



Research Article

Diagnostic Accuracy of Gray-Scale Ultrasonography and Colour Doppler in Differentiating Benign and Malignant Thyroid Nodules: An FNAC-Correlated Study

Krishna Raja Sadasiveni¹, Abburi Anil Kumar², Surya Lakshmi Spandana Pilli³, S. Sai Charan⁴, Gopireddy Vedantha Reddy^{5*} and Rupesh Mandava⁶

¹⁻²Assistant Professor, Department of Radiodiagnosis, Great Eastern Medical School and Hospital, Srikakulam, Andhra Pradesh, India

³⁻⁵Senior Resident, Department of Radiodiagnosis, Great Eastern Medical school and Hospital, Srikakulam, Andhra Pradesh, India

⁶HOD and Professor, Department of Radiodiagnosis, Great Eastern Medical School and Hospital, Srikakulam, Andhra Pradesh, India

***Corresponding Author**

Dr. Gopireddy Vedantha Reddy
(gvedanthareddy@gmail.com)

Article History

Received: 11.06.2026

Accepted: 15.09.2026

Published: 30.09.2026

doi: 10.61336/jdme.0703.14

Abstract: Background: Thyroid nodules are common, but only a small proportion are malignant. Gray-scale ultrasonography is the principal imaging modality for thyroid evaluation, while colour Doppler provides complementary information regarding nodule vascularity. Accurate ultrasound-based risk assessment can improve the selection of nodules requiring FNAC. **Aim:** To determine the diagnostic accuracy of gray-scale ultrasonography and colour Doppler in differentiating benign and malignant thyroid nodules, using FNAC as the reference standard. **Materials and Methods:** This hospital-based analytical cross-sectional diagnostic-accuracy study included 65 patients with thyroid lesions evaluated in the Departments of Radio-diagnosis and Pathology, Navodaya Medical College and Hospital, Raichur, from December 2017 to August 2019. Gray-scale ultrasonography assessed composition, echogenicity, shape, margins, calcification, halo and cervical lymphadenopathy. Colour Doppler assessed the presence and distribution of vascularity. Ultrasonographic diagnoses were correlated with FNAC findings. Sensitivity, specificity, PPV, NPV, accuracy, odds ratios and Cohen's kappa were calculated with 95% CIs. Fisher's exact test or the Fisher-Freeman-Halton exact test was applied, with $p < 0.05$ considered statistically significant. **Results:** FNAC classified 63 (96.9%) lesions as benign and 2 (3.1%) as malignant papillary carcinomas. Ultrasonography correctly differentiated all benign and malignant lesions, yielding sensitivity, specificity, PPV, NPV and binary accuracy of 100%. However, the sensitivity and PPV had wide 95% CIs of 15.8-100.0%. Both malignant nodules had mixed composition, irregular shape, irregular margins and microcalcifications. Irregular shape and irregular margins were significantly associated with malignancy (OR 86.43; 95% CI 3.45-2,167.36; $p = 0.005$ each). Microcalcification demonstrated the strongest association (corrected OR 635.00; 95% CI 10.33-39,038.61; $p < 0.001$). Colour Doppler patterns differed significantly between benign and malignant lesions ($p = 0.001$); combined intranodular and perinodular vascularity occurred in both malignant nodules ($p < 0.001$). Overall exact lesion-specific accuracy was 92.3% (95% CI 83.0-97.5%), with almost perfect agreement between ultrasonography and FNAC ($\kappa = 0.905$, $p < 0.001$). **Conclusion:** Combined gray-scale ultrasonography and colour Doppler demonstrated excellent agreement with FNAC. Irregular shape, irregular margins, microcalcifications and combined intranodular-perinodular vascularity were important indicators of malignancy. However, because only two malignant nodules were included, the apparently perfect binary diagnostic performance should be interpreted cautiously and validated in larger histopathology-correlated studies.

Keywords: Thyroid Nodules, Colour Doppler Ultrasonography, Fine-Needle Aspiration Cytology

Copyright @ 2026: This is an open-access article distributed under the terms of the Creative Commons Attribution license which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use (NonCommercial, or CC-BY-NC) provided the original author and source are credited.

INTRODUCTION

Thyroid nodules are common clinical and radiological findings, particularly among women and older adults. Their detection has increased substantially with the widespread use of high-resolution ultrasonography. Although most thyroid nodules are benign, approximately 5-10% may be malignant; therefore, accurate differentiation between benign

and malignant nodules is essential to guide fine-needle aspiration, surgery and follow-up while avoiding unnecessary invasive procedures [1]. Clinical examination alone cannot reliably characterize thyroid nodules or detect small, non-palpable lesions. Gray-scale ultrasonography is consequently the primary imaging modality for thyroid evaluation because it is readily available, inexpensive, non-ionizing and capable of providing detailed information regarding gland size, echotexture, nodule composition and cervical lymph nodes [2].

