is only about 4–6.5% [7].

Approximately 41% of healthcare spending is directed toward patients who have just been diagnosed. This stage involves an initial visit to a general practitioner, followed by consultation with a specialist (e.g., oncologist or endocrinologist), ultrasound imaging, and FNAB, and possibly a subsequent cytological assessment if the initial FNAB result is inconclusive. All of these steps significantly raise the costs associated with managing thyroid nodules, both benign and malignant [14]

One meta-analysis conducted in the United States showed that as many as 68.8% of thyroid nodules that underwent surgical removal were benign, leading to unnecessary and excessive healthcare spending [15]

From the patient's perspective, the financial burden is also substantial, as many individuals cover these costs out of pocket despite partial reimbursement. This includes diagnostic tests, transportation, and income loss due to work absence or the need to quit a job. In extreme cases, this can result in debt due to incurred medical expenses. Consequently, some patients interrupt treatment or opt out of recommended diagnostic procedures, which adversely affects their health status. Even patients whose disease is in remission still experience fear of recurrence and must continue to bear the cost of follow-up monitoring [15].

Patient stress:

Thyroid biopsy, as well as the uncertainty associated with the nature of the lesion, may contribute to increased emotional distress.There is not much research in the literature specifically focused on stress related to thyroid biopsy. However, Chinese researcher Li et al. [16]conducted a study among adult patients, analyzing sleep disturbances occurring simultaneously with, among other things, thyroid nodule screening procedures. Their observations indicate that individuals with unidentified thyroid nodules had poorer sleep quality and decreased psychological well-being.

4. Artificial Intelligence in Thyroid Imaging Diagnostics

4.1.TI-RADS Classification

A standardized system for the classification and description of focal thyroid gland lesions in ultrasound imaging was introduced for image analysis, developed by the American College of Radiology (ACR), known as the TI-RADS (Thyroid Imaging Reporting and Data System) [17]. In addition to this system, other systems also exist[18]:

- EU-TIRADS (European Thyroid Imaging Reporting and Data System)
- K-TIRADS (Korean Thyroid Imaging Reporting and Data System),
- C-TIRADS (Chinese TI-RADS).

The highest sensitivity is demonstrated by K-TIRADS, which is characterized by 86% diagnostic accuracy, whereas ACR TI-RADS presents higher specificity at 64%. In the above-mentioned classification systems, tissue elasticity and vascularity are not assessed, therefore a multimodal scoring system was developed – the French TIRADS model. The Thyroid Multimodal Imaging Comprehensive system surpasses the diagnostic performance of the American model, as it performs broader analyses of ultrasound image features [19]

The aim of integrating the TI-RADS system with artificial intelligence (AI) is to further enhance the automation of ultrasound image analysis using computer algorithms based on TI- RADS guidelines. The system performs analysis based on trained datasets and features, and then proposes the nature of the lesion – whether it is benign or potentially malignant. The use of these methods will contribute to: reducing the number of fine-needle aspiration biopsies (FNAB), increasing the effectiveness of thyroid nodule recognition, and minimizing discrepancies in interpretation between specialists[20]

The standard ACR TI-RADS classification involves the evaluation of characteristic features in the ultrasound image of the thyroid and assigning individual points for a given feature, which may increase the risk of malignancy. The final sum of all points determines the nodule's final category in the TR1–TR5 range.

- TR1 is the normal category (0% risk), which means that the malignancy risk is negligible and no intervention is required.
- TR2 is a benign lesion (malignancy risk up to 2%) and requires ultrasound follow-up only if other concerning symptoms accompany the lesion.
- TR3 represents a low-risk nodule, as the malignancy risk is approximately 5%, and also requires periodic follow-up examinations.
- TR4 is associated with an increased risk, as the malignancy rate is around 5–20% according to the classification, and a thyroid cell biopsy should be considered if the lesion is larger than 1.5 cm, whereas smaller lesions should be monitored.
- The TR5 nodule classification is assigned more than 6 points. The set of morphological features in the ultrasound image indicates a high malignancy risk (over 20%) of the diagnosed thyroid nodule, and biopsy is recommended for nodules larger than 1 cm.

If the nodule is smaller than 5 mm and classified within TI-RADS, it is most likely not a malignant lesion. However, if the lesion exceeds 2.5 cm in TR3, it becomes concerning[17][18].

