INSIGHTS-JOURNAL OF
HEALTH AND REHABILITATION

ASSESSMENT OF THYROID NODULES ULTRASOUND FEATURES UTILIZING TIRADS CLASSIFICATION; CONSISTENCY AMONG OBSERVERS AND COMPARISON WITH HISTOPATHOLOGICAL FINDING IN PAKISTANI POPULATION

Original Research

Swera Saif[1]*, Muhammad Adnan Hafeez[2], Fatima Mahrukh[3], Izza Javaid[2]
[1]Student, MS-AHS, Faculty of Allied Health Sciences, Superior University, Lahore, Pakistan.
[2]Faculty of Allied Health Sciences, Superior University, Lahore, Pakistan.
[3]CMH Lahore, Medical College and Institute of Dentistry, Lahore, Pakistan.
Corresponding Author: Swera Saif, Student, MS-AHS, Faculty of Allied Health Sciences, Superior University, Lahore, Pakistan. sweracheema1212@icloud.com
Acknowledgement: The authors express gratitude to Cancer Care Hospital, Lahore, for their support in conducting this study.

Conflict of Interest: None Grant Support & Financial Support: None

ABSTRACT

Background: Thyroid nodules are a common clinical finding, with the majority being benign, but a subset requiring precise evaluation to rule out malignancy. Ultrasonography (USG) serves as a non-invasive, widely used imaging modality for assessing thyroid nodules, focusing on features such as echogenicity, margins, and calcifications. The Thyroid Imaging Reporting and Data System (TIRADS) standardizes these assessments, enhancing diagnostic accuracy and reducing interobserver variability. Fine-needle aspiration cytology (FNAC) remains the gold standard for histopathological confirmation, guiding appropriate clinical management.

Objective: To assess thyroid nodules using the TIRADS classification system, evaluate interobserver consistency, and compare findings with histopathological diagnoses in the Pakistani population.

Methods: This cross-sectional study was conducted at Cancer Care Hospital, Lahore, including 73 patients with thyroid nodules. Ultrasound imaging was performed using a high-frequency linear transducer (7–14 MHz), with nodules classified according to the TIRADS system by two independent radiologists. Interobserver agreement was assessed using kappa statistics. FNAC or post-surgical histopathology served as the reference standard. Statistical analysis was conducted using SPSS version 25.0, with sensitivity, specificity, and accuracy calculated for TIRADS classification in predicting malignancy.

Results: Interobserver agreement was highest for shape (86%, kappa = 0.860) and margins (85%, kappa = 0.792), followed by echogenicity (76.2%, kappa = 0.879), echogenic foci (74%, kappa = 0.450), and composition (50.2%, kappa = 0.220). The TIRADS system showed overall consistency in 80.2% of cases (59/73). FNAC confirmed 50 benign and 23 malignant nodules. TIRADS 4 and 5 had the highest malignancy correlation. Sensitivity, specificity, and accuracy of TIRADS in predicting malignancy were 62.96%, 98.0%, and 85.7%, respectively.

Conclusion: The study confirms that ultrasound, combined with the TIRADS classification system, is a reliable tool for evaluating thyroid nodules, with strong interobserver consistency. While highly specific, moderate sensitivity suggests the need for FNAC confirmation, particularly in intermediate-risk nodules. The findings reinforce TIRADS as an effective system for malignancy risk stratification, aiding in clinical decision-making.

Keywords: Biopsy, Fine-Needle Aspiration; Neoplasms, Thyroid; Sensitivity and Specificity; Thyroid Nodule; Thyroid Ultrasonography; TIRADS Classification; Ultrasonography.

