

THE USE OF THE THYROID IMAGING REPORTING AND DATA SYSTEM (TIRADS) IN THE DIAGNOSIS OF THYROID NODULES

Ali Mohammed Asiri^{1*}, Mohammed Abdullah Asiri², Abdullah Ali Al-Asmari³, Abdulatif Saeed Alqahtani⁴, Abdulrahman Zaid Fatani⁵, Mohammad Saud Alahmadi⁶

¹Radiologist, Radiology Department, Armed Forces Hospital – Southern Region, Ministry of Defense, Saudi Arabia. Email: ali1.asiri@MODHS.MED.SA

²Radiographer, Radiology Department, Armed Forces Hospital – Southern Region, Ministry of Defense, Saudi Arabia. Email: moh.assiri@MODHS.MED.SA

³Radiographer, Radiology Department, Armed Forces Hospital – Southern Region, Ministry of Defense, Saudi Arabia. Email: ab1.alasmari@MODHS.MED.SA

⁴Radiographer, Radiology Department, Armed Forces Hospital – Southern Region, Ministry of Defense, Saudi Arabia, Email: abdulatef.alqahtani@modhs.med.sa

⁵Bachelor's degree in Biochemistry, College of Science; Master's degree in Technical Anatomy and Histology, College of Medicine, Email: aftani@taibahu.edu.sa

⁶Medical Secretary Technician in the Occupational Health and Safety, Department Medical Secretary Diploma, msalahmdy@taibahu.edu.sa

*Corresponding Author: Ali Mohammed Asiri, ali1.asiri@MODHS.MED.SA

ABSTRACT

Thyroid nodules represent one of the most common diagnoses in medicine, having been found to be present in as many as two-thirds of adults upon the application of high-resolution ultrasound scanning. Though the vast majority turn out to be benign, the lack of ability to distinguish the relatively small number of clinically meaningful malignancies from those of low significance in terms of treatment has led to large numbers of patients undergoing fine needle aspiration (FNA) biopsies, diagnostic operations, and an over-diagnoses of thyroid cancer pandemic across the world. This is the challenge which the TIRADS system was designed to overcome. This narrative review aims to explore the history, architecture, efficacy, and application of TIRADS schemes, with special attention given to the most widely used classification scheme, the American College of Radiology TIRADS (ACR TI-RADS) and alternative classifications, namely EU-TIRADS, K-TIRADS, and C-TIRADS. In this review, using the results of recent meta-analyses and validations across multiple institutions, the performance of these systems is evaluated in relation to their sensitivity, specificity, discriminative ability, and potential for avoiding unnecessary biopsies. Common problems such as inter-rater variability, uncertain diagnosis of follicular lesions, and management of indeterminate cytology are reviewed in detail, as well as the increasing use of machine learning in improving the accuracy and efficiency of feature extraction and risk stratification. There is significant evidence that TIRADS, and the ACR TI-RADS system in particular, has greatly increased the specificity of nodules assessment and avoided low-yield interventions, and that convergence between competing classification systems and long-term outcome assessments are essential needs.

KEYWORDS: *thyroid nodule , ultrasonography , TIRADS , risk stratification , fine-needle aspiration , thyroid cancer ,ACR TI-RADS ,artificial intelligence*

1. INTRODUCTION

The thyroid nodule is one of the most frequently encountered abnormalities in contemporary clinical practice. As neck ultrasonography has become ubiquitous, both as a dedicated thyroid examination and as an incidental consequence of imaging performed for unrelated indications, the rate at which nodules are detected has risen sharply. The central clinical challenge posed by these lesions is deceptively simple to state yet difficult to resolve: although nodules are extremely common, only a small minority harbour clinically significant malignancy, and the great majority will never threaten a patient's health. The diagnostic task is therefore not merely to detect nodules but to discriminate, as reliably and reproducibly as possible, the few that warrant intervention from the many that can be safely observed [1].

The basis of such discrimination was the impression of the examiner as a whole, based largely on the extensive use of fine-needle aspiration (FNA) cytology. Such an approach had two important drawbacks connected to each other. First, the individual impression created significant inter-observer differences in diagnoses and, consequently, provided inconsistency in diagnostics. Secondly, the low bar of biopsy resulted in the enormous number of biopsies performed most of which turned out to be unnecessary. In this way, overdiagnosis became a real phenomenon associated with the excessive use of this method, the development of which led to an epidemic of diagnosis of small papillary carcinomas which would never be diagnosed clinically. Overdiagnosis of this sort has been established as a serious public-health issue [2,3].

In this context, ultrasound-based risk stratification algorithms were developed. Drawing a parallel with the Breast Imaging Reporting and Data System (BI-RADS), which revolutionised the field of mammographic diagnosis,

researchers suggested a new Thyroid Imaging Reporting and Data System (TIRADS) in which combinations of sonographic characteristics result in a categorised assessment of malignancy risk. The benefits of using TIRADS include three main points: the creation of standard language for describing the nodules, assigning a clear numerical malignancy risk to each category, and transforming this risk into specific recommendations on whether or not to perform a biopsy. Such a system would be able to solve both parts of the diagnostic dilemma [4,5].

Indeed, during the last 15 years, several TIRADS variations have appeared, created by different professional societies or research groups; they differ from each other in terms of criteria, scoring, classification, and cut-off for biopsy recommendation. While such diversity of approaches is undoubtedly a result of active research interest in this topic, it has also generated confusion since one can get a different classification and a different approach to a particular thyroid nodule depending on which TIRADS system is used. The aim of this narrative literature review is to offer a holistic view on the topic. Specifically, the review will discuss the historical background and conceptual evolution of the TIRADS approach; analyse the structure of the most commonly used TIRADS systems; examine the evidence comparing diagnostic utility and clinical impact of TIRADS systems; consider current limitations of the approach that reduce its effectiveness; and finally, consider recent contributions of AI. As a narrative rather than systematic review, the present work draws selectively but critically on the most authoritative primary sources, society guidelines, large validation cohorts, and contemporary meta-analyses, with an emphasis on publications that have shaped clinical practice and on the most recent evidence available. The structure of the review reflects a deliberate logic. It begins with the epidemiological problem that TIRADS was created to solve, because the value of any risk-stratification tool can only be judged against the clinical burden it is meant to relieve. It then proceeds historically, since the contemporary systems are best understood as successive refinements of a shared conceptual ancestry rather than as independent inventions. The detailed treatment of ACR TI-RADS that follows reflects both its prominence in the literature and its didactic value as an exemplar of feature-based scoring, after which the competing systems are examined comparatively. While the preceding parts are devoted to description, subsequent parts take on evaluation, comparing systems with each other as well as to the inherent flaws that they may have, before proceeding to discuss how new technology such as artificial intelligence and integration with cytology will be impacting the field in the future. The common denominator throughout will be the issue of sensitivity versus specificity in balancing danger detection and over-diagnosis/over-treatment.

2. Thyroid Nodules and the Clinical Imperative for Risk Stratification

Epidemiology provides the framework within which the issue of risk stratification gains importance. While palpation can identify thyroid nodules in just a minority of individuals in the community, high-resolution ultrasound demonstrates nodules at much greater rates, anywhere between one fifth and two-thirds of unselected patients, depending upon the particular population sample and quality of technology employed. The prevalence of nodules increases with age, is more common among females, and also in geographic areas with iodine deficiency or radiation exposure. What is important about this scenario is that the extremely high background rate of occurrence is associated with relatively lower risk for cancer, where, in many studies, the rate of malignancy in the nodules encountered ranges from 7 to 15 percent, and a significant number of cases represent small papillary carcinoma that may have indolent course [1].