As mentioned earlier, focal lesions in ultrasound imaging can be assessed based on the analysis of characteristic morphological features, which are assigned specific points in the TI-RADS classification:

1. Composition - we can distinguish different types:

- Cystic nodules,
- Spongi formnodules,
- Mixed nodules,
- Solid (dense),

Cystic and spongiform nodules are not relevant in diagnostics, as they are benign in nature. Mixed nodules may consist of a solid and a cystic part. The proportion of the solid to the cystic component matters even when the nodule is large; if the solid part is suspicious in nature, a fine-needle aspiration biopsy should be performed. A suspicious lesion may be hypoechoic, contain microcalcifications, have an acute angle at the wall of the nodule, or have irregular margins.

2. Echogenicity

- Anechoiclesions,
- Hyperechoiclesions,
- Hypoechoiclesions,
- Isoechoiclesions.

In the examination, the echogenicity of the thyroid gland is compared to that of the surrounding muscles. If the nodule has reduced echogenicity and is similar to the surrounding muscles, it may result in a higher TI-RADS score by 3 points. In turn, anechoic, meaning fluid-filled cysts, are extremely hypoechoic on ultrasound and are classified as benign lesions and receive 0 points.

3. Shape

A specific dimension of the nodule, the measured shape of the nodule is taller than wide in the transverse plane, which may suggest malignancy and increases the TI-RADS score by 3 points. A taller-than-wide dimension is usually of no concern, although it does not exclude the severity of the lesion.

4. Margins

- Irregular,
- Regular,
- Infiltrative.

Regular margins, meaning well-circumscribed, receive 0 points. Irregular margins, meaning ragged edges with protrusions into the adjacent parenchyma, receive 2 points. Sometimes, the thyroid capsule may be breached, and the dimensions of the lesion extend beyond the thyroid (minimal or extensive infiltration) into the surrounding tissues or vessels. Additionally, if the echogenic line of the capsule disappears, or bulging of the contour is visible, it may indicate high-grade malignancy. Sometimes it is difficult to define the margin of the lesion, so 0 points are assigned due to its location.

5. Echogenic Foci

- “Comet-tail” artifact,
- Microcal cifications,
- Macrocal cifications,
- Peripheral calcifications.

The so-called “comet-tail” artifact on ultrasound appears as hyperechoic bands forming a V- shape and generally indicates a benign lesion (0 points). On the other hand, microcalcifications up to 1 mm, which have characteristic hyperechoic dots without acoustic shadowing, are usually malignant (they may indicate papillary carcinoma) and require more detailed diagnostics. If macrocalcifications with acoustic shadowing and hyperechogenicity are present, such lesions may be malignant, but not necessarily (may be benign), and receive 1 point. Discontinuous peripheral calcifications at the margin of the lesion may be concerning when they appear as hyperechoic foci and protrude into the thyroid tissue - 2 points are assigned in the classification [18].

4.2. Examples of AI-based systems

Despite the introduction of the TI-RADS classification, differences in the interpretation of thyroid ultrasound findings between various radiologists still occur in clinical practice. These observations result from the subjective assessment of the physician regarding the classification of the nodule, which contributes to performing a higher number of fine-needle aspiration biopsies (FNAB) in order to clarify the nature of the lesion through histopathological examination. Therefore, AI is gaining increasing popularity, supported by deep learning algorithms and the analysis of visual similarities in ultrasound images, which contributes to supporting the analysis of ultrasound images in patients. As a result, this will support the diagnostic decision-making process and reduce uncertainty regarding the nature of the lesion, especially among less experienced specialists[21].

Examples of artificial intelligence–based systems can be divided into:

- Deep neural networks (Deep Learning): ResNet-50, DenseNet121, EfficientNet B3/B7, MobileNetV2, VGG16, InceptionV3, InceptionResNetV2, GoogleNet[22][23]
- Classical algorithms (Machine Learning): Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Random Forest (RF), Logistic Regression (LG), GlmNet, K-Nearest Neighbors (K-NN), XGBoost, and KNN [24].

All of the above-mentioned classical algorithms demonstrate similar effectiveness in the diagnosis of thyroid nodules; however, LG achieved better performance compared to the other classifiers. DenseNet169 showed the highest diagnostic accuracy at 95.96%, followed by ResNet101 at 94.74%, and EfficientNetB1 at 86.14%.

These algorithms demonstrate that artificial intelligence may offer more accurate analysis than standard diagnostic approaches, which are based on the subjective interpretation of the clinician[22][23][24].