Several gray-scale sonographic characteristics are associated with an increased risk of malignancy, including solid composition, marked hypoechogenicity, irregular or infiltrative margins, microcalcifications, a taller-than-wide shape and evidence of extrathyroidal extension. Conversely, purely cystic or spongiform nodules, smooth margins and comet-tail artifacts generally suggest benignity. Standardized risk-stratification systems based on these features have improved the reproducibility of ultrasonographic reporting and have helped identify nodules that require cytological assessment [2-3]. Nevertheless, considerable overlap exists between the appearances of benign and malignant nodules, and no individual gray-scale feature provides complete diagnostic certainty.

Colour Doppler ultrasonography supplements gray-scale imaging by demonstrating the distribution and degree of vascularity within and around a thyroid nodule. Predominantly peripheral vascularity is more frequently encountered in benign nodules, whereas marked intranodular vascularity, irregular vascular patterns and penetrating vessels may raise suspicion of malignancy. However, vascularity alone has limited specificity and should be interpreted along with the morphology of the nodule [4].

Fine-needle aspiration cytology (FNAC) is a minimally invasive, safe and cost-effective procedure for assessing thyroid nodules. It plays a central role in distinguishing nodules that require surgery from those that can be managed conservatively. Ultrasound guidance improves sampling accuracy, particularly in small, deeply situated, partially cystic or heterogeneous nodules [1,5]. Correlation of gray-scale and colour Doppler findings with FNAC can therefore determine the diagnostic performance of ultrasonography and identify the imaging characteristics most strongly associated with malignancy. The present study was undertaken to evaluate gray-scale and colour Doppler features of thyroid nodules and determine their diagnostic accuracy in differentiating benign from malignant nodules, using FNAC as the reference standard.

Aim

To determine the diagnostic accuracy of gray-scale ultrasonography and colour Doppler in differentiating benign and malignant thyroid nodules, using FNAC as the reference standard.

Objectives

- To describe the gray-scale ultrasonographic characteristics of thyroid nodules, including composition, echogenicity, shape, margins and calcification
- To assess the colour Doppler vascularity patterns of benign and malignant thyroid nodules
- To correlate ultrasonographic diagnoses with FNAC findings and calculate sensitivity, specificity, positive predictive value, negative predictive value and diagnostic accuracy

MATERIALS AND METHODS

Source of Data

The study population consisted of patients with clinically palpable thyroid swellings or thyroid nodules detected during ultrasonography who attended or were referred from the outpatient and inpatient departments of Navodaya Medical College and Hospital, Raichur. Patients were referred to the Department of Radio-diagnosis for ultrasonographic evaluation and subsequently underwent FNAC in coordination with the Department of Pathology.

Study Design

This was a hospital-based analytical cross-sectional diagnostic-accuracy study. Gray-scale ultrasonography and colour Doppler constituted the index tests, while FNAC was considered the reference standard for classifying thyroid nodules as benign or malignant.

Study Location

The study was conducted jointly in the Department of Radio-diagnosis and the Department of Pathology, Navodaya Medical College and Hospital, Raichur, Karnataka.

Study Duration

The study was carried out from December 2017 to August 2019.

Sample Size

A total of 65 eligible patients with thyroid lesions were included. The minimum sample size was calculated using:

$$n = \frac{4pq}{d^2}$$

where p was the anticipated prevalence of the relevant diagnostic outcome (74%), $q = 100 - p = 26\%$, and the allowable error was taken as 15% of the anticipated prevalence, approximately 11.1%.

$$n = \frac{4 \times 74 \times 26}{(11.1)^2} \approx 62$$

The calculated minimum sample size was 62. It was increased to 65 to account for possible incomplete or non-diagnostic observations.

Sampling Technique

A convenience sampling technique was used. All eligible patients presenting during the study period were approached and enrolled until the required sample size of 65 was achieved.

Inclusion Criteria

- Patients aged 10-75 years
- Patients of either sex with a clinically palpable thyroid swelling or a thyroid lesion detected on ultrasonography
- Patients with symptomatic or asymptomatic thyroid nodules, irrespective of thyroid hormone status
- Patients who underwent both gray-scale/colour Doppler ultrasonography and FNAC
- Patients who provided written informed consent; assent and parental or guardian consent were obtained for eligible minors

Exclusion Criteria

- Patients with a known bleeding or coagulation disorder
- Patients receiving anticoagulant therapy when FNAC could not be safely performed
- Patients who refused ultrasonography or FNAC
- Patients whose FNAC specimen was inadequate or non-diagnostic and in whom repeat sampling could not be performed
- Patients with incomplete clinical, ultrasonographic or cytological records

Data Collection

After approval from the Institutional Ethics Committee, eligible patients were informed about the nature and purpose of the study. Written informed consent was obtained before enrolment. Information was recorded on a predesigned and pretested data-collection proforma.

The collected information included age, sex, presenting symptoms, duration of swelling, pain, dysphagia, hoarseness of voice, weight changes, menstrual disturbances, previous thyroid disease, relevant treatment history and available thyroid-function test results. Findings of clinical examination, gray-scale ultrasonography, colour Doppler and FNAC were recorded using unique study identification numbers. Personal identifiers were kept confidential.

Procedure and Methodology

Each patient was examined in the supine position with the neck mildly extended by placing a pillow beneath the shoulders. Ultrasonography was performed using a GE LOGIQ P6 ultrasound system equipped with a high-frequency 3-12 MHz linear-array transducer.