INTRODUCTION

Thyroid nodules (TNs) are a prevalent clinical finding, with their detection rates increasing significantly over recent decades. The American Thyroid Association defines a thyroid nodule as a discrete lesion within the thyroid gland that is distinct from the surrounding thyroid parenchyma (1). These nodules, which can be benign or malignant, are commonly encountered in clinical practice, with their prevalence varying based on the detection method. They are identified in approximately 4–7% of cases through palpation, 19–35% via ultrasound, and 8–65% in autopsy studies (2). The thyroid gland, the first endocrine gland to develop in humans (3), plays a crucial role in regulating metabolism, growth, and other essential physiological processes. Structurally, it is a butterfly-shaped gland located in the lower anterior neck, extending from the fifth cervical to the first thoracic vertebrae. It comprises two lobes connected by the isthmus and produces vital hormones, primarily thyroxine (T4) and triiodothyronine (T3), which influence numerous bodily functions (4,5). The increasing detection of thyroid nodules is largely attributed to the widespread use of diagnostic imaging modalities, often performed for unrelated medical evaluations (6). Despite the high prevalence of thyroid nodules, the incidence of thyroid malignancy remains relatively low, at approximately 3.2 per 100,000 individuals (7). This suggests that while detection rates are rising, the overall impact on mortality remains limited (8). Risk factors contributing to the development of thyroid nodules include obesity, diabetes, and insulin resistance, further emphasizing the need for effective diagnostic and management strategies (2,9).

Among the available diagnostic tools, high-resolution ultrasound (US) is considered the gold standard for detecting thyroid nodules due to its superior sensitivity compared to physical palpation (10). However, despite its effectiveness in identifying nodules, its diagnostic performance in differentiating benign from malignant nodules remains suboptimal. Certain ultrasound characteristics, such as microcalcifications, hypo echogenicity, taller-than-wide shape, and irregular margins, are associated with malignancy but have limited predictive value (11). This diagnostic uncertainty often necessitates further assessment using fine-needle aspiration biopsy (FNAB), a procedure with high sensitivity and specificity for detecting malignancy. However, FNAB has limitations, including inconclusive results and the potential for overdiagnosis, which may lead to unnecessary interventions (12). To standardize thyroid nodule assessment and risk stratification, the Thyroid Imaging Reporting and Data System (TIRADS) was introduced by Horvath et al. in 2009, modeled after the Breast Imaging Reporting and Data System (BI-RADS). The American College of Radiology (ACR) later modified TIRADS to improve its clinical applicability (13). The adoption of TIRADS and histopathological findings has enhanced diagnostic accuracy, aiding clinicians in decision-making and reducing unnecessary procedures (14). Additionally, ensuring inter-observer consistency among radiologists remains a challenge, as variability in image interpretation can impact clinical decisions. Addressing this issue is crucial for improving diagnostic precision and determining the necessity for further evaluations, such as FNAB or post-surgical pathology (15). Given the rising prevalence of thyroid nodules and the diagnostic challenges associated with distinguishing benign from malignant lesions, this study aims to evaluate current diagnostic approaches and explore advancements that enhance diagnostic accuracy. By investigating the role of ultrasound, FNAB, and risk stratification systems like TIRADS, this research seeks to provide insights that optimize thyroid nodule assessment, minimize unnecessary interventions, and improve patient outcomes (16).