The imbalance between an abundance of nodules and the absence of lethal diseases is evident in certain epidemiological observations. Thyroid cancer incidence rates standardized for age have risen manifold within the past few decades in many countries, but thyroid cancer deaths have stayed relatively constant at low levels. This paradox is typical of overdiagnosis, which refers to cases of cancer tumors which conform to pathological criteria but will never pose a threat of symptoms or lethality if they are left undetected. Cancer registry analyses involving dozens of countries suggest that a significant proportion of this so-called epidemic is related to diagnostic screening and that over one million patients may have experienced overdiagnosis in a single decade [2,3]. Overdiagnosis is evident in the graph illustrated below showing the rise in incidence without a concomitant rise in mortality rates; the shaded part of the graph indicates overdiagnosis.

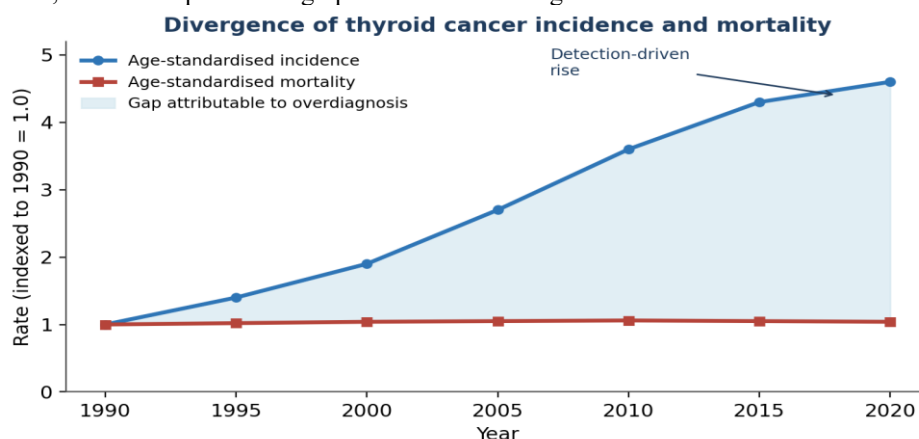


Figure 1. Schematic representation of the divergence between rising age-standardised thyroid cancer incidence and stable mortality, illustrating the burden attributable to overdiagnosis. Trends are indexed to 1990 and synthesised from the global epidemiological literature [2,3].

In the process of diagnosis, ultrasonography plays an important part. Ultrasonography is non-invasive, readily available, relatively cheap, and does not use ionising radiation, hence becoming an obvious choice when it comes to defining the characteristics of a nodule after it has been detected. The drawback with ultrasound lies in its being subjective; what makes a nodule either benign or malignant is defined based on subjective criteria, and there is significant variability both in obtaining the images and interpreting them. What seems to raise concerns to one radiologist may seem quite harmless to another, leading to completely different decisions on biopsies [4]. Fine needle aspiration cytology, the step to follow whenever a biopsy has been recommended, is itself imperfect since there are cases where the obtained sample does not provide a conclusive answer.

The very concept of risk stratification protocols was designed specifically to bring structure to such an inherently qualitative procedure. Defining which ultrasonic characteristics need to be evaluated, how they ought to be described, and how their integration allows estimating the likelihood of malignancy, TIRADS guidelines aim to standardize the process, making sure that biopsy indications match the underlying likelihood of disease occurrence and facilitate interprofessional communication. In other words, the fundamental problem to be solved is keeping the sensitivity to those thyroid cancers that really require intervention while simultaneously limiting the number of unnecessary procedures resulting from high nodule prevalence [4,5].

3. Conceptual Origins and Historical Evolution of TIRADS

The conceptual heritage of TIRADS extends back directly to the BI-RADS criteria, for which by the end of the 1990s there was already evidence that an organised system of categorical reporting could create uniform terminology in describing the imaging findings, relate the descriptive features to probability of malignancy and management guidelines. It was this achievement that served as inspiration for the very first proposed version of the TIRADS system, put forward by Horvath et al in 2009. This method employed the concept of an atlas comprising ten distinctive ultrasound patterns, each one associated with a specific likelihood of being malignant, based on empirical data; a lesion was assigned its category according to correspondence with the most suitable pattern. Although innovative, the approach was impractical, due to lack of conformity with the great diversity of thyroid nodular morphology in reality [6].

In 2011, Kwak et al. offered a more practical model wherein the number of suspicious sonographic characteristics within a nodule, including solid composition, marked hypoechogenicity, microlobulation, microcalcification, and a greater height than width, was tallied, and then a risk category determined accordingly. The system developed by Kwak and collaborators proved easier to apply at the point-of-care, and further showed that risk for malignancy increases monotonically with the number of suspicious characteristics—a concept that would go on to have staying power, informing several later models [7]. In developing such early systems, the important decisions were identified, decisions still influencing subsequent developments to this day—whether pattern-based classification or a tally/count/weighted sum of characteristics, and whether mapping that to risk of malignancy on a graded scale.

The period of 2016 and 2017 was characterized by the emergence of some standardized and society-recognized systems. The American Thyroid Association adopted the pattern-based sonographic risk stratification system and included it in their management guidelines where five levels of suspicion were established including those with the probability of malignancy and cut-offs for biopsy based on size were defined [1]. The Korean Society of Thyroid Radiology introduced the Korean version of TIRADS that combined the use of the pattern-based approach and took into consideration the composition and echogenicity along with the suspicion score [9]. A group of experts gathered by the European Thyroid Association developed the European TIRADS that consisted of five levels based on the previous experiences [8].

Publication of the American College of Radiology TI-RADS in 2017 under the chairmanship of Tessler was especially significant. Instead of using a holistic pattern matching process, ACR assigns scores for individual characteristics in five domains and then adds them up to derive a risk score, a method which brings the precision of a feature score while incorporating precise rules [5]. The most recent example is the Chinese Thyroid Imaging Reporting and Data System in 2020, which tries to unify the various classification systems used in China through a counting system based on large multicentre data [10]. This progression is summarized in Table 1 below, and the variety of methods employed here hints at future comparison problems discussed later in this paper.

Table 1: Chronological development and core design features of the principal thyroid ultrasound risk-stratification systems [1,5–10].

System	Year	Design principle	Category structure
Horvath TIRADS	2009	Pattern atlas	Ten patterns mapped to TIRADS 2–5
Kwak TIRADS	2011	Count of suspicious features	TIRADS 3, 4a, 4b, 4c, 5
ATA 2015	2016	Sonographic pattern	Benign, very low, low, intermediate, high
K-TIRADS	2016/2021	Pattern (composition + echogenicity)	K-TIRADS 1–5
EU-TIRADS	2017	Pattern with risk tiers	EU-TIRADS 1–5

System	Year	Design principle	Category structure
ACR TI-RADS	2017	Weighted points summed	TR1–TR5 (points-based)
C-TIRADS	2020	Counting model (+/- features)	C-TIRADS 1–5 (with 4a–4c)

4. The Anatomy of ACR TI-RADS

As it has been used more than any other TI-RADS, the ACR TI-RADS should be described in detail. The main feature of this system lies in its points-based structure. The radiologist assesses the lesion based on the five categories and awards the appropriate number of points for each category depending on the features displayed, and the sum of points will provide the final classification of the lesion. The five categories include composition, echogenicity, shape, margin, and echogenic foci. They have been chosen since they represent independent risk factors for the presence of malignancy [5].