4.3. Comparison of Diagnostic Performance: Radiologists vs Artificial Intelligence in Thyroid Ultrasound Assessment

As already stated, accurate interpretation of thyroid ultrasound findings is dependent on the physician's skills and experience; therefore, the evaluation of sonographic features may vary between clinicians. This contributes to discrepancies in diagnosis, as less experienced specialists tend to make more errors, which in turn leads to an excessive number of referrals for thyroid FNAB, especially in individuals with a low risk of malignancy[25].

In the context of these challenges, a study conducted by Weng et al. in 2023 [26]aimed to compare the accuracy of an AI algorithm with radiologists' interpretation of thyroid ultrasound images, in order to classify thyroid nodules using the receiver operating characteristic (ROC) curve analysis and its corresponding area under the curve (AUC) metric. Based on a dataset of 1,278 thyroid nodules, the deep learning algorithm was trained and then tested on an independent set of 99 cases. Additionally, to evaluate the generalizability of the AI system, ultrasound images from various types and models of devices were used to allow for performance comparison across different ultrasound equipment manufacturers. In total, 378 ultrasound images were included in the evaluation. Concurrently, four radiologists (with 2 to 22 years of clinical experience) analyzed the same 378 cases, assigning each lesion a category and characteristics, as well as determining whether FNAB was indicated.The results for the AI algorithm showed an AUC of approximately 0.69, while the AUC values for radiologists ranged from 0.63 to 0.66 (AUC scale ranges from 0 to 1.0). A higher AUC value indicates better diagnostic performance. The deep learning algorithm demonstrated slightly better accuracy in distinguishing benign from malignant nodules, whereas radiologists achieved comparable values in their routine clinical assessments. This suggests that AI tools may offer objective diagnostic support in the evaluation of thyroid nodules, particularly in settings with limited access to radiologists, or as an aid for less experienced physicians.

Another study confirming the comparable diagnostic sensitivity of AI to that of radiologists was conducted by Choi et al.[27], involving 89 patients with 59 benign and 43 malignant thyroid nodules. A computer-assisted diagnostic system achieved an overall diagnostic accuracy of 90.7%, which was similar to the sensitivity demonstrated by radiologists (88.4%). Furthermore, the agreement between AI and radiologists included consistent evaluations of features such as echogenicity and spongiform appearance of the nodules.

Additional confirmation came from a study led by Zhao et al.[28], which analyzed 723 thyroid nodules in 536 patients. The findings again demonstrated similar sensitivity in differentiating thyroid lesions between AI and radiologists, although specificity was reduced in the AI group.

Therefore, the implementation of artificial intelligence in thyroid ultrasound diagnostics has the potential to enhance the accuracy and efficiency of image interpretation, reduce the time required for image analysis by physicians across specialties, support radiologists in achieving more precise diagnoses, and decrease their workload[17]. However, further large-scale meta-analyses are necessary to improve diagnostic precision and to establish AI as a reliable support tool in clinical practice.

5. Reduction of Unnecessary Fine-Needle Aspiration Biopsies in Thyroid Nodule Diagnostics Through the Use of Artificial Intelligence

Research on the application of artificial intelligence (AI) in thyroid imaging diagnostics highlights its potential to reduce the number of unnecessary fine-needle aspiration (FNA) biopsies, and in some cases, even to outperform human experts in the interpretation of medical images. In the context of thyroid nodule assessment, numerous studies have compared the classification performance of AI systems with that of clinicians. Pilot studies have demonstrated that machine learning-based systems often achieve higher accuracy in detecting malignant lesions than traditional methods based on the American College of Radiology's TI-RADS classification. Furthermore, the introduction of deep learning techniques into thyroid imaging has led to improved diagnostic outcomes, often surpassing the performance of experienced radiologists, as confirmed by several independent investigations[28].

The use of AI in differentiating thyroid nodules represents a major step forward in enhancing diagnostic accuracy and reducing the frequency of unnecessary biopsies. Several computer-aided diagnosis (CAD) systems such as S-Detect, AmCAD-UT, Koios DS, and Medo Thyroid—have received FDA approval and have demonstrated promising performance in analyzing thyroid ultrasound images. In particular, the S-Detect system, which employs a semi-automated workflow, serves as a valuable decision-support tool for clinicians by reducing the likelihood of overlooking malignant changes. The latest iterations of this system are capable of identifying calcifications as a key diagnostic feature[29].