METHODS

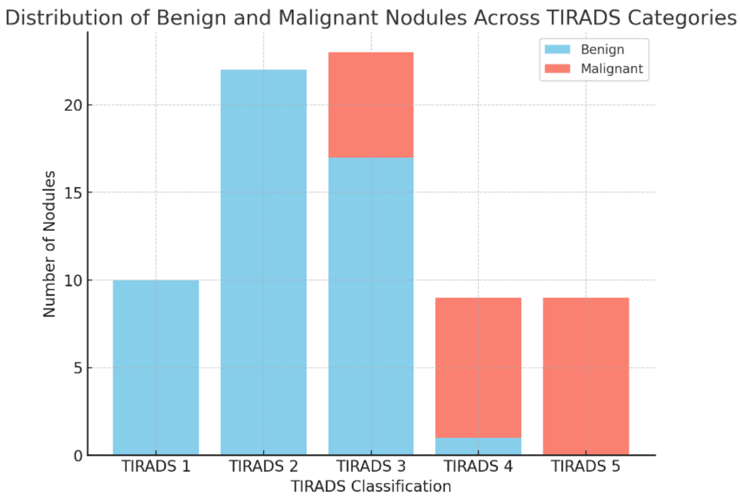
This cross-sectional study was conducted at Cancer Care Hospital, Lahore, including a total of 73 patients. The objective was to assess thyroid nodules using ultrasound features categorized under the TIRADS classification, evaluate inter-observer agreement, and compare findings with histopathological diagnoses in the Pakistani population. Ethical approval was obtained from the institutional review board, and informed consent was obtained from all participants before their inclusion in the study. Patients were enrolled based on predefined inclusion and exclusion criteria. Those with thyroid nodules identified on ultrasound and subsequently undergoing fine-needle aspiration cytology (FNAC) or surgical resection were included. Patients with a history of thyroid surgery, incomplete clinical or imaging data, or non-thyroidal neck lesions were excluded to maintain the reliability of the study results (4,15). Ultrasound imaging was performed using a high-frequency linear transducer ranging from 7 to 14 MHz. Two independent radiologists, one senior with more than ten years of experience and one junior with less than five years of experience, assessed the thyroid nodules using the TIRADS classification system. To minimize bias, the radiologists were blinded to histopathological outcomes while performing ultrasound evaluations. Inter-observer agreement was assessed using the kappa coefficient to determine consistency in TIRADS classification between the two observers.

Patient demographics, including age, gender, and risk factors, were recorded and summarized using measures of central tendency. Histopathological evaluation, considered the gold standard, was performed on samples obtained via FNAC or post-surgical resection to determine the nature of the nodules. Statistical analysis was conducted using Microsoft Excel and SPSS version 25.0, applying descriptive statistics to summarize patient characteristics and inferential statistical methods to determine the correlation between ultrasound-based classifications and histopathological findings. The study ensured that the experience levels of the radiologists were explicitly detailed to assess their potential impact on inter-observer agreement. Additionally, the blinding of radiologists to histopathological results eliminated the possibility of bias in ultrasound assessments, improving the reliability of the diagnostic performance evaluation. While the study sample size was limited, the methodology provides a structured framework for assessing the accuracy of the TIRADS classification system in clinical practice, with an emphasis on reducing observer variability and improving diagnostic reliability.

RESULTS

The study included a total of 73 patients with thyroid nodules, which were assessed based on ultrasound features categorized under the TIRADS classification system. The results demonstrated strong reliability and consistency across multiple ultrasound parameters. Among the evaluated features, echogenicity showed consistent classification in 76.2% (56 out of 73) of cases, with an inter-observer agreement kappa value of 0.879. The shape of the nodules was consistent in 86% (63 out of 73) of cases, with a kappa value of 0.860. Composition exhibited lower consistency at 50.2% (37 out of 73), with a weak inter-observer agreement of 0.220. The margin classification was consistent in 85% (62 out of 73) of cases, with a kappa value of 0.792, while echogenic foci showed agreement in 74% (54 out of 73), with a moderate kappa value of 0.450. Overall, the TIRADS classification was consistent in 80.2% (59 out of 73) of cases, indicating that ultrasound is a reliable tool for thyroid nodule assessment, particularly in parameters such as echogenicity, shape, and margins. The study further examined the correlation between FNAC findings and TIRADS classification. Among the 50 benign nodules confirmed by FNAC, 10 were classified as TIRADS 1, 22 as TIRADS 2, 17 as TIRADS 3, and only 1 as TIRADS 4. In contrast, among the 23 malignant nodules, the majority were categorized as higher-risk TIRADS classifications, with 9 in TIRADS 5, 8 in TIRADS 4, and 6 in TIRADS 3. Notably, no malignant nodules were classified as TIRADS 1 or TIRADS 2, reinforcing the reliability of the TIRADS system in differentiating between benign and malignant nodules. The overall distribution indicated that lower TIRADS categories were primarily associated with benign nodules, whereas higher categories corresponded to a higher likelihood of malignancy, aligning with established diagnostic criteria.