Composition refers to the ratio of solid and cystic components of the nodule. Nodules that are predominantly cystic or purely cystic, along with spongiform nodules, score no points because they are considered reliably benign; on the other hand, nodules that are predominantly solid receive the most points in the category. Echogenicity refers to the comparative echo appearance of the nodule relative to its thyroid gland parenchyma and adjacent strap muscle; isoechoic or hyperechoic nodules score few points, hypoechoic nodules score medium points, and markedly hypoechoic nodules score the most points since they are highly related to cancerous nodules. The third feature is the shape, which refers to the tallest-to-widest aspect seen on transverse sonograms indicating growth through normal tissue planes, a known feature for malignant thyroid nodules. As for margin, scoring for smooth or vaguely defined margins would score a low number of points, while lobulated or irregular margins, or frankly extrathyroidal margins, score high points. Echogenic foci refers to differentiating benign artefacts and peripheral macrocalcifications from punctate echogenic foci that indicate microcalcifications [5,11].

The total points assigned in each category, tabulated in Table 2 below, are thereafter added up. One major advantage of this classification scheme is that since points are assigned on an additive basis, a nodule with multiple suspicious findings quickly escalates into a higher category even without the presence of a single specific suspicious finding. The total number of points correlates with one of the five risk categories, referred to as TR1–TR5, shown in Table 3 below. TR1 and TR2 stand for benign nodules and not-suspicious nodules respectively; TR3 is mildly suspicious; TR4 is moderately suspicious; and TR5, equivalent to 7 points or more, is highly suspicious. As evident from Table 3, the malignancy risk increases sharply at each category [5,12].

ACR TI-RADS point-based scoring

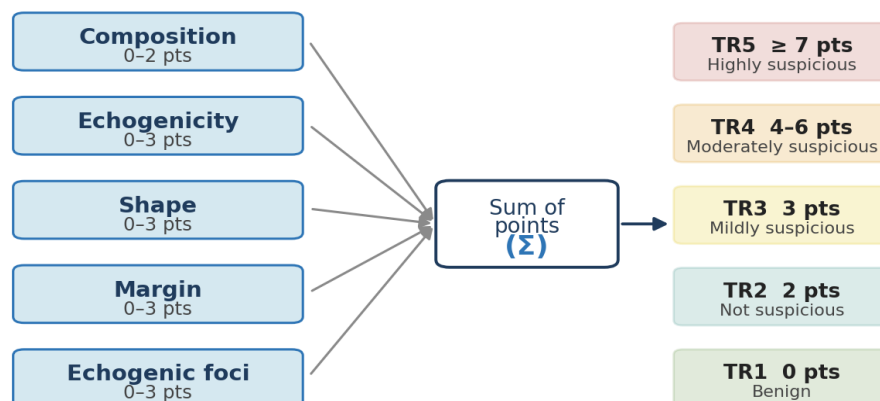


Figure 5. The ACR TI-RADS scoring logic. Points awarded across five feature categories are summed to yield a total that maps onto five risk levels (TR1–TR5) with progressively higher estimated malignancy risk [5].

Table 2: Point assignment for the five feature categories of ACR TI-RADS. Points are summed across all categories to determine the risk level [5,11].

Feature category	Descriptor	Points
Composition	Cystic / spongiform	0
Composition	Mixed cystic and solid	1
Composition	Solid or almost completely solid	2
Echogenicity	Anechoic / hyperechoic / isoechoic	0–1

Feature category	Descriptor	Points
Echogenicity	Hypoechoic	2
Echogenicity	Very hypoechoic	3
Shape	Wider-than-tall	0
Shape	Taller-than-wide	3
Margin	Smooth / ill-defined	0
Margin	Lobulated or irregular	2
Margin	Extrathyroidal extension	3
Echogenic foci	None / comet-tail artefact	0
Echogenic foci	Macrocalcifications	1
Echogenic foci	Peripheral (rim) calcifications	2
Echogenic foci	Punctate echogenic foci	3

It is not only the risk category by itself but the interaction between the risk category and the size of the nodule that determines the value of the clinical utility of the system. Thus, ACR TI-RADS gives specific size cut-offs above which biopsy becomes a recommendation, and other cut-offs above which sonographic follow-ups instead of biopsy become a recommendation, with the size cut-off values gradually becoming lower with increasing risk level. A TR5 nodule is recommended to be biopsied when its maximal diameter reaches 1 centimetre, but TR3 and TR4 nodules have higher size cut-offs, while TR1 and TR2 nodules do not need biopsy and even monitoring in the majority of cases. These specific recommendations are based on the logic that allows for concentrating attention on suspicious and relatively large nodules and still allow monitoring of smaller ones that cannot produce any medical value through biopsy. Follow-up is an important component of ACR TI-RADS because it guarantees the detection of the exceptional case of a significant cancer [11,12].

Table 3: ACR TI-RADS risk levels, point ranges, interpretation, and size-dependent fine-needle aspiration (FNA) thresholds [5,11].

Level	Total points	Interpretation	FNA threshold (max diameter)
TR1	0	Benign	No FNA
TR2	2	Not suspicious	No FNA
TR3	3	Mildly suspicious	FNA if ≥ 2.5 cm; follow if ≥ 1.5 cm
TR4	4–6	Moderately suspicious	FNA if ≥ 1.5 cm; follow if ≥ 1.0 cm
TR5	≥ 7	Highly suspicious	FNA if ≥ 1.0 cm; follow if ≥ 0.5 cm

This biology behind the weights assigned to each feature is crucial to understand since it not only highlights the reasons behind the advantages but also predicts the inherent limitations of this approach. Composition is considered of primary importance since cystic and predominantly cystic masses are invariably benign with their colloid and fluid content representing degeneration or hyperplasia and not neoplasia. Hypoechogenicity indicates a high cellular content with little or no colloid present in the mass as seen in malignant lesions, especially papillary carcinoma, which have a different acoustic response compared to normal follicular structures. Taller-than-wide aspect ratio denotes uncontrolled growth beyond the normal planes of the gland, a property more commonly seen in malignancy rather than benign expansile growths. Irregular and lobulated margins are indicative of an infiltrating nature of the lesion, while extra-thyroid extension is a direct sign of aggressive behaviour. The presence of punctate echogenicities is correlated with psammoma bodies, calcified concretions associated with papillary carcinoma; distinction between these entities and benign comet-tailed artefacts of inspissated colloid represents one of the most difficult decisions made by the classifier [5,11].

The points-based design is advantageous from a certain perspective when evaluating nodules that combine both favorable and unfavorable characteristics. Unlike systems that rely on choosing one characteristic above all others for classification purposes, the point-based approach allows adding all the abnormalities that exist in the nodule and obtaining an overall result. Thus, while the solid, hypoechoic nodule is assigned a medium number of points based on its characteristics, the addition of such features as irregular borders or microcalcifications automatically assigns it to a high-risk category. This method reflects the process of building suspicion regarding the benign or malignant nature of the nodule, taking into account multiple aspects at once. In contrast, the additive system is

vulnerable to follicular nodules, which usually do not have any of the factors leading to an increase in the points number.

5. The Landscape of Competing Risk-Stratification Systems

Even though ACR TI-RADS has been validated more extensively than any other classification, it still competes among numerous other classifications that differ based on the priorities of the region where the classification was developed. This is why an understanding of similarities and differences between existing classifications is necessary since it significantly influences how a nodule is classified. The most significant difference between existing systems consists of whether the system classifies nodules according to a pattern based on ultrasound or a scoring system based on characteristics that define the nodule. While the first option is generally faster, it may have problems when a nodule contains elements that do not fit a specific pattern; the latter is more accurate and reliable.