Both thyroid lobes and the isthmus were examined in transverse and longitudinal planes. The following characteristics were documented:

- Thyroid gland size and overall echotexture
- Homogeneous or heterogeneous background parenchyma

- Presence of solitary or multiple nodules
- Site and maximum dimensions of the nodule
- Solid, cystic or mixed composition
- Echogenicity relative to the normal thyroid parenchyma
- Shape, including wider-than-tall or taller-than-wide configuration
- Regular, irregular or ill-defined margins
- Presence and type of calcification
- Halo around the nodule
- Cystic or degenerative changes
- Extrathyroidal extension and suspicious cervical lymph nodes

Nodules measuring less than 5 mm were recorded but were not subjected to detailed characterization unless clinically indicated. Nodules measuring at least 5 mm were systematically evaluated.

Colour Doppler examination was subsequently performed with optimized gain, pulse-repetition frequency and wall-filter settings. Vascularity was categorized as absent, predominantly peripheral, predominantly intranodular or combined peripheral and intranodular. The overall sonographic impression was classified as benign or malignant based on the combined gray-scale and Doppler characteristics. In patients with multiple nodules, the dominant or most suspicious nodule was selected for correlation with FNAC.

FNAC Procedure

The selected nodule was localized under ultrasonographic guidance. The patient remained supine with the neck extended, and the skin was cleaned with povidone-iodine solution using aseptic precautions. The patient was instructed to avoid swallowing or moving during needle placement.

A sterile 23-27-gauge needle attached to a 10-mL disposable syringe was inserted into the target nodule using a parallel or perpendicular approach to the transducer. The needle tip was continuously visualized to avoid adjacent vessels and ensure sampling of the solid or most suspicious component. At least two passes were made when necessary to obtain an adequate sample. Local anaesthesia with lignocaine was used when clinically required.

Sample Processing

The aspirated material was expelled onto clean, labelled glass slides and immediately spread into thin smears. Wet-fixed smears were placed in 95% ethanol and stained using the Papanicolaou method. Additional smears were prepared and stained according to the routine protocol of the pathology laboratory when required.

The slides were examined by a pathologist and categorized as benign, suspicious or malignant according to cytomorphological findings. Colloid nodules, colloid cysts, multinodular goitre, adenomatous or hyperplastic nodules and thyroiditis were considered benign. Cytologically malignant lesions, including papillary carcinoma and other definite malignant neoplasms, were classified as malignant. Inadequate samples were repeated whenever feasible. Follicular-patterned lesions were interpreted cautiously because cytology cannot reliably distinguish follicular adenoma from follicular carcinoma without demonstrating capsular or vascular invasion.

Statistical Methods

Data were entered into Microsoft Excel and analysed using IBM SPSS Statistics. Continuous variables were summarized as mean and standard deviation or median and interquartile range, as appropriate. Categorical variables were presented as frequencies and percentages.

Gray-scale and colour Doppler findings were compared between FNAC-confirmed benign and malignant nodules using the chi-square test or Fisher's exact test. A two-by-two diagnostic table was constructed by considering malignancy on FNAC as disease-positive. The following diagnostic indices were calculated with 95% confidence intervals:

$$\begin{aligned}\text{Sensitivity} &= \frac{TP}{TP + FN} \times 100 \\ \text{Specificity} &= \frac{TN}{TN + FP} \times 100 \\ \text{Positive predictive value} &= \frac{TP}{TP + FP} \times 100 \\ \text{Negative predictive value} &= \frac{TN}{TN + FN} \times 100 \\ \text{Diagnostic accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\end{aligned}$$

Agreement between ultrasonographic and FNAC classifications was evaluated using Cohen's kappa coefficient. All statistical tests were two-sided, and $p < 0.05$ was considered statistically significant.

RESULTS

Among the 65 thyroid lesions, FNAC classified 63 (96.9%) as benign and 2 (3.1%) as malignant. Combined gray-scale ultrasonography and colour Doppler correctly identified all 63 benign lesions and both malignant lesions, with no false-positive or false-negative diagnoses. The association between ultrasonographic and FNAC classifications was statistically significant (Fisher's exact $p < 0.001$), with a Phi coefficient of 1.000. Sensitivity, specificity, PPV, NPV and diagnostic accuracy were each 100.0%. The corrected diagnostic odds ratio was 635.00 (95% CI 10.33-39,038.61), while the negative likelihood ratio was 0.00. However, the 95% CIs for sensitivity and PPV were wide (15.8-100.0%) because only two malignant lesions were included Table 1.