One notable analysis [30]involved a retrospective, multicenter diagnostic accuracy study based on data from the Open AI Dataset Project. The dataset comprised digital images obtained from three university hospitals and 215 other institutions. Images underwent standardized quality control during case selection, scanning, labeling, and review. Multilayer images were acquired using three different scanners, and patches were extracted from whole-slide images (WSIs) before undergoing focus fusion and color normalization. Among the six AI models tested, Inception ResNet v2 achieved the highest performance and was subsequently used to analyze the complete dataset. In the second phase of the study, the model's performance was compared to that of expert cytopathologists using 1,031 randomly selected image patches. Expert evaluations were repeated after reviewing the AI-generated results.The model was trained on 10,332 image patches derived from 306 thyroid FNA cases, including 78 malignant cases (papillary thyroid carcinoma) and 228 benign cases, collected from 86 centers. The Inception ResNet v2 model achieved accuracies of 99.7% on the training set, 97.7% on the validation set, and 94.9% on the test set. Corresponding sensitivities were 99.9%, 99.6%, and 100%, while specificities were 99.6%, 96.4%, and 90.4%, respectively. In direct comparison with cytopathologists, the AI model demonstrated significantly superior performance: accuracy (99.71% vs. 88.91%), sensitivity (99.81% vs. 87.26%), and specificity (99.61% vs. 90.58%). When expert assessments were repeated with access to the AI output, diagnostic accuracy improved for all evaluators (96%, 95%, and 96%, respectively), and interobserver agreement increased from a kappa value of 0.64 to 0.84. These findings suggest that the integration of AI models into thyroid FNA cytology workflows can substantially improve diagnostic accuracy and reduce interobserver variability.

The application of AI-particularly machine learning (ML) and deep learning (DL) algorithms-provides significant value to thyroid nodule diagnostics. Among the key advantages is high diagnostic efficacy: numerous studies have shown that AI models achieve sensitivity and specificity comparable to, or exceeding, those of experienced radiologists, particularly in distinguishing between benign and malignant lesions. This capability facilitates a reduction in the number of unnecessary FNA biopsies, thereby lowering patient burden and reducing healthcare costs.

AI also supports less experienced clinicians by shortening the learning curve and improving diagnostic decision-making accuracy. The automation of ultrasound image evaluation accelerates the diagnostic process and helps standardize image interpretation, thereby reducing interobserver variability.

Despite these advantages, AI technologies present important limitations. Many models are trained on hospital-based datasets, which may limit their generalizability to broader populations or screening contexts. Furthermore, model performance is highly dependent on the quality and diversity of training data; a lack of such diversity can lead to misclassification, particularly in rare lesion subtypes. Another major challenge is the so-called "black box" problem, in which the decision-making process of AI models is not fully transparent, making it difficult for clinicians to interpret or trust the model's output[31].

Additionally, most currently available models remain in the validation or clinical evaluation stage, and their effectiveness in real-world clinical practice requires further large-scale, multicenter, prospective studies involving diverse patient populations[31].

6. Exploring the Advancements in Thyroid Nodule Diagnostics through Artificial Intelligence

6.1. AI Models in the Diagnosis of Thyroid Nodules

The increasing prevalence of thyroid nodules and the associated burden on healthcare systems necessitate the search for new solutions aimed not only at improving the diagnostic process but also at supporting therapeutic decision-making. A new and promising approach is the use of artificial intelligence (AI) systems. Depending on the model, AI can perform various functions: from assisting in the classification of nodules based on malignant features in ultrasound examinations, through the development of clinical decision support systems (CDSS), to the standardization of fine-needle aspiration biopsy (FNAB) results. The development of hybrid models, which combine different data analysis methods processed in separate stages, has enabled the integration of AI with physician experience, clinical data, and laboratory results.

An example of such a model is Halbrid+, based on the ResNet101 algorithm, which is used to extract features from ultrasound images of thyroid nodules. This model utilizes TIRADS classification features, and the use of the Global Attention Feature Map (GAFM) has allowed better consideration of the interactions between image features and clinical data. The model was trained on a dataset of ultrasound images of thyroid nodules and patient demographic data. Evidence of its precise ability to distinguish between malignant and benign nodules is the obtained AUC value of 0.92[32]. The study found that the standalone CADx model, a computer-aided diagnosis system, achieved a slightly better result than an experienced radiologist (AUC ~0.87 vs. ~0.86). It is worth noting that the ResNet101 model used in this study is dependent on the quality of input data if the data quality is low, diagnostic effectiveness will be significantly reduced.