5.1 The European EU-TIRADS

EU-TIRADS developed by the European Thyroid Association is based on a pattern approach and comprises five categories. EU-TIRADS 1 is used when there are no nodules, EU-TIRADS 2 indicates benign findings including pure cystic/spongiform lesions with very low risk of malignancy, whereas EU-TIRADS 3, 4, and 5 indicate low-, intermediate-, and high-risk, respectively. High-risk nodules are defined as having any one of four features strongly indicative of malignancy including marked hypoechogenicity, irregular margin, non-ovoid shape, and microcalcification. Importantly, this classification system provides an estimate of malignancy risk for each category and biopsy criteria according to nodule size and the objective of increasing uniformity of reporting in the culturally varied European setting [8]. Sensitivity rates for EU-TIRADS have been found to be in the range of 80%–85% while specificity rates have been reported to vary widely across populations [4].

5.2 The Korean K-TIRADS

The latest version of K-TIRADS was revised in 2021. It uses criteria based on nodule composition and echogenicity, along with the presence of suspicious features. Any solid hypoechoic nodule with suspicious features will fall into the category of high suspicion (K-TIRADS 5), and combinations of increasingly less suspicious features result in K-TIRADS 4, 3, and 2 categories, representing intermediate, low suspicion, and benign nodules, respectively. One key difference between K-TIRADS and ACR TIRADS has been its tendency to biopsy more nodules. The 2021 revision attempted to limit biopsies of benign lesions while maintaining sensitivity towards small cancers. Size thresholds were also revised [9]. Meta-analysis showed that K-TIRADS is highly sensitive at the biopsy threshold, though not specific when compared to the ACR criteria [13,14].

5.3 The Chinese C-TIRADS

The C-TIRADS, established in 2020, is meant to establish a uniform approach among many scattered classification models existing in China before. In this classification method, a scoring approach is adopted where malignant-appearing features (i.e., vertical shape, solidity, marked hypoechogenicity, irregular margin, and microcalcifications) score positively while a benign characteristic is scored negatively in the count, which corresponds to C-TIRADS 2 to 5, with further subdivision of the fourth grade into 4a, 4b, and 4c. This new model has been calibrated using data from multiple centres collected through a national ultrasound group, and its performance has been found to be promising in several studies that systematically reviewed its performance as compared with the other classifications [10,15].

5.4 The American Thyroid Association classification

The ATA ultrasound-based classification scheme, which is part of the association's larger management guidelines, is classified based on pattern and is still extensively used, especially in the field of endocrinology. Due to the sensitivity of the ATA classification, it has been observed to be quite more sensitive but not very specific as compared to the ACR TI-RADS system and hence recommends more biopsies [1,16]. The table below summarizes the major classifications along important criteria that will be used in our comparative analysis.

*Table 4. Comparative overview of major TIRADS systems. *Performance ranges are drawn from comparative studies and meta-analyses and vary with the chosen category threshold and the case mix of the underlying cohort [4,13–16].*

System	Design	Categories	Reported sensitivity*	Reported specificity*
ACR TI-RADS	Points-based	TR1–TR5	~63–79%	~72–90%
EU-TIRADS	Pattern	1–5	~83–86%	~32–79%
K-TIRADS	Pattern	1–5	~88–93%	~57–63%
ATA 2015	Pattern	5 tiers	~70–93%	~85–87%
Kwak TI-RADS	Feature count	3–5	~90–96%	~60–64%
C-TIRADS	Counting model	2–5	~68–82%	~71–81%

6. Diagnostic Performance: Evidence from Meta-Analyses

With more and more validation studies becoming available, meta-analytic comparisons of the systems have become possible. The interpretation of these data, however, calls for consideration of a number of subtle methodological issues. In any test system, sensitivity and specificity crucially depend on the definition of positivity in terms of the cut-off value; specifically, a decrease in the cut-off increases the former, while reducing the latter. System performance depends not only on the prevalence and distribution of pathology but also on the proportion of follicular-patterned nodules, as well as on the reference standard used, either cytologic or histologic. For this reason, it is preferable to base the assessment of a system's discriminative power on an integrated measure, such as AUC or the diagnostic odds ratio, rather than the individual components, like sensitivity or specificity [13,14].

The results of the meta-analysis which comprised sixty-seven studies with a total number of seventy-six thousand plus nodules demonstrated a high performance of five different systems for the discrimination between benign and malignant lesions along with establishing the optimum cut-off points. The study showed a very high level of AUC for all of the systems, with maximum values of 0.90, 0.88, 0.86, and 0.84 being reached for the Korean, ACR, ATA, and EU-TIRADS respectively, whereas the Kwak system demonstrated an even higher level, reaching around 0.93. In addition to the above, the same study also estimated the characteristic dependence between sensitivity and specificity that appeared at each system's biopsy cut-off point. Thus, for the Korean and Kwak systems, a very high level of sensitivity was reached at the cut-off point with low specificity; on the contrary, the ACR system demonstrated high specificity [13].

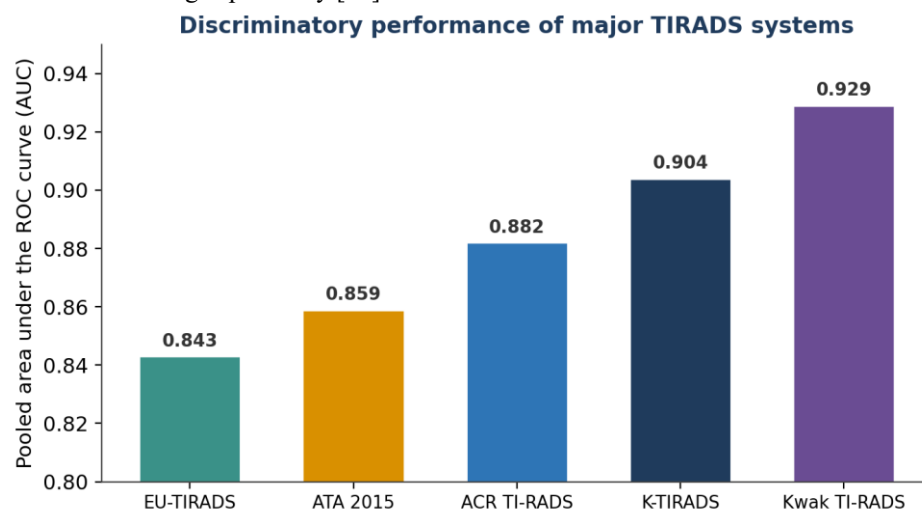


Figure 2. Pooled area under the ROC curve (AUC) for the major risk-stratification systems, reflecting overall discriminatory accuracy. Values represent the highest reported summary estimates from a large multi-system meta-analysis [13].