Gray-scale characteristics differed between benign and malignant lesions. Among the 63 benign lesions, 43 (68.3%) were solid, 10 (15.9%) were cystic and 10 (15.9%) had mixed composition. Both malignant lesions had mixed composition, although the overall association between composition and malignancy was borderline and not statistically significant ($p = 0.053$). Irregular shape and irregular margins were each observed in both malignant lesions but in only 3 (4.8%) benign lesions. Both characteristics were significantly associated with malignancy, with an OR of 86.43 (95% CI 3.45-2,167.36; $p = 0.005$). Predominant hypoechogenicity was present in all malignant lesions and 18 (28.6%) benign lesions, but the association was not significant (OR 12.30; 95% CI 0.56-268.68; $p = 0.091$). Heterogeneous echotexture was also present in both malignant lesions and 35 (55.6%) benign lesions without a significant association ($p = 0.502$). Although any calcification was present in both malignant lesions, its association with malignancy was borderline (OR 17.07; 95% CI 0.77-375.99; $p = 0.058$). Microcalcification was present in both malignant lesions and

Table 1: Diagnostic Accuracy of Combined Gray-Scale Ultrasonography and Colour Doppler for Differentiating Benign and Malignant Thyroid Lesions, Using FNAC as The Reference Standard (N = 65)

A. Diagnostic Cross-Tabulation

Ultrasonographic diagnosis	Benign on FNAC, n (%)	Malignant on FNAC, n (%)	Total, n (%)
Benign	63 (100.0)	0 (0.0)	63 (96.9)
Malignant	0 (0.0)	2 (100.0)	2 (3.1)
Total	63 (100.0)	2 (100.0)	65 (100.0)

Fisher's exact test: $p < 0.001$; Phi coefficient = 1.000

B. Diagnostic-Performance Measures

Diagnostic measure	Estimate, %	95% CI
Sensitivity	100.0	15.8-100.0
Specificity	100.0	94.3-100.0
Positive predictive value	100.0	15.8-100.0
Negative predictive value	100.0	94.3-100.0
Diagnostic accuracy	100.0	94.5-100.0
Positive likelihood ratio	Not estimable	Specificity was 100%
Negative likelihood ratio	0.00	
Corrected diagnostic odds ratio†	635.00	10.33-39,038.61

†Haldane-Anscombe correction was applied because the false-positive and false-negative cells were zero

Table 2: Gray-Scale Ultrasonographic Characteristics of Benign and Malignant Thyroid Lesions According to FNAC (N = 65)

Gray-scale characteristic	Benign (n = 63), n (%)	Malignant (n = 2), n (%)	OR for malignancy (95% CI)†	Test statistic	p value
Composition				Fisher-Freeman-Halton exact test	0.053
Cystic	10 (15.9)	0 (0.0)			
Solid	43 (68.3)	0 (0.0)			
Mixed	10 (15.9)	2 (100.0)			
Irregular shape	3 (4.8)	2 (100.0)	86.43 (3.45-2,167.36)	Fisher's exact test	0.005
Irregular margin	3 (4.8)	2 (100.0)	86.43 (3.45-2,167.36)	Fisher's exact test	0.005
Predominantly hypoechoic	18 (28.6)	2 (100.0)	12.30 (0.56-268.68)	Fisher's exact test	0.091
Heterogeneous echotexture	35 (55.6)	2 (100.0)	4.01 (0.19-87.01)	Fisher's exact test	0.502
Any calcification	14 (22.2)	2 (100.0)	17.07 (0.77-375.99)	Fisher's exact test	0.058
Microcalcification	0 (0.0)	2 (100.0)	635.00 (10.33-39,038.61)	Fisher's exact test	<0.001
Hypoechoic halo present	15 (23.8)	2 (100.0)	15.65 (0.71-343.77)	Fisher's exact test	0.065
Cervical lymphadenopathy	17 (27.0)	2 (100.0)	13.29 (0.61-290.75)	Fisher's exact test	0.082

†Odds ratios and their 95% CIs were calculated using the Haldane-Anscombe correction when a zero cell was present

Table 3: Colour Doppler Vascularity Patterns of Benign and Malignant Thyroid Lesions According to FNAC (N = 65)

A. Distribution of Doppler Patterns

Colour Doppler pattern	Benign (n = 63), n (%)	Malignant (n = 2), n (%)	Total, n (%)
No detectable vascularity	41 (65.1)	0 (0.0)	41 (63.1)
Vascularity present, pattern not further specified	20 (31.7)	0 (0.0)	20 (30.8)
Predominantly peripheral vascularity	2 (3.2)	0 (0.0)	2 (3.1)
Combined intranodular and perinodular vascularity	0 (0.0)	2 (100.0)	2 (3.1)
Total	63 (100.0)	2 (100.0)	65 (100.0)

Fisher-Freeman-Halton exact test: p = 0.001

B. Association of selected Doppler Findings with Malignancy

Doppler characteristic	Benign (n = 63), n (%)	Malignant (n = 2) n (%)	OR for malignancy (95% CI) [†]	Test statistic	p value
Any detectable vascularity	22 (34.9)	2 (100.0)	9.22 (0.42-200.55)	Fisher's exact test	0.133
Predominantly peripheral vascularity	2 (3.2)	0 (0.0)	4.92 [‡] (0.18-133.65)	Fisher's exact test	1.000
Combined intranodular and perinodular vascularity	0 (0.0)	2 (100.0)	635.00 (10.33-39,038.61)	Fisher's exact test	<0.001

[†]Haldane-Anscombe correction was applied to zero cells. [‡]Odds ratio expressed in favour of a benign diagnosis for predominantly peripheral vascularity