In recent years, the development of multimodal models which, unlike hybrid models, integrate various data sources has advanced. One such model is described in the study by Pathak et al. [33] It uses natural language processing (NLP) to interpret ultrasound examination reports and incorporates the patient's clinical data (age, sex) and thyroid disease history. The study focused on the automatic extraction of relevant features of thyroid nodules from ultrasound reports. Among the models used, GatorTron, which was trained on over 90 billion words, achieved the highest performance. Its F1 score of 0.9321 demonstrates the high effectiveness and precision of this approach. At the same time, it should be noted that a limitation of both this model and the previously described one is their reliance on datasets whose quality can significantly influence the results. Another challenge is limited generalization due to demographic differences, stemming not only from ethnic diversity but also from dietary habits and lifestyle. Therefore, AI models must continue to be tested to enhance their utility in clinical practice.

Research by Yuet al. [34] focused on a model using histopathological, genomic, and immunological data. The goal was to create a model capable of detecting features of tumor molecular heterogeneity, which currently pose a major diagnostic challenge. The model also assessed the predictability of lymph node metastasis and recurrence-free survival in patients with papillary thyroid cancer. Achieving an AUC of 0.864, it confirmed previous findings that AI-based models significantly improve diagnostic effectiveness.

The research results using AI models are promising. However, further testing is essential to increase their accuracy and better utilize their potential in clinical practice. It is also important to consider challenges related to the implementation of such solutions, including ethical issues, patient data protection, and the need for continued technological development.

6.2. The Role of AI in Cytological Interpretation

In a study conducted by Slabaughet al. [35] attention was drawn to the challenges pathomorphologists face when evaluating cytological images, which often result in significant diagnostic discrepancies. The authors also noted that one of the main reasons for the growing interest in AI is the increasing shortage of qualified specialists in this field. According to the researchers, AI can not only improve diagnostic accuracy and introduce greater standardization of results but also help relieve physicians and support the education of junior specialists. Similarly, in the study, the results of which were already discussed in the previous chapter[30], assessed the influence of AI on the diagnostic precision of fine-needle aspiration cytology (FNAC) and interobserver variability. These results suggest that artificial intelligence can have a crucial impact on enhancing the accuracy of thyroid FNAC diagnoses, contributing to greater standardization of diagnostic outcomes-a significant step toward improving the quality of work in pathological diagnostics.

6.3. Personalized Medicine

Artificial intelligence (AI) enables a more precise approach to diagnostics and treatment by creating individualized risk profiles for patients. AI's ability to integrate genomic (e.g., BRAF, RAS mutations), cytological, molecular, and clinical data facilitates the personalization of diagnostic and therapeutic decisions. In studies exploring the use of AI in the diagnosis and prognosis of thyroid diseases—especially in the context of thyroid nodules—the concept of "patient trajectory models" is increasingly emerging. These are advanced predictive models that analyze clinical, imaging, and laboratory data to forecast the individual course of disease in patients. Thanks to AI development, it is now possible to quickly identify patients at high risk for developing thyroid nodules. In the study of Liang et al. [36], the authors focused on developing ML models that estimate the 3-year risk of thyroid nodule occurrence based on demographic features (age, sex) and laboratory test results. Among the models tested, the XGBoost algorithm achieved the highest sensitivity (66%). The most influential parameters were HDL levels, age, fasting glucose, and creatinine levels. One limitation of the study is that the data came from a single center and the number of variables was limited, excluding other risk factors such as smoking.

6.4. Cost Reduction with AI Use

The application of AI in healthcare can lead to significant cost reduction by decreasing the number of unnecessary FNAC procedures[29][37]. Reducing unnecessary interventions leads to fewer hospitalizations, anesthetic procedures, and recovery costs. In the long term, AI may contribute to relieving specialist centers and improving access to high-quality diagnostics. Automated systems enable faster diagnosis and ensure result standardization across different centers, improving the efficiency of the healthcare system as a whole. The educational value of AI models is also noteworthy, serving as invaluable support for young specialists.