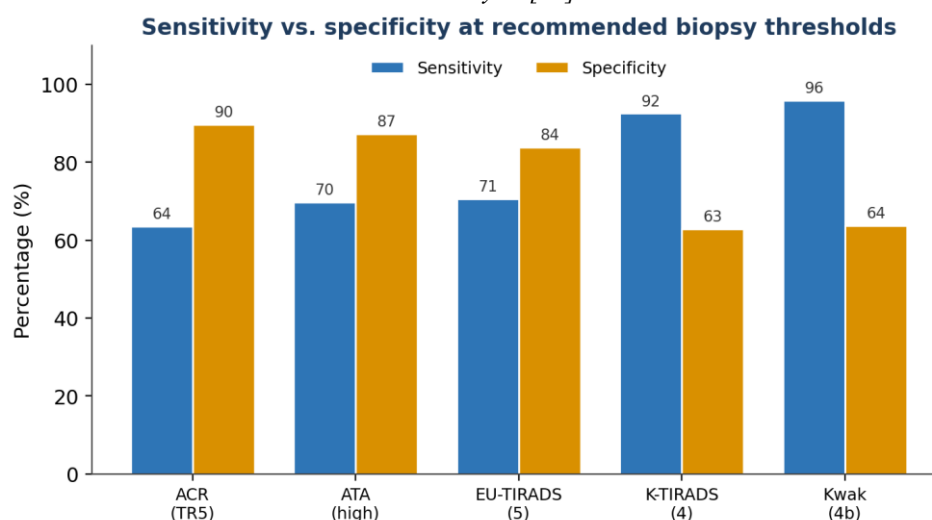


Figure 3. Sensitivity and specificity of major systems at commonly recommended biopsy thresholds, illustrating the characteristic inverse relationship between the two measures across systems [13].

In addition to the above, another network meta-analysis that assessed six different risk stratification systems using direct comparison between any two systems within the same population made it evident that among the six systems evaluated, the one that had the best overall diagnostic accuracy for diagnosing thyroid cancer via ultrasound was the ACR TI-RADS system [16]. It was noted that this information could assist in the decision-making process and

potentially influence modifications of the system, as well as the editorial of this research, which supported the conclusion that evidence suggests the superiority of ACR TI-RADS to other competing models for risk stratification [16]. This finding is congruent with the fact that the main benefit of the ACR TI-RADS system is related to its better specificity, which is crucial for avoiding unnecessary biopsies [12].

There are several studies comparing the performance of the two systems ACR and Kwak, showing excellent results of both and determining the best cut-off points for each; for the ACR system, it was found that the fourth risk level serves as a reasonable cut-off point, but that the exact level may be modified depending on which aspect, sensitivity or specificity, is prioritised [14]. There have also been some investigations of the more recently developed systems as well as computer-assisted approaches to thyroid nodules classification, with one study concluding that among the five most recently suggested diagnostic systems, including C-TIRADS from China, the latter provides the best trade-off between diagnostic efficacy and accuracy [15]. The common theme throughout all of these studies is not that one of the systems is better overall, but that the different systems have distinct characteristics, and the distinguishing feature of the ACR system lies primarily in high specificity and a reduced rate of unnecessary biopsies.

It is important to discuss how two seemingly similar methods can produce different outcomes, since it helps us understand how we are to interpret the results obtained from the studies in question. The crucial factor that plays an essential role here is the category that acts as the cut-off point for a positive result. A test performed using a low cut-off point will demonstrate extremely high sensitivity but relatively low specificity while a test performed with a high cut-off point will have the opposite effect. Due to the difference in category definition and probability of malignancy implied by the categories used by the two methods, the cut-offs must match in order to conduct fair analysis [16]. The second factor is the make-up of the test group used. Populations selected from surgical cases have high malignant incidence rates and challenging nodule patterns leading to skewed statistics, but a population taken from consecutive outpatient populations has a lower malignant prevalence rate that more accurately mirrors practice. The third factor is the incidence of lesions with follicular characteristics, which are poorly characterized by all of the systems so that a group with many follicular lesions would reduce the calculated sensitivity of any system. An understanding of these parameters protects the user from an overly simplistic interpretation of sensitivity and specificity and instead focuses on clinically relevant measures of threshold integration [13,14,16].

7. Clinical Impact: Reducing Unnecessary Biopsies

However, in terms of clinical relevance, the final rationale for a risk-stratification protocol does not lie in its mathematical sophistication, but in its practical impact on patient management, and in this case, the most compelling data are the numbers of avoided fine needle aspirations. One notable study that employed retrospective analysis of the biopsy criteria of various systems on a set of nodules that had been diagnosed determined that such systems could help avoid from seventeen to fifty-three percent of performed biopsies. The highest percentage was obtained by the ACR TI-RADS, as it categorised more than half of performed biopsies as unnecessary with extremely low rates of false negatives and high negative predictive value. As the goal of research was viewed as the creation of the 'right' TIRADS, avoiding unnecessary interventions at the expense of missing important findings, the ACR TI-RADS came close enough to that target [12].

This benefit has been validated by later studies. In an updated report about the ACR classification system, it was found out that compared to other systems, it lowers the incidence of unnecessary aspiration of non-malignant nodules by around twenty to forty-seven percent and, more importantly, the possibility of missed cancer cases is lessened by the recommendations for follow up for cases where the criterion for immediate biopsy is not fulfilled [11]. Practical implementations of these methods confirmed the advantages observed retrospectively; pre and post comparison of practices using ACR TI-RADS showed significant declines in the rate of referral for biopsies and increased monitoring of low-risk cases [17]. This is shown in figure 4 below.

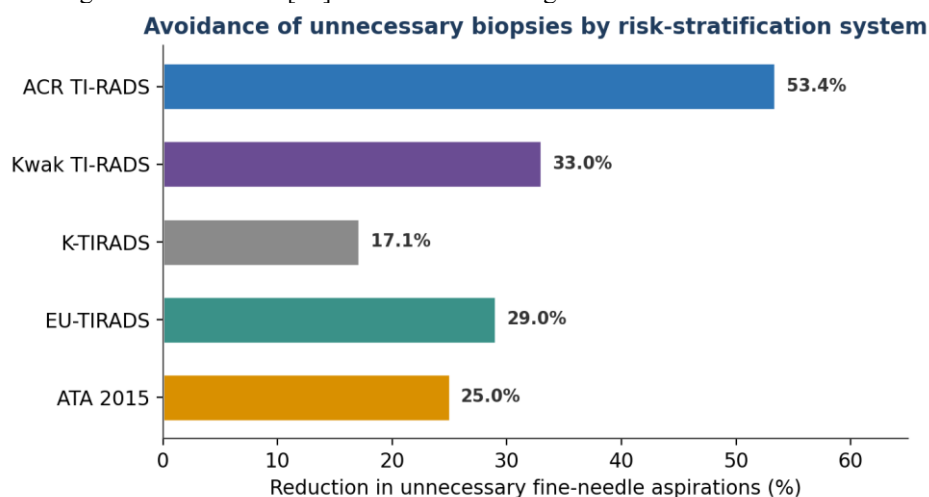


Figure 4. Comparative reduction in unnecessary fine-needle aspirations achieved by the major systems when their biopsy criteria are applied. ACR TI-RADS consistently yields the largest reduction [11,12].

The implications of these changes are significant. First of all, each skipped biopsy means that the patient does not have to go through an unnecessary procedure and all its stress and, in addition, avoids the risk of having his/her nodule being classified as non-diagnostic or indeterminate, resulting in more tests and examinations and, finally, in the possibility that the diagnostic cascade ends in surgery aimed at eliminating a benign nodule. From the perspective of public health, proper selection of nodules for biopsy is one of the rare means of battling overdiagnosis outlined above, since this approach prevents the incidental nodule detected during the sonographic screening from being transformed into cancer [2,11]. At the same time, it is necessary to admit that the effectiveness of the method depends greatly on its consistent application, including strict compliance with follow-up instructions, as well as on the type of nodules to be examined.