Table 4: Correlation Between Specific Ultrasonographic and FNAC Diagnoses and Category-Wise Diagnostic Performance (N = 65)

Lesion category	USG positive/FNAC positive, n	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)	Accuracy, % (95% CI)	Cohen's κ	p value [†]
Colloid cyst	10/10	100.0 (69.2-100.0)	100.0 (93.5-100.0)	100.0 (69.2-100.0)	100.0 (93.5-100.0)	100.0 (94.5-100.0)	1.000	<0.001
Colloid/hyperplastic/adenomatous nodule	12/12	100.0 (73.5-100.0)	100.0 (93.3-100.0)	100.0 (73.5-100.0)	100.0 (93.3-100.0)	100.0 (94.5-100.0)	1.000	<0.001
Colloid nodule with cystic degeneration	10/10	100.0 (69.2-100.0)	100.0 (93.5-100.0)	100.0 (69.2-100.0)	100.0 (93.5-100.0)	100.0 (94.5-100.0)	1.000	<0.001
Multinodular goitre	9/11	81.8 (48.2-97.7)	96.3 (87.3-99.5)	81.8 (48.2-97.7)	96.3 (87.3-99.5)	93.8 (85.0-98.3)	0.781	<0.001
Hashimoto's thyroiditis	16/18	88.9 (65.3-98.6)	93.6 (82.5-98.7)	84.2 (60.4-96.6)	95.7 (85.2-99.5)	92.3 (83.0-97.5)	0.811	<0.001
Graves' disease	1/2	50.0 (1.3-98.7)	100.0 (94.3-100.0)	100.0 (2.5-100.0)	98.4 (91.6-100.0)	98.5 (91.7-100.0)	0.660	0.031
Papillary carcinoma	2/2	100.0 (15.8-100.0)	100.0 (94.3-100.0)	100.0 (15.8-100.0)	100.0 (94.3-100.0)	100.0 (94.5-100.0)	1.000	<0.001
Overall exact lesion diagnosis	60/65					92.3 (83.0-97.5)	0.905	<0.001

[†]Fisher's exact test was used for each one-versus-all diagnostic comparison

absent from all benign lesions, producing the strongest association with malignancy (corrected OR 635.00; 95% CI 10.33-39,038.61; p < 0.001). Hypoechoic halo and cervical lymphadenopathy were also observed in both malignant lesions, but their associations did not reach statistical significance (p = 0.065 and p = 0.082, respectively) Table 2.

Colour Doppler demonstrated significantly different vascularity patterns between benign and malignant thyroid lesions (Fisher-Freeman-Halton p = 0.001). Among the benign lesions, 41 (65.1%) had no detectable vascularity, 20 (31.7%) had vascularity without further pattern specification and 2 (3.2%) showed predominantly peripheral vascularity. Both malignant lesions demonstrated combined intranodular and perinodular vascularity, a pattern not observed in any benign lesion. Although any detectable vascularity was present in all malignant lesions compared with 22 (34.9%) benign lesions, this association was not statistically significant (OR 9.22; 95% CI 0.42-200.55; p = 0.133). Predominantly peripheral vascularity was not associated with malignancy (p = 1.000). In contrast, combined intranodular and perinodular vascularity was strongly associated with malignancy, with a corrected OR of 635.00 (95% CI 10.33-39,038.61; p < 0.001) Table 3.

The overall association between the seven-category USG and FNAC diagnoses was $\chi^2 = 268.57$, p < 0.001. Ultrasonographic diagnoses showed strong agreement with FNAC across the different lesion categories. Colloid cysts, colloid/hyperplastic/adenomatous nodules and colloid nodules with cystic degeneration demonstrated 100.0% sensitivity, specificity, PPV, NPV and accuracy, with perfect agreement ($\kappa = 1.000$, p < 0.001). For multinodular goitre, sensitivity and PPV were 81.8%, specificity and NPV were 96.3%, and accuracy was 93.8% (95% CI 85.0-98.3%), with substantial agreement ($\kappa = 0.781$, p < 0.001). Hashimoto's thyroiditis showed a sensitivity of 88.9%, specificity of 93.6%, PPV of 84.2%, NPV of 95.7% and accuracy of 92.3% (95% CI 83.0-97.5%), with strong agreement ($\kappa = 0.811$, p < 0.001). Graves' disease had a lower sensitivity of 50.0%, but its specificity and PPV were 100.0%, with an accuracy

Table 5: Overall Lesion-Wise Agreement

USG diagnosis	Correctly confirmed by FNAC, n/N (%)	Misclassified, n
Colloid cyst	10/10 (100.0)	0
Colloid/hyperplastic/adenomatous nodule	12/12 (100.0)	0
Colloid nodule with cystic degeneration	10/10 (100.0)	0
Multinodular goitre	9/11 (81.8)	2
Hashimoto's thyroiditis	16/19 (84.2)	3
Graves' disease	1/1 (100.0)	0
Papillary carcinoma	2/2 (100.0)	0
Overall	60/65 (92.3)	5

of 98.5% and substantial agreement ($\kappa = 0.660$, $p = 0.031$). Both papillary carcinomas were correctly identified, producing 100.0% diagnostic indices and perfect agreement, although the confidence intervals for sensitivity and PPV were wide. Overall, ultrasonography provided the exact lesion diagnosis in 60 of 65 cases, corresponding to an accuracy of 92.3% (95% CI 83.0-97.5%) and almost perfect agreement with FNAC ($\kappa = 0.905$). The overall association was statistically significant ($\chi^2 = 268.57$, $p < 0.001$) Table 4-5.