Inter-observer agreement varied across different ultrasound parameters, with the highest consistency observed in echogenicity and shape, while composition showed weaker agreement. This discrepancy suggests that certain ultrasound features may be more subject to interpretation, highlighting the need for further standardization in evaluating thyroid nodules. The findings emphasize the role of ultrasound as a non-invasive and reliable method for monitoring thyroid nodules, particularly when combined with FNAC to enhance diagnostic accuracy and assist in clinical decision-making. The analysis of the TIRADS classification in comparison to FNAC findings revealed an overall diagnostic accuracy of 85.7%. Sensitivity, representing the ability of TIRADS to correctly identify malignant nodules, was 62.96%, while specificity, reflecting the ability to correctly classify benign nodules, was 98.0%. These results indicate that while TIRADS is highly specific in ruling out malignancy, its moderate sensitivity suggests that some malignant nodules may be misclassified into lower-risk categories, potentially leading to underdiagnosis. The inter-observer reliability of the final TIRADS classification was not explicitly analyzed; however, given the strong agreement in echogenicity, shape, and margins but lower consistency in composition and echogenic foci, variability in interpretation could impact the accuracy of malignancy prediction. The findings support the utility of TIRADS in differentiating benign from malignant thyroid nodules, though its sensitivity limitations highlight the need for complementary diagnostic approaches, such as FNAC, particularly for nodules in intermediate-risk categories.



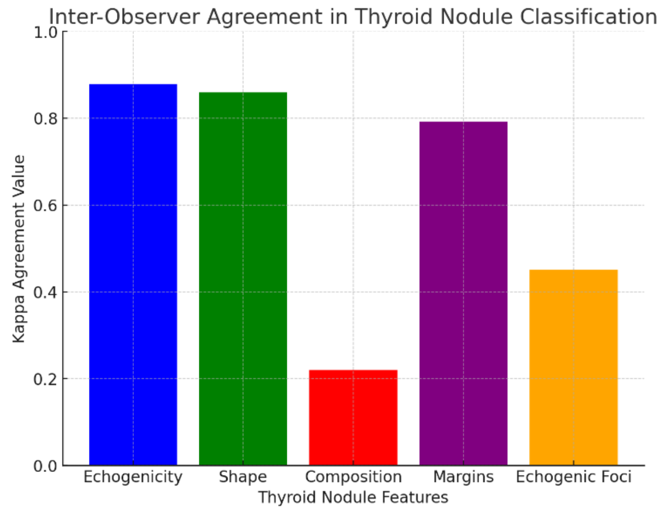


Table No 1: Categories of Thyroid Nodules

	Observer 1	Observer 2
Echogenicity		
Anechoic	12 (16.44) %	11 (15.07) %
Hyperechoic	42 (57.53) %	44 (60.27) %
Hypoechoic	19 (26.03) %	18 (24.66) %
Total	73	73
Measure of agreement (kappa values) b/w obs. 1 and obs. 2	0.879	
Shape	Observer 1	Observer 2
Taller than wider	65 (89.04) %	65 (89.04) %
Wider than taller	8 (10.96) %	8 (10.96) %
Total	73	73
Measure of agreement (kappa values) b/w obs. 1 and obs. 2	0.860	
Composition	Observer 1	Observer 2
Cystic	17 (23.29) %	19 (26.03) %
Spongy	14 (19.18) %	13 (17.81) %
Mixed	38 (52.05) %	38 (52.05) %
Solid	4 (5.48) %	3 (4.11) %
Total	73	73
Measure of agreement (kappa values) b/w obs. 1 and obs. 2	0.220	
Margins	Observer 1	Observer 2