8. Limitations and Diagnostic Pitfalls

Despite their proven advantages, there are a number of inherent weaknesses of the TIRADS approaches which limit the reliability and repeatability of the assessments. First of all, this is the issue of interobserver variability. Even though a structured approach has been developed in order to avoid this problem, the subjective assessment of each sonographic feature still cannot be avoided. The identification of the degree of hypoechogenicity (mild vs marked), the description of margins (smooth vs subtle lobulation), and even the detection of punctate microcalcifications as opposed to artifacts depend upon both the quality of the image and the skills of the observer. Another relevant issue is the evaluation of follicular patterned tumors. The criteria used for TIRADS classification were established based on the characteristics of the papillary thyroid carcinoma, such as its hypoechoic nature, irregular borders, and microcalcifications. In contrast, follicular carcinomas and the follicular variant of papillary carcinoma often have none of these characteristics but instead appear as relatively featureless, iso- or hyperechoic masses with well-defined margins, resulting in low scores within the TIRADS system. For this reason, it has been noted that follicular malignancy is often under-detected with the current TIRADS score systems and therefore a low TIRADS score should not be used as assurance of a benign lesion when the thyroid fine needle aspiration results in a follicular pattern [18].

Additional limitations occur at the margins of validity of each system. Most TIRADS systems have been validated on patients above a certain age with lesions larger than one cm, and their performance on lesions smaller than this size, on children and adolescents, and in the context of diffuse thyroid diseases like Hashimoto thyroiditis remains unclear. The quantification of some variables can be affected by technical aspects like the plane on which the taller-than-wide sign was examined. Moreover, the very diversity of scoring schemes represents a limitation in and of itself, as the same nodule can fall into different risk categories and receive diverging management advice in different scoring systems, defeating the very aim of achieving consistency [9,18]. The main limitations and their real-world consequences are summarised in Table 5 below.

Table 5. Principal limitations of TIRADS systems, their underlying mechanisms, and clinical implications [4,9,18].

Limitation	Mechanism	Practical implication
Interobserver variability	Subjective feature assessment	Discordant categories and biopsy decisions
Follicular lesions	Lack papillary sonographic hallmarks	Malignancy under-detected at low categories
Subcentimetre nodules	Limited validation below 1 cm	Uncertain thresholds for small nodules
Special populations	Paediatric / diffuse disease under-studied	Reduced generalisability
System multiplicity	Divergent definitions and thresholds	Same nodule classified differently

It is recognition of these issues that has led to renewed research interest in ancillary modalities that may provide additional information to enhance grey-scale interpretation. One promising area of investigation is ultrasound elastography, which assesses tissue stiffness on the rationale that malignancy often correlates with increased hardness of nodules compared to benign lesions, with some studies suggesting improvement in specificity in particular when used with TIRADS to identify intermediate-risk lesions. The technique of contrast-enhanced ultrasound, which measures microvascular perfusion of the lesion, has also shown promise as an auxiliary modality for more accurate risk stratification, with the development of specific contrast-enhanced reporting systems being considered. Neither of these modalities has as yet found its way into the standard TIRADS scheme as a necessity due to technical and expertise limitations as well as unclear incremental advantage over meticulous grey-scale scoring. Nevertheless, these are promising examples of a potentially fruitful approach to improving structured risk stratification through objective quantifiable measurements rather than subjective assessment of features [4,18].

Another aspect worth taking into account is the implications of risk stratification on costs and organization. On the one hand, risk stratification helps save money due to less need to conduct biopsies and to place low-risk nodules into surveillance, with no unnecessary operations and no repeating tests and operations for diagnostic

reasons and savings related to reduced treatments of overdiagnosed malignancies. On the other hand, implementing systematic reporting means allocating considerable resources in terms of physicians' work on feature detection and follow-ups. Generally speaking, the benefits outweigh the costs involved, but it happens only because the program is implemented effectively and systematically, since neglecting surveillance can mean trading risk for risk. Indeed, a system that is able to decrease biopsies but unable to perform proper follow-ups will not be effective enough. It should be mentioned that these organizational factors show that the TIRADS is more than just a diagnostic tool; it is a complex system aimed at reducing cancer-related risks [11,12,17].

9. Integration with Cytology and the Diagnostic Pathway

The operation of the TIRADS method is not independent but occurs in tandem with the fine-needle aspiration cytology, where the results are presented using The Bethesda System for Reporting Thyroid Cytopathology. The interdependence of both systems is mutual. The TIRADS protocol helps to determine those nodules that will go on to undergo aspiration, after which, based on the results, the Bethesda system categorizes the aspirate according to the degree of malignancy. There are six groups into which the aspirate is divided by Bethesda; each group entails certain risks of malignancy, namely, non-diagnostic, benign, atypia of undetermined significance, follicular neoplasm, suspicious for malignancy, and finally, malignant. The third revision of the Bethesda system was published in 2023 [19].

The relationship between sonographic and cytological risk is especially important for the interpretation of indeterminate cytology, which remains the Achilles' heel of thyroid diagnostics. In the case of an indeterminate aspirate, the pre-test sonographic risk represented by the TIRADS score could contribute to refining the risk of malignancy and deciding on further actions – whether to re-aspirate, perform molecular analyses, undergo diagnostic surgery or observe. An encouraging TIRADS category of nodules with indeterminate cytology would reduce the post-test risk of malignancy, while a very suspicious sonographic appearance would increase the post-test risk; several papers have evaluated the ability of risk stratification systems in the detection of cancer in this population. The combination of standardized imaging, standardized cytology, and more recently, molecular tests, constitutes the cutting-edge method for evaluating thyroid nodules and a paradigmatic example of the trend towards a more integrated probabilistic approach to diagnosis [19].

10. Artificial Intelligence and the Future of Risk Stratification

The limitations inherent in the use of TIRADS by humans, particularly interobserver variability and the time needed for careful assessment of the characteristics, make the classification system a ripe candidate for artificial intelligence (AI). Two approaches can be identified. In the first case, machine learning is used to optimize the algorithm, through modification of the scores attributed to certain characteristics, in an effort to enhance specificity while not compromising sensitivity. An important research finding involved the application of AI to a large group of biopsy-verified thyroid lesions, and the derivation of an improved AI-enhanced ACR TI-RADS classification that achieved the same high level of diagnostic accuracy while cutting down on unnecessary biopsies [20].

The second method is based on deep learning with the application of convolutional neural networks, and it focuses on analyzing ultrasonography images directly. These machine learning algorithms can be applied either to reproduce the classifications of TIRADS features and hence standardize the identification of nodule features like composition, echogenicity, shape, margin, and echogenic foci or directly for predicting the likelihood of malignancy. A deep learning algorithm that was specifically tailored toward classifying ACR TI-RADS features performed better in differentiating between malignant and benign lesions than junior radiologists, and its accuracy was comparable to that of expert readers, interestingly, the algorithm trained on TIRADS feature categories performed even better compared to one trained on the actual histological outcome, indicating an added benefit by the use of the structured intermediate representation [21]. Other deep learning frameworks have been successful in enhancing the ability of moderate and highly suspicious nodules' discrimination in terms of their being malignant or not and help inexperienced readers in performing similarly to expert ones [22,23].

These changes indicate that AI will serve more as a complement rather than a substitute for radiologists, in terms of standardisation of features' assessment, speed-up of reporting, and a replicable second opinion. Some validation studies of AI-based risk-stratification systems have shown their effectiveness in independent cohorts, and the use of large language models able to make TIRADS stratifications based on free text reports seems to be the next step towards full automation [24]. Nevertheless, there are some caveats to take into consideration. For example, models that have been trained using data from specific hospitals do not always generalise well, and deep learning interpretability is still an issue. Furthermore, prospective evidence about the improvement in clinical outcomes due to AI-based risk stratification is only just beginning to appear. Nonetheless, it is evident where things are headed and artificial intelligence will continue to grow in importance in the evaluation of thyroid nodules [22,24].