7.Conclusions

Artificial intelligence appears to be a promising tool that enables not only a reduction in the number of biopsies but also the standardization of ultrasound results. The use of hybrid and multimodal models in the diagnosis of thyroid nodules may improve access to specialists in regions affected by workforce shortages. Furthermore, these systems may serve as auxiliary tools and educational support for less experienced physicians. AI allows for the creation of highly effective predictive models that can assist in decisions regarding observation, recurrence risk after treatment, repeat biopsies, or surgical intervention. AI is already transforming nodule diagnostics, and further development of this technology could make diagnosis more precise, less invasive, and more accessible.

Despite promising results, the implementation of AI models in thyroid nodule diagnostics comes with challenges, such as the need to collect large, diverse datasets and ensure algorithm interpretability. A significant limitation is the dependence of diagnostic effectiveness on input data quality. Future research should focus on improving data quality and minimizing error risks due to AI system imperfections, while also addressing ethical and regulatory considerations. The deployment of AI systems requires appropriate technical infrastructure and personnel training.

Disclosure
Author'scontribution: Patrycja Jędrzejewska-Rzezak
Conceptualisation: Agnieszka Kasprzak
Methodology: Katarzyna Oświeczyńska
Software: Agnieszka Zaleszczyk
Check: Angelika Malec
Formal: Katarzyna Oświeczyńska
Investigation: Agnieszka Zaleszczyk, Angelika Malec,Agnieszka Kasprzak
Resources: Agnieszka Zaleszczyk, Angelika Malec,Patrycja Jędrzejewska-Rzezak
Datacuration: Patrycja Jędrzejewska-Rzezak,Katarzyna Oświeczyńska
Writing-Rough Preparation: Angelika Malec
Writing-Review and Editing: Patrycja Jędrzejewska-Rzezak, Katarzyna Oświeczyńska
Visualisation: Agnieszka Zaleszczyk
Supervision: Agnieszka Kasprzak
ProjectAdministration: Agnieszka Kasprzak

All authors have read and agreed with the published version of the manuscript.

Funding statement: The study did not receive special funding.
Institutional review board statement: Not applicable.
Informed consent statement: Not applicable.
Data availability statement: Not applicable.
Conflict of interest: The authors declare no conflict of interest.

REFERENCES

1. Pitoia, F., & Trimboli, P. (2024). New insights in thyroid diagnosis and treatment. *Reviews in Endocrine and Metabolic Disorders, 25*, 1–3. https://doi.org/10.1007/s11154-023-09859-5
2. Antonia, T. D., Maria, L. I., & Ancuta-Augustina, G. G. (2023). Preoperative evaluation of thyroid nodules: Diagnosis and management strategies. *Pathology - Research and Practice, 246*, Article 154516. https://doi.org/10.1016/j.prp.2023.154516
3. AlSaedi, A. H., Almalki, D. S., & ElKady, R. M. (2024). Approach to thyroid nodules: Diagnosis and treatment. *Cureus*. https://doi.org/10.7759/cureus.52232
4. Lotter, W., Hassett, M. J., Schultz, N., Kehl, K. L., Van Allen, E. M., & Cerami, E. (2024). Artificial intelligence in oncology: Current landscape, challenges, and future directions. *Cancer Discovery, 14*, 711–726. https://doi.org/10.1158/2159-8290.CD-23-1199
5. Lebrun, L., & Salmon, I. (2024). Pathology and new insights in thyroid neoplasms in the 2022 WHO classification. *Current Opinion in Oncology, 36*, 13–21. https://doi.org/10.1097/CCO.0000000000001012
6. Ali, S. Z., Baloch, Z. W., Cochand-Priollet, B., Schmitt, F. C., Vielh, P., & VanderLaan, P. A. (2023). The 2023 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid*. https://doi.org/10.1089/thy.2023.0141
7. Bagıs, M., Can, N., Sut, N., Tastekin, E., Erdogan, E. G., Bulbul, B. Y., et al. (2024). A comprehensive approach to the thyroid Bethesda category III (AUS) in the transition zone between 2nd edition and 3rd edition of the Bethesda System for Reporting Thyroid Cytopathology: Subcategorization, nuclear scoring, and more. *Endocrine Pathology, 35*, 51–76. https://doi.org/10.1007/s12022-024-09797-1
8. Zahid, A., Shafiq, W., Nasir, K. S., Loya, A., Abbas Raza, S., Sohail, S., et al. (2021). Malignancy rates in thyroid nodules classified as Bethesda categories III and IV: A subcontinent perspective. *Journal of Clinical and Translational Endocrinology, 23*, Article 100250. https://doi.org/10.1016/j.jcte.2021.100250
9. Lind, P., Jacobson, A., Nordenström, E., Johansson, L., Wallin, G., & Daskalakis, K. (2024). Diagnostic sensitivity of fine-needle aspiration cytology in thyroid cancer. *Scientific Reports, 14*, Article 24216. https://doi.org/10.1038/s41598-024-75677-7
10. Poller, D. N., Johnson, S. J., & Bongiovanni, M. (2020). Measures to reduce diagnostic error and improve clinical decision making in thyroid FNA aspiration cytology: A proposed framework. *Cancer Cytopathology, 128*, 917–927. https://doi.org/10.1002/cncy.22309
11. Machała, E., Sopiński, J., Iavorska, I., & Kołomecki, K. (2018). Correlation of fine needle aspiration cytology of thyroid gland with histopathological results. *Polish Journal of Surgery, 90*, 1–5. https://doi.org/10.5604/01.3001.0012.4712