	Observer 1	Observer 2
Smooth	27 (36.9) %	27 (36.9) %
Ill define	38 (52.05) %	37 (50.68) %
Lobulated	6 (8.22) %	6 (8.22) %
Extra thyroidal Extension	2 (2.74) %	3 (4.11) %
Total	73	73
Measure of agreement (kappa values) b/w obs. 1 and obs. 2	0.792	
Echogenic Foci	Observer 1	Observer 2
Commit tail	26 (35.62)	25 (34.25) %
Macro calcification	23 (31.51) %	29 (39.37) %
Peripheral	17 (23.29) %	16 (21.92) %
Punctate	7 (9.59) %	3 (4.11) %
Total	73	73
Measure of agreement (kappa values) b/w obs. 1 and obs. 2	0.450	

Table No 2: Comparison of FNAC Results and TIRADs Classification for Benign and Malignant Thyroid Nodules

FNAC			TIRADsO2					Total
			TIRADs 1	TIRADs 2	TIRADs 3	TIRADs 4	TIRADs 5	
Benign	TIRADsO1	TIRADs 1	10	0	0	0	0	10
		TIRADs 2	0	19	0	2	1	22
		TIRADs 3	1	4	10	2	0	17
		TIRADs 4	0	0	0	1	0	1
	Total		11	23	10	5	1	50
Malignant	TIRADsO1	TIRADs 3	2	0	0	4	0	6
		TIRADs 4	0	2	0	6	0	8
		TIRADs 5	0	0	0	0	9	9
	Total		2	2	0	10	9	23
Total	TIRADsO1	TIRADs 1	10	0	0	0	0	10
		TIRADs 2	0	19	0	2	1	22
		TIRADs 3	3	4	10	6	0	23
		TIRADs 4	0	2	0	7	0	9
		TIRADs 5	0	0	0	0	9	9
	Total		13	25	10	15	10	73

INSIGHTS-JHR
INSIGHTS-JOURNAL OF HEALTH AND REHABILITATION

DISCUSSION

The findings of this study demonstrated a high level of interobserver agreement in key ultrasound features used for thyroid nodule assessment, reinforcing the reliability of the TIRADS classification system in clinical practice. Echogenicity showed a strong consistency of 76.2%, with most nodules classified as hypoechoic. The observed variation between radiologists highlights the need for standardized interpretation, as inconsistencies in echogenicity evaluation could influence malignancy prediction. Previous studies have reported similar trends, emphasizing the strong correlation between hypoechoic nodules and malignancy, alongside high interobserver agreement in their classification. This study further validated these findings, supporting hypo echogenicity as a crucial parameter in malignancy risk stratification. Nodule shape exhibited strong interobserver consistency (86%), particularly in identifying irregular or taller-than-wide nodules, which have been widely associated with an increased risk of malignancy. The findings aligned with previous literature, where taller-than-wide nodules consistently demonstrated a strong association with malignant pathology. Margins were also highly consistent (85%), with ill-defined or spiculated margins being more prevalent in nodules diagnosed as malignant. The reliability of margin assessment has been emphasized in previous studies, where spiculated or irregular margins were considered significant predictors of malignancy. The high level of agreement in margin classification in this study underscores its importance in thyroid nodule assessment.

Composition exhibited lower consistency (50.2%), indicating greater observer variability in differentiating solid, mixed, or predominantly cystic nodules (17). This discrepancy reflects the challenge in accurately categorizing nodules with heterogeneous compositions, a factor noted in prior research where composition assessment showed moderate agreement among radiologists. While solid nodules have been consistently linked to malignancy, the variability in classifying mixed or cystic nodules suggests that further standardization and training are required to improve accuracy in composition evaluation (18). Echogenic foci, including microcalcifications, exhibited moderate interobserver consistency (74%). Microcalcifications have been strongly associated with papillary thyroid carcinoma, and this study corroborated previous findings that identified microcalcifications as a highly specific indicator of malignancy (19). Despite the observed consistency, the identification of punctate echogenic foci demonstrated some variability between observers, which has been similarly reported in previous research. The moderate agreement highlights the need for improved training in the identification of echogenic foci to enhance diagnostic reliability (20).