11. Future Directions

A number of priorities have emerged from this review. First, there is the priority of standardization. As several systems exist using different criteria and cutoffs, it is difficult to achieve the aim of standardization that inspired the development of TIRADS in the first place, and a compelling argument can be made for unification at an international level in terms of terminology, and even scoring system, where regional biopsy cut-offs are still adjustable. Mapping of these systems to one another and recognizing the elements common to all has been a step towards this goal [4,16].

Secondly, research on generating prospective outcome data needs to be considered. The available evidence is mostly retrospective and based on cytological or histopathological proof of malignancy. These are problematic endpoints because both clinically important and clinically irrelevant cancers are counted. The development of systems requires solid evidence on clinically important outcomes, which are the avoidance of unnecessary surgical treatment and the detection of cancer that otherwise would not have been detected. Third, there is a need to extend and validate TIRADS in patient groups where no evidence currently exists, including children, sub-centimeter nodules, and diffuse thyroid diseases, along with the use of auxiliary methods, such as elastography and contrast-enhanced ultrasonography when it is useful. Lastly, the correct application of artificial intelligence technology and its thorough external validation, in particular with respect to generalizability and interpretability, is key in this next step in the evolution of systems [11,22,24].

12. CONCLUSION

The Thyroid Imaging Reporting and Data System can be considered as one of the most promising initiatives in implementing an evidence-based standardisation approach within an area which has been characterised by subjective assessment for years. Thanks to this system's ability to correlate the sonographic image of the nodule with its probability of being cancerous, as well as providing clear guidance on how to handle the nodule, TIRADS has succeeded in increasing the consistency of nodule assessment and improving nodule selection for biopsy, as well as allowing for effective control of the number of low-yield interventions resulting from the high prevalence of benign nodules. The results obtained in a number of meta-analyses show that while some systems demonstrate high discriminatory abilities, ACR TI-RADS outshines the others in terms of specificity, and in its ability to decrease unnecessary biopsies to a considerable extent without increasing the rate of false negatives [12,13,16]. However, these gains have been limited. Variability in interobserver agreement still exists, follicular lesions remain difficult to detect reliably, validation is incomplete in certain populations, and the existence of multiple systems works against the goal of standardisation, for which TIRADS was created. The linking of the TIRADS system with the Bethesda system for cytology, as well as new innovations in molecular diagnosis, has improved the diagnostic process, and there is potential for artificial intelligence to remove much of the subjective nature of lesion scoring. However, the lasting legacy of TIRADS is less one of technique than of concept, as TIRADS shifted the approach to the diagnosis of thyroid nodules from one that seeks to establish tissue diagnoses to a risk-adjusted probabilistic decision process. That concept will survive whatever tools emerge.

REFERENCES

- [1] Haugen, B. R., Alexander, E. K., Bible, K. C., Doherty, G. M., Mandel, S. J., Nikiforov, Y. E., Pacini, F., Randolph, G. W., Sawka, A. M., Schlumberger, M., Schuff, K. G., Sherman, S. I., Sosa, J. A., Steward, D. L., Tuttle, R. M., & Wartofsky, L. (2016). 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*, 26(1), 1–133. <https://doi.org/10.1089/thy.2015.0020>
- [2] Vaccarella, S., Franceschi, S., Bray, F., Wild, C. P., Plummer, M., & Dal Maso, L. (2016). Worldwide thyroid-cancer epidemic? The increasing impact of overdiagnosis. *New England Journal of Medicine*, 375(7), 614–617. <https://doi.org/10.1056/NEJMp1604412>
- [3] Lin, Y., Wu, Y., Hu, Z., Chen, L., Lin, Z., & Lin, Y. (2023). Trends in incidence and overdiagnosis of thyroid cancer in China, Japan, and South Korea. *Cancer Science*, 114(10), 4052–4062. <https://doi.org/10.1111/cas.15909>
- [4] Floridi, C., Cellina, M., Buccimazza, G., Arrichiello, A., Sacchini, A., Arrigoni, F., Pompili, G., Barile, A., & Carrafiello, G. (2019). Ultrasound imaging classifications of thyroid nodules for malignancy risk stratification and clinical management: State of the art. *Gland Surgery*, 8(Suppl 3), S233–S244. <https://doi.org/10.21037/gs.2019.07.01>
- [5] Tessler, F. N., Middleton, W. D., Grant, E. G., Hoang, J. K., Berland, L. L., Teefey, S. A., Cronan, J. J., Beland, M. D., Desser, T. S., Frates, M. C., Hammers, L. W., Hamper, U. M., Langer, J. E., Reading, C. C., Scoutt, L. M., & Stavros, A. T. (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(5), 587–595. <https://doi.org/10.1016/j.jacr.2017.01.046>
- [6] Horvath, E., Majlis, S., Rossi, R., Franco, C., Niedmann, J. P., Castro, A., & Dominguez, M. (2009). An ultrasonogram reporting system for thyroid nodules stratifying cancer risk for clinical management. *Journal of Clinical Endocrinology & Metabolism*, 94(5), 1748–1751. <https://doi.org/10.1210/jc.2008-1724>
- [7] Kwak, J. Y., Han, K. H., Yoon, J. H., Moon, H. J., Son, E. J., Park, S. H., Jung, H. K., Choi, J. S., Kim, B. M., & Kim, E. K. (2011). Thyroid imaging reporting and data system for US features of nodules: A step in establishing better stratification of cancer risk. *Radiology*, 260(3), 892–899. <https://doi.org/10.1148/radiol.11110206>
- [8] Russ, G., Bonnema, S. J., Erdogan, M. F., Durante, C., Ngu, R., & Leenhardt, L. (2017). European Thyroid Association guidelines for ultrasound malignancy risk stratification of thyroid nodules in adults: The EU-TIRADS. *European Thyroid Journal*, 6(5), 225–237. <https://doi.org/10.1159/000478927>
- [9] Ha, E. J., Chung, S. R., Na, D. G., Ahn, H. S., Chung, J., Lee, J. Y., Park, J. S., Yoo, R. E., Baek, J. H., Baek, S. M., Cho, S. W., Choi, Y. J., Hahn, S. Y., Jung, S. L., Kim, J. H., Kim, S. K., Kim, S. J., Lee, C. Y., Lee, H. K., ... Choi, M. (2021). 2021 Korean Thyroid Imaging Reporting and Data System and imaging-based management