DISCUSSION

The present FNAC-correlated study evaluated the diagnostic utility of gray-scale ultrasonography and colour Doppler in 65 patients with thyroid lesions. FNAC categorized 63 lesions as benign and two as malignant papillary carcinomas. Combined ultrasonography correctly classified all lesions as benign or malignant, resulting in sensitivity, specificity, PPV, NPV and accuracy of 100%. However, the sensitivity and PPV had wide 95% confidence intervals because only two malignant lesions were encountered. Therefore, the apparently perfect performance should be interpreted as an estimate from a small and markedly imbalanced sample rather than evidence that ultrasonography can replace cytological examination.

Haugen *et al.* [1] recognized ultrasonography as the primary imaging investigation for thyroid nodules but emphasized that imaging findings should guide, rather than replace, FNAC. Similarly, Tessler *et al.* [2] developed ACR TI-RADS to improve risk stratification and reduce unnecessary FNAC. Russ *et al.* [3] proposed EU-TIRADS and demonstrated that combinations of suspicious gray-scale characteristics provide better diagnostic discrimination than any isolated feature. The excellent binary accuracy in the present study was consistent with the high sensitivity and specificity reported for high-suspicion ultrasound categories in these systems. Nevertheless, the low malignant prevalence of 3.1% was lower than that reported in most referral-based studies and increased the observed NPV and overall accuracy.

Ha *et al.* [4], after applying seven professional-society guidelines to 2,000 thyroid nodules, found that diagnostic performance differed according to the sonographic classification and FNAC threshold. Grani *et al.* [5] also showed that structured risk-stratification systems could maintain high sensitivity while decreasing unnecessary biopsies. Ha *et al.* [6] observed that biopsy rates depended on both the selected guideline and nodule-size threshold. These studies support the use of systematic feature-based assessment rather than an unstructured overall impression. In the present study, ultrasonography achieved an overall exact lesion-specific agreement of 92.3%, with Cohen's $\kappa = 0.905$, indicating almost perfect agreement with FNAC.

The gray-scale findings showed that both malignant lesions had mixed composition, irregular shape, irregular margins and microcalcifications. Irregular shape and irregular margins were significantly associated with malignancy, each producing an OR of 86.43 and $p = 0.005$. Microcalcification demonstrated the strongest association, with a corrected OR of 635.00 and $p < 0.001$. These findings agree with recognized thyroid ultrasound risk-stratification systems, in which irregular margins, punctate echogenic foci and non-oval or taller-than-wide shape are major indicators of malignancy [2,3,7]. Castellana *et al.* [7] reported that the major ultrasound-based systems offered useful sensitivity for selecting nodules for FNAC, although their specificity and biopsy rates varied. Kim *et al.* [8] likewise found good overall diagnostic performance across four commonly used ultrasound risk-stratification systems.

Both malignant nodules in the present study were predominantly hypoechoic and heterogeneous. However, hypoechogenicity did not reach statistical significance ($p = 0.091$), and heterogeneous echotexture was also non-significant ($p = 0.502$). This was probably related to the presence of hypoechogenicity and heterogeneity in benign inflammatory conditions, particularly Hashimoto's thyroiditis. Ha *et al.* [9] emphasized that hypoechogenicity alone has limited specificity and should be interpreted with shape, margins, echogenic foci and composition. The absence of statistical significance in the present study therefore did not imply that hypoechogenicity lacked clinical importance; rather, it indicated overlap between malignant nodules and benign diffuse thyroid disease.

Any calcification was found in 22.2% of benign lesions and both malignant lesions, with a borderline association ($p = 0.058$). When the type of calcification was considered, microcalcification was present in both papillary carcinomas

and absent from benign lesions, producing a highly significant relationship. This distinction is clinically important because coarse or peripheral calcification can occur in benign degenerative nodules, whereas punctate echogenic foci corresponding to psammoma bodies are strongly associated with papillary carcinoma [2,9]. Trimboli *et al.* [10] reported that current risk-stratification systems had been evaluated predominantly against papillary thyroid carcinoma, whose characteristic ultrasound features include microcalcifications and irregular margins. Thus, the strong association observed here may partly reflect the fact that both malignant lesions were papillary carcinomas.

A hypoechoic halo and cervical lymphadenopathy occurred in both malignant lesions but did not reach statistical significance. The corresponding estimates had very wide confidence intervals, reflecting the small malignant group. Moreover, a thin, complete halo is generally associated with benign encapsulated nodules, whereas a thick or incomplete halo may be suspicious. Likewise, reactive cervical lymph nodes can accompany Hashimoto's thyroiditis and other benign inflammatory conditions. Consequently, the morphology of the halo and lymph nodes is more informative than their presence alone.