12. Sripodok, S., & Benjakul, N. (2023). Interobserver variability in inconclusive diagnostic categories of thyroid fine needle aspiration cytology: An urban-based tertiary hospital experience. *Annals of Diagnostic Pathology, 63*, Article 152083. https://doi.org/10.1016/j.anndiagpath.2022.152083
13. Ha, E. J., Suh, C. H., & Baek, J. H. (2018). Complications following ultrasound-guided core needle biopsy of thyroid nodules: A systematic review and meta-analysis. *European Radiology, 28*, 3848–3860. https://doi.org/10.1007/s00330-018-5367-5
14. Uppal, N., Collins, R., & James, B. (2023). Thyroid nodules: Global, economic, and personal burdens. *Frontiers in Endocrinology, 14*. https://doi.org/10.3389/fendo.2023.1113977
15. Tamhane, S., & Gharib, H. (2016). Thyroid nodule update on diagnosis and management. *Clinical Diabetes and Endocrinology, 2*. https://doi.org/10.1186/s40842-016-0035-7
16. Li, R., Li, G., Wang, Y., Bao, T., Lei, Y., Tian, L., et al. (2021). Psychological distress and sleep disturbance throughout thyroid nodule screening, diagnosis, and treatment. *Journal of Clinical Endocrinology & Metabolism, 106*, E4221–E4230. https://doi.org/10.1210/clinem/dgab224
17. Tessler, F. N., Middleton, W. D., Grant, E. G., Hoang, J. K., Berland, L. L., Teefey, S. A., et al. (2017). ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White paper of the ACR TI-RADS Committee. *Journal of the American College of Radiology, 14*, 587–595. https://doi.org/10.1016/j.jacr.2017.01.046
18. Ludwig, M., Ludwig, B., Mikuła, A., Biernat, S., Rudnicki, J., & Kaliszewski, K. (2023). The use of artificial intelligence in the diagnosis and classification of thyroid nodules: An update. *Cancers, 15*. https://doi.org/10.3390/cancers15030708
19. Shreyamsa, M., Mishra, A., Ramakant, P., Parihar, A., Singh, K., Rana, C., et al. (2020). Comparison of multimodal ultrasound imaging with conventional ultrasound risk stratification systems in presurgical risk stratification of thyroid nodules. *Indian Journal of Endocrinology and Metabolism, 24*, 537–542. https://doi.org/10.4103/ijem.IJEM_675_20
20. Zhao, C. K., Ren, T. T., Yin, Y. F., Shi, H., Wang, H. X., Zhou, B. Y., et al. (2021). A comparative analysis of two machine learning-based diagnostic patterns with thyroid imaging reporting and data system for thyroid nodules: Diagnostic performance and unnecessary biopsy rate. *Thyroid, 31*, 470–481. https://doi.org/10.1089/thy.2020.0305
21. Gruson, D., Dabla, P., Stankovic, S., Homsak, E., Gouget, B., Bernardini, S., et al. (2022). Artificial intelligence and thyroid disease management: Considerations for thyroid function tests. *Biochemia Medica, 32*. https://doi.org/10.11613/BM.2022.020601
22. Sureshkumar, V., Jaganathan, D., Ravi, V., Velleangiri, V., & Ravi, P. (2024). A comparative study on thyroid nodule classification using transfer learning methods. *Open Bioinformatics Journal, 17*. https://doi.org/10.2174/0118750362305982240627034926
23. Amer, H. M., Nasr, S. A., Abdel-Fattah, H. M., Abdelsalam, M. M., & El-Din Moustafa, H. (2023). An accurate deep learning based framework for detection of thyroid cancer using ultrasound images. *Vol. 24*.
24. Liang, J., Pang, T., Liu, W., Li, X., Huang, L., Gong, X., et al. (2023). Comparison of six machine learning methods for differentiating benign and malignant thyroid nodules using ultrasonographic characteristics. *BMC Medical Imaging, 23*. https://doi.org/10.1186/s12880-023-01117-z