The TIRADS classification system demonstrated a strong interobserver agreement of 80.2%, reinforcing its clinical utility in stratifying thyroid nodules based on malignancy risk. Higher TIRADS categories, particularly TIRADS 4 and 5, showed a clear association with malignancy, consistent with previous studies that validated the predictive accuracy of TIRADS. The integration of TIRADS with FNAC findings further strengthened its diagnostic value. The overall sensitivity of TIRADS in predicting malignancy was 62.96%, while specificity was 98.0%, with an accuracy of 85.7%. The high specificity of TIRADS supports its role in reducing unnecessary FNAC procedures for benign nodules, although the moderate sensitivity indicates that some malignant nodules may still be under-classified, warranting further evaluation through FNAC or adjunctive imaging techniques. One of the strengths of this study was the systematic assessment of multiple ultrasound parameters across two independent observers, providing a comprehensive evaluation of interobserver agreement. The use of FNAC as a reference standard enhanced the reliability of the findings, ensuring objective malignancy confirmation. Additionally, the study contributed to the limited body of research on the Pakistani population, offering valuable insights into the applicability of the TIRADS classification system in this demographic.

Despite its strengths, this study had certain limitations. The relatively small sample size may have restricted the generalizability of the findings, and the single-institution setting could introduce selection bias. Although blinding of radiologists to FNAC results minimized bias, variability in observer experience may have influenced classification differences, particularly in more subjective features such as composition and echogenic foci. The absence of long-term follow-up data on nodule progression limited the ability to assess the true clinical impact of TIRADS over time. Additionally, while sensitivity, specificity, and accuracy were analyzed, further studies incorporating advanced imaging techniques such as elastography or contrast-enhanced ultrasound could enhance diagnostic precision. The findings of this study reinforce the clinical utility of TIRADS in differentiating benign from malignant thyroid nodules, supporting its role in guiding FNAC recommendations and reducing unnecessary invasive procedures. However, moderate sensitivity underscores the need for a multimodal approach, integrating ultrasound features with cytological and molecular markers to improve malignancy prediction. Future research should focus on multicenter studies with larger sample sizes to further validate these findings and refine the accuracy of ultrasound-based thyroid nodule assessment.

CONCLUSION

In conclusion, ultrasound imaging, in conjunction with the TIRADS classification system, proves to be a reliable and effective tool for evaluating thyroid nodules. The study demonstrated strong interobserver consistency across key ultrasound parameters, reinforcing the utility of standardized classification in clinical decision-making. The findings highlight the value of ultrasound in stratifying malignancy risk, while FNAC remains essential for definitive diagnosis. By enhancing diagnostic accuracy and reducing unnecessary invasive procedures, the integration of ultrasound and TIRADS supports more precise and efficient thyroid nodule management.

Author Contribution

Author	Contribution
Swera Saif*	Substantial Contribution to study design, analysis, acquisition of Data Manuscript Writing Has given Final Approval of the version to be published
Muhammad Adnan Hafeez	Substantial Contribution to study design, acquisition and interpretation of Data Critical Review and Manuscript Writing Has given Final Approval of the version to be published
Fatima Mahrukh	Substantial Contribution to acquisition and interpretation of Data Has given Final Approval of the version to be published
Izza Javaid	Contributed to Data Collection and Analysis Has given Final Approval of the version to be published

REFERENCES

1. Grani G, Sponziello M, Pecce V, Ramundo V, Durante C. Contemporary Thyroid Nodule Evaluation and Management. J Clin Endocrinol Metab. 2020;105(9):2869-83.

2. Alexander EK, Cibas ES. Diagnosis of thyroid nodules. Lancet Diabetes Endocrinol. 2022;10(7):533-9.

3. Durante C, Hegedüs L, Na DG, Papini E, Sipos JA, Baek JH, et al. International Expert Consensus on US Lexicon for Thyroid Nodules. Radiology. 2023;309(1):e231481.