- of thyroid nodules: Korean Society of Thyroid Radiology consensus statement and recommendations. *Korean Journal of Radiology*, 22(12), 2094–2123. <https://doi.org/10.3348/kjr.2021.0713>
- [10] Zhou, J., Yin, L., Wei, X., Zhang, S., Song, Y., Luo, B., Li, J., Qian, L., Cui, L., Chen, W., Wen, C., Peng, Y., Chen, Q., Lu, M., Chen, M., Wu, R., Zhou, W., Xue, E., Li, Y., ... Zhan, W. (2020). 2020 Chinese guidelines for ultrasound malignancy risk stratification of thyroid nodules: The C-TIRADS. *Endocrine*, 70(2), 256–279. <https://doi.org/10.1007/s12020-020-02441-y>
- [11] Hoang, J. K., Middleton, W. D., & Tessler, F. N. (2021). Update on ACR TI-RADS: Successes, challenges, and future directions, from the AJR special series on radiology reporting and data systems. *American Journal of Roentgenology*, 216(3), 570–578. <https://doi.org/10.2214/AJR.20.24608>
- [12] Grani, G., Lamartina, L., Ascoli, V., Bosco, D., Biffoni, M., Giacomelli, L., Maranghi, M., Falcone, R., Ramundo, V., Cantisani, V., Filetti, S., & Durante, C. (2019). Reducing the number of unnecessary thyroid biopsies while improving diagnostic accuracy: Toward the “right” TIRADS. *Journal of Clinical Endocrinology & Metabolism*, 104(1), 95–102. <https://doi.org/10.1210/je.2018-01674>
- [13] Castellana, M., Castellana, C., Treglia, G., Giorgino, F., Giovannella, L., Russ, G., & Trimboli, P. (2020). Performance of five ultrasound risk stratification systems in selecting thyroid nodules for FNA: A meta-analysis. *Journal of Clinical Endocrinology & Metabolism*, 105(5), 1659–1669. <https://doi.org/10.1210/clinem/dgz170>
- [14] Kang, Y. J., Stybayeva, G., Lee, J. E., & Hwang, S. H. (2022). Diagnostic performance of ACR and Kwak TI-RADS for benign and malignant thyroid nodules: An update systematic review and meta-analysis. *Cancers*, 14(23), 5961. <https://doi.org/10.3390/cancers14235961>
- [15] Li, W., Sun, X., Liu, B., Zhang, J., Zhang, Z., Wang, Y., & Du, P. (2023). Diagnostic efficiency among Eu-/C-/ACR-TIRADS and S-Detect for thyroid nodules: A systematic review and network meta-analysis. *Frontiers in Endocrinology*, 14, 1227339. <https://doi.org/10.3389/fendo.2023.1227339>
- [16] Kim, D. H., Kim, S. W., Basurrah, M. A., Lee, J., & Hwang, S. H. (2023). Diagnostic performance of six ultrasound risk stratification systems for thyroid nodules: A systematic review and network meta-analysis. *American Journal of Roentgenology*, 220(6), 791–803. <https://doi.org/10.2214/AJR.22.28556>
- [17] Watkins, L., O'Neill, G., Young, D., & McArthur, C. (2021). Comparison of British Thyroid Association, American College of Radiology TIRADS and artificial intelligence TIRADS with histological correlation: Diagnostic performance for predicting thyroid malignancy and unnecessary fine needle aspiration rate. *British Journal of Radiology*, 94(1123), 20201444. <https://doi.org/10.1259/bjr.20201444>
- [18] Grani, G., Lamartina, L., Cantisani, V., Maranghi, M., Lucia, P., & Durante, C. (2018). Interobserver agreement of various thyroid imaging reporting and data systems. *Endocrine Connections*, 7(1), 1–7. <https://doi.org/10.1530/EC-17-0336>
- [19] 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*, 33(9), 1039–1044. <https://doi.org/10.1089/thy.2023.0141>
- [20] Wildman-Tobriner, B., Buda, M., Hoang, J. K., Middleton, W. D., Thayer, D., Short, R. G., Tessler, F. N., & Mazurowski, M. A. (2019). Using artificial intelligence to revise ACR TI-RADS risk stratification of thyroid nodules: Diagnostic accuracy and utility. *Radiology*, 292(1), 112–119. <https://doi.org/10.1148/radiol.2019182128>
- [21] Chen, Y., Gao, Z., He, Y., Mai, W., Li, J., Zhou, M., Li, S., Yi, W., Wu, S., Bai, T., Zhang, N., Zeng, W., Lu, Y., & Liu, H. (2022). An artificial intelligence model based on ACR TI-RADS characteristics for US diagnosis of thyroid nodules. *Radiology*, 303(3), 613–619. <https://doi.org/10.1148/radiol.211455>
- [22] Wu, G. G., Lv, W. Z., Yin, R., Xu, J. W., Yan, Y. J., Chen, R. X., Wang, J. Y., Zhang, B., Cui, X. W., & Dietrich, C. F. (2021). Deep learning based on ACR TI-RADS can improve the differential diagnosis of thyroid nodules. *Frontiers in Oncology*, 11, 575166. <https://doi.org/10.3389/fonc.2021.575166>
- [23] Yang, J., Page, L. C., Wagner, L., Wildman-Tobriner, B., Bisset, L., Frush, D., & Mazurowski, M. A. (2023). Thyroid nodules on ultrasound in children and young adults: Comparison of diagnostic performance of radiologists' impressions, ACR TI-RADS, and a deep learning algorithm. *American Journal of Roentgenology*, 220(3), 408–417. <https://doi.org/10.2214/AJR.22.28231>
- [24] Wildman-Tobriner, B., Yang, J., Allen, B. C., Ho, L. M., Miller, C. M., & Mazurowski, M. A. (2024). Simplifying risk stratification for thyroid nodules on ultrasound: Validation and performance of an artificial intelligence thyroid imaging reporting and data system. *Current Problems in Diagnostic Radiology*, 53(6), 695–699. <https://doi.org/10.1067/j.cpradiol.2024.07.006>
- [25] Pizzato, M., Li, M., Vignat, J., Laversanne, M., Singh, D., La Vecchia, C., & Vaccarella, S. (2022). The epidemiological landscape of thyroid cancer worldwide: GLOBOCAN estimates for incidence and mortality rates in 2020. *The Lancet Diabetes & Endocrinology*, 10(4), 264–272. [https://doi.org/10.1016/S2213-8587\(22\)00035-3](https://doi.org/10.1016/S2213-8587(22)00035-3)
- [26] Huang, J., Ngai, C. H., Deng, Y., Pun, C. N., Lok, V., Zhang, L., Yuan, J., Xu, W., Zheng, Z. J., & Wong, M. C. S. (2023). Incidence and mortality of thyroid cancer in 50 countries: A jointpoint regression analysis of global trends. *Endocrine*, 80(2), 355–365. <https://doi.org/10.1007/s12020-022-03274-7>
- [27] Lu, K., Wang, L., Lai, S., Chen, Z., Zhu, Q., Cong, S., Gan, K., Chen, X., Huang, C., & Kuang, J. (2025). Optimizing the screening process for TIRADS could reduce the number of unnecessary thyroid biopsies. *Endocrine Connections*, 14(5), e250097. <https://doi.org/10.1530/EC-25-0097>

- [28] Na, D. G., Baek, J. H., Sung, J. Y., Kim, J. H., Kim, J. K., Choi, Y. J., & Seo, H. (2016). Thyroid imaging reporting and data system risk stratification of thyroid nodules: Categorization based on solidity and echogenicity. *Thyroid*, 26(4), 562–572. <https://doi.org/10.1089/thy.2015.0460>
- [29] Zhu, H., Yang, Y., Wu, S., Chen, Y., Wang, H., & Zhou, J. (2021). Diagnostic performance of US-based FNAB criteria of the 2020 Chinese guideline for malignant thyroid nodules: Comparison with the 2017 ACR, 2015 ATA, and 2016 KTA guidelines. *Quantitative Imaging in Medicine and Surgery*, 11(8), 3604–3618. <https://doi.org/10.21037/qims-20-1365>
- [30] Cibas, E. S., & Ali, S. Z. (2017). The 2017 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid*, 27(11), 1341–1346. <https://doi.org/10.1089/thy.2017.0500>