25. Cao, C. L., Li, Q. L., Tong, J., Shi, L. N., Li, W. X., Xu, Y., et al. (2023). Artificial intelligence in thyroid ultrasound. *Frontiers in Oncology, 13*. https://doi.org/10.3389/fonc.2023.1060702
26. Weng, J., Wildman-Tobriner, B., Buda, M., Yang, J., Ho, L. M., Allen, B. C., et al. (n.d.). Deep learning for classification of thyroid nodules on ultrasound: Validation on an independent dataset.
27. Choi, Y. J., Baek, J. H., Park, H. S., Shim, W. H., Kim, T. Y., Shong, Y. K., et al. (2017). A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of thyroid nodules on ultrasound: Initial clinical assessment. *Thyroid, 27*, 546–552. https://doi.org/10.1089/thy.2016.0372
28. Zhao, W. J., Fu, L. R., Huang, Z. M., Zhu, J. Q., Ma, B. Y., & Tarantino, G. (2019). Effectiveness evaluation of computer-aided diagnosis system for the diagnosis of thyroid nodules on ultrasound: A systematic review and meta-analysis. *Medicine, 98*. https://doi.org/10.1097/MD.0000000000016379
29. Xia, M., Song, F., Zhao, Y., Xie, Y., Wen, Y., & Zhou, P. (2023). Ultrasonography-based radiomics and computer-aided diagnosis in thyroid nodule management: Performance comparison and clinical strategy optimization. *Frontiers in Endocrinology, 14*. https://doi.org/10.3389/fendo.2023.1140816
30. Lee, Y., Alam, M. R., Park, H., Yim, K., Seo, K. J., Hwang, G., et al. (2024). Improved diagnostic accuracy of thyroid fine-needle aspiration cytology with artificial intelligence technology. *Thyroid, 34*, 723–734. https://doi.org/10.1089/thy.2023.0384
31. Cece, A., Agresti, M., De Falco, N., Sperlongano, P., Moccia, G., Luongo, P., et al. (2025). Role of artificial intelligence in thyroid cancer diagnosis. *Journal of Clinical Medicine, 14*. https://doi.org/10.3390/jcm14072422
32. Jia, X., Ma, Z., Kong, D., Li, Y., Hu, H., Guan, L., et al. (2022). Novel human artificial intelligence hybrid framework pinpoints thyroid nodule malignancy and identifies overlooked second-order ultrasonographic features. *Cancers, 14*. https://doi.org/10.3390/cancers14184440
33. Pathak, A., Yu, Z., Paredes, D., Monsour, P., Rocha, A. O., & Brito, J. P., et al. (n.d.). Extracting thyroid nodules characteristics from ultrasound reports using transformer-based natural language processing methods.

34. Yu, Y., Ouyang, W., Huang, Y., Huang, H., Wang, Z., Jia, X., et al. (2024). AI-based multimodal multi-tasks analysis reveals tumor molecular heterogeneity, predicts preoperative lymph node metastasis and prognosis in papillary thyroid carcinoma: A retrospective study. *International Journal of Surgery*. https://doi.org/10.1097/js9.0000000000001875
35. Slabaugh, G., Beltran, L., Rizvi, H., Deloukas, P., & Marouli, E. (2023). Applications of machine and deep learning to thyroid cytology and histopathology: A review. *Frontiers in Oncology, 13*. https://doi.org/10.3389/fonc.2023.958310
36. Liang, Q., Qi, Z., & Li, Y. (2024). Machine learning to predict the occurrence of thyroid nodules: Towards a quantitative approach for judicious utilization of thyroid ultrasonography. *Frontiers in Endocrinology, 15*. https://doi.org/10.3389/fendo.2024.1385836
37. Peng, S., Liu, Y., Lv, W., Liu, L., Zhou, Q., Yang, H., et al. (2021). Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: A multicentre diagnostic study. *The Lancet Digital Health, 3*, e250–e259. https://doi.org/10.1016/S2589-7500(21)00041-8