Colour Doppler patterns differed significantly between benign and malignant lesions ($p = 0.001$). Both malignant nodules demonstrated combined intranodular and perinodular vascularity, whereas predominantly peripheral vascularity occurred only among benign nodules. Combined intranodular and perinodular vascularity was strongly associated with malignancy ($p < 0.001$). Nevertheless, the presence of vascularity of any type was not significant ($p = 0.133$). Khadra *et al.* [11], in a systematic review and meta-analysis, concluded that increased intranodular or peripheral vascularity alone could not reliably predict thyroid malignancy. This supports the present observation that the distribution of blood flow was more informative than the simple presence or absence of vascularity. Colour Doppler should therefore be considered an adjunct to gray-scale morphology and not an independent basis for recommending FNAC.

The lesion-specific analysis showed complete agreement for colloid cysts, colloid/hyperplastic/adenomatous nodules, colloid nodules with cystic degeneration and papillary carcinoma. Diagnostic accuracy was slightly lower for multinodular goitre (93.8%) and Hashimoto's thyroiditis (92.3%), although agreement remained substantial. Misclassification occurred mainly between multinodular goitre and Hashimoto's thyroiditis because both can produce glandular heterogeneity, multiple pseudonodules and altered vascularity. Graves' disease had a sensitivity of only 50.0%, although its specificity was 100%. The wide confidence interval for sensitivity reflected the presence of only two FNAC-confirmed cases.

Kim *et al.* [12] found that the rate of unnecessary FNAC differed substantially across ultrasound risk-stratification systems, illustrating the trade-off between maintaining sensitivity and improving specificity. Dobruch-Sobczak *et al.* [13] similarly concluded that ACR TI-RADS, EU-TIRADS and K-TIRADS all provided useful malignancy stratification but differed in their categorization and biopsy recommendations. Durante *et al.* [14] recommended integrating clinical history, ultrasound risk category, nodule size and cytological findings when determining management. Piticchio *et al.* [15] also found meaningful differences among American, European and Asian TI-RADS systems in head-to-head comparisons. Collectively, these findings suggest that high diagnostic performance depends on standardized reporting and combined interpretation of multiple features.

FNAC was an appropriate minimally invasive comparator for this study and remains central to clinical decision-making. Cibas and Ali [16] standardized cytological reporting through the Bethesda System and linked each category with malignancy risk and management recommendations. However, FNAC cannot reliably distinguish follicular adenoma from follicular carcinoma because capsular or vascular invasion requires histopathological examination. The absence of histopathological verification in the present study may therefore have introduced reference-standard misclassification.

Overall, the findings demonstrated that gray-scale ultrasonography provided the major discriminatory features, while colour Doppler supplied supportive vascular information. Irregular shape, irregular margins and microcalcifications were the most important malignant features. The observed 100% binary accuracy and 92.3% exact lesion-specific accuracy were encouraging, but external validation in a larger sample with more malignant nodules and histopathological confirmation is required.

CONCLUSION

Combined gray-scale ultrasonography and colour Doppler demonstrated excellent agreement with FNAC in differentiating benign and malignant thyroid lesions. All 63 benign lesions and both papillary carcinomas were correctly classified, producing 100% sensitivity, specificity, PPV, NPV and binary diagnostic accuracy. Exact lesion-specific agreement was 92.3%, with almost perfect agreement with FNAC ($\kappa = 0.905$, $p < 0.001$).

Irregular shape, irregular margins and microcalcifications were significantly associated with malignancy. Combined intranodular and perinodular vascularity was also strongly associated with malignant nodules, whereas the simple

presence of vascularity was not independently significant. Gray-scale morphology therefore remained the principal basis for malignancy assessment, with colour Doppler serving as a complementary tool.

Despite the favourable results, the very small number of malignant nodules produced wide confidence intervals. Ultrasonography should consequently be used for risk stratification and selection of nodules for FNAC rather than as a replacement for cytological or histopathological diagnosis.

Limitations

- The study was conducted at a single tertiary-care hospital; therefore, the findings might not be generalizable to other populations or healthcare settings
- The sample size was relatively small, with only 65 participants
- Only two malignant nodules were identified. This produced wide confidence intervals for sensitivity, PPV and malignancy-associated odds ratios
- The very low prevalence of malignancy resulted in spectrum imbalance and may have inflated specificity, NPV and overall diagnostic accuracy
- Convenience sampling was used, which may have introduced selection bias
- FNAC rather than histopathological examination was used as the reference standard for all lesions
- FNAC could not reliably distinguish follicular adenoma from follicular carcinoma because assessment of capsular and vascular invasion requires histopathology
- Surgical histopathological confirmation was unavailable for cytologically benign nodules; therefore, false-negative FNAC findings could not be completely excluded
- The study did not employ a contemporary standardized scoring system such as ACR TI-RADS, EU-TIRADS or K-TIRADS
- Doppler vascularity was incompletely subclassified in some benign lesions, limiting detailed comparison of peripheral, intranodular and mixed vascular patterns
- Ultrasonography is operator-dependent, and interobserver and intraobserver agreement were not evaluated
- The radiologist and cytopathologist might not have been completely blinded to the patients' clinical or imaging information, creating a possibility of observer bias
- Advanced techniques such as elastography, contrast-enhanced ultrasonography and quantitative vascular indices were not assessed
- The study did not perform long-term follow-up of cytologically benign nodules to identify subsequent growth or malignant transformation