4. Kaltoft M, Stage MG, Todsen T, Hahn CH. [Not Available]. Ugeskr Laeger. 2023;185(9).

5. Chen DW, Lang BHH, McLeod DSA, Newbold K, Haymart MR. Thyroid cancer. Lancet. 2023;401(10387):1531-44.

6. Grani G, Sponziello M, Filetti S, Durante C. Thyroid nodules: diagnosis and management. Nat Rev Endocrinol. 2024;20(12):715-28.

7. Alexander LF, Patel NJ, Caserta MP, Robbin ML. Thyroid Ultrasound: Diffuse and Nodular Disease. Radiol Clin North Am. 2020;58(6):1041-57.

8. Escalante DA, Anderson KG. Workup and Management of Thyroid Nodules. Surg Clin North Am. 2022;102(2):285-307.

9. AlSaedi AH, Almalki DS, ElKady RM, ElKady R. Approach to thyroid nodules: diagnosis and treatment. Cureus. 2024 Jan 13;16(1).

10. Alexander EK, Doherty GM, Barletta JA. Management of thyroid nodules. The Lancet Diabetes & Endocrinology. 2022 Jul 1;10(7):540-8.

11. Grani G, Sponziello M, Filetti S, Durante C. Thyroid nodules: diagnosis and management. Nature Reviews Endocrinology. 2024 Aug 16:1-4.

12. Shi M, Nong D, Xin M, Lin L. Accuracy of Ultrasound Diagnosis of Benign and Malignant Thyroid Nodules: A Systematic Review and Meta-Analysis. International Journal of Clinical Practice. 2022;2022(1):5056082.

13. Mistry R, Hillyar C, Nibber A, Sooriyamoorthy T, Kumar N. Ultrasound classification of thyroid nodules: a systematic review. Cureus. 2020 Mar;12(3).

14. Fresilli D, David E, Pacini P, Del Gaudio G, Dolcetti V, Lucarelli GT, Di Leo N, Bellini MI, D'Andrea V, Sorrenti S, Mascagni D. Thyroid nodule characterization: How to assess the malignancy risk. update of the literature. Diagnostics. 2021 Jul 30;11(8):1374.

15. Dai W, Cui Y, Wang P, Wu H, Zhang L, Bian Y, Li Y, Li Y, Hu H, Zhao J, Xu D. Classification regularized dimensionality reduction improves ultrasound thyroid nodule diagnostic accuracy and inter-observer consistency. Computers in Biology and Medicine. 2023 Mar 1;154:106536.

16. Ahmed MO, Ayad CE, Gareeballah A, Arafat M, ALAli A, Odeh KA, Abdallah E. Assessment of Thyroid Nodules using TIRADS Classification: Interobserver Agreement and Correlation with Histopathology. Int J Biomed. 2022;12(3):385-90.

17. Hekimsoy İ, Öztürk E, Ertan Y, Orman MN, Kavukçu G, Özgen AG, Özdemir M, Özbek SS. Diagnostic performance rates of the ACR-TIRADS and EU-TIRADS based on histopathological evidence. Diagnostic and Interventional Radiology. 2021 Jul;27(4):511.

18. Rago T, Vitti P. Risk stratification of thyroid nodules: from ultrasound features to TIRADS. Cancers. 2022 Jan 30;14(3):717.

19. Soyer Güldoğan E, Ergun O, Taşkın Türkmenoğlu T, Yılmaz KB, Akdağ T, Özbal Güneş S, Durmaz HA, Hekimoğlu B. The impact of TI-RADS in detecting thyroid malignancies: A prospective study. La radiologia medica. 2021 Oct;126(10):1335-44.

20. Cozzani, F., Bettini, D., Rossini, M., Bonati, E., Nuzzo, S., Loderer, T., Pedrazzi, G., Zaccaroni, A. and Del Rio, P., 2021. Thyroid nodules with indeterminate cytology: Association between nodule size, histopathological characteristics and clinical outcome in differentiated thyroid carcinomas—A multicenter retrospective cohort study on 761 patients. *Updates in Surgery*, *73*(5), pp.1923-1930.