REFERENCES

1. Haugen, B. R. *et al.* "2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer." *Thyroid*, vol. 26, no. 1, 2016, pp. 1-133. <https://doi.org/10.1089/thy.2015.0020>.
2. Tessler, F. N. *et al.* "ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee." *Journal of the American College of Radiology*, vol. 14, no. 5, 2017, pp. 587-595. <https://doi.org/10.1016/j.jacr.2017.01.046>.
3. Russ, G. *et al.* "European Thyroid Association Guidelines for Ultrasound Malignancy Risk Stratification of Thyroid Nodules in Adults: The EU-TIRADS." *European Thyroid Journal*, vol. 6, no. 5, 2017, pp. 225-237.
4. Ha, E. J. *et al.* "US Fine-Needle Aspiration Biopsy for Thyroid Malignancy: Diagnostic Performance of Seven Society Guidelines Applied to 2,000 Thyroid Nodules." *Radiology*, vol. 287, no. 3, 2018, pp. 893-900. <https://doi.org/10.1148/radiol.2018171074>.
5. Grani, G. *et al.* "Reducing the Number of Unnecessary Thyroid Biopsies While Improving Diagnostic Accuracy: Toward the 'Right' TIRADS." *The Journal of Clinical Endocrinology & Metabolism*, vol. 104, no. 1, 2019, pp. 95-102. <https://doi.org/10.1210/jc.2018-01674>.
6. Ha, S. M. *et al.* "Diagnostic Performance of Practice Guidelines for Thyroid Nodules: Thyroid Nodule Size versus Biopsy Rates." *Radiology*, vol. 291, no. 1, 2019, pp. 92-99. <https://doi.org/10.1148/radiol.2019181723>.
7. Castellana, M. *et al.* "Performance of Five Ultrasound Risk-Stratification Systems in Selecting Thyroid Nodules for FNA." *The Journal of Clinical Endocrinology & Metabolism*, vol. 105, no. 5, 2020, pp. 1659-1669. <https://doi.org/10.1210/clinem/dgz170>.
8. Kim, P. H. *et al.* "Diagnostic Performance of Four Ultrasound Risk-Stratification Systems: A Systematic Review and Meta-Analysis." *Thyroid*, vol. 30, no. 8, 2020, pp. 1159-1168. <https://doi.org/10.1089/thy.2019.0812>.
9. Ha, E. J. *et al.* "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*, vol. 22, no. 12, 2021, pp. 2094-2123. <https://doi.org/10.3348/kjr.2021.0713>.
10. Trimboli, P. *et al.* "The Ultrasound Risk-Stratification Systems for Thyroid Nodules Have Been Evaluated against Papillary Carcinoma: A Meta-Analysis." *Reviews in Endocrine and Metabolic Disorders*, vol. 22, no. 2, 2021, pp. 453-460. <https://doi.org/10.1007/s11154-020-09592-3>.
11. Khadra, H. *et al.* "Is Vascular Flow a Predictor of Malignant Thyroid Nodules? A Meta-Analysis." *Gland Surgery*, vol. 5, no. 6, 2016, pp. 576-582. <https://doi.org/10.21037/gs.2016.11.03>.

12. Kim, P. H. *et al.* "Unnecessary Thyroid Nodule Biopsy Rates under Four Ultrasound Risk-Stratification Systems: A Systematic Review and Meta-Analysis." *European Radiology*, vol. 31, no. 5, 2021, pp. 2877–2885. <https://doi.org/10.1007/s00330-020-07384-6>.
13. Dobruch-Sobczak, K. *et al.* "Summary of Meta-Analyses of Studies Involving TI-RADS Classifications in Evaluating the Malignant Potential of Focal Thyroid Lesions." *Journal of Ultrasonography*, vol. 22, no. 89, 2022, pp. 121–129. <https://doi.org/10.15557/JoU.2022.0020>.
14. Durante, C. *et al.* "2023 European Thyroid Association Clinical Practice Guidelines for Thyroid Nodule Management." *European Thyroid Journal*, vol. 12, no. 5, 2023, article e230067. <https://doi.org/10.1530/ETJ-23-0067>.
15. Piticchio, T. *et al.* "Head-to-Head Comparison of American, European and Asian TI-RADSs in Thyroid Nodule Assessment: A Systematic Review and Meta-Analysis." *European Thyroid Journal*, vol. 13, no. 2, 2024, article e230242. <https://doi.org/10.1530/ETJ-23-0242>.
16. Cibas, E. S. and S. Z. Ali. "The 2017 Bethesda System for Reporting Thyroid Cytopathology." *Thyroid*, vol. 27, no. 11, 2017, pp. 1341–1346. <https://doi.org/10.1089/thy.2017.0